Application of Machine Learning Algorithms for the Integrated Solid Waste Management in Guwahati city

Thesis submitted in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

by

Tinka Singh

Centre for the Environment

Indian Institute of Technology Guwahati

Guwahati–781039, India

Application of Machine Learning Algorithms for the Integrated Solid Waste Management in Guwahati city

Tinka Singh

Application of Machine Learning Algorithms for the Integrated Solid Waste Management in Guwahati city

Thesis submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

by

Tinka Singh
Roll No.: 176152101

Centre for the Environment

Indian Institute of Technology Guwahati

Guwahati – 781039, India

December 2023

Dedicated to my Parents, my relatives and to those scholars who are trying to make differences

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

Centre for the Environment

DECLARATION

I here with declare that the matter embodied in this thesis is the result of investigations carried out by me in the Centre for the Environment, Indian Institute of Technology Guwahati, India, under the guidance of Prof. Ramagopal V. S. Uppaluri, Centre for the Environment and Department of Chemical Engineering, Indian Institute of Technology Guwahati, India. Considering the compliances of scientific observations and analysis, wherever applicable, acknowledgements for the findings of other investigators have been made in the thesis document and relevant publications.

To the best of my knowledge, no portion of the thesis has been submitted elsewhere for a diploma or a degree.

18th Dec 2023 **(Tinka Singh)**

IIT Guwahati

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

Centre for the Environment

CERTIFICATE

This is to certify that Ms. Tinka Singh worked under my supervision as a regular registered PhD student. Her thesis entitled **"Application of Machine Learning Algorithms for the Integrated Solid Waste Management in Guwahati city"** is an authentic record of the obtained results and subsequent analysis of the research work carried out by the author under my supervision in the Centre for the Environment, Indian Institute of Technology Guwahati. I hereby certify that she has fulfilled all the requirements with respect to the investigations embodied in this thesis.

Prof. Ramagopal V. S. Uppaluri
Thesis Supervisor
Professor (Higher Academic Grade)
Centre for the Environment &
Department of Chemical Engineering
Indian Institute of Technology Guwahati
Assam-781039, India

Acknowledgements

My research odyssey would be incomplete without acknowledging the inspiration and support that kept me going even in adverse situations. Foremost, I convey my gratitude and love to my family for bringing me into the person I am now. My Maa and Paa made the endless sacrifices to enable me to reach my present state of higher academic realization and excellence. I owe everything to them. I also express my heartfelt gratitude to 'God-The Almighty' and my late grandparents. Their invisible hand helped me to tide over the hardships and complete my work.

I would like to express my gratitude to my thesis supervisor Prof. Ramagopal V.S. Uppaluri for his unwavering support, patience, constant guidance, tremendous encouragement, knowledgeful insights, continuous moral support and for fostering my research interest during all the ups and down in my thesis work. I am overwhelmed and indebted to my supervisor for bringing out the honest researcher in me and equipping me to face the life ahead. I will be forever grateful to him for he believed in me when no one else did. I could not have imagined having a better advisor and mentor for my PhD journey. I am deeply thankful to Mrs. Krishna Priya Ma'am (Prof. Uppaluri's wife) for her kind words, appreciation and support.

Besides my advisor, I would like to extend my sincere thanks to the honorable members of my doctoral committee, Prof. Animes Kr. Golder, (Department of Chemical Engineering, IIT Guwahati), Prof. Rajashree Bedamatta, (Department of Humanities and Social Sciences, IIT Guwahati) and Prof. Pranab Kr. Ghosh (Department of Civil Engineering, IIT Guwahati), for their motivation, valuable suggestions, and enlightening comments during my seminars. I express my warmest thanks to Prof. Animes Kr. Golder for his inspiring and uplifting suggestions since my SOAS days, to Prof. Rajashree Bedamatta for she nurtured me since my coursework days and always helped me to further strengthen my PhD work with her apt and intellectual support and to Prof. Pranab Kr. Ghosh for his thought-provoking queries and useful suggestions.

I am grateful to the former Head of the Centre (HOC, Centre for the Environment) Prof. Mihir Kr. Purkait and the current HOC, Prof. Utpal Bora for providing the necessary facilities. It is due to their kind cooperation that I successfully completed my PhD tenure. I owe my gratitude towards Mr. Partha Pratim Bakal, and Dr. Deepmoni Deka for their ceaseless, unbiased and unwavering help and suggestions during the entire tenure of my PhD thesis. I would like to express my sincere thanks to Mr. Kaustabh Rakshit, Mr. Mridul Das for their continuous support in the entire duration of my PhD thesis work.

I will take the opportunity to thank Vikas Kr. Sharma for his unconditional help, moral support, strength, guidance, suggestion and affection throughout my PhD journey. His faith on me have always pushed me to march beyond my limit and achieve the best in my academic pursuits.

The account of colleagues to thank for their endless and impartial support and motivation is long. Firstly, I render my deepest thanks to my dear friends Kamalini Acharjee for her unconditional support, love and motivations to keep me going, Saranya Mohan, Shweta Kumari, Junali, Archana and Darshana for their needful support and utmost care. My special and wholehearted thanks to Priyanka Panigrahi and Pratibha Maurya for their unconditional care, motivating words, affection, moral boosts that always allowed me to overcome my frustration and make me feel my stay at IIT Guwahati like that of my home. I express my gratitude to Tasrin Shehnaz, Khyati Raina, Roushni, Udangshree, Nilotpola, Emte and Tanmayee, for their positivity, encouragements and kind gestures. Thank you all for giving me the countless beautiful memories during the stay here. Your friendship is irreplaceable.

I am also thankful to Manash, Joy, Mir, Suchisloka, Maithili, Nilotpol, Neha, Juhi, Shweta K., Soumya, Rituparna, Soshmita, Srijon, Bhairab, Mir, Debajeet, Arunima, Sukanya, Sutapa baa to my seniors Rajneesh Sir, Jayakrishnan bhaiya, Somnath Da, Juna Probha baa, to my juniors Smitom, Himadree, Prangan, Devoleena, Sumona, Ida, Abhipsa, Krishn Murari, Amit and to my labmates. Besides, I would like to acknowledge my gratitude towards my relatives, my cousins, my nephews and nieces for their constant moral support and love.

Also, I am pleased to mention the pivotal role of my teachers Dr. Rabinder Prasad, Dr. Subhasish Banerjee, Dr. Swarnendu Kr. Chakraborty, Mr. Abhjeet Boruah, Dr. Manash P. Saikia, Dr. Deepak Gupta in enlightening me in the path of science I followed. Finally, I am truly grateful to the Centre for the Environment, IIT Guwahati, and MHRD, Government of India, for providing me with the opportunity and fellowship support during my PhD tenure. Still, the list is incomplete without the names of those who have directly or indirectly helped me and contributed to my success.

Tinka Singh
aitiinka@gmail.com

Abstract

Rapid urbanization and enhanced consumption patterns lead to substantial growth in the municipal solid waste (MSW) generation rates. Social and environmental challenges are ever increasing due to world-wide enhancement in the per capita waste generation. Since 1990s, developed nations prioritized waste reduction and recycling. While this has been the case, challenges in resources, infrastructure, and awareness hindered developing nations like India to achieve similar environmental benchmarks and standards. To bridge this gap, investments and advances in sustainable waste management are required along with international collaboration that promote eco-friendly practices. However, developing nations such as India could not march to realize such environmental compliances. Accordingly, technological interventions have fostered the integrated city-wide solid waste management in several important yet mandatory themes. In these challenges, precise prediction and forecasting of total solid waste generation rate, compost and biogas production rates, and rate of emissions of greenhouse gases and particulate matters from landfills and incinerators are prominent. Further, it is also important to emphasize research studies on the techno-economic analyses of composting and anaerobic biogas digester systems. With such efforts, environmental planning of solid waste managerial infrastructure can be established through effective planning strategies.

Rapid urbanization, increased consumption and per capita waste generation and disparities in environmental compliance are few prominent features that are responsible for environmental issues. Considering these aspects, the PhD thesis endeavors to comprehensively tackle municipal solid waste management (MSWM) in the Guwahati City. The research emphasis has been on waste characterization, generation rate prediction, GHG emissions, and compost and biogas production. It involves a survey-based quantification of waste composition among different categories and classification of waste data based on socio-demographic groups and seasons. Machine learning (ML) and time series modeling are to be targeted to model and forecast total MSW generation, organic fraction, and recyclables. Considering the relevance of the modelling efforts for the city of Guwahati, the PhD thesis also explores the effectiveness of alternate machine learning techniques for the prediction and forecasting of greenhouse gas (GHG) emissions and particulate matters (PMs), which are generated from landfill and incineration processes. Additionally, meteorological parameters and MSW organic waste (OW) data are to be utilized to predict and forecast compost and biogas production (BP) rates using artificial neural networks (ANN) and hybrid ML modeling approaches. The research also targets the development of techno-economic calculators for alternate BP designs and compost facility scales for the realization of optimal prediction and long-term forecasting of BP and compost rates. Further, a techno-economic modelling framework have been enabled as well to judge upon the feasibility of anaerobic bio-gas and composting plants for a city-wide waste recycling strategy.

The PhD thesis fulfilled eight major objectives namely (a) survey-based quantification and seasonal classification of waste among socio-demographic groups in the Guwahati City; (b) modelling waste generation rate using ensemble tree-based ML algorithms and forecasting with moving average methods in the Guwahati City; (c) efficient prediction of organic and recyclable waste generation with supervised ML algorithms and autoregressive integrated moving average models (ARIMA) in Guwahati city; (d) ML-based prediction and forecasting of GHG emissions and particulate matters from MSW landfill and incineration sites in Guwahati City; (e) prediction and forecasting of composting rates using meteorological parameters and OW data with ANN and hybrid ML for Guwahati City; (f) prediction and forecasting of anaerobic digestion rates using meteorological parameters and OW data with ML and hybrid ML models for Guwahati City; (g) development of a techno-economic calculator for biogas production based on alternate designs and optimal forecasting of biogas generation rate in Guwahati City and (h) development of a techno-economic calculator for composting facilities based on alternate scales and optimal forecasting of compost production in Guwahati City. A statistical tool namely SPSS software has been used to achieve relevant statistical results. Also, Python version 3.8 with NumPy, Scikit-learn, Pandas, and Matplotlib libraries for data analysis, modelling, and automated data processing on an i7-4790 CPU @ 3.60 GHz processor has been used.

The historical data such as socio-economic and demographic data were collected from open government and public data such as Census of India data, Ministry of Statistics and Planning Implementation (MOSPI) data, meteorological data from Modern-Era Retrospective analysis for Research and Applications MERRA-2 dataset and Iowa Environmental Mesonet (IEM), waste generation data from Central Pollution Control Board (CPCB) and GHG emissions data from Emissions Database for Global Atmospheric Research (EDGAR) portal.

The overall modelling framework implementation involved various phases such as data gathering, pre-processing, modelling, and analysis. Several data pre-processing stages were required for the preparation and translation of raw secondary data into variables suitable for modelling and analysis. These include secondary data extraction, data loading into appropriate data structures, secondary data transformation, outlier removal using filtering processes, and integration into consolidated datasets.

The first objective was to conduct a survey-based quantification and seasonal classification of waste among socio-demographic groups in Guwahati City. The investigation revealed that the total solid waste in Guwahati comprised 51.62 % OW, 24.22 % packaging waste (papers and plastics), 8.13 % combustible waste (clothes, diapers, rugs, bags, shoes), 4.62 % park and garden waste, 3.54 % metals and glass waste, and 7.87 % miscellaneous waste. Notably, higher income levels were positively correlated with increased solid waste generation. Per capita generation rates were obtained as 0.17 kg capita^{-1} day^{-1} for A3 (poor income), 0.18 kg capita^{-1} day^{-1} for A1 (middle income), 0.23 kg capita^{-1} day^{-1} for A2 (high income), and 0.516 kg capita^{-1} day^{-1} for the commercial

area sub-groups. The practical implications of these findings are opportunities for custom designing of waste management strategies based on socio-demographic characteristics, resource allocation facilitation, effective guidance for the environmental impact assessments of solid waste supply chain system, strengthening of waste management policies, and promotion of community engagement for the effective and inclusive waste management practices in the Guwahati city.

The targeted second objective was to model waste generation rates in Guwahati City using ensemble tree-based ML algorithms and forecast with moving average methods. The second objective findings demonstrated that among the considered tree-based ML models, Gradient Boosting (GB) exhibited the highest model accuracy, achieving 97% for predicting the total solid waste in Guwahati. The GB model outperformed other models, displaying superior performance metrics such as R = 0.94, R^2 = 0.99, RMSE = 3.01, MAE = 2.86, and IoA = 0.94. In comparison to Random Forest (RF) and Decision Tree (DT) models, the GB model showcased enhanced accuracy (R = 0.90, R^2 = 0.98, RMSE = 83.21, MAE = 74.84, IoA = 0.72) and (R = 0.82, R^2 = 0.97, RMSE = 325.82, MAE = 302.20, IoA = 0.45), respectively. The findings include the optimization of waste management strategies through accurate prediction models, aiding municipal authorities in resource allocation and planning for effective waste disposal and recycling initiatives in the Guwahati City.

The third objective aimed to enhance the efficiency of predicting organic and recyclable waste generation in Guwahati city with the supervised machine learning (ML) algorithms and autoregressive integrated moving average models (ARIMA). The findings revealed that the artificial neural network (ANN) approach demonstrated superior performance with an impressive 92% overall accuracy rate. However, in due course of the specific assessment of the plastics and paper diversion cases, the ANN model exhibited a notable error rate of 18.21% in the sample data. To address this limitation, the multilayer perceptron (MLP) model emerged as a viable alternative, showcasing commendable performance metrics with an R^2 value of 0.92 and an RMSE of 10.43 for OW prediction, and similar confidence assuring findings (R^2 = 0.92, RMSE = 18.21) for the prediction of recyclables (plastics and papers) waste generation rate. These outcomes provide valuable insights for waste management practitioners, and infer that while ANN is effective for overall waste prediction, the MLP could be even more suitable for specific waste streams. Such findings on plastics and papers waste generation rate contribute to the betterment of accurate and tailored waste management strategies in urban areas such as those of the Guwahati city.

The fourth objective of the thesis focused on the effective utility of ML for the prediction and forecasting of GHG emissions and particulate matters from MSW landfill and incineration sites in Guwahati City. The findings revealed that, employing stepwise regression with a cubic configuration, the model achieved high accuracy with R^2 and RMSE values of 0.94 and 23.2 for CH_4, 0.96 and 9.5 for CO_2, and 0.85 and 45.6 for PMs fluxes, respectively. Notably, the multilayer

perceptron (MLP) algorithm outperformed all other ML algorithms, and demonstrated exceptional accuracy with R^2 = 0.97, RMSE = 0.32, IoA = 0.94, MAE = 28.32, MAPE = 49.21, and SSE = 0.21. For forecasting purposes, the ARIMA model with a configuration of (1,1,0) × (0,1,1) proved to be the most effective, affirming promising R^2 and RMSE values of 0.89 and 5.67, respectively. These results have significant practical implications for waste management and environmental planning in Guwahati and can henceforth provide a robust framework for the prediction and mitigation of the environmental impact of MSW landfill and incineration activities.

The fifth objective aimed at the prediction and forecasting of composting rates in Guwahati city. For this purpose, meteorological parameters and OW data were modelled with the ANN and hybrid ML models. The findings revealed that, among the considered ML models, GB exhibited exceptional performance with R^2 and RMSE values of 0.99 and 0.757, respectively. Subsequently, the MLP model demonstrated good performance with R^2 of 0.938 and RMSE of 13.37. Through the implementation of a randomization technique, the MLP model outperformed other models with improved RMSE (9.32), R2 (0.95), and IoA (0.87) and thereby achieved an overall model accuracy of 90%. The study also identified ideal conditions for compost generation during the summer season (April to June) with a temperature (T_{max}) of 30 ± 2°C and relative humidity (RH) above 80%. These results can be applied for waste management and urban planning, and thereby offer a reliable framework for the prediction of composting rates based on meteorological parameters and OW data. Thereby, the findings facilitate the betterment of the sustainable waste management practices in the Guwahati City.

The thesis's sixth objective of the study focused on the prediction and forecasting of anaerobic digestion rates in Guwahati City. For this purpose, meteorological parameters and OW data were modelled through ML and hybrid ML models. The research aimed to assess upon the sensitivity of five parameters (T, P, RH, Pr, and OW) for the accurate prediction of BP rates. The data analysis studies conveyed that T and OW have been the most sensitive variables, and assured coefficients of 0.81 and 0.65, respectively. Thereby, the parameters sensitively influenced the BP rate variations. The MLP model demonstrated superior performance in comparison to the GB algorithm and henceforth ascertained 84% accuracy in the prediction of BP rates. These findings contribute scientifically to understand the key parameters affecting anaerobic digestion rates. Thereby, practical applications can seek to optimize waste-to-energy processes. The economic evaluation case study can further enhance confidence levels for the assurance of sustainable waste management practices in the urban areas in the cities of a developing nation such as India.

The seventh objective targeted the development of a techno-economic calculator for BP rates. To do so, alternate designs were considered for the optimal prediction and forecasting of biogas generation rates in Guwahati City. The findings involved the determination of AD calculator that facilitates the estimation of BP rates for both short-term and long-term AD operations. The model validation efforts assured effective calculator performance for the prediction of BP rates within a

±10% range with those presented in the literature. Notably, the Deenbandhu digester emerged as the most effective among alternate designs, and thereby produced highest BP rate of 60 m^3 day^{-1}. An increase in the selling price of compost from $0.50 to $0.75 has a significant positive impact on net profit. At 0.25 $ m^{-3} biogas selling price, a negative net profit of Rs. -254,398 ($ -3,051 USD) was obtained, which eventually increased to Rs. 2,62,920 ($3,154 USD) for 0.50 $ m^{-3} biogas selling price and to Rs. 7,19,445 ($8,629 USD) at 0.75 $ m^{-3} biogas selling price. Thus, digester plant upgradation ascertained improved profitability and biogas price alteration assured a significant improvement in the overall profitability. Thus, price subsidy appears to be the most relevant strategy to ensure upon the deployment of such sustainable waste processing strategies. The practical applications of the conducted research convey that the tool can serve robustly for biogas production optimization through the appropriate selection of alternate AD designs. Such efforts can significantly contribute to sustainable waste-to-energy initiatives in urban environments such as those in the Guwahati city.

The eighth objective of the thesis targeted the development of a techno-economic calculator for composting facilities in Guwahati City. To do so, alternate scales and optimal forecasting of compost production can be targeted. The final findings of the PhD thesis presented technical and economic performance indices for three alternate compost facilities with daily processing capacities of 25 tons for high scale, medium scale, and yard waste units). The projected annual revenue from compost sales reached $900,000. Notably, the High Scale Compost Facility (HSCF) stood out with a cash flow $0.45 million, and a payback period of 6 years. Thereby, it ensured financial viability of the system. The HSCF compost system demonstrated significant profitability potential, for the case of a market value of the compost ranging from $10 to $25 ton^{-1}. These results provide a valuable tool for decision-makers in waste management, and thereby offered insights into the economic feasibility and optimal scale of composting facilities. All these contribute to sustainable waste management practices and economic viability in the Guwahati City.

In summary, the PhD thesis affirmed that ML techniques, time series and hybrid ML models can accurately predict and forecast MSW generation patterns. Thus, they can assist in the efficient waste management in the Guwahati city. These models also demonstrated their potential to optimize composting and biogas processes under different climatic conditions. The integration of an AD calculator, ML techniques, and GIS data enhanced the pragmatic characteristics of the modelling efforts for relevant and optimal biogas and compost facilities. Such modelling framework can assure improved operational efficiency, increased energy generation, and cost-effective operations. Considering the thesis findings, it can be concluded that the ISWM of Guwahati city can be best realized through the application of machine learning algorithms for solid waste rate prediction and subsequent planning of OW recycling strategies. Such efforts can enable the policy makers of the city for put forth a feasible and practical sustainable waste management plan in the forthcoming decade that is about to witness rapid economic growth, urbanization and associated complexities in the solid waste management challenges on a city-wide basis.

Novelty Statement

The novelty of this thesis lies not only in its methodology but also in its focus on addressing critical gaps in existing literature. It critically examines prevalent assumptions and challenges established theories, aiming to unveil blind spots and highlight areas that require further exploration. By tackling these gaps head-on, the research seeks to generate valuable knowledge that can be applied to real-world problems on solid waste industry and contribute to the advancement of ISWM as a whole. This thesis is driven by a passion for pushing the boundaries of knowledge and making meaningful contributions that has shaped the future trajectory of MSWM research.

- Existing studies have primarily focused on solid waste characterization and quantification, neglecting the potential influence and implications of sampling strategies. This thesis aims to fill this crucial research gap by implementing the stratified random sampling based seasonal influence, offering fresh insights into the broader understanding of solid waste characteristics.
- Previous research has predominantly examined varied socio-economic factors but within a narrow context, failing to consider the influential factors that shape its dynamics. Thereby, this thesis bridges this research gap by incorporating socio-economic parameters such as gross district domestic product (GDDP), working population (WP) and literate population (LP) integrated case studies for the ML based prediction and forecasting of total MSWG.
- The PhD thesis addresses the K-cross validation-based refinement in the ML based modelling framework for the prediction and forecasting of total, organic and recyclables waste generation rate to overcome overfitting issues and enhance the accuracy of ML based prediction models.
- This thesis quantifiably contributes by rigorously modelling and comparing solid waste disposal methods, offering valuable insights into environmental implications and pioneering advanced prediction models for simultaneous CH_4, CO_2, N_2O, SO_2 fluxes, and PMs emissions from landfill and incineration sites, addressing a critical research gap and enhancing understanding of sustainable waste management.
- This thesis introduces a novel approach by leveraging meteorological parameters integrated modelling for the ML based prediction and forecasting of biogas and compost generation rate. Subjective aspects of the novelty refer to the utilization of historical datasets for the sensitive

influence analysis of the seasonal aspects on the complex seasonal patterns of the biogas and compost generation rates.

- This thesis introduces a novel dimension by developing AD calculator for Indian biogas and compost plants and techno-economic analysis integrating ML algorithms for performance optimality.

CONTENTS

List of Tables

Lists of Figures

Nomenclature

ABBREVIATIONS

AD	Anaerobic Digestion
AHP	Analytic Hierarchy Process
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
ANP	Analytic Network Process
ARIMA	Auto Regressive Integrated Moving Average
BP	Biogas Production
CAPEX	Capital Expenditure
CNN	Convolutional Neural Network
CP	Compost Production
CPCB	Central Pollution Control Board
DES	Directorate of Economics and Statistics
DT	Decision Tree
EDGAR	Emissions Database for Global Atmospheric Research
EDU	Education level
ELM	Extreme Learning Machine
EMA	Exponential Moving Average
EPA	Environmental Protection Agency
FELs	Front-end loaders
GA	Genetic Algorithm
GB	Gradient Boosting
GBRT	Gradient Boosted Regression Tree
GDDP	Gross District Domestic Product
GDP	Gross Domestic Product
GHG	Green House Gas
GI	Germination Index
GIS	Global Information System
GMC	Guwahati Municipal Corporation
HDI	Human Development Index
HH	Household
HH inc	Household Income
HH size	Household size
HHW	Household waste
HPO	Hyper Parameter Optimization
HSCF	High scale composting facility
IEM	Iowa Environmental Mesonet

IoA	Index of Agreement
IQR	Interquartile Range
IRENA	International Renewable Energy Agency
IRR	Internal Rate of Return
***k*NN**	k-Nearest Neighbour
L1	Least Absolute Shrinkage and Selection Operator (LASSO)
L2	RIDGE
LCI	Life Cycle Inventory
LP	Literate Population
LR	Linear regression
LR	Linear Regression
LULC	Land Use and Land Cover
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC	Moisture Content
MCDA	Multi-Criteria Decision Analysis
MEDA	Maharashtra Energy Development Agency
MERRA-2	Modern-Era Retrospective analysis for Research and Applications
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear regression
MNRE	Ministry of Natural and Renewable Energy
MNRE	Ministry of New and Renewable Energy
MoP	Ministry of Power
MOHUA	Ministry of Housing and Urban Affairs
MOSPI	Ministry of Statistics and Programme Implementation
MSA	Measure of the Sample Adequacy
MSCF	Medium scale composting facility
MSE	Mean Square Error
MSW	Municipal Solid Waste
MSWG	Municipal Solid Waste Generation
MSWM	Municipal Solid Waste Management
NCA	Neighbourhood Component Analysis
NPV	Net Present Value
OGD	Open Government Data
OPEX	Operating Expenditure
OW	Organic waste
P	Pressure
PBP	Payback Period
PCA	Principal Component Analysis
PM	Particulate Matters
PO	Parameter Optimization
Pr	Precipitation
RA	Regression Analysis
RDA	Redundancy Analysis
RF	Random Forest

RH	Relative Humidity
RMSE	Root Mean Square Error
RT	Retention Time
RVS	Reduced Volatile solid
sARIMA	seasonal-Auto Regressive Integrated Moving Average
SC	Significance Coefficient
SMA	Simple Moving Average
SP	Selling Price
SSE	Sum of Squared Error
SVM	Support Vector Machine
SVR	Support Vector Regression
SWG	Solid Waste Generation
SWM	Solid waste management
T	Temperature
TEA	Techno-economic assessment
TEA	Techno-Economic Assessment
TS	Total solid
TSD	Total solid Discharge
WD	Wind Direction
WLC	Weighted Linear Combination
WMA	Weighted Moving Average
WNN	Wavelet Neural Network
WP	Working Population
WS	Wind Speed
WTC	Waste-to-Compost
WtE	Waste-to-Energy
YWCF	Yard waste composting facility

NOTATIONS

POP	Population
HH_{size}	Household sizes
$GDDP_{per\ capita}$	Gross District Domestic Product per capita
$Q_1\ and\ Q_2$	First and the third Quartile
t	tons
ε	error
C/N	Carbon Nitrogen Ratio
$RMSE$	Root Mean Square Error
r	Correlation coefficient
$k\ (x_{new})$	k nearest neighbours of (x_{new})
$V_{BG,i}$	Estimated biogas production rate associated with feed material i
$m_{TS,i}$	mass flow rate of TS contained in feed material i
V_{gs}	Volume of gas storage chamber
Q	influent

$C_{CH4,i}$	Biogas content associated with feed material i (%)
$V_{CH4,i}$	Methane production rate of feed material i (m^3 - CH_4/day)
C	Penalty parameter
V_{CH4}	Total methane production rate (m^3 CH_4/day)
Y	Yield factor
S	Initial concentration of volatile solids (VS)
R^2	Coefficient of determination
T (max)	Maximum Temperature
T_{atm}	Atmospheric pressure
T	Temperature
t d^{-1} or tpd	tons per day
hp	House power
kg d^{-1}	Kilogram per day
D	Diameter of the digester
$tanh$	activation function
kWh^{-1}	Kilo Watt hour
H	Height of the digester
C_{RE}	covariance matrix
λ_k	eigenvalues of the corresponding RDA's k^{th} axis
R	Response variable
E	Exploratory variable
C_{EE}^{-1}	inverse variance of variance-covariance matrix
u_k	Canonical analysis produces ordinations of *y* that are constrained linearly to a set of variables E
N or n	number of observations
$\bar{y}_{D_1} and\ \bar{y}_{D_2}$	mean predictions of the training set for D_1 and D_2 respectively
$\widehat{Y}_i$	model predicted value
Y_i	actual (data) value,
Y	mean value of waste generation rate
X_i	true value
W	weighting value
A	Average in period
S_t	the smoothed value
s	$smoothing\ factor$
EMA_C	EMA for the current time period
C_T	Current time-period
Z_i	output of the MLP model
b_i	bias term for the i^{th} neuron
w_{ij}	weight between the i^{th} and j^{th} neurons
p_i	input to the i^{th} neuron
$W(x)$	current generation of weights and biases
λ	learning rate
$\Delta W(x)$	change in weights and biases based on the fitness function

a	weight vector in SVR
b	input variable in SVR
α_i	Lagrange multiplier for the ith training data point
$K(x_i, x)$	kernel function
c	bias term in SVR
α_i	Lagrange multiplier for the ith training data point
$K(x_i, x)$	kernel function
c	bias term in SVR
y_t	goal to be forecasted
α	constant
$\beta_1 y_{t-1}$ and $\theta_1 \varepsilon_{t-1}$	linear combination of lags Y ranging from lags 0 to a and 0 to c respectively
V_d	Digester volume (in m^3)
L	length of the digester
Y_p	potential biogas yield
V_f	total feedstock volume (m^3 day^{-1})
CF_n	Cash flow for n year
R	Revenue
i	waste component
$(dry\ mass_1)_i$	dry mass waste component i entering the facility on the tipping floor
$moist_i$	MC of waste component i at tipping floor and measured in weight %
$(mass_1)_i$	wet mass of component i entering the facility
vol_1	total volume flow rate of wastes entering the tipping floor
$mass_3$	total waste wet mass flow rate entering the composting pad after shredding
$dry\ mass_1$	initial dry mass flow rate of incoming waste streams at tipping floor
$dry\ mass_3$	total dry mass flow rate entering the composting pad after shredding
$mass_4$	wet mass flow rate of MSW at the end of the curing pad
$compost_{moist}$	moisture of MSW at the end of curing
tph	tons per hour
cfm	cubic feet per minute

Chapter 1
Introduction and Literature Review

Chapter 1

Introduction and Literature Review

In this chapter, section 1.1 presents an overview of few important aspects related to solid waste generation. The section highlights the growing concern of non-holistic waste disposal (particularly in Indian cities like Guwahati) and associated challenges for the effective management of solid waste such as financial constraints, lower awareness levels, and unique waste characteristics in the region. Further, section 1.2 provides statistical data of global and Indian solid waste generated quantities and thereby highlights on the sensitive influence of population growth, consumption patterns, and urbanization in waste quantification estimates and future trends. Section 1.3 devotes to the affirming of a comprehensive and integrated approach for solid waste management (SWM), by considering the limitations of the conventional reductionist approach. Thereby, the need for the adoption of Integrated Solid Waste Management (ISWM) paradigm has been inferred. Following this, section 1.4 addresses the available prior art in the specific field of technical research devoted to large scale (city wide) SWM and associated mitigation strategies in developing countries. Thereafter, section 1.5 elucidates upon the scope for further research in few topics that ensure effective large scale SWM systems. Finally, section 1.6 of the chapter presents the objectives of the Ph.D. thesis followed with thesis organization related details in section 1.7.

1.1 An Overview of Solid Waste

The ever-increasing threat to the environment and humanity due to non-holistic, poor and non-hygienic waste disposal needs a comprehensive and pragmatic framework to resolve all major issues. Both urban sprawls resulting from population growth and changing consumer habits have propelled a substantial rise in the per capita waste generation. Consequently, municipal corporations face significant challenges in the management of solid waste. Thereby, similar issues exist in all major cities and towns across India, including Guwahati. Thus, considering the involved diverse socio-economic and demographic factors, comprehensive solutions are to be sought to address complex waste disposal and management problems.

While pragmatic SWM systems exist in developed countries, their adaptation to resolve similar issues in Indian scenario and for local conditions is daunting. Such a challenge is further amplified due to the current level of poor awareness among citizens in most Indian cities. Implementing sophisticated waste management methodologies of the developed countries in a developing country framework has several

constraints. These includes financial constraints, different user demographics, lower general education and awareness levels regarding waste management in communities, higher costs associated with the operation of advanced waste plants, and the nature and quantity of waste that needs to be managed. Additionally, limited financial resources and competing priorities further complicate waste management efforts in developing countries like India.

Waste management tends to have a lower priority in developing countries. This is due to their greater emphasis on issues such as hunger, health, water, employment, and civil unrest (Ferronato and Torretta, 2019). Municipalities in India often face financial difficulties, which hinder their ability to provide the desired level of public services. The poor financial status of municipal corporations has been attributed to the challenges faced in urban centers (Guerrero et al., 2013; Cohen, 2006; Sharholy et al., 2008; MoEF, 2015). A significant portion (around 75-85%) of the budget for municipal solid waste management (MSWM) is allocated to waste collection and transportation processes (Kumar et al., 2017). In summary, the current situation highlights the urgent need to comprehensively address waste disposal and management challenges. Tailor-made solutions must be developed by considering specific local issues related to garbage disposal and treatment.

1.2 Solid Waste Statistics

In 2020, the global production of municipal solid waste (MSW) was estimated to be 2.24 billion tons, and it is projected to increase to 3.88 billion tons by 2050 (World Bank, 2019). According to a World Bank report, the amount of solid waste in urban areas of East Asia alone will rise from 760,000 tons day^{-1} to 1.8 million tons day^{-1} in a time span of 25 years. Additionally, waste management costs are expected to nearly double from US$ 25 billion to US$ 47 billion by 2025. In India, the annual generation of urban MSW is approximately 40 million tons, and it is expected to increase at an annual rate of 1.33% (EIA, 2013). The quantity of waste produced is sensitively influenced with factors such as population size, living standards, income levels, economic growth, consumption patterns, and institutional frameworks (Nguyen et al., 2020; Kumar et al., 2017; Sharma et al., 2018). In 1991, only 25.78% of India's population was urban and this was accounted for 297 million people. By 2031, the urban population is projected to exceed 48% with over 620 million people (UNDP Report, 2008). The

daily per capita generation of MSW in India ranges from about 100 grams in small towns to 500 grams in large towns (Unnikrishnan and Singh, 2010). It is estimated that the per capita MSW generation will increase annually at a rate of 1-1.33% (Tripathi et al., 2022; Singhal et al., 2022). Several studies have conveyed that the MSW generation (MSWG) rates are lower in small towns in comparison to metro cities, and the per capita generation rate in India ranges from 0.2 to 0.5 kg day^{-1} (Kumar et al., 2017; Sharholy et al., 2008; Jha et al., 2011; Central Pollution Control Board, CPCB, 2014). Among 50 metro cities in India, approximately 45,000 tons of waste are generated daily. However, Class I cities generate about 72,000 tons day^{-1} (MOHUA, 2016).

In India, the solid waste characteristics data convey that the organic fraction of the waste varies widely (40-85% of the waste) (National Solid Waste Association of India, 2017). This is due to its sensitive dependence on income and lifestyle of the population. The production and consumption of plastic increased over 180 times between 1950 and 2018 (Rafey and Siddiqui, 2021). Per capita generation of waste in Guwahati altered from 400 to 500 grams day^{-1} (GMC, 2019) and was dependent mostly on the income level group and seasonal alterations. The primary constituents of MSW comprised organic materials (42.2%) and plastic waste (25.2%) (Singhal et al., 2022).

For Guwahati city, Fig. 1(a) depicts the upward trajectory of total MSWG in tons and the corresponding population in millions during 1980-2010 period. The figure highlighted the parallel increase in waste generation with population growth. Also, it was estimated that out of the total plastic generated from household (HH), 60% is recyclable plastic and 1500 rag pickers engage themselves to operate only one MSW dumping site at West Boragaon in the city (Kashyap et al., 2010). Fig. 1(b) illustrates the MSWM in Guwahati from 1990 to 2020, and showcases the employed waste management practices such as landfilling, recycling and incineration.

Guwahati city faces a grave problem in terms of the MSWM. Few researchers in the past did cite MSWM problems in Guwahati (Goswami and Sarma, 2008; Pradhan et al., 2012). The city has only one landfill site in West Boragaon. A study highlighted that no proper disposal method has been existent in the West Boragaon dumpsite (Singhal et al., 2022). Due to disposal without any processing, the local people complained with respect to health ailments. Such non-holistic and non-scientific

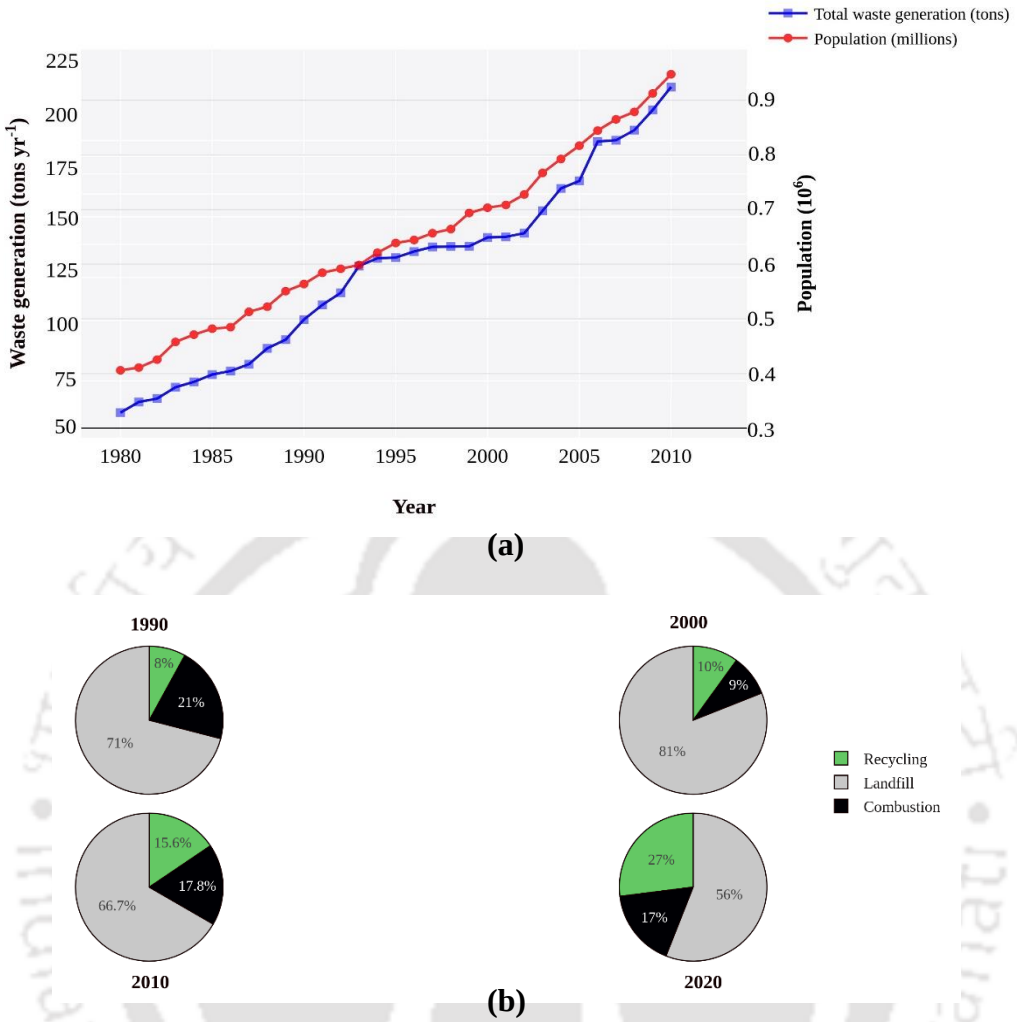


Fig. 1 (a) Total MSWG rate and Population growth trends for Guwahati city in the year range of 1980–2010 **(b)** Solid Waste distribution characteristics for Guwahati city (1980–2010). Source: World Bank data

disposal further enhances air, water and soil pollution. At Deepor Beel, the open dumping of the total waste of Guwahati lead to detrimental effect on the aquatic habitat. Accordingly, due to relevant threat to the water body, the Guwahati Municipal Corporation (GMC) has been unable to execute a scientific solid waste disposal system due to insufficient waste disposal facilities and capacities.

As per the guidelines of MSW Act 2016, the current practice of waste collection and disposal in Guwahati city is far away from a system designed with a scientific basis. Scarcity of land for a waste management facility to handle such a huge quantity of waste for the city's size is a major bottleneck to resolve and achieving success in this direction. Heaps of waste lying on the roadside is a common occurrence in Guwahati. Few residents litter their immediate surroundings and cause difficulties for the

resident community. While few researchers addressed certain issues for the resolution of the MSWM (Chen et al., 2010; Minghua et al., 2009), but not much research work was conducted to find out precise solutions for the scientific disposal and processing of the waste generated from Guwahati city. Not much research has been evident to opine upon alternate feasible options for the waste management of Guwahati city by considering the contemporary sensitive and detrimental issues. Hence, the design and development of a waste management model which could cater to environmental sustainability and economic viability including the processing and disposal of solid waste in Guwahati city is the need of the hour as it has profound impact on the competent strategies of the MSWM. The development of a sustainable working comprehensive model that definitely addresses resolution of key issues of MSWM in the city will be extensively focused in the thesis.

1.3 Need for a Systemic Approach

Effective waste management necessitates upon the implementation of suitable technical solutions, adequate organizational capacity, and collaboration among diverse stakeholders. Seadon (2010) underscores the intricate nature of MSWM, and inferred upon the consideration of interdisciplinary and multi-sectoral frameworks. These considerations encompass various aspects such as manufacturing, transportation, urban development, land use patterns, public health, etc. The author emphasizes the interaction and complexity between the physical components of the waste management system and the conceptual components that encompass social and environmental spheres. However, the conventional approach for effective MSWM is based on a reductionistic framework and henceforth could not address and resolve inherent complexities. It involves dividing interacting systems and their elements into increasingly smaller parts. System processes, such as waste generation, collection, and disposal operations, are often treated to be independent to one another despite their interlinkages and mutual influence (Seadon, 2010). Similarly, waste itself is not regarded as a single entity due to restrictive resolution of fragmented parts with numerous primary and secondary classifications. Waste streams from different sectors, such as residential and commercial, are often treated separately (Seadon, 2010). Consequently, techniques commonly focus upon addressing the one type of waste at a time. This

translated into the emphasis on individual technologies but not the comprehensive waste management systems (Dijkema et al., 2000).

The current paradigm of SWM, known as ISWM, gained widespread acceptance in developed countries due to policy shift from landfilling as a disposal option. Thus, the growing demand for a more comprehensive approach in the 1990s propelled the ISWM. In contrast to the engineering-focused modern SWM practices of the 1970s that mainly dealt with technical problems using technical solutions (van de Klundert and Anschutz, 2001), the concept of ISWM aims to strike a balance between three key dimensions of waste management. These refer to environmental effectiveness, social acceptability, and economic affordability. Henceforth, implementing ISWM in developing nations is essential for several reasons. ISWM promotes environmental sustainability through the minimization of the environmental impact of waste and conservation of resources. It improves public health through appropriate waste handling and disposal mechanism and subsequent reduction of disease transmission risks. ISWM also creates economic opportunities through resource recovery and recycling, support for local industries and reduction of the reliance on virgin materials. It fosters social equity by creating jobs in waste management sectors and empowering marginalized communities. Additionally, ISWM contributes to climate change mitigation by reducing greenhouse gas (GHG) emissions. With their strict compliance to international standards, the developing nations enhance their reputation to eventually attract further investments. Moreover, ISWM can result in long-term cost savings through the minimization of environmental damage and omittance of expensive clean-up operations.

ISWM also emphasizes upon the integration of various interconnected processes and entities within a waste management system (McDougall et al., 2001). Through the integration of waste materials, sources, and treatment methods, ISWM aims to reduce environmental impacts and lower costs. It advocates for a market-oriented approach in which energy and materials derived from waste have viable end uses. Additionally, ISWM systems can be subjected to continuous improvement (McDougall et al., 2001). Also, considering community goals, the ISWM aims to fulfill targeted needs of the stakeholders. Further, considering the local context, it integrates technical aspects (such as waste characteristics) with the cultural, social, environmental, and economic factors. The optimal combination of prevention,

reduction, recovery, and disposal methods is determined based on the available and appropriate options, which are customized to meet specific community requirements (Kollikkathara et al., 2009; McDougall et al., 2001; van de Klundert and Anschutz, 2001). In a nutshell, ISWM encompasses all major aspects including socio-economic, technical, environmental, and techno-economic frameworks.

1.3.1 Socio-economic Perspectives

Various studies explored the influence of socio-economic and demographic factors on waste generation. Factors such as income level, HH size, GDP, education (EDU) level, and environmental attitudes have been commonly identified to critically influence waste generation rate (Buenrostro et al., 2001; Banar and Ozkan, 2008; Afon, 2007). However, it is important to note that these effects can differ across locations, including countries, cities, and even within different zones of a city. Thus, the prediction of the level of association between these factors on a larger scale is highly challenging. Thus, what could be a significant predictor in one context may not be true in another context (Banar and Ozkan, 2008). SWM is as well influenced by social pressure within a community. Thereby, it highlights the need to cater to specific HH requirements (Barr et al., 2003; Lansana, 1993). Thus, a detailed understanding into the diverse needs across different housing types and areas is crucial to assure upon the effectiveness of waste management systems and identification of variations in generated waste characters among socio-demographic groups. MSWG varies across locations due to varied consumption patterns due to environmental, demographic, and socio-economic factors (Singh et al., 2011; Sjöström and Östblom, 2010). Thus, to develop effective MSWM systems for reduced waste generation, the mentioned factors must be examined and considered (Wang et al., 2011).

1.3.2 Technical Perspectives

Technical aspects involve the application of appropriate technologies and engineering solutions for generated solid waste, transportation, treatment and disposal. Among these, few relevant themes refer to the design and operation of waste management strategies by adopting modelling concepts such as machine learning (ML) to competently ensure the efficiency and effectiveness of waste management processes. For this purpose, various models have been developed to predict waste generation at city or national scales. In few prior arts, authors deployed predictive models that incorporated temporal

features such as seasonality and historical trends. For example, Rimaityte et al. (2012) inferred that the combination of auto regressive and integrated moving average (ARIMA) model with seasonal exponential smoothing model yielded effective predictions of weekly waste generation. ML methods, including neural networks, have also been employed to predict time dependent waste generation rates. Zade and Noori (2008) adopted a feed-forward artificial neural network (ANN) to predict weekly waste generation in Mashhad, Iran, and achieved promising results. Antanasijevic et al. (2013) deployed ANNs to predict MSWG rate at the national level in Bulgaria and Serbia. Few studies integrated spatial and temporal models to enhance prediction accuracy. Dyson and Chang (2005) developed a system dynamics model that integrated population, median HH income, HH size, and historical waste generation data for waste generation rate prediction in San Antonio, Texas. Johnson et al. (2017) developed a spatiotemporal model with gradient boosted regression tree (GBRT) algorithm. The authors considered weather, urban morphology, and socio-economic and demographic information to predict weekly MSWG in different administrative sections of New York City. These studies demonstrate the potential of predictive models for the accurate estimation of waste generation rates based on historical data and relevant factors.

1.3.3 Environmental Perspectives

Environmental aspects focus upon the minimization of the environmental impact of the generated solid waste. These include reduction of GHG emissions, mitigation of air and water pollution, prevention of soil contamination, and promotion of resource conservation and sustainable utilization schemes. In atmospheric science, ANN models, particularly multilayer perceptron (MLP), gained significant prominence for the prediction and forecasting of atmospheric GHG concentrations. Studies conducted by Gardner and Dorling, (2000), Chelani et al. (2002), Jiang et al. (2004), Karacan (2008), and Radojevic´ et al. (2019) did widely adopt ANN approaches. Hence, the authors inferred their superior performance in comparison with conventional statistical models for the prediction of complex air pollution parameters. MLP neural network models were extensively considered in various GHG studies. Ozkaya et al. (2007) deployed a backpropagation MLP model to predict methane (CH_4) fraction in a bioreactor landfill in Turkey. Li et al. (2011) customized a backpropagation neural network model to forecast temperature (T), CH_4, carbon dioxide (CO_2), and oxygen concentrations in a large landfill

system in southern California. Considering parameters such as average air temperature (T_{atm}), Abushammala et al. (2014) proposed a feedforward backpropagation MLP neural network to predict methane oxidation fraction in the landfill cover soil in Malaysia.

Also, Global Information System (GIS) based approaches have been inferred to be useful tools for the determination of suitable locations for solid waste facilities. This is due to their ability to integrate spatial data and consider key factors such as population density, land use, transportation, and environmental sensitivity. Accordingly, informed decision-making and sustainable waste management practices can be achieved (Sener et al., 2006; Sumathi et al., 2008; Wang et al., 2009; Sultana and Kumar, 2012). Such methodologies enabled optimal site selection, minimal negative impacts on the environment and communities, and supported long-term planning and monitoring of solid waste facilities. GIS custom serves the optimal site selection protocol by considering factors such as proximity to waste sources, transportation routes, accessibility, and environmental sensitivity (Moeinaddini et al., 2010). It also enables the assessment of potential environmental impacts and the identification of sensitive areas that shall be avoided or mitigated. Moreover, GIS-based approaches facilitate stakeholder engagement through the visualization and communication of spatial information. Thus, they lead to enhanced inclusive decision-making processes. The framework also supports long-term planning and management of solid waste facilities by providing tools for monitoring, analysis, and emergency response planning (Chiueh et al., 2008; Chang et al., 2009). In conclusion, GIS-based approaches provide a systematic and data-driven methodology for the selection of suitable locations. They promote sustainable waste management practices, and minimize negative influence on the environment and communities.

1.3.4 Techno-economic Perspectives

Techno-economic considerations involve the assessment of the financial viability and economic feasibility of waste management strategies. This includes the evaluation of the costs and benefits associated with alternate waste prominent options such as compost processes and anaerobic digestion (AD), conduct of cost-benefit analyses, and consideration of the potential for revenue generation through resource recovery or recycling schemes. Techno-economic assessment (TEA) strategies play a

vital role in the evaluation of the feasibility and sustainability of biogas and compost production (CP) systems. These strategies integrate various analytical approaches to assess upon the technical and economic aspects of considered waste to value added product conversion processes. In terms of cost analysis, studies devoted to the evaluation of capital costs, operational expenses, and maintenance costs of the commissioned and functioning of biogas and compost facilities (Jones and Salter, 2013 and Anderson et al., 2013; Wang et al., 2022; Lin et al., 2019). Revenue estimation involves an assessment of the potential income generated from biogas sales based on market prices and the value derived from the sale of the compost products (De Clercq et al., 2019). Feasibility analysis involves the examination of the technical aspects including feedstock availability, infrastructure requirements, energy conversion efficiency, and waste management regulations (Emery et al., 2007). These comprehensive assessments provide valuable insights into the viability and potential benefits of the integration of biogas and compost yield within waste management systems.

1.4 Targeted perspectives

This thesis aims to provide a comprehensive understanding of MSWM in Guwahati City and explore various aspects related to waste characterization, generation rate prediction, GHG emissions, and the production of compost and biogas. A survey-based quantification will be conducted on the waste composition of different entities in MSW, followed by the classification of waste data among socio-demographic groups based on income levels and seasons. To model and forecast the total MSWG, organic fraction and recyclables varied ML and time series modelling will be deployed. The thesis will also focus on ML-based prediction and forecasting of GHG emissions and particulate matters (PMs) from MSW landfill and incineration processes. Furthermore, meteorological parameters and MSW organic waste (OW) data will be used to predict and forecast compost and biogas production (BP) rates using ANN and hybrid ML modeling techniques. The research will extend to the development of techno-economic calculators for alternate BP designs and compost facility scales. Accordingly, it enables optimal prediction and long-term forecasting of BP and CP rates. Overall, this thesis will provide targeted perspectives on waste management, renewable energy generation, and environmental impacts in Guwahati City. Thereby, it will provide valuable insights for sustainable waste management strategies and policy-making in the North-east India region.

1.5 Prior Art Summary

1.5.1 Solid Waste Quantification and Characterization

1.5.1.1 HH Survey based Solid Waste Characterization

HH solid waste constitutes a significant portion of MSW and is an inherent aspect of the daily life routines of city residents (Hering, 2012). Henceforth, the available prior art emphasizes upon the quantification of the waste composition from HH level survey. In a study conducted by Ogwueleka (2013), the author examined the daily HH solid waste production of 74 HHs in Abuja, Nigeria for a period of seven days. The results revealed that the average rate of HH solid waste generation was 634.0 grams capita^{-1} day^{-1}. The study also indicated that the majority waste constituted compostable materials. These were evaluated as 63.60% of the total waste along with 9.70% for paper and 8.70% for plastic wastes. A survey conducted in Hanoi City, Vietnam, involved an analysis of the HH solid waste generation rate and composition (Phuong et al., 2021). The study sampled 110 HHs in regions with variant population density and HH scale. The results conveyed an average waste generation rate of 0.63 kg person^{-1} day^{-1}, which was marginally higher rates in rural areas. Food and garden waste comprised the largest proportion and as 78.9%. Thereafter, plastic and paper constitute larger volumes. However, relevant sampling technique was not incorporated in the HH stratification. As a well-known technique, stratified random sampling is vital in HH surveys to achieve representativeness, reduce variability, allocate resources efficiently, and provide accurate estimates for specific population subgroups. The method divide the population into distinct strata based on relevant characteristics, and thereby ensures proportional representation in the sample (Beigl et al., 2008). Ojeda-Benitez et al. (2008) examined HH solid waste generation by analyzing the garbage bags produced by 125 families and in a time period of eight days. Three family typologies (nuclear, extended, and mono-parental) were studied to infer upon the composition and quantity of generated solid waste. The results affirmed that the waste generation varied with respect to family typology and socioeconomic stratum. While the nuclear family was analyzed with a per capita daily waste generation of 1.10 kg, the extended family produced 0.782 kg, and the mono-parental family generated 1.35 kg of capita^{-1} daily waste.

A study was also conducted to determine the quantity and composition of HH waste. Along with these, the driving forces for the waste generation were explored along with potential mitigation (Zhang et al.,

2018). Accordingly, a survey involving 418 HHs in Shenzhen City was conducted. The survey revealed that HH size and income were significant factors to influence HH waste generation. Another research group conducted waste characterization surveys for HH residents' and SWM program operators. Field investigations and on-site waste measurements and associated characterizations were targeted in Nablus district of Palestine (Al-Khatib et al., 2010). Among 1068 surveyed HHs, majority (71.8%) were single-family residences. For the case, the average family size was 6.5 persons. The highest waste percentage (37.9%) was obtained for families with 4-6 members and the lowest percentage (11.9%) was achieved for families with over ten members. The respondents were predominantly residents of villages (52.1%) and the city (41.2%), and only a small portion (6.7%) resided in refugee camps. In terms of income, most HHs (37.7%) received a monthly income ranging from 376 to 750 USD, and only 11% earned above 1250 USD a month. The results conveyed that on a percentage weight basis, the compostable organics (garden and food waste), paper and card, and plastic were 65.1%, 9.1% and 7.6% respectively. In the mentioned theme of HH survey based solid waste quantification, a summary of the most relevant literature has been presented in Table 1.1.

Table 1.1 A summary of the survey-based findings of HH level based solid waste quantification.

S. No.	Scale	Sample size	Output	Waste generation	Critical Findings	References
1	HH level (Abuja, Nigeria)	74 HHs (Seven days)	Waste generated (kg capita^{-1} day^{-1})	Total waste generation 0.654 kg capita^{-1} daily and an average of 0.648 kg capita^{-1} daily	• The correlation between HH income and consumption patterns directly influences the composition and quantities of HH waste, leading to alterations in waste generation. • HH size negatively correlated with daily per capita waste generation in high-income group, slightly significant in medium-income group, and not significant in low-income group. • The average bulk density of the HH waste from all three-income groups was 240 kg m^{-3}.	Ogwueleka, (2013)
2	HH level (Hanoi city, Vietnam)	110 HHs (One week)	Waste generated (kg HH^{-1} day^{-1})	Average HH solid waste generated: 2.35 kg HH^{-1} day^{-1} and 0.67 kg person^{-1} day^{-1}	• OW constituted a significant portion of the HH solid waste, indicating the potential for effective waste management strategies. • Variations in waste composition across different residential areas, indicating the importance of tailored waste management approaches based on local characteristics and needs.	Phuong et al., 2021
3	HH level (Mexicali, Mexico)	125 families (Eight days)	Waste generated (kg capita^{-1} day^{-1})	Per capita waste generated in high-, middle- and low-income groups are respectively: 1.058 kg, 1.04 kg, 0.886 kg.	• The family unit significantly influences the perception and behavior of its members regarding consumption patterns and waste generation and handling. • Negative impact of HH size on per capita quantities of solid waste • Waste generated in family typology, 1.10 kg for the nuclear: 0.782 kg for the extended and 1.35 kg for the mono-parental.	Ojeda-Benitez et al., 2008
4	HH level (Shenzhen & Tianjin, China)	418 HHs (One month)	Waste generated (Kiloton)	HH food waste: 1,673 ± 144 Kt	• The quantity of HH food waste generated is influenced by HH size and income. • Investigated the characteristics of food waste, encompassing its quantity as a percentage of the food consumed, the ratio of avoidable and non-avoidable waste, and the composition of the generated food waste.	Zhang et al., 2018
5	HH level (Nablus district, Palestine)	1068 HHs (Two months)	Waste generated (kg person^{-1} day^{-1})	The average daily waste generation rate ranges between 0.68 and 1.02 kg person^{-1} day^{-1}	• The average daily generation rate of solid waste ranges 0.82 person^{-1} day^{-1} greater than rural area and refugee camps. • Significant contamination issues due to increased HH size leading to inadequate handling of large volumes of waste. • Average mean value of waste generated 0.82 kg person^{-1} day^{-1}	Al-Khatib et al., 2010

S. No.	Scale	Sample size	Output	Waste generation	Critical Findings	References
6	HH level (Uttarakhand)	47 HHs (One month)	Waste generated (kg capita^{-1} day^{-1})	The average (HSW) generation rate was 0.26 kg/c/d and it was composed of OW (57.3%), plastics (14%), paper (10.9%), and glass and ceramic (1.3%) and other materials (16.5%	• 329 solid waste samples from 47 HHs, involving a total population of 220 were collected. • A total of 374.263 kg of waste is collected during the survey. • The results of descriptive statistics of the waste samples shows that the average generation of waste from a HH was 0.26 ± 0.08 kg capita^{-1} day^{-1} with an average 4.68 residents/HH. • The minimum and maximum values of generated waste are 0.134 and 0.528 kg capita^{-1} day^{-1}, respectively, which indicate significant variability of per capita waste generated by different HH. • The largest component, which accounted to 57.3% was found to be the food or OW.	Rawat and Daverey, 2018
7	HH level (Dehradun)	144 HHs (Three months)	Waste generated (g capita^{-1} day^{-1})	The average HH waste quantity in HHs was estimated: 267.17 g/day (SD = 38.13, n = 144).	• The food/kitchen waste was the major constituent (≥80% of total weight) of HH waste in the Dehradun city followed by polythene and plastic (≈7%), paper (≈6%), cardboard (≈2%), glass/ceramic scrap (≈1%). • The maximum average quantity of HW was 680.83 ± 564.19 g capita^{-1} day^{-1} in HH with 8 family members followed by 342.93±314.49 g capita^{-1} day^{-1} in HH with 6 family members. • 240.74 ± 381.67 g capita^{-1} day^{-1} in HH with 5, 236.13 ± 114.05 g capita^{-1} day^{-1} in HH with 7, 136.97 ± 118.0 g capita^{-1} day^{-1} in HH with 4 and 129.16 ± 118.75 g capita^{-1} day^{-1} in HH with 2 family members	Suthar and Singh, 2014
8	HH level (Beijing, China)	113 families (Ten days)	Waste generated (kg persons^{-1} day^{-1})	The generation rate of HH wastes was 0.23 kg persons^{-1} day^{-1}. HH waste consisted of mostly of kitchen waste, followed by paper/cardboard, plastics with each approximately 69.3%, 10.3%, and 9.8% respectively	• A total of 856.3 kg or 0.8563 tons of solid waste was collected, involving a population of 368. • The total bulk density of wastes was 221 kg/m^3. • Also, the HH size was negatively related to daily per capita generation of total HH wastes, kitchen wastes, paper and plastics.	Qu et al., 2009

9	HH level (Chittagong City)	75 HHs (Six months)	Waste generated (kg persons^{-1} day^{-1})	The residential waste generation rate by a person in the study area (Bangladesh) was 0.25 kg day^{-1}.	• The study area comprises a population of almost 3500 persons. • Almost 875 kg of solid waste was generated per day. • The study showed the per household waste generation rate is 1.3 kg day^{-1} and 0.25 kg persons^{-1} day^{-1}. This finding varies from the value (0.15 persons^{-1} day^{-1}) that was recorded by The World Bank (1999).	Sujauddin et al., 2008
10	HH level (Mekong Delta city)	100 HHs (Two months)	Waste generated (g capita^{-1} day^{-1})	The average household solid waste generation rate was 285.28 g capita^{-1} day^{-1}. The compostable and recyclable shares respectively accounted for 80.02% and 11.73%.	• The average daily generation rate of solid waste ranges 0.82 person^{-1} day^{-1} greater than rural area and refugee camps. • The average of total HSW generation in Can Tho city was 285.28 g capita^{-1} day^{-1} (including 283.10 g capita^{-1} day^{-1} for dry season and 287.46 g capita^{-1} day^{-1} for rainy season) for an average of 4.41 residents per HH. • The food waste accounted for the largest part of the total (84.18e85.10%), followed by plastic (6.37e7.15%), and paper (4.73e4.75%).	Thanh et al., 2010
11	HH level (Semarang)	102 HHs (Seven days)	Waste generated (kg persons^{-1} day^{-1})	The average waste generated in Semarang was 0.385 kg persons^{-1} day^{-1}, which is relatively low compared with waste generation in other developing countries.	• The average density of the waste was about 295 kg m^{-3}. • Putrescible contributed to about 70% of the total solid waste an estimated 340 t were generated daily. • Plastics were the second largest component that contributed significantly to the Semarang waste stream in terms of its volume (30%) and weight (10.6%). • Paper products accounted for 10.2% of the weight, or 25% by volume. • The maximum and minimum waste HH^{-1} generation day^{-1} weight (kg) was respectively 6.11 and 0.13	Supriyadi et al., 2000
12	HH level (East Coast, Malaysia)	338 HHs (Five months)	Waste generated (kg capita^{-1} day^{-1})	The results of the study revealed that 74.3 % of households disposed of food debris as waste and 18.3% disposed of plastic materials as waste.	• The average bulk density of the HH waste from all three-income groups was 321 kg m^{-3}. • The study also showed that 50.3% of the households segregate their waste while 49.7% did not.	Fadhullah et al., 2022

1.5.1.2 Seasonal Influence on Solid Waste Quantification and Characterization

For the purpose, a two-stage survey was conducted in Can Tho city, Vietnam. It involved 100 HHs being surveyed during dry and rainy seasons (Thanh et al., 2010). The HH solid waste was collected from each HH and was classified into 83 subcategories. The study revealed that the average HH solid waste generation in Can Tho city was 285.28 gm capita^{-1} day^{-1}. Thus, parameter was marginally higher during the rainy season (287.46 gm capita^{-1} day^{-1}) in comparison to the dry season (283.10 gm capita^{-1} day^{-1}). The per-capita generation of HH hazardous waste was approximately 0.53 gm day^{-1}. Compostable materials accounted for 80.02% of the waste, and the recyclable materials represented only 11.73% of the total waste. Another research group carried out a comprehensive four-stage systematic tracking survey by considering a size of 240 HHs for one week in each season during the period of the summer of 2011 to the spring of 2012 (Gu et al., 2015). The results conveyed that Suzhou's HH solid waste generation rate was 280.5 gm capita^{-1} day^{-1}. Accordingly, an annual generation of HH solid waste of about 568 thousand tons was apparent. Notably, 89.3% of the generated waste constituted compostable and recyclable materials. The studies conducted by another research group (Gómez et al., 2009) involved the characteristics of MSW composition in Chihuahua city, Mexico, and for various seasons. During three periods in 2006 and 2007 (April, August, and January), the authors compared the findings of solid waste characterization and generation from HHs in the city. For all considered seasons, the average daily waste generation in Chihuahua, was obtained as 0.592 kg capita^{-1}. The results conveyed that the winter season ascertained the lowest waste generation rate and the waste constitution refereed to largest proportion of OW (45%), followed by paper (17%) and other waste categories (16%). A study conducted in Lahore, Pakistan, revealed that more plastic and textile wastes are generated in winter season, whereas food and yard waste are more predominant in spring season (Kamran et al., 2015). During winter months, the lowest income group generated the minimum amount of MSW, and with an average of 0.39 kg capita^{-1} day^{-1}. In comparison, the high-income and middle-income group generated 1.1 kg capita^{-1} day^{-1} and 0.56 kg capita^{-1} day^{-1} respectively. Based on these limited studies reported till date on seasonal variations influencing the volume and composition of solid waste, Table 1.2 presents a brief summary of the adopted research methodology and associated outcomes.

Table 1.2 A summary of the prior art investigations addressing seasonal influence on the solid waste characterization and quantification.

S. No.	Author (s)	Study area	Sample size	Duration	Key findings
1	Thanh et al., 2010	Mekong Delta city, Vietnam	100 HHs	Two-stage survey (dry season and rainy season in 2009)	• The rate of MSWG is influenced by expenditure and population density level of a HH. • The average HH waste generated was 285.28 gm capita^{-1} day^{-1}. • The compostable and recyclables accounted for 80.02% and 11.73% respectively • Organic mainly (kitchen waste) was determined in the rainy (May through November) and dry seasons (December through April) with not much change.
2	Gu et al., 2015	Suzhou city in East China	140 HHs (67 in the old district and 73 in the new district)	Summer of 2011 to the spring of 2012 (one week survey for each season)	• For high income people, their consumer demands tend to diversify, resulting in the disposal of materials that are increasingly complex in nature. • Winter season generates more waste (355 gm capita^{-1} day^{-1}) and summer the least 248 gm capita^{-1} day^{-1}.
3	Gómez et al. in 2009	Chihuahua city, Mexico	87 HHs	Three periods (April and August, 2006 and January, 2007)	• OW constituted approx. 45% of all generated MSW. • The average amount of MSW generated per capita was 0.516, 0.628 and 0.631 kg capita^{-1} day^{-1} for levels I, II and III, respectively. • Amount of waste (%) for levels I, II and III, respectively, OW with 43%, 49% and 44%, then paper, with 13%, 18% and 19%, followed by plastics with 15%, 12% and 12%. • There was a decrease in waste quantity across all three socio-economic levels in January, during the low temperature season. • OW dominated waste composition at all three socio-economic levels, especially during April. However, the organic fraction decreased in August and January across all levels. • The highest income level (Level III) generated the largest quantities of uneaten food and cardboard, likely due to their consumption of packaged food products.

1.5.2 Prediction of Solid Waste Generation Rate

Solid waste generation rate prediction using ML modeling offers several significant advantages. Firstly, ML models leverage upon historical data and adapt advanced algorithms to analyze complex patterns and relationships. This assures highly accurate predictions. This accuracy is crucial for waste management authorities as it enables them to effectively plan and allocate resources, minimize inefficiencies and reduce costs. Secondly, ML models are scalable, and allows for various geographic regions and time periods. Such scalability emphasizes that waste management strategies can be tailored to specific areas and adjusted as per the need. Additionally, ML models can provide real-time insights based on the latest data, and thereby enable waste management agencies to make prompt informed decisions and efficiently respond to dynamic waste generation patterns efficiently (Kontokosta et al., 2018). Such real-time capability enhances the overall effectiveness of waste management systems.

The prediction of MSWG rate is important to plan and realize effective waste management and other mandatory targets such as the minimization of environmental impact, optimization of resource utilization, and promotion of sustainable practices. The associated methodology supports evidence-based decision-making, facilitates policy development, and enables the implementation of waste management strategies that align with environmental, social, and economic goals (Kannangara et al., 2018). In this field of research, the adopted methodologies primarily refer to the identification of competent socio-economic and demographic factors and evaluation of competent ML algorithms for solid waste prediction. A brief account of these approaches in the prior art has been presented in the following sub-sections.

1.5.2.1 Selection of Socio-Economic and Demographic Factors for Improved Prediction of Solid Waste Generation Rate

Several socio-economic and demographic variables that measure economic affluence and life style parameters have been found to enhance the MSWG rate of a city or a village (Ghanbari et al., 2021). These primarily refer to geographical location, population, literacy rate, food habits, culture and beliefs, number of HHs in a specific geographical area and various economic parameters such as gross domestic product (GDP) (Grazhdani, 2016) and employment status (Sankoh et al., 2012; Kontokosta et al., 2018).

Furthermore, HH waste generation rate is highly heterogeneous and is widely dependent upon the socio-economic status of the HH (Suthar and Singh, 2015; Khalil et al. 2022; Miezah et al. 2015; Sankoh et al., 2012). Thus, socio-economic and demographic variables that account for the economic affluence and life style parameters have been analyzed to critically influence the MSWG rate. Notably, these refer to population income (Sankoh et al., 2012), consumption expenditure level (Daskalopoulos et al. 1998) and purchasing power parity (Abbas et al. 2022; Zhu & Rahman 2020). All these relatively influenced the MSWG rate.

A study conducted by Nguyen et al. (2020) targeted the influence of socioeconomic alterations on the MSW characteristics in Taiwan. The study suggested a strong correlation between economic conditions and the generated recyclable waste. In a study conducted by Kolekar et al. (2016), the authors summarized the influence of economic, socio-demographic, and management-oriented factors on the MSWG rate. Accordingly, the authors examined the interplay of these factors were examined to infer upon the critical influence on the MSWG patterns. Similarly, Ma et al. 2020 emphasized upon the socioeconomic influential factors that affect the physical composition of MSW in China. Precisely, the authors investigated the influence of city size and economic development level on the composition of MSW. Namlis and Komilis, (2019) examined the relationship between the GDP, human development index (HDI), unemployment rate, and the generation rates of thirteen solid waste streams in ten European countries. The findings of the study inferred that a positive correlation exists between GDP and waste generation rate. Accordingly, it was suggested that the higher GDP levels infer upon increased waste generation rates. A brief account of the important findings in this research theme has been presented in Table 1.3.

Table 1.3 A summary of the literature findings on the socio-economic and demographic factors based solid waste characterization and quantification.

S. No.	Author(s)	Factors		Data type	Approach	Data Sources	Key findings
		Socio-economic	***Demographic***				
1	Grazhdani, 2016	EDU level, income	Housing structure, mean building age and waste management policy	Primary	Regression models	Survey (questionnaires)	• A robust positive correlation exists between GDP and the level of HH participation in waste generation and recycling activities. • For the annual per capita waste generation: R^2 = 0.92, Adjusted R^2 = 0.78, chi^2 (1) = 1.01 and Prob > chi^2 = 0.3687 • For the annual per capita waste disposal: R^2 = 0.89, Adjusted R^2 = 0.71, chi^2 (1) = 9.54 and Prob > chi^2 = 0.0087
2	Nguyen et al. 2020	HH characteristics, marital status, public safety, labor and employment, income status	Population size	Secondary	Statistical data analysis	12-year worth of self-reported dataset (Taiwan)	• The new Taipei city administrative unit generated total waste: 159,865 ± 144,635 (in tons) (mean ± std); kitchen waste: 121,088 ± 37,676, paper and plastics (recyclables): 239,957 ± 52,632 and 41,383 ± 12,752 respectively. • Waste generation rate is 0.82 kg/ person/ day • Plastic and glass waste generation decrease when the GDP increases.
3	Namlis and Komilis, 2019	GDP, HDI, Unemployment Rate	--	Secondary	Statistical data analysis	7-year 13 solid waste streams database from ten European countries, Eurostat (2008 to 2015)	• Norway: 23 kg cap^{-1} y^{-1} and a GDP: 100,000 \$ cap^{-1} y^{-1} • Romania, with a GDP less than 10,000 \$ cap^{-1} y^{-1} and waste generated: 5 kg cap^{-1} y^{-1}. • HDI is positively while unemployment rate is negatively correlated
4	Suthar and Singh (2015)	Number of residents per sampled HH, total income of the HH Waste management factor: recycling and disposal options	--	Primary	Statistical data analysis	Waste sampling of 144 houses and questionnaire approach for 3 months (March to May 2011)	• HH waste generation rates in the city ranged from 24.5 to 4147.1 g/day. • The average quantity of HW in HHs was estimated to be 267.17 g/day (SD = 38.13, n = 144). • Food/kitchen waste constituted the majority, accounting for at least 80% of the total weight of HH waste. • Polythene and plastic comprised approximately 7% of the HH waste weight. • Paper and cardboard accounted for around 6% and 2% of the HH waste weight respectively.

1.5.2.2 ML Algorithms-based Prediction

The available literature for the prediction of the MSWG rate refers to a wide range of empirical and abstract modelling techniques (Molina-Gómez et al. 2021). These can be as simple as application-based pedagogies or as complex as sophisticated technologies with thorough academic insights (Chung, 2010). Abbasi and El Hanandeh (2016) examined waste generation in Logan, Australia for a longer period of twelve months. Accordingly, the authors deployed short-term prediction models such as support vector regression (SVR), ANN, adaptive neuro fuzzy inference system (ANFIS), and k-Nearest Neighbour (*k*NN) for MSWG rate prediction. Among all models, the study reported highest accuracy for the ANFIS model (R^2 and RMSE of 0.98 and 175.18 respectively). Kumar et al. (2018) investigated the influence of EDU, occupation, income, and house type on the weekly plastic waste generation rate prediction in Dhanbad, India. Accordingly, short-term prediction models, such as ANN, SVR, and random forest (RF), were deployed for a dataset comprising 120 observations. The ANN model demonstrated highest accuracy among the mentioned models (R^2 of 0.75 and an RMSE of 9.53). Abbasi et al. (2019) conducted a short-term study to predict monthly waste generation. Accordingly, the authors used nine input parameters, namely GDP, rain, maximum temperature (T_{max}), population, HH size, educated man and women, and the unemployment rate. For the study, radial basis function (RBF)-ANN, SVR, and ANFIS models were deployed to model a data size of 252 sets. The RBF-ANN model achieved an R^2 of 0.68 and an RMSE of 0.12 for monthly prediction. Also, the conducted study in the Tehran city, Iran included a feature correlation analysis.

Ayeleru et al. (2021) deployed four input parameters, namely population, formally employed, unemployed, and the number of family units for the prediction of MSWG rate with 30-year dataset in Johannesburg, South Africa. The authors employed long-term ANN and SVR models and achieved the best accuracy (R^2 of 0.99) with ANN. An overview of the important findings in this research theme has been mentioned in the Table 1.4.

Table 1.4 A summary of prior art findings for the MSWG rate prediction with ML algorithms.

S. No.	Input features	Output	ML algorithms	Forecasted period	Data size	Performance evaluation (Best model)	Demogra -phic scale	Critical findings	References
1	Population, employed and unemployed numbers, family size	30 years MSW generation	ANN, SVR	Long-term	30	ANN (R^2: 0.99)	City level	• The average annual prediction of 1.782 million tons. • The forecasted period (2021-2050): 1614800-1951139 tons respectively of MSW.	Ayeleru et al. (2021)
2	Population counts over 45 years, income, employment rate	Annual MSW generation	ANN, decision tree (DT)	Mid-term	1553 (MSW Census) and 1867 (Bluebox/ paper-Census)	ANN (R^2: 0.72)	City level	• Use feature selection techniques • Total MSWG: 5,00,000 (tons yr^{-1}) • MSWG per HH: 900kg HH^{-1} yr^{-1} • Paper waste and blue box waste per HH: 105 and 150 kg HH^{-1} yr^{-1} respectively	Kannangara et al. (2018)
3	HH types, income, employees, EDU level	Weekly municipal waste generation	GBRT, ANN	Short-term	-	GBRT (R^2: 0.87 RSME: 0.034)	Building	• Analyze feature importance • The total waste generation model performed the best with R^2 value of 0.87 and 85.3% of accuracy • R^2 for the refuse model is 0.87 with 82.3% accuracy 20% MAE • The metal/glass/plastics and paper recycling models showed lower prediction accuracy with R^2 values of 0.73 and 0.78	Kontokosta et al. (2018)
4	Waste generation in past twelve month	Monthly MSW generation	SVR, ANN, ANFIS, *k*NN	Short-term	-	ANFIS (R^2: 0.98 RMSE: 175.18)	City	• High accuracy • The performances indices for ANFIS model (MAE, RMSE, MAPE, R^2) for train and test score are 0.001, 0.002, 3.39E−06, 0.99 and 52.16, 175.18, 0.008, 0.98 respectively. • The five-year prediction results indicated 1,270,000 kg of waste increased by the year 2020.	Abbasi and El Hanandeh (2016)

1.5.3 Prediction of GHG and PMs Emissions Rate from Incineration and Landfill Sites

Accurate GHG emission predictions using ML models facilitate the implementation of effective mitigation strategies and waste management practices. Singh et al. (2022) examined the GHG emissions from MSW incineration site in a WtE plant in Delhi, India. The plant processed 1300 tons of MSW on a daily basis and generated about 700-750 kWh ton^{-1} of electricity of MSW. It was inferred that the GHG emissions from the plant were about 910 kg CO_2 e per ton^{-1} of processed MSW. For the case, the CO_2 emissions accounted about 83% of the total GHG emissions. Further, CH_4 and N_2O emissions respectively accounted about 11% and 6% of total GHG emissions. These findings highlight upon the importance of waste management strategies that target upon the reduction of organic content in the MSW. Accordingly, GHG emissions from the incineration process can be significantly minimized. Few authors investigated the GHG emissions from MSW incineration in Kathmandu, Nepal (Dhakal et al., 2010). The study found that the carbon content of the MSW had a significant influence on GHG emissions, and higher carbon content lead to higher CO_2 emissions.

Fallah et al. (2020) conducted a study on the prediction of landfill gas emissions. For the purpose, the authors deployed an auto-regressive neural network model. For a dataset of 1883 samples, the considered input parameters were mean T_{atm}, maximum pressure (P), relative humidity (RH), maximum wind speed (WS). Accordingly, for the mentioned four input parameters, CH_4 generation rate was targeted for prediction. To evaluate the performance of the model, the mean absolute percentage error (MAPE) was deployed as the metric. The model achieved a MAPE of 3.03%, and thus affirmed a relatively low error rate for the prediction of CH_4 generation rate. Abushammala et al. (2014) deployed an ANN model to predict landfill gas behavior. For a dataset of 121 samples, five input parameters namely soil and T_{atm}, soil moisture content (MC), CH_4 concentration, and O_2 concentration were considered to target the prediction of CH_4 oxidation rate. Accordingly, the model was inferred to provide a promising level of correlation with the actual CH_4 oxidation rates, (R^2 value of 0.937). Confirming an accurate prediction, the mean squared error (MSE) value of 0.0082 was reported.

However, in the precise field of ML modelling-based prediction and forecasting of emissions and particulate matters, there has been a dearth of literature for open dumps and incinerators. This conveys

Table 1.5 A summary of the key findings of the city-wide GHG and PMs emissions rate investigations.

S. No.	Application	Input feature(s)	Output	ML approach	Data size	Error metrics	Key findings	References
1.	Landfill gas prediction	Average T_{atm}, P, RH, WS	Daily CH_4 generation rate	Auto-regressive neural network	1883	MAPE 3.03%	• CH_4 prediction with the average index of agreement (IoA) of 0.92 and co-efficient of correlation, R: 0.85 • The model proved to be effective with 84.3% reduction (6.4 times) in MSE during the testing phase. • CH_4 daily generation: 150,868.913 m^3 day^{-1}	Fallah et al. (2020)
2.	Landfill gas prediction	T_{atm}, soil temperature, soil MC, CH_4 and O_2 concentration	CH_4 Oxidation	ANN	121	R^2: 0.937 MSE: 0.0082	• R value ranged from 0.81 to 0.99 and efficiency of the model was 0.8978. • MSE in the training phase was 7.0766×10^{-7} • As the experimental results were expressed in removal efficiency, which falls between 0 and 100%, the transfer functions utilized at the hidden and output layers were linear. These functions were well-suited to the logsig function.	Abushammala et al. (2014)
3	Pollutants emission monitoring	Injected frequency, amount of activated carbon, the concentration of hydrogen chloride in flue gas, mixing chamber's temperature	Dioxin emission	ANN	63	R^2: 0.99	• For testing dataset, RMSE, MAE and R^2 values are respectively 0.0189, 0.0140 and 0.6273 • The test set revealed that the model's predicted results were higher than the actual observed values, suggesting that the model did not meet the necessary standards.	Bunsan et al. (2013)
4	Landfill gas prediction	Total waste landfilled, temperature, annual precipitation, landfill age, depth, and cover.	CH_4 generation rate	LandGEM and Fuzzy logic model		R^2 of 0.951	• LandGEM, the fuzzy based model showed better results. • When plotting the LandGEM and Fuzzy model values to the actual measured data, the fuzzy model resulted in a better fit to actual data than the	Mohsen and Abbassi, 2019

that a significant research gap exists and prompts upon the need for studies that leverage upon ML methodologies for the accurate prediction of greenhouse gas and particulate matter emissions. Bridging upon this knowledge gap is very important to advance into trans-disciplinary activity that promote sustainable waste management practices in the realm of solid waste disposal. Table 1.5 summarizes key findings in the mentioned field of research.

1.5.4 Compost Generation Rate Prediction with ML Algorithms

To further augment upon the ML modelling applications, a comprehensive analysis of the influence of meteorological parameters and the volume of OW on compost maturity is verily required for compost systems. Such a knowledge paradigm can be leveraged to optimize composting processes and thereby ensure high-quality CP. All these contribute to sustainable waste management practices. Gao et al. (2007) conducted a study on compost maturity prediction and with wavelet neural network (WNN) model. For the study, the dataset referred to 500 samples. The WNN model was customized with five input parameters i.e., T_{max} duration, MC, volatile solids (VS), the value of fecal bacteria, and germination index (GI). The targeted maturity level variable was further classified into four categories namely full maturity, preferable maturity, general maturity, and immaturity. The performance of the model was evaluated with the MSE index. Accordingly, an MSE value of 0.066 was obtained, and ascertained upon good prediction accuracy.

The composting maturity prediction with the ANN model was investigated by Kujawa et al. (2014). Utilizing 1536 samples-based dataset, the ANN model was built with color and texture features as input parameters (14 to 49 range). For the prediction, compost maturity level was targeted. This was classified into two categories (early maturity or not early maturity). The model performance was evaluated in terms of classification error (1.56% reported value). Such a low value of the classification error affirmed the accuracy of the ANN model to predict the maturity level. Similarly, Kujawa et al. (2020) predicted compost maturity by deploying convolutional neural network (CNN) model. With the deployed dataset of 1312 RGB images, the CNN model being designed for a 2-level classification task, was targeted for the prediction of the probability of the maturity classes. Accordingly, the classification error altered from 0.51% to 17.77%, and thereby confirmed upon very good accuracy of the CNN model to predict

Table 1.6 A summary of the findings associated to the prediction of composting process with ML algorithms.

S. No.	Variables		ML approach	Size of the dataset	Performance metrics	Key findings	Authors
	Independent	***Dependent***					
1.	T, MC, VS, the value of fecal bacteria and GI	4-classification: full, general and preferable maturity, immaturity of the compost	WNN	500	MSE: 0.066	• 500 samples were divided into 5 groups, and each time 80:20 ratio of sample data is divided as training and testing samples • Training, testing error and average iteration times were 0.066, 0.125 and 351 respectively. • The training error and testing error were calculated as the averages of five iterations each.	Gao et al. (2007)
2.	RGB images of the compost	Compost maturity	CNN	1300	Classification error: 0.51% to 17.77%	• Five models were employed (CNN16 VIS, CNN4 UVA1s, CNN8 UVA5s, CNN16 UVA10s and CNN16 MIX) • The MIX-based model, incorporating 16 convolution filters, achieved the highest classification accuracy with an error rate of only 0.51% • The optimal VIS model utilized 16 convolution filters and achieved a classification error rate of 4.57%. • The model exhibited an error rate of 6.60% in misclassifying cases from the test set.	Kujawa et al. (2020)
3	Colour and texture features of compost	Early compost maturity prediction	ANN	1530	Error prediction: 1.56%	• The most effective models for the VIS, UV-A, and MIX light variants were identified as MLP 31-50-2 VIS, MLP 30-50-2 UVA10s, and MLP 35-50-2 MIX, respectively. • Their classification error for the test set achieved the values of 4.43%, 3.39% and 1.56% for MLP 31-50-2 VIS, MLP 30-50-2 UVA10s and MLP 35-50-2 MIX respectively.	Kujawa et al. (2014)

the RGB images-based compost maturity. Table 1.6 summarizes upon the most relevant literature findings in this research theme.

1.5.5 Biogas Generation Rate Prediction with ML Algorithms

Very few literatures elucidated upon the ML algorithm application for biogas generation prediction. Thus, limited studies were devoted towards the influence of meteorological parameters. Ugwu and Enweremadu (2019) investigated the kinetics of biomethane production from vegetable waste. The transference model exhibited the highest accuracy, with R^2 and Adj. R^2 values of 0.983 and 0.982, respectively. The experimental methane yield was 270.98 mL g^{-1} VS^{-1}, all the theoretical yield ranged from 342.06 to 444.48 mL g^{-1} VS^{-1}. The simulation model prediction difference altered from 0.03% to 1.28%. The growth functions, particularly the transference and first-order models, accurately predicted the kinetics of biomethane production.

Wannasek et al. (2017) conducted lab scale experimentation to infer upon the efficacy of biogas yield from maize silage in terms of T, precipitation (Pr), and biomass dry matter yields. The study found that the biomass dry matter yields ranged from 3.65 to 20.67 tons ha^{-1}. Remarkably, the biogas $hectare^{-1}$ yields exceeded 6000 Nm^3, and thereby indicated a high potential for BP. The prediction of biomethane yield from agricultural residue, forest residue, and urban solid waste was investigated using ML approaches, and with ANN, DT, Multiple linear regression (MLR), SVR, and *k*NN (Wang et al., 2022). A dataset of 277 samples was used in the study. Among the alternate models, the *k*NN model exhibited the best performance (R^2 value of 0.75 and an RMSE of 38.56 mL g^{-1} VS^{-1}). Zhang et al. (2022) investigated the accurate prediction of biogas yield using hybrid ML models. It was inferred that influential operational parameters such as pH and T are the vital factors to affect the performance of the AD processes. A MAPE of 2.15% was achieved for biogas yield prediction. A brief account of important findings in the mentioned research theme is presented in Table 1.7.

Table 1.7 Key findings summary for the bio-gas generation rate prediction with ML algorithms.

S. No.	Variables		Approach	Size of the dataset	Performance metrics	Key findings	Authors
	Independent	***Dependent***					
1.	Total solids, VS, ash content MC, pH, T	Biomethane production	Kinetic models	Experimental data	R^2, Adj. R^2 and RMSE values are respectively 0.983, 0.982 and 9.209 (Transference model)	• The experimental methane yield was 270.98 mL g^{-1} VS^{-1}, • Theoretical methane yields were 444.48 mL g^{-1} VS^{-1} and 342.06 mL g^{-1} VS^{-1} and model simulation ranged from 267.5 to 270.89 mL g^{-1} VS^{-1}. • With a prediction difference of 0.03–1.28%, all growth functions acceptably predicted the kinetics of vegetable waste. • Transference and first-order models best predicted with a prediction difference of 0.03% and 0.31%, respectively. • The statistical indicators of model fitness ranged from 0.946 to 0.983, 0.941 to 0.982 and 9.209 to 16.54 for R^2, Adj. R^2 and RMSE, respectively	Ugwu and Enweremadu, 2019
2.	T, Pr, maize silage	Biogas yield	Lab scale experimentation	Experimental data	-	• The biomass dry matter yields of all harvests ranged from 3.65 to 20.67 tons ha^{-1} • Biogas $hectare^{-1}$ yields surpassing 6000 Nm^3. • Maize silage is better option for biogas production since it has a high specific CH_4 yield.	Wannasek et al. (2017)
3	Agricultural residue, forest residue, and urban solid waste, total solids (TS), VS	Biomethane yield prediction	ML approach (ANN, DT, MLR, SVR, *k*NN)	277 samples	R^2: 0.75 and RMSE: 38.56 mL g^{-1} VS^{-1} (*k*NN)	• The optimal *k*NN model exhibited a MAE of 30.2 mL g^{-1} VS^{-1}. • The best model demonstrated a generalization performance with an average relative error of 10.05%.	Wang et al, 2022

1.5.6 TEA of Biogas Plant

1.5.6.1 Techno-economic Anaerobic Digestion (AD) Calculator and Optimality of Process Parameters

In this field of study, the literature documents two freely available AD calculators (Jones and Salter, 2013 and Anderson et al., 2013). Both AD calculators were developed to address customized application domains. In the first prior art, the scenario involved farmland being operated as a utility to support the AD unit. In this model, the farmers can judge upon the farmland's operability to achieve optimal profit for the AD unit. In the later prior art, the emphasis was upon the biogas yield from the AD unit. Accordingly, the unit's operating parameters were sought to realize the localized demand aspects. Thus, profit maximization was customized for the mentioned perspective. However, for such scenarios, the biogas yields have been often considered as fixed values in both AD calculators. This was the case despite the parameter being sensitively influenced with operational factors such as retention time (RT) and T (Lohani and Havukainen, 2018; Mao et al., 2015).

The above-mentioned articles also mentioned that the AD unit has not been a primary business component of the farmers. It was a supplementary facility being used for the processing of farm's biodegradable waste into biogas. In many such cases, the farmers or plant owners do not have expertise in the AD process. Such a know-how is relevant for India, as the AD utilizers in Indian farms do not have in depth technical expertise. Also, the prior art reported calculators (Jones and Salter, 2013; Anderson et al., 2013) are also not applicable for Indian AD units. This is due to the consideration of a fixed choice of biogas yields. Such an assumption is unfair due to the critical dependence of yield on operating conditions that critically vary with seasonal influence and other parameters. Table 1.8 summarizes key aspects of both mentioned calculators and adopted research pedagogy.

Table 1.8 A summary of the key findings of the AD calculator-based TEA of biogas plants.

Schmack Biogas (The Andersons Centre, 2010)														
		Techno-analysis						**Cost/ Economic analysis**						
Feed Inputs		**Scale & RT (days)**	**BP (m^3 yr^{-1})**	**Methane Content in Biogas (%)**	**Biogas Utilization Efficiency (%)**	**Biomethane Production (m^3 day^{-1})**	**Digester Volume (m^3)**	**Capital Invest-ment ($)**	**Annual Operating & Maintenance Cost ($)**	**Bio-methane Selling Price ($/$m^3$)**	**Total Revenue ($)**	**Total Operating & Maintenance Cost ($)**	**Net Present Value (NPV) ($)**	**Payback Period (PBP, years)**
Material	***Volume***													
Grass silage Cattle slurry Maize silage	4000 t yr^{-1} 8000 t yr^{-1} 4500 t yr^{-1}	Plant based and 55 days	1,740,000	45	64	364	250	1,000,000	50,000	0.2	5,865,900	750,000	4,115,900	4.5
Ribe CAD (The Andersons Centre, 2010)														
Agricult-ural Residue	3250 t yr^{-1}	Farm based and 62 days	73,000	60	63	160	150	500,000	42,000	0.4	1,16,400	550,000	1,10,400	4
Copys Green Farm (RASE, 2011)														
Cattle slurry Maize silage	2500 t yr^{-1} 2500 t yr^{-1}	Farm based and 45 days	1,680	50	55	28	50	100,000	15,000	0.3	503.70	60,000	59,496.3	2
Lodge Farm (RASE, 2011)														
Cattle manure Bio-waste	4100 t yr^{-1} 137 t yr^{-1}	Farm based and 50 days	36,500	40	50	80	180	450,000	28,000	0.2	85,000	87,600	37,600	6

1.5.6.2 Seasonal Influence on AD Unit Performance

Seasonal variations have a notable impact on anaerobic digestion units. The performance and efficiency of these units can be influenced by changes in T, substrate availability, and microbial activity. During colder seasons, such as the winter, the lower T negatively affect the AD process, and lead to reduced microbial activity and slower degradation of organic matter. This can result in reduced BP and longer RT. On the other hand, warmer seasons tend to enhance microbial activity and promote faster decomposition. This leads to improved performance and higher biogas yields. Additionally, seasonal variations in substrate availability, such as fluctuations in agricultural or food waste inputs, can affect the overall stability and efficiency of AD units.

An in-depth understanding into the sensitive influence of seasonal variations on AD units is necessary to ascertain upon their optimal performance, process stability, and assurance of maximal potential benefits of the AD. In tropical and subtropical regions, the digester conditions are often psychotropic or psychrophilic during the winter season. For this reason, the BP is lowered. Till date, many prior art investigations (Sabbir et al., 2021; Rahman et al., 2019) targeted the effect of seasonal variations on AD units and for only laboratory scale systems. Thus, medium to large scale AD units have not been significantly delineated in the literature. Furthermore, limited information exists with respect to the influence of T_{atm} on the T of full-scale AD processes. For such systems, an increase in BP with respect to the T_{atm} has been reported (Kirkegaard et al., 2017; Carrere et al., 2016). Sposob et al. (2020) studied the seasonal fluctuations in the microbial community dynamics and CH_4 yield of two full-scale ADs treating food waste. The authors evaluated the influence of organic load rate (OLR) on the system performance. Also, Kalia et al. (2000) conducted a study on the performance of a dome biogas plant and observed that digester T reduced during the winter season and thereby affirmed crucial influence of the seasonal alterations on the BP profiles. Table 1.9 summarizes the critical findings of the mentioned research theme.

Table 1.9 A summary of the findings associated to the seasonal influence on biogas yield.

S. No.	Technology	Waste type	Seasonal parameters		Digestion period (days)	Biogas yield (m^3 $tons^{-1}$)	Capacity	Authors
1	Full-scale AD plants	Food waste	*Feedstock:* TS (%): Spring: 9.9 ± 0.2, Summer: 11.2 ± 0.0, Fall:11.3 ± 0.6, Winter:10.8 ± 1.1 VS (%): Spring: 8.6 ± 0.5, Summer: 9.7 ± 0.3, Fall: 9.8 ± 0.6, Winter: 9.2 ± 0.9 pH: Spring: 4.5 ± 0.3, Summer: 4.0 ± 0.2, Fall: 4.1 ± 0.2, Winter: 4.4 ± 0.1	*Reactor performance:* OLR (kg VS m^{-3} d^{-1}): Spring: 4.2 ± 0.3, Summer: 5.4 ± 0.1, Fall: 4.6 ± 0.1, Winter: 3.9 ± 0.2 VS removal (%): Spring: 90.1 ± 6.5, Summer: 90.7 ± 4.8, Fall: 79.5 ± 10.0, Winter: 83.7 ± 4.1 pH: Spring: 7.9 ± 0.2, Summer: 8.1 ± 0.4, Fall: 8.0 ± 0.0, Winter: 8.3 ± 0.1	65	CH_4 yield (m^3 CH_4/kg VS) Spring: 0.43 ± 0.2, Summer: 0.34 ± 0.6, Fall: 0.40 ± 0.1, Winter: 0.39 ± 0.1 CH_4 (%): Spring: 63.7 ± 4.0, Summer: 63.7 ± 6.1, Fall: 63.7 ± 3.0, Winter: 63.7 ± 4.1	1000 tons yr^{-1})	Sposob et al. (2020)
2	Fixed dome Janata biogas plant	Agricultural waste	*Ambient temperature:* • During winter months the average ambient temperature is decrease from approx. 25-26°C in summer to 9-10°C, led to a reduction in digester T from around 22-23°C to 13-14°C. • T dropped further to the psychrophilic range of 13-14°C.		55	• 55-60% methane gas produced • The rate of gas production during the peak summer quarters exhibited a decrease of 34%, whereas the decline during the peak winter quarters was comparatively lower at 13%.	3 m^3	Kalia et al. (1998)
3	Fixed-dome AD unit	Slurry	*Ambient temperature:* • The average ambient temperature in the autumn, late autumn, and winter was 29.05, 22.90, and 17.64 °C, respectively. The average slurry T in the autumn (30.38 °C) was higher than in the late autumn (29.36 °C) and in the winter (25.76 °C).		20	CH_4 production in the AD was on an average of 66% of the produced biogas	350 m^3	Sabbir et al. (2021)

1.5.6.3 ML Algorithms-based Prediction of Biogas Yield

Till date, ML models exhibited a great potential to accurately predict biogas yield. ML models utilize advanced algorithms to analyze and learn from large datasets. Accordingly, they can identify complex patterns and relationships between various factors influencing BP. The ML models can consider inputs such as feedstock characteristics, environmental conditions, process parameters, and microbial activity to effectively predict biogas yield (De Clercq et al., 2019). Thus, through the leveraging of ML techniques, researchers and operators can enhance the understanding of complex BP dynamics and optimize process performance. This not only aids in resource planning and system optimization but also contributes to the development of sustainable and efficient BP technologies (Wang et al., 2021).

ML algorithms can ascertain upon the improvement in the enhanced efficiency, sustainability and profitability of BP systems (De Clercq et al., 2019). Till date, the ML has emerged as a powerful tool for the predictive modelling-based BP studies. ML primarily offers promising features such as the ability to accurately predict the BP process, optimize plant performance, and enhance efficiency. Thus, ML algorithms enables biogas producers to identify critical patterns and trends in the data. Such identified patterns can be used to forecast BP based on a combinatorial yet complex alterations of input parameters, such as feedstock composition (Li et al., 2022), T, and pH (Xu et al., 2022). With respect to traditional modelling techniques, the ML were beneficial in terms of speed and accuracy (De Clercq et al., 2019). The ML allows for efficient decision-making and can lead to significant cost savings and increased profitability. Moreover, ML can aid in the optimization of plant performance and through the identification of fields for improvement through research intervention and associated alterations in the operational parameters. This can lead to reduced downtime, increased plant availability, and improved product quality (Zhang et al., 2022). In this field, few investigations can be summarized (Table 1.10). A brief account of these is as follows.

Table 1.10 A summary of the ML approaches-based findings on the prediction of biogas generation rate.

S. No.	Algorithms	Dependent variable	Parameters	Performance metrices	Findings	Authors
1	Logistic regression, SVR, RF, Extreme Gradient Boosting (XGBoost), and *k*NN	Biomethane prediction	Municipal faecal residue, kitchen food waste, food and vegetable waste, and chicken litter	• The best model was *k*NN with R^2 value 0.87. • *k*NN model performance on the Hainan dataset, with an R^2 of 0.86 on the training data and 0.85 on the test data • Model accuracy: 86.01 %	The BP rate was categorized into different bins based on the production amount: "low," "medium," and "high," roughly corresponding to the ranges of 0 to 5033 m^3, 5034 m^3 to 10,067 m^3, and 10,068 m^3 to 15,100 m^3, respectively.	De Clercq et al. (2019)
2	GA, hybrid extreme learning machine (ELM) model coupled with GA	Biogas yield	Feed volume = 23–45 m^3 and total volatile fatty acids of AD = 1750–3000 mg L^{-1})	• ELM: R^2 (0.987); MAE (55.18) • GA-ELM: R^2 (0.990); MAE (47.71)	• The best ELM model had a good prediction for validation data (R^2 = 0.972). • Prediction error of 2.15 %.	Zhang et al. (2022)
3	Linear regression (LR), RF, ANN, SVR, and XGBoost	Biogas generation	Feed amount, digester properties, and feedstock properties: TS, chemical oxygen demand (COD))	• The RF and XGBoost models excelled on the training set with high R^2 values (0.94 and 0.90) and low RMSE values (0.26 and 0.26). • The RF and XGBoost on the test set with higher R^2 values (0.61-0.62) and lower RMSE values (0.62-0.64).	• The average daily biogas yield was 14,508 m^3, displaying a significant variation ranging from 3,077 m^3 to 21,406 m^3. • CH_4 concentration exhibited a relatively stable trend within a narrow range of 50% to 62%.	Li et al. (2022)

1.5.7 TEA of Compost Facilities

1.5.7.1 Site Selection with Geographic Information Systems (GIS)

The suitability and optimality of a waste to compost conversion facility depends upon several factors such as environmental, social, and economic factors. For this purpose, the multi-criteria decision analysis (MCDA) based approach is promising due to the consideration of several mentioned factors. Accordingly, values are assigned to determine the corresponding relative weighted values. Among many alternate schemes, analytic hierarchy process (AHP) is a widely accepted MCDA method that applies a pairwise comparison of multiple criteria and a multi-level hierarchical structure to obtain the relative weight of each considered criterion. Further, the integration of GIS features and AHP techniques (a GIS-based MCDA approach) enhances the customization by utilizing spatial data and relative weighted criteria. Accordingly, more valuable spatial data (i.e., spatial data containing more information that can be further used) can be proposed to ascertain upon the critical decision-making process. Considering this promising aspect, the GIS-based MCDA approach was investigated in several studies for landfill siting (Nas et al., 2010; Sener et al., 2006; Sumathi et al., 2008; Wang et al., 2009; Sultana and Kumar, 2012). However, limited research was devoted to the GIS-based siting of solid waste conversion facilities (Aragones-Beltran et al., 2010; Tavares et al., 2011; Khan et al., 2016). Aragones-Beltran et al. (2010) applied the analytic network process (ANP) and considered the twenty-one economic, legal, and environmental criteria and as well grouped them into clusters for MSW plant site selection. Chiueh et al. (2008) and Chang et al. (2009) proposed GIS-based systems for the allocation of compensatory funds for existing solid waste incinerators. Also, the research as well evaluated the environmental impact assessments. Also, Moeinaddini et al. (2010) presented a siting methodology by incorporating GIS with the AHP for a solid waste incineration plant. Thus, a methodology that utilizes existing data is to be addressed for the affirming of a sustainable waste management infrastructure in a particular region. Table 1.11 presents a summary of the GIS-based methodologies for landfill sites with the GIS-based tools.

Table 1.11 A summary of the findings associated to the GIS-based tools for MSWM.

S. No.	GIS based tools	Salient features	Critical findings	Authors
1	GIS and MCDA	• Offers a range of methods and processes for organizing decision-related issues. • Enables the integration of geographic data and subjective assessments to generate decision-making information. • The approach provides a wide range of methodologies to structure decision problems, design and evaluate alternative options, and prioritize decisions.	• A combined approach using multi-criteria decision analysis and inexact mixed integer linear programming for solid waste management. • From an initial pool of 17 potential sites, further analysis and refinement based on specific micro-level factors resulted in the identification of the top 3 optimal locations for landfill construction. • Based on area availability, the top-ranked sites for landfill construction were determined. Sites 1, 5, and 13, covering areas of 0.36 km^2, 0.11 km^2, and 0.06 km^2, respectively, were selected as the most suitable options.	Sumathi et al. (2008)
2	GIS and Weighted Linear Combination (WLC)	• The use of this technique involves employing GIS to generate composite maps based on a decision rule. • In the decision-making process, this technique provides greater flexibility compared to the Boolean approaches. • The approach produces a continuous suitability image, posing challenges for selecting specific landfill sites or making other allocations.	• Utilizing the WLC and AHP methodology within a GIS environment for siting MSW landfills. • The factors considered for assessing technical suitability of an area include land-specific technical aspects, distance from the electrical grid, waste feed rate, proximity to waste generation centers, population growth, distance to secondary markets, and distance to waste sorting centers.	Moeinaddini et al. (2010)
3	GIS with AHP	• It provides an objective approach by identifying and assigning weightage to each criterion. • It enables decision makers to conduct decision analysis functions, including ranking alternatives to determine the optimal choice. • It aids in the identification and prioritization of selection criteria, analysis of collected data, and streamlining the decision-making process by utilizing a combination of empirical data and the subjective judgment.	• Combining AHP with GIS for landfill site selection. • The data within the study area were categorized into four suitability classes: high, moderate, low, and very low. • These classes represented 3.24%, 7.55%, 12.70%, and 2.81% of the study area, respectively. • A total of 73.70% of the study area was identified as completely unsuitable for a landfill site.	Şener et al. (2010)

1.5.7.2 Sizing and Costing based TEA

In this field of study, the literature conveys that many studies were devoted to solid waste utilization techniques and some of these elucidated upon the energy and economic assessment of specific technologies (Bonk et al., 2015; Emery et al., 2007). The techno-economic analysis (TEA) of the composting process involved a comprehensive approach that combines technical, environmental, and economic factors for the evaluation of the feasibility of composting process. TEA provides insights into the costs and benefits of alternate composting methods. Accordingly, it can assist to identify the most cost-effective and environmentally sustainable option for a given situation. Wang et al. (2022), considered the TEA of four alternate composting methods (Solar-assisted (SAC), Bio-enhanced (BEC) and Heat-dewatering composting (HDC)). The study found that the solar-assisted composting method had the highest economic feasibility due to its promising features such as high compost quality and shorter composting time. Samarasinghe and Wijayatunga, (2022) evaluated the economic feasibility of organic municipal waste for the production of compost. The study inferred that for a 10.25-year period, the net present value (NPV) was $ 48,736,517 and was economically feasible.

Another study by Yoshizaki et al. (2012) detailed upon the TEA of composting food waste with a continuous flow system. The study concluded that for a 10-year tenure, the composting was economically feasible with a NPV of 4.4 million Malaysian ringgit. The study also inferred that the compost could be produced with high nutrient content and was suitable for its utility as a fertilizer. Lin et al. (2019) evaluated the techno-economic feasibility of the co-composting of yard trimmings. The co-composting yielded a $1.8 million NPV over 5 years, and also reduced landfill waste and produced high-quality compost for soil amendment application. A customized techno-economic calculator for compost for is crucial for comprehensive financial analysis, cost evaluation (infrastructure, equipment, labour, energy consumption, raw materials, maintenance), and strategic planning, investment decisions, risk assessment, and policy development in composting. However, there is a lack of research on the development of a techno-economic calculator being customized for varied scale composting facilities (large, medium to small scale). A brief summary of the prior art has been presented in Table 1.12.

Table 1.12 A summary of the profitability studies of the compost facilities.

Economic and profitability analysis												
S. No.	**Facility type(s)**	**Feedstock**	**Capacity**	**Equipment cost ($)**	**Annual output & selling price (SP) of organic fertilizer**	**Capital Cost ($ ton^{-1})**	**Operational Cost ($ ton^{-1})**	**PBP (yr)**	**Net Present Value (NPV) ($)**	**Internal rate of return (IRR) (%)**	**Critical findings**	**Authors**
1	Solar-assisted (SAC) Bio-enhanced (BEC) Heat-dewatering composting (HDC)	Kitchen waste	1 ton day^{-1} 1 ton day^{-1} 1 ton day^{-1}	46410 61880 61880	109.5 tons annum^{-1} 109.5 tons annum^{-1} 90 tons annum^{-1} and SP: 77.35 $ ton^{-1}(each)	83538.00 82764.50 100725.17	32.68 42.75 79.17	4.44 4.72 16.24	32133.11 25035.93 −62606.90	18.36 16.65 −7.97	• If the total fixed capital investment is reduced by 20%, the corresponding decrease in NPV would be 44% for SAC, 56% for BEC, and 32% for HDC. • Net profit $ annum^{-1}: 18824.94 17544.02 6203.57 • The total operating cost for SAC was estimated to be 24% lower compared to BEC and 59% lower compared to HDC and 59 % lower than that for HDC. • Income from the sale of organic fertilizer was 24.80 $ for SAC; 32.83 $ for BEC and 22.63 $ for HDC.	Wang et al. (2022)
2	Large-scale commercial composting facility (modelling)	Yard trimmings	5700 MT yr^{-1}	The equipment purchase costs were estimated using the Built-in Cost Model in SuperPro Designer software	--	$84 MT^{-1}	$21 MT^{-1}	4.9	1.8 million	33	• Annual net profit was 342,000 $ yr^{-1} • Net production cost 31.55 $ MT^{-1} • Net return on investment 20.31%	Lin et al. (2019)
3	Windrow composting	Municipal OW (kitchen waste, wood, grass, paper)	550	7,82,250	Compost sales 25 $ ton^{-1}	32,312,500	1,718,750	10.25	48,736,517	15	• Weighted average cost of capital was 3.7 % • Total investment 34,375,000 $ • Nitrogen fertilizers by compost: 326,133 $ yr^{-1}	Samarasinghe and Wijayatunga, (2022)

1.5.7.3 ML Algorithms-based Prediction of Compost Generation Rate

Several prior arts targeted the modelling of organic composting processes. Mostly, they involved experimental verification (Kishimoto et al., 1987; Kaiser, 1996; Hall, 1998; Lin et al., 2008). Accordingly, the models consider a multitude of independent factors, such as T, MC, total bulk density, oxygen concentration, CO_2 content, energy flow, and substrate properties in the modelling efforts (Hall, 1998; Hamelers, 2004). In addition, sub-processes of the composting process, including respiratory activities, heat transfer, and process kinetics, have been considered in the modelling efforts (Das and Keener, 1997; Haug, 1993; Mason and Milke, 2005).

Mathematical formulations involving the quantification of biochemical and physical processes in composting process have been also addressed for the prediction of composting dynamics. The inclusion of theoretical rather than empirical stoichiometric coefficients ensured greater universality of the modelling efforts in conjunction with the previous works (Kaiser, 1996; Lin et al., 2008; Sole-Mauri et al., 2007). Another research group reported empirical models for the compost generation rate or compost maturity. These evaluations were carried out in terms of T, MC, cumulative energy, and airflow rate (Kishimoto et al., 1987).

The possible scope for ML techniques to predict and forecast compost generation rate needs to be targeted. This is due to the CP rate being an important parameter in the assessment of the techno-economic viability of composting processes. Efforts till date have limitations in their ability to predict compost generation rate. In ML methods, techniques such as ANN, LR, *k*NN, SVR, tree-based ML approaches (Moncks et al., 2022; Ding et al., 2022) and convolutional neural networks (CNNs) were reported to predict CP and volume (Kujawa et al., 2020; Gao et al., 2007; Xue et al., 2019). However, the utility of ML in composting is still in its early stages. Hence, further research needs to be devoted to develop accurate and practical models that can be used to improve the efficiency and effectiveness of alternate composting processes. Needless to analyze, data sharing and integration will be necessary to realize the greater potential of ML in waste management and thereby transform waste treatment into highly intelligent operational and management activities. The available prior art in the field of ML-based prediction of CP rate has been summarized in Table 1.13.

Table 1.13 A summary of the studies targeting ML based prediction of composting process performance.

S. No.	Feedstock	Independent parameters	Prediction	ML models	Performance	Key findings	References
1	Sewage sludge & agro-industrial waste	Time, T, MC, T_{atm}, RH, moisture from Gravimetry	CP	*k*NN, LR, MLP	For *k*NN: R^2=98.79 MLP: R^2=96.18 and LR: R^2=98.79	• The standard deviation and standard error for *k*NN model was respectively 11.36 % and 6.56 % showing best results. • Mean absolute errors for MLP: 2.594 and RMSE: 3.0288. • The trial-and-error method is employed for hyperparameter tuning in various ML models. • No model optimization.	Moncks et al. (2022)
2	Kitchen waste	Time, T, pH, MC, C:N, ammonia, nitrogen, organic matter, nitrate, TOC, GI	Composting maturity	LR, *k*NN, DT, SVR, RF	LR: R^2= 0.73; RMSE= 9.41; MAE= 7.06 *k*NN: R^2= 0.87; RMSE= 6.57; MAE= 4.33 DT: R^2= 0.76; RMSE= 8.89; MAE= 5.16 SVR: R^2= 0.69; RMSE= 10.21; MAE= 7.09 RF: R^2= 0.85; RMSE=7.09 MAE= 3.94	• Mass ratio of mixing kitchen waste and wood chips was 9:1 • The optimal range of T, MC, and pH was 30-45°C, 55–65% and 6.3–8.0 respectively with C:N = 25. • After composting, T: 0.42 ± 0.02 and a GI value: 104.61 ± 4.84 %. • No model optimization.	Ding et al. (2022)
3	Sewage sludge & rapeseed straw	Compost samples (.jpeg images)	Early-stage compost maturity	CNN	Classification error=0.51% (CNN 16MIX)	• Validation set: 1.02; Test set: 0.51 Epoch to get best network: 145 • For CNN16 MIX correctly classified cases: 196. • For CNN16 MIX the value of area under the curves is 0.999	Kujawa et al. (2020)

1.6 Research gaps

1.6.1 Sampling Strategies for Enhanced Precision in Waste Composition Estimates

The relevant prior art refers to the contributions of Ojeda-Benitez et al. (2008), Al-Khatib et al. (2010), Ogwueleka, (2013), Zhang et al. (2018), Phuong et al. (2021), Gu et al. (2015), Gómez et al. (2009), Thanh et al. (2010). Among these, the subjective depth was well addressed by only few authors. Phuong et al., 2021 conducted a survey on HH solid waste generation and physical composition. The study sampled 110 HH in specific areas with different population density and HH scale. In another study by Zhang et al., 2018, a HH level survey (418 HH) was conducted by targeting income and HH sizes. Adopting a questionnaire-based study, the authors quantitatively analyzed the amount, composition and environmental impact of generated MSW. However, the authors did not implement the sampling stratification process in the conducted waste characterization studies. To create an effective decentralized waste collection and management system for any specific study area, an area-wise characterization of HH waste (HHW) is necessary along with effective sampling schemes.

Also, a holistic understanding of the influence of seasonal variation on the solid waste quantification and characterization shall be also targeted. This is due to the reason that the seasonal fluctuations in waste generation and composition do critically influence the optimality of waste collection schedules, resource allocation, and implementation of targeted waste reduction measures. Very limited integration of seasonal variation in waste management strategies were addressed in the prior art. Gu et al., 2015 elucidated on the seasonal patterns of waste generation and proposed strategies for waste management that account for these variations. However, the authors considered only HH size, income and EDU level as the critical factors affecting the MSW characters and their sensitive variations. The role of factors such as HH types, number of HH, age etc., were not studied in the prior art. Thus, based on available limited prior art, seasonal dynamics with effective sampling strategies along with the consideration of above-mentioned factors shall be considered for the survey-based quantification of the MSW characters in the Guwahati city. This shall be regarded as a foremost objective of this Ph.D. thesis.

1.6.2 Sensitivity Analysis of Key Socio-economic Factors for Total Waste Generation Rate Prediction

In this research theme, the pertinent prior art refers to the works of Ghanbari et al. (2021), Grazhdani, (2016), Sankoh et al. (2012), Kontokosta et al. (2018), Ghinea et al. 2016, Debrah et al. (2021), Suthar and Singh, (2015), Khalil et al. (2022), Miezah et al. (2015); Sankoh et al. (2012), Johnson et al. 2017, Daskalopoulos et al. (1998), Abbas et al. (2022), Zhu & Rahman, (2020), Nguyen et al. (2020), Kolekar et al. (2016), Ma et al. (2020), Namlis and Komilis, (2019). An extensive prior art in the chosen field highlighted few socio-economic factors as the most significant contributors to the MSWG rate (Ghinea et al. 2016; Johnson et al. 2017). A study conducted by Nguyen et al. (2020) in Taiwan, inferred that population growth, supply-demand conditions, and lifestyle were significant to serve as waste generation predictors. In a study by Namlis and Komilis (2019), socioeconomic factors such as income, EDU, and family size were analyzed for their critical influence on waste generation rates. The authors observed that both income and EDU had a positive correlation with waste generation rates. Further, few prior art infer that socio-economic factors such as age (Debrah et al. 2021) and GDP (Grazhdani, 2016) did potentially influence the MSWG rate. Despite conducting a detailed case study on administrative levels in the mentioned prior art, gross district domestic product (GDDP), working population, literate population were not considered as important parameters to infer upon their criticality towards the prediction of solid waste generation rate. In summary, the relevance of above-mentioned socio-economic indicators in the prediction of MSWG rate shall be considered as the second objective of the Ph.D. Thesis.

1.6.3 K-fold Cross Validation and Hybrid ML Modelling-based Prediction and Forecasting of Solid Waste Generation Rate

The relevant literature refers to the works of Molina-Gómez et al. (2021), Chung (2010) Abbasi and El Hanandeh (2016) Kumar et al. (2018) Ayeleru et al. (2021), Kontokosta et al. (2018), Abbasi et al. (2019) as notable contributions. Targeting the prediction of total solid waste generation using ML algorithms, Kannangara et al. (2018) predicted annual MSW generation using decision tree and ANN in Ontario, Canada. For the purpose, the authors used feature selection techniques. However, the model suffered with low accuracy. Also, while the models performed exceptionally well for the training data,

they failed to generalize themselves to unseen data and affirmed poor predictive performance. Similarly, Kumar et al. (2018); Kontokosta et al. (2018); and Abbasi et al. (2019) compared the performance of alternate ML models at the city level and evaluated the accuracy of predictions. However, the accuracy of the considered ML models suffered with overfitting issues. Such overfitting issues needs to be reduced so that the models can better capture the true underlying patterns, improve their generalization capabilities, and provide more accurate predictions for new and unseen data. Further, Ayeleru et al. (2021) forecasted long term (30 years) MSWG rate with ANN and support vector machine (SVM) methods. Despite ensuring higher model accuracy, the models could not capture the complex relationships between various predictors and solid waste generation rate. In other words, a useful and detailed methodology has not been apparent in the systematic evaluation of the accuracy of the ML models.

To refine the predictive performance and the final forecast, fine tuning of the models with K-cross validation efficacy of best ML prediction model with the time-series component of ARIMA models needs to be addressed. Time series analysis (moving average; MA, and ARIMA) is a superior approach for forecasting as it captures temporal patterns, handles time-dependent relationships, incorporates lagged variables, and accounts for noise and randomness. By considering the sequential nature of data, time series models identify and model patterns and trends that occur over time. This results in more accurate predictions. Thus, the models can effectively handle dependencies, such as autocorrelation, and incorporate past values of the variable being forecasted to account for historical behaviour and patterns. Time series analysis also excels in filtering out noise and capturing the underlying signal. Diagnostic tools facilitate model performance, assessment adapt themselves allowing for updates with new data, and forecast uncertainty estimates. All these aid in better decision-making and risk assessment. Also, to improve upon long-term accuracy, hybrid modelling (coupled with ML and ARIMA models) shall be targeted. Thereby, improved accuracy can be achieved through the integration of ML's ability to capture nonlinear relationships and proficiency of ARIMA model in modelling time-dependent dynamics. The hybrid model's flexibility allows for the inclusion of additional predictors, and thereby enhances forecasting accuracy. It handles data limitations, such as missing values and

outliers, and with scalability and automation, it is relevant for large-scale applications. Overall, hybrid ML-ARIMA modelling approach offers a scientifically sound approach for the forecasting of accurate MSWG rate, supported informed decision-making and resource planning. Thus, various ML-based tools for improved model accuracy using cross validation techniques shall be considered to evaluate the MSWG dynamics for classified ward wise data representation. To elucidate on the critical aspects of organic and recyclables waste generation rates, classified ward wise data representation will be more effective. Also, in mentioned prior art, the authors primarily ignored the forecasting method for long-term MSWG forecasting. Thus, considering the long-term forecasting with time series analysis shall be the third objective of the Ph.D. thesis.

1.6.4 Robust and Accurate Prediction of GHG and PMs Emissions from Landfill and Incineration Sites based on Large Dataset and Additional Meteorological Parameters

The research theme was addressed by Singh et al. (2022), Dhakal et al., 2010), Bunsan et al. (2013), Abushammala et al. (2014), Fallah et al. (2020), Mandal et al. (2020). Among these, subjective depth was targeted by Fallah et al. (2020), Abushammala et al. (2014) and Mandal et al. (2020). Fallah et al. (2020) deployed an ANN model to predict CH_4 fluxes, using input variables such as T, RH, Pr, and WS. However, to enhance the robustness of the model, additional meteorological parameters, such as atmospheric pressure (P_{atm}), wind direction shall be considered as they can provide valuable insights into the atmospheric conditions that influence GHG emissions from solid waste landfill and incineration processes. Accordingly, the model can better capture the complex interactions between waste management practices and meteorological conditions, and can perform with greater accuracy and reliability. In a study by Abushammala et al. (2014), the ANN model achieved a strong correlation between the predicted and actual CH_4 emissions. Despite exhibiting good performance accuracy for CH_4 emissions prediction, due to limited data size, the model suffered high variance and overfitting issues. Mandal et al. (2020) retrospectively assessed daily average $PM_{2.5}$ exposure in Delhi, India, from 2010 to 2016. Multiple data sources and ensembled averaging approaches were used by the authors. The study deployed a multi-stage modelling exercise and accordingly incorporated satellite data, land use variables, and population density. To counter the limited ground monitoring data, a calibration

regression was used to model the $PM_{2.5}$: PM_{10} relationship. Six different learners, including generalized additive models, elastic net, SVR, RF, ANN, and XGBoost, were deployed to model the spatiotemporal patterns of $PM_{2.5}$. However, the conducted study lacked the integration of auxiliary data. Thus, the ML models can be better utilized through the incorporation of additional data sources, such as meteorological data, landfill operation parameters, and waste characteristics, and population statistics of the study area. Thereby, improved accuracy for $PM_{2.5}$ and PM_{10} can be realized.

Also, using CO_2 content, and meteorological parameters such as T_{atm}, P_{atm} and cumulated rain, Scozzari (2008) predicted the superficial gas flux (the measurement or estimation of GHG emissions released into the atmosphere per unit area or time) from the landfill site using ANN. However, the modelling was based on a small dataset. Thus, realizing accurate and extensive datasets for training and ML model validation remains a challenge due to smaller datasets. Based on the limited prior art available on landfill emissions (CH_4, CO_2, N_2O, SO_2, $PM_{2.5}$ and PM_{10}) and no prior art in the field of emissions from incineration of solid waste, an accurate estimation of emissions with ML-based models shall be investigated. Thus, with a larger dataset and meteorological parameters and socio-economic and demographic factors, a comprehensive quantification of the carbon footprint of landfill and incineration processes shall be targeted through modelling studies. Additionally, ML-based models can play a crucial role in supporting the monitoring efforts of GHG emissions. Thus, the prediction and forecasting of GHG and particulate matters (PMs) from landfill and incineration sites of the Guwahati city with ML and hybrid ML methods shall be addressed as the fourth objective of the Ph.D. thesis.

1.6.5 Meteorological Parameters and OW Volume-based Prediction of Compost Generation Rate with ML Algorithms

In this domain of study, the relevant prior art primarily refers to the works of Gao et al. (2007), Kujawa et al. (2014) and Kujawa et al. (2020). For the purpose, Gao et al. (2007) utilized WNN that considered input features such as T_{max} duration, MC, volatile solids, the value of faecal bacteria and GI for the compost maturity prediction. Also, Kujawa et al. (2014) found a novel route to predict the compost maturity (based on colour and texture features) with a CNN. However, the authors did not emphasize upon the influence of the external parameters (meteorological) and organic content on the composting

process and the maturity of compost. In summary, monitoring these parameters and making necessary adjustments based on the volume of OW can profoundly improve the realization of compost maturity and associated financial dividends. Based on no available prior art in this field of research, the compost generation rate modelling with ML models and with mentioned parameters shall be considered as the fifth objective of the Ph.D. thesis.

1.6.6 Sensitive influence of meteorological parameters and OW volume on the ML algorithms-based prediction of biogas generation rate

In the field of study, the research works of Ugwu and Enweremadu (2019) Wannasek et al. (2017) and Wang et al. (2022) targeted the criticality of meteorological parameters on BP rate. However, the authors elucidated only upon the influence of optimal T, RH and P on the biogas generation rate. Thus, the authors did not delineate upon other important meteorological parameters such as Pr and, volume of OW. In summary, the influence of meteorological parameters and the volume of OW as additional parameters on the production of biogas during anaerobic digestion for the ML based prediction of biogas generation rate shall be considered as the sixth objective of the Ph.D. Thesis.

1.6.7 Economic Feasibility of Alternate Biogas Digesters based on Techno-economic and ML Modelling Strategies

In the mentioned field of study, the articles contributed by Jones and Salter (2013), Anderson et al. (2013), Wu et al. (2016), Castano et al. (2014), Ihara et al. (2020), Nandi et al. (2020), Rahman et al. (2019) have been significant. Wu et al. (2016) reported a techno-economic calculator being developed to support operators in the day-to-day and long-term engagement for the United Kingdom's farm-based AD unit operation. The authors considered sensitivity of T, RT, and mixing effects but for only farm-based AD digester system. Also, the authors did not delineate upon the role of the feedstocks, plant capacity, and scalability. All these can have a sensitive influence on the economic viability and suitability of the digester type selection for a specific project. Also, the effect of seasonal variations on AD units for their efficient and reliable operation was not considered by the authors. Till date, many prior art investigations (Nandi et al., 2020; Rahman et al., 2019) targeted the effect of seasonal variations on AD units for laboratory scale systems. Thus, medium to large scale AD units were not

studied. In addition, little information is available in terms of the influence of ambient temperature on the T of the full-scale AD process. Castano et al. 2014 and Ihara et al. 2020 reported an improvement in BP rate through an increase in ambient temperature in a small-scale digester. Kalia et al. 1998 examined the performance of the dome biogas plant under seasonal variations and found that the declining temperatures from summer to winter translated into reduced production rates. While the AD based BP system at mesophilic and thermophilic conditions was well studied, its performance in psychrophilic and psychotropic temperatures (conditions prevalent in North-east India) were studied in a limited context.

Also, to improve the efficiency, sustainability and profitability of BP, the concept of ML needs to the further considered (De Clercq et al., 2019). Neto et al. (2021) investigated the relationship between food waste and BP rate and with an ANN model being trained with backpropagation. The input parameters for the model were pH, RT, and T, for the predicted BP rate parameter. Beltramo and Hitzmann, 2019, examined the substrate characteristics of corn straw and grass silage in relation to biogas flow rate. For this, the authors deployed the MLP model. However, the authors did not study the patterns and trends in data, and prediction accuracy and optimization of plant performances. Therefore, considering feedstock types, plant capacity, scalability, seasonal variation on AD units and application of ML algorithms, the seventh objective of the Ph.D. Thesis shall aim upon the enhanced ML model based predictive performance, and economic viability of BP.

1.6.8 Techno-economics and ML Algorithms-based Assessment of Alternate Composting Facilities

For the research theme, the relevant literature refers to Molina-Gómez et al. (2021), Chung (2010) Abbasi and El Hanandeh (2016) Kumar et al. (2018) Ayeleru et al. (2021), Kontokosta et al. (2018), Abbasi et al. (2019) as notable contributions. Dadhich et al. (2021) applied techno-economic assessment approach for food waste and yard waste silage as feedstocks and thereby evaluated the CP volumes. However, the authors did not consider relevant factors such as RT, T, MC, equipment and operational, compost price, and cost overhead etc. All these can profoundly affect the financial viability of a composting operation. Thus, the techno-economic strategies are important in scale variant composting

facilities (large, medium and yard scale compost facilities) as they provide a comprehensive understanding of the financial viability, resource optimization, process improvement, market dynamics, and risk mitigation. Also, Lin et al. (2019) conducted sensitivity analysis on the TEA of compost systems. However, the authors considered limited parameters such as feedstock costs, labour costs, compost price, RT for composting and tipping fee. Also, despite addressing subjective aspects in TEA, the cited literature lacks standardization of economic analysis and scalability assessment. Without standardization, it is difficult to compare the outcomes of alternate composting studies. Also, no study has been addressed till date for the solid waste to compost facility siting decision based on a detailed waste assessment procedure. Such studies need to consider waste availability, existing landfill locations, and an actual road network as important biases. Thus, research efforts need to address these and improve comparability and consistency across studies. While most studies involved small scale single composting facility investigations, the scalability of alternate composting methods can ensure upon their feasibility on a scalability basis. In summary, the techno-economic assessment on composting facilities, and the sensitivity analysis-based insights for the assessment of influencing factors on CP volume estimates shall be considered as the last objective of the Ph.D. thesis.

1.7 Objectives of the Ph.D. Thesis

In accordance to the summarized research gaps in section 1.5 of the chapter, the Ph.D. thesis targets the following objectives in the field of city wide ISWM of Guwahati city.

(i) Survey based quantification of the constitution of various entities in municipal solid waste (MSW) and subsequent seasonal classification of the waste data among socio-demographic groups (low, middle and high income) for the Guwahati City.

(ii) Modelling total solid waste generation rate prediction using ensembled tree-based ML approaches and forecasting with MA methods (simple, weighted and exponential) for the Guwahati City.

(iii) Secondary data based efficient prediction of organic and recyclables solid waste generation rate using supervised ML algorithms and ARIMA (time series modelling) approaches for the Guwahati City.

(iv) ML based prediction and forecasting of GHG emissions (CH_4, CO_2, N_2O, SO_2) and particulate matters ($PM_{2.5}$ and PM_{10}) from MSW landfill and incineration process sites of the Guwahati city.

(v) Meteorological parameters and OW data-based prediction and forecasting of CP rate (conventional composting of MSW) with ANN and traditional ML algorithms and hybrid ML modelling respectively for Guwahati city.

(vi) Meteorological parameters and OW data-based prediction and forecasting of BP rate (anaerobic digestion of OW) using ML and hybrid ML modelling approaches respectively for the Guwahati city.

(vii) Development of a techno-economic calculator for alternate design (floating drum, fixed dome and balloon digesters) based BP and subsequent optimal prediction and long-term forecasting of biogas generation rate in the Guwahati city.

(viii) Development of a techno-economic calculator for alternate scale (yard, medium and high scale compost facilities) based CP and subsequent optimal prediction and long-term forecasting of CP rate in the Guwahati city.

1.8 Organization of the Ph.D. Thesis

With the realization of the above-mentioned objectives, the Ph.D. thesis work has been organized in eight chapters. A brief account of the chapters is as follows:

Chapter 1 elucidates upon the need for research in the field of MSWM approaches to handle the complexity of the issues to be resolved in the context of a developing country. Particularly the nexus of eco-social systems, shall be explored for effective MSWM and decision-making process. The relevant studies were elaborated based on the available prior art. Thereafter, customized research gaps were identified to converge upon the objectives of the Ph.D. thesis. Finally, the organization of the thesis has been summarized.

Chapter 2 details upon the experimental and modelling approaches adopted for all investigations addressed in the thesis. These include (a) waste characterization and quantification along with seasonal

study for the income groups in Guwahati city, (b) MSWG rate (total, organic and recyclables) prediction and forecasting using ML-based and time series analysis (ARIMA and MA) approaches respectively in Guwahati city, (c) GHG emissions (CH_4, CO_2, N_2O, SO_2) and particulate matter ($PM_{2.5}$, PM_{10}) prediction and forecasting from MSW landfill and incineration sites in Guwahati city using ML and ARIMA techniques, (d) deployment of ML and hybrid ML models for the prediction and forecasting of CP and BP rate in Guwahati city, and with meteorological parameters and MSW data and (e) TEA studies for alternate biogas digester designs and compost facilities and process parametric sensitivity assessment towards prediction and long-term forecasting of biogas and compost generation rate in Guwahati city.

Chapter 3 details upon a survey-based approach being deployed to quantify the constitution of MSW in Guwahati City and classify the waste data across socio-demographic groups. Adopting a random stratified clustering sample, HH interviews, waste audits, and direct observations, data with respect to the composition of MSW, including OW, paper, plastic, glass, metals, and others was collected for the Guwahati city. The chapter also delineates upon the findings associated to comprehensive insights into the seasonal variation influence on the MSW composition.

Chapter 4 elucidates upon the sensitivity of demographic and socio-economic parameters for the fair prediction and forecasting of the total MSWG rate. ML models were formulated by mapping solid waste quantities at the municipal level and with socio-economic and demographic variables of Guwahati city. Tree-based ML algorithms, DT, RF and GB, were applied to build the models with a data size of 1936. The MA approaches were adapted for the forecasting of the MSWG rate. Performance indices such as MAE, R^2, RMSE were used for model validation. ML and MA approaches were deployed to model historical data (Census data, CPCB, Ministry of Statistics and Programme Implementation, MOSPI, GMC data) to effectively serve the dynamic needs of the solid waste industry.

Chapter 5 addresses the prediction and forecasting of MSWG (organic and recyclables) with five ML (MLP, genetic algorithm (GA), SVR, gradient boosting (GB) and *k*NN) and hybrid modelling (ARIMA-ML) approaches. Socio-demographic factors that influence waste generation have been considered. These refer to population, HH information, age, and EDU level. The methodology was targeted to create

useful tools for regional SWM and planning. This involved the integration of data from various public domain sources (GMC, Census of India). This was followed by pre-processing and modelling strategies to effectively meet the targeted objectives.

Chapter 6 focuses on the relevance of nine ML techniques to predict and forecast (ARIMA) GHG emissions (such as CH_4, CO_2, N_2O, SO_2) and particulate matters ($PM_{2.5}$ and PM_{10}) from the landfill sites and incineration processes at a solid waste dump site in the city of Guwahati. The emissions data were collected from the Emissions Database for Global Atmospheric Research (EDGAR), meteorological data from the Meteorological reanalysis (MERRA-2) and Iowa Environmental Mesonet (IEM). While EDGAR provides reliable information on GHG emissions, MERRA-2 offers global meteorological data since 1970 (by integrating satellite observations, surface measurements, and weather balloon data). Further, IEM provides meteorological measurements specific to Iowa.

Chapter 7 delineates upon the prediction and forecasting of compost maturity as a function of climatic parameters and OW content using ML models. The data size contained approximately 864 sample points records of the meteorological parameters (MERRA-2) (T, RH, Pr, WS), OW (CPCB) and compost yields (Open Government Data) data for the year range of 2010–2021. The modelling efforts involved the consideration of MLP and traditional ML algorithms (*k*NN, GB, RF) for the prediction of compost generation rate and ARIMA model supplemented ML models for the long-term forecasting of the compost generation rate.

Chapter 8 encompasses upon a novel methodology for the prediction and forecasting of BP rates with ML models. The biogas generation rate was modelled by considering meteorological parameters (T, P, RH, Pr) and OW content. It involved 728 sample points, and accounted for the records such as meteorological parameters, OW, and biogas yields (year range of 2015 to 2021). The modelling process targeted the application of MLP, GB, RF, *k*NN and SVR algorithms. Additionally, ARIMA and MA models were utilized to realize long-term forecasting of the BP rate.

Chapter 9 addresses the development of a techno-economic calculator for alternate biogas digester designs, namely floating drum, fixed dome, and balloon digesters. The calculator considers various

factors such as design parameters, input feedstock, and operational conditions for the evaluation of TEA of alternate designs. Furthermore, ML approach has been deployed to optimize the prediction of biogas generation rate for the Guwahati city. The long-term forecasting of BP rate was also performed by utilizing mentioned techniques. Overall, the mature approach enabled informed decision-making and supported the implementation of efficient and sustainable BP systems in Guwahati city.

Chapter 10 elaborates upon the TEA of the CP system at different scales (yard, medium, and high-scale compost facilities). The calculator considers various parameters such as facility size, input materials, and operational conditions to assess upon the techno-economic viability of each process scale. Additionally, ML models were utilized to optimize the prediction of CP rates for the Guwahati city. Long-term forecasting of CP rates was performed with advanced techniques. Thereby, accurate projections were achieved. Thus, the approach targeted the informed decision-making for the pragmatic implementation of efficient and sustainable composting systems in the Guwahati city.

Chapter 11 summarizes the most relevant conclusions from the studies conducted in the Ph.D. thesis. Thereafter, possible directions for future research have been elaborated to accelerate and realize pragmatic ISWM strategies.

Chapter 2
Materials and Methods

Chapter 2

Materials and Methods

In this chapter, a brief account of the adopted methods in the Ph.D. thesis has been presented. The adopted methodologies for various studies involved the following salient features of the Guwahati city as the study area: (a) predictive modelling using ML-based approaches for MSWG prediction, OW generation rate prediction, GHG emissions modelling, and compost and biogas generation rate minimization, (b) implement time series analysis and hybrid modelling for efficient long term forecasting of the best performing predictive models (c) realize efficient approaches for the optimization of model inputs and its efficacy and (e) finally TEA for biogas and compost facilities to evaluate economic viability, resource optimization, and risk analysis, aided informed decision-making and sustainable waste management. For the conducted studies, Python version 3.8 with NumPy, Scikit-learn, Pandas, and Matplotlib libraries for data analysis, modelling, and automated data processing on an i7-4790 CPU @ 3.60 GHz processor was deployed.

2.1 Study Area

Guwahati, one of the oldest urban centres of India, occupies the most dominant position in administration and trade in the entire north-east India. The city lies between 25°5′ N to 26°12′ N latitude and 91°34' E to 91°5′ E longitude and covers a large portion of the Kamrup Metropolitan district of Assam. The spatial expansion of Guwahati has been due to the huge population explosion over a short period of time. According to the census reports, its population growth increased from 43,615 in 1951 to 9,68,549 in 2011. The average density of the population growth of the city is as high as 4468 persons km^{-2} as per the 2011 census. The city presently has 31 municipal wards and recently included a new ward (Narengi) in the Kamrup Metropolitan district (Census 2011–19). The city centre, old city, business centres as well as the newly developed residential areas affirm highest population growth density. This is indicative of the growing compactness and complexities in the land use of the city with time (Bhattacharyya, 2001). The highly heterogeneous functional characteristics of the city led to the excessive population growth influx and rapid urban development. Similar to any other growing city, the lifestyle of the people is in a phase of transition due to rapid urban development. This resulted in higher waste generation from varied sources in the city. Accordingly, considering the local municipal ward-list record, the Guwahati city was divided

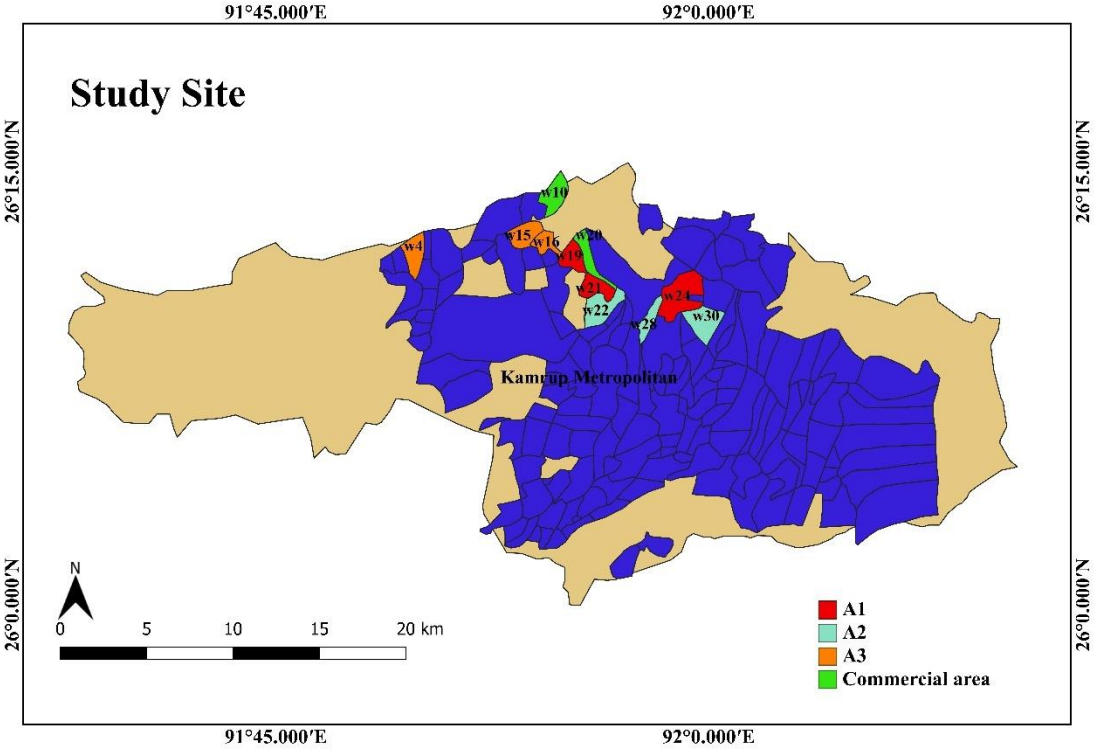


Fig. 2.1 Representation of A1, A2, A3 and Commercial sub-areas of Guwahati city

into three sub-areas (A1, A2, A3) and a commercial area. In total, 31 wards were selected to study the HHW characterization. The demographical representation of these areas is presented in Fig. 1.

2.2 Urban Solid Waste Characterization

2.2.1 Sampling Approach

Each ward in Guwahati city has a unique blend of people and classes. Hence, waste generation considerably varies among them. The city's geographically complex distribution of demographic parameters infuses additional complexity into the city's waste stream characteristics. For the conducted study, the city is divided into six zones namely Central (14, 15, 18 & 9); East (19, 20, 21, 22, 23 & 24) South (10, 11, 12 & 13); West (1, 2, 3, 4, 5, 7 & 8); Dispur (25, 26, 28, 29, 30 & 31) and Lokhra (6, 16, 17 & 27) and served as the strata for the study. These zones were chosen for initial convenience of the sampling and henceforth facilitated better collection of socio-economic and demographic data for the effective parse MSW characteristics. Accordingly, the study analyzed the role of commercial levels and socio-economic groups in solid waste generation in the month of February (2021), June (2022), and August (2022). Due to the large pool of HH in GMC (229716) (GMC, 2011) and associated logistical

constraints for sampling frame identification within each stratum, random stratified sampling (Neyman, 1992) was deployed.

2.2.2 Survey Summary

The targeted population growth of the study area constituted about 229716 HH in the Kamrup district in 2011 (GMC report, 2011). A survey was designed and administered with the officials in the GMC, Directorate of Economics and Statistics DES, ward-wise HH, and commercial areas. A simple and structured questionnaire was prepared and pre-tested (Appendix A). The questionnaire was targeted to collect information about residents' socio-economic characteristics, attitudes towards waste, waste management behaviours (disposal and waste separation), the quantitative ability to afford for collection services and the problems that exist in the current management system. For seasonal analysis, the surveys took place for about one week at the mass level in the city. The data was collected for the months of February (2021), June (2021) and August (2022). Each of these respectively represent winter, summer and the monsoon seasons. The six zones that served as the strata for this study were surveyed individually for the correct classification of each sector by HH income. The study inferred that each zone being represented by its wards, and clusters of HH were unique in the physical layout of HH and distribution of income. Accordingly, the survey was successfully conducted in 227 chosen HH and thereby enabled data collection with respect to the income, demographic and socioeconomic conditions.

2.2.3 Socio-demographic Categorization-based Population Growth Classification

As per 2011 records, the Guwahati city has a population growth of 957,351 inhabitants (Census of India, 2011) and 229,716 HHs (GMC Report, 2011). The secondary and tertiary sectors mostly account for the city's economic activity. According to DES, 2018, 33% of the population growth (main workers) were economically active, 6% were partially active (marginal workers) and 61% were non-workers. Such a classification of the active sector of the population growth was based on the number of minimum wage earners, and subsequent identification of several minimum wage ranges. Accordingly, the study classified its subjects based on their income. In due course of time, Guwahati gradually evolved from an agricultural society to a post-industrial society. Rapid industrialization and urbanization were

responsible for such a trend. The city's GDDP per capita increased from ₹109,439 ($1,345 USD) in 2001 to ₹ 237,212 ($2920 USD) in 2011

Table 2.1 A summary of the socio-economic indicators for the Guwahati city.

	2001-2011
Population (lakhs) [(1)]	957,351
Yearly population growth rate (%) [(1)]	1.73
HH information [(1)]	
- HH number	229716
- Good	108918
- Livable	84940
- Dilapidated	35858
Working Population (WP) [(1)]	
- Main workers	327058
- Marginal workers	44001
- Non-workers	586292
GDDP per capita (₹) [(2)]	₹237,212 (US$2,920)
GDDP growth rate (%) [(2, 3)]	5.3
GDDP by sector (%) [(2, 3)]	
- Agriculture	32.6
- Industry	19.8
- Services	47.6

(1) Census of India, 2011
(2) Ministry of Statistics and Planning Implementation, 2018
(3) Directorate of Economics and Statistics, 2015

(Table 2.1). Thus, the GDDP contributed by the agriculture-related sectors rapidly decreased from 48.2% in 2001 to 32.6% in 2011, and the decadal GDDP growth rate in the industry sector was about 19%. Such a structural change of the economic system in the Guwahati city significantly altered consumer behaviour as well as their garbage disposal patterns. Accommodating such alterations, the Government of Assam, India initiated various MSW management practices in the past few years (Singhal et al., 2022). Moreover, the HH information and WP also affirmed that the middle-income group constituted the majority case.

2.2.4 Data Collection

Following the city's settlement plan and demographical data analyses, the survey areas and HH waste distribution were identified for the city. As per the census data (Census of India 2011), the population growth of these wards varies from 21,462 to 39,116. The data collected from the GMC, Assam was in a detailed format. This referred to structure of the HH sizes, the income level, waste generation rate. To assess upon the socio-economic status of the HH, the information with respect to annual income of the HH, house or building structure, locality of colony, available facilities in house, type of vehicles in

house, other luxury facilities in houses, etc. was also collected for the ward numbered as 4, 15, 16, 19, 21, 22, 24, 28 and 30 (Table 2.2). For the commercial area, 10 and 20 ward numbers were chosen. The collected

Table 2.2 Notation, Socio-economic levels, household information and MSW amounts data of the sampled wards of Guwahati city.

Notation	Ward Distribution	Household counts	Population growth	MSW, tonsday^{-1}
A1 (Middle income)	w_{19}, w_{21}, w_{24}	27433	110891	139.42
A2 (High income)	w_{22}, w_{28}, w_{30}	23275	95706	181.73
A3 (Low income)	w_{4}, w_{15}, w_{16}	20154	86591	124.36
Commercial area	w_{10}, w_{20},	-	-	142.41

datasets were eventually classified into three economic classes namely low-income, middle-income and high-income groups. Such a grouping of HHs in the six zones was necessary as the waste generation characteristics were anticipated to alter significantly among different classes of people. Accordingly. the study targeted to correlate the alterations in waste generation due to altered economic strata. Few photographs of the main waste fractions being realized during the waste fraction classification process have been shown in Fig. 2.2.

2.2.5 Characterization and Waste Fractions Quantification

Based on its composition, the MSW was classified into six major fractions. The sorting was based on the fractions being found very often. Such a classification was also reported in a previous study (Singhal et al., 2022). After sorting, each waste fractions were weighed with a calibrated digital scale.

In summary. the garbage was divided into following main categories:

- Kitchen waste/food waste – peeling waste, discarded vegetables, food waste, discarded food, seeds, etc.
- Paper – paper scrapes, packing papers, discarded papers from students' bags, etc.
- Plastic and polythene bags – plastic goods, polythene and other items made of primarily plastic.
- Glass and ceramic scrap – scrape of glass, bottles, glass containers, broken kitchen items made of glass and ceramics, etc.
- Textiles and rubbers – fabric, diaper, shoes, slippers, pillow, carpet, rug, bag
- Cardboards – non-recyclable paper, cardboards, cartons, etc.
- Miscellaneous – metallic items, can, rubber, textile, leather, jars of metal, soiled paper, wood, saw dust, leaf litter, garden pruning, dirt and other inert material.

After screening, the garbage was disposed safely to the waste collection facilities for further disposal of the garbage.

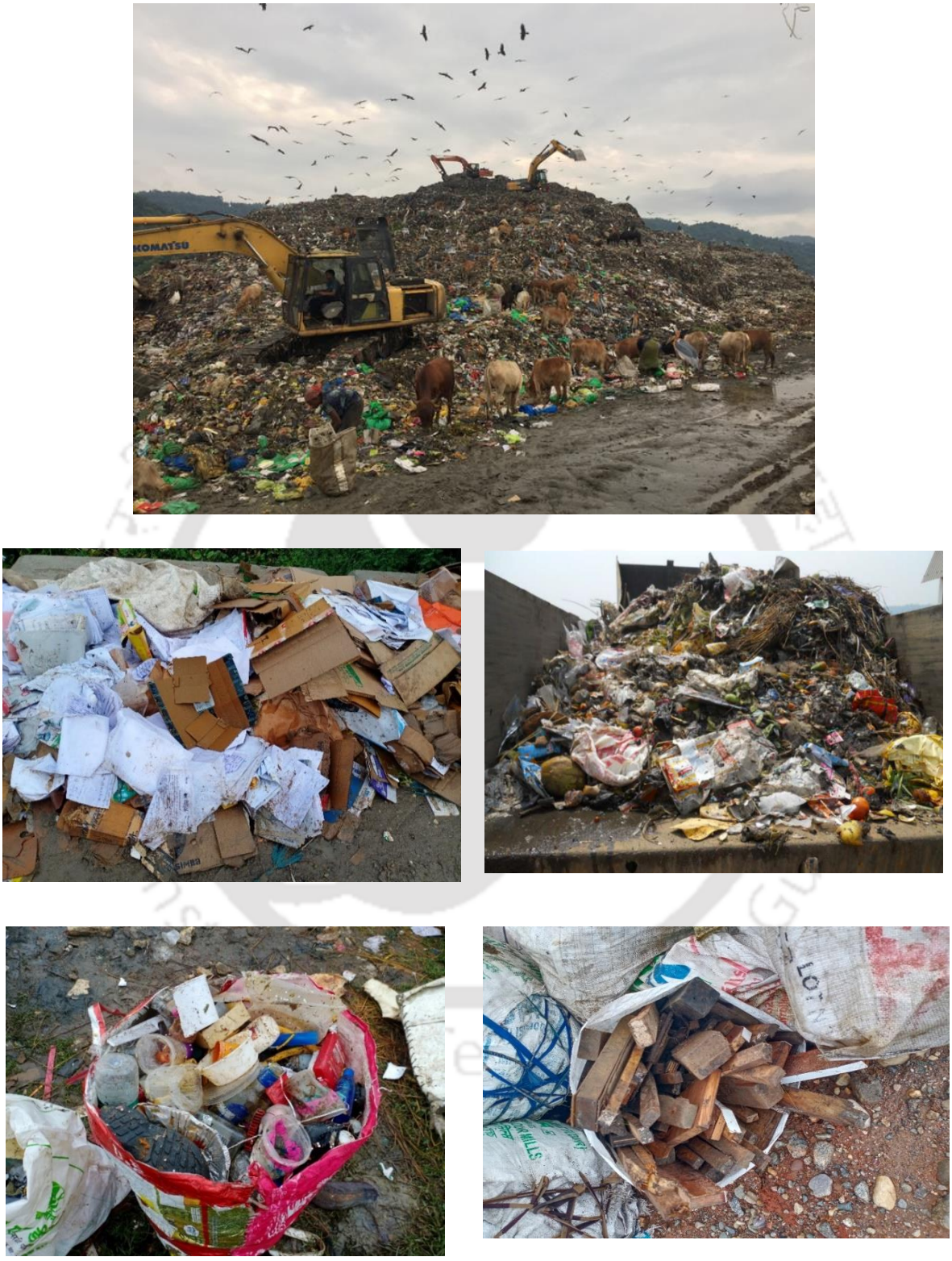


Fig. 2.2 Photographs of the waste fractions achieved during solid waste characterization studies of the Guwahati city

2.2.6 Analysis of the Findings

The analysis of variability (ANOVA) technique was adopted to infer upon the significant variations in the rates of waste generation among various HH and socioeconomic levels (including commercial area). The data was subjected to descriptive statistical analysis to eventually obtain useful statistical parameters such as median, standard deviation, variance, etc. For data analysis, the SPSS (Window Version 21.0) was then implemented. The data in the study had an acceptable p value (<=0.05). Using simple statistical analysis, the relationship between HH numbers, HH waste generation rates of various socioeconomic groups, and alternate seasonal cases was determined.

The altered social, economic and demographic factors may get accompanied with one other. Thereby, they may cause multi-collinearity problems during data analysis. Henceforth, the correlation analysis was conducted as a first step to ignore highly correlated factors and associated problems of collinearity. Especially, whenever the absolute term's Pearson correlation coefficient (*r*) was more than or equal to 0.7, one of the components was deleted. Otherwise, both were considered in the sample and were treated as independent terms. Accordingly, the preserved variables were subjected to the principal component analysis (PCA).

PCA can reduce the dimensionality of the original data, and henceforth serves as an efficient multivariate statistical technique (Dunteman, 2008). In the PCA, coordination systems are rotated to realize that the top two components account for the majority of variance. Since they contribute to the lion share of the total variance, only a few components should be considered for interpretation. Further, a dataset's socioeconomic variables do get diminished during PCA analyses (Vicente and Reis, 2008). In order to evaluate the annual alterations of the MSW quantities, cluster analysis was carried out (Zhou et al., 2015). Also, a correlation between the assessed water quality's temporal and spatial variation was reported in few research investigations that adopted correlation and PCA analysis (Gomez et al., 2008; Hatik and Gatina, 2017). The methodology converted associated variables into uncorrelated principal components, and thereby allowed the most variance among features to be saved. The main benefit of the PCA is that it improves the results through a reduction in the number of correlated features (too many can cause overfitting). Thereby, data labelling was not required. Being a method for the reduction

of unsupervised dimensionality, the PCA involves a smaller number of correlated features. Thus, both computational speed and visualization can be improved.

The exploratory factor demonstrated the relationship between variables and utilised statistical method. Thus, a stronger connection was affirmed by a higher number of components. To validate the output of the PCA, the Kaiser-Meyer-Olkin (KMO) and Bartlett's test values were computed. Accordingly, the KMO value shall exceed 0.5, and the p-value of Bartlett's test shall be less than 0.05 for the PCA result to be acceptable (Kaiser, 1974). Additionally, principal components with eigenvalues greater than one were selected. Accordingly, variables were identified in cases where cross-loaded variables were present to ensure that the least absolute loading of a principal component was 0.75 for each case (Ngai et al., 2007). To confirm upon the acceptability of the PCA results, a measure of the sample adequacy (MSA) and communalities, was also considered. These represented the sum of squared factor loadings. For both criteria, it was important to confirm that they were within an acceptable range (values greater than 0.5). To conduct the analysis, the PCA method with Varimax rotation and Kaiser normalization was utilized in the IBM SPSS Statistics version 19 software.

An RDA, or redundancy analysis, have been deployed in this study. RDA can be thought of as a limited variation of PCA. Rao presented the approach in 1964 (Rao, 1964), and Van Den Wollenberg developed it in 1977 (van den Wollenberg, 1977). RDA is frequently used in environmental and ecological sectors. RDA utilizes an eigen-analysis equation to determine the relationships between response and explanatory variables (Huang et al., 2015; Chen et al., 2017; You et al., 2018; Hira et al., 2018; Chuang et al., 2019). The equation for RDA can be expressed as follows:

$$(C_{RE} C_{EE}^{-1} C_{RE}^{'} - \lambda_k . I).u_k = 0 \qquad \textbf{(2.1)}$$

where C_{RE} is the covariance matrix among the response variables R and explanatory variable E; C_{EE}^{-1} is the inverse variance of variance-covariance matrix of the standardized exploratory variables; I is an identity matrix; λ_k are the eigenvalues of the corresponding RDA's kth axis, and u_k represents the canonical analysis produces ordinations of *y* that are constrained linearly to a set of variables E in such a way that either Euclidean distance or eigenvectors is displayed. A biplot or triplot diagram is then used to display the RDA's graphical output. Triplots distinguish between the ordination diagrams for

the response and explanatory factors, whereas diagrams interpret the ordination of sites, objects, and observations with respect to Y and X. on the contrary, a triplot has three components: explanatory variables (E), response variables (R), and sites/observations (Legendre and Legendre, 2012). In this study, RDA is conducted using Origin for Windows 10® software. The results are performed using scaling type 2 with species normalized scores.

2.3 ML-based prediction and forecasting of total MSWG rate

2.3.1 Overall methodology

For prediction and forecasting studies, this study targeted comprehensive modelling in the entire study area of Guwahati city (Fig. 2.3 (a)). The spatial expansion of Guwahati was due to a huge population growth explosion in a short period of time. The findings provided valuable insights into the spatial patterns and characteristics of Guwahati city. Thereby, urban dynamics and distribution were highlighted. Fig. 2.3 (b) depicts the MSWG per capita in the Guwahati region (Data source: GMC, 2019). It can be observed that the standard deviation of MSWG per capita of municipalities with a lower population growth density (rural municipalities) was lower than that of areas with more densely populated areas (urban municipalities). Thus, MSWG per capita in urban areas had a wider range. For instance, an area of 14.98 km^2 (ward 1) with a population density of 2223 (33305 persons) generated of 160 tonnes of waste per day. Also, the most densely populated ward refers to 49,021 persons occupying an area of area 25.97 km^2. With respect to annual MSWG per capita, the increment was approximately three-fold, (from 2.84 kg in 1991 to 6.21 kg $person^{-1}$ in 2019).

Python version 3.8 was deployed to analyse the data and implement the prediction models. To automate loading, pre-processing and integration of data, python scripts were developed. NumPy, Scikit-learn, Pandas and Matplotlib library packages were considered to implement modelling efforts on a i7-4790 CPU @ 3.60 GHz processor configuration. Fig. 2.4 outlines the overall methodology being adopted. The figure depicts a data collection and preprocessing phase and a modelling and analysis phase in the overall methodology. For the preparation and transformation of raw secondary data into processed data for appropriate modelling and analysis, several data pre-processing steps were needed. These refer to the extraction of data (secondary), data loading into proper data structures, derivation of socio-economic

factors, data transformation, outliers removal through filtering mechanisms and finally integration of data into consolidated datasets. Finally, after retrieval of the combined datasets, the data were fed to

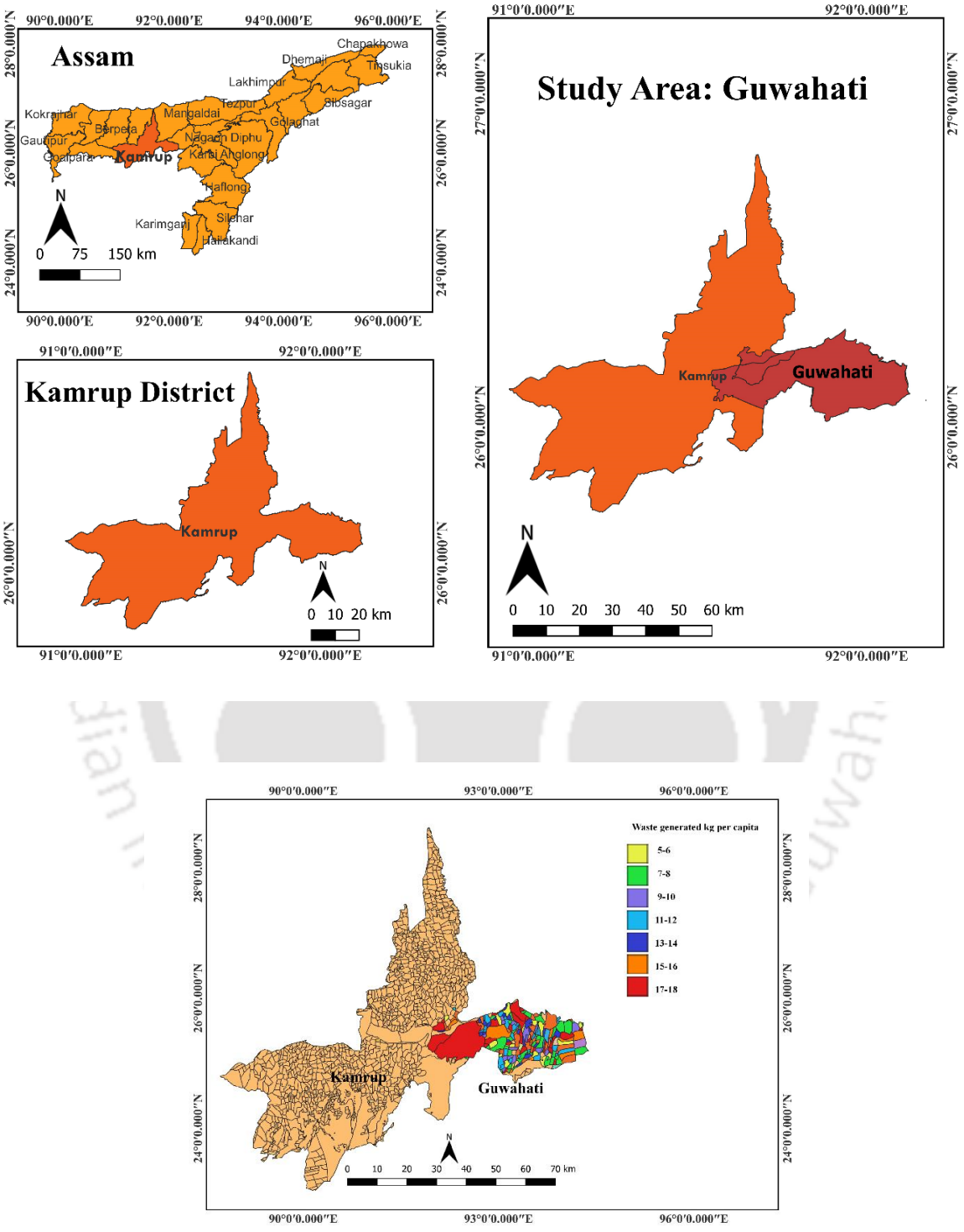


Fig. 2.3 (a) Location map of the study area and **(b)** Ward-wise generation per capita of Guwahati city (Data source: GMC, 2018)

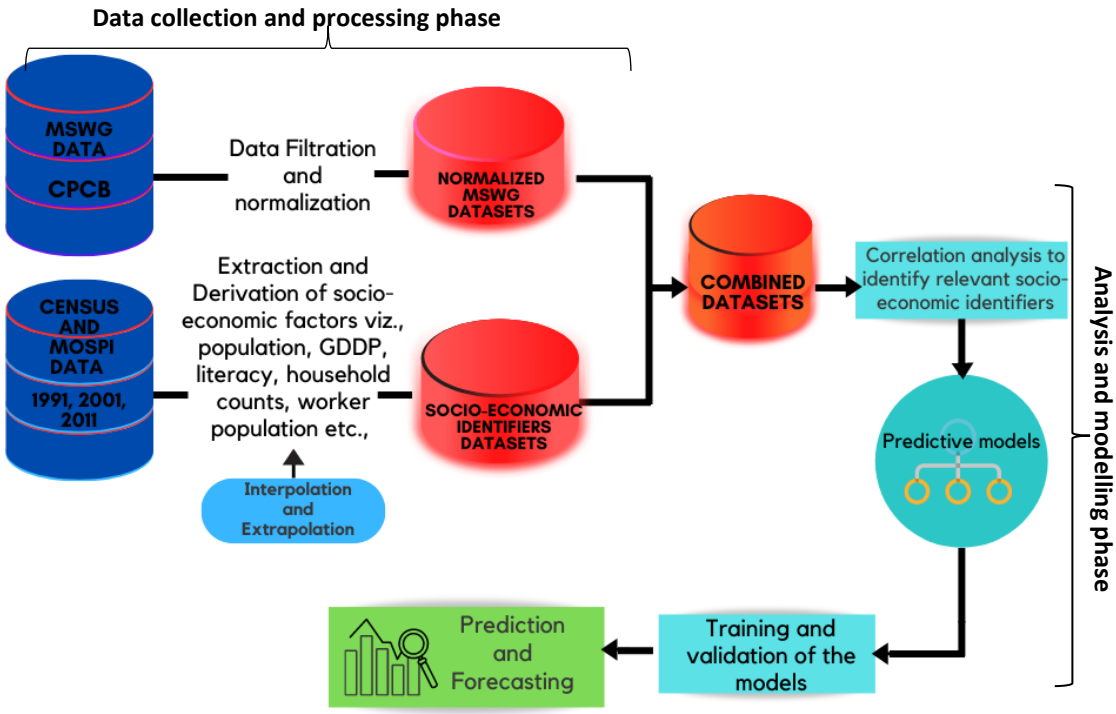


Fig. 2.4 Overall methodology for the ML algorithms-based prediction and forecasting of total MSWG rate

ML algorithms and for the respective functionalities of the training, validation and forecasting of the alternate models.

2.3.2 Data consolidation and Selection of Socio-economic Parameters

The consolidated data for the conducted investigations refer to population growth, GDDP, GDDP per capita, LP, WP and HH information (Table 2.3). For these data, the best data sources correspond to the Indian census data (a decennial census performed nationwide), and the annual data were collected from the MOSPI (for GDDP). Also, the GDDP per capita was obtained by dividing the district's (Kamrup) GDDP with its total population growth. For HH data, the number of HH and HH sizes were considered. Also, the number of HH was further subcategorized into good, living and dilapidated conditions. Accordingly, HH size was classified as 1–3, 4–8 and above 9. An extensive prior art in the chosen field highlighted socio-demographic factors and datasets as the most significant contributors to the total MSWG rate (Benítez et al. 2008; Ghinea et al. 2016; Johnson et al. 2017; Noori et al. 2009; Tchobanoglous & Kreith 1994). The population growth has been unarguably a dominant parameter to critically influence the total waste generation rate. Hence, it was considered as a key variable to customize the performance of long-term time series prediction models (Hockett et al. 1995; Katsamaki

et al. 1998; Navarro-Esbrí et al. 2002; Rimaityte et al. 2012). Most socio-economic factors that influence the MSWG rate such as education level (Debrah

Table 2.3 Data sources summary for the total MSWG prediction and forecasting study.

Parameters	Data Sources	Available Years
	Independent variables	
Population growth	Census Data	1991-2011
LP	Census Data	1991-2011
GDDP	MOSPI Data	1991-2011
HH Data (HH Count: *Good, Living, Dilapidated* & Sizes: *HH 1-3, 4-9, Above 9*)	Census Data	1991-2011
WP (main, marginal & non-WP)	Census Data	1991-2011
	Dependent variable	
MSWG	CPCB	1991-2016

et al. 2021), employment status (Bandara et al. 2007), total HH and HHSize (Suthar & Singh 2015) were provided in the national census data and at 10-year intervals. The annual data for GDDP (Grazhdani 2016) were obtained from MOSPI at the municipal level. This was as per the norm that the MOSPI data at subdivisions referred to that of the municipality level. The same was verified through a comparison of population growth projections being reported in waste generation and census data. Further, such a level-based data classification in the census data evaluates the competence of the desired socio-economic factors in various alternate methods. Also, over the years, the level-based data varied substantially. Accordingly, the socio-economic factors were interpreted in such a way that they can be invariably computed with the data structures of the census data of 1991, 2001 and 2011. To evaluate the annual values from 1991 to 2011, linear interpolation and extrapolation were adopted.

To provide adequate consistency for short-length data imputation issues, linear interpolation was reported to be effective (Sun & Chungpaibulpatana 2017). Additionally, an analysis of the relevant socio-economic factors in the census reports affirmed progressive but not random variations.

The MSWG rate data for validation were procured from the CPCB, Government of India. The CPCB presents annual consolidated reports being developed with the forwarded reports of the SWM from the local bodies and then to the State Pollution Control Boards. Thereby, CPCB is responsible for coordination and assigns resources for waste management programmes being deployed by multilevel

municipal governance bodies. Thus, CPCB information with respect to waste generation rates is available in the public domain.

Despite generating a compiled waste generation data from the municipalities, the data can have errors due to the self-reporting strategies. Thus, such data may not be free from errors due to inaccurate records, over- or under-evaluated waste quantity data and misinterpreted survey records. For this reason, since obtained process data reconciliation and data validation are prone to possess erroneous information and outliers for the carried investigations, it was assumed that to describe cross-sectional differences among municipal bodies, the data possessed adequate quality and variation. Accordingly, the data can be applied for the development of better predictive models.

To assist modelling efforts, preliminary screening of socio-economic factors was implemented between MSWG rate and socio-economic factors. To do so, *r* being applied in earlier investigations as a ranking technique to select ML algorithm features was adopted (Iguyon & Elisseeff, 2003). Accordingly, *r* between each variable pair were calculated with the expression:

$$r = \frac{N\left(\sum xy\right) - \left(\sum x\right)\left(\sum y\right)}{\sqrt{\left[N\sum x^2 - \left(\sum x\right)^2\right]\left[N\sum y^2 - \left(\sum y\right)^2\right]}} \quad \textbf{(2.2)}$$

where, r is the correlation coefficient between variables x and y, and N is the number of observations respectively. Thus, using correlation coefficients, the criticality and sensitivity of various socio-economic parameters was addressed with a considerable improvement in the prediction accuracy as a bias.

2.3.3 Data Pre-processing

To prepare and transform collected data into refined data, various data pre-processing steps were necessary for appropriate modelling and subsequent analysis. These refer to the structures deriving socioeconomic factors, transforming data, filtering to remove outliers, and finally integrating the data into combined datasets.

For the detection and removal of outliers from datasets, interquartile range (IQR) filtering was adopted. Outliers can result into inaccurate estimates or significant changes in waste generation patterns can

report divergences from the municipal bodies. The upper and lower limits of the valid data range per municipality were determined as $Upper\ Limit = Q_3 + IQR \times 1.5$ and $Lower\ Limit = Q_1 - IQR \times 1.5$. In these expressions, Q_1 and Q_3 correspond to the first or lower and third or upper quartiles of the collected data, respectively (Kannangara et al., 2018). Subsequently, the data points which exist beyond the upper and lower quartiles were considered as outliers and were filtered.

2.3.4 Machine Learning Techniques

Regression has been a dominant and a major statistical practice in ML and can be applied in economics, psychology, geography, and so forth (Sammut and Webb, 2017). To evaluate the dependence or relation between random variables of interest and deduce mathematical functions, regression analysis (RA) can be applied (Wisniewski & Rawlings, 1990). RA requires two real-valued variables being represented as target/dependent (y) and independent variables (x). The RA's main objective is to map a function such that $y = f(x) + \varepsilon$, where ε corresponds to the error (Draper & Smith, 1981; Freund & Wilson 1998). Multi-regression problems arise due to the occurrence of more than one independent variable. For such systems, regression expression transforms to $y = f(x_1, x_2, x_3 ..... x_n) + \varepsilon$ where $(x_1, x_2, x_3 ..... x_n) \in x$. Several independent variables namely population growth, LP, GDDP and $GDDP_{per\ capita}, HH,\ HH_{size},\ WP$ and one target variable MSWG were considered in the modelling effort of the thesis. To estimate relevant mathematical functions, regression algorithms were applied, to represent MSWG as a function of all considered seven core independent variables. Mathematically, the objective function can be expressed as:

$$MSWG = f(POP, LP, GDDP, GDDP_{percapita}, HH, HH_{size}, WP) + \varepsilon$$

2.3.4.1 Decision Tree Algorithm

Among ML methods, the tree-based models are popular. This is due to their iterative divide-and-conquer nature. In such models, non-parametric method is followed to recognize complex patterns associated to the classification of tasks being characterised with several pattern types and a large number of attributes (Kannangara et al., 2018). Such models are efficient and easy to implement but with intensive computation. A set of highly explicable and logical (if-then) conditions is constructed by the

tree-based ML models. This was done recursively by subdividing the decision space into smaller subspaces with training data presented to the decision process in the form of a tree. Thus, the algorithms can discreetly implement feature selection and can be applied for classification and regression on datasets with larger variables (Quinlan, 1999; Solano Meza et al., 2019). Recently, DT techniques have been applied on a wider scale to predict the waste generation rate and for long-term waste prediction (Breiman et al., 2017).

Numerous techniques exist to construct a tree. However, regression tree was constructed with a famous yet most widely applied framework called classification and regression trees (CART) (Breiman et al., 2017). The CART involves the specific logical tests being conducted for the entire datasets (*D*) that exist at the root or internal node of the tree. Thereafter, the DT partitions the test data into two groups D_1 and D_2 such that the partitions mutually satisfy the minimization of overall SSE. Accordingly, the SSE can be evaluated as

$$SSE = \sum_{i \in D_1} (y_i - \bar{y}D_1)^2 + \sum_{j \in D_2} (y_j - \bar{y}D_2)^2 \quad \textbf{(2.3)}$$

where $\bar{y}_{D_1}$ and $\bar{y}_{D_2}$ refer to the mean predictions of the training set for D_1 and D_2 respectively. The process is iterated until the convergence analysis conditions are met with no more possible splits. Finally, the predicted values of the target variables result from the decision nodes (last nodes) (Hastie et al., 2009; Johnson et al., 2017; Kannangara et al., 2018).

2.3.4.2 Random Forest Algorithm

Using random sub-spaces and bagging (Nisbet et al., 2009), the research group of Breiman were the first to delineate upon the RF model (bagging algorithm) as a classification-based model (Breiman et al., 2017). Being considered as one of the most dominant ML algorithms, the RF incorporates an ensemble of tree predictors. In these, each tree relies on the values of an arbitrary vector being sampled individually and with the same distribution. Thereby, RF generates final output as an average of all the tree predictions (Shi & Horvath, 2006).

2.3.4.3 Gradient Boosting Algorithm

The GB model demonstrated the ability to model complex non-linear relationships between variables and on the basis of DT regression model's concept. It was reported to be promising to achieve higher prediction accuracy than that being achieved through conventional time-series approaches (Johnson et al., 2017). The DT utilizes a bifurcation approach and accordingly customizes various linear models to fit each region (Breiman et al., 2017; Hastie et al., 2009). Thereafter, the process is recursively performed by involving the determination of the split point (maximum deduction for residual sum of squares, RSS) at each stage. Such a methodology would result in a single tree-like structure that best depicts the underlying correlation between variables in a dataset.

The thesis addresses two ensemble procedures namely RF and GB (extensions of bagging and boosting respectively). Bagging (also known as bootstrap aggregation) and boosting are two widely used ensemble learning paradigms (models that result as a combination of multiple simple models) in an ML algorithm. The core idea behind these DT-based ensemble strategies is to build several DT and consolidate their predictions through the average (in regression) or voting (in classification) methodologies. Accordingly, the variance gets reduced but enhances the prediction accuracy. Thus, bagging constructs individual DT and allocates equal weight to all DT. On the other hand, in due course of the boosting of the new DT, the performance of the prior ones gets influenced and a weight based assignment on the trees' performance supports the computation procedures (Friedman, 2001; Sutton 2005).

In the traditional approaches such as statistical and material flow models, the application of the heterogeneous data and minimization error and uncertainties are not possible. Hence, prediction and forecasting accuracies could not be enhanced. As a viable alternative, the ML has an added advantage to include all possible information in the database and filed survey data. Thus, due to non-priori selection of the variable, the ML based methods are more successful than the traditional ones. The approach discussed will assist the decision makers to accurately estimate the MSWG rate for the effective planning of the SWM facilities. This is a limitation in the traditional approach for MSWG rate prediction. Also, among ML, except DT, RF and GB, other approaches have limitations in terms of

identification and inclusion of other influencing explanatory variables, sensitivity to outliers, lower accuracy etc., However, the tree-based models do possess an advantage by considering all the available information in the particular field. Henceforth, greater prediction accuracy of the model can be realized.

2.3.5 Model Training

The poor performance of the ML algorithms with unseen data samples, results in overfitted models. Thus, model testing is mandatory with the unseen data. To do so, the dataset was divided into training and testing datasets. While the training set was to build the model, the testing set was deployed to verify the over-fitness of the data. In the relevant prior art, the training to testing partition ratios of 80:20 or 60:40, 90:10, 85:15 were deployed (Pao & Chih, 2006). Due to unavailability of new data, usually training and test datasets were split with 70 to 30 ratio. This is due to the precise ratio enabling average model error.

2.3.6 Model Testing and Validation

To analyze the regression algorithms' capabilities for the set objective, the trained models were tested with new data not utilized for the training phase. The validation stage is based on a two-step hierarchy namely hyper parameter optimization (HPO) phase and a training or prediction phase (Kroese et al., 2019). During HPO, to infer upon appropriate parameters for the data with ten cross-fold validation for all tree models, a grid search was implemented simultaneously. The data collected from 1991-2011 were split into ten subsets (K = 10 folds). In the first iteration, the first fold corresponding to the year 1991 - 1992 was used to test ML models and the remaining folds (1993 - 1994, 1995 - 1996, 1997 – 1998, 1999 – 2000, 2001 – 2002, 2003 – 2004, 2005 – 2006, 2007 – 2008, and 2009 - 2011) were used for the training datasets. In the next iteration, the second fold (1993 - 1994) served for the testing stage and the rest of the data (1991 - 1992, 1995 - 1996, 1997 – 1998, 1999 – 2000, 2001 – 2002, 2003 – 2004, 2005 – 2006, 2007 – 2008, and 2009 - 2011) was deployed for the training. This process was repeated upto the final iteration stage. Similarly, for waste data (CPCB), the first fold corresponding to the year 1991 - 1993, was used to test ML models and the remaining folds (1993 - 1996, 1996 - 1999, 1999 – 2002, 2002 – 2005, 2005 – 2008, 2008 – 2010, 2010 – 2012, 2012 – 2014, and 2014 - 2016) were used to the training phase. Thus, the final iteration was not included in the cross-validation

analysis. Also, it shall be noted that the ratios between training/predicting data did enhance from one iteration to the next (the proportion of predicting data was 3%, 7%, 12%, 16%, 21%, 28%, 34%, 45%, and 51% in the 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th and the 9th iterations respectively). As a result, the 10-fold cross validation can be considered as a procedure being performed to estimate the ability of ML models for the prediction of missing data. Also, the following step involved the application of parameters inferred from the optimization step, and subsequent comparison through error computation strategies being summarized as follows.

Firstly, the train scores and the test scores were determined to infer upon the model accuracy. The predictive performance of the ML model was quantified in terms of MSE, MAE and coefficient of determination (R^2) (Breiman et al., 2017). These can be expressed as:

$$MAE = \frac{\sum_{i=1}^{n}(Y_i - X_i)^2}{n} \tag{2.4}$$

where Y_i = predicted value, X_i =true value and n= total number of datapoints

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2 \tag{2.5}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2} \tag{2.6}$$

where n is the number of observations, $\hat{Y}_i$ is the model predicted value, Y_i is the actual (data) value, $\bar{Y}$ is the mean value of waste generation rate. The SSE was also evaluated as a percentage value and was achieved with the RMSE. For both training and testing datasets, the SSE and R^2 were determined and thereby affirmed performance indices. Due to the adjustment of model parameters and structure during the phase, the training error is usually found to be lower than the testing error.

2.3.7 Forecasting Models

Using moving average (MA) technique, the forecasted MSWG was determined to understand pertinent data trends with alternate best fit models. Thus, the predictions are an average of any subset of numbers. The MA is appropriate for the forecasting of long-term trends and can be applied for any time period range. In the thesis, three MA approaches namely SMA, WMA, and EMA were considered for their incorporation into the long-term forecasting prediction models (Droke, 2001).

The SMA is a simple and straightforward technical indicator and is evaluated as the summation of the recent data points in a given dataset and their division with the total number of time periods. Thus, the SMA is best expressed as:

$$SMA = \frac{(A_1 + A_2 + ...A_n)}{n} \tag{2.7}$$

where A =Average in period n =Number of time-periods

The WMA involves the assignment of heavy weights to the more current data points due to their greater relevance than those data corresponding to the distant past. Thus, different weights have been assigned at diverse points of the sample window during WMA. Mathematically, the WMA is the complexity of the data for a fixed weighting function and is expressed as:

$$WMA = \frac{\sum_{i=1}^{n} W_i * D_i}{\sum_{i=1}^{n} W_i} \tag{2.8}$$

where W =weighting value, D = Data values, and n = number of time-periods

EMA customizes greater weight assignment to the most recent data points. Thereby, EMA is more responsive in the forecasting methodology, and involves the following four steps:

Step 1: Compute the SMA over a particular time period.

Step 2: Calculate the EMA weighting multiplier (also known as the "exponential smoothing") as per the following norm:

For any time period t, the smoothed value S_t can be determined with the expression:

$$S_t = \alpha y_{t-1} + (1-\alpha) S_{t-1} \quad \textbf{(2.9)}$$

$$0 < \alpha \leq 1 \qquad t \geq 3$$

The above expression is the basic equation of exponential smoothing. In this expression, the parameter α is termed as the smoothing constant and *y* refers for the original observation.

Step 3: Use both smoothing factor and previous EMA and reach to the current value

Step 4: Assign higher weight to recent data points with the expression:

$$EMA_c = \left(\left[C_T X \left(\frac{s}{1+N}\right)\right]\right) + EMA_p X \left(\left[1-\left(\frac{s}{1+N}\right)\right]\right) \quad \textbf{(2.10)}$$

where $EMA_P =$ EMA for the previous time period,

$EMA_C =$ EMA for the current time period, $s = smoothing$, N = number of time-periods and $C_T =$ Current time-period

2.4 ML algorithms-based prediction and forecasting of organics and recyclables solid waste generation rate

2.4.1 Site selection and data collection

The study area was divided into three sub-zones (Z1, Z2, Z3). This was based on the local municipal ward-list record. To investigate the HH waste characteristics, 31 wards were chosen. The demographic data for these areas has been depicted graphically in Fig. 2.5. In terms of the database, government data sources were used to collect social, economic, and demographic data from 2001 to 2011 (Census of India, MOSPI, CPCB, GMC). Thereby, correlation analysis was conducted. According to census data (Census of India, 2011), the population growth of these wards altered from 21,462 to 39,116. The stratified random sampling method was deployed. The data was obtained from the GMC, Assam, Census of India, and included a detailed account of the number of HH, HH sizes, income levels, and waste generation rates.

Between 2001 and 2011, in about 229,716 HHs in the city, 108,918 were found to be in good condition, 84,940 were deemed livable, and 35,858 were classified to be in dilapidated conditions. Each zonal division has been schematically represented in Fig. 2.6 (a). Similarly, the information on HH sizes is

divided from 1 member to more than 8 membered families (Fig 2.6 (b)). The MSW fractions were sorted into six primary categories. This was based on frequent characterization and prior research. These are food and kitchen waste, paper residues, polycarbonate and plastic bags, glassware and porcelain scrap, rubber and textiles, and cardboards. Additionally, a variety of materials such as sawdust, leaf litter, garden waste, dirt, metal jars, cans, and other inert objects were also sorted. The largest quantities of waste generated were 46%, 48%, 35% for OW; 11%, 19%, 12% for plastics; and 12.5%, 21.5%, 13% for papers in the sub-zones Z1, Z2, and Z3 respectively Fig 2.6 (c). The percentage waste fraction for the OW, papers and plastics (recyclables) have been depicted in Fig 2.6 (d).

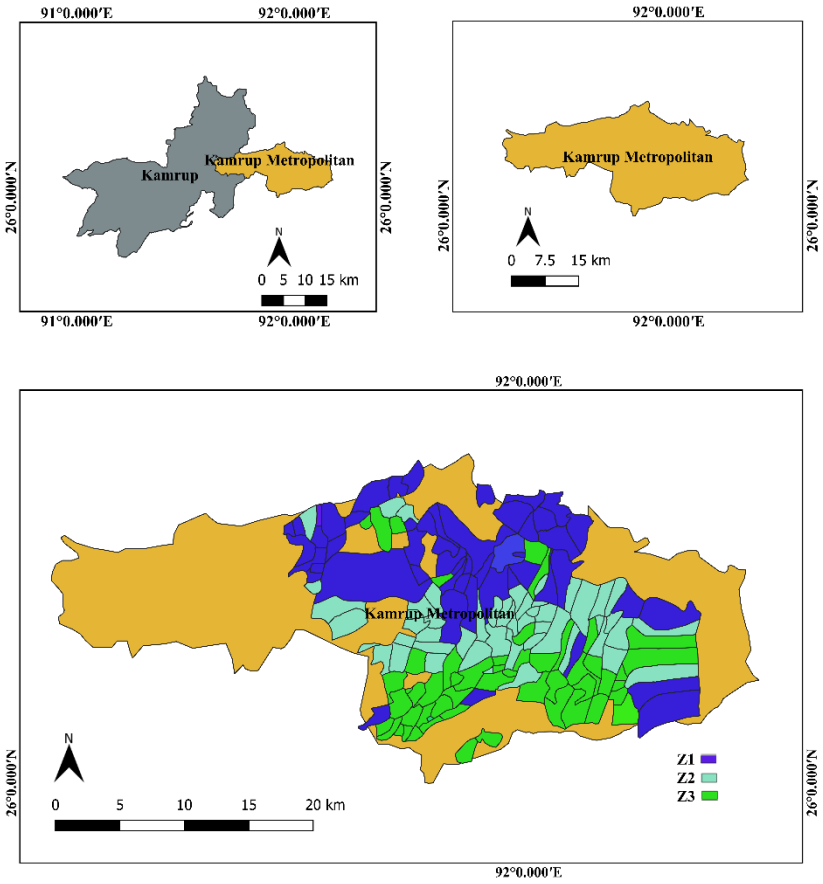
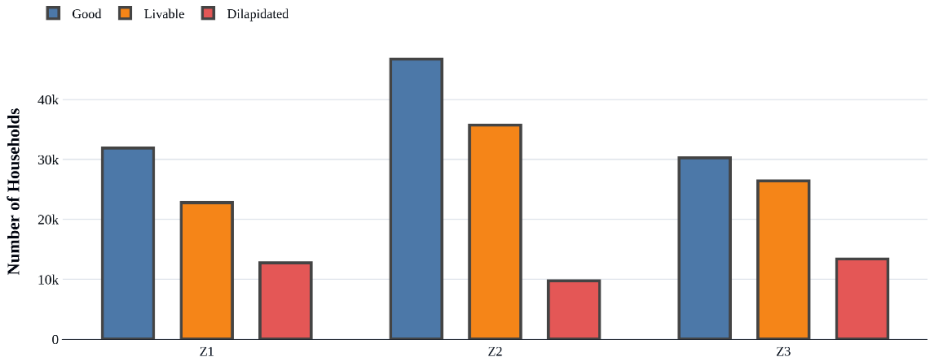
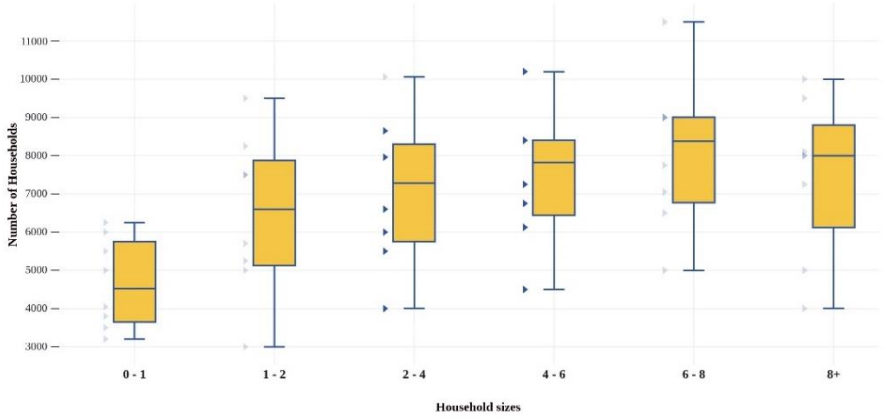


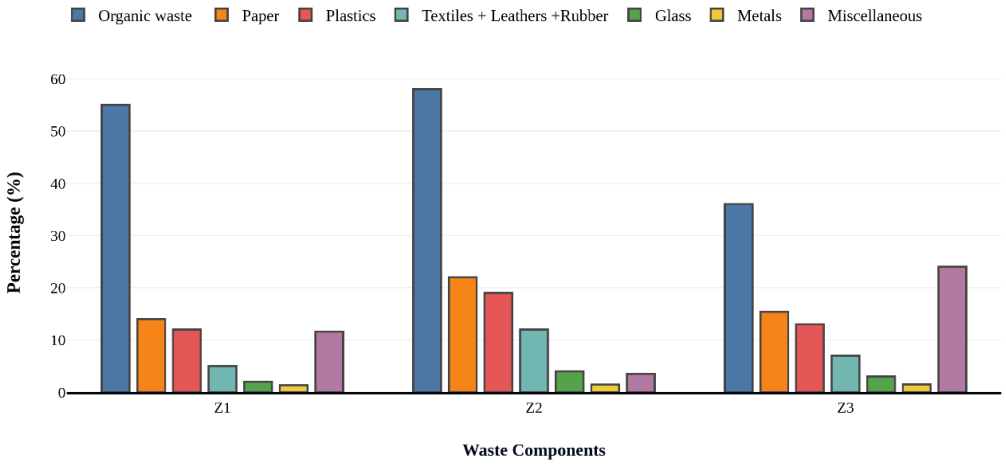
Fig. 2.5 Zone-wise representation (Z1, Z2 and Z3) of the study area in the recyclables and OW generation rate study



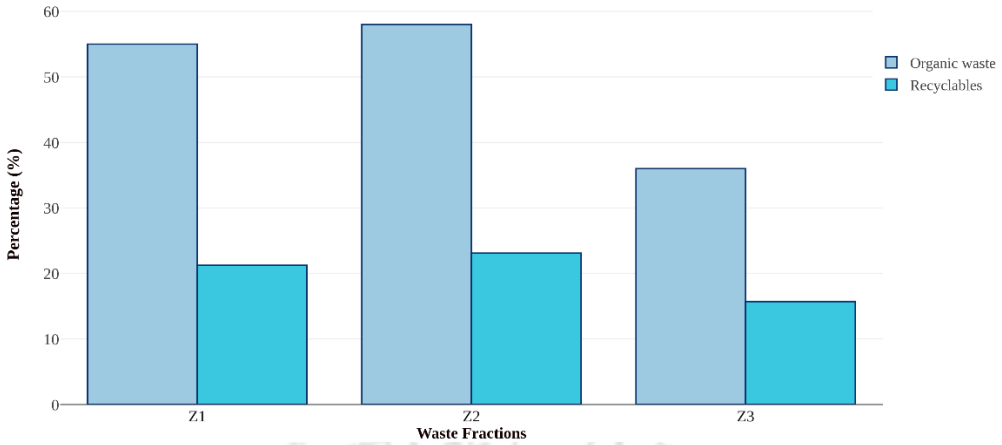
(a)



(b)



(c)



(d)

Fig. 2.6 Zone wise (Z1, Z2 and Z3) distributions of **(a)** Number of Households (HH) **(b)** HH_{sizes} **(c)** percentage distribution of waste components and **(d)** percentage distribution of waste fractions in the recyclables and OW generation case study

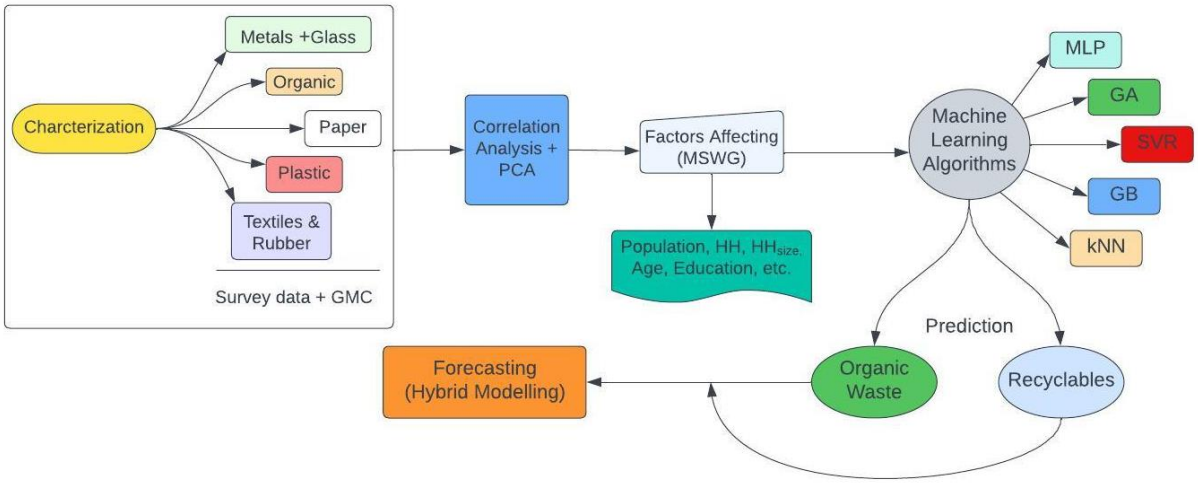


Fig. 2.7 Overall modelling methodology for the prediction and forecasting of recyclables and OW generation rate

2.4.2 Overall ML based Methodology

The study followed a two-phase methodology, (as demonstrated in Fig. 2.7) and thereby involved a phase of data collection and pre-processing, and then with a phase of modelling and analysis. The analysis and modelling phase was implemented with Python scripts. To ensure that the raw secondary data is suitable for modelling and analysis, various pre-processing measures have been employed. These measures include data extraction, loading into appropriate data structures, derivation of socioeconomic factors, data transformation, outlier removal with filtering mechanisms, and consolidation of data into

combined datasets. The merged datasets were subsequently utilized to train, validate, and predict alternative models using ML algorithms.

2.4.3 Waste Data Availability and Selection of Economic Criteria

Among the reported waste quantities, residential MSW were considered as the most reliable entities (Data sources: GMC). The amount of waste dumped was used to determine the total MSW of the HH. The municipal bodies provided good data of glass, metal, paper, plastic, yard/leaf debris, and kitchen OW. The remaining information was obtained from the Census Statistics (Government of India). It was repeatedly discovered that economic factors which measure economic growth and living conditions do influence garbage generation. Several studies, including those by few authors (Abdoli et al., 2011; Gomez et al., 2008), have demonstrated the importance of population growth as a critical predictor in long-term time series waste forecasting models, as it significantly impacts overall waste generation. In cross-sectional studies, population growth was combined with other waste volume variables to derive per capita waste, which assumes that per capita waste production is not influenced by population size (Hockett et al, 1995). To standardize the amount of waste generated by HHs in a municipality, the total number of HHs were used as the normalization variable. This allowed the determination of the number of residences serviced by waste collection services on an annual basis. Few authors (Ghinea et al., 2016) found that demographic age structure has a significant impact on garbage generation. In cross-sectional studies, population growth has been incorporated as a component of dependent waste quantity variables. This resulted in per capita waste calculations. The assumption is that per capita trash production is independent of population growth size.

In a related prior art (Gomez et al., 2008), age structure has been inferred to have a significant impact on most societal characteristics, including spending, saving, employment, and taxation. As an ideal parameter, it is important for the prediction purpose as it recognises moderate patterns and gradually tends to vary. Another important factor that has been mentioned is the size of the HH. Empirical garbage survey studies demonstrated a negative correlation between population growth size and waste generation (Bandara et al., 2007; Benítez et al, 2008). This is contradictory to the assumption of a constant per capita waste production rate. Prior research also conveyed that the waste generation rates are also affected by socioeconomic factors such as income and education level (Grazhdani, 2016). Also,

the data pre-processing, data integration, outlier removal has been duly followed as per the methodology elaborated in section 2.3.2 and 2.3.3.

2.4.4 Modelling Algorithms

2.4.4.1 Multilayer Perceptron Algorithm

As a neural network model, the MLP uses a mathematical neuron model called a perceptron for loops prevention by establishing connections between nodes. It is a feed-forward network that consists of multiple linked perceptron layers. In these layers, each perceptron possesses an activation function and weights are assigned for every input. The MLP aims to achieve maximum prediction accuracy by deriving these weights from the available data. Along with one input and output layers, it can have multiple hidden layers.

The basic functioning of the simple MLP comprises of five steps. Firstly, the input data is initialised and distributed randomly. Secondly, the network is trained to predict outcomes. Thirdly, the cost of error is calculated, in which the mean squared error (MSE) is typically used for regression. Fourthly, the weights are recalculated based on the error and the process of repeating steps (b) and (c) continues until the error is minimal. MLPs are versatile modelling techniques for the effective handling of diverse types of data. Deep learning, which is a vast area of ML that focuses on neural networks, is considered to be a subset of MLP (Yang, 2019; Zaki and Meira, 2019). The MLP model can be represented by the following equation:

$$Z_i = f\left(b_i + \sum (w_{ij} \times p_i)\right) \quad \textbf{(2.11)}$$

where Z_i is the output of the MLP model, b_i is the bias term for the i^{th} neuron, w_{ij} is the weight between the i^{th} and j^{th} neurons, and p_i is the input to the i^{th} neuron. The activation function f is typically a sigmoid function that maps the input to a value between *0* and *1*.

2.4.4.2 Genetic Algorithm

The GA is a class among evolutionary algorithms, and mimics Darwin's theory of evolution. In the GA, an initial population growth of solutions to the GA is represented by chromosomes (Bautista and Pereira, 2006). Selected solutions of the population growth are eventually processed to create a new

population growth. The motivation for the new population growth is its ability to outperform the existing population growth. The fitness of solutions ensures chosen population growth for newer solutions (offspring). Thus, the more suitable a solution is, the better equipped it will be to pass on its characteristics to the following generation. The GA is often carried out as an iterative process until a predetermined criterion was met (for instance, the number of population growth or an improvement in the best solution). The GA algorithm can be represented by:

$$W(x+1) = W(x) + \lambda(\Delta W(x)) \quad \textbf{(2.12)}$$

where $W(x)$ is the current generation of weights and biases, λ is the learning rate, and $\Delta W(x)$ is the change in weights and biases based on the fitness function.

2.4.4.3 Support Vector Regression Algorithm

The SVR algorithm aims to realize a hyperplane that best fits the training data through error minimization. The hyperplane being defined as a linear function to produce an output variable with input variables is expressed as:

$$Y = a^n.b + c \quad \textbf{(2.13)}$$

where:

- Y is the predicted output variable
- *b* is the input variable
- *a* is the weight vector
- *c* is the bias term.

Thus, during the training phase, the SVR learns the weight vector and bias term during training using support vectors that exist closest to the hyperplane. Accordingly, the primary objective of SVR is to maximize the margin between positive and negative support vectors. However, it allows some data points to fall within the margin or even onto the wrong side of the hyperplane. In non-linear SVR, a kernel function is used to transform input variables into a higher-dimensional space. Accordingly, a linear regression model can be applied (Cao et al., 2006). The equation that governs non-linear SVR is as follows:

$$Y = \sum_{i=1}^{n} \alpha_i K(x_i, x) + b \quad \textbf{(2.14)}$$

where *Y* represents the predicted output variable, *x* represents the input variable, n is the number of training data points, α_i is the Lagrange multiplier for the i[th] training data point, $K(x_i, x)$ is the kernel function that maps the input variables to a higher-dimensional space, and b is the bias term. The Lagrange multipliers α_i are learnt during training and are used to determine the support vectors that define the hyperplane. The algorithm's performance is significantly influenced with the selected kernel function. Hence, careful tuning is required to achieve the best results.

2.4.4.4 k-Nearest Neighbour Algorithm

The *k*NN is a non-parametric method and is used to forecast MSWG. The tuning parameter *k* in the *k*NN algorithm is important for the precise prediction. It identifies a collection of *k* samples (training data) that correspond to the unknown samples of the most closely related test data. Considering the median of the response variable, unknown samples are ascertained by the *k* samples (Hastie et al., 2009; Sammut and Webb, 2011; Kuhn and Johnson, 2013). Thereby, the elbow criterion method is used to calculate the value of the *k* parameter. The *k*NN algorithm can be formally written as:

$$f(p) = \arg\max_c \left(\left| \{ i : q_i = c, i \in I \} \right| \right) \quad \textbf{(2.15)}$$

In the above expression, the function *f* considers a vector *p* and classifies it to belong to a class c. This is achieved by finding the most frequently occurring class in the nearest neighbours to *p*. Thereby, it is stated as:

$$I = \{ i : p_i \in N(p; k) \} \quad \textbf{(2.16)}$$

Also, the ML approach for GB technique was implemented for the case (outlined in sub-section 2.3.3.3). For MSWG prediction, MLP, GA, SVR, *k*NN, and GB were chosen due to the fact that the MSWG rate involves complex patterns and non-linear relationships among various factors that influence waste production. Models such as MLP, SVR, *k*NN, GA and GB are well known for their ability to capture and learn such complex non-linear relationships in the data. Accordingly, they are well suited for the

desired task. Additionally, waste generation data may exhibit skewed distributions or may contain outliers. Thus, for the case, other considered models must demonstrate improved robustness and generalization in comparison to RF and DT models. The flexibility and adaptability of these models in handling diverse input variables, categorical data, and high-dimensional feature spaces further support their suitability for solid waste generation prediction. Lastly, the inclusion of different models enables the utilization of ensemble learning techniques, and eventually enhance overall prediction performance, accuracy and mitigate the risks associated with the limitations that exist due to the selection of a single model. Further, to implement model training, testing and validation phase (HPO), the methodology as outlined in sections 2.3.4 and section 2.3.5 respectively were duly followed.

2.4.5 Forecasting Models

For the forecasting purpose, time series analysis is a promising ML technique. Deploying a single independent feature (as input feature), targeted dependent features are predicted by the time series analysis and in the form of a univariate regression. Since forecast numbers depend solely on historical trends, they cannot be guaranteed. Nevertheless, a rough estimate of the events can assist researchers to gain a better understanding of the pertinent issues. As one of the most commonly used forecasting techniques for time series forecasting, the ARIMA was deployed. The ARIMA was proven to be an effective statistical method for time series data based future trends prediction (Abraham and Ledolter, 1986; Granger and Ramanathan, 1984). The methodology adopts differences between values but not the actual values. The ARIMA model has three components namely (i) moving average (MA), which computes successive averages over increasing time frames to represent a previously available time series, (ii) integrated structure, which replaces data values based on the difference between data values and previous values, and (iii) autoregression (AR), which models the changing values of variables based on their previous values (Abraham and Ledolter, 1986; Granger and Ramanathan, 1984). ARIMA uses a *abc* equation for forecasting, where parameter *a* represents the number of lagging observations, parameter *b* represents the number of times the raw observations get alteration, and parameter *c* represents the size of the MA window. In the AR model, the value of a variable is determined by its

own lagged value. The MA model calculates the values with lagged forecast errors. As a result, the generic ARIMA (model) equation is as follows:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_n Y_{t-p} \varepsilon_t + \varepsilon_1 + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} \qquad \textbf{(2.17)}$$

where y_t is the goal to be forecasted, α is a constant, $\beta_1 y_{t-1}$ and $\theta_1 \varepsilon_{t-1}$ represent the linear combination of lags Y ranging from lags 0 to *a* and 0 to *c* respectively.

In this study, a hybrid methodology was considered to incorporate both ARIMA and considered ML models. In this regard, Box and Jenkins (1976) suggested that a combination of hybrid models could potentially improve the accuracy of predictions and henceforth overcome limitations associated with a single model. Theoretical and empirical studies conveyed that the combination of various models lead to hybrid models that assured reduced generalisation variance or error. The combination efforts targeted risks associated to the reduction of a single, potentially inaccurate model, by simultaneously utilizing multiple models within the framework of hybrid models. Thus, models are sought for a combination scheme due to the difficulties that arise in the identification of the true data in the modelling of the mentioned process due to the individual models that have inherent limitation.

ARIMA statistical methodology was used in this study to forecast MSWG for both OW and recyclables. Since ARIMA-MLP performed the best, studies were further considered for the case. The data from the ARIMA model (model.predict test.csv) matched the 280 datasets and upto 2011 year. Using this dataset, trash generation for recyclables and OW was predicted, and ARIMA-MLP exhibited more accurate forecasting than other conventional ML methods. The ml prediction data frame was saved as a.csv file, and thereby created the dataset. The trained model was used to predict the label for the expected data with the method model.predict(). After obtaining the predicted data with the model.predict(X test) function, it was transformed into a 2-dimensional DataFrame in an array format. The column ids of the test DataFrame were retrieved and were combined with the projected data. A specific column or multiple columns were chosen to serve as indexes in the DataFrame. The set index() function was applied, and thereby passed the column or list of column labels as a parameter to configure the index

of the DataFrame. Finally, the index was set with the in-place parameter to avoid creating a new copy. The resulting DataFrame was then saved as a .csv file.

To resolve the issue of autocorrelated errors in the prediction equation, one period's worth of lag was added for the dependent variable or the initial difference of X was regressed onto itself. The ARIMA model that resulted from the analysis of this study was represented by the equation ARIMA (1,1,0). This indicates that while no moving average (MA) terms were included in the model, one autoregressive (AR) lag was considered, and one order of differencing was used to transform the series into a stationary one. In light of this, the following prediction equation can be realized:

$$Y_t = \gamma + Y_{t-1} + \theta(Y_{t-1} - Y_{t-2}) \quad \textbf{(2.18)}$$

In the above model, the output is predicted with a constant and slope coefficient. Thus, ARIMA (1,1,0) model is a first-order autoregressive model with one order of non-seasonal differencing.

However, for a given trained model, model.predict() predicted the label of predicted data. This method accepted one argument (the predicted data model. predict(X_test)), and returned the learned label for each object in the array. Thereafter, the predicted data (array format) was converted into a dataframe (a 2-dimentional tabular data structure). Eventually, column ids from the test_dataframe were taken and concatenated dataframe with predicted data. A specific column or multiple columns as an index were set in the DataFrame to eventually create a list of column labels being used to set an index. Thereby, the column or list of column labels were passed as an input to the DataFrame.set_index() function and henceforth set it as an index of DataFrame. Parameter inplace was used to set the index in the existing DataFrame but to create a new copy. Finally, the dataframe was exported into the .csv file.

For exporting ML prediction into .csv file the pseudo code is as follows:

- *model.predict(X_test) gives you an array of the prediction result → pred = model.predict(X_test).*
- *convert it to dataframe → pred = pd.DataFrame(pred,columns=[column_name_in_submission_sample]).*
- *take id column from test_data(given test.csv) and concatenate with pred →*
- *submission = pd.concat([test_data.id,pred],axis=1).*
- *Set column id as index →sub.set_index('id',inplace=True).*
- *Make csv of above dataframe sub → submission.to_csv(f" model.predict_test.csv")*

2.5 ML based Prediction of GHG and PMs Emissions Rate from Incineration and Landfill Sites

The GHG investigations were targeted for Guwahati city, the gateway city for North-east India and a region of strategic importance. The study also identified the predictive accuracy of all models and their computational time. Further, the strengths and weaknesses for GHG prediction were analyzed. The Kamrup solid waste disposal site is in Guwahati city and is managed by GMC. The site serves as the primary landfill for the city's municipal waste. The landfill site is situated in a low-lying area near the Brahmaputra River (Fig. 2.8). The site is divided into different cells, which are used for the disposal of different types of waste. Garbage trucks are used to bring the waste to the site. Thereafter it is unloaded at the tipping area. A small incineration plant exists at the site.

The term open dump conveys the open field dumping of solid waste in the Guwahati city. The engineered landfill refers to the landfill designated by the Guwahati Municipal Corporation (GMC) for the collective filling of solid waste generated in the entire Guwahati city. Guwahati, India, presently employs open dumps and one engineered landfill for solid waste disposal. A well-known open dump near the Ramsar Site, Deepor Beel, poses serious environmental risks. Henceforth, the city aims to facilitate the transition from open dumps to engineered landfills. This conveys the inadequacy of current waste management practices. Due to pertinent challenges of the weak soil, the West Boragaon landfill site has been adequately addressed with the geocells constructed with high-density polyethylene or novel polymeric alloy strips. Such scientific approaches stabilize the foundation, prevent premature dumping, and align with the best practices of the worldwide municipalities. All these measures reflect Guwahati's commitment to sustainable waste management.

Fig. 2.8 Solid waste disposal site of the Kamrup Metropolitan segment

2.5.1 Measurements and Data Collection

The measurements were carried out at two plots (approximately 50 acres of landfill area) and in the year range of 1970-2018. The emission data were collected from Emissions Database for Global Atmospheric Research (EDGAR). The EDGAR GHG emission gridmaps represent an independent and reliable source of information to support the analysis and development of territorial policies at sub-national level in the field of climate action. Other measurements such as the meteorological data (atmospheric T, RH, Pr, WS and WD) were taken from Modern-Era Retrospective analysis for Research and Applications (MERRA-2) and Iowa Environmental Mesonet (IEM). MERRA-2 is a NASA atmospheric reanalysis dataset that provides global meteorological data from 1970 year onwards. It is a comprehensive dataset that includes a wide range of meteorological variables such as T, Pr, wind, RH, and atmospheric composition. The MERRA-2 dataset is based on a state-of-the-art atmospheric model that combines satellite observations, surface measurements, and weather balloon data to produce a consistent and continuous record of the timely atmospheric conditions data (Gelaro et al., 2017). The dataset is produced at a high spatial resolution (horizontal resolution of 0.5 degrees longitude by 0.625 degrees latitude and for 72 vertical levels). The IEM is a weather network operated by the Department

of Agronomy at Iowa State University. The network provides real-time weather information and climate data for the state of Iowa and surrounding regions. The IEM network consists of over 1,200 weather stations located throughout Iowa, and in neighbouring states such as Minnesota, Nebraska, and Illinois. The stations provide a range of meteorological data, including T, Pr, WS and WD, RH, and barometric pressure. In addition to real-time weather data, the IEM website also provides access to historical climate data, weather radar imagery, and other tools and resources for meteorological research and analysis. Weather data can significantly influence the emissions of GHGs from solid waste landfills and incinerators in several ways.

For landfills, weather conditions such as T and Pr do affect the rate of OW decomposition and subsequent CH_4 production (Woon and Lo, 2013). T_{max} can enhance the decomposition rate. This leads to higher CH_4 emissions, and vice versa trends exist for wetter conditions that witness decelerated decomposition (Spokas et al., 2011). Wind patterns also affect the dispersion of CH_4 emissions, and due to this, they influence local air quality. For the incinerators case, weather conditions can influece the efficiency of combustion and the resulting emissions of GHGs and other pollutants (Mavrotas et al., 2015). T and RH do affect the combustion process. Thus, higher T lead to more complete combustion and lower emissions. Wind patterns can also affect the dispersion of emissions from the incinerator and potentially affect the air quality. Accurate weather data can be used to model the sensitive influence of these factors on GHG emissions from landfills and incinerators. With such pedagogues, the design and operation of these facilities can be improved to minimize emissions (Talaiekhozani et al., 2018).

Also, the amount of waste generated is a key factor to influence GHG emissions from solid waste landfills and incineration (Xu et al., 2019). Landfills are responsible for the emission of GHG gases such as CH_4, CO_2, and other volatile organic compounds. The amount of waste in a landfill affects the rate and duration of CH_4 production (a potent GHG with a warming potential 28 times higher than CO_2 over a 100-year time horizon). Therefore, larger landfills with more waste tend to produce more methane emissions. Similarly, incineration of waste generates GHG emissions such as CO_2, NO_x, and SO_2. The amount of incinerated waste directly affects the amount of generated GHG emissions.

Moreover, the transportation of waste to the landfill or incineration facility also contributes to GHG emissions, as the deployed vehicles as well emit CO_2.

Population growth and GDP information are important factors to influence GHG emissions rate from solid waste landfills and incineration (Abdoli et al., 2011; Daskalopoulos et al., 1998; Beigl et al., 2004). With increasing population growth values, the amount of waste generated also increases. Thus, more waste needs to be managed, and this contributes to increased GHG emissions. Landfills and incineration processes produce GHG emissions such as CH_4 and CO_2. Therefore, with greater population growth, greater demand exists for waste management. This leads to increased GHG emissions. The GDP is a measure of the economic output of a country. With increasing GDP (nation) or gross district domestic product, and GDDP (district), the consumption and production of goods and services also increases. This leads to an increase in waste generation and thereby leads to an enhanced GHG emissions from waste management activities such as landfilling and incineration.

For the study, available census data is from 2011 (previous data were from 1961-2001). The census reporting level for subdivisions was found to match the municipality level data of MSW. This was confirmed through a comparison of the population growth values in both datasets. Also, the annual GDP data were gathered from the MOSPI. To validate the MSWG rate data, the study obtained information from the Government of India. The CPCB collects annual consolidated reports from local bodies regarding solid waste management and coordinates and allocates resources for waste management programs implemented by multi-level municipal governance bodies. Given CPCB's mandate for the endurance of proper waste management practices, the waste generation rate data of the CPCB is publicly available.

A detailed account of the data preprocessing approach, outlier removal data standardization to ensure equal scales and variances, and application feature selection techniques for the reduction of the number of variables based on their predictive ability and significance has already been discussed in 2.3.2 and 2.3.3 sub-sections of the thesis. All these elements improved quality of data analysis and henceforth enhanced the accuracy and reliability of subsequent analyses.

2.5.2 Overall Methodology

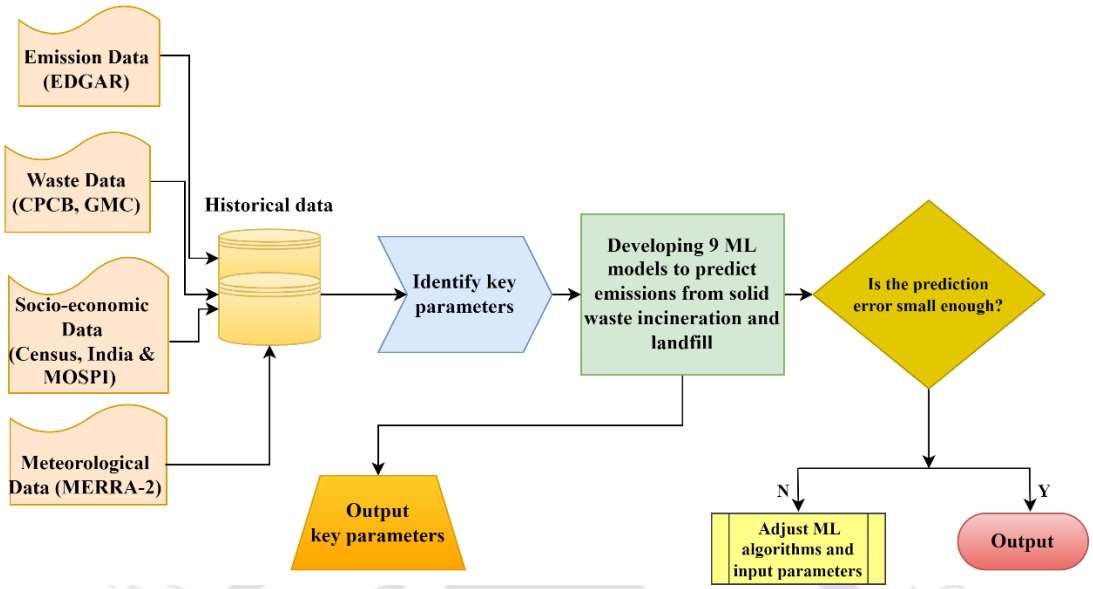


Fig. 2.9 Overall ML based methodology for the prediction of GHG & PMs emissions rate prediction

A comprehensive modelling framework was employed (illustrated in Fig. 2.9). An objective comparison between nine ML models (tree-based ML approaches, GA, MLP, *k*NN, SVR, ridge and lasso) was conducted to ascertain the most efficient model for a combination of GHG (CH_4, CO_2, N_2O, $PM_{2.5}$, PM_{10}, SO_2) emissions data (1970-2018) from EDGAR, waste data (GMC, CPCB), GDDP data (MOSPI), population growth data (Census, India) and meteorological data (T, RH, Pr, WS and WD from MERRA-2). For data analysis and prediction and forecasting purposes, python scripts were deployed.

2.5.3 Optimization of Model Inputs

An initial investigation was carried out to identify the most sensitive input parameters based on their significant statistical characters for the forecasting of GHG emissions and for the training period ranging from 1970-2018. Such an examination encompassed techniques such as Pearson correlation analysis, stepwise regression (Wilkinson, 1979), and neighbourhood component analysis (NCA) (Roweis et al., 2004). Among these, Pearson correlation allows a measurement of the linear relationship between variables through the determination of correlation coefficient. Thus, a high positive or negative correlation index infers upon influential variables for the targeted variable. In other words, through the selection of inputs with a high correlation, the most influential features can be addressed. This improves

model performance and interpretability. The stepwise regression approach iteratively selects and removes variables based on statistical significance. While forward selection adds variables based on individual contributions, the backward elimination removes variables based on their impact on the model fitness. Such a selection process can lead to improved interpretability, reduced complexity, and enhanced generalization ability of the model. NCA is a dimensionality reduction technique that functions to maximize the discriminability of the targeted variable. It analyzes neighbourhood relationships to determine informative features for the class separation. Thus, through the selection of inputs based on their discriminative capabilities, the NCA helps to reduce noise, improve class separability, and identify the most relevant variables in the modelling effort. Henceforth, results with better prediction accuracy and interpretability of the model can be achieved.

2.5.4 Deployed ML Algorithms

Four categories of ML models (Fig. 2.10) were targeted in this study. These were classical regression approaches (Ridge, Least Absolute Shrinkage and Selection Operator (LASSO), SVR, and *k*NN); tree-based ML approaches (DT, RF and GB); a feed forward ANN model (MLP) and evolutionary algorithm (GA). The regression algorithms applied in this study attempt to estimate functions elucidating upon the effect of all four core independent variables or datasets on the GHG emissions (CH_4, CO_2, SO_2, N_2O, $PM_{2.5}$, PM_{10}). Mathematically, the objective function is expressed as

$$Landfill_{emission} = f(population, waste\ generation, weather\ data, GDDP) + \varepsilon$$

$$Incineration_{emission} = f(\ population, waste\ generation, weather\ data, GDDP) + \varepsilon$$

2.5.4.1 Regularization Techniques

Regularization is a technique often deployed to control overfitness. It discourages the regression weights to reach large values. To do so, it reduces the variance at an expense of a marginal increase in the bias. Depending upon the weights' norm, the regularization adds a penalty term to the cost function (Flach, 2012; Kroese et al., 2019; Hastie et al., 2009; Bishop, 2006). Two alternate regularization techniques have been applied in this study namely LASSO (L1) and Ridge (L2) approaches. These have

subtle yet significant differences. Among these, while LASSO applies the L1 (Manhattan) norm, the Ridge applies the L2 (Euclidean) norm. For this reason, L2 regression weights may remain small yet

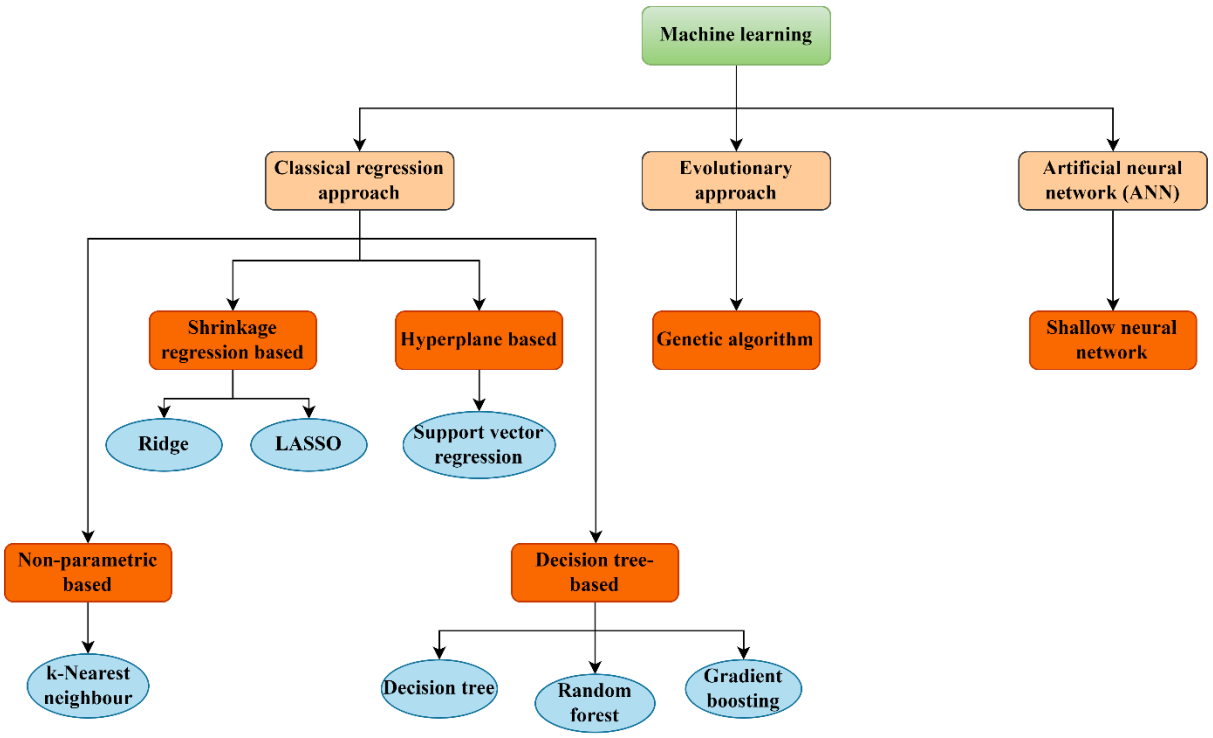


Fig. 2.10 Flowchart depicting deployed ML models for the GHG and PMs emissions case study

non-zero elements. However, the L1 favours sparse model and drives many weights to the zero value. Thereby, it acts as a feature extraction method for the case of multi-regression (Flach, 2012; Zaki and Meira, 2019; Hastie et al., 2009; Bishop, 2006). The regularized cost functions for L1 and L2 can be expressed as:

$$CF_{L1} = \sum_{i=1}^{n}(w.p_i + b - q_{act_i})^2 + c|w| \quad \textbf{(2.19)}$$

$$CF_{L2} = \sum_{i=1}^{n}(w.p_i + b - q_{act_i})^2 + c\,.w^2 \quad \textbf{(2.20)}$$

where $c \geq 0$ is the regularization constant that controls the trade-off between the regularization and SSE components of the regularized cost function, b is the bias/intercept and w is the slope/weight) to a set of data points (p) to predict future values of q for given values of p, *and* the actual value q_{act}. For the case of c being zero, no regularization exists and the model will have low bias and possibly high variance. However, when $c \to \infty$, all the weights tend to become zero. Thereby, they produce a low variance and high bias overfitted model. Thus, the c variation enables a finer balance between bias

and variance for an optimal predictive model. A small yet positive value of *c* always guarantees a solution (Flach, 2012; Sammut and Webb, 2011; Mohri et al., 2018; Hastie et al., 2009; Bishop, 2006). L1 and L2 are simple LR models with a normed cost function. Henceforth, the fitness method remains the same.

2.5.4.2 Tree, Evolutionary and Feed Forward ANN Approaches

A detailed explanation of DT, RF, GB, MLP, GA, SVR and *k*NN has been respectively elucidated in sub-sections 2.3.4.1, 2.3.4.2, 2.3.4.3, 2.4.4.1, 2.4.4.2, 2.4.4.3 and 2.4.4.4 of the Ph.D. thesis.

2.5.5 Model Evaluation

To evaluate the ML model's predictive capability, R^2, RMSE, index of agreement (IoA), MAPE, MAE, SSE were applied (Kazemi et al., 2021; Kubat et al., 2018). The R^2, RMSE and MAE related details have been discussed in section 2.3.5. The relevant expressions for SSE, MAPE and IoA are as follows:

$$SSE = \sum_{i=1}^{n} (Y_i - X_i)^2 \tag{2.21}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{Y_i - X_i}{Y_i} \right) \tag{2.22}$$

where Y_i = predicted value, X_i =true value and n= total number of datapoints, and $\overline{X_i}$ is the mean of the true value. The SSE was also determined as a percentage value with the RMSE method. The SSE and R^2 were calculated for both training and testing datasets, to accordingly confirm upon the performance indices. The training error has been frequently found to be smaller than the testing error due to the adjustment of model parameters and structure during the phase. The IoA is necessary in model evaluation as it quantitatively measures the agreement between observed and predicted values. Accordingly, it captures both systematic and random errors. It provides a robust and widely accepted metric for the comparison and selection of models based on their performance.

$$IOA = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} \left(\left| P_i - \overline{P} \right| + \left| O_i - \overline{O} \right|^2 \right)}, 0 \le IOA \le 1 \tag{2.23}$$

where O_i and $\overline{O}$ are the observation value and the average observation value respectively, P_i and $\overline{P}$ are the predicted and average predicted values respectively. For model testing, validation (HPO), and forecasting with time series approach (ARIMA), the methodology elucidated in the respective sections of 2.3.6 and 2.4.5 of the Ph.D. thesis were duly followed.

2.6 Modelling of Climatic Conditions-based Prediction of Compost Production Rate

2.6.1 Data Consolidation

The investigations focused on the influence of the climatic parameters on compost production. Accordingly, the model incorporated T, RH, Pr, WS and OW. Thus, a very good climatic database is required to assist modelling efforts. For this, MERRA-2 was used as a data source in the specific subjects of meteorological, ocean, land and atmospheric entities. The MERRA-2 (Version 2) is a multi-decadal reanalysis product released by the NASA Global Modelling and Assimilation Office (GMAO) in 2017. Monthly MERRA-2 data for Guwahati city was extracted in the year range of 2010-2021 and with the MERRA-2 tavg1_2d_slv_Nx: 2d database.

It is well known that the composting being a complex biochemical process is regulated by several variables. Notable among these are carbon-nitrogen (C/N) ratio, MC, interaction between oxygen and aeration, T, pH and raw material size. Invariably, compost production assessment requires an introspection into the influence of climatic parameters on the composting process performance. Among mentioned variables, physico-chemical factors such as C/N ratio, aeration and pH are often addressed in experimental investigations. Hence, these were not been considered as influencing variables in the conducted studies. Thus, the studies targeted the criticality of the climatic or meteorological factors in influencing the composition production rate or yield prior to the available historical data. Accordingly, the influence of seasonal variation on the CP rate was assessed using ML techniques.

CPCB, India does not directly provide the OW data. However, it provides the MSWG data nationwide. Therefore, OW data was achieved with the MSWG rate along with a factor that accounts for the maximum weight percent of OW in the overall MSWG (ranging from 34 to 53 wt%) (Singhal et al., 2022). This was based on the literature approach (Singhal et al., 2022) in which the OW data was obtained using the MSWG rate and a multiplication factor that accounts for the maximum weight

percent of OW in the overall MSWG. In most literature, the OW value varied from 34 – 53 wt%. Considering higher consumption of organic content by the city dwellers, the present study considered an OW multiplication factor of 50 wt% for the MSW generated in Guwahati city. Such an assumption was justified with the fact that the organic fraction of the generated MSW is highly complex and is often achieved through a detailed survey. However, contemplating on the fact that the multiplication factor can influence the optimal performance of alternate ML algorithms, the thesis also considered a case study that targeted random alteration of OW in the range of 40 – 45%. In summary, it was observed that the OW critically influenced the prediction accuracy of the ML algorithms. Moreover, monthly compost data was collected from OGD in the year range of 2010-2021. Data pre-processing, data integration and outlier detection and removal were elucidated in the previous sections of the thesis (2.3.2 and 2.3.3). The pre-processed data was segregated for three alternate seasons i.e., summer (April-June), monsoon (July-September) and winter (December-February).

2.6.2 Overall Methodology

The overall modelling framework implemented in this study has been depicted in Fig. 2.11. This involved various phases such as data gathering, pre-processing, modelling, and analysis. Several data pre-processing stages were required for the preparation and translation of raw secondary data into variables suitable for modelling and analysis. These include secondary data extraction, data loading into appropriate data structures, secondary data transformation, outlier removal using filtering processes, and integration into consolidated datasets. The MERRA-2 data were compared with the data collected from CPCB and OGD, and the data was compiled for the year range of 2010-2021. A random train-test split of a 70/30 ratio was adopted for the data. Accordingly, a data size of approximately 604 and 259 samples was achieved as training and testing datasets respectively. These datasets were deployed for model development and validation. Accordingly, the study targeted four alternate ML algorithms namely MLP, *k*NN and tree-based ensemble techniques (RF and GB) to predict the CP. Mathematically, the objective function CP of this approach is represented as:

$$CP = f(T, RH, P, WS, OW) + \varepsilon$$

To assist modelling efforts, preliminary screening of the parameters was implemented among CP, climatic parameters (T, RH, P, WS) and OW factor. To do so, the correlation coefficient being applied

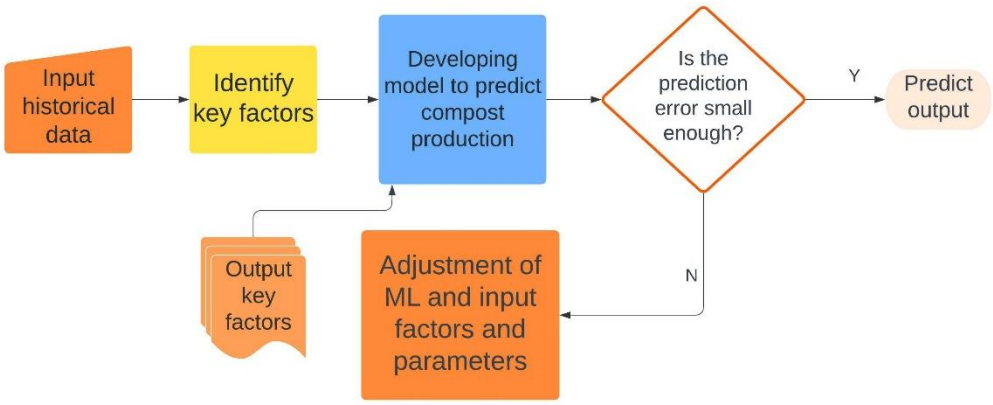


Fig. 2.11 Overall methodology for climate data-based CP rate prediction

in earlier investigations as a ranking technique to select ML algorithm features was adopted (Iguyon and Elisseeff 2003).

2.6.3 Deployed Algorithms, Model Evaluation and Forecasting Models

The modelling efforts involved the consideration of MLP (sub-section: 2.4.4.1) and traditional ML algorithms namely *k*NN (sub-section: 2.4.4.4), GB (sub-section: 2.3.4.3) and RF (sub-section: 2.3.4.2) for prediction and ARIMA (sub-section: 2.4.5) model supplemented ML models (hybrid modelling) for long-term forecasting of the compost generation rate. For model evaluation, performance indices such as R^2, RMSE, MAE and IoA were deployed. Also, for better accuracy, 10-fold cross validation technique was implemented (discussed in section 2.3.6 of the thesis).

2.7 Meteorological parameters and ML algorithms-based modelling of biogas generation rate prediction

2.7.1 Data selection

In the conducted investigations, emphasis was on the influence of meteorological variables (T, P, RH, Pr) and OW on BP rate. MERRA-2, a comprehensive reanalysis product by NASA, was utilized to obtain climatic data at a resolution of 0.5° × 0.625° and sin the year range of 2010 to 2021 and for Guwahati city. MSWG data from CPCB, India was used to estimate OW data along with a maximum OW weight % factor. Biogas data from 2015 to 2021 was collected, and data pre-processing techniques, including outlier removal using interquartile range filtering, were applied. The pre-processed data was

categorized into three seasons namely winter (December-February), monsoon (July-September), and summer (April-June) seasons.

The overall modelling framework being implemented in the conducted investigation has been depicted in Fig. 2.12. The MERRA-2 data were compared with the data collected from CPCB and OGD, and the data was compiled for the year range of 2015-2021. A random train-test split of a 70/30 ratio was performed on the data. Accordingly, a data size of approximately 510 and 218 samples was achieved as training and testing datasets respectively. These datasets were deployed for model development and validation. The conducted studies targeted five alternate ML algorithms namely MLP, SVR, *k*NN and tree-based ensemble techniques (RF and GB) for the BP rate prediction. The objective function has been represented as:

$$BP = f(T, P, RH, Pr, OW)$$

where BP, T, P, RH, Pr and OW refers to BP rate, T, atmospheric pressure, RH, precipitation and OW respectively. The significance and sensitivity of various meteorological factors and OW were analyzed using *r*, and with a considerable improvement in the prediction accuracy as a bias.

2.7.2 Deployed ML Models, Model Evaluation and Forecasting Model

The modelling efforts involved the consideration of MLP (sub-section: 2.4.4.1), SVR (2.4.4.3), *k*NN (sub-section: 2.4.4.4), GB (sub-section: 2.3.4.3) and RF (sub-section: 2.3.4.2) for prediction and ARIMA (section: 2.4.5) model supplemented ML models (hybrid modelling) for long-term forecasting of the BP rate. Also, for a more comprehensive analysis and comparison in terms of accuracy and reliability of the forecasted results, both MA (section 2.3.7) and ARIMA models were deployed. The MA model is useful for quicker insights and for initial forecasting pattern that especially involved the data with good stability and limited variation. On the other hand, the ARIMA model offers a more comprehensive approach that captures complex patterns and dynamics. Thus, it is suitable for forecasting in scenarios in which the data is non-stationary or exhibits trends and seasonality. For model evaluation, performance indices namely R^2, RMSE, MAE and IoA were employed. Also, for better accuracy, 10-fold cross validation technique was implemented (discussed in section 2.3.6 of the thesis).

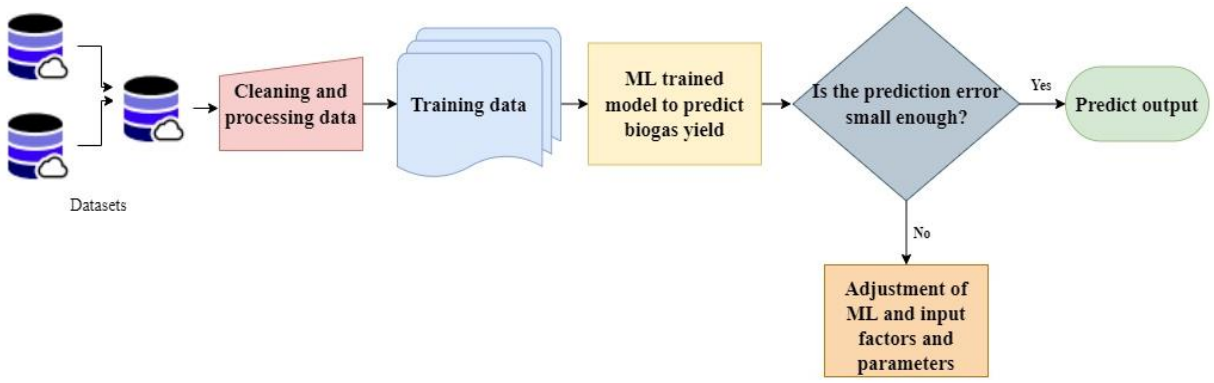


Fig. 2.12 Overall methodology for biogas generation rate prediction

2.8 Sizing and Costing Models for the Evaluation of Alternate Biogas Digesters and Predictive Modelling of Biogas Generation Rate

2.8.1 Overview of AD Calculator

The AD calculator comprised of two primary components. These refer to a biogas yield model and a cost or economic model. To estimate biogas yield, the BP model in the AD calculator utilized the VS rate. Few prior arts based on empirical studies (Rao et al., 2000; Li et al., 2018), inferred VS as a function of RT and T. Hence, the model proposed in this context predicts VS by considering the influence of T and RT as key factors. Such an alteration to the VS model was proposed by few other researchers as well (Rao et al., 2000; Li et al., 2018). Hence, the VS value used for computation was the weighted average of the two estimated VS values. Both costs and revenue of the two alternative approaches have been taken into account during analysis. Many site-specific details are needed for this module, including the employee numbers, gate fees, feed material transit expenses, etc. However, the tariffs and government incentives are subject to change and can be adjusted by the user. After estimating economic measures, they can be utilized to approximate the future costs and revenues of the AD unit in the long run. Thus, obtaining all relevant data, a decision can be made with respect to the operational status of the AD unit.

2.8.2 Targeted Perspectives

The AD calculator proposed in this study has been specifically designed to cater to the needs of AD units. Table 2.4 illustrates the typical configurations of commonly deployed AD units. A simplified

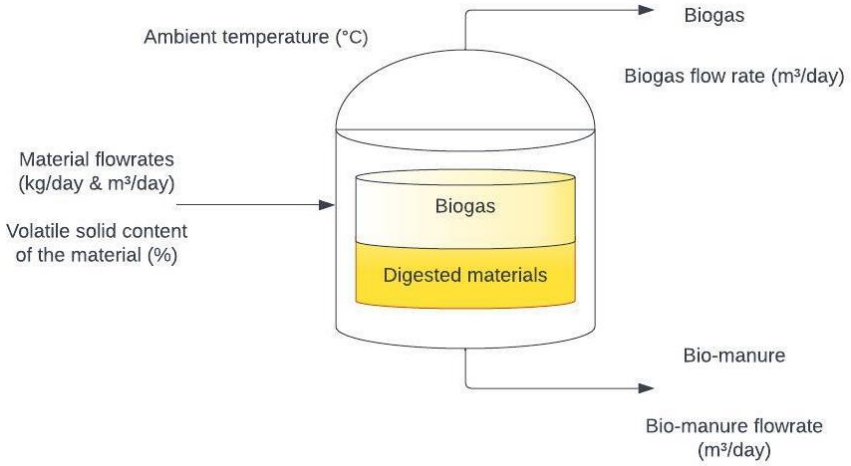


Fig. 2.13 Schematic of an AD unit and associated process variables

Table 2.4 A summary of the typical Indian AD unit configurations.

Design outline	Common approaches in India	Alternative (s)
Operating temperature	Mesophilic (25–45 °C)	Thermophilic (50–60 °C)
Wet or dry	Wet (5–20% dry matter in the digester)	Dry (>18% dry matter in the digester)
Flow of feed material	Continuous	Batch cycles
Number of digesters	Single/double	Multiple
Tank design	Floating drum/ Fixed dome	Balloon digesters

illustration of an AD unit with a typical combination of process inputs, outputs and parameters has been depicted in Fig. 2.13. In the digester tank, on a frequent basis, only few process variables could be measured. Thus, it is not financially feasible to install more measurement devices or laboratory equipment in such systems. This severely restricts the utility of more complex available BP models such as kinetic models that have been documented in the literature. The main focus of this thesis is to develop and apply a BP model that relies on readily available macroscopic data of the small to large-scale AD units. The concept behind the biogas yield evaluation is straightforward. Accordingly, process parameters can be accounted or modified to achieve alternate biogas yield values.

To aid preliminary calculations, the considered BP model reported data in terms of potential biogas yield. The accuracy of the uncalibrated model was evaluated with the reported case studies of AD units. For a period of one year, a farm-based company in Sonapur, Assam provided sample data. This was utilised to evaluate the BP model's appropriateness as a tool to benchmark BP for daily operations. The reported AD unit considers a variety of feed materials which includes rice straw, and OW mixture.

2.8.3 A Summary of Indian Biogas Generation Rate Models

Several alternate designs ranging from small to large scale biogas digesters are often deployed. Thus, the selection with respect to digester type is important to estimate its capacity, and especially where in cases that involve partly underground commissioning of the digester. Fixed dome plants are widely utilized in India as a cost-effective and efficient technology for the treatment of OW and producing biogas. They are especially relevant in rural areas in which abundant agricultural waste exists. The technology is recognized and supported by the Indian government and as a means to promote sustainable development and rural electrification. Floating drum and balloon digesters are two other types of biogas digesters commonly used in India for the treatment of OW. Floating drum digesters consist of a cylindrical digester with a floating gas holder that rises and falls with the BP rate. Balloon digesters, on the other hand, are made of a flexible material that expands as biogas is produced. These systems are popular in rural areas in which abundant agricultural waste exists along with the demand for renewable energy. A summary of the mentioned digesters is presented in Table 2.5.

Depending upon the type of substrate being digested, T, RT, pH, and other factors the BP rate varies. Therefore, it is important to determine the appropriate BP rate for the specific conditions of the digester being used. Daily BP can be estimated with the expression (Fulford, 2015):

$$BG_P = \frac{Y \times V_d \times S}{1000} \quad \textbf{(2.24)}$$

Table 2.5 A summary of the commonly deployed Indian biogas digestors.

Characteristic	Fixed dome digester	Floating drum digester	Balloon digester
Construction material	Concrete or brick	Mild steel	Plastic
Digester shape	Dome-shaped	Cylindrical	Spherical
Gas storage	Under the dome	In the drum	In the balloon
Mixing mechanism	Natural convection	Mechanical stirring	Natural convection
Maintenance requirement	Low	High	Low
Installation cost	Low	High	Low
Temperature range	20-45°C	20-40°C	20-45°C
Feedstock suitability	Solid and semi-solid	Solid and semi-solid	Solid and semi-solid
BP rate	Moderate	High	Moderate
Applicability	Suitable for small to medium-sized farms	Suitable for large farms and industrial use	Suitable for small to medium-sized farms

where BG_P is the biogas yield (in m^3 day^{-1}), V_d is the digester volume (in m^3), S is the initial concentration of VS in the slurry (kg m^{-3}) and Y is a yield factor based on T and the RT. The literature commonly reports fixed biogas yields, which are often used in preliminary calculations. However, it is acknowledged that operating parameters such as T (Risberg et al., 2013) and RT (Trisakti et al., 2015; Converti et al., 2009) can affect BP and hence the biogas yield. Fixed biogas yields are determined under specific operating conditions. Their utility presumes that the process operates with conditions similar to those reported. Also, limited case studies exist to verify the accuracy of AD calculators, and such studies are promising due to the variant operating conditions of AD units.

Several other empirical studies inferred that the biogas yield is influenced by RT (Rahimi et al., 2019; Singh et al., 2020) and a reduced volatile solids (RVS) rate serving as an intermediate term. RVS refers to the percentage of VS entering the digester and eventually consumed by the microbes. Typically, a plot of biogas yield against RT depicts a logarithm relationship. It is assumed that RVS and BP have a linear correlation. Another approach is to assume a theoretical maximum biogas yield (potential biogas yield), with RVS being a fraction of this maximum yield. This leads to the following formula for biogas yield:

$$Y = Y_p \times RVS \tag{2.25}$$

where Y_p represents the potential biogas yield that can be obtained from the feed material in terms of m^3 biogas kg^{-1} VS and RVS is the reduced VS (%).

The biogas yield determination is a simple concept to apply, and is easy for non-experts. However, the biogas yield is dependent on RVS, which is in turn affected by operating temperature. RT is determined with volume feedstock rate and the active volume of the digester tank, and as follows:

$$RT = \frac{V_d}{V_f} \tag{2.26}$$

where V_d is the digester volume (m^3) and V_f is the total feedstock volume (m^3 day^{-1}).

Few published literatures did not consider the effects of operating temperature on BP in the biogas yield modelling effort. However, kinetic models did consider temperature effects and are more sophisticated than the biogas yield models. Additional terms have been included in kinetic models to account for the concentrations of various inhibiting compounds (Angelidaki et al., 1999), feed materials represented as organic component groups instead of VS (Tomei et al., 2009), and the effect of operating temperature. While kinetic models are useful for academic purposes, it is challenging to apply them to AD units due to the lack of continuous measurements within the digester. While literature-reported parameter values can be used for preliminary estimations, they may not be applicable to situations sensitive to climate, T, and feedstock variations. Furthermore, kinetic models are more complex and difficult for non-experts to comprehend and maintain.

2.8.4 Comparative Studies with other AD Calculators

The general aim of the AD calculators is to provide a simple tool for the unit owners to assess the alternate modes of operation. Thereby, the most profitable options can be explored. However, differences exist in BP estimates and in the associated sensitivity of process variables. The notable feature in literature proposed model is that it was achieved through the inclusion of the farmland data. Thus, the farm business was centred around the AD unit, and the farmland functioned as a support facility to the AD unit. Comparatively, the AD calculator reported in this article has similar utility of farmland in comparison to another prior art (Jain and Sharma, 2011). However, the AD model has been facilitated with animal or agricultural waste feedstock at the farm or nearby areas.

2.8.5 Biogas Yield Estimation

2.8.5.1 Sizing of Biogas Plant

It is important to estimate the dimensions of biogas plants and determine their total volume for alternate sizes and types. Also, it is important to determine the division between the digester and gas storage volumes. Table B1 in the appendix provides a summary of average values for these volumes, and with the multiplication factors being used to estimate the total plant volume based on the rated daily gas production of different types of plants. The table assumes that the gas storage volume is typically designed to hold 60% of the rated daily gas production. However, these values may have to be adjusted based on local biogas plant designs and expert knowledge.

i. Fixed Dome Plant (Deenbandhu Design)

The Deenbandhu digester design is a combination of a hemisphere and a curved base. According to specifications provided by Kalia and Singh (2010) for various sizes of Deenbandhu digesters, the depth of the curved base (k) is typically around 40% of the digester radius (r). To determine the volume of a Deenbandhu digester, it is only necessary to measure the maximum diameter (D) of the digester using the same method described for the hemisphere design.

ii. Floating Drum Plant

Floating drum biogas plants are typically circular in shape and can be designed as the volume of two cylinders. While the larger cylinder represents the digester, which can be partly or entirely underground, the smaller cylinder represents the gas storage tank, which is situated on top and within the structure. To calculate the volume of a floating drum biogas plant, both diameter (D) and height (H) of the digester must be measured. The diameter of the digester can be measured directly if a tape can be placed across the top section of the gas storage tank, or alternatively, it can be calculated by measuring the circumference of the digester or gas storage tank and dividing it by 3.14 (π). If the digester is partially or completely underground, the height of the digester can be estimated by multiplying the height of the gas storage tank with a factor of 2.3. The gas storage volume in a floating drum design typically accounts for 30% of the total plant volume (Kabyanga et al., 2018). Therefore, it is necessary to calculate both digester and gas storage volumes to accurately estimate the total plant volume. If the digester is entirely above ground, the height of the digester should be measured from its base to the top of the digester wall.

iii. Balloon Digester

The shape of a fully expanded balloon digester is similar to that of a cylinder. The total plant volume can be calculated as the volume of a cylinder, and the digester volume is typically around 75% of the total plant volume, (25% gas storage volume). To determine the volume of a balloon digester, the diameter (D) and length (L) of the digester must be measured (Kabyanga et al., 2018 and Kumar et al., 2015). The diameter can be measured from top to bottom and across one of the ends of the digester, or across the top of the bag from seam to seam, and then divide it by 1.57. The length should be measured from one end to the other, excluding any parts that would form the ends of the cylinder upon full

expansion. The volume calculations of the considered digesters are summarized in Table B2 in the appendix.

2.8.5.2 Base Model Configurations

The model for biogas yield proposed in this study relies on empirical models previously reported in research studies (Shailendra et al., 2016; MEDA, 2012). These studies also report a logarithmic correlation between RT and RVS. Other published articles have shown that the relationship between biogas yield and RT is also logarithmic, but with a notable curve at low RTs (lesser than 5 days) (Zhang et al., 2018; Wu et al., 2016). The disparity in results is likely due to the organic composition of the material being digested, as it can affect the breakdown of certain organic constituents.

2.8.5.3 Temperature Sensitivity

The temperature inside a digester plays a crucial role in determining the rate of gas production per day and the total amount of biogas that can be generated from a given feedstock. Microbes that produce biogas function best at a temperature range of 20-45°C in small-scale biogas plants. To maximize daily gas production, the size of a biogas plant is often designed accordingly. In colder climates, the slurry is heated to maintain the temperature within this range. For this purpose, at times, produced biogas is used.

2.8.5.4 Summary of Model Parameters Values

In the preceding sections, multiple model parameters have been introduced. Summarized parameter values and the reference for these parameters are presented in Table B3 in the appendix and the parameters can be adjusted by the user to determine Y (refer to Equation 2.25). For preliminary estimations, Table B4 in the appendix includes literature-reported values for potential biogas yield. However, it is recommended to verify these values to eliminate practical uncertainties in the calculations. Thus, it is important to note that the properties of the feed material can vary widely from one site to another and may require prompt adjustments.

2.8.6 Estimation of Economic Parameters

The latter portion of the AD calculator focuses on the utilization of the generated biogas and provides estimations of the associated operational costs and revenue for each scenario. This section outlines the

inputs considered for the calculations and the subsequent evaluations. The following sub-sections delineate upon relevant parameters evaluation.

2.8.6.1 Feed Cost

The feed material value refers to the cost incurred in acquiring the feed material (also includes the transportation cost). Generally, the feed material is sourced from the farm, but it may also include biodegradable waste from nearby facilities. In some instances, AD owners receive a gate fee to process certain feeds (can be considered as a form of income). Thus, the estimated feed cost for such cases would be negative.

2.8.6.2 Total Capital Cost

To estimate the total capital investment required for the setting up of a biogas plant, the fixed capital expenditure (CAPEX) and working capital investment were added. Working capital investment typically represents 15% of the CAPEX and is required to initiate the plant (Towler and Sinnott, 2008). The CAPEX includes various costs associated with plant construction, such as equipment purchase, installation, piping, and other expenses related to the building and electricity. The factorial method was used to estimate the CAPEX. This involved multiplication of the purchased equipment price with different factors based on the commissioning expense categories of the plant. Manufacturing costs for each plant were determined using specific expressions:

$$Cost_{Manu} = \frac{\text{Rate of construction}}{\text{Surface area of the plant}} \quad \textbf{(2.27)}$$

$$CAPEX = \text{Cost of land} + Cost_{Manu} + \text{Labour cost} \quad \textbf{(2.28)}$$

$$\text{Labour cost} = \text{Number of workers} \times \text{working rate of labour} \times \text{number of days} \quad \textbf{(2.29)}$$

2.8.6.3 Operating Cost

The production operational costs comprise fixed and variable costs. Fixed costs include maintenance, operating labor, supervision, direct salaries, general overheads, land and building rent, property taxes, insurance, and environmental charges. According to Zupančič et al. (2012), a large-scale biogas plant requires six operators (three of whom are truck drivers). Their services are provided during working

hours, with one person on call in the evening and night, as reported by Thompson et al. (2019). The operator-related costs were estimated based on a 12-hour working day and a salary of ₹55 h^{-1} ($0.6 h^{-1}) (suggested by Ministry of New and Renewable Energy, MNRE (2018)). The relevant expressions are:

$$OPEX = \text{Actual working cost} \times \text{Average working time} \times 365 \text{ (days)} \quad \textbf{(2.30)}$$

$$\text{Actual working cost} = CAPEX + OPEX \quad \textbf{(2.31)}$$

2.8.6.4 Revenues

The government provides financial support to biomethane through a feed-in-premium subsidy, which is a fixed add-on to the market price. The selling price of biomethane can be estimated by taking into account the old subsidy and the natural gas price. The primary objective of this study is to determine the minimum selling price of biomethane and the corresponding subsidy. The revenues (R) were obtained by multiplying the volume of biomethane produced ($V_{Biomethane}$) with the price of biomethane that includes the subsidy ($P_{Biomethane}$).

$$R = V_{Biomethane} \times P_{Biomethane} \quad \textbf{(2.32)}$$

The computational evaluations were based on the presumption that the biomethane being evaluated as a fraction of the biogas would be available for commercial applications. While such a presumption may be relevant for large digesters with affordable auxiliary technologies, the same may not be true for the small digesters, given the fact that hydrogen sulphide produced from small digesters are harmful due to enhanced SOx emissions. Thus, it is presumed that either relevant affordable technologies are available for the conversion of biogas into biomethane for the small digesters or the small digester combust the biogas for commercial applications such as community kitchen cooking purpose. Either way, the small digester issue needs to be resolved at the experimental R&D level and is beyond scope of the Ph.D. thesis.

2.8.6.5 Profitability Analysis

Six metrics were used to evaluate the profitability of the plant, which are net profit, benefit-cost ratio, net present value (NPV), payback period (PBP), gross benefit, and the value of gas. The NPV represents the present economic value of the project. It considers the time value of money during the lifetime of the project. The cash flow for year *n* (for a maximum of N) is denoted as CF_n in the equation. A positive NPV indicates that the project is economically feasible. The relevant expression is:

$$NPV = -TCI + \sum_{n=0}^{N} \frac{CF_n}{(1+r)^n} \tag{2.33}$$

The PBP expresses the necessary period for full investment recovery. The gross benefit, net profit and benefit cost ratio are evaluated with expressions:

$$PBP = \frac{\text{Actual working cost}}{\text{Net profit}} \tag{2.34}$$

$$\text{Gross Benefit} = \frac{\text{Value of gas}}{\text{Additional value of fertiliser}} \tag{2.35}$$

$$\text{Net profit} = \text{Gross Benefit} - OPEX \tag{2.36}$$

$$\text{Benefit cost ratio} = \frac{\text{Net profit}}{OPEX} \tag{2.37}$$

2.8.7 ML based Prediction and Forecasting of Biogas Generation Rate

A comprehensive range of ML models, including hybrid model (ARIMA-MLP), MLP, and SVR were considered for prediction and were discussed in sub-sections 2.4.5, 2.4.4.1 and 2.4.4.3 respectively. Additionally, for the long-term forecasting of biogas yield, hybrid modelling (ARIMA-MLP) as discussed in section in 2.4.5 was deployed along with the MLP. The model evaluation deployed performance indices such as R^2 and RMSE (Breiman et al., 2017). Furthermore, for improved accuracy, the 10-fold cross-validation technique, discussed in section 2.3.6, was implemented.

2.9 Techno-economics of Alternate Compost Facilities and Prediction and Forecasting of Compost Generation Rate

2.9.1 Scope of the Study

In order to effectively develop waste conversion facilities and devise a sustainable waste management strategy, it is essential to consider economic competitiveness as a key factor. This requires a thorough analysis of the economic feasibility and profitability of waste conversion technologies and management practices. Such methodologies would likely involve the evaluation of several factors such as waste conversion facilities, capital and operating costs, revenue streams, market demand for waste-derived products, potential environmental impacts, and regulatory compliance requirements. With the incorporation of the economic competitiveness into decision-making processes, the stakeholders can ensure that the chosen waste conversion technologies and management practices are cost-effective and financially sustainable. Such an approach can assist in the maximum resource recovery and reduce waste disposal costs. Along with this, it will minimize negative environmental impacts and enables compliance with relevant regulations. Thus, in the end, an emphasis on economic competitiveness can lead to the development of more efficient and effective waste conversion facilities and sustainable waste management strategies.

The goal of the conducted study was to develop a framework for the development and demonstration of a comprehensive techno-economic model and thereby assist municipal decision makers to quickly establish waste conversion facilities. The methodology involves needful information for the design of a compost calculator tool and henceforth allows facility owners and businesses to estimate the amount of compostable materials they generate and the potential compost volume they could produce. To aid preliminary calculations, the CP model considers the data on potential compost yield. The accuracy of the uncalibrated model was evaluated for the reported case studies. The sample data was collected from a large-scale composting plant being operated by GMC. The data of the plant was utilised to evaluate the competence of the CP model and as a benchmark tool for CP. The reported compost plant considers a variety of feed materials which includes HH's OW, mixed vegetable waste and paper waste.

2.9.2 Site Selection

Along with waste abundance, the suitable and optimal location of a waste conversion facility depends upon several factors such as environmental, social, and economic factors. In this study, site selection was performed in two stages that involved an exclusion and preference analysis (Sultana and Kumar, 2012). Considering social and environmental factors, the exclusion analysis screens out unsuitable lands from the study area. These have been outlined in Table 2.6. For every 8 constraints, a buffer zone was created and the areas inside and outside the buffer zones were assigned *0* and *1* values, respectively. Accordingly, a binary map was generated for every constraint. A final constraint map was prepared by multiplying all binary values from the whole maps.

The subsequent preference analysis was performed to identify the relative preference of various regions within the study area. Eight factors were considered to find the most preferable sites for a waste conversion facility plant. These refer to: (i) protected region (ii) railways (iii) buildings (iv) land-use (v) urban settlement (vi) waterways and (vii) roadways. The appropriate selection was based on expert opinion in the relevant literature (Sultana and Kumar, 2012; Tavares et al., 2011). The weights of the preference factors were determined with the AHP (Saaty, 2000). The AHP involves the generation of a hierarchy of decision criteria and alternatives, and subsequent assignment of numerical values to each criterion and alternative based on their relative importance or performance. The process involves pairwise comparisons between criteria and alternatives. To do so, a scale to measure the relative importance or performance of each item in comparison to others was used in the same level of the hierarchy. After completing the pairwise comparisons, the AHP generates a set of weights for each criterion and alternative. These can be used to evaluate and rank alternatives as per their overall performance or desirability. Using the AHP, the preference factors were compared with one other and each assigned factor was assigned a value on a 9-point scale. The weight of each factor was evaluated with these assigned values.

Thereafter, multiple buffer zones were created around each preference factor, and scores (on a scale of 0–10) were assigned to the buffer zones. Such an assignment was dependent on their distance from the corresponding factor. Later, they were multiplied with the corresponding weights to determine the

relative preference of the corresponding region of the study area. In this study, places with a suitability index (a value that indicates the suitability of each location in the map by considering the criteria entered into the model) of 7, 8, 9, and 10 were opined as suitable sites for a waste conversion facility. Fig. 2.14 depicts the distribution of land cover types (such as forests, grasslands, wetlands, urban areas, etc.) and land use activities (such as agriculture, forestry, mining, recreation, etc.) in the Kamrup region. Such pictorial representation assists the decision-makers to understand the current status and trends in land use and land cover (LULC), identify areas of high ecological value or potential environmental risk, and develop strategies for sustainable land use and management for waste conversion facilities.

Table 2.6 Identified constraints and corresponding buffer zones for the site selection of composting facility.

Criteria	Specifications	References
Rivers, lakes, and other water bodies	Greater than 50-200 m from small to large water bodies	Liu et al. (2017)
Rural and urban areas	Greater than 1-2 km from residential and urban areas	Sumathi et al., 2008
Industrial zones	Greater than 1-2 km from industrial zones	Sumathi et al., 2008
Environmentally sensitive areas (ESA) (conservation areas, habitat sites)	Greater than 1 km from ESAs	Eskandari et al. (2012)
Park and recreational areas	Greater than 500 m	Karimi et al., 2019
Wetlands	Greater than 200 m	Karimi et al., 2019
Roads	Greater than 30 m	Karimi et al., 2019
Railways	Greater than 300 m	Karimi et al., 2019

2.9.2.1 Overview of the Compost Calculator

The compost calculator constitutes the following four primary modules:

i. *Input data module:* The module requires information with respect to the feedstocks used, such as the quantity and composition of organic materials.
ii. *Process module:* In this module, the composting process has been modelled. Thus, key parameters such as temperature in the entire composting cycle have been predicted.
iii. *Output module:* The module allows the determination of the quantity and quality of the final compost product.
iv. *Economic module:* The costs and revenues associated with CP rate, including labour, equipment, and marketing factors have been estimated in this module.

The proposed model in this context takes into account the effects of T and RT as important factors for CP volume prediction. The analysis considers both costs and revenue of two alternative approaches. To accurately estimate economic measures, several site-specific details such as employee numbers, gate

fees, and feed material transit expenses must be taken into account. However, the module allows the user to adjust for the alterations in tariffs and government incentives. After estimating economic measures, future costs and revenue of the compost facility unit during the long term can be determined. Through the assimilation and analysis of all relevant data, the module can assist in the decision making of the operational status of the compost facility.

2.9.2.2 A Summary of Indian Compost Production Systems

Alternate and variegated CP facilities operate in India. These range from small-scale HH composting systems to large-scale industrial composting systems. The capacity of a compost facility can vary widely due to various factors such as the size of the facility, the types and amounts of OW being processed, the composting method used, and the processing time required to produce high-quality compost. The solid waste composting facility was based on two typical designs for the MSW. These are a medium scale MSW composting facility (MSCF) and a high scale MSW composting facility (HSCF). In addition, a typical yard waste composting facility (YWCF) was also considered. Apart from compost utility for the soil amendment, the MSCF and HSCF produces a by-product that can be used as a landfill cover or can be directly landfilled in a *final storage* or *zero emission* landfill. However, for the direct landfilling of the composted product, indirect benefits still exist, as waste volume reduction is achieved and the landfill with significantly reduce potential for emissions. Such a system will probably require lower construction and monitoring costs in comparison to a typical MSW landfill. In order to produce high quality compost, the MSW was assumed to consist organic components such as food waste, mixed paper and yard wastes.

The HSCF produces compost for utility in agriculture, landscaping, and land reclamation operations. The facility assumes that the preprocessing did already remove non-compostable items and recyclables. The composting process begins with a horizontal hammermill and water is added to achieve 50% moisture. Composting takes place in windrows, and the material is turned with a windrow turner. An odour-control system using biofiltration is utilized, and the compost is cured in piles (approximately 2.7 meters high). A postprocessing trommel screen is used to produce a finer compost fraction suitable for application. Reject materials are temporarily stored and then transferred to a landfill.

The purpose of MSCF is to reduce the volume of MSW and the amount of organic matter that can quickly degrade prior to landfilling. This reduces emissions of gas and leachate associated with landfilling. The process involves the utility of a trommel screen to remove bags and large items. This is followed by a horizontal hammermill that shreds the smaller fraction. Also, water is added to achieve a MC of 50% (by weight), and the waste is composted in windrows that are turned with a windrow turner. Biofiltration is used to control odour. Thus, the resultant composted MSW and reject materials are transported directly to a landfill and without curing. The YWCF accepts yard waste from residents or from dedicated vehicles. Plastic bags with leaves and grass clippings are manually opened and removed. The facility uses a tub grinder to shred branches and no water is added as yard waste usually has high MC. Composting and curing takes place on a windrow pad, which is turned monthly with a front-end loader. A postprocessing trommel screen produces a fine fraction for the potential marketing of the compost. The composting pad is uncovered, and no odour-control system is installed.

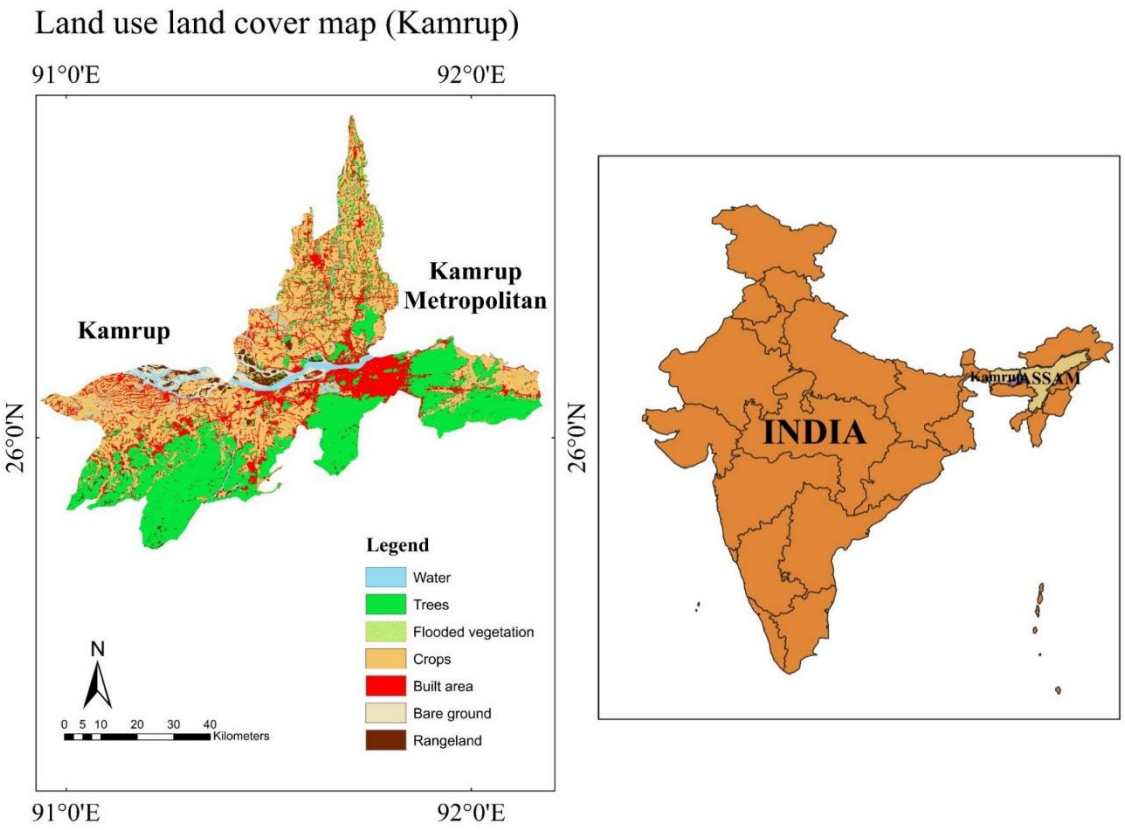
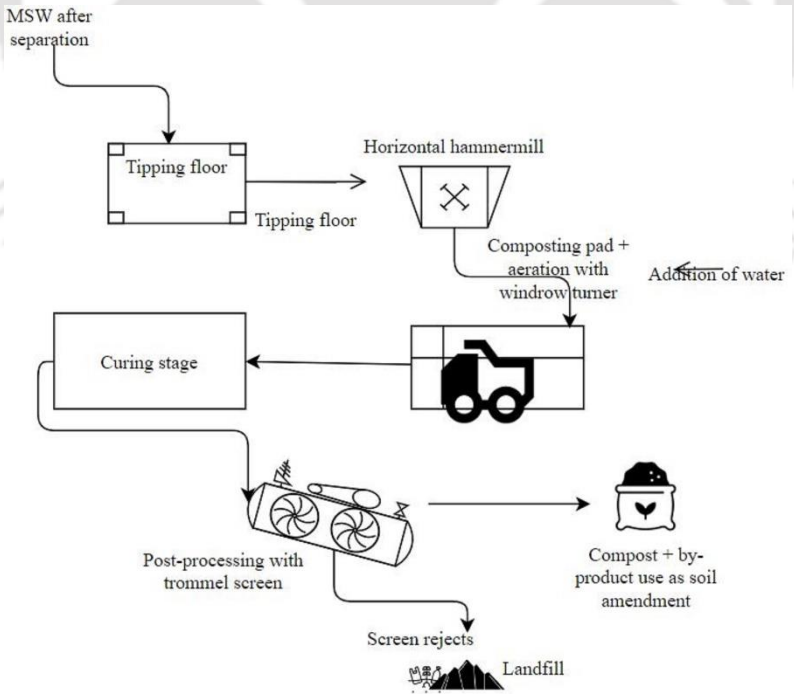


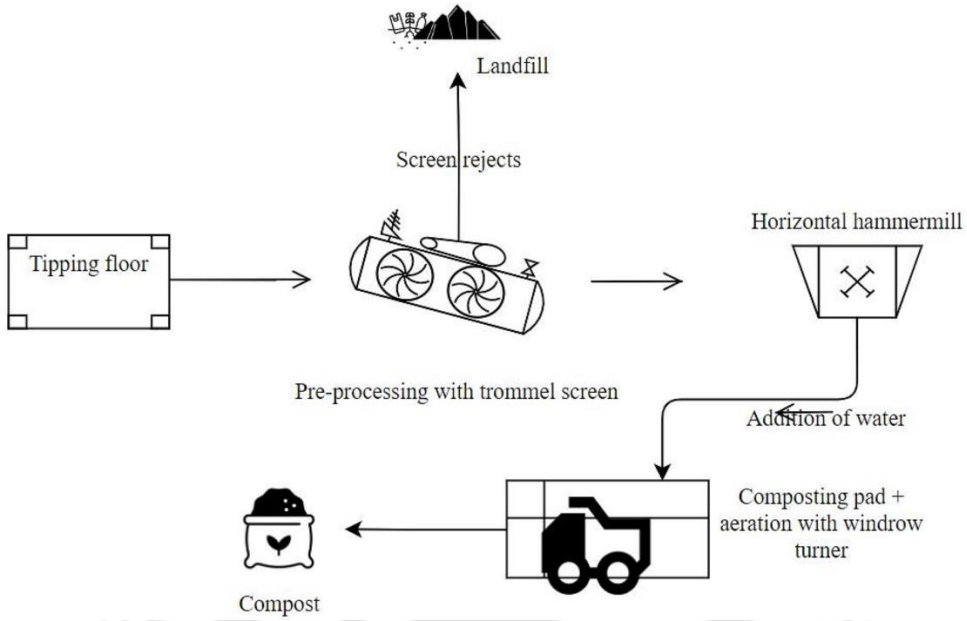
Fig. 2.14 LULC map for Kamrup, Assam

The process flow diagrams for the mentioned three types of facilities being considered in this study have been shown in Fig. 2.15 (a-c). The designs were partly based on relevant prior art (Zurbrügg et al., 2004) and can be regarded as typical windrow composting facilities. The overall facility area comprised the tipping floor, treatment area (for trommeling and shredding), the composting pad, the curing area, buffer zone, offices, roads, and storage of reject material and equipment (Diaz et al. 2007). Design guidelines in the literature for compost windrow turner vendors were used to calculate the composting pad area.

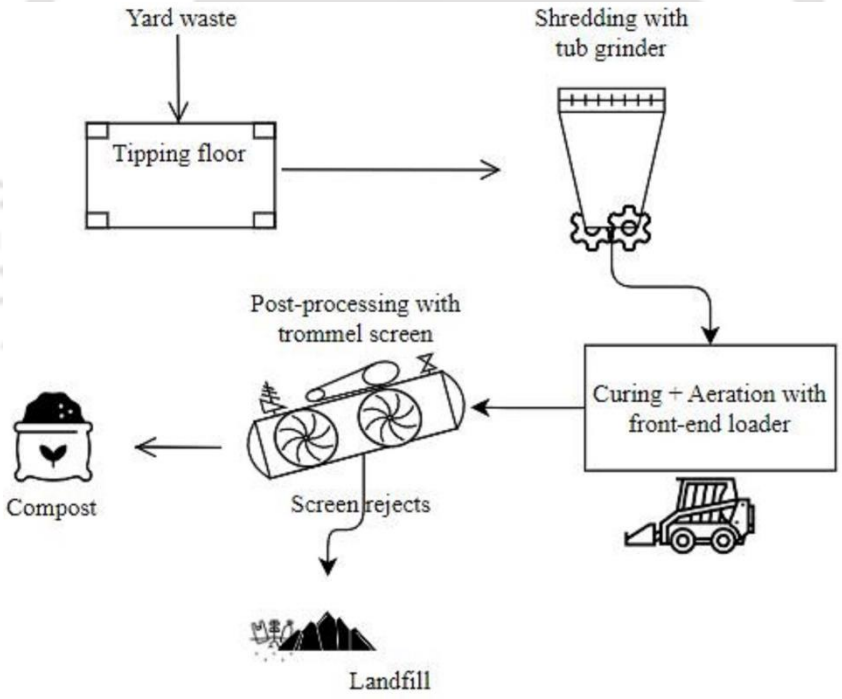
The composting pad was enclosed for both types of MSW composting facilities. Such a system has been facilitated with the ventilation (fans continuously directing the air to an odour control system). The number of selected fans quantify their capacity to draw air from the enclosed composting pad as well as the tipping floor. For the purpose, a building height of 4.5 m was used and an air exchange rate of 12 times daily (every 120 min) was considered. The power requirements were based on Kong et al. (1996), who provided design parameters for a biofilter to treat styrene latent industrial emissions. The odour control system was assumed to operate 24 h a day.



(a)



(b)



(c)

Fig. 2.15 Process flow diagrams of **(a)** HSCF **(b)** MSCF and **(c)** YWCF composting systems

2.9.2.3 Comparative Study with other Compost Calculators

The primary objective of the compost calculator module is to serve a straightforward tool to the facility owners for the evaluation of alternate operating schemes. Thereby, they determine the most profitable

option. However, in due course of such analysis, variations can exist in the estimation of CP and in the sensitivity of process variables. The proposed (uncalibrated) model has been compared with three existing calculators in Indian scenario (Compost Calculator Tool, IIT Delhi; Composting Financial Calculator; MNRE and Solid Waste Management Composting Model, Center for Environmental Planning and Technology (CEPT) University in India) and other prior data. The proposed tool aids in the selection of specific composting project (small scale to large scale). This is not the case in the existing calculators that customize decisions for only large-scale facilities. Also, the model considers organic feedstock to achieve high quality compost. It also considered the adaptation to the local conditions in India, including factors such as feedstock availability, mass flow, types of equipment, various cost estimations and revenues. Thereby, fair and accurate estimates have been targeted.

2.9.3 Evaluation of Compost Production Rate

2.9.3.1 Design of Specific Elements in the Compost Facilities

The design approach for the key elements in all three types of compost facilities has been presented briefly in the following sub-sections. While differences exist in the designs of these facilities, the approach allowed for the part-time usage of the equipment and assumed a linear correlation (with an intercept) of all design parameters with the waste flow rate. As a result, irrespective of the waste flow rate, a minimum number of units was assumed to exist for each type of facility, and to facilitate more accurate comparisons with data from existing solid waste composting facilities. Such assumptions prompted the precise evaluation of the design parameters and ensured the applicability of the resultant design for a wider range of waste flow rates.

i. Material Flows

Material flows for compost facilities do vary and depend upon the specific processes and equipment being used along with the types of feedstocks. Generally, incoming waste is received at a tipping floor and was sent to the composting pad after sorting, screening and shredding. At the pad, windrow turners or front-end loaders enable mixing and aeration of the waste and thereby promote the decomposition process. After an appropriate retention time, the compost is moved to a curing area, that facilitates maturity of the compost. Reject material, such as plastic and other non-compostable items, is typically removed and either sent to a landfill or further processed for potential recovery. The final product i.e.,

high-quality compost, can be used for various purposes, including soil amendment and erosion control (Diaz et al., 2002; Singh et al., 2018; Chen et al., 2019). At the tipping floor, the relevant material balance expression is:

$$\left(dry\ mass_1\right)_i = (1 - moist_i) \times \left(mass_1\right)_i \quad \textbf{(2.38)}$$

where $\left(dry\ mass_1\right)_i$ is the dry mass waste component *i* entering the facility on the tipping floor (tpd); $moist_i$ MC of waste component *i* at tipping floor and measured in weight %; $\left(mass_1\right)_i$ is the wet mass of component *i* entering the facility tons^{-1} day^{-1} (tpd); and *i* is the waste component, where *i* refers to either OW or paper waste or yard waste. The dry mass entering the facility can be computed as:

$$dry\ mass_1 = dry\ mass_{organicwaste} + dry\ mass_{paperwaste} + dry\ mass_{yardwaste} \quad \textbf{(2.39)}$$

where $dry\ mass_1$ is the initial dry mass flow rate of incoming waste streams at tipping floor (tpd). The total moisture balance expression is:

$$moist_1 = \left(\left(mass_1\right)_{organicwaste} \times moist_{organicwaste} + \left(mass_1\right)_{paperwaste} \times moist_{paperwaste} + \left(mass_1\right)_{yardwaste} \times moist_{yardwaste}\right) \quad \textbf{(2.40)}$$

where $moist_1$ is the initial MC of daily waste entering the facility. Other balance expressions are:

$$vol_1 = \left(vol_1\right)_{organicwaste} + \left(vol_1\right)_{paperwaste} + \left(vol_1\right)_{yardwaste} \quad \textbf{(2.41)}$$

$$mass_3 = \left(mass_3\right)_{organicwaste} + \left(mass_3\right)_{paperwaste} + \left(mass_3\right)_{yardwaste} \quad \textbf{(2.42)}$$

$$dry\ mass_3 = \left(dry\ mass_3\right)_{organicwaste} + \left(dry\ mass_3\right)_{paperwaste} + \left(dry\ mass_3\right)_{yardwaste} \quad \textbf{(2.43)}$$

$$mass_4 = \frac{dry\ mass_4}{1 - compost_{moist}} \quad \textbf{(2.44)}$$

where vol_1 is the total volume flow rate of wastes entering the tipping floor (m^3 d^{-1}); $mass_3$ is the total waste wet mass flow rate entering the composting pad after shredding (tpd); $dry\ mass_3$ is the total dry mass flow rate entering the composting pad after shredding (tpd); $mass_4$ is the wet mass flow rate of

MSW at the end of the curing pad (tpd) and $compost_{moist}$ is the moisture of MSW at the end of curing (default value of 40% wet weight basis (Diaz et al., 2002)).

ii. Trommel Screens

The MSCF utilized a precomposting trommel screen with a 12-cm opening to remove large items. In the HSCF and YWCF systems, since recycling and preprocessing of wastes were carried out prior to the waste entry into the facility, no precomposting screen was used. Thereby, the wastes were directly shredded by the hammermill prior to the composting. For each of the three components, precomposting trommel screen efficiencies were determined in terms of the particle size range (Tchobanoglous et al., 1993) and experimental data from Alter (1983). The prescreening efficiencies were determined for each MSW component and as the fraction of waste entering the trommel leaving as fines. The composting efficiencies, MC, and bulk densities for mixed paper (10.2, 95, 58%), yard waste (60.0, 122, 79%), and food waste (70.0, 594, 79%) were also considered. The postscreening efficiencies for the other two facilities were based on the data given in a relevant prior art (Chanakya and Sreesha, 2012). Diaz et al. (2002) proposed a gross specific energy consumption of 1.1 kWh ton^{-1} for precomposting trommel screens and 0.8 kWh ton^{-1} for the screening of the light fraction of solid wastes (assumed for the postprocessing trommel screen). In this work, for the evaluation of energy requirements, the former value was applied to the preprocessing screens of the MSCF, and the latter value for the postprocessing screen of the HSCF and YWCF. At least one trommel screen unit was utilized in all facilities, and a coefficient of 0.0025 units (tpd) was assumed for the estimation of the partial number of units required for different input flow rates, (assuming that one trommel screen unit is necessary for the processing of 50 $tons^{-1}$ $hour^{-1}$ tph at 8 hours day^{-1}). Thereby, relevant expressions in the system are:

$$Pre\text{-}trommel_{hp} = Pre\text{-}trommel_{hp\_coeff} \times \left(\frac{mass_2}{operating\ hours\ day^{-1}} \right) \quad \textbf{(2.45)}$$

$$Post\text{-}trommel_{hp} = Post\text{-}trommel_{hp\_coeff} \times \left(\frac{mass_4}{operating\ hours\ day^{-1}} \right) \quad \textbf{(2.46)}$$

where $Pre\text{-}trommel_{hp}$ and $Pre\text{-}trommel_{hp\_coeff}$ are the preprocessing and postprocessing screening required horsepower (hp) respectively. $Pre\text{-}trommel_{hp\_coeff}$ and $Post\text{-}trommel_{hp\_coeff}$ are the unit gross energy requirements per ton of waste treated (hp.tph) (with default value of 1.47 hp.tph and 1.07 hp.tph respectively (Diaz et al., 2002)).

iii. Hammermill/Tub Grinders

Typically, the operating hammermill hp requirement for municipal solid waste (MSW) shredding is about 20 hp h ton^{-1} (Tchobanoglous et al., 1993). To achieve a final product size of 2 inches, an additional energy requirement of 1.64 times the mentioned value is necessary (Tchobanoglous et al., 1993). This is due to its optimal particle size for the organics composting process (Diaz et al., 2007). For the case of HSCF, the energy requirement has been reduced through the multiplication of the above energy requirement product (20 x 1.64) with a factor of 0.65. The factors refer to the MSW entering such facilities in comparison to the MSCF (Tchobanoglous et al., 1993). Thus, less energy is required to shred the pre-sorted MSW to the same particle size in comparison to the shredding of unsorted MSW. For MSW facilities, at least one hammermill unit is used, and a coefficient of 0.0029 units (tpd) was assumed to estimate the partial number of units required for different input flow rates. This is based on the assumption that 1 hammermill unit is required for the processing of 42.5 tph and for 8 hours day^{-1}. The maintenance cost for a hammermill is relatively inexpensive since it involves the weekly build-up of the hammers (Diaz et al., 2002). For tub grinders, the energy requirements have been determined to be 13.7 hp h ton^{-1}. This was based on a linear regression of hp and the corresponding input mass flow rates (in tph) from various tub grinder models. Accordingly, power expressions for the system is:

$$Hammer_{hp} = 20 \times \left(\frac{mass_3}{operating\ hours\ day^{-1}} \right) \times input_{coeff} \times size_{coeff} \qquad \textbf{(2.47)}$$

where $Hammer_{hp}$ is the required net power of hammermill (hp); 20 is the power coefficient (hp.tpd^{-1}), $input_{coeff}$ is the input material factor (default value of 1.00 and 0.65 for the MSCF and HSCF, respectively and the values correspond to municipal solid wastes and presorted municipal solid wastes, respectively); $size_{coeff}$ is the product size factor (1.64 for both facilities). Also, $Hammer_{coeff}$ coefficient

refers to the number of hammermills to incoming waste flow rate (default of 0.00294 units.tpd). The number of hammer mills has been estimated with the expression:

$$Number\ of\ \ operating\ \ hammer = Hammer_{coeff} \times mass_1 + 1 \quad \textbf{(2.48)}$$

The number of operating tub grinder and their power can be expressed as:

$$Number\ of\ \ operating\ \ tubgrinder = Grinder_{coeff} \times mass_1 + 1 \quad \textbf{(2.49)}$$

$$Tubgrinder_{hp} = 13.7 \times \left(\frac{mass_1}{operating\ \ hours\ \ day^{-1}} \right) \quad \textbf{(2.50)}$$

where $Grinder_{coeff}$ is the coefficient that relates the number of tub grinders with waste flow rate (default of 0.00385 for the YWCF); $Tubgrinder_{hp}$ is the yard waste composting facility's tub grinders required gross hp and 13.7 is the coefficient that relates hp to waste capacity (hp.tph)

iv. Windrow Turner

To establish a correlation between the waste turning capacity (in tph) of a windrow turner and its hp, four values ranging from 177 hp to 450 hp were utilized. These correspond to waste handling capacities ranging from 500 tph to 2000 tph. A coefficient of 0.173 hp h ton^{-1} was assumed as the basis for the evaluation of the energy requirements of the equipment. It should be noted that the typical range of waste flow rates reported for windrow turners is based on the mass of wastes that exist on the composting pad. Thus, the composting pad input mass flow rate was not considered. In other words, to relate the input mass of wastes entering the facility to the windrow turner power, the flow rate, retention time, and turning frequency were deployed. The relevant modelling expressions for the system are:

$$Turner_{reqd} = \frac{Mass_{compost} \times Turner_{freq}}{operating\ \ hours\ \ day^{-1}} \quad \textbf{(2.51)}$$

$$Turner_{hp} = 0.173 \times Turner_{reqd} \quad \textbf{(2.52)}$$

$$Number\ of\ \ windrow\ turners = \frac{Turner_{reqd}}{1700 + 1} \quad \textbf{(2.53)}$$

where $Turner_{reqd}$ is the mass of compost required to be turned on an operational hourly basis (tph); $Turner_{freq}$ is the compost piles turning frequency per week (default of 1 and 3 for the MSCF and HSCF, respectively); $Turner_{hp}$ is the required windrow turner (hp); 0.173 is the coefficient (hp.tph); and 1,700 refers to the turning capacity of typical windrow turner (tph).

v. Front-End Loaders (FELs)

FELs were presumed to be deployed for material transfer activities within the facility. These refer to hauling of the reject material, transportation of the compost to the curing pad, and handling of the wastes at the tipping floor. For the YWCFs case, FELs have been also utilized for the turning of the compost windrows. To determine the number and requirements of FELs, several assumptions were relevant. For the facilities handling 50, 300, and 1,000 tpd of wastes, 1, 2, and 3 FELs were respectively assumed. It was further assumed that each FEL unit had a 150 hp. These assumptions were deemed reasonable, and were partially based on the data given for the MSW composting facility located near Pune city. Based on a linear regression of the aforementioned data, coefficients of 0.003 units tpd and 0.50 hp tpd were achieved. For this, it shall be noted that at least one FEL was utilized in all facilities. The modelling expressions can be summarised as:

$$Number\ of\ \ FEL = 0.003 \times mass_4 + 1 \quad \textbf{(2.54)}$$

$$\mathrm{FEL}_{hp} = 0.5 \times mass_1 \quad \textbf{(2.55)}$$

where $Number\ of\ \ FEL$ is the number of front-end loaders; 0.003 is the coefficient; FEL_{hp} is the required FEL (hp); and 0.5 is the coefficient refers to hp.tpd.

vi. Odour-Control System

For both MSCF and HSCF, it was assumed that the composting pad would be enclosed. A fan system continuously ventilates and directs air to the odour-control system. The number of fans selected would have the capacity to draw air from both the enclosed composting pad and the tipping floor. A building height of 4.5 m was used, and an air exchange rate of 12 times daily (every 120 min) was assumed. A coefficient of 0.00278 hp.cfm^{-1} was utilized. This was based on a linear regression of biofilter flow rates

to total energy requirements (Kong et al., 1996). Also, the odour-control system was assumed to operate 24 h day^{-1}. The relevant expressions are:

$$Total_{flow} = \frac{Vol_{air}}{120} \quad \textbf{(2.56)}$$

$$Odour_{hp} = 0.00278 \times Total_{flow} \times (24/8) \quad \textbf{(2.57)}$$

where $Total_{flow}$ is the flow through biofilters (cfm); 120 as the time period (min) during which the whole building air has been exchanged; $Odour_{hp}$ is the total required fan (hp); 0.00278 as the coefficient based on linear regressions of the operating data hp.cfm^{-1} (cubic feet per minute); 24/8 as the coefficient being used to correct the continuous daily operation (24 h daily) of the odour control system.

vii. Area Requirements

The facility area was assumed to encompass several components, such as the tipping floor, treatment area, composting pad, curing area, buffer zone and offices. The tipping floor was designed based on a maximum RT of 2 days, an average waste height of 1.8 m, and a manoeuvrability factor of 2.0. The tipping floor area requirements were calculated based on the daily flow rate and the bulk densities of each waste component. The composting pad was designed based on guidelines for equipment turning clearance, space between windrows, side clearance, windrow height, windrow base, windrow crown, and number of windrows in parallel. For the YWCF, windrow geometry was assumed to be the same as the MSCF and HSCF. However, a manoeuvrability factor of 2.5 was used for the composting pad total area determination. For the HSCF, a curing pile height of 2.5 m was used with a base-to-height ratio of 2 and a manoeuvrability factor of 2.0.

The office space was determined based on a coefficient of 250 square feet person^{-1}. The reject material was stored in piles of the same geometry as the curing piles for the HSCF and, with a maximum storage period of 2 days. The composting pad and tipping floor were enclosed, and ventilation was provided by fans that continuously direct the air to an odour-control system. The area evaluation expressions are:

$$Area_{staging} = (Area_{tipping} \times Tipping_{man}) + (Area_{screening} \times Screening_{man}) \quad \textbf{(2.58)}$$

$$Curring\ pad = Area_{curing} \times Curing_{man} \quad \textbf{(2.59)}$$

$$Area_{office} = Office_{coeff} \times number\ of\ employee \quad \textbf{(2.60)}$$

$$Area_{rejected} = \text{Re}\ ject\_pile_{lenght} \times Height_{\text{Re}\ ject\_pile} \times \text{Re}\ ject\_pile_{ratio} \quad \textbf{(2.61)}$$

$$Area_{buffer} = 4 \times Dis\tan ce_{buffer} + 2 \times Dis\tan ce_{buffer} \times (Lenght_{facility} + Width_{facility}) \quad \textbf{(2.62)}$$

$$Area_{facility} = Area_{staging} + Compost_{padding} + Curring\ pad + Area_{rejected} + Area_{office} + Area_{staging} + Area_{buffer} \quad \textbf{(2.63)}$$

where $Area_{staging}$ is the total staging area (m^2); $Area_{tipping}$ is the tipping floor area (m^2); $Tipping_{\text{man}}$ is the tipping floor manoeuvrability factor (default value 2.0); $Area_{screening}$ is the area for screening, shredding, and storage of equipment (m^2); $Screening_{\text{man}}$ is the screening area manoeuvrability factor (default value 2.0); $Curring\ pad$ is the total required curing stage area (m^2); $Area_{curing}$ is the area required for curing of wastes in HSCF (m^2); $Curing_{man}$ is the curing area manoeuvrability factor (default of 2.0); $Area_{office}$ is the total area of office space (m^2); $Office_{coeff}$ is the office space coefficient (default of 20 m^2 per employee); $Area_{rejected}$ is the area required for temporary storage of rejected wastes (m^2); $Width_{road}$ is the road width (default 5 m); $Area_{buffer}$ is the buffer zone area (m^2); $Dis\tan ce_{buffer}$ is the buffer zone distance (varies with local legislation and is a function of the location of the composting facility and adjacent sites (e.g., rivers, lakes, wells, ESAs etc., since odour control systems are used for the MSCF and HSCF, a default value of 153 m will be used; a 60 m buffer distance will be used for the YWCF) and $Area_{facility}$ is the total area of the facility, m^2 (Karmakar et al., 2007; Komilis and Ham, 2004).

viii. Electrical Energy Requirements

Electricity is required for various operations in composting facilities, such as mixing, shredding, screening, and operations of fans for odour control. The total electrical energy requirements for the MSW composting facilities were estimated based on the facility size, equipment types, and process flow rates. The electrical energy consumption for different processes was estimated based on the power

ratings of the equipment and the estimated operating hours. Thereafter, the total energy consumption was determined as the sum of the energy consumed by each process. Typically, the electrical energy consumption ranges from 10 to 40 kWh ton^{-1} of input waste, and this alteration is the facility size and process flow rates (Kumar and Samadder, 2017; Li et al., 2018). The energy consumption for fans and pumps is the highest, and contributes about 50% of the total energy consumption (Kumar and Samadder, 2017). Various energy-saving measures can be employed to reduce the electricity consumption of composting facilities. These include variable frequency drives for fans and pumps, material throughput optimization, and utilization of renewable energy sources (Kumar and Samadder, 2017; Li et al., 2018). The net electrical power consumption expression is:

$$elect_{hp} = hammer_{hp} + trommel_{hp} + odour_{hp} \quad \textbf{(2.64)}$$

Table 2.7 A summary of the energy requirements for various composting process equipment.

Equipment	YWCF		MSCF		HSCF	
	No. of units	***Energy requirements***	***No. of units***	***Energy requirements***	***No. of units***	***Energy requirements***
Trommel Screens	1-2 units/10-50 tpd	5-10 kW unit^{-1}	2-3 units/50-100 tpd	10-20 kW unit^{-1}	3-4 units/100-200 tpd	20-30 kW unit^{-1}
Hammermill/Tub Grinders	1 unit/10-50 tpd	5-10 kW unit^{-1}	1-2 units/50-100 tpd	100-200 kW unit^{-1}	2-3 units/100-200 tpd	200-300 kW unit^{-1}
Windrow Turner	1 unit/10-50 tpd	10-20 kW unit^{-1}	1-2 units/50-100 tpd	20-30 kW unit^{-1}	2-3 units/100-200 tpd	30-50 kW unit^{-1}
Front-End Loaders (FELs)	1 unit/10-50 tpd	Diesel-powered	1-2 units/50-100 tpd	Diesel-powered	2-3 units/100-200 tpd	Diesel-powered
Odor-Control System	1 unit/10-200 tpd	1-5 kW unit^{-1}	1-2 units/200-500 tpd	5-10 kW unit^{-1}	2-3 units/500-1000 tpd	10-20 kW unit^{-1}
Area Requirements	1-2 acres		2-5 acres		5-10 acres	
Electrical Energy Requirements	1 unit/10-50 tpd	50-100 kW	1-2 units/50-100 tpd	100-200 kW	2-3 units/100-200 tpd	200-300 kW
Material flow	500-1000 tons yr^{-1}		1000-5000 tons yr^{-1}		5000-10000 tons yr^{-1}	

Sources: Diaz et al., 2007; Epstein, 2017; Environmental Protection Agency (EPA) report 2013

where $elect_{hp}$ is the electrical power requirements due to equipment operation (hp). The equations for design of composting facilities have been summarized in Appendix C.

The number of units and energy requirements for YWCF, MSCF and HSCF systems have been summarized in Table 2.7.

2.9.3.2 Cost Estimation

To enable a fair estimation of composting facilities, several key parameters are to be considered. These include construction, land acquisition, engineering, equipment, operating, capital, and total costs. Construction cost refers to the expenses associated with the commissioning of the compost model infrastructure. These refer to the composting facility, storage areas, and processing equipment. They also include the cost of materials, labour, and equipment rented or purchased. Land acquisition cost refers to the amount of money needed to purchase or lease the site at which the compost plant will be stationed. This cost varies and depends on the location, size, and zoning regulations of the land. Engineering cost includes the cost of hiring engineers to design and oversee the construction of the compost model. It can include tasks such as creating blueprints, conducting feasibility studies, and ensuring compliance with regulations.

Equipment cost refers to the expenses associated with the purchasing or leasing of the equipment needed to run the compost model. These primarily refer to shredders, compost turners, and screening equipment. Operating cost includes ongoing expenses associated with the running of the compost model. Its prime entities are labour, utilities, and maintenance costs. Capital cost refers to the total amount of money required to start the compost model. It includes all the above mentioned costs. Finally, the total cost of the compost model is the sum of all the above costs. Estimating the total cost of a compost model is highly complex and depends upon various factors such as the size and complexity of the operation, location, and regulations. Fair cost estimation of composting facility is mandatory for budget and finance related decisions (Komilis and Ham, 2004). The expressions for the plant can be presented as:

$$\mathrm{Cost}_{cap} = \mathrm{Cost}_{constr} + \mathrm{Cost}_{land} + \mathrm{Cost}_{engg} + \mathrm{Cost}_{equip} \quad \textbf{(2.65)}$$

$$\mathrm{Cost}_{\mathrm{constr}} = \mathrm{Cost}_{\mathrm{grad}} + \mathrm{Cost}_{\mathrm{pav}} + \mathrm{Cost}_{\mathrm{fence}} + \mathrm{Cost}_{\mathrm{build}} \quad \textbf{(2.66)}$$

$$\mathrm{Cost}_{land} = land\ unit\ \mathrm{cost} \times Area_{fac} \quad \textbf{(2.67)}$$

$$Cost_{engg} = Const_{\%} \times Cost_{const} \quad \textbf{(2.68)}$$

$$Cost_{equip} = Cost_{turner} + Cost_{hammer} + Cost_{grinder} + Cost_{trommel} + Cost_{FEL} \quad \textbf{(2.69)}$$

$$Cost_{labour} = \frac{Employee\ wage \times employee\ number}{mass_1 / operating\ hours\ per\ day} \quad \textbf{(2.70)}$$

$$Feedstock\ Cost = Cost\ per\ unit\ of\ feedstock \times Amount\ of\ feedstock\ used \quad \textbf{(2.71)}$$

2.9.3.3 Revenue Estimation

The revenue generated by a compost facility is dependent upon several factors such as the quality and quantity of compost produced, market demand for the compost, and the SP of the compost. The revenue generated by the compost facilities have been calculated with the following equation:

$$Revenue = Quantity\ of\ Compost \times Selling\ Price \quad \textbf{(2.72)}$$

Also, the PBP of a compost facility refers to the time taken by the facility to recover its initial investment through the revenue generated from the sale of compost. The equation for the determination of PBP is:

$$Payback\ Period = \frac{Initial\ Investment}{Annual\ Cash\ Inflows} \quad \textbf{(2.73)}$$

2.9.4 Sensitivity Analysis

In order to assess upon the profitability of two alternate composting systems, a sensitivity analysis was conducted to evaluate the impact of various input parameters on the operating costs and revenues. The sensitivity analysis involved the variation of each parameter by ±10% from the base case value for the subsequent assessment in the resulting change in profitability. The parameters considered in the sensitivity analysis included plant size, capital investment, electricity consumption, equipment costs, feedstock prices, transportation costs, RT, and compost selling price. For each parameter, only changes in profitability greater than ±\$2.5 ton^{-1} of compost were considered to be significant and were included in the analysis. Accordingly, with the sensitivity analysis, the most influential parameters for each composting system were identified. Such an information can be used to optimize the design and

operation of the composting systems and thereby maximize profitability. Additionally, the sensitivity analysis can help to identify areas in which more accurate or detailed data is needed to further enhance the accuracy of the cost and profitability estimates.

2.9.5 ML-based Prediction of Compost Production Rate

ML is a relevant framework for CP prediction. This is due to its ability to optimize and automate various aspects of the production process. This leads to enhanced efficiency, profitability, and quality control. In the thesis, a predictive modelling approach with ML has been deployed to predict the compost production volumes. ML models can be trained to predict CP as the output variable with various input variables such as feedstock type and quality, composting parameters, equipment, labour, location, and market demand.

2.9.5.1 Overall Methodology and Deployed ML Models

The overall ML based prediction framework involves data pre-processing, modelling, and analysis. To do so, several data pre-processing stages were first required for the preparation and translation of raw primary data (from GMC compost facility) into variables suitable for modelling and analysis. Needless to convey, they involve outlier removal with the filtering processes, and integration of consolidated datasets (2016-19). Such details have been presented in an earlier article (Singh and Uppaluri, 2022). For the composting integrated dataset, a random train-test split of a 70:30 ratio was considered to achieve 140 and 60 samples as training and testing datasets respectively. Thereafter, the datasets were deployed for model development and validation. With three alternate ML algorithms namely SVR, GA and MLP (have been discussed earlier), the ML addresses a multi-regression system in which one dependent variable (CP volume) is predicted in terms of dependent variables (feedstock quantity, MC, RT).

2.9.5.2 Model Testing and Validation and Forecasting Models

To enhance the robustness and model accuracy, a similar approach as discussed in the previous sections has been deployed in the training and validation phase (HPO with grid-based K-fold cross validation approach). The accuracy of the ML model was determined through a combination of the train and test scores, with the subsequent evaluation of the RMSE and R^2 based predictive performance. The metrics were used to assess upon the accuracy of the model for CP volume prediction. Thereby, the optimization

of production processes and the improvement of overall profitability can be achieved. Also, ARIMA-ML was deployed as the forecasting model (discussed in section 2.4.5 of the Ph.D. thesis).

Chapter 3
MSW Characterization & Quantification

Chapter 3

MSW Characterization and Quantification and Seasonal Classification for Sociodemographic Groups in Guwahati City

This chapter presents the findings associated to MSW quantification and seasonal classification of socio-demographic elements in the Guwahati city. After a brief introduction in section 3.1, sections 3.2 and 3.3 respectively detail upon the current methods for HHW disposal followed by the findings of the HH survey. This survey was conducted with a well-designed questionnaire to gather data on socio-demographic characteristics, waste attitudes, disposal behaviors, financial capacity, and existing management issues of the HH. Section 3.4 elucidates upon the interpretation of the Pearson correlation and PCA analysis to identify patterns and relationships among the considered socio-economic and demographic parameters. Characterization studies were also conducted to quantify and infer upon the waste compositions. These were addressed in section 3.5 of the Ph.D. thesis. Following this, section 3.6 delineates upon literature data and associated comparative assessment.

Overview

A detailed survey-based quantification of MSW was targeted in the study. Also, based on the collected data from 31 wards of Guwahati city, the seasonal classification of waste characters for various sociodemographic groups has been addressed. Population classification through sociodemographic categories, data collection techniques, and information gathering were deployed to characterize and evaluate the HHW fraction. Statistical approaches such as correlation analysis and PCA have been implemented to identify dominant waste constituents and assess upon seasonal influence and associated sensitivities.

3.1 Introduction

In this chapter, area wise characterization of HHW has been targeted to devise upon a pragmatic decentralised waste collection and management system at the general population level in Guwahati city. The primary emphasis of the carried-out investigations was to ascertain upon the influence of area, HH income, and seasonal effect on the MSW rate alteration in the city of Guwahati. Three distinct times of the year (February, 2021, June, 2021 and August, 2022) were used to define the MSW produced in the Guwahati city's commercial arena, sub-areas A1, A2, and A3, and for people belonging to various income levels. Population data along with HH socioeconomic status, awareness for general population, education and training with respect to environment factors were considered to target and design

[#] Published article: Singh, T., & Uppaluri, R. V. (2023). Characterization of municipal solid waste and seasonal classification for various socio-demographic groups in Guwahati city. *Material Cycle and Waste Management.*

centrally controlled HHW management system in urban areas. The chapter aims to address few critical issues such as HHW quantitative measurement and concentration, evaluation of HHW amount and rate of generation in various socioeconomic groups in the Guwahati city along with the potential scope for HHW utility as a precious commodity for the generation of energy. All these can translate into the well documented integrated solid waste management strategy for the Guwahati city and with minimal environmental threats.

3.2 Current MSW Disposal Methods

As per GMC record, the Guwahati city is divided into several subzones or colonies. In total, 11 different wards were selected for HHW characterization study. Table 3.1 shows the census information for these regions. As per census data, the population of these subzones ranged from 21,443 to 39,350 (Census of India, 2011). The GMC is primarily liable for the collection, public transit, and strategic planning of the city's MSWM. In some locations, several non-profit organizations as well as NGOs also help municipal staff with HHW transportation and collection services. With these additional MSW collection locations in these regions, the GMC also offered other facilities and services. Separate mechanisms were used in the waste disposal system for the disposal of open rubbish, nonferrous metals, and cement. Throughout the subzones, specific community containers (Plate 1A) have been placed for the purpose of collecting HHW. The HHW is then transported to the city's adjacent disposal locations for the waste prior to it being dumped at the last dumpsite (located close to Boragaon). The number of public trash cans required is determined by the population and number of homes in each block or colony. A tractor-trailer method is used to collect MSW in regions with community containers. In order to avoid trekking of the landfill, few residents of the blocks dump their HHW on the sides of the roads, in public areas, on waste land, etc. and in community bins positioned nearby to their residences. Despite GMC employees visiting frequently (every 5–10 days) such secondary community disposal sites, occasional delays do happen. People are frequently forced to dispose their HHW in an arena closer to an overflowing community receptacle. Scavengers will occasionally collect HHW fractions like food scraps or vegetable trash from communal bins (street dogs, cows, birds, etc.). Moreover, ragpickers collect recyclable waste from public garbage bins. These include polythene bags, plastic scrap, polythene bottles, metals, paper, and glass. Such interactions provide serious challenges for the GMC

in terms of efficient trash collection and disposal techniques. These practises needlessly strain GMC staff members. In the city, a formal facility for HHW separation, energy recovery, or composting doesn't exist. MSWM in the city still heavily relies on landfilling. The landfill location lies between N 26.1158° and E 91.7086° and covers an area of around 3.7 hectares. Since September 2018, it has served as the city's brand-new landfill. A JCB eventually settles the discharged MSW with 700 tonnes of garbage on a daily basis.

3.3 Household Survey Findings

The majority of the 227 HH respondents (62.8%) were single families (with 1-3 members who lived alone). The typical family size in the residential zones was 6.5 people. Among HH, about ten members had the least percentage share (11.9%), and the group of 4-6 members had the next highest percentage share (27.9%). Apart from the residential areas, few commercial establishments (restaurants, shops etc.) were also considered. Regarding income, the majority of HHs (43.7%) made between 483.66 and 846.17 US$ per month, and just 13% made more than 1,209.16 US$. The results of the HH survey regarding socioeconomic status have been shown in Fig. 3.1. For MSWM services, the majority of HHs said they could only pay 0.60 - 0.97 US$ (47.2%) and 1.21 - 2.42 US$ (33.7%) per month. On the contrary, only 19.1% could afford 3.63 – 6.04 US$ for such services. The concept of enhanced sustainable MSWM is generally welcomed. In a public awareness effort, almost 25% of respondents said that they would be interested to volunteer.

Table 3.1 A summary of survey-based ward wise data of population, area and disposal facilities in the Guwahati city.

Wards	**Population (Census, 2011)**	**Area (km^2)**	**Waste disposal facilities**			
			Concrete	***Metallic***	***Open***	***Total***
Ganeshguri Area (w_{19})	27125	3.53	4	5	8	17
Chandmari Area (w_{21})	33057	3.36	4	4	6	14
Narengi Housing Colony (w_{24})	35524	5.15	3	6	Nil	9
Noonmati Refinery Area (w_{22})	38081	13.25	5	3	Nil	8
Beltola Area (w_{28})	39350	5.30	4	4	2	10
Khanapara Area (w_{30})	33460	6.28	6	5	Nil	11
Pandu Nath Basti (w_4)	35524	3.64	2	1	3	6
Fatasil Harijan Colony (w_{15})	28419	1.28	3	3	4	10
Itabhata Harijan Basti (w_{16})	37806	2.50	3	5	3	11
Panbazar Area (w_{10})	21443	2.19	3	6	4	13
Christian Basti Area (w_{20})	31722	3.46	5	4	6	15

In addition, a significant majority of respondents (26.4 percent of HHs) indicated that they were willing to separate the five main categories of solid garbage, including paper, glass, metal, plastics, and organic biodegradable waste. 11.3% of HHs indicated that they would separate their garbage if a little financial reward was provided. Lack of time, disease-related reluctance, and dislike to handle waste were the biggest barriers for the participation of the remaining 63.2% of respondents. 1-2% of respondents composted food waste, which constitutes a sizeable portion of their HH solid waste. Otherwise, they feed it to animals to resolve garbage disposal issue. Almost every respondent threw their food waste in the ordinary garbage bins. The location type and the individual in-charge for HH disposal were analyzed to be statistically significant (p-value = 0.25). According to respondents, the father and mother in the questioned families are responsible for waste disposal and, with respective percentages of 28.3% and 19.8%. The percentage share for the father was 21.2%, 9.9%, and 18.5% in A1, A2 and A3 regions. On the other hand, the respective percentage share for the mother was 15.9%, 2.1%, and 22.1%. The Guwahati city municipality had in total 345 trash cans to accommodate 957,351 people (population in 2011). Their capacities range from $1m^3$ (82 percent of garbage cans) to $20m^3$ (0.8 percent of cans). During survey of the issue encountered with the closest located trash can for solid waste, 52% of respondents mentioned that the presence of insects and rats (49%) along with filthiness of the trash cans were the most pressuring issues (53%). Other pressing issues mentioned were the absence of a container on the homeowner's street (58%) and the occasional emptying of trash cans (53%).

3.4 Correlation Analysis Findings

Table 3.2 lists the Pearson correlations between many important socioeconomic and waste generation variables. Number of HH (0.91), population (0.87), HH size (0.74), education level (0.65) and number of employees (0.63) possessed a high and positive correlation coefficient. On the contrary, the age group (-0.21) had negative correlation. Also, correlation coefficients for the pairs, number of employees-population (0.34), age-education level (0.54) and number of employees-age (0.29), have all affirmed positive but weak correlation. Accordingly, targeting the subduing of the associated multi-collinearity issues, the respective model pairs were omitted during the subsequent modelling efforts.

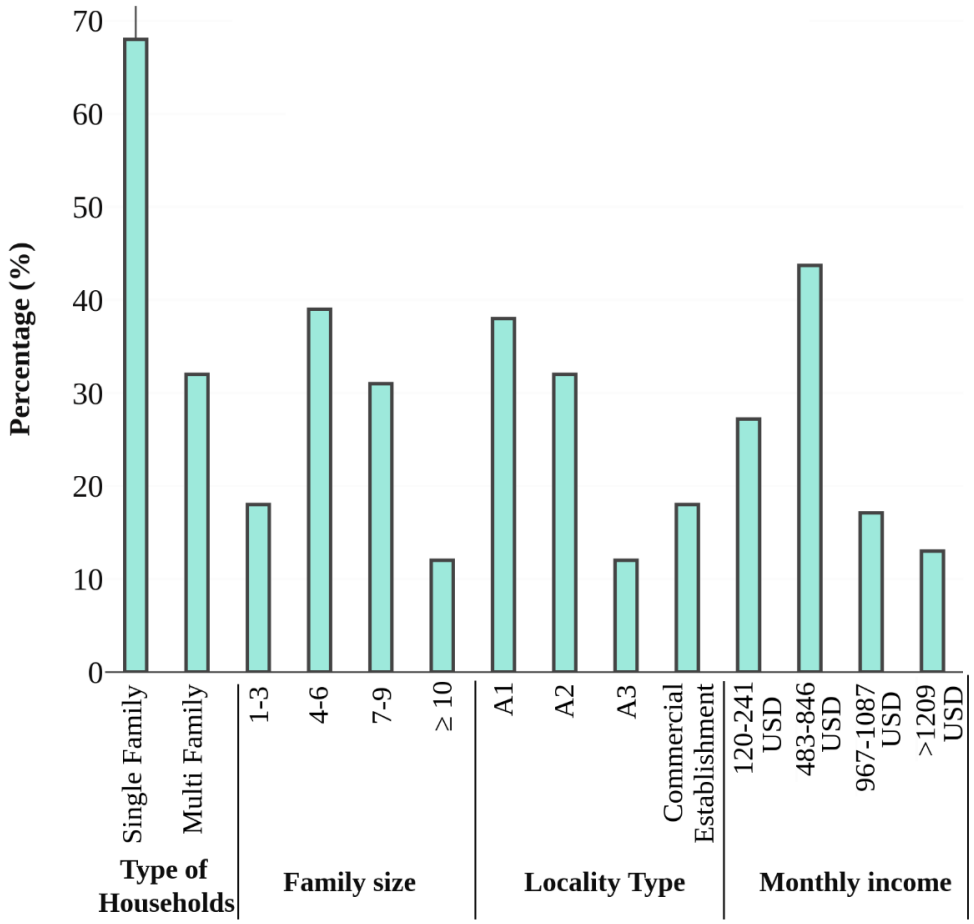


Fig. 3.1 Percentage contribution of socio-economic groups towards the survey based MSW quantity in the Guwahati city

A table representation with factor scores or weights for each variable was produced with the PCA analysis (Table 3.3 (a-c)). In general, a variable with a greater socioeconomic position is correlated with it, and vice versa. Working population and educational attainment have been given positive weights in several PCA analyses (Kannangara et al., 2018). This suggests that a home with a larger proportion of working-age adults and a higher level of education will be ranked higher in terms of their socioeconomic status than a HH without those features. The working population's stronger correlation with predicted factors being related to lower socioeconomic level could be the cause of this outcome. Poor housing and sanitary circumstances are two examples of unfavourable impactor influences. These results are applicable to circumstances in which indices have been built for the larger Guwahati region. In these

situations, the asset may signify wealth in some areas of the city but not in others due to economic disparities.

In Table 3.3(a) each PC axis PC 1 or PCA 1, each axis is an eigenvector. So, in this case the first two eigenvectors first two axes explain only 64.20 % of the total variance of this data set meaning that most of the variance is coming from factors that we didn't account for or didn't measure. The analysis is optimized as in general if the first two PC axes don't explain at least 60 % of the total variance then the PCA will not be useful (Noori et al., 2009). In summary, the data shows the contribution of each eigenvalue to the total variance in the dataset. The cumulative column represents the cumulative percentage of variance explained up to that eigenvalue. The higher the eigenvalue and its corresponding percentage of variance, the more significant the associated principal component or factor is in capturing the variation in the data.

Table 3.2 Pearson correlation indices of various socio-economic factors in the survey-based waste quantification case study.

	Waste generation	**Population growth**	**HH**	**HHtype**	**HHsize**	**Age**	**EDU Level**	**HH Inc**	**Number of Employees**
Waste generation	1								
Population growth	0.87	1							
HH	0.91	0.78	1						
HH type	0.64	0.38	0.67	1					
HHsize	0.74	0.64	0.81	0.65	1				
Age	-0.21	0.33	0.17	-0.11	0.11	1			
EDU Level	0.68	0.42	0.37	0.32	0.26	0.54	1		
HH inc	0.72	0.43	0.33	0.51	0.64	-0.12	0.33	1	
Number of Employees	0.63	0.34	0.22	0.59	0.31	0.29	0.73	0.88	1

Table 3.3 (a) Eigenvalues of the Correlation Matrix.

Eigenvalue	**Percentage of Variance**	**Cumulative**
8.01389	44.52%	44.52%
3.54153	19.68%	64.20%
2.64835	14.71%	78.91%
1.76881	9.83%	88.74%
1.26012	7.00%	95.74%
0.49104	2.73%	98.47%
0.17431	0.97%	99.43%
0.10194	0.57%	100.00%

Table 3.3 (b) RDA for response variables.

	Response Variables							
	Coefficients of RDA1	**Coefficients of RDA2**	**Coefficients of RDA3**	**Coefficients of RDA4**	**Coefficients of RDA5**	**Coefficients of RDA6**	**Coefficients of RDA7**	**Coefficients of RDA8**
OW	1.31671	2.08134	0.35543	0.61966	0.37307	0.44696	0.14477	0.07666
Paper	2.93023	0.98828	0.82293	0.42485	0.15011	0.30209	0.12075	0.04272
Plastics	2.68438	0.32616	0.19913	0.0695	0.05295	0.17137	0.00632	0.02084
Metals	0.12142	0.16353	-0.13392	-0.23757	0.25749	0.01881	-0.04907	0.22503
Glass	0.78169	-0.19766	-0.04658	-0.62148	0.18826	0.25663	0.23465	-0.08584
Rubbers/ Leathers	0.98876	-0.64044	0.75981	-0.08033	0.25858	-0.29095	-0.1728	-0.00186
Textiles	0.80089	-0.86864	0.89257	-0.20078	0.38718	0.14921	-0.01206	-0.02869
GW	0.12432	1.3062	0.82682	0.02297	-0.03491	-0.09852	0.17286	0.06517
Wood scraps	-1.30133	-0.19976	-0.32684	0.68979	0.67859	-0.14468	0.13186	0.00151
Misc.	0.46902	-0.02986	0.39828	0.80111	-0.10718	0.52381	-0.05756	-0.04631

Table 3.3(c) RDA for explanatory variables.

	Explanatory Variables							
	Coefficients of RDA1	**Coefficients of RDA2**	**Coefficients of RDA3**	**Coefficients of RDA4**	**Coefficients of RDA5**	**Coefficients of RDA6**	**Coefficients of RDA7**	**Coefficients of RDA8**
Population growth	0.33346	0.83731	0.22662	0.20719	0.23417	0.12076	0.13274	0.07995
HH	0.20241	0.50226	0.2802	0.64699	0.09118	0.46032	0.09377	0.06395
HHtype	0.15595	0.8749	0.21563	0.26856	0.21726	0.12899	0.02688	0.09458
HHsize	0.13706	0.38974	0.00894	0.39159	0.1398	0.04007	0.09373	0.30267
EDU	0.42466	0.53866	-0.10047	-0.07673	-0.70674	-8.06E-04	-0.01246	-0.11786
HH Inc	0.27013	0.91578	-0.08766	0.08726	0.0946	0.01075	0.22151	0.12227
AGE	-0.12956	0.82706	-0.15811	-0.33141	-0.37971	0.05484	0.07929	-0.1043
EMP	0.84399	0.40863	-0.21984	0.14002	0.20344	5.31E-04	0.05257	0.00889

Results of the RDA are summarized in Table 3.3(b, c) and shown in Fig. 3.2 (a-b). An RDA will use those characteristics for each observation and try to relate them to explain the relationships between response variables (Table 3.3 (b, c)). The response variable analysis measures the outcome of the study and exploratory variables depicts that outcome such as population growth, HHsizes, HHinc and employee counts which have positive association with respect to RD axes. The scree plot (Fig. 3.2(a)) is the big drop-off between PC 1 to PC 10. In the figure it can be seen that there is a significant drop between PC 1 to PC 2. Similarly, for PC 2 to PC 3 and PC 4 the drop-off is significant. A flat line indicates ineffective PCA analysis (Noori et al., 2009).

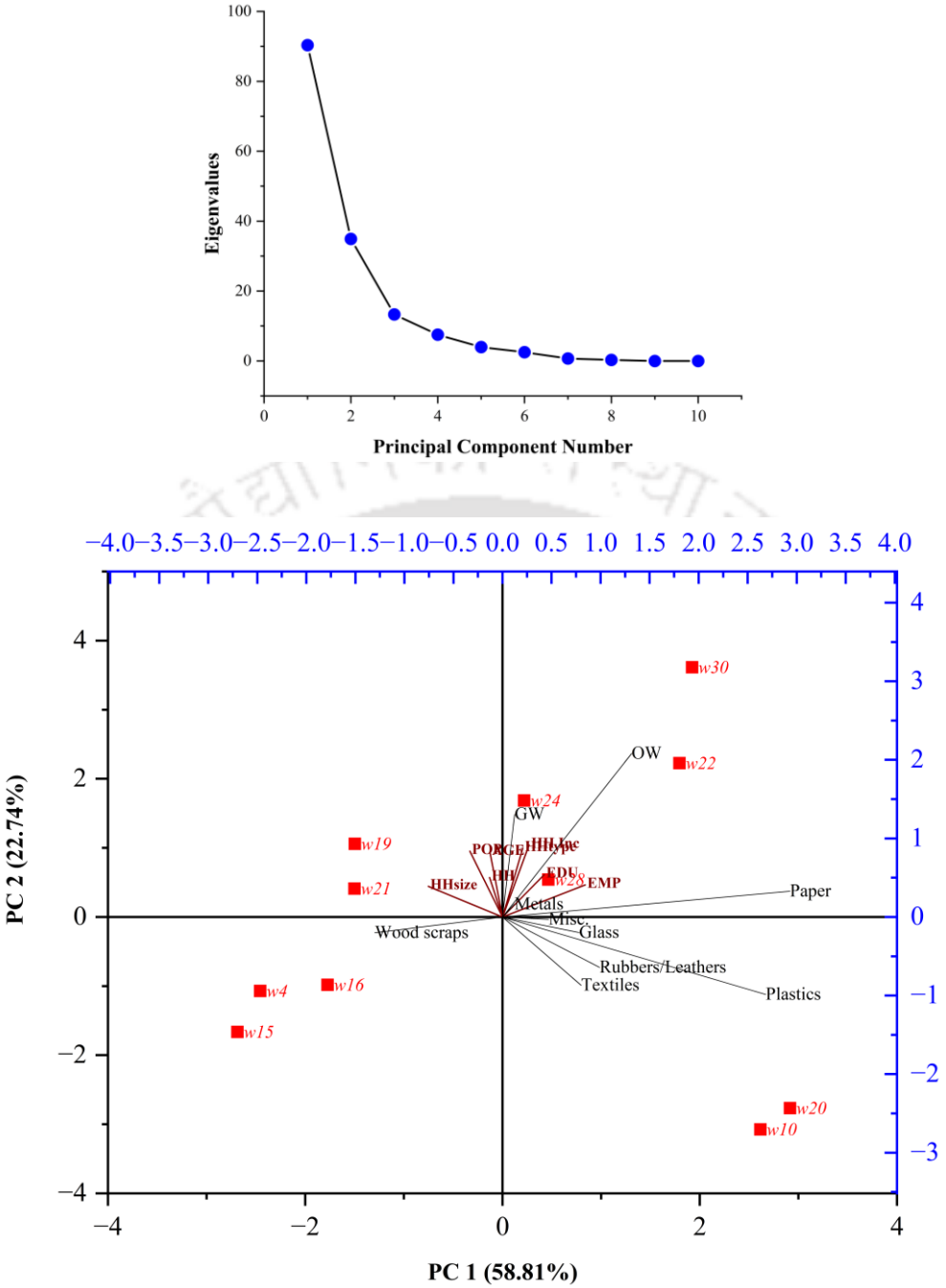


Fig. 3.2(a) Scree plot between the eigenvalues and principal axes **(b)** RDA for the socioeconomic, demographic factors and waste compositions in the municipal wards of Guwahati city

The variables selected in the analysis are represented in the triplot by thick arrows, which point in the direction of maximum change in the value of the associated variable (Fig. 3.2(b)). Fig 3.2(b) displays the relationships of integrated factors (identified by the PCA analysis) and waste generation rate for the year 2021-2022. Each point represents data of one ward placed in RDA axes coordinates. The figure also displays in MSWG rate in the nine considered wards grouped by social-economic status. The strength of the association between MSW rate and administrative wards in groups is measured by the

projection of wards objects onto the arrows of these MSW. However, based on the length of arrows, there are only OW in the three studied years and plastics and glass wastes in 2021-22 represented well by the first two axes. OW appears to be highly related to sub-area A1 and A2 representing w22, w24, and w28. The figure depicts a distinct pattern between these wards. Judged by the angle between arrows, it is clearly that OW is influenced equally by the two social-economic groups (A1 and A2). This means OW has been generated considerably in these wards. Similarly, paper waste seems to be more generated in the A2 sub area in 2021. Other concerned wastes (metals, garden waste (GW), glass) have rather short arrows indicating these wastes are not well explained by the first two axes so will not be mentioned for correlation explanation. Fig. 3.2(b) plots and compares the data of w30, w4 and w19 in order to examine how MSW composition change in spatial and temporal dimensions. The figure shows that plastics waste, and glass, are generated more in these wards. Also, plastics waste, textiles, rubbers, glass, metals are much more prevelant in commercial habitat (w10, w20). Redundancy analysis (Table D1 in the appendix) with eigenvalues indicate the amount of variation explained by each RDA axis. The percentage of inertia shows the proportion of total variation explained by each axis, while cumulative inertia represents the cumulative amount of variation explained by all the preceding axes. In this case, the first axis (RDA1) explains 52.12% of the total variation, and the cumulative inertia gradually increases as more axes are added. Standardized scores for the PC axes (Table D2 in the appendix) refer to the mean values of the variables after they have been transformed or standardized based on the PCA.

Working population and educational level have been given positive weights in several PCA analyses (Kannangara et al., 2018). This suggests that a home with a larger proportion of working-age adults and a higher level of EDU will be ranked higher in terms of their socioeconomic standing than a HH without those features. The working population's stronger correlation with predicted factors being related to lower socioeconomic level could be the cause of this outcome. Poor housing and sanitary circumstances are two examples of unfavourable impactor influences. These results are applicable to circumstances in which indices have been built for the larger Guwahati region. In these situations, the asset may signify more income in some areas of the city but not in others due to economic disparities. From the considered

socio-economic factors, the MSA and communality values for retained factors (explanatory variables) population growth, HH counts, HHsize and employee number are 0.65, 0.79; 0.71, 0.83; 0.75, 0.87 and 0.63, 0.78 respectively. The results convey that for population growth and HH characteristics (number of HHs and HH size), all durable assets have good factor scores. On the contrary, EDU level and number of employees have a positive but low factor score. Also, it is interesting to note that the study's age group had negative factor scores across the board. A dependent variable with a mean equal to zero and a standard deviation equal to one can then be created for each HH and with the factor scores from the first main component as weights. Thus, the mentioned dependent variable can be considered as the socio-economic score for each HH. Thereby, it can be affirmed that higher socioeconomic score of the HH, higher socio-economic status of the HH was apparent. A higher socioeconomic score suggests better income, education, and resources for a HH. This can lead to improved socioeconomic status through increased economic opportunities, better access to education and networks, and enhanced health and well-being (Nguyen et al., 2020). Few authors (McKenzie et al., 2011) elucidated how upon HH size in the literature. Filmer et al. (2001), did not account for HH size on the basis that the advantages of the applied indicators are already present at the HH level. According to the interpretations of the weights for the situation under consideration, A2 would receive a better socioeconomic status score due to its assets, number of HH size, population growth, and educational level. The age group findings for A1, A2, and A3 have been similar and with low scores. This implies a lower socioeconomic status score for all cases. The justification for the selection of representative elements is closely related to the fulfilment of a few requirements in the subsequent analytical stage including KMO, MSA, and communalities.

In conclusion, eight parameters in total have been considered in the PCA methodology. KMO in the investigation produced a result of 0.72, and thereby conveyed that the PCA is suitable for the dataset. The HH size and population growth have a maximum value set for A1 (0.70, 0.65), A2 (0.83, 0.78) and A3 (0.72, 0.68) respectively. The PCA results for A1 provide valuable insights into the underlying patterns and relationships among the variables. While population growth size exhibited a moderate contribution (mean: 0.694, SD: 0.461), the number of HH had a relatively lesser influence. HH size

conveyed a moderate contribution (mean: 0.314, SD: 0.463), the age demonstrated a contrasting pattern with a negative factor score (-0.238) and a mean of 0.053 (SD: 0.224). The EDU level moderately influenced the patterns (mean: 0.410, SD: 0.491), and the number of employees had a lesser influence (mean: 0.006, SD: 0.076).

3.5 Household Data based MSW Characteristics

Table 3.4 presents the number of HH, property owners, and proprietors who took part in the study. Along with these, the table also presented the total quantity of waste generated in each sub-area. The weighted mean was used to compute the annual average values of MSW composition per capita. With a temperature range of 12 to 25 °C, the results for the month of February were selected as the benchmark. The maximum temperature range for the months of June and August, (altered from 20 to 31 °C and 26 to 34°C, respectively) were also taken into consideration. Even though there were no discernible differences amongst the sub-areas, the A3 (the group with the least income) had the least annual average garbage production rate. In the survey conducted for A1 and A2 in the month of February, the number of HH surveyed was 27 and 28, and the total waste generated during this period was 105.2 kg and 104.8 kg respectively. For the month of June, the survey results revealed a HH count of 26 and 25, and the generated waste increased to 129.5 kg and 159 kg respectively. Finally, in August, the number of surveyed HH was 27 and 28, and the generated waste was 150.3 kg and 117 kg respectively. Henceforth, it can be inferred that the A2 (higher economic group) generated more waste in comparison to the A1 case.

Similarly, for the A3, the survey conducted in February recorded a count of 23 HH, and for a total generated waste of 91.5 kg. In June, for a total HH of 21, the generated waste reduced to 73.9 kg. In August, for a total HH count of 22, the waste generated increased marginally to 81.1 kg. The surveyed commercial establishments also revealed that for the 29 commercial surveyed establishments, a total of 177.7 kg of waste was generated in February.

Fig. 3.3 depicts the garbage production for each time period and at various economic levels. Guwahati produced 0.201 kg of garbage per person per day (weighted average). The table makes it very evident that in comparison to the A1 and A2 groups, the A3 group (least income) contributed very less generated

waste quantity. Furthermore, waste production reduced for all three level cases in February (with low temperatures). Thus, the fresh foods and vegetables consumption conveyed lesser waste production in

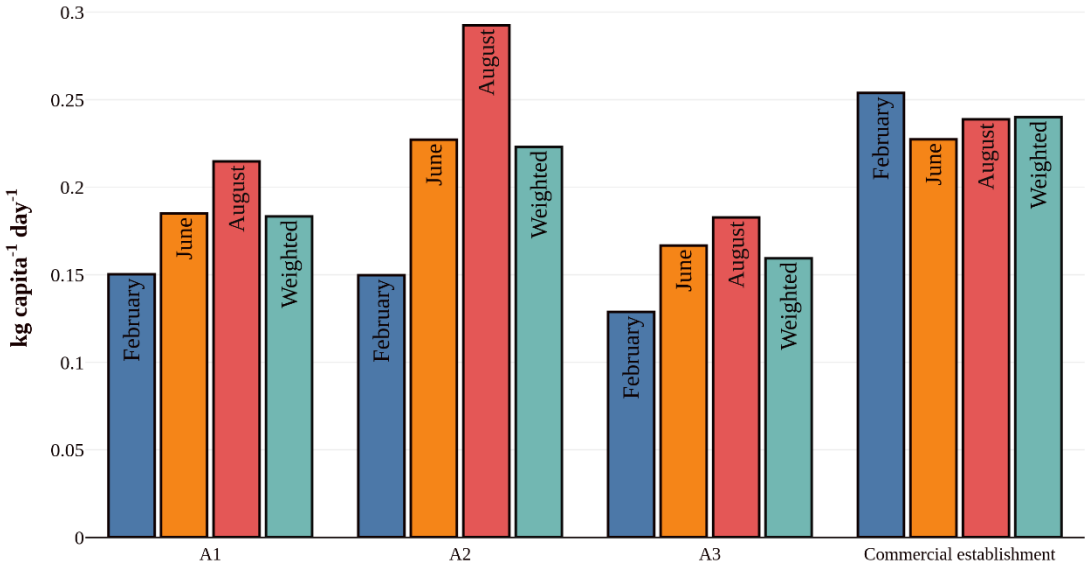


Fig. 3.3 Bar chart depicting total MSW rate per capita of alternate sub-areas and seasons in the Guwahati city

winter and henceforth human habitat alteration. The relevant tendency of higher income groups producing more waste is due to continuous and quicker economic rise. This is consistent with research findings reported for Abu Dhabi and Morelia (Abu Qdais et al., 1997; Buenrostro et al., 2001; Cao et al., 2006). The authors discovered that both location and season did affect the waste content.

3.6 Household Waste Characteristics

The average percentage share of generated HHW in the Guwahati city and for the study period were 99.5±0.9 %, 99.4±0.3 % and 98±0.7 % for A1 A2 and A3 respectively and 98.3±1.0 % for commercial establishments. Table 3.5 display the findings for each fraction collected in each sub-area during the three seasons and as an average percentage. For all three levels, and particularly in August, the organic content share contributed the highest waste quantities. This fraction reduced for each of the four sub-areas in February. A2 (highest income) had the most commercially purchased uneaten food as well as uneaten products that were disposed. This is due to prevalent consumption dates. The percentage of newspapers and magazines per person is greater in June in all three sub-areas. This could be due to the

holiday season. The A2 (highest income) group indicated highest cardboard and plastic waste. These HHs generate more cardboard for packaging due to their higher incomes and ability to buy a wider choice of products. For all three sub-areas, the proportion of plastic containers increased in August. The use of beverages in plastic containers (water, fruit juice, dairy products, flavoured drinks made with artificial flavours, etc.) is linked to the high temperatures. Tetra Pak items were categorised in this portion as they are frequently used for fruit juice and dairy packaging. In June and August, more materials were disposed.

Table 3.5 also presents the significant weighted mean of OW (54±3.0 %), papers (12±5.8 %) and plastics (19±7.2 %) for the commercial area. In this region, OW and paper exhibited significant rise during the month of August. The organic (kitchen) waste is a good indicator of the alteration from summer to winter. The average percentage of kitchen waste was 47.3% in the summer and subsequently fell to 45% in the winter. This is due to the reason that people consumed more fruits and vegetables in the summer season due to high temperatures. It's interesting to note that the amount of OW produced in the commercial sector was also very high due to many establishments, shops, restaurants, and other small businesses in this region.

Additionally, Guwahati city currently lacks a procedure to collect used and obsolete batteries. As a result, batteries are also disposed along with other fractions. There weren't many disposals collected for the remaining component, which includes apparel and footwear. These items are frequently donated or sold again as used goods. The study did not account for bulky waste, which includes furniture and electrical and electronic equipment, as they are collected only on specific days by the Guwahati municipal service, and their future quantity could not be estimated. Typically, bulky waste is sent to a landfill in which waste pickers dismantle and sell the parts individually. The waste fractions sorted in this study were divided into six groups: OW, paper waste (including newspapers, magazines, and cardboard), metal waste (including metal and aluminium cans), glass waste (including both clear and coloured glass), plastic waste (including plastic and plastic containers), and other items (such as fine debris, batteries, synthetic and natural fibers, and various other items).

The waste composition characters for the investigated business establishments and the three socioeconomic categories have been displayed in Fig. 3.4(a-d) respectively. These have been presented as a proportion of the six principal fractions. The largest quantities of waste generated were 46%, 48%, 35%, and 39% for OW, 18%, 22%, 23%, and 28% for plastics, and 15%, 17%, 18%, and 26% for papers and in sub-areas A1, A2, A3 and commercial enterprises respectively. The rates of metal and glass creation were found to be similar across all three sub-areas. These figures provide insights into the relative distribution of different waste types. OW appears to be the most prevalent waste category, followed by plastics and papers (Al-Khatib et al., 2010). Also, the percentage distribution of papers and plastics have been higher in commercial establishments (Miezah et al., 2015). The ANOVA analysis affirmed no significant differences between the sub-areas. Fig. 3.5 depicts the characterization values for various fractions of urban solid waste. The organic component accounted for the largest proportion of waste (at around 54%), followed by the plastic component (accounted for about 17% of all the waste).

Table 3.6 summarises the findings in comparison to the literature findings. According to a study, high income groups produce more garbage than other socioeconomic categories (Gomez et al., 2009). The choice of significant socio-economic elements serves as the primary criterion for the determination of the waste characterisation study. In conclusion, this study showed the effectiveness of particular criteria and how it affected waste formation patterns. The type of garbage produced in Guwahati city varies with the season and location due to economic and social variables. In order to develop and put into practise efficient treatment strategies for a reduced amount of discarded trash, it is essential to understand the composition of solid waste. The findings indicate that the solid packaging waste could be economically and potentially recovered with careful and effective MSWM planning. Increasing separate collection efforts in the area in accordance with the techniques used in the research region would not only improve the volume of package garbage collected and the profit margins associated with authorised enterprises, but would also reduce the overall volume of disposed waste. As a result, the cost of garbage disposal as a whole will reduce significantly and the ability to exploit trash as a supplementary raw resource will increase. Hence, the most crucial part of integrated waste management is the evaluation of waste quantities in a category format. These investigations need to be carried out

more frequently in order to assess advancements and related practical management techniques for the time-bound reduction in the MSW landfill content.

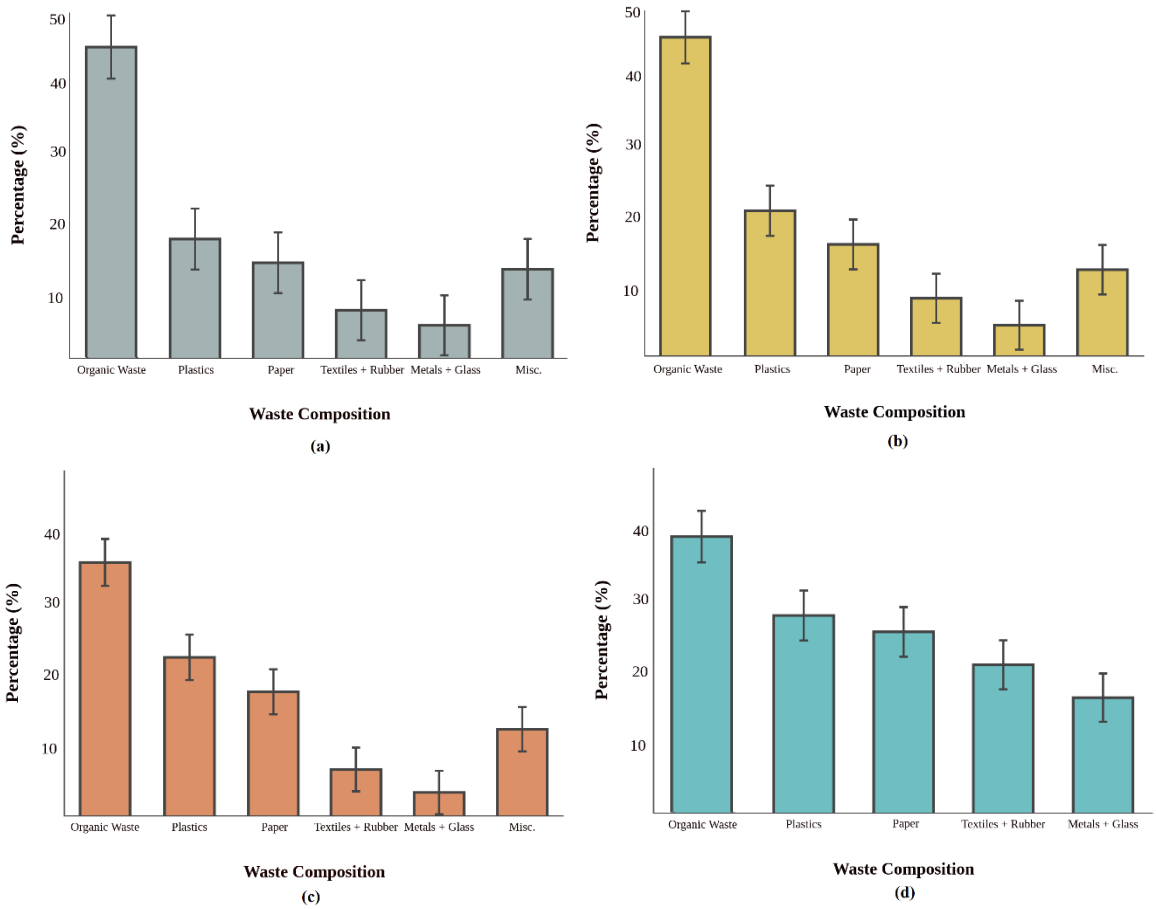


Fig. 3.4 Bar chart depicting category wise percent average solid waste compositions of four sub-areas of the Guwahati city **(a)** A1 **(b)** A2 **(c)** A3 and **(d)** commercial area

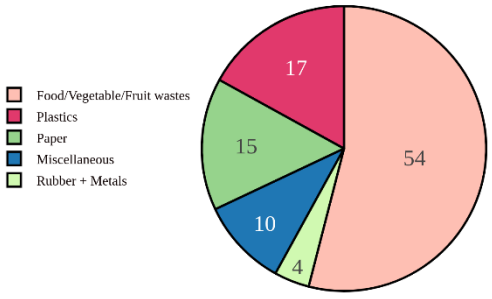


Fig. 3.5 Pie-chart depicting overall average percent composition of MSW in the Guwahati city

Table 3.4 Survey based waste type and quantity data for three sub-areas and three seasons of the Guwahati city.

	A1						A2						A3						Commercial Establishment					
	Feb		Jun		Aug		Feb		Jun		Aug		Feb		Jun		Aug		Feb		Jun		Aug	
	HH	***kg***	***HH***	***kg***	***HH***	***kg***	***HH***	***kg***	***HH***	***kg***	***HH***	***kg***	***HH***	***kg***	***HH***	***kg***	***HH***	***kg***	***Estd.***	***kg***	***Estd.***	***kg***	***Estd.***	***kg***
	4	15.3	5	20.0	5	23.5	4	13.1	2	15.1	5	18.5	4	11.7	3	14.3	3	11.4	4	22.5	5	34.1	3	19.6
	3	18.8	2	14.4	3	27.5	5	21.7	2	21.9	4	15.8	3	13.9	3	11.1	2	13.3	5	24.4	3	23.4	2	15.5
	6	13.1	2	17.2	5	24.2	3	15.4	5	29.5	6	16.7	3	11.2	2	8.0	5	15.2	4	26.8	3	22.7	4	29.2
	3	20.4	4	19.9	2	18.0	4	14.6	4	24.8	3	21.3	2	14.1	2	10.3	3	9.0	3	22.8	4	21.8	5	34.7
	2	8.3	5	21.5	4	21.2	5	17.3	5	26.4	2	14.1	4	9.4	4	9.1	3	11.2	5	33.1	2	17.0	3	25.4
	4	11.6	6	27.2	5	15.1	2	6.8	3	19.0	4	18.4	2	14.6	5	13.2	2	8.4	5	29.2	4	31.6	4	21.9
	5	17.7	2	9.3	3	20.8	5	15.9	4	22.3	4	12.2	5	16.6	2	8.9	4	12.6	3	18.9	3	16.0	4	27.4
Total	**27**	**105.2**	**26**	**129.5**	**27**	**150.3**	**28**	**104.8**	**25**	**159**	**28**	**117**	**23**	**91.5**	**21**	**74.9**	**22**	**81.1**	**29**	**177.7**	**24**	**166.6**	**25**	**173.7**

Table 3.5 A summary of the percentage distribution of alternate solid waste type fractions among four sub-areas and three seasons of the Guwahati city.

Waste Fraction (%)	**A1**				**A2**				**A3**				**Commercial Area**			
	Feb	**Jun**	**Aug**	**Weighted**	**Feb**	**Jun**	**Aug**	**Weighted**	**Feb**	**Jun**	**Aug**	**Weighted**	**Feb**	**Jun**	**Aug**	**Weighted**
Organic	49±2	51±1.3	54±1.8	51±1.1	50±1.4	52±2	55±0.3	52±1.2	36±1.1	39±0.9	40±0.5	38±0.8	52±3.4	54±2.8	58±2.6	54±3.0
Paper	13±3	15±1.5	12±2.6	13±1.3	14±1.3	14±1.9	16±1.1	14±1.2	11±0.4	13±0.3	12±0.5	12±0.4	11±5.7	12±6.8	14±4.9	12±5.8
Plastics	18±0.5	19±1	16±2.1	17±0.8	21±0.6	22±0.7	18±0.8	20±0.7	15±0.2	13±1.2	12±0.3	13±0.6	19±6.2	21±8.8	20±6.7	19±7.2
Textiles + Leather + Rubber	9±0.2	7±0.5	9±1.3	8±0.9	10±1.9	5±1.3	4±2.5	6±1.9	12±0.4	8±0.2	4±1.9	8±0.8	11±3.5	6±5.8	5±3.8	7±4.3
Glass	2±0.3	1±0.6	1.5±0.3	1.5±0.4	2±1.3	3±1.1	2.5±0.9	2.5±1.1	6±3.5	8±0.6	7±0.8	7±0.6	3±1.5	2±1.9	2±1.8	2.3±1.7
Metals	0.3±0.8	0.4±0.9	0.4±0.5	0.36±0.8	0.4±0.9	0.3±0.7	0.2±1	0.3±0.8	7±0.6	6±0.2	9±0.3	7±0.4	2±1.9	3±2	3±2.2	2.6±2.0
Miscellaneous	8±0.7	6±0.3	7±0.4	7±0.5	2±1.4	4±2	4±1.9	3.3±1.6	12±0.8	12±0.3	14±0.4	12±0.5				
Total waste generation	**99.3±1.0**	**99.4±1.1**	**99.9±0.8**	**99.5±0.9**	**99.4±0.4**	**99.3±0.2**	**99.7±0.7**	**99.4±0.3**	**99±0.6**	**99±0.9**	**99±0.8**	**99±0.7**	**98±0.9**	**98±1.3**	**99±0.8**	**98.3±1.0**

Table 3.6 A summary of best data reported in this work and prior art in the research theme of survey based solid waste characterization.

S. No.	Study area and Approaches	Participants	Inferences	Authors
1	Guwahati, India Survey, correlation and statistical analysis	227 HHs (middle, high, low-income groups) and commercial area (restaurants, shops)	• HH characteristics, age, education level, number of employees as well as seasonal variations were studied. • Number of HH and population resulted strong correlation for waste generation • At all three socioeconomic categories, a daily average of 0.201 kg of garbage was produced per person. • From August to January, waste creation was seen to be on the decline, before slowly beginning to rise in the month of June.	This work
2	Dehradun, India Survey analysis	144 internal residential settlement plans and workouts	• The rates of HHW production varied from 24.5 to 4147.1 g day^{-1}. • HH sizes, consumption patterns were not discussed in the study	Suthar and Singh (2015)
3	Lahore, Pakistan Sample collections	84 samples, self-classified monthly income and family members	• For groups with high, middle, and low incomes, respectively, waste generation is 1.1, 0.56, and 0.39 kg $capita^{-1}$ day^{-1}. • Amount of urbanisation, HH income, and population density weren't taken into account.	Kamran et al. (2015)
4	Veles, Macedonia Survey	56 families	• Influential parameters such as seasonal characterization was not considered	Hristovski et al. (2007)
5	Chihuahua, Mexico Sample collections	80 HHs from three socio-economic groups	• Average urban waste generated was 0.676 kg $capita^{-1}$ day^{-1} • A small number of criteria were adopted as a result of small-scale surveys (conducted at the level of HH or city units).	Filmer et al. (2001)

3.7 Summary

In this chapter, the sensitive influence of geographical area, HH income, and season on the generated MSW variation in Guwahati city has been analyzed. Three specific time periods, namely February 2021, June 2021, and August 2022, were chosen to analyze the amount of MSW produced in different areas of Guwahati, including the commercial district and sub-areas A1, A2, and A3, along with the residents

with different income levels. During these times of the year, interviews were conducted for 227 HHs. Notable findings of the study are as follows. The waste disposal classification analysis conveyed that 51.62% of the solid waste was OW, 24.22% of it was package waste (papers and plastics), 8.13% of it was burnable waste (clothes, diaper, rug, bag, shoes, slippers, pillow, carpet), 4.62% of it was park and garden waste, 3.54% of it was metals and glass waste, and 7.87% of it was miscellaneous. It was analyzed that the OW rates were adversely correlated with the income level after accounting for alterations in waste components as per the income level. The case with the greatest income level had the highest concentration of OW (A2).

For all three socioeconomic categories, a daily average of 0.201 kg of garbage was produced per person. The findings show that more wealth leads to more produced solid waste. They were 0.18, 0.23, 0.17, and 0.516 kg capita^{-1} day^{-1} for A1 (middle income), A2 (high income), and A3 (poor income) and commercial area, respectively. It was found that the generated waste reduced from August to January and then gradually increased upto the month of June. This was due to seasonal alterations. Since results for February (low temperature season) were 18.76% lower than in June, it can be inferred that there is a considerable difference between the waste flows sampled at various dates in two years for the Guwahati city. The best available reported research (Hristovski et al., 2007; Filmer et al., 2001; Suthar and Singh, 2015; Kamran et al., 2015) did not consider factors such as seasonal variations, HH sizes, consumption patterns, the extent of urbanization, HH income, and population density. Additionally, random stratified sampling was not explored in previous studies. Therefore, the fulfilled objective improved the know-how of the subjective findings by considering the mentioned factors. Thereby, the findings enhanced validity and comprehensiveness of the reported data trends and paved the way for a pragmatic MSWM.

Chapter 4
Prediction and Forecasting MSWG Rate using ML Algorithms

Chapter 4

Prediction and Forecasting of MSWG rate in Guwahati city with Machine Learning Algorithms

This chapter elaborates upon the findings related to the application of ML algorithms for the accurate estimation of current and future solid waste generation rates of the Guwahati city. The ML models were formulated by mapping solid waste quantities at the municipal level with socio-economic and demographic variables of the Guwahati city. After presenting a brief introduction in section 4.1, the chapter delineates upon the data attributes of the independent and dependent variables after data pre-processing and outlier removal strategies in section 4.2. Section 4.3 elucidate upon the correlation analysis to identify and prioritize socio-economic parameters that exhibit strong correlation with the dependent variable. Sections 4.4 and 4.5 respectively elucidate upon the findings of the predictive modelling and forecasting results of MSWG in the city. Each of these sections aim to identify the best fit model for the prediction of MSWG rate. A comparative assessment with respect to literature data and practical implication of the modelling study have been presented in sections 4.6 and 4.7 respectively. Finally, section 4.8 summarizes the key findings associated to the realization of the second objective of the thesis[#].

Overview

ML-based prediction of the total MSWG rate have been targeted to simultaneously minimize prediction uncertainty and enhance prediction vigor. Firstly, Among the socio-economic parameters, population growth, HH counts, LP, and GDDP indicated strong positive correlation index. Secondly, among all algorithms, the RF and GB with response train and test scores of 0.90 and 0.94 respectively performed well and among these two, the GB had a model accuracy of 97%. Overall, the GB model ($r = 0.94$, $R^2 = 0.99$, RMSE = 3.01, MAE = 2.86, and IOA = 0.94) surpassed the RF model ($r = 0.90$, $R^2 = 0.98$, RMSE = 83.21, MAE = 74.84, and IOA = 0.72), and DT ($r = 0.82$, $R^2 = 0.97$, RMSE = 325.82, MAE = 302.20, and IOA = 0.45) performance. Thirdly, the R^2 and RMSE values for EMA-GB and SMA-GB were 2.12 and 3.83 respectively and confirmed their statistical excellence. On the contrary, the RMSEs for EMA-DT and WMA-DT were 4.22 and 11.45 respectively. This, affirmed indifference among the models and, the GB exhibited the best performance to predict and forecast total MSWG rate for the Guwahati city.

4.1 Introduction

This chapter targets the efficacy of tree-based models (GB, RF, and DT) and time series ML modelling algorithms for city wide (Guwahati) MSWG prediction and forecasting respectively. The central objective was to develop a robust model and facilitate a comparative analysis and long-term model

[#] Published article: Singh, T., & Uppaluri, R. V. S. (2022). Machine learning tool-based prediction and forecasting of municipal solid waste generation rate: a case study in Guwahati, Assam, India. International Journal of Environmental Science and Technology. https://doi.org/10.1007/s13762-022-04644-4

performance evaluation. The specific novelty being targeted was improved accuracy, and enhanced capabilities for the prediction and forecasting of MSWG rates. These algorithms leverage the inherent structure and relationships within the data to provide more robust and efficient predictions. Accordingly, innovative and valuable tools in waste management research can be sought. Comparatively, the effectiveness of the ML methods with respect to the classical statistical methods was also sought. Python version 3.8 was deployed to analyze the data and implement the prediction models. To automate loading, pre-processing and integration of data, python scripts were developed. NumPy, Scikit-learn, Pandas and Matplotlib library packages were deployed to implement modelling efforts on a i7-4790 CPU @ 3.60 GHz processor configuration. To analyze the data, DT was used as a first model that deployed ML as a non-parametric algorithm to model data separation limitations based on the learning decision rules on the input characteristics of the model. Following this, RF and GB were considered as the second and third predictive models for the exploration of associated correlations among socio-economic parameters. Moreover, to address the model overfitting issues, the chosen models serve better towards the prediction of MSWG rate. Socio-economic factors, namely population growth, GDDP, LP, total HH, HH_{sizes}, and WP were considered along with time to influence the MSWG rate prediction. The forecasting was carried out using simple moving average (SMA), weighted moving average (WMA), and exponential moving average (EMA) approaches. The model performance was assessed in terms of five metrics namely r, R^2, RMSE, MAE and IoA. For better accuracy results, the models were fine-tuned with 10-fold cross validation approach.

4.2 Data Characteristics

Table 4.1 summarizes the datasets before pre-processing. Prior to this, the MSWG data from CPCB constituted 408 data points. Also, census and MOSPI data referred to 360 data points. After pre-processing, the CPCB, census and MOSPI data were reduced to 376 and 312 respectively. The MSWG rate of Guwahati city increased steadily since 1991 (Fig. 4.1(a)) and affirmed a steady but not large variation in the waste generation pattern. This is in agreement with the literature reported hypothesis that the population rise and habitat improvisation together contribute to the proportional enhancement in the MSWG rate. To be specific, the annual volume of MSWG in Guwahati city increased approximately two-fold from 1.48 MT (metric ton) in 1991 to 2.13 MT in 2016.

Table 4.1 Conditioned and integrated datasets summary for the ML-based prediction of total MSWG rate.

Dataset	*Dependent Variable*	*Independent Variable(s)*	*Number of Datapoints*
MSWG	CPCB (1991-2016)		376
Population growth		Census Report (1991-2011)	312
LP		Census Report (1991-2011)	312
GDDP		Ministry of Statistics and Program Implementation (1991-2011)	312
HH (Count & Size)		Census Report (1991-2011)	312
WP		Census Report (1991-2011)	312

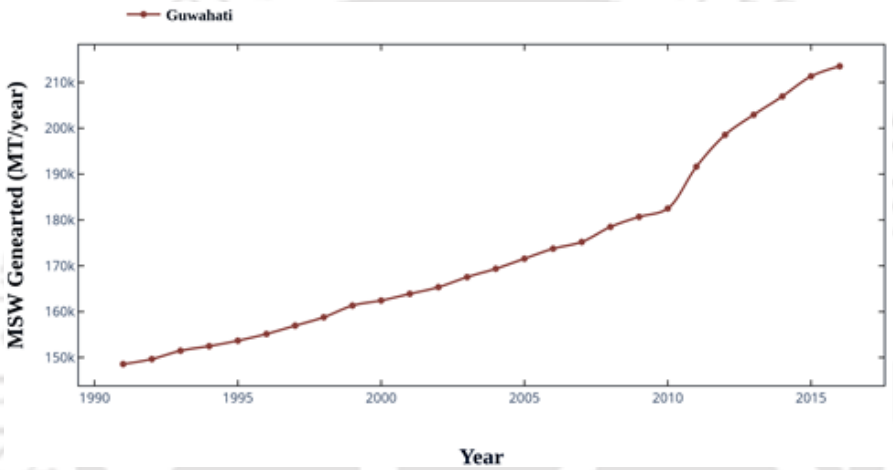


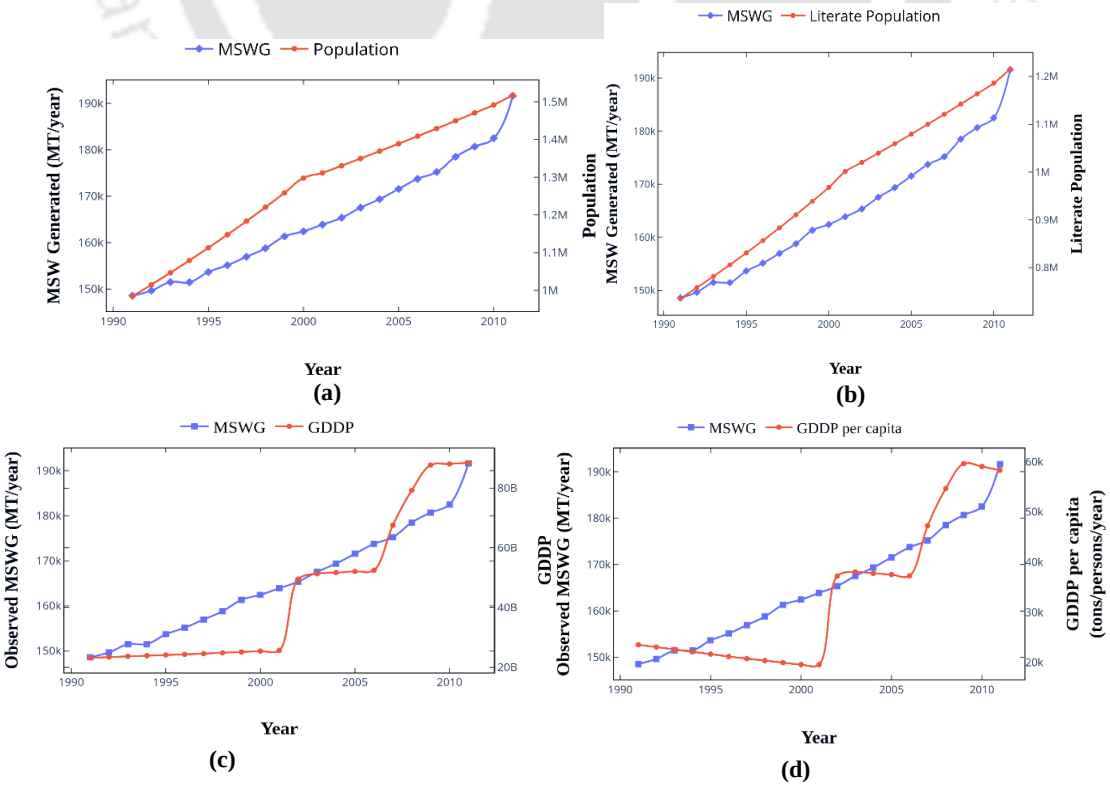
Fig. 4.1 Graph depicting CPCB data based MSWG rate trends of the Guwahati city

After data conditioning and integration, the socio-economic parameters were reduced to 312 data points for each parameter. Fig. 4.2 (a - g) depict the respective socio-economic factor variation with year along with the MSWG rate. The plots depicting MSWG rate alteration along with population parameters namely population growth and LP illustrated that a steep increase in the MSWG rate was apparent and with a positive trend (Fig. 4.2 (a-b)). In the recent past, due to economic growth, enhanced urban population increased the dramatic urban growth and land-use change and henceforth increased the MSWG rate (Mahini & Gholamalifard, 2006).

Fig. 4.2 (c) and (d) respectively depict the MSWG trend along with GDDP and $GDDP_{per\ capita}$ for the Guwahati city. The figures clearly illustrated that for 2002 and 2006-2008, a high non-linear variation exists for the GDDP and $GDDP_{per\ capita}$. However, the MSWG rate enhanced steadily during this period but enhanced steeply in 2010-2011. These variations were due to the pertinent impact of

globalization past the Indian economy on the city's income demographics. In the chapter, the GDDP was considered as an indicator of economic growth due to the fact that it is likely to be sensitive to affirm financial capacity for the payment towards environmental improvement.

The comparative trends of MSWG rate and HH indicators have been shown in Fig. 4.2 (e) and (f), respectively. All figures affirmed that while HH counts and sizes enhanced steadily, the MSWG also enhanced steadily but with a steep and non-linear increase in the year 2010. The correlation indicates that with the total increase of HH counts and HH_{Size}, the MSWG rate increases. Further, the HH in good condition affirm more waste generation than those with livable and dilapidated conditions. Similarly, HH sizes with 4-8 members direct more waste generation than HH_{Size} 1-3 and above 9. Fig. 4.2 (g) depicts the employment scenarios for Guwahati city along with the MSWG rate. The data confirms more non-WP in the city. This includes students, pensioners, dependents, beggars, vagrants, etc. However, for this parameter, the declining trend can be seen since 2001, along with a marginal reduction trend in the MSWG rate. Despite all this, the steep and non-linear enhancement in MSWG rate existed in 2010 and this is due to strong variations in habitat and lifestyle parameters. The MSWG rate can be analyzed to be positively correlated with the employment status statistics. However, more solid waste has been generated by the main WP to better relaxation in the economic constraints.



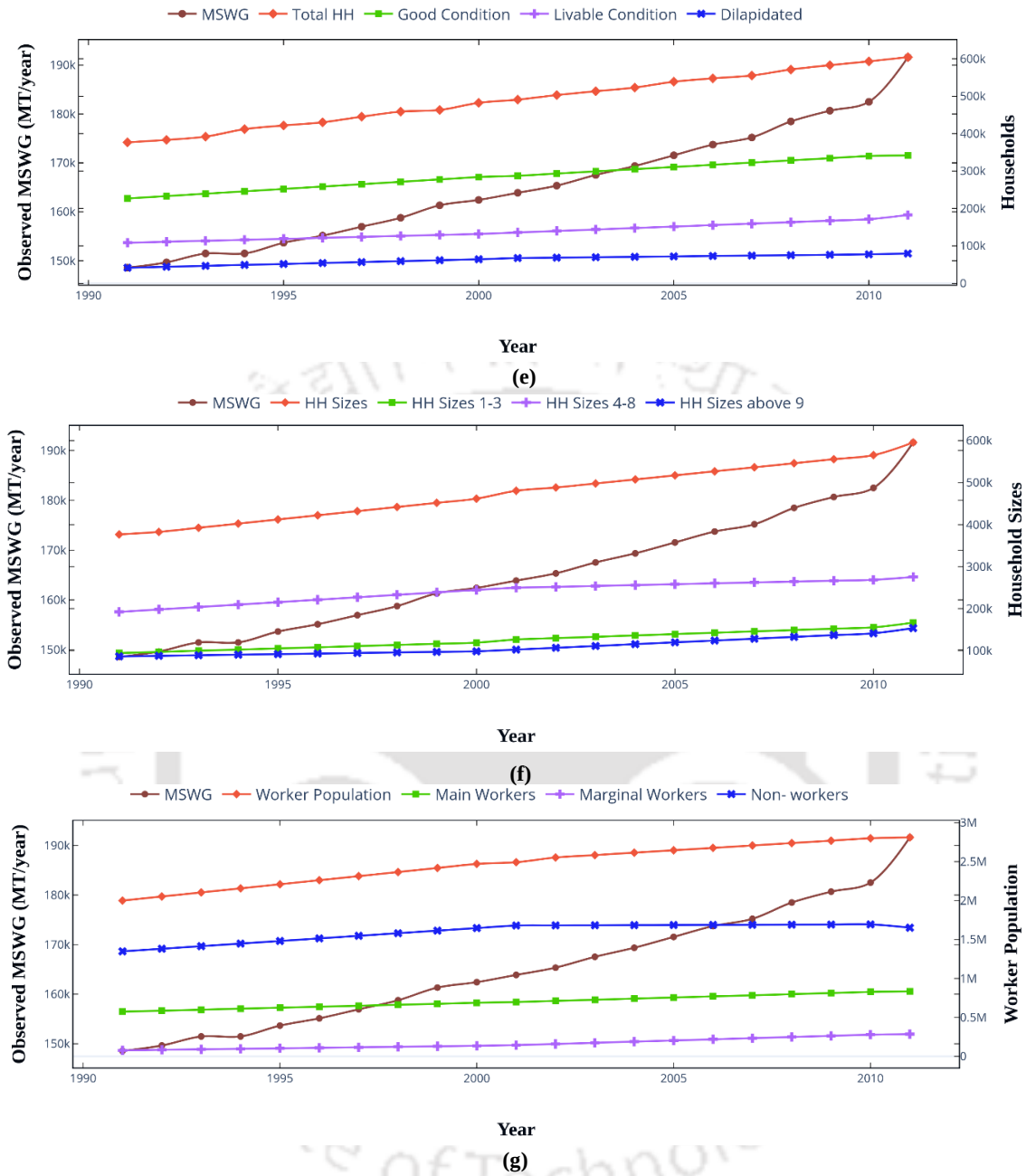


Fig. 4.2 (a - g) Time dependent variation of MSWG and socio-economic factors for the Guwahati city

4.3 Correlation Analysis of the Data

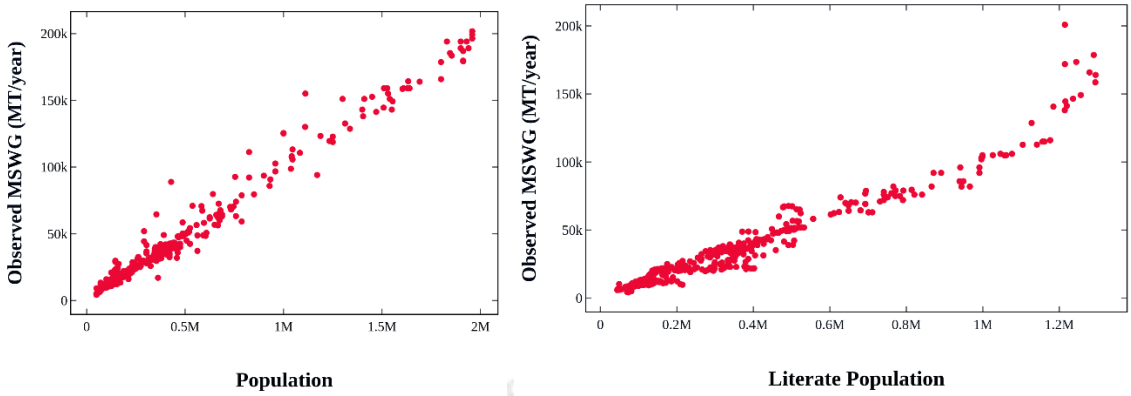
The correlation analysis was primarily conducted to determine and rank the socio-economic parameters with high correlation indices having dependent variables. Accordingly, those parameters were isolated with weak or no correlation with each other. Such an analysis would be helpful to obtain best relevant information of the modelling efforts. To do so, a correlation matrix that estimates the correlation index between each variable pair was determined and analyzed.

For the case, Table 4.2 summarizes the correlation matrix for MSWG and mentioned socio-economic parameters. The first column in the table confirms that among all parameters, the population growth affirmed a maximum *r* of 0.87. Fig. 4.3 conveys this positive correlation between population growth and MSW to accordingly infer that the waste generation rate got positively influenced with the population. The positive trends of GDDP-MSWG and WP-MSWG have been consistent with the inferences indicated in the relevant prior art (Bandara et al., 2007). Also, LP-MSWG, GDDP-MSWG, $GDDP_{per\ capita}$-MSWG, HH, HH_{Size} and WP with respect to MSWG have all affirmed positive correlation. Further, GDDP-WP, HH-HH_{Size} , GDDP-$GDDP_{per\ capita}$, LP-WP parameter pairs have been observed to be significantly correlated with one another. Thereby, to subdue the multicollinearity issues, such model pairs were avoided in due course of the modelling effort. Scattered plots of population growth, LP, GDDP, HH counts and sizes vs. MSW further illustrate such pairing effect (Fig. 4.3 (a)-(m)).

Following this, HH, LP and HH_{Size} can be analysed to possess higher correlation coefficients of 0.85, 0.82 and 0.81 respectively. Moreover, the number of HH and its sizes were found to be positively correlated with MSWG. Accordingly, this observation validated the hypothesis that the size of a city influences per HH waste quantities or considered socio-economic factors. Similarly, the WP also exhibited a positive correlation with respect to the MSWG rate. Altogether the table reflected an overall positive correlation on all variables with the lowest $GDDP_{per\ capita}$ (r = 0.72).

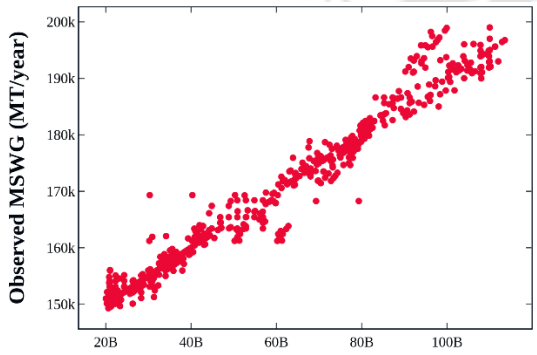
Table 4.2 Correlation Matrix data of socio-economic parameters and total MSW generation rate per household for the Guwahati city.

	MSWG	***Population growth***	***LP***	***GDDP***	$GDDP_{per\ capita}$	***HH***	HH_{Size}	***WP***
MSWG	1							
Population growth	0.87	1						
LP	0.82	0.65	1					
GDDP	0.76	0.69	0.68	1				
$GDDP_{per\ capita}$	0.72	0.64	0.61	0.75	1			
HH	0.85	0.84	0.54	0.64	0.66	1		
HH_{Size}	0.81	0.65	0.51	0.61	0.58	0.66	1	
WP	0.74	0.68	0.71	0.75	0.76	0.61	0.56	1

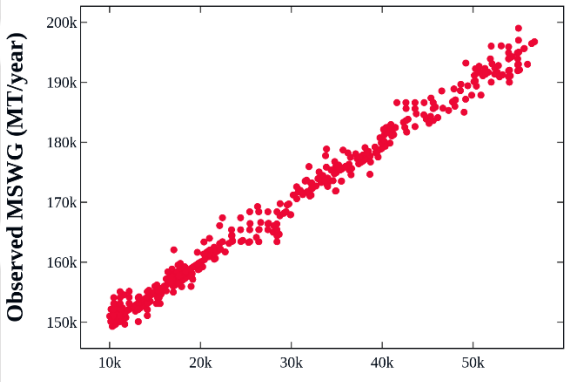


(a)

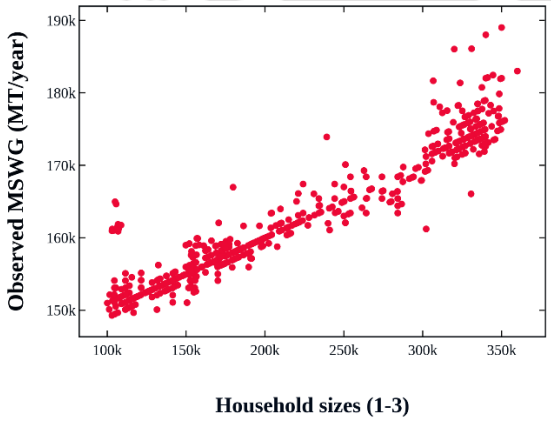
(b)



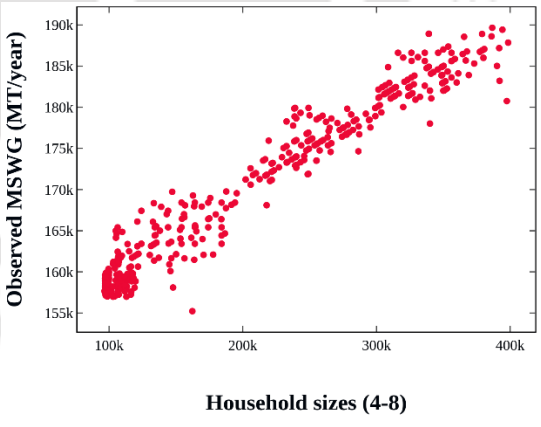
(c)



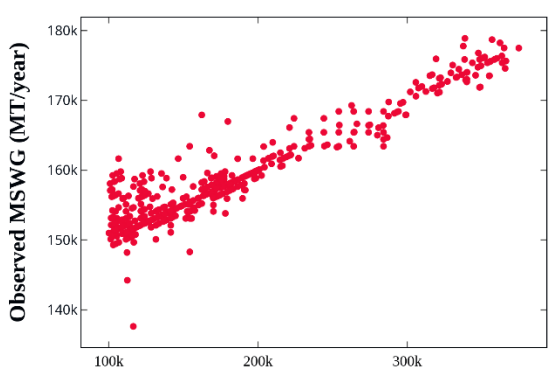
(d)



(e)

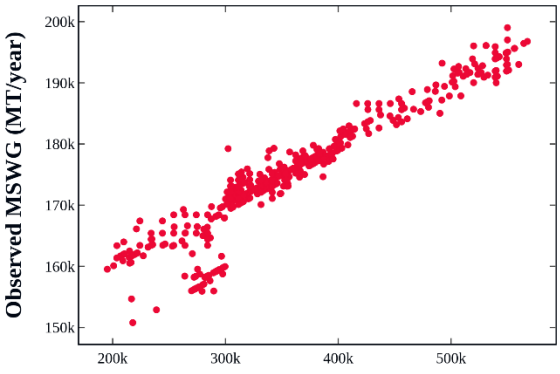


(f)



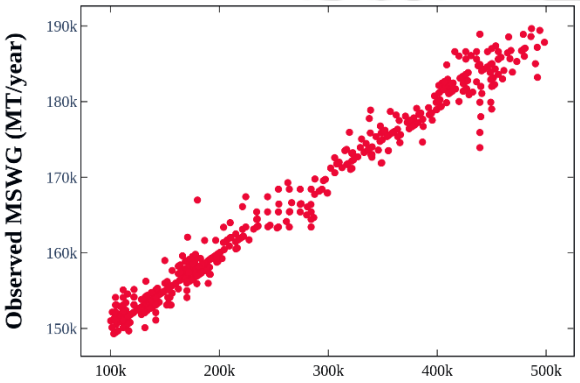
Household sizes (Above 9)

(g)



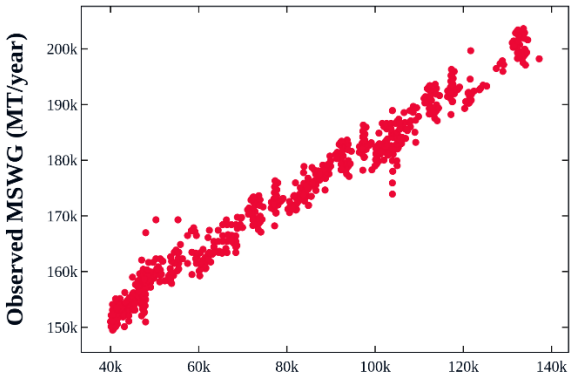
Number of Households (Good Condition)

(h)



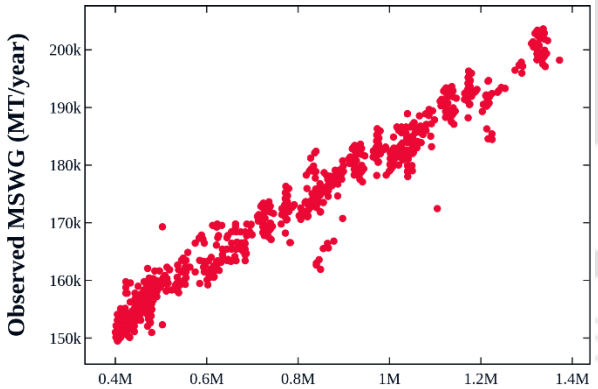
Number of Households (Livable Condition)

(i)



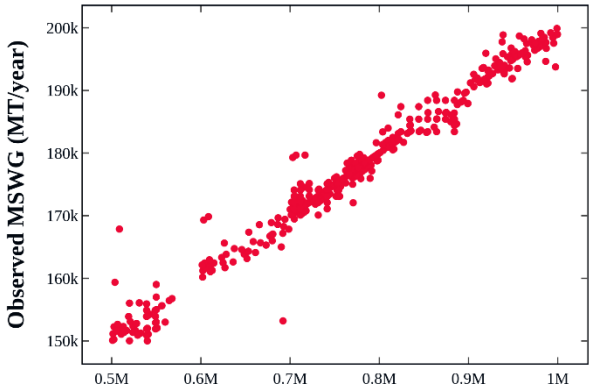
Number of Households (Dilapidated Condition)

(j)



Worker Population (Main Worker)

(k)



Worker Population (Marginal Worker)

(l)

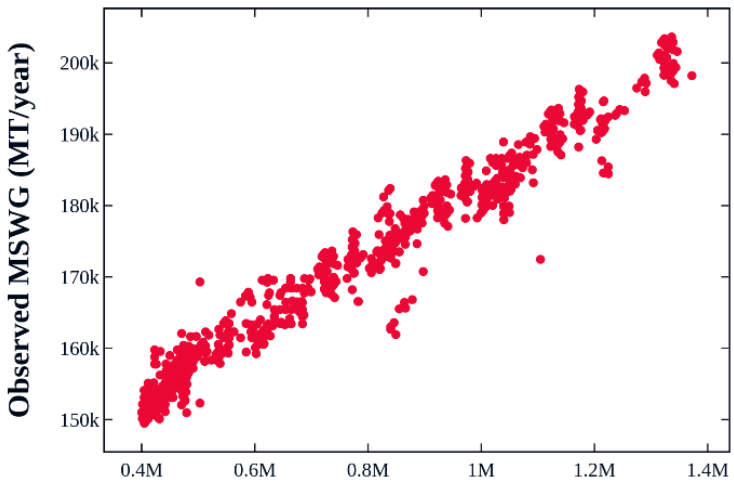


Fig. 4.3 (a)-(m) Scatter plots depicting the strength of correlation indices of total MSWG rate with respect to socio-economic parameters and other factors of Guwahati city. Abbreviations and acronyms. MT: metric ton

4.4 Modelling and Prediction of MSWG Rate

According to the ranking, the socio-economic and demographic factors were first to be incorporated one-by-one in the model. Table 4.3 summarizes the RMSE variation for training and testing cases for various combinations of considered socio-economic and demographic factors in the model. For the case studies, error analysis was obtained with a step wise LR based modelling approach. For the GB, the training RMSE reduced to 198.04 to 154.19 for a consideration of mentioned socio-economic parameters but not HH_{size} *and WP*. Corresponding trends in the testing model RMSE indicated a reduction from 111.60 – 104.56 but without considering HH_{size} *and WP*. Similar inferences can be obtained for the RMSE trends of RF and DT models and for both training and testing model cases. In summary, the training error reduced with the inclusion of more and more socio-economic parameters and this occurred due to adequate enhancement in the model complexity. However, the testing error ceased to reduce for the testing model case that refers to the model performance evaluation for new data after consideration of HH parameter. Thus, it can be inferred that the MWSG prediction rate is sensitively influenced with population growth, LP, GDP, $\text{GDDP}_{\text{per capita}}$ and HH and inclusion of other additional socio-economic parameters does not improvise the model performance.

Table 4.3 Training and testing RMSE data summary for total MSWG rate prediction for various combinations of independent variables.

Model	Training RMSE (MT year^{-1})			Testing RMSE (MT year^{-1})		
	DT	***RF***	***GB***	***DT***	***RF***	***GB***
Population growth	287.17	212.48	198.04	264.97	130.19	111.60
Population growth, LP	277.82	203.55	189.46	250.98	123.86	109.37
Population growth, LP, GDDP	273.19	185.18	176.01	235.79	115.41	105.87
Population growth, LP, GDDP, $GDDP_{per\ capita}$	265.81	177.56	162.81	235.33	113.18	105.21
Population growth, LP, GDP, $GDDP_{per\ capita}$ ***, HH***	252.94	173.24	155.19	234.83	110.97	104.56
Population growth, LP, GDP, $GDDP_{per\ capita}$ ***, HH,*** HH_{size}	251.19	170.17	154.03	234.67	110.53	104.22
Population growth, LP, GDP, $GDDP_{per\ capita}$ ***, HH,*** HH_{size} ***,WP***	250.62	169.15	153.23	234.45	110.19	104.12

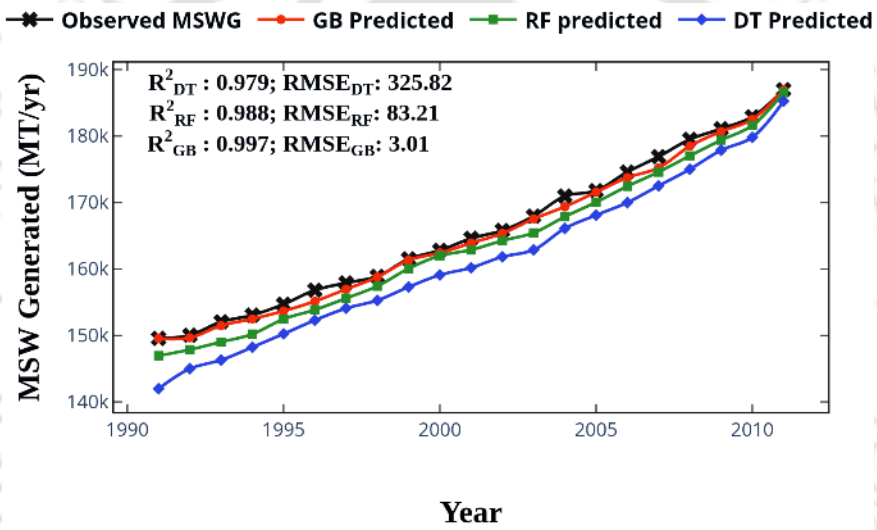


Fig. 4.4 Line plots depicting the performance of DT, RF and GB algorithms for total MSWG rate prediction

To assess upon the developed model accuracy, a comparison of secondary and simulated data was conducted for the considered observation period (1991–2011). In general, the obtained results indicate good accuracy of the modelling effort and infer upon its validity for the forecasting of the MWSG. For the Guwahati city, Fig. 4.4 depicts a comparative trend analysis plot for the secondary data and DT, RF and GB model predicted data. The results indicate that in comparison with the DT, the GB and RF fairly track historical developments in the MSWG patterns.

The grid search based ten-fold cross validation for all models was evaluated for its efficacy (70:30 for training : testing cases of the dataset) or the considered DT, RF and GB tree-based models, quality results have been obtained. The training and testing case studies for DT, RF and GB, indicated scores

of 1.00, 0.98 and 0.99 and 0.92, 0.95 and 0.98, respectively. The maximum depth of DT, RF and GB during the training phase were 28, 57 and 48 respectively. Among these, the DT produced the best train score. However, the variation between train and test scores has been observed to be higher for the DT (0.6 value) in comparison with the RF (0.2) and GB (0.1) models. This is due to the fact that since DT is capable towards overfitness associated to the application of greedy algorithms, the optimal tree may not be found for few cases. In other words, the DT model learnt the training data too well and could not be therefore generalized towards the pertinent needs. On the contrary, RF and GB have been capable to fix the overfitness issue that exists in the DT and this is confirmed in the observed trends. Fig. 4.4 depicts the MSWG rate prediction by DT, RF and GB models in conjunction with the secondary data of Guwahati city. The trends affirmed the closer vicinity of data trends of RF and GB in comparison with the DT. Further, among all models, the GB exhibited a promising performance with an R^2 and model error (calculated using data partitioned ratio) value of 0.997 and 3.01 respectively.

Few statistical indicators have been used to assess upon the performance of the investigated models DT, RF, and GB. Table 4.4 and Fig. 4.5 (a-c) exhibit and illustrate the relevant findings and the fit charts. Being the least complex model in the tree-based ML algorithms, DT performed average with good (R^2 = 0.97), and with a linear fitting relationship. Based on the DT, the ensemble learning algorithms RF and GB exhibited good performances. GB performed superior to other models with an R^2 = 0.99. Thus, as a state-of-the-art ML model, it had promising performance due to parallelization, sparse data handling capacity, and ability to avoid overfitting. With an R^2 of 0.98, RF also performed well. As a result, the GB can be used for further predictions, and other models can be considered as supplementary. The IoA for the models were also calculated. This refers to the ratio between the MSE and the potential error.

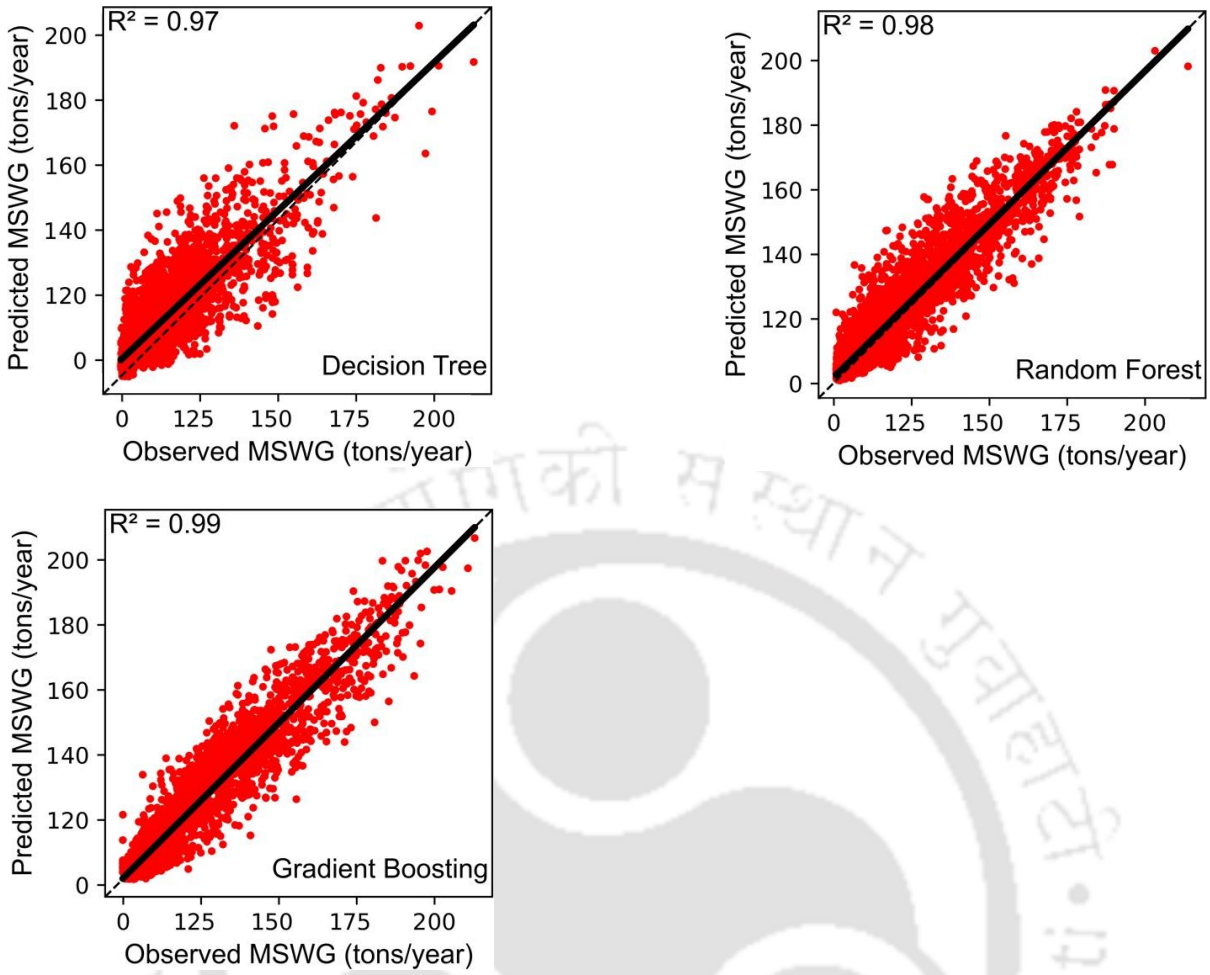


Fig. 4.5 Scatter plot depicting alternate ML algorithms performance **(a)** DT, **(b)** RF and **(c)** GB

Among the mentioned tree-based models, DT had a low train (0.86) and test score (0.81) and had a R^2, RMSE, MAE and IOA of 0.97, 325.82, 302.20 and 0.45 respectively. GB resulted in higher train (0.94) test score (0.91) and reduced RMSE (3.01). Hence, it was able to provide an accurate prediction (Table 4.4). Along with the train and test scores, the RMSE (83.21) and MAE (74.84) have been better for RF as well in comparison to actual values but are high with respect to other models. The overall evaluation and comparison of the three ML algorithms confirm that GB is the best-suited ML model.

Table 4.4 A summary of alternate ML models performance indices for total MSWG rate prediction.

Algorithms	Train Score	Test Score	R^2	RMSE	MAE	IoA
DT	0.86	0.81	0.97	325.82	302.20	0.45
RF	0.90	0.85	0.98	83.21	74.84	0.72
GB	0.94	0.91	0.99	3.01	2.86	0.94

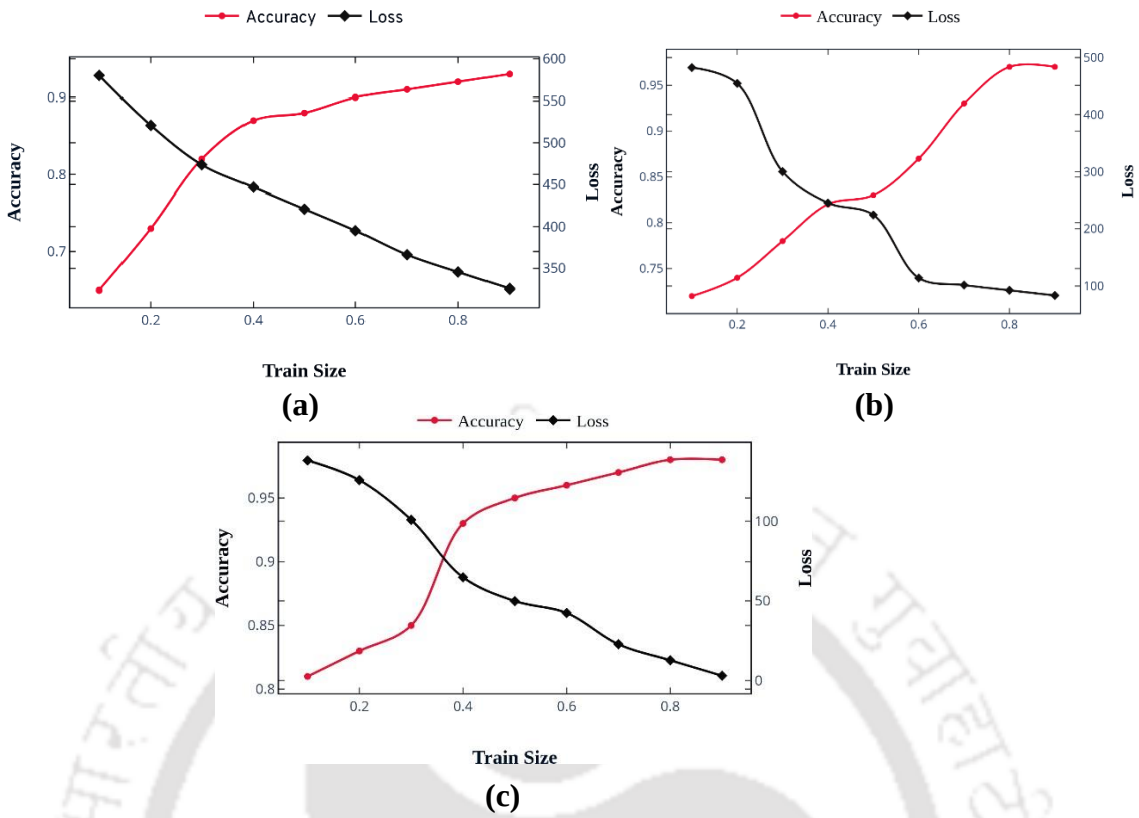


Fig. 4.6 (a-c) Accuracy and loss curves of **(a)** DT **(b)** RF and **(c)** GB algorithms for MSWG rate prediction

Fig. 4.6 (a-c) depict accuracy and loss graphs for the performance evaluation of the suggested model. The accuracy and loss have been represented as the y-axes in these plots that considered the sample size (percentage) as the x-axis. The number of training cycles being adopted for the complete dataset or a percentage of training sets have been used to represent the x-axis. For a variation in train size, the y1-axis for the accuracy varied as 0.65-0.93 (DT); 0.72-0.97 (RF) and 0.81-0.98 (GB) and the loss curves y2-axis decreased for DT (580.01-325.82); RF (482.55-83.21) and GB (138.09-3.01). A deeper introspection into the accuracy graph reveals that, for a given set of small samples, during initial stages, each model curve grew quickly. Accordingly, DT, RF, and GB each had an accuracy of accuracies of 93%, 97%, and 98%, respectively. Also, both accuracy curves confirmed an upward trend with percentage samples. In fact, the first through third sample sizes support the scenario of a significant expansion. The growth rate then gradually declined until it achieved a stationary profile. Similarly, loss measurements convey upon the model limitations and henceforth its subpar performance. The statistics indicate that the loss (error) has been decreasing and hence higher model performance existed with steady increase in the fitness. Even though the larger time frame indicated that there have not been many

peaks and valleys, the reduction in loss over time inferred that the model has been successful to adapt and perform well.

Involving many attributes, the DT is a non-parametric technique being often applied for the recognition of complex patterns in even smaller data sets. The DT algorithm reduced the uncertainty associated to the identification of an unknown pattern. In comparison, the RF algorithms constitutes multiple DTs in which the DT model forms an ensemble with bagging. Thus, the RF model comparatively reduced the data variance and prevented the DT model's greater dependence on highly influential variables. Lastly, the boosting approach enabled the addition of newer models for their sequential ensembling. In essence, boosting attacks the bias-variance trade-offs through an initiation with a weak model. For instance, the DT model had only a few splits and was therefore effective for the good performance (in terms of sequential boosting) through its continued effort to build new trees. Thus, the GB's performance efficiency enhanced profoundly in comparison with the RF and DT.

4.5 Statistical Model-based Forecasting of the MSWG Rate

The MA methodology was considered in this work for the MSWG forecasting with statistical models. This was considered to facilitate noise minimization and outliers removal. MA is a statistical tool being adopted for the forecasting and analysis of time-series data. The method functions through the acquisition of long-range correlations (Molugaram, & Rao, 2017). For the GB, RF and DT, Fig. 4.7 (a)-(c) respectively illustrate the forecasted MSWG rate plots for the Guwahati city in the year range of 2012-2050. Three MA methods namely EMA, SMA and WMA were considered to obtain useful insights and analysis of the forecasted MWSG rate trends for the DT, RF and GB. For GB, the MSWG rate has been forecasted to vary as 194152-277035, 192750-274105 and 193580-270646 MT $year^{-1}$ for EMA, SMA and WMA cases, respectively and, in the year range of 2012-2050. Similarly, for RF, the MSWG forecasted rate varied as 189231-261759, 184209-256749 and 178973-252646 MT $year^{-1}$, and for DT, it was 187477-274012, 181704-269549 and 179934-263157 MT $year^{-1}$, respectively for EMA, SMA and WMA and in the year range of 2012-2050. For all models, it was observed that the EMA could better forecast the MSWG rate followed with SMA and then WMA. Also, to better evaluate the model accuracy, the MAE and RMSE evaluations were also considered in addition to the correlation

coefficient. Overall, a positive correlation exists for the waste generation with time. From the figures, the projected statistics convey that by the year 2050, approximately 277035 MT of waste would be generated in the city (daily per capita waste generation in the city is expected to increase by approximately 20% or more).

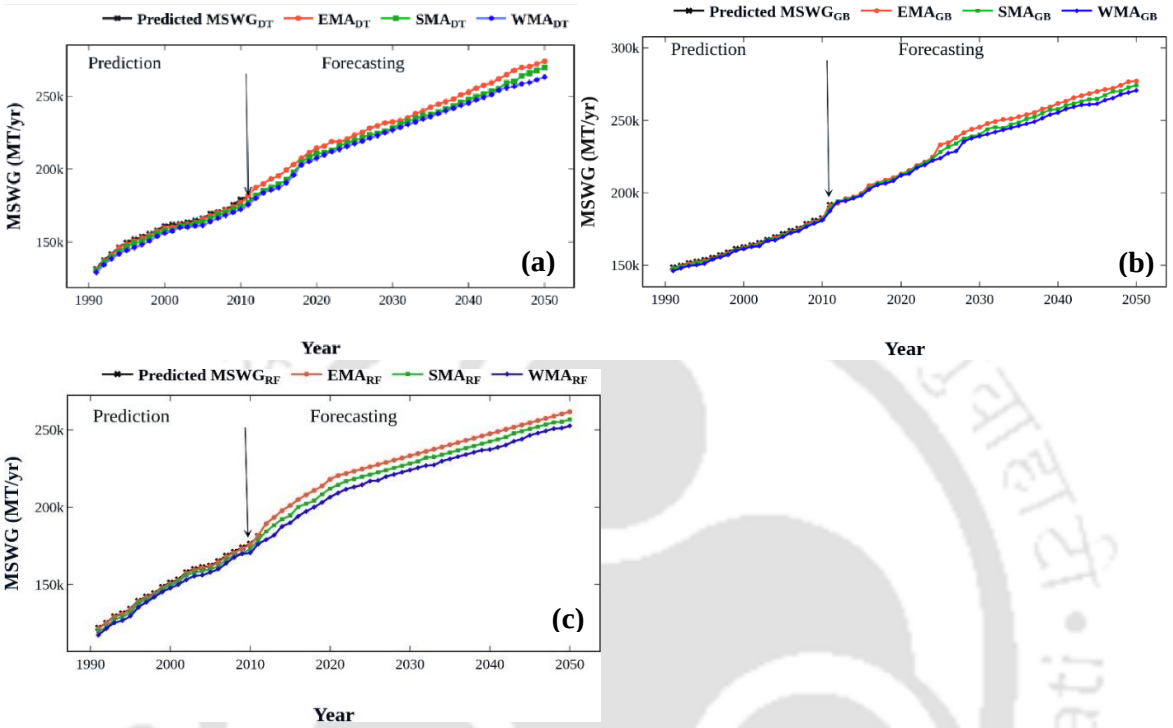


Fig. 4.7 EMA, SMA and WMA based predicted and forecasted total MSWG rate trends for **(a)** DT **(b)** GB and **(c)** RF algorithms

Fig. 4.8 (a-d) presents a comparison of the train and test scores for the models. For GB, the training score of 0.94 was obtained and for the RF and DT, the values are 0.90 and 0.86, respectively. Similarly, the test scores for GB, RF and DT were 0.91, 0.90 and 0.81, respectively. Fig. 4.8 (b), (c) and (d) depict the comparison of MAE, RMSE, and R^2 between observed and predicted MSWG respectively. The RMSE values for DT, RF and GB were 325.82, 83.21 and 3.01, respectively. The MAE results depict a similar trend for the RMSE (302.20, 74.89 and 2.88 respectively for DT, RF and GB). Additionally, the R^2 value for GB, RF and DT were 0.997, 0.988 and 0.979 respectively.

Thus, the results affirmed that the GB performed better than the RF and DT. Since RMSE is a measure of the goodness of the fit, lower MSE indicates lower error and hence better fit. Thus, the overall conclusion is that the GB is the best-suited algorithm to predict and forecast the MSWG rate. Also,

while RF and GB provided best results for all parameters, the RF at times repeated closer data prediction for few sets. On the contrary, the GB did not indicate such issues.

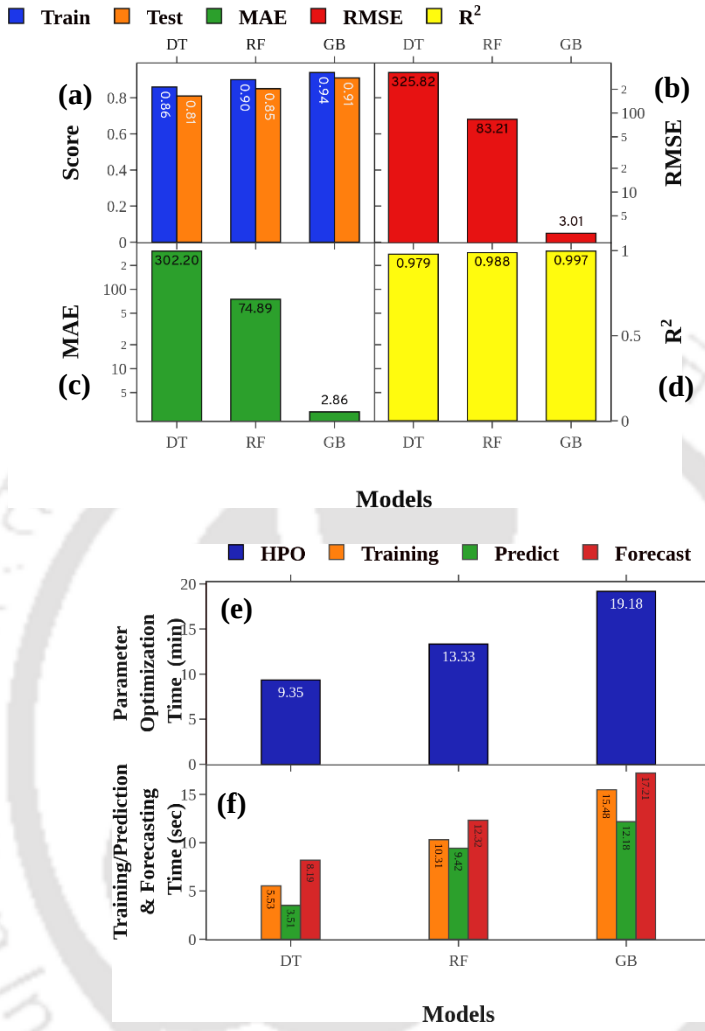


Fig. 4.8 Comparative depiction of **(a)** train and test scores, **(b)** RMSE, **(c)** MAE, **(d)** R^2 **(e)** HPO and **(f)** computational time (training, prediction and forecasting) for alternate tree-based models for prediction of total MSWG rate

The forecasted data were compared with the primary data obtained from the GMC of Guwahati city. The obtained data referred to the MSWG pertinent rate for the year range of 2016-2019. Table 4.5 (a) summarizes the model error in conjunction with the primary data. For model validation, the MSWG data source was collected from GMC and in the year range of 2016-2019. Thus, it can be easily understood that the GB indicated lowest forecasting errors (1.22 – 1.57%, for the year ranging from 2016-2019) in comparison with the RF and DT. Thus, while RF and DT models confirmed almost

similar error values, the RF was better stationed to predict the MSWG rates (also affirmed by its R^2 values). Henceforth, the validation confirms the better performance of GB and RF.

Table 4.5 (a) A summary of alternate ML model forecasting errors for the total MSWG rate forecasting case study.

Year	*MSWG*	*GB forecasted MT year^{-1}*	*GB Model error*	*RF forecasted MT year^{-1}*	*RF Model error*	*DT forecasted MT year^{-1}*	*DT Model error*
2016	207521	204985	1.22 %	203120	2.12 %	202172	2.57 %
2017	209812	206722	1.47 %	206115	1.76 %	205448	2.07 %
2018	212825	208825	1.87 %	207543	2.48 %	206433	3.00 %
2019	213761	210387	1.57 %	209215	2.12 %	208300	2.55 %

In comparison to the literature reported error values, the GB model was efficient to better predict the MSWG rate trends (Johnson et al., 2017). However, the DT model (with forecasting error range of 2.57-2.55%) did not perform well in comparison with the GB and RF. This may be attributed to the ability of the bagging and boosting algorithms for the learning of complex non-linear behaviour and subsequent optimization of model parameters. Due to its non-parametric nature, the DT model was able to build a single model for a given training data. On the contrary, the bagging (RF) and boosting techniques (GB) can produce multiple trees that efficiently combine their output. Hence, better optimization could be targeted to find the best model.

The ML solution needs to be highly reliable and accurate. To resolve the issues associated to the achievement of higher accuracy, the appropriate selection of several parameters is mandatory for the subsequent training of a model. To achieve these tasks, HPO is often targeted. The HPO is an important step towards the customized real-time application of the ML algorithms. Often, HPO involves significant time consumption towards effective training of the model and appropriate selection of correct parameters to achieve best accuracy score of the tested model. This work customized grid search-based determination of best parameters for each considered model. This involved the creation of several runs using different parameters with specified transformations and estimators. For the transformation step, the parametric combination yielding best results was chosen.

Along with the criticality of the HPO computation time, Fig. 7 (e-f) depicts the computational time demand for each considered model. The HPO for lesser complex computational models was lower than

10 min and the HPO for more complex high-end ML algorithms was about 20 min. For the GB, it was observed to consume the most computation time (19.18 min) was apparent due to associated training complexities. On an interval, the training time was noted to be closer to 5 s for the three tree-based models. Thus, DT being a computational model with lower complexity demanded only 9.35 minutes of computational time for the parametric optimization. For both DT and RF, the training time was lower than 15 s. Comparatively, the GB took 15.48 s for the training process. For the forecasting investigations, the computational times were similar to those being reported for the training scenarios. The assessment of computation time conveys no added advantage for the high-end ML algorithms in terms of the accurate outcomes with increased computation time. On the other hand, dependencies of the computational time can be observed on dataset size and the available computational power.

During forecasting studies, the socio-economic and demographic factors were first incorporated according to their rank and with a one-by-one strategy in the model. Thereby, population growth, LP, GDP, $GDDP_{\text{per capita}}$, HH have been observed to improve the model performance considerably. Table 4.5 (b) summarizes and compares the forecasting performance of the models for their optimal complexities in terms of RMSE and R^2 for all the three MA approaches. The optimal GB and DT models for EMA approach affirmed an RMSE and R^2 values of 2.12; 4.22 and 0.981; 0.967 respectively. The obtained trends were similar to those being discussed for the training cases.

Table 4.5 (b) Forecasting performance of optimally trained models in terms of RMSE and R^2 indices for the total MSWG rate case study.

Approach	*GB*		*RF*		*DT*	
	R^2	*RMSE*	R^2	*RMSE*	R^2	*RMSE*
EMA	0.981	2.12	0.972	3.63	0.967	4.22
SMA	0.977	3.83	0.966	5.72	0.956	7.81
WMA	0.951	5.91	0.959	8.21	0.942	11.45

4.6 Generic Characteristics of the Modelling Framework

The tree-based models were tested for their ability to predict in general the MSWG rates. Accordingly, higher R^2 values was observed for both RF and GB models during the training and testing phases. Thereby, it can be inferred that the applied ML models so developed were neither overfitted nor over-trained. The modelling effort encountered two common issues. These refer to a model that involved too

well fit training data and a model with too low fit testing data. To mitigate the overfitness issues, resampling procedure was incorporated through a 10-fold cross-validation in the modelling effort. However, the predictive analysis for the MSWG conveyed that the R^2 values was found to be lower than 0.85 during the testing phase despite affirming very high R^2 values (>0.90) during the training phase. The comparative study conveyed that the GB exhibited best performance among all considered tree-based algorithms (train and test score of 0.99 and 0.95, respectively). The RMSE, MAE and R^2 for the GB are 3.01, 2.86 and 0.997, respectively. For the prediction of MSWG rate, the ANN and DT models affirmed testing phase R^2 values of 0.72 and 0.54, respectively (Kannangara et al., 2018). However, the authors divided the dataset into 60:40, 70:30, 80:20, 85:15 and 90:10 training:testing ratios and each partition ratio generated a total of hundred arbitrary partitions.

Few authors also reported good prediction performances for the model with relevant R^2 (0.96); RMSE (1500.12) (Dissanayaka & Vasanthapriyan, 2019) and R^2 (0.97); RMSE (201.6) values (Nguyen et al., 2021). However, the DT models for both the studies exhibited lower accuracy and influential socio-economic variables were not been considered. With a smaller data size of 232, Johnson et al. (2017) reported the prediction model accuracy of GB model of R^2 (0.88) and RMSE (21.6). With good performance metrics of RMSE (0.174), the DT model suffered overfitting issues. Thus, an efficient testing phase modelling effort strategy has not been reported in the literature and hence the reported trends of lower *r* values were apparent. Contrary to this, as a critical finding the chapter reported good correlation values even during the testing phase. Table 4.6 conveys a comparison of the best data with the data reported in the prior art. In summary, the chapter demonstrated the efficient performance of chosen ML algorithms for MSWG prediction rate and with smaller datasets.

Table 4.6 A summary of best data obtained in this work and in the prior art for the prediction and forecasting of MSWG with ML algorithms.

Sl No.	Authors	ML Models	Dataset	Dataset partition ratio		Prediction performance		Inference
				Training	*Testing*	R^2	*RMSE*	
1	This work	DT RF GB	312, 376 Data from 1991-2011 (Census	70	30	0.979 0.988 0.997	325.82 83.21 3.01	• RF and GB models showed high R^2 values during both training and testing. It means that the

			of India, CPCB)					developed ML models were not over-fitted or over-trained. • The RMSE for the GB model shows significantly better results than RF and DT. • Socio-economic parameters such as Industries & HH age and genders can be included to check the performance of waste generation.
2	Nguyen et al. (2021)	kNN RF	189 (2015-2017)	80	20	0.96 0.97	202.3 201.6	Accuracy of the model was low due to the size and diversity of the data, including factors like lack of data at lower administrative levels
3	Rathod et al. (2020)	DT	Data collected from 200 regions of Akola city	70	30	0.504	0.174	The model suffered overfitting issues
4	Dissanayaka & Vasanthapriya n, 2019	LR ANN RF	Data from 2009-2017	80	20	0.697 0.992 0.960	2706.7 622.08 1500.1	Influential variables such as literacy, expenditures have not been studied to increase the efficiency in waste management.
5	Kannangara et al., 2018	ANN DT	1553 Data from 2001-2014	80	20	0.72 0.54	20 23	Accuracy 72% ML can produce accurate models for prediction if sufficient socio-economic explanatory variables are given.
6	Johnson et al., 2017	GB	232 (Historical data from 2005-2011)	80	20	0.88	21.6	Weather conditions as an external feature would further improve the robustness of the model.

4.7 Practical Implications

The outcome of the study has numerous practical implications for researchers, policymakers, and environmental protection groups. Foremost, the estimated MSWG in any region can be implied to investigate urban metabolism in order to develop and apply circular economy concepts. Urban metabolism is broadly used to describe how a city as an eco-system consumes material, food, energy, and water to support its growth and reproduction, and accordingly generates its products and by-products (e.g., GHG, pollutants, and waste) (Lu et al., 2021). The amount of solid waste generated is a critical metric for useful insights into the urban eco-metabolism and particularly in the industrial sector (Zhang et al., 2018). It is also a useful index to apprehend the efficiency of a circular economy system (MacArthur, 2013). Accordingly, it attempts to repurpose certain waste resources for more circular applications.

Secondly, the predicted amount MSWG can be employed for a chain of evidence-based policy-making. For instance, it can be used to plan a region's waste management capacity, such as landfill space and existing and expected 3R capacities. In due course of the activity implementation, planners frequently encounter issues such as no data or limited data. Policymakers can create adequate arrangements for the incentives to support recyclers and award penalties for polluters. This will be based on the severity of the problem and waste management capabilities. Subsidies, tax deductions, and low-cost land usage have all been widely reported in the past to assist recyclers for their enhanced profitability. The predicted result can also be utilized to coordinate inter-regional planning and co-ordination. An accurate estimation of MSWG will be one of the most significant pieces of information for such policy-vision and efforts.

Lastly, the generated MSW can be deployed in various engagement activities. The urgency of the problem can be better perceived by the general public through MSWG volume estimates, recycling ability and landfills. As a consequence, it could be more effective in persuading stakeholders to avoid the Not-In-My-Back-Yard (NIMBY) mentality (Bao et al., 2019) and to actively pursue a circular economy (Ruiz et al., 2020). Such an estimate will evolve a longitudinal data set, for a better affirming

of the SWM performance trends. This will assist people to achieve a virtual circle between built environment development and natural environment protection being performed on a regular basis.

4.8 Summary

This chapter provided useful insights into the ML-based modelling perspectives and forecasting approach for the MSWG rate based on available literature data and pertinent pedagogy. MSWG data at the regional level is very crucial to develop effective SWM planning. However, in many regions, especially emerging economies lack reliable data. This chapter adopted limited, publicly available data and adequate data analytics for the prediction of MSWG rate in Guwahati city. Seven factors such as population growth, GDDP, $GDDP_{per\ capita}$, HH, HH_{Size}, LP and WP were considered. The results of the data analysis depict that these factors can explain the majority of the variations in MSWG (due to coefficients of determination (R^2) of 0.75 or higher). To describe the MSWG rate variation, socio-economic and demographic parameters from census data were not sufficient. This is due to their nationwide availability status. Thus, the data can be used to develop waste management strategies such as urban metabolism monitoring of input data (e.g., materials, energy), output data (e.g., waste), and planning of waste management facilities (e.g., recycling plants or landfills). Such an approach for municipal waste prediction can be used as a guide for other regions by considering relevant development and environmental issues.

Three tree-based models namely DT, RF and GB as popular and powerful ML models were chosen and compared based on their strengths and drawbacks. The GB and RF delivered promising results in comparison with the DT. Among the two algorithms, the GB exhibited better performance with R^2 and RMSE values of 0.997 and 3.11, respectively. The DT conveyed higher errors of 3.00 % (for the year 2018) and lower R^2 values (0.967-0.942) in comparison with GB and RF models. The GDDP data appears to be most critical for the effective prediction of the MSWG rate. Moreover, the RMSE for EMA-GB forecast was the best (2.12) and outperformed the RF (3.63) and DT (4.22) models.

In summary, the comprehensive major findings are as follows. Firstly, the HPO fine-tuned tree-based ML models had higher test scores and confirmed superior model predictive accuracy values. Secondly,

hypertuned DT exhibited linear prediction and the RF model performance improved with the enhancement of the learning speed. Thirdly, the input data has its own limitations due to survey-based restrictions. Hence, for any city, further data in terms of ward-wise information will further complicate model performance and the HPO has been anticipated to meet such stringent needs. Thus, the suggested methodology is generic in nature and can be applied suitability for any city and for much complex input datasets through an appropriate modification of influential parameters. Also, the accuracy graph-based analysis confirmed that for a given set of small samples, each model's curve grew quickly during the initial stages, and the DT, RF, and GB had accuracies of 93%, 97%, and 98%, respectively. Fourthly and finally, the quest for generalized application of ML algorithms for MSWG rate prediction has been complimented to target GB for its prediction speed and accuracy to handle complex datasets with greater ease.

The research confirmed that the estimation of MSWG rate is very crucial for the subsequent system planning of MSW management from both short and long-term perspectives. Using statistical data in different cities, the presented generic approach can be applied for MSWG rate prediction in any city in the world. To do so, influencing parameters need to be investigated through careful introspection and analysis of the complex data. Best findings of this work can foster the Guwahati metropolitan region to initiate and develop an integrated decentralised community-based SWM approach. This can be targeted through enhanced recycling and composting practices (bio-waste) for the subsequent realization of circular bio-economy. According to this study, HH with higher incomes produce more waste. Therefore, the study conveyed the need for the implementation of the decentralized community-based SWM through the micro-management strategies in the municipality. For this, co-operative working culture between government and other local entities is the need of the hour. With such strategies, MSWM can gradually reduce associated environmental pollution and health risks in the city.

Chapter 5
Machine Learning & Time Series Models for Prediction & Forecasting of Organic Waste and Recyclables

Chapter 5

Enhancing the Efficacy of Solid Waste Generation Prediction through Supervised ML and Time Series Modelling: A Case Study in Kamrup, Guwahati

In this chapter, the findings associated to the accuracy enhancement in the prediction of OW recyclables generation rate with supervised ML techniques and time series modeling have been elucidated. Section 5.1 presents the relevant background of the work. Following this, section 5.2 of the chapter is dedicated for the examination of the data attributes of both independent and dependent parameters. In section 5.3, prediction performance was evaluated (with RMSE and R^2), for alternate ML models. Additionally, long-term forecasting of solid waste was conducted to estimate future waste generation trends and the findings have been discussed in section 5.4. Finally, section 5.5 presents a summary of the critical findings of the addressed objectives in the chapter[#].

Overview

The prediction and forecasting of OW and recyclables were targeted with ML algorithms and, socio-demographic factors that influence waste generation. The performance of MLP, evolutionary algorithm (GA), SVR, GB and kNN, were compared, and thereby conveyed that the MLP followed by GA resulted in better performance and with an R^2 values of 0.92 and 0.87 and RMSE values of 10.43 and 35.32 respectively. The hybrid ARIMA-MLP model was also developed for long-term forecasting. This resulted in a standard error of 5.67, and enhanced prediction accuracy from 92-96%. These important findings can aid to create useful tools and assist in the regional solid waste management and planning in the Guwahati city and North-east India.

5.1 Introduction

The chapter elaborates upon the findings associated to develop OW and recyclables generation prediction based on ten years of residential waste data (2001-2010) collected from the GMC. In the previous chapter, the influence of socio-economic and demographic factors on total MSWG were investigated with tree-based ML and statistical approaches. However, in this chapter, the predictive performance of five ML techniques, namely MLP, GA, SVR, GB, and *k*NN have been compared to develop a more robust model and facilitate comparative analysis and long-term performance evaluation for the prediction of organic and recyclable solid waste generation rate. Such a scientific approach allows for more comprehensive and accurate predictions of OW and recyclable waste generation rate

[#] Article communicated: Singh, T., & Uppaluri, R. V. S. (2023). Enhancing the efficacy of solid waste generation prediction through supervised ML and time series modelling: A case study in Kamrup, Guwahati.

and thereby enables better-informed waste management strategies and decision-making. Socioeconomic and demographic indicators such as population growth, HH information, age, and education level have been employed as predictor variables to create the models. The modelling was based on an integrated dataset consisting of socioeconomic and demographic information from the Indian census, annual HH solid waste generation (organic) data for the HH level (in the Kamrup metropolitan under GMC), and data on the diversion of plastics and paper. The primary goal of the considered models was to predict waste data in multiple municipalities and aid in waste management planning, and especially in areas that have scarce waste datasets. Also, it shall be noted that hybrid modelling (ML-time series) was employed for long-term forecasting. The adopted approach leverages the strengths of both methodologies, and thereby allowed a more robust and accurate prediction of future trends and patterns in waste generation. Through the incorporation of ML algorithms for the capture of complex relationships and time series modeling that capture temporal dependencies, the hybrid modeling approach was eventually targeted to enhance the accuracy and reliability of long-term forecasts. Thereby, better planning in SWM can be ascertained.

5.2 Data Attributes

After pre-processing, the datasets for the case have been summarised in Table 5.1. Data from CPCB, GMC, and the Census of India were used to compile MSW, paper, plastics, and kitchen OW data. Data with respect to socioeconomic factors were obtained from the GMC and the Indian census (2001-2011). The most comprehensive waste data for the reporting years were found in the MSW dataset. After data conditioning and integration processes, it had 300 datasets. Since 2001, the overall MSWG continuously increased, (illustrated in Fig. 5.1). The average MSW per family, however, shows a steady reduction with time. Also, much significant variation in waste generation trends was not apparent. After data conditioning and integration, the dataset for the plastic and paper diversion had 263 datasets.

Table 5.1 A summary of datasets after conditioning and integration for the prediction and forecasting of organic and recyclables waste generation rate.

Dataset	Independent variables	Dependent variable	Data source and number of datapoints
MSW-Census	Socio-economic variables (population, age, HH, HH	Residential MSW generation per HH (2001–2011)	CPCB, GMC and Census of India; 300

	size, Education level, Employee details (2001-2011)		
OW	Socio-economic variables (population, age, HH, HH size, Education level, Employee details	Disposal of domestic organic kitchen waste by HH (2001–2011)	Census of India, GMC; 263
Recyclables (paper+plastics)	Socio-economic variables (population growth, age, HH, HH size, EDU level, Employee details	HH trash diversion from plastics and paper per HH (2001–2011)	Census of India, GMC; 263

Fig. 5.2 (a-d) illustrates positive correlation among population, HH, HH size and level of education. Thereby, it can be inferred that the waste generation rate is significantly influenced by HH and population growth. The upward trends in MSWG per HH-population growth (0.84) and MSWG per HH-HH (0.88) are consistent with the conclusions drawn from the prior Pearson correlation result. Also, correlation coefficients for the pairs MSWG per HH-HH size (0.72) and MSWG per HH education level (0.52), have all affirmed positive but weak correlation. Hence, targeting the subduing of the associated multi-collinearity issues, the model pairs MSWG per HH-age and number of employees were omitted during the modelling effort.

5.3 Prediction Characteristics

5.3.1 Multilayer Perceptron

The MLP model consistently affirmed higher ratings during test (0.85) and train (0.82) phases. The MLP was trained with three hidden layers, each with a size of 5, *tanh* as the activation function, and a few auxiliary parameters. The RMSEs for OW and recyclable were 10.43 and 18.21, respectively. In comparison to other examined models, the R^2 values for organic and recyclable materials were 0.92 and 0.92, respectively. Due to the intricacy of the model, the computing time was estimated to be 8 minutes. No outliers affected the anticipated data. This is due to the grid search HPO, and the scatter plot (Fig. 5.3 (a) and (f)) conveys the same. As a result, the MLP can act as a potential and regressively trained model, and can eventually encourage the development of more accurate models.

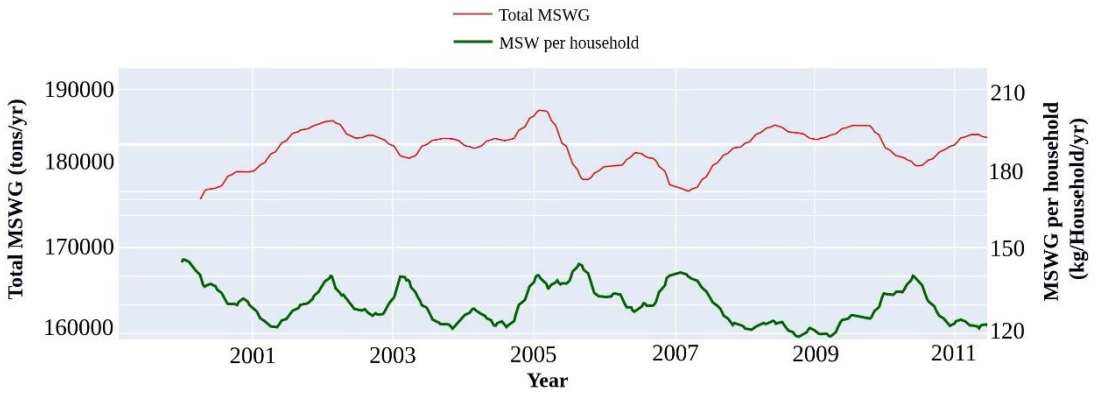


Fig. 5.1 Time dependent trends of total MSWG rate and average MSW per HH for Kamrup metropolitan area

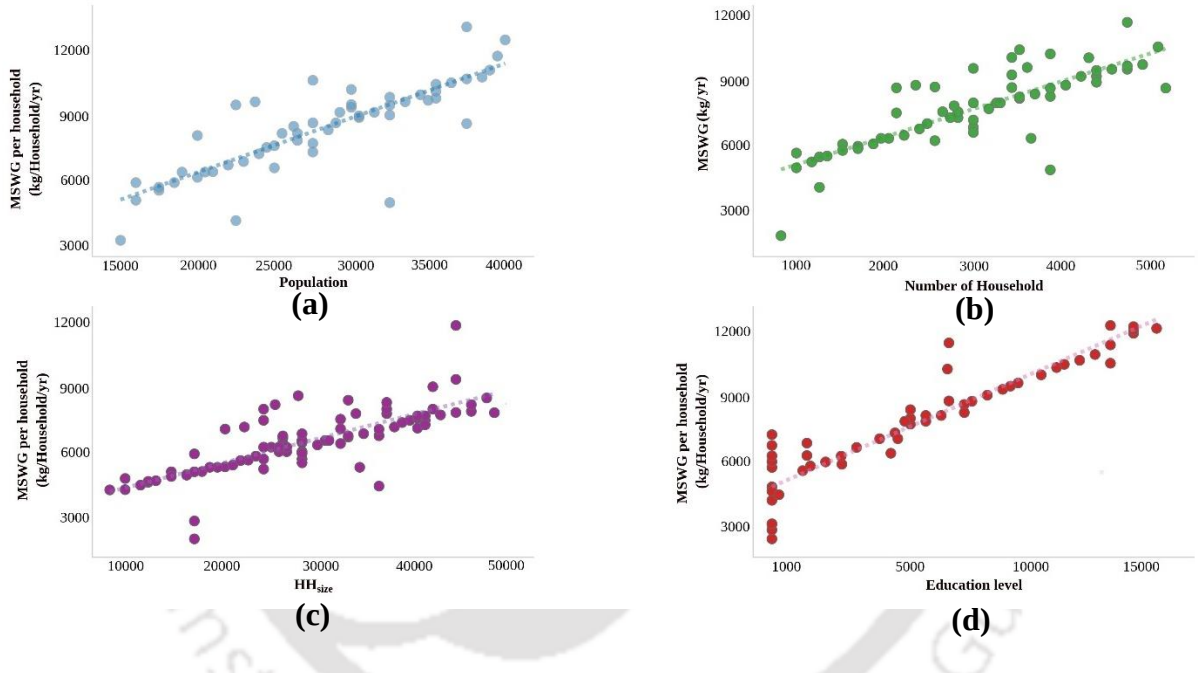


Fig. 5.2 (a-d) Correlation graphs depicting the interdependence of various socio-economic variables in the organics and recyclables case study

5.3.2 Genetic Algorithm

In the genetic algorithm, chromosomes are produced through the random selection of location points from a pre-processed dataset and with a random selection method. Each chromosome possessed 20 bits in size (each bit being a gene). The size of the primary population, which has 100 chromosomes, did not change over the course of the algorithm. The population growth-based fitness measure was used to choose the top chromosomes. As a result, the best chromosome is the one that has a repeating mean value throughout multiple generations. In conclusion, the mean and 100 best chromosomes from 100 generations were considered along with the fitness value. The chromosomes were altered to boost

fitness value after a certain point at which the fitness value became saturated. The scatter plot of 100 generations has been shown in Fig. 5.3 (b) and (g) for OW and recyclables respectively. The best chromosomes have been chosen based on the findings of the fitness measure. For OW and recyclables, the best chromosomes globally produced the best results (RMSE (35.32); R^2 (0.87) and RMSE (40.03); R^2 (0.83) respectively). The computation time was approximately 3 mins and was better than the MLP computational time.

5.3.3 Support Vector Regression

The SVR model was trained with the parameter values of 1 and 0.2. A train score of 0.79 and a test score of 0.75 was confirmed by SVR. As a result, the model resulted in higher RMSE (72.55) and R^2 (0.78) for recyclables and similarly values for; R^2 and RMSE were (0.80 and 64.17) for OW indicating lower RMSE than the recyclables (Fig. 5 (c) and (h)). The SVR was trained with an ε of 0.2 and C (penalty parameter) of 1. The computation time was evaluated to be about 2 mins. However, despite affirming acceptable R^2 value, the model was unable to provide better prediction in comparison with MLP and GA.

5.3.4 GB Model

For the GB, the test and train scores were 0.78 and 0.76, respectively. Trees with a maximum depth of 28 were used to train the GB. The overfitting problem has been successfully resolved by the RF and GB. The scatter plots for GB for recyclables and OW, respectively, have been shown in Fig. 5.3 (d) and (i) respectively. The RMSE (94.61) and R^2 (0.75) of GB, OW have been better in comparison to the RMSE (102.56) and R^2 (0.76) of GB recyclables. This is most likely as the GB after being taught for a single tree at a time, was able to fix the errors of the prior trees for each new tree. Moreover, the set sequence and order of the GB trees' execution could not be changed. Also, the computation time was approximately 35 sec which was the best in comparison to all other considered models.

5.3.5 *k*NN Algorithm

The *k*NN performed as an average model, and this was evident from the train and test scores of 0.81 and 0.79, respectively. The *k*NN model was trained for k=2. With enhanced *k*, the model's accuracy

declined. The RMSE and R^2 values for OW and recyclables cases were 142.19 and 0.72 and 191.89 and 0.71, respectively. The correlation between the model's predicted values and the actual values can be assured with a line fit and can be explained without being influenced by outliers or data noise. Fig. 5.3 (e) and (j) depict the same. However, certain restrictions exist in terms of the utility of conspicuous restriction and in terms of non-generalization to other datasets. The training dataset's k-nearest values were aggregated by the *k*NN model with increasing fitness. However, it was unable to explain the distribution of the underlying data structure or the relationship between dependent and independent variables. The model will therefore become incorrect with the inclusion of additional testing datasets from various in-situ measurements that considered an MSWG data with respect to population and HH range. Another significant problem is that *k*NN required practically no training. The testing phase, which requires expensive computing, was when most operations took place (6 minutes computation time). For this reason, despite a fair performance, the *k*NN did not affirm a good fit.

One by one and in order of importance, the socioeconomic characteristics were included to the model. The variation in RMSE between training and testing according to the incorporated socio-economic characteristics has been shown in Table 5.2. With more inclusion of socioeconomic indicators, the training error reduced. As a result, the models were much more sophisticated. The training and testing error, which assesses upon the model's performance for new data, increased marginally after the age parameter was included. Consequently, it can be stated that it is sufficient to incorporate population growth, number of HH and HH_{size} in MSWG predictive models. Thus, inclusion of other variables did not increase the model performance. With 10 neurons in the hidden layer, a neural network technique was used for this investigation. Other considered algorithms produced similar results.

Fig. 5.4 (a-b) depict the accuracy and loss graphs for the best model (the MLP). These charts show the training epoch on the x-axis and the accuracies and losses for both models on the y-axis (sample size). The training cycle count for the entire dataset has been represented as a percentage on the x-axis. The early curve profiles for some samples were steep and were in charge of 92% of the accuracy of OW and recyclables, respectively (according to the accuracy graph, which can be seen in more detail). Also, for

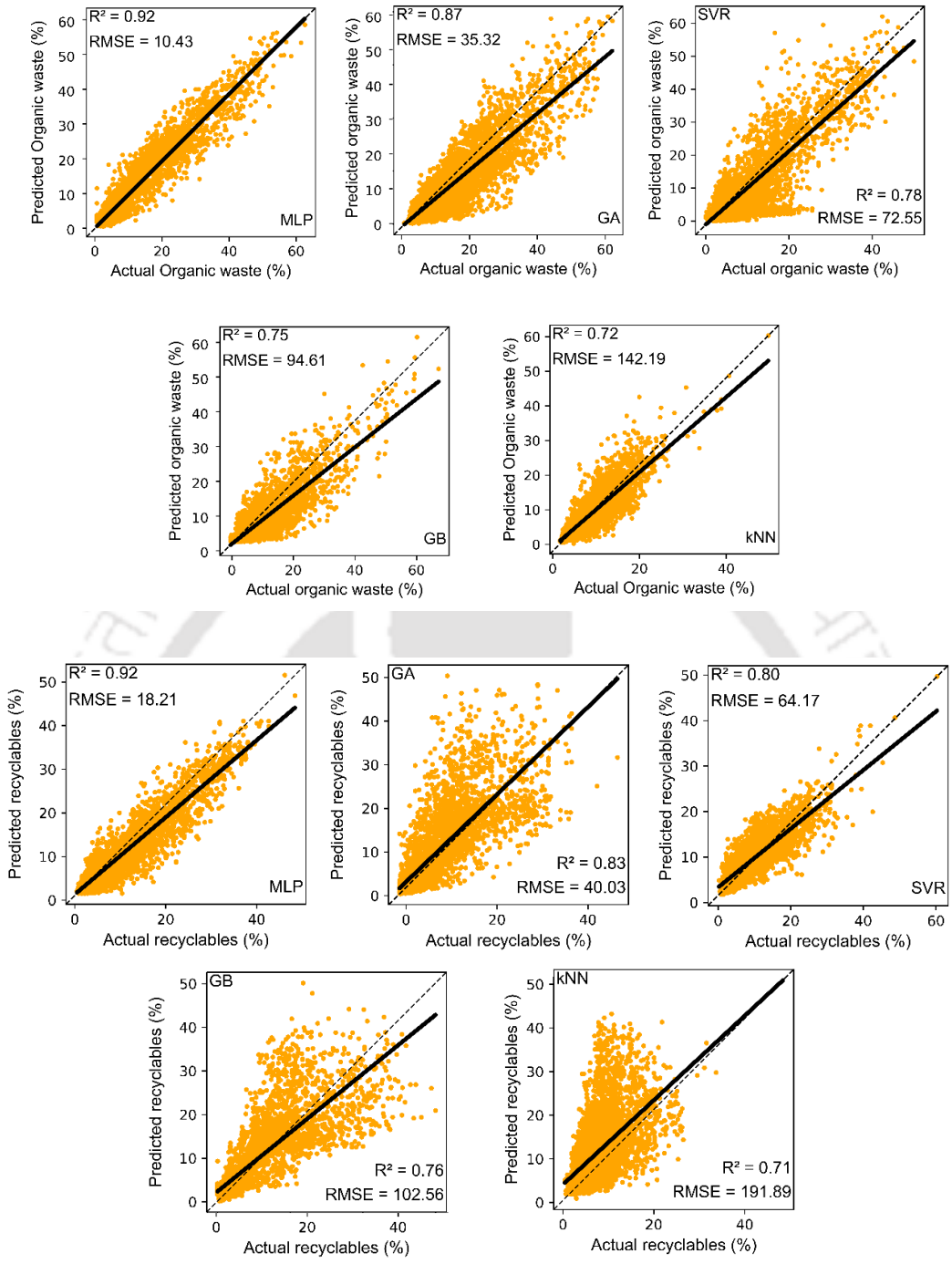


Fig. 5.3 Parity plots depicting the performance of: **(a, f)** MLP, **(b, g)** GA**, (c, h)** SVR**, (d, i)** GB, and **(e, j)** *k*NN for the prediction of OW and recyclable generation rate respectively

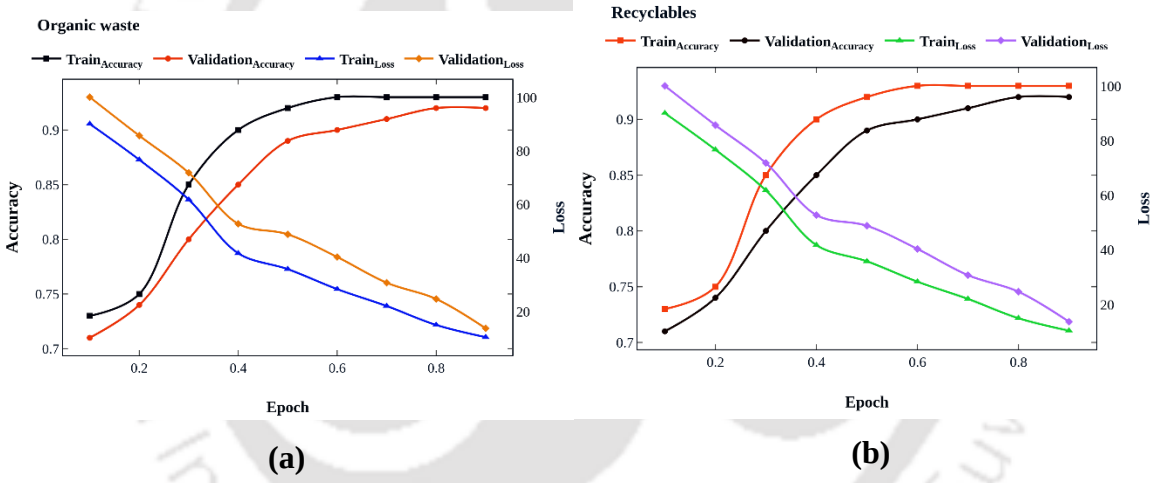
enhanced sample size, both accuracy curves exhibit an upward trend. In fact, a significant growth scenario can be shown up to four sample sizes. Thereafter, the growth rate gradually reduced until it reached a stationary profile. Loss measurements describe the poor performance of the model and in a manner similar to error measures. The numbers depict that the loss (error) is eventually decreasing. In

Table 5.2 Training and testing RMSE error alteration with increasing complexity of the MLP model.

Model	Training RMSE (% $HH^{-1}year^{-1}$)	Testing RMSE (% $HH^{-1}year^{-1}$)
Population growth	121.16	141.24
Population growth, Number of HH	115.83	136.79
Population growth, Number of HH, HHsize	92.42	103.44
Population growth, Number of HH, HHsize, EDU level	88.65	98.83
Population growth, Number of HH, HHsize, EDU level, Age	89.33	101.56

Table 5.3 Summarized training and testing data partition ratios.

Training and testing data partition ratio	Training RMSE (% $HH^{-1}year^{-1}$)	Testing RMSE (% $HH^{-1}year^{-1}$)
70:30	56.81	74.67
80:20	54.92	82.11
90:10	87.20	102.49



(a) **(b)**

Fig. 5.4 (a-b) Accuracy and Loss curves for the MLP-based prediction of OW and recyclables generation rate

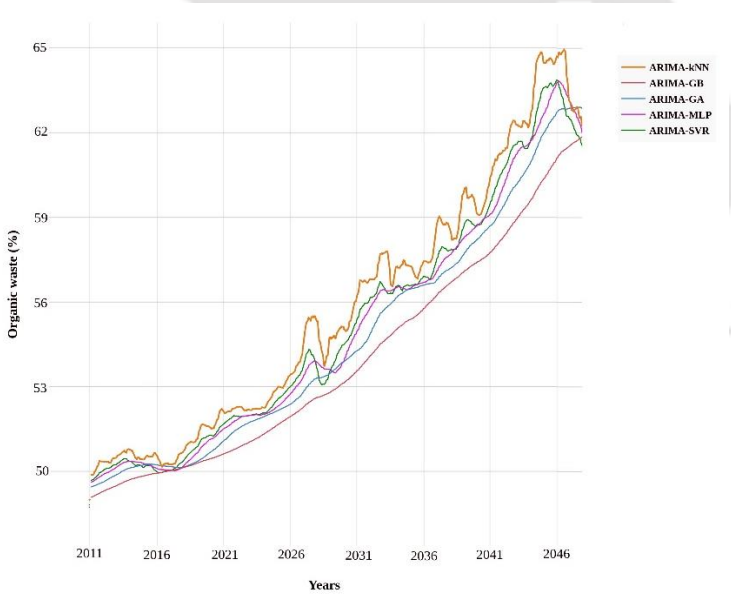
Table 5.4 A summary of standard error rate for alternate forecasting parameter configurations and ML models.

ARIMA		**Parameter(a,b,c) configuration**	
Models	***(5,1,0)***	***(3,1,0)***	***(1,1,0)***
MLP	68.49	14.97	8.32
GA	82.11	37.67	12.34
SVR	72.53	48.18	25.50
GB	91.51	49.72	29.61
***k*NN**	167.04	98.63	66.60

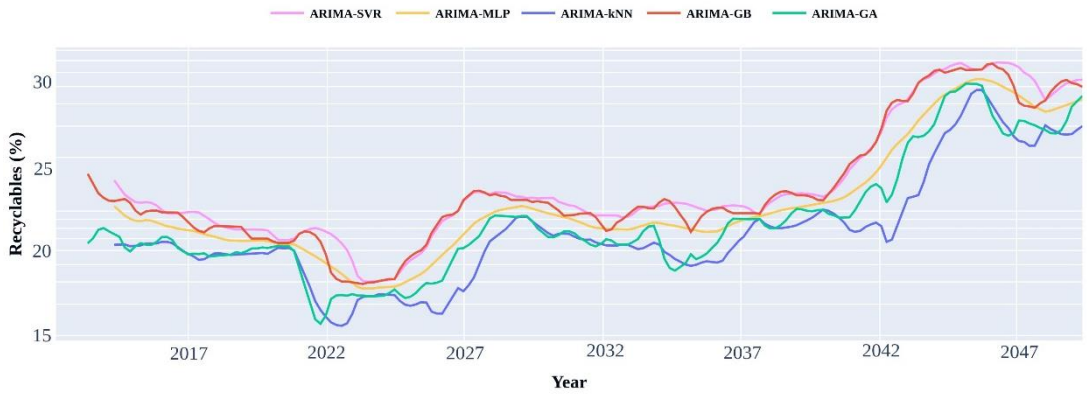
other words, with increasing epoch, model performance improved. The reduction of loss over time, despite the presence of a few peaks and valleys, proves that the MLP models learnt and performed well.

5.4 Long-term Forecasting of OW and Recyclables Generation Rate

The 38 years' waste generation data were subjected to forecasting with the ARIMA (1, 1, 0) model. As a result, each model's projected and expected results for the year range of 2012–2014 were correlated. Fig. 5.5 (a-b) shows that the ARIMA-MLP outperformed all other models in terms of results. These graphs show how future trends for the city of Guwahati can be predicted with expected previous data for the period (2012–2050). The optimal parameter configuration, determined by ARIMA (1, 1, 0) and with the lowest standard error, demonstrated that the ARIMA-MLP (8.32) and ARIMA-GA (12.34) had the best configurations for the forecasting of future values (Table 5.4). The validation findings of the forecasted OW (actual: 45–53%) and that of recyclables (actual: 25–32%) for the year range of 2012–14 has been shown in Table 5.5. The comparison findings of data with the prior art has been presented in Table 5.6. This highlights the efficient performance of the selected ML algorithms in predicting the OW and recyclables rate. In summary, this study has shown that these algorithms can effectively predict the MSWG rate with limited historical datasets.



(a)



(b)

Fig. 5.5 (a-b) Temporal series results depicting the forecasting trends of **(a)** OW **(b)** recyclables generation rate in the year range of 2012-2050

The relationship between OW percentage and GDP is complex. Elevated GDP often correlates with increased consumption, particularly of packaged and processed foods. These contribute to higher OW generation. Urbanization, commonly associated with economic growth, tends to intensify OW production due to population density and lifestyle changes. Improved waste management infrastructure accompanying rising GDP levels can enhance tracking and reporting of the OW generation patterns. Additionally, higher GDP may foster better consumer awareness and education. This promotes responsible OW disposal practices. Overall, the correlation is multifaceted and context dependent.

Guwahati's increasing GDP correlates with heightened consumption of packaged foods, elevating OW, notably food scraps and yard waste. During the survey of waste characterization in Guwahati city, it was observed that high-income groups generated more waste. This contributed to the overall increase in OW. Such an observation aligns with the concurrent rise in GDP. This conveys that the economic affluence corroborates with the waste generation patterns. The correlation underscores the influence of income levels, intertwined with GDP growth, on the composition of OW in the city.

Table 5.5 Alternate ML based model performance indices for the forecasting of OW and recyclables generation rate for the year 2012-14.

Forecast for OW				
Models	***Forecasts (%)***	***Standard Error***	***95% confidence interval***	
MLP	49.5-50.5	5.67	48-50	50-51
GA	48-49	7.83	47-49	49-51
SVR	49-50	6.11	48-50	50-52
GB	47-50	10.34	46-49	50-52
***k*NN**	52-54	18.97	51-53	53-55
Forecast for Recyclables				
Models	***Forecasts***	***Standard Error***	***95% confidence interval***	

MLP	23-24	10.93	22-25	24-26
GA	21-22	15.32	20-23	22-24
SVR	23	24.97	21-23	22-24
GB	23	24.65	20-23	22-23
***k*NN**	21.5-22	36.77	22-24	23-26

Table 5.6 A summary of best findings of this work and prior art data for the ML-based prediction and forecasting of SWG rate.

S. No.	Algorithms	Study area	Dataset Size	Accuracy Score	R^2	RMSE	Inference	References
1	MLP SVR GA GB *k*NN Hybrid Model (ARIMA-MLP)	Kamrup metropolitan, Guwahati Forecasting	826	92% 96%	0.92 0.78 0.87 0.75 0.72 -	10.43 72.55 35.32 94.61 142.19 -	• MLP resulted the best model fit for prediction • To increase the accuracy ARIMA model is deployed with ML approaches. • HPO technique was incorporated to mitigate the over fitting issues of each model.	This work
2	SVR, *k*NN, MLP	Vietnam	120	75.0%	*k*NN: 0.92	*k*NN: 0.14	The size and diversity of the data, as well as the lack of data at lower administrative levels, resulted a lower accuracy of the model.	Nguyen et al., 2021
3	GA, MLP	Delhi City, India	498	86.1%	MLP: 0.77	MLP: 0.43	The model suffered overfitting issues	Soni et al., 2019
4	GB	New York City	672	93.8%	0.906	22.059	Variables that may have a significant impact, such as literacy and age, have not been examined	Johnson et al., 2017
5	*k*NN, SVR, ANN	Logan city, Australia.	461	91.2%	SVR: 0.93	SVR: 300.70	The available socio-economic explanatory variables may not have been adequate for a	Abbasi and Hanandeh 2016

							comprehensive analysis	
6	MLP	Mashhad City, Iran	219	92.5%	0.87	0.14	The data size was too small to predict the accuracy	Abdoli et al., 2011

5.5 Summary

The conducted study demonstrated that the adopted ML based methodology has been efficient to detect the sensitive social-economic factors for organic and recyclables waste generation rate prediction. An examination of the consumption pattern is a resourceful technique to enhance the effectiveness of the command-and-control tool as alterations in consumption patterns are partially influenced by policy execution. For developing nations in which MSWM is not well established, such analysis will be highly potentiative to drive a transformation in human habitat towards ecological conservation (Mmereki et al., 2015). To perform policy comparison and appropriately select management policies, decision makers can choose cities or similar social, economic, and demographic factors. Dynamic policy interventions introduce complexity to data patterns. Thus, meticulous data analysis is crucial as an initial step for newly implemented models and predictions. Rigorous scientific scrutinization efforts ensures accurate model calibration and enhances the reliability of policy outcomes. Effective data analysis serves as the foundation for evidence-based decision-making in dynamic policy environments.

The ML methodologies are designed to capture intricate data patterns. Such approaches ensure them to be invaluable for the analysis of multifaceted scenarios such as Extended Producer Responsibility (EPR) and the ban on single-use plastics. To address these issues, ML techniques should be applied within a microscopic framework that allows a comprehensive understanding of complex interdependencies. For such analyses, large sets of data prior to and after the implementation of such single and multi-effect policy alterations shall be collected. Otherwise, machine learning approaches may not function well due to poor learning of inadequate data patterns.

Thus, despite affirming promising approaches for the current observations, inherent challenges exist for the incorporation of policy interventions, which are inherently qualitative for their non-integration into

numerical models. Historical data, while potentially lacking sensitivity to recent policy changes, remains policy-relevant and serves as a valuable foundation for the affirmation of future interventions.

The analysis of monthly, weekly, or daily MSW data provides a temporal granularity, a feature that is essential for the identification of crucial patterns and dynamics for effective policy interventions. The incorporation of such temporal resolutions enhances the precision of trend analyses in the conducted research. Thereby, it fosters an enhanced and nuanced understanding of MSWM and sustainability measures. Further research in this direction holds the potential to refine ML models and through the accommodation of the unique challenges posed by policy-driven variables. Such research strategies profusely contribute to enhanced practices in informed decision-making approaches.

The conclusion of the conducted study is that the alteration of socioeconomic parameters did result in significant alteration in the type of wastes. These waste categories vary depending upon the social and economic circumstances in each subdivision and in business regions. Henceforth, based on these modifications, decision-makers can set up technologies and processes for the recycling of paper, glass, and other goods. The consumption and waste generation patterns of the major socio-economic categories in the city of Guwahati can be considered along with comparable yet important socio-economic aspects in the future studies.

Also, the study evaluated alternate model performance with R^2, RMSE, and computational time as performance metrics. The results indicated a strong positive correlation between population growth, HH information, and EDU level. The MLP algorithm had train and test scores of 0.85 and 0.82, respectively, and performed well with a model accuracy of 95%. The overfitting issues of GB and MLP were precisely mitigated with a grid-based ten-fold cross-validation approach-based search and forecasting with the MA-based approach. This study contributed to the overcoming of overfitting issues for the modelling of nonlinear data and lowering of learning speed with the HPO. Overall, the MLP model for OW (R^2=0.92, RMSE=10.43) and that of recyclables (plastics and papers) (R^2=0.92, RMSE=18.21) outperformed all other index considered algorithms. The hybrid ARIMA-MLP models had 5.67 standard error (OW) and similar for recyclables (10.93) and were statistically distinct with 95% confidence interval.

The ML models deployed in this study can help to predict the amount of waste being generated in a given time period. This was based on factors such as population growth, economic development, and lifestyle changes in any municipality, county or region. Through the analysis of these data, city authorities can take proactive measures to manage waste more efficiently. These include development of better waste management infrastructure, implementation of recycling programs, and promotion of sustainable practices. The MLP and GA models have been proven to be effective in waste management prediction, as they can handle both linear and non-linear data relationships. On the contrary, the SVR model can be used to optimize the waste management process, by finding the best combination of parameters for the system. The GB model is suitable to handle large datasets and has been effective to predict waste generation in cities with high accuracy.

Chapter 6
Prediction and Forecasting of GHG and PMs Emissions from MSW Landfill & Incineration

Chapter 6

ML-Based Prediction and Forecasting of GHG Emissions and Particulate Matters from MSW Landfill and Incineration in Guwahati city

In a comparative framework, the chapter addresses the findings of the deployed nine ML models for the prediction of GHG emissions and particulate matters pollutants from MSW landfill and incineration in Guwahati city. These ML models were evaluated based on their performance in capturing the complex relationships between various input variables and emissions fluxes. After summarizing relevant information in section 6.1, section 6.2 of the chapter addressed the workflow involved with the optimization of model inputs (correlation study, stepwise regression and NCA approach). Section 6.3 addresses the predictive performances (R^2, RMSE, IoA, SSE, MAPE and MAE) of the models for the prediction of parameters such as CH_4, CO_2, N_2O and SO_2 fluxes, and Particulate Matters (PMs). In section 6.4, an in-depth analysis of the cross-validation results have been discussed to assess upon the generic nature of the models (with the partitioning of the data into training and validation subsets). Computation time, another aspect in the analysis has been addressed in section 6.5 to evaluate the time required for model training, feature selection, and prediction generation. Furthermore, section 6.6 delves into forecasting analysis based on the trained model's ability to predict future CH_4 and CO_2 fluxes, as well as PM concentrations. Accordingly, comparative studies with respective prior art have been elucidated. Such forecasts provide valuable insights into the potential trends and patterns of these variables. Finally, section 6.7 summarizes the key research findings of the fulfilled objective of the Ph.D. thesis[#].

Overview

Considering nine ML algorithms, their prediction accuracy was assessed with six indices for CH_4, CO_2, N_2O, SO_2 fluxes, and PMs emissions. The influential parameters (Tatm, RH, P, Pr) were determined through correlation, stepwise regression, and NCA analyses. Stepwise regression (cubic) achieved good R^2 values of 0.94 (CH_4), 0.96 (CO_2), 0.85 (PMs), 0.78 (N_2O), and 0.61 (SO_2) with similar RMSE trends. The MLP algorithm exhibited higher accuracy (R^2 = 0.97, RMSE = 0.32) and the ARIMA (1,1,0) × (0,1,1) was suitable to forecast with a good significance (R^2 = 0.89, RMSE = 5.67).

6.1 Introduction

The effectiveness of ML models to predict GHG emissions (CH_4, CO_2, N_2O, $PM_{2.5}$, PM_{10}, SO_2) from incineration processes and landfills has been investigated in this chapter. To this end, an objective comparison between nine ML models (tree-based ML approaches, GA, MLP, *k*NN, SVR, ridge and

[#] Article communicated: Singh, T., & Uppaluri, R. V. (2023). ML-Based Prediction and Forecasting of GHG Emissions and Particulate Matters from MSW Landfill and Incineration in Guwahati city

lasso) was conducted to ascertain upon the most efficient model with GHG emissions data (1970-2018) from EDGAR (Emissions Database for Global Atmospheric Research), waste data (GMC, CPCB), GDDP (Ministry of Statistical Programme Implementation), population data (Census, India) and meteorological data (T, RH, Pr, P, WS and WD) from MERRA-2. The investigations were devoted to Guwahati city, the gateway city for North-east India and a region of strategic importance. The chapter also assured all models' predictive accuracy and computational time and further analysed their strengths and weaknesses for GHG prediction. Also, a clear seasonal pattern was identified for each pollutant. Also, $PM_{2.5}$, PM_{10} was studied including their extreme episodes.

6.2 Model Inputs Optimization

An initial investigation was carried out to identify the criticality of the input parameters based on their significant statistical impact on the forecasting of GHG emissions during the training period and in the year range from 1970-2018. The examination involved techniques such as Pearson correlation analysis, stepwise regression, and NCA. The input predictor parameters were Tatm, Pr, and RH. These were analyzed to study their influence on WS, WD, and P and with the socio-economic and demographic factors such as population, GDDP and waste generation (WG). T_{atm} was categorically included to highlight upon the impact of environment temperature. The Pearson correlation coefficient (*r*) results depicted in Fig. 6.1 conveyed that the T_{atm} exhibited strong positive correlation with CH_4-flux (0.81) and lowest significance coefficient (SC) or p-values 2.1×10^{-15} followed by CO_2-flux (0.65) with SC value 3.5×10^{-13}. This affirmed that the T_{max} enhanced decomposition rates and emissions and henceforth increased both CH_4 and CO_2 fluxes (Tokida et al., 2010; Dijkstra et al., 2012). However, the corresponding correlation values for N_2O (0.22) and SO_2 (0.14) fluxes as well as PMs were low (0.14). WS and WD also affect the dispersion of both landfill gas emissions and odour and from the incinerator stack. RH levels do also affect the moisture content of the waste and henceforth the combustion efficiency. For the case, CH_4 (0.35), N_2O (0.85), CO_2 (0.57) and SO_2 (0.51) fluxes as well as PMs (0.73) exhibited strong positive correlations. Also, the studies inferred that a strong positive correlation existed between the CH_4-CO_2 flux (0.74). This conveyed that higher CH_4 emissions also enhanced CO_2 release. Similar positive correlations were

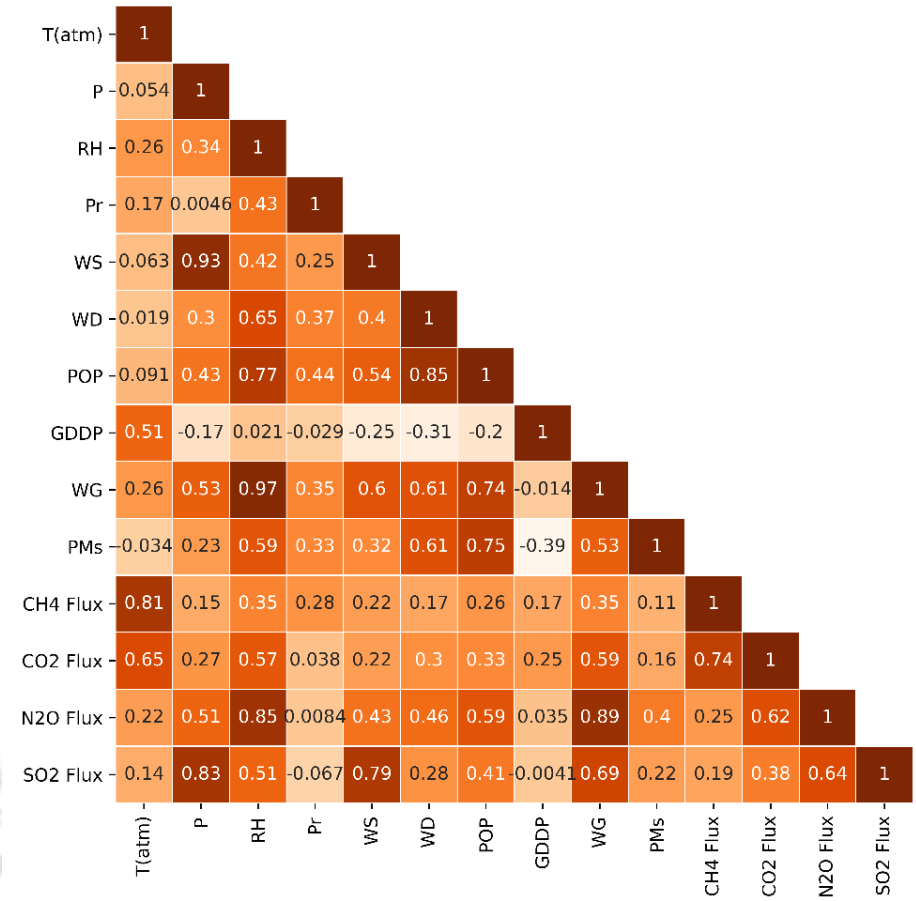


Fig. 6.1 Seaborn heat map for the predictions in the GHG emissions fluxes rate case study

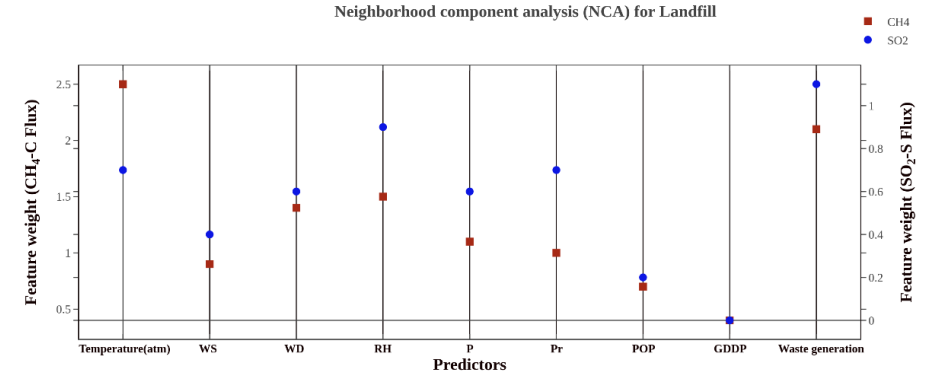
obtained for SO_2, N_2O fluxes and the PMs. The WG showed strong positive correlation with population growth (0.74), N_2O (0.89) and SO_2 (0.69) and also positive correlation with CH_4 (0.35), CO_2 (0.59) and PMs (0.53).

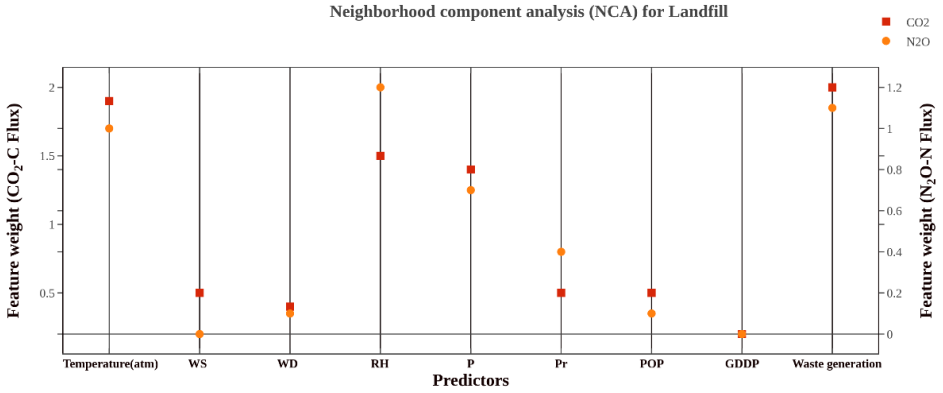
In Table 6.1, the outcome of the stepwise regression procedure being deployed as a selection technique has been presented for the evaluation of the effect of all predictors on GHG fluxes. The analysis confirmed that both T_{atm} and RH have been highly influential parameters for the emitted GHG fluxes. This corroborates with the inferences presented in the relevant prior art (Ngwabie et al., 2019; Zawadzka et al, 2010). The best R^2, and RMSE values were achieved by considering all predictors/inputs (R^2 = 0.94; RMSE = 23.21 mg m^{-2} day^{-1} and R^2 = 0.96; RMSE = 9.56 mg m^{-2} day^{-1} for CH_4 and CO_2 emissions fluxes from solid waste landfills, respectively). Similarly, for the solid waste incineration case, the R^2 and RMSE values for CH_4 and CO_2 fluxes were R^2 = 0.95; RMSE = 18.46 mg m^{-2} day^{-1} and R^2 = 0.98; RMSE = 4.31 mg m^{-2} day^{-1} respectively.

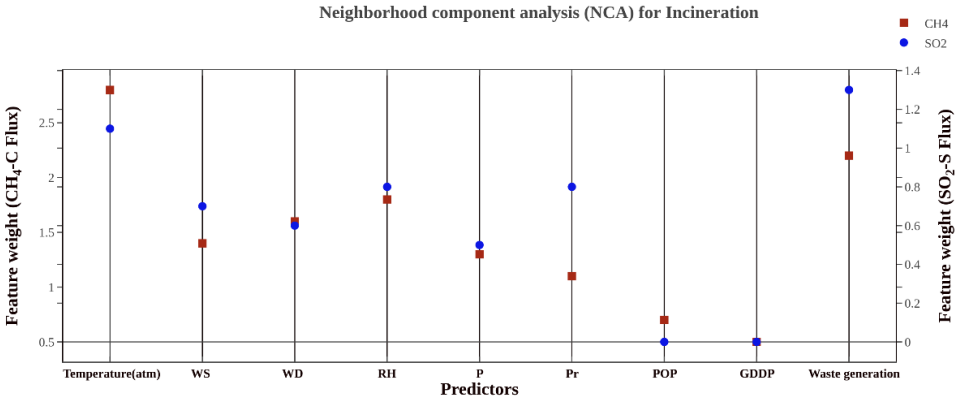
Table 6.1 A summary of R^2 and RMSE data associated to the variable selection phase of GHG and PMs fluxes rate case study.

Stepwise regression	**R^2**	**RMSE**		**R^2**	**RMSE**
Landfill			**Incineration**		
CH_4 flux (mg m^{-2} day^{-1})			**CH_4 flux (mg m^{-2} day^{-1})**		
Linear	0.65	35.6	*Linear*	0.68	32.7
Quadratic	0.82	29.8	*Quadratic*	0.78	25.4
Cubic	0.94	23.2	*Cubic*	0.95	18.4
CO_2 flux (mg m^{-2} day^{-1})			**CO_2 flux (mg m^{-2} day^{-1})**		
Linear	0.54	48.5	*Linear*	0.63	36.7
Quadratic	0.84	21.2	*Quadratic*	0.89	18.6
Cubic	0.96	9.5	*Cubic*	0.98	4.3
N_2O flux (mg m^{-2} day^{-1})			**N_2O flux (mg m^{-2} day^{-1})**		
Linear	0.50	65.8	*Linear*	0.61	71.3
Quadratic	0.65	43.1	*Quadratic*	0.68	65.6
Cubic	0.78	31.4	*Cubic*	0.75	60.3
SO_2 flux (mg m^{-2} day^{-1})			**SO_2 flux (mg m^{-2} day^{-1})**		
Linear	0.21	112.5	*Linear*	0.31	103.6
Quadratic	0.52	89.4	*Quadratic*	0.42	97.4
Cubic	0.61	49.2	*Cubic*	0.67	33.2
PMs (mg m^{-3})			**PMs (mg m^{-3})**		
Linear	0.67	67.9	*Linear*	0.70	66.4
Quadratic	0.76	56.4	*Quadratic*	0.82	35.7
Cubic	0.85	45.6	*Cubic*	0.91	22.4

The final results from the feature selection NCA method for both landfill and incineration cases have been illustrated in Fig. 6.2. The classified feature weights (from the highest to the lowest value) obtained from the NCA analysis of landfill case suggested that T_{atm} (2.5), RH (1.5), P (1.2), Pr (0.9), WS (0.8), WD (1.3) and WG (2.1) mg m^{-2} day^{-1}were the most influencing factors for the CH_4 emission flux data. Further, T_{atm} (1.8), RH (1.5), P (1.3), Pr (0.7) and WG (2.0) influenced the most for CO_2 flux emissions respectively. According to the NCA analysis, besides T_{atm} and RH, the P to a lesser degree was able to influence the N_2O emissions data (McBain et al., 2005; Zhang et al., 2008; Hensen et al., 2013). NCA analysis for the incineration case affirmed similar findings and inferred that T_{atm}, RH, and P were the most influential factors for the effective representation of CH_4 and CO_2 fluxes respectively. However, for the CH_4 case, the Pr has been additionally important to influence the CH_4 fluxes. These findings corroborated with those conveyed in the relevant prior art (McBain et al., 2005; Sutthasil et al., 2020; Jha et al., 2008). Also, the population growth as a predictor had a minimal effect on the GHG fluxes.







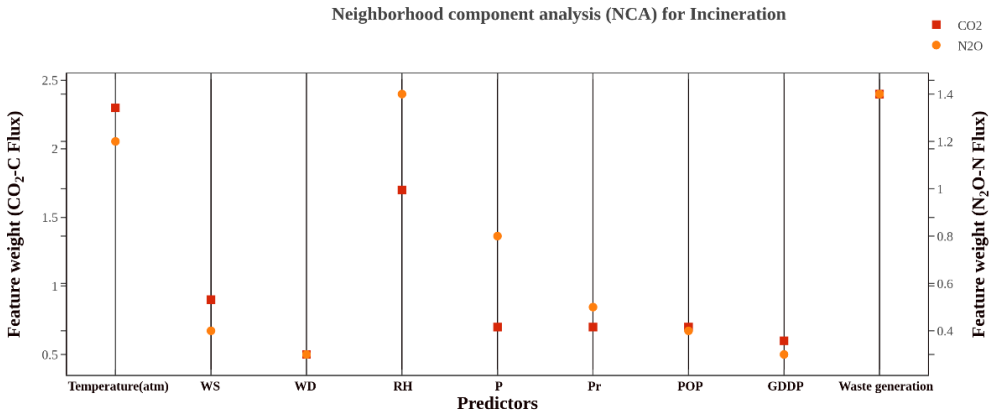


Fig. 6.2 Schematic representing NCA based feature selection analysis findings in the GHG and PMs fluxes rate case study

The NCA was also deployed to ascertain upon the significance of each of the aforementioned variables for the effective prediction of PM concentrations. The inclusion of temperature as an input variable aims to comprehensively capture these complex thermal effects and their implications for PM dynamics (Fig. 6.3 a-b). Temperature variations can be observed in these figures to critically influence the concentrations of PM. Seasonal variations in PMs concentrations follow distinct patterns. Winter months exhibit higher PM levels due to increased heating emissions ($PM_{2.5}$ = 96.22; PM_{10} =138.73 mg m^{-3}). PM concentrations gradually reduce during spring and reach their lowest levels in the months of May and June ($PM_{2.5}$ = 64.67; PM_{10} = 83.32 mg m^{-3}). Summer months generally have lower PM concentrations due to favourable weather conditions ($PM_{2.5}$ = 25.62; PM_{10} = 38.16 mg m^{-3}). Autumn ($PM_{2.5}$ = 58.33; PM_{10} = 76.45 mg m^{-3}) may witness a marginal increase in the PM levels. However, the winter season reinstated higher concentrations due to heating demands. RH as well influenced PM_{10} and $PM_{2.5}$ concentrations at solid waste landfill and incineration sites. Fig 6.3 (c) depicts that the lower RH levels could enhance particle agglomeration and henceforth increase the hygroscopic growth of particles. This resulted in the elevation of PM_{10} and $PM_{2.5}$ levels.

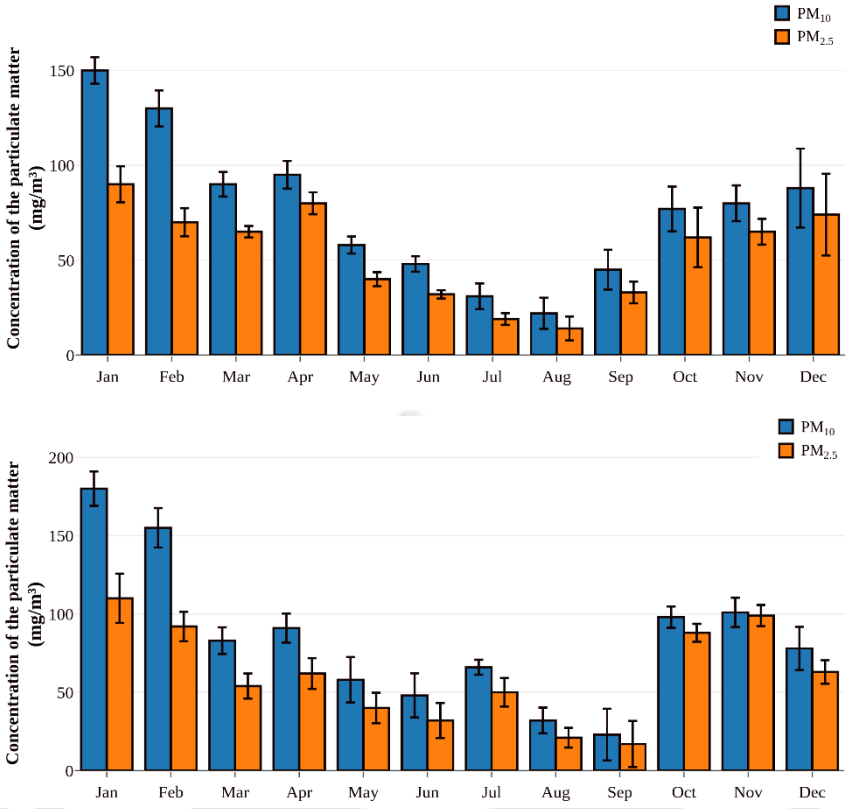
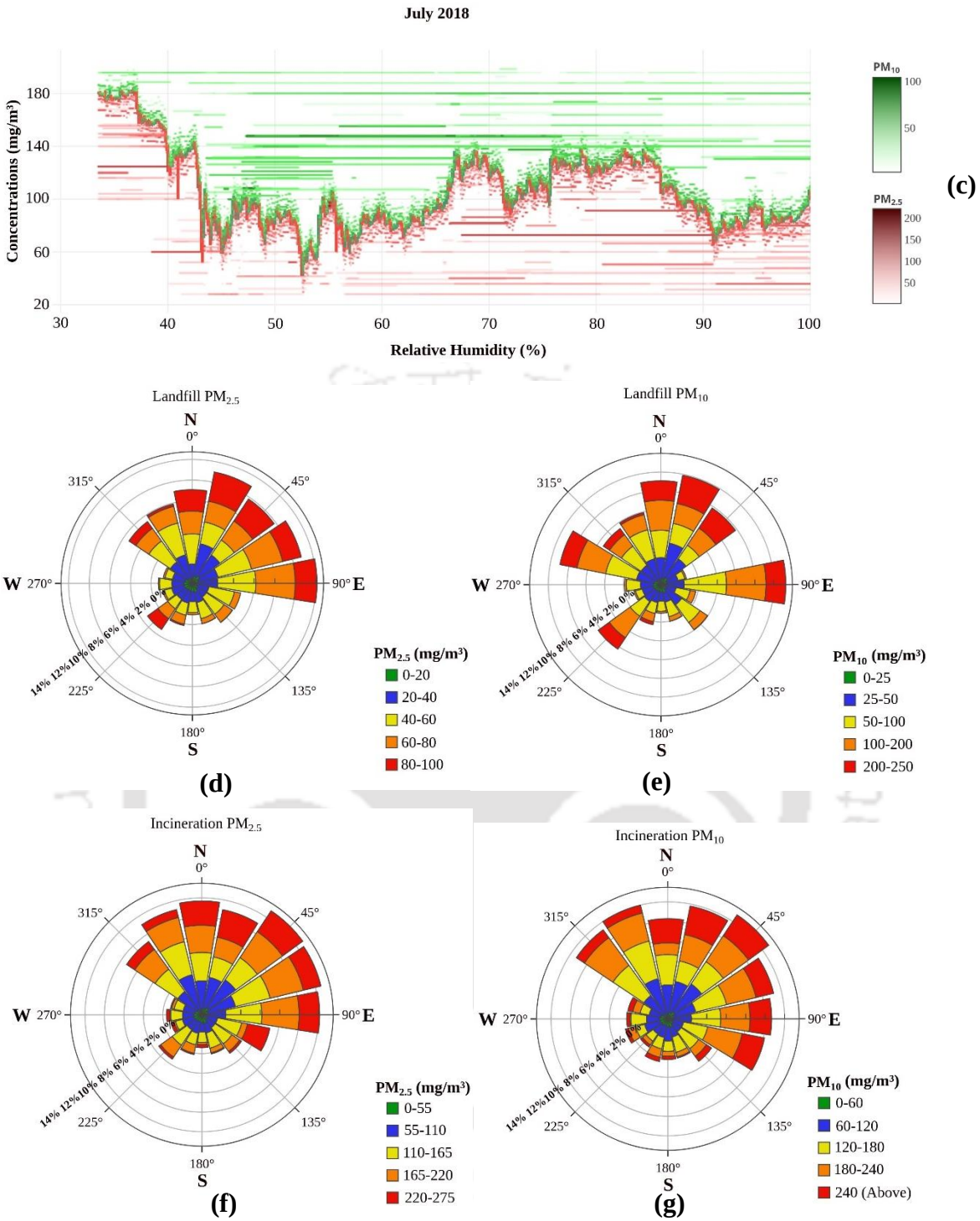


Fig. 6.3 Month wise $PM_{2.5}$ and PM_{10} fluxes trends for **(a)** landfill and **(b)** incineration sites of Guwahati city in 2018 year

Fig. 6.3 (d)-(e) and 6.3 (f)-(g) depict the respectively windrose patterns of landfill and incineration sites for the Guwahati city. For the case involving WS greater than 3 m s^{-1}, the concentration of $PM_{2.5}$ (average approximately 60-70 mg m^{-3}) increased for the landfill sites, and for the S-SW, W-NW, NW, N-NW, N, N-NE, NE, E-NE directions. However, for an increase in WS, the relevant directions were W-NW, NW, N-NW, N, N-NE, NE, E-NE, E for the incineration sites. Similar observations have been evident for PM_{10} concentrations (90-110 mg m^{-3}) and for both landfill and incineration sites respectively.

The overall influence of Pr on $PM_{2.5}$ and PM_{10} concentrations is critically influenced with factors such as rainfall intensity, duration, and the nature of particle sources in a specific location. Fig. 6.3 (h-i) depict the trends and relationships between the Pr and the PMs. The discerned valleys and peaks confirm upon the positive and negative relationships between the parameters. As shown in Fig. 6.3 (h), a weak



but positive association was identified between $PM_{2.5}$/Pr (R = 0.469) and PM_{10}/Pr (R = 0.412). This conveys that the enhanced Pr could only enhance the PM levels marginally. Pr can contribute to the release of PM from waste materials, especially during runoff events. However, the correlation strength is weak due to other factors such as waste management practices, emission controls, and meteorological conditions that as well critically influence PM concentrations. In view of these findings, input

parameters (i.e. T_{atm}, RH, P, Pr, WS, WD, population growth, GDDP) were selected for the ML predictive analysis

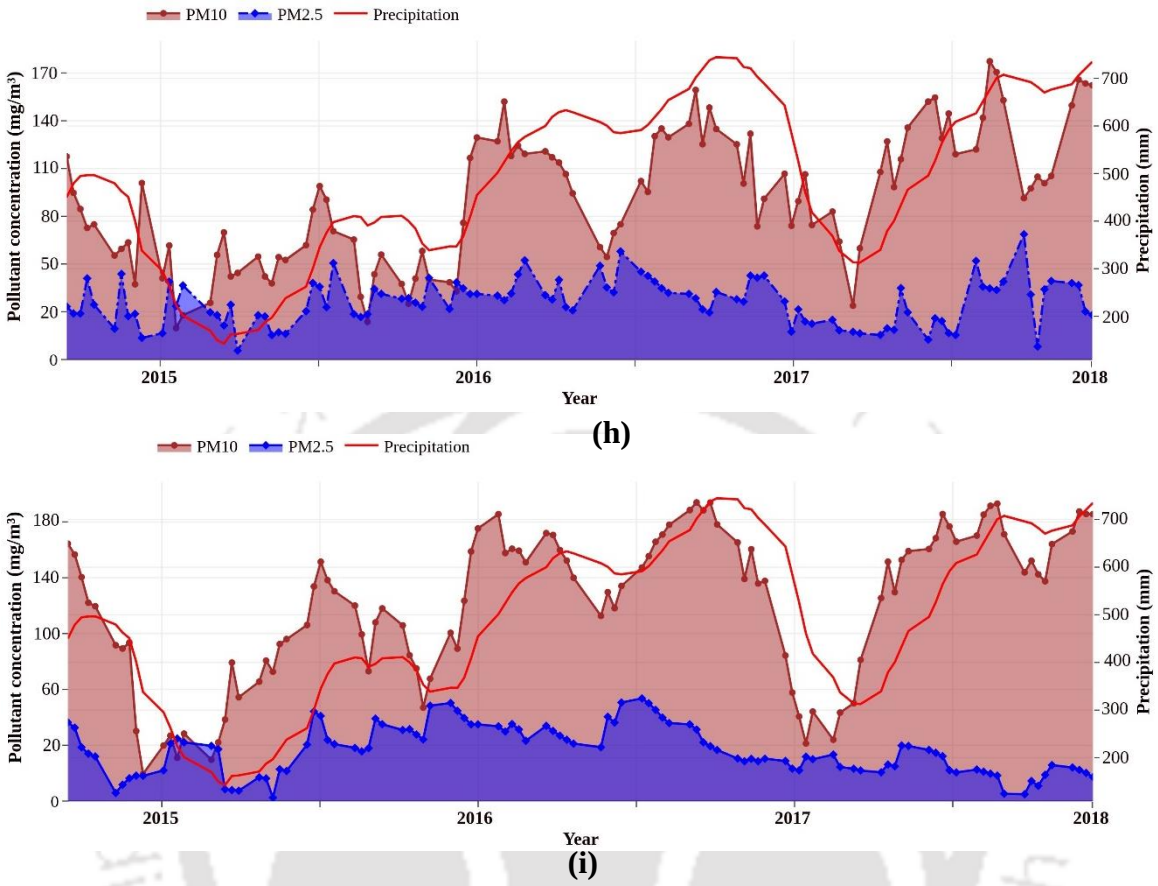


Fig. 6.3 (c) $PM_{2.5}$ and PM_{10} concentrations alteration with RH in 2018; WS windrose diagram for **(d)** $PM_{2.5}$ and **(e)** PM_{10} concentrations and landfill case WS windrose diagram for **(f)** $PM_{2.5}$ and **(g)** PM_{10} concentrations and incineration case **(h)** PMs concentration and precipitation trends of Guwahati city in 2014-2018 year range for the landfill case and **(i)** Precipitation and PM_{10} and $PM_{2.5}$ concentrations for landfill and incineration PMs concentration and precipitation trends of Guwahati city in 2014-2018 year range for the incineration case

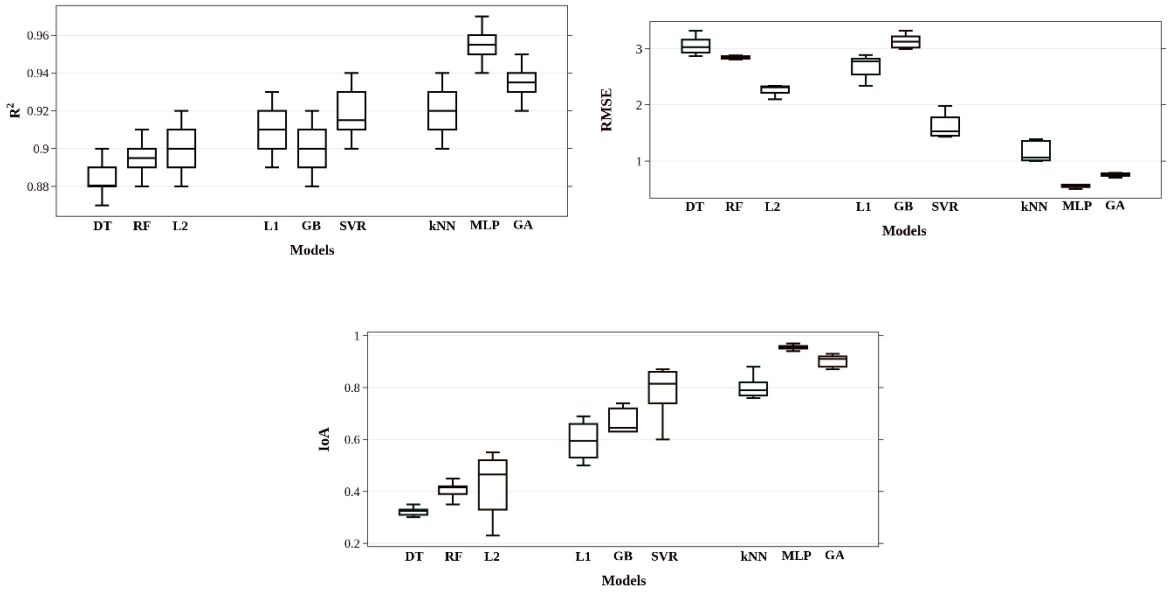
of GHG emissions. The predictor GDDP has been selected due to the significant growth of economic and technological indices in the targeted study area.

6.3 Predictive Performance of Alternate ML Models

6.3.1 CH_4 Fluxes

Fig. 6.4(a) depicts the performance metrics for CO_2 fluxes training and prediction sub-cases and for nine alternate ML models. The box and whisker plots indicate the median value and variability of the performance metrics. The results clearly demonstrate that the MLP performed the best among all ML models. For the case, R^2 values for both training and prediction phases were 0.99 and 0.97. Also, RMSE (0.32) and the IoA (0.94) for the MLP affirmed a stronger agreement between the

predicted and observed values. The range of variability for the MLP is in terms of its adaptability to diverse datasets. The model adjusted and adapted to the specific characteristics and complexities



(a)

Fig. 6.4 (a) Training performance indices of alternate ML models for CH_4 flux prediction

of different datasets. Accordingly, the model's performance increased potentially. In this study, five hidden layer neurons were deployed and with forward propagation approach. Accordingly, the MLP applied the learned weights and biases to the new input data, and produced a predicted output by utilizing activation functions and the inter-connections between neurons in the network. Thus, the MLP enabled to generalize its learned knowledge for effective and accurate predictions with respect to new and unseen datapoints. The MAE, MAPE and SSE performance indices also confirmed the best performance of the MLP. Also, as the evolutionary metaheuristic model, the GA resulted in a near-optimal solution due to its robustness and adaptability traits (R^2=0.95; RMSE=0.63; IoA=0.89).

Further, the *k*NN (R^2 = 0.92 and 0.86) and SVR (R^2 = 0.92 and 0.84) also achieved good predictive performances in both training and validation phases. The *k*NN's ability to adjust predictions assisted it to maintain accurately anomaly detection even in dynamic environments that witnessed timely variation of emissions pattern due to varied meteorological conditions. On the other hand, the SVR had lower

sensitivity towards outliers and for this reason, it was controlled to a limited context for its adaptability to the hyperparameters such as the regularization parameter (C) and the kernel function choice. However, these parameters gave flexibility to fine tune the model according to the specific problem at hand, and through a balanced trade-off between bias and variance for optimal performance. The tree-based models performed well during the training phase, but did not comparatively perform better during the prediction phase. Accordingly, the relevant performance indices were achieved for DT (R^2 = 0.96 and 0.90); RF (R^2 = 0.98 and 0.91) and GB (R^2 = 0.94 and 0.92). This is due to their lower effectiveness for learning in the data prediction tasks (Magazzino et al., 2020; Guo et al., 2021; Zhang et al., 2022). The RMSE and IoA values for the models were 3.34 and 0.35; 2.21 and 0.46; 3.15 and 0.74 respectively. It was further observed that the shrinkage regression-based models (i.e., Ridge, L2 and LASSO, L1) performed better in comparison to the tree based classical regression approach. This was especially during the prediction phase. Among these classical regression models, L1 exhibited better performance metrics (R^2 = 0.93; RMSE = 2.93 and IoA = 0.69) than L2 (R^2 = 0.92; RMSE = 2.28 and IoA = 0.57). Ridge regression has been effective to deal with datasets possessing higher multicollinearity. Hence, the algorithm performed well and retained all predictors in the model in comparison to LASSO.

6.3.2 CO_2 Fluxes

Unlike the case for CH_4 fluxes, a precise challenge with CO_2 predictions exists for the ML model's ability to capture the peak CO_2 fluxes. Fig. 6.4 (b-d) clearly depict that MLP has been the only ML model to address the mentioned challenge and was able to effectively capture the pattern. Accordingly, it outperformed all reported trends for other ML models (R = 0.98; RMSE = 0.63; IoA = 0.98) and with lesser variability in comparison to the previously reported CH_4 predictions. MLP utilizes activation functions such as sigmoid and *tanh*. Such activation functions introduce non-linear mappings between the input and output of each neuron. Thus, they enable the network to learn and represent complex non-linear relationships in the data. These reported findings corroborate with those presented in the relevant prior art (Huang and Dmitry, 2021; Mehrdad et al., 2021). The GA has been ranked as the second-best algorithm due to its good predictive performances in both training and prediction phase (R^2 = 0.97 and 0.95, respectively). The GA has been able to handle noisy and uncertain environments. The populations

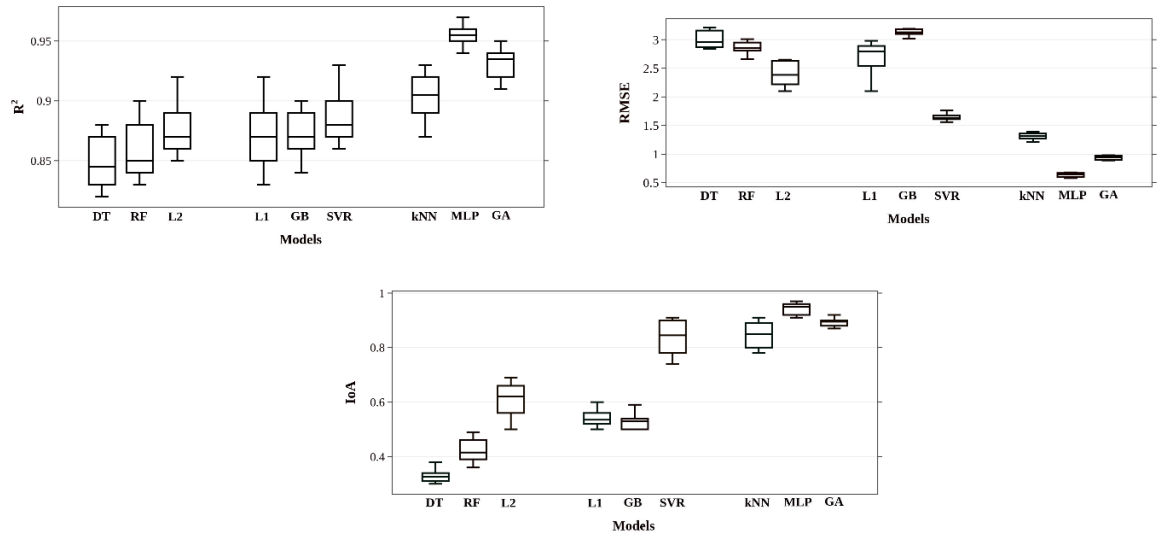
functioned and evolved the through iterations to handle the pertinent challenges and were able to adapt and converge towards good solutions even in presence of noise and outliers.

*k*NN and SVR performed marginally better in the prediction phase (R^2 = 0.94 and 0.93, respectively) than in the training phase (R^2 = 0.92 and 0.90, respectively). *k*NN has been robust to noisy training data and functioned effectively with large training data. Thus, good evaluation with RMSE = 1.32, IoA = 0.91 were obtained. The hyperparameters tuned for the *k*NN were the *k* neighbors (in this case the number of measurements for CO_2 fluxes) and the distance function was used to estimate a new data point. The value for *k* was tuned and was carefully chosen due to the fact that a lower value can lead to the overfitness of the data, and a larger *k* value can lead to the underfitness of the data (Adjuik and Davis, 2022). The GridSearchCV module was used to test a range of *k* (from 1 to 20) and the optimal *k* was obtained as 5. Accordingly, the training stage of the SVR algorithm was able to find the optimal hyperparameters for the best generalization of the model to predict CO_2 flux for the test dataset. Despite the fact that the SVR can function with many hyperparameters, the study considered four hyperparameters (C, ε, γ and the kernel function) for the tuning of the GridSearchCV module. Comparatively, the SVR model in this study was able to generalize fairly well to the data not being deployed for the training of the model (RMSE and IoA values of 1.75 and 0.90 respectively).

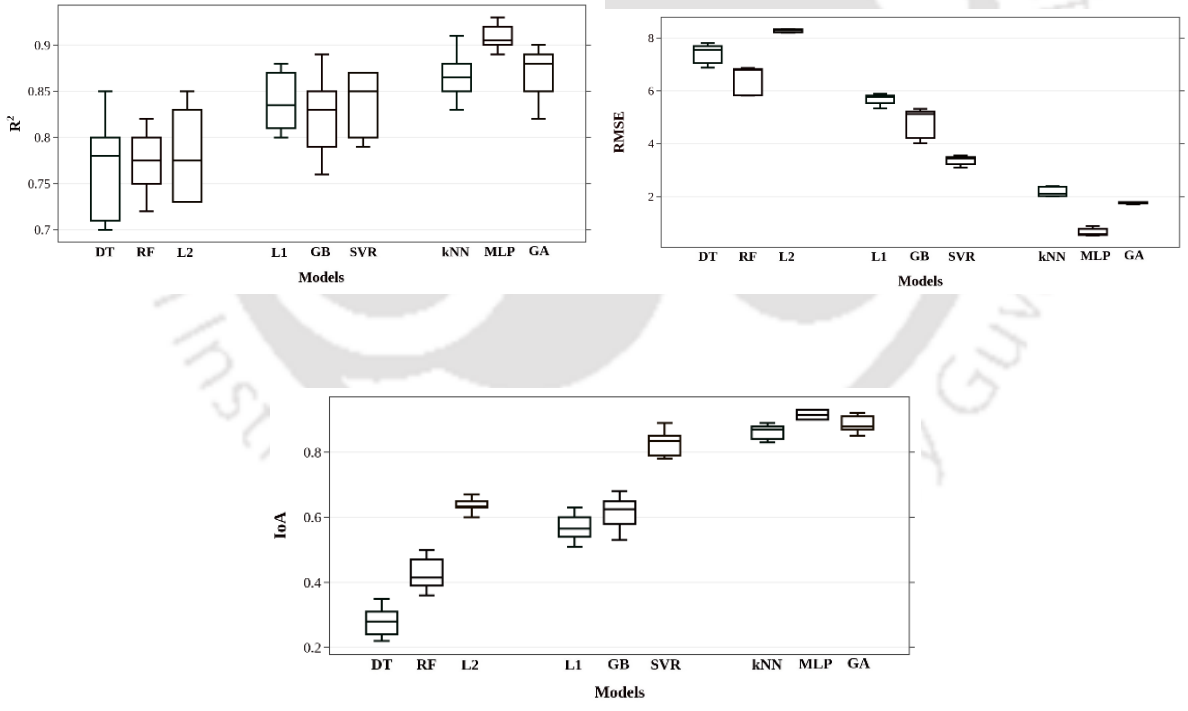
Tree-based ML models followed the shrinkage-based regression ML approach. Accordingly, L2 outperformed other shallow ML model L1 in the prediction stage (R^2 = 0.926). Characterized with a compact support, the L2 model had better ability to fit the data series for the pattern prediction of the CO_2 peak fluxes. In summary, except MLP, all ML models could not ensure a good prediction of the CO_2 fluxes.

For the case, other performance indices (MAE, MAPE and SSE) have been depicted for all nine algorithms in Fig. 6.4 (e-g). The performance indices of the MLP for both CH_4 and CO_2 fluxes have been better in comparison to those achieved for other models. A lower MAE, MAPE, and SSE confirms better model performance in terms of accuracy and precision. Accordingly, the MLP achieved an MAE, MAPE and SSE of 28.32, 49.21 and 0.12 mg m^{-2} day^{-1} respectively for CH_4 fluxes and 32.15, 52.19 and

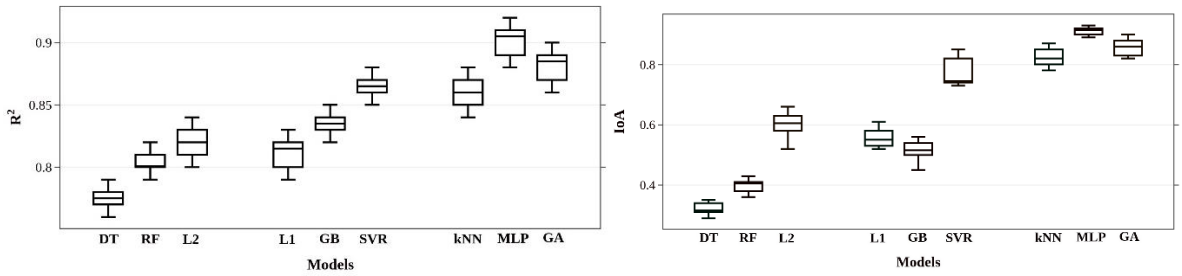
0.15 mg m^{-2} day^{-1} respectively for CO_2 fluxes. This confirmed a relatively small average deviation from the true values. Similar results were obtained for the considered ML models and for N_2O and SO_2 fluxes.

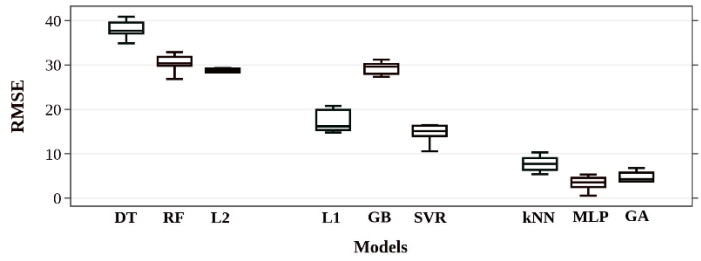


(b)

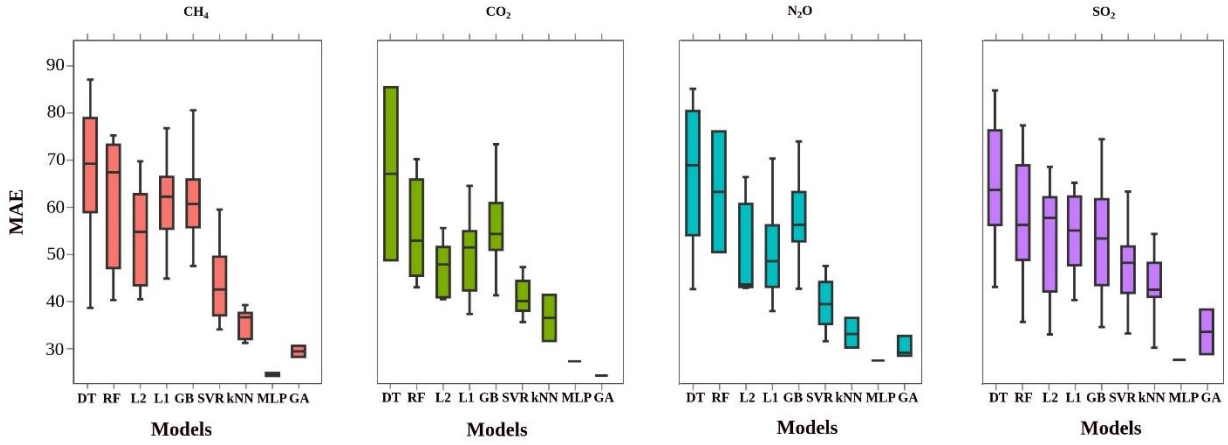


(c)

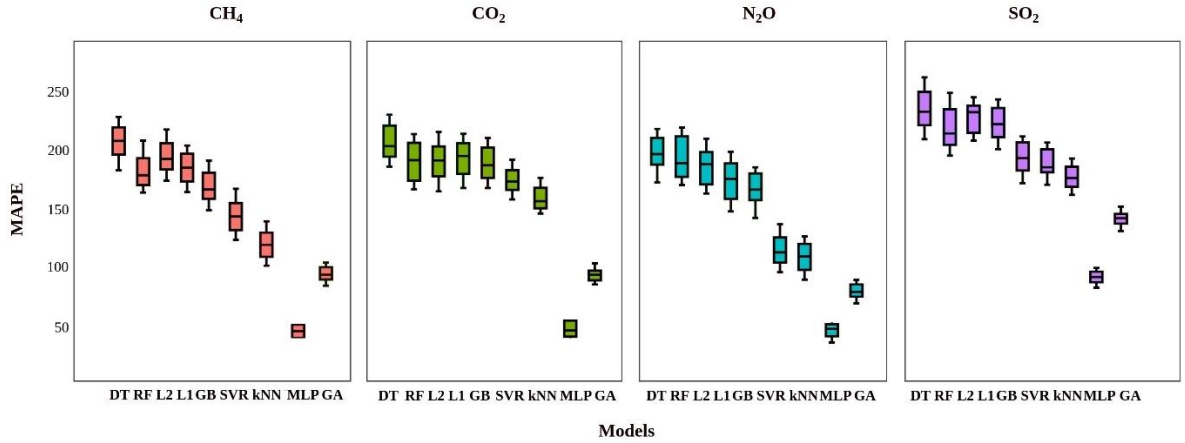




(d)



(e)



(f)

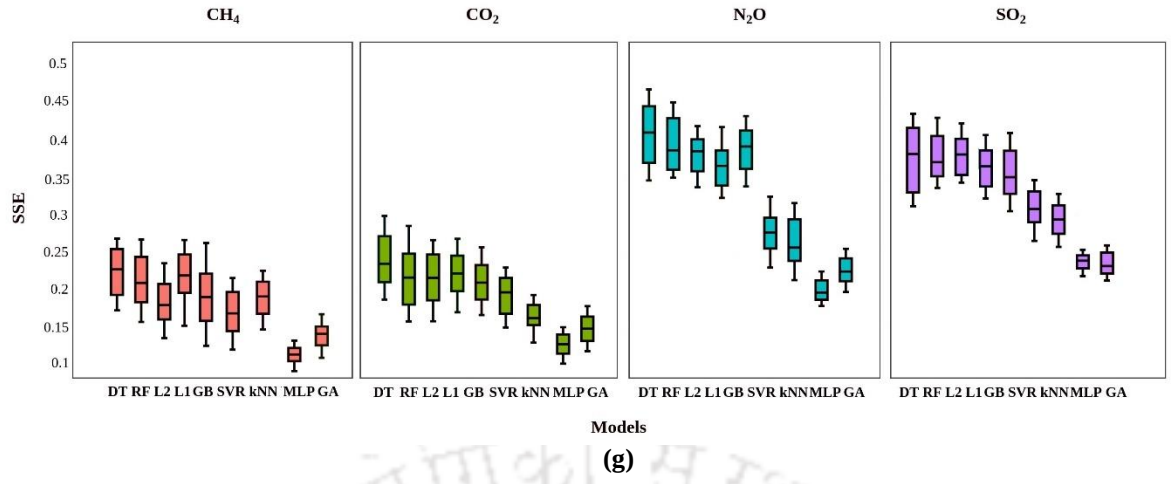


Fig. 6.4 Training performance indices of alternate ML models for **(b)** CO_2 **(c)** N_2O and **(d)** SO_2 cases respectively **(e)** MAE **(f)** MAPE and **(g)** SSE performance indices of alternate ML models for the prediction of CH_4, CO_2, N_2O and SO_2 emissions fluxes

For CH_4 and CO_2 flux prediction (in the year 2018), the MLP can be analyzed to the best (CH_4: 24.8 gm m^{-2} day^{-1}; CO_2: 0.40 gm m^{-2} day^{-1}) in comparison to other considered models (Fig. 6.5 (a-b)). Considering all mentioned predictors, it can effectively handle large feature spaces without necessitating upon explicit feature selection or dimensionality reduction. The GA could be considered as the second-best optimal algorithm with CH_4 and CO_2 emission values of 22.4 gm m^{-2} day^{-1} and 0.37 gm m^{-2} day^{-1} respectively. Also, *k*NN and SVR were effective to capture complex interactions and non-linear relationships among the predictors. Thus, they were suitable for CH_4 and CO_2 fluxes prediction. The DT model performed poorly despite affirming its ability to analyze relevant environmental variables and converge upon the decisions related to the flux levels. However, it possessed higher variance (landfill: 3.25 gm m^{-2} day^{-1} and incineration: 6.16 gm m^{-2} day^{-1} for CH_4 flux and landfill: 0.09 gm m^{-2} day^{-1} and incineration: 0.23 gm m^{-2} day^{-1} for CO_2 flux) and hence overfitness of the training data.

6.3.3 Particulate Matters (PMs)

For the evaluation of the models, RMSE, and R^2 were deployed as key metrics to predict PMs model prediction (Fig. 6.6 (a-b). For $PM_{2.5}$ and PM_{10}, the MLP achieved an RMSE of 8.26 and 5.92 mg m^{-3} respectively. This affirmed an acceptable standard deviation for the case. The consistent outcomes observed across multiple evaluation metrics conveyed a robust and reliable performance of the MLP for the prediction of PMs concentrations. Additionally, the R^2 value of $PM_{2.5}$ (0.99) and PM_{10} (0.98)

mg m^{-3} inferred that the MLP accounted for a variance above 95% in the PM concentrations. This conveyed

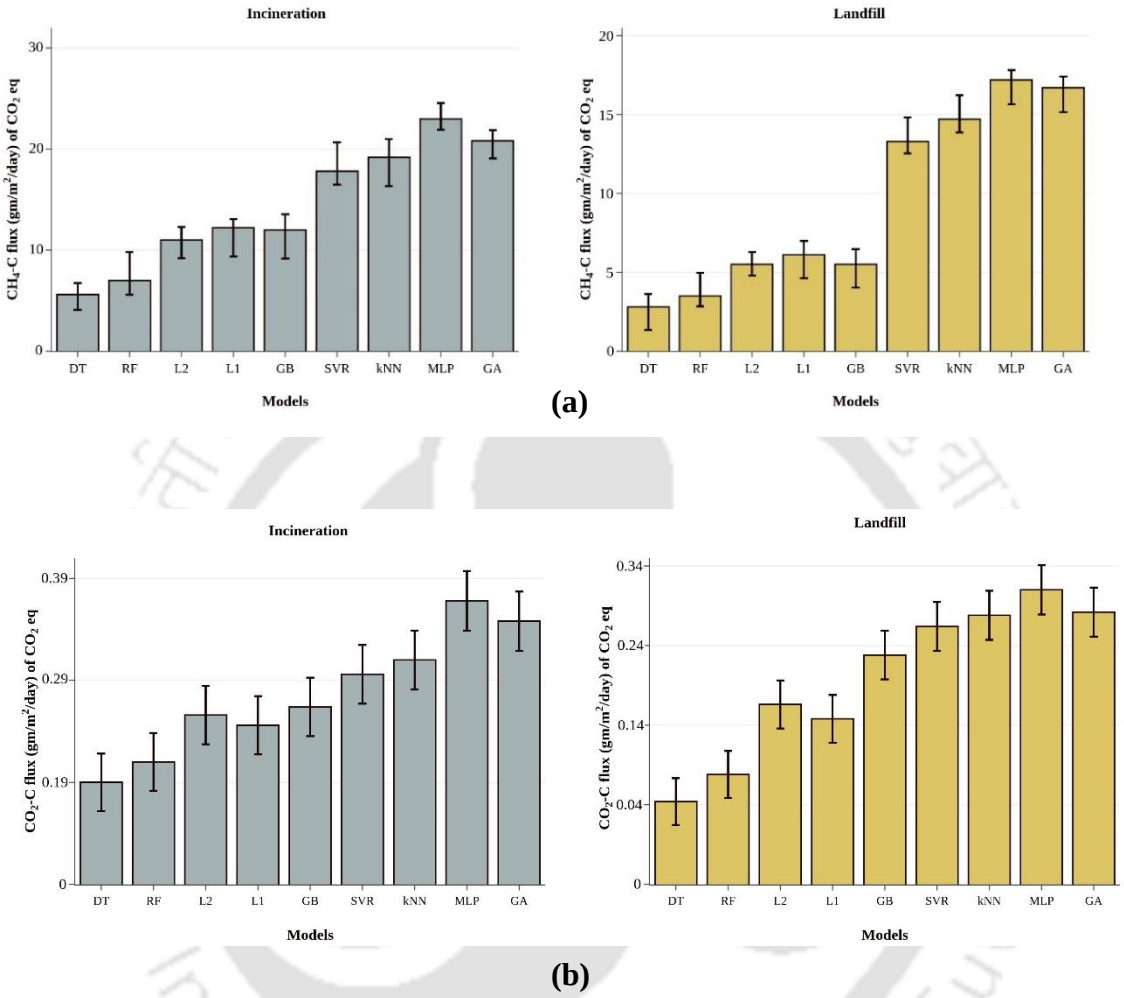
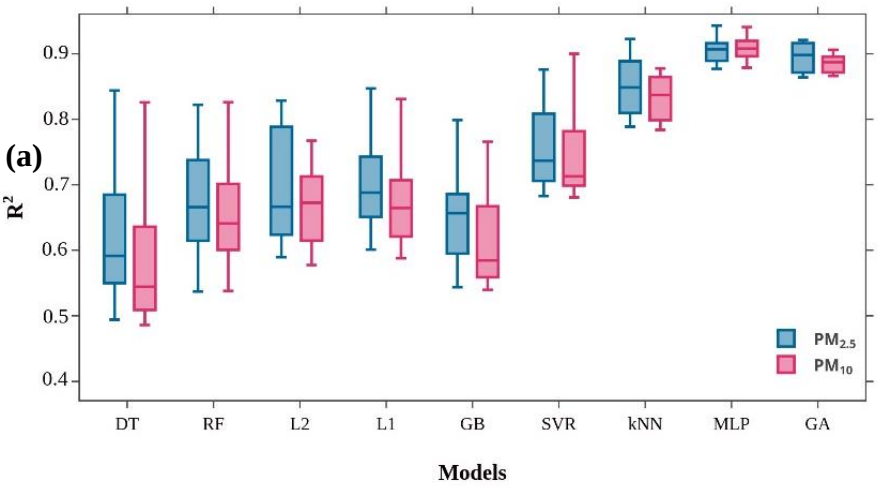


Fig. 6.5 Bar chart depicting the efficacy of alternate ML models for the prediction of **(a)** CH_4 and **(b)** CO_2 emissions fluxes



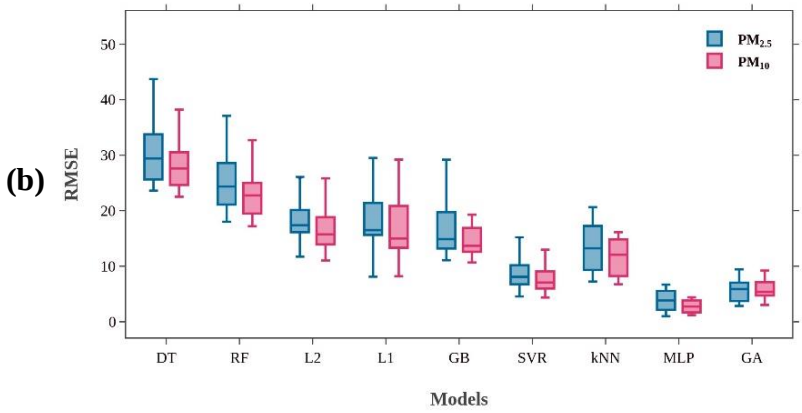


Fig. 6.6 (a) R^2 and **(b)** RMSE performance indices for $PM_{2.5}$ and PM_{10} emissions fluxes cases

a reasonably good fit. Similarly, the $PM_{2.5}$ and PM_{10} for GA models were 11.33 and 11.56 mg m^{-3} for a corresponding RMSE values of 0.93 and 0.91 mg m^{-3} respectively.

6.4 Cross-validation Analysis Findings

In the previous sub-sections that involved the ML model predictive competence studies being addressed for the GHG, the entire was split into 47 years training set (1970 to 2017) and one year prediction or test data set (2018). However, such an analysis can be opined to be partially reliable as the predictive performance obtained for only one test set can be different to that obtained for another test set. Thus, cross-validation studies were conducted in which the original dataset (training & testing) was partitioned. Such a partitioning approach ensured that both training and testing sets can be subjected to greater analysis through their appearance for the testing methods.

For the model evaluation, K-fold cross validation was deployed. The data collected from 1970-2018 were split into ten sub- sets (K = 10 folds). In these, each subset corresponds to one growing season. In the first iteration, the first fold corresponding to the year 1970 - 1974, was used to test ML models and the remaining folds (1975 - 1979, 1980 - 1984, 1985 – 1989, 1990 – 1994, 1995 – 1999, 2000 – 2004, 2005 – 2009, 2010 – 2014, and 2015 - 2018) were used for the training purpose. In the next iteration, the second fold (1975 - 1979) served for the testing stage and the remaining data (1970 - 1974, 1980 - 1984, 1985 – 1989, 1990 – 1994, 1995 – 1999, 2000 – 2004, 2005 – 2009, 2010 – 2014, and 2015 - 2018) was utilized for the training phase. This process was repeated until the final iteration could be

reached. Thus, the last iteration refers to the same case addressed in the prior predictive analysis. Here, the GHG data from the growing season of 1970 were predicted. Hence, the final iteration was omitted in the cross-validation studies. It should also be noted that the ratios between training/predicting data enhanced from the first iteration and upto the last iteration (the proportion of predicting data was 3%, 7%, 12%, 16%, 21%, 28%, 34%, 45% and 51% in the 1st, 2nd, 3rd, 4th, 5th, 6th, 7th, 8th and 9th iterations respectively). As a result, the 10-fold cross validation studies enabled a greater insight into the ML model's ability to predict missing GHG data.

The prediction stage findings in terms of the best R^2 and RMSE values for all ML models have been presented in Table 6.2. Once again, these results confirm upon the superiority of the MLP model (with the highest R^2 and the lowest RMSE) for CH_4, CO_2 fluxes and PMs prediction and for various combinations and complexities of the training/testing dataset. Thus, the best mean values were obtained with the MLP to predict CH_4, CO_2 fluxes and PMs (R^2 = 0.87, 0.85 and 0.79). These findings herewith affirm greater potential and utility of the MLP model to simulate and predict GHG emissions.

Table 6.2 A summary of the K-cross validation analysis findings for the ML based prediction of CO_2, CH_4, $PM_{2.5}$ and PM_{10} emissions fluxes.

K-fold cross validation (train:test)	Models								
	Best RMSE (R^2)								
	DT	**RF**	**L2**	**L1**	**GB**	**SVR**	***k*NN**	**MLP**	**GA**
CH_4 flux									
1st iteration (97-3%)	117.21 (0.46)	110.32 (0.41)	90.63 (0.55)	93.79 (0.58)	101.33 (0.55)	84.14 (0.57)	81.95 (0.58)	76.37 (0.59)	79.48 (0.57)
2nd iteration (93-7%)	112.11 (0.48)	104.16 (0.52)	81.05 (0.60)	73.16 (0.61)	95.88 (0.67)	74.38 (0.61)	70.76 (0.64)	68.68 (0.63)	71.12 (0.66)
3rd iteration (88-12%)	90.56 (0.50)	98.76 (0.57)	75.55 (0.63)	65.99 (0.64)	83.12 (0.70)	69.59 (0.66)	66.18 (0.69)	53.46 (0.74)	61.36 (0.69)
4th iteration (84-16%)	81.91 (0.65)	76.25 (0.64)	63.81 (0.72)	56.73 (0.69)	76.32 (0.74)	57.62 (0.73)	53.09 (0.72)	42.63 (0.78)	50.68 (0.76)
5th iteration (79-21%)	76.12 (0.72)	70.64 (0.74)	56.65 (0.75)	45.32 (0.73)	66.53 (0.77)	43.22 (0.79)	46.63 (0.76)	28.09 (0.81)	43.51 (0.82)
6th iteration (72-28%)	56.65 (0.74)	56.83 (0.79)	44.93 (0.80)	35.63 (0.78)	51.67 (0.81)	39.55 (0.78)	37.06 (0.82)	19.39 (0.85)	30.83 (0.83)
7th iteration (66-34%)	42.73 (0.77)	35.18 (0.83)	32.52 (0.85)	22.92 (0.82)	43.45 (0.82)	29.43 (0.85)	28.11 (0.85)	13.44 (0.89)	24.72 (0.87)
8th iteration (55-45%)	30.20 (0.79)	28.34 (0.85)	23.52 (0.87)	19.56 (0.84)	31.73 (0.84)	19.57 (0.86)	15.68 (0.89)	9.33 (0.90)	11.42 (0.90)
9th iteration (49-51%)	25.13 (0.82)	19.29 (0.86)	16.49 (0.88)	15.52 (0.89)	23.19 (0.86)	13.31 (0.90)	9.02 (0.91)	2.43 (0.95)	6.61 (0.93)
Mean	70.29 (0.65)	66.64 (0.69)	53.90 (0.73)	47.62 (0.73)	63.69 (0.75)	47.86 (0.71)	45.38 (0.76)	34.86 (0.79)	42.19 (0.78)
CO_2 flux									
1st iteration (97-3%)	120.55 (0.48)	115.54 (0.44)	102.19 (0.57)	100.32 (0.56)	103.80 (0.54)	93.17 (0.53)	90.21 (0.56)	73.95 (0.55)	85.52 (0.54)
2nd iteration (93-7%)	112.73 (0.44)	108.34 (0.53)	93.32 (0.60)	93.23 (0.63)	97.44 (0.65)	84.73 (0.65)	80.43 (0.62)	67.24 (0.65)	77.48 (0.66)

3rd iteration	99.82	102.97	85.82	85.84	86.20	78.55	76.32	57.13	65.43
(88-12%)	(0.52)	(0.56)	(0.68)	(0.62)	(0.72)	(0.64)	(0.66)	(0.73)	(0.68)
4th iteration	85.12	85.83	73.81	74.77	77.22	72.62	63.77	52.31	53.81
(84-16%)	(0.63)	(0.64)	(0.71)	(0.69)	(0.72)	(0.75)	(0.70)	(0.80)	(0.76)
5th iteration	79.95	76.28	66.92	67.94	68.74	60.22	56.18	48.22	45.25
(79-21%)	(0.68)	(0.74)	(0.72)	(0.73)	(0.75)	(0.77)	(0.74)	(0.83)	(0.82)
6th iteration	67.88	64.93	54.64	55.13	52.83	50.55	46.84	30.51	34.63
(72-28%)	(0.71)	(0.79)	(0.75)	(0.78)	(0.79)	(0.79)	(0.81)	(0.85)	(0.85)
7th iteration	53.03	53.15	46.15	42.73	45.76	41.09	38.12	24.42	29.93
(66-34%)	(0.74)	(0.83)	(0.80)	(0.82)	(0.80)	(0.80)	(0.83)	(0.87)	(0.87)
8th iteration	48.34	42.74	35.26	29.56	34.91	32.95	25.52	11.72	17.85
(55-45%)	(0.79)	(0.85)	(0.83)	(0.84)	(0.84)	(0.86)	(0.86)	(0.92)	(0.90)
9th iteration	36.22	32.80	29.11	25.16	23.19	21.45	12.66	3.12	11.82
(49-51%)	(0.83)	(0.84)	(0.87)	(0.88)	(0.85)	(0.88)	(0.90)	(0.93)	(0.91)
Mean	78.18	75.84	65.25	63.85	65.56	59.48	54.45	40.95	46.85
	(0.64)	(0.69)	(0.72)	(0.72)	(0.74)	(0.71)	(0.74)	(0.79)	(0.77)
$PM_{2.5}$									
1st iteration	36.21	30.62	42.55	37.92	41.77	43.81	36.72	33.61	32.13
(97-3%)	(0.61)	(0.55)	(0.54)	(0.56)	(0.57)	(0.55)	(0.57)	(0.75)	(0.56)
2nd iteration	31.53	35.82	45.21	40.88	32.54	40.38	31.83	30.71	35.88
(93-7%)	(0.64)	(0.62)	(0.62)	(0.63)	(0.59)	(0.62)	(0.66)	(0.72)	(0.64)
3rd iteration	36.32	31.83	37.51	35.61	34.61	35.59	27.22	25.03	30.91
(88-12%)	(0.70)	(0.65)	(0.64)	(0.65)	(0.67)	(0.65)	(0.70)	(0.77)	(0.71)
4th iteration	34.09	25.42	30.61	31.71	33.67	31.62	24.17	22.82	27.42
(84-16%)	(0.73)	(0.64)	(0.73)	(0.67)	(0.70)	(0.70)	(0.77)	(0.84)	(0.77)
5th iteration	31.12	19.73	28.91	26.93	30.53	27.22	21.63	20.80	30.73
(79-21%)	(0.79)	(0.72)	(0.77)	(0.69)	(0.75)	(0.74)	(0.72)	(0.85)	(0.83)
6th iteration	26.42	25.83	31.81	35.63	28.11	24.55	18.16	18.61	28.83
(72-28%)	(0.82)	(0.75)	(0.78)	(0.77)	(0.80)	(0.77)	(0.85)	(0.92)	(0.88)
7th iteration	18.61	17.83	23.82	32.81	21.85	21.24	14.53	15.34	23.15
(66-34%)	(0.83)	(0.81)	(0.88)	(0.84)	(0.83)	(0.82)	(0.82)	(0.90)	(0.85)
8th iteration	12.94	19.51	15.71	12.73	19.73	19.12	13.68	12.45	11.42
(55-45%)	(0.82)	(0.86)	(0.84)	(0.85)	(0.88)	(0.89)	(0.88)	(0.88)	(0.92)
9th iteration	16.52	11.62	11.43	14.81	15.19	13.26	12.05	8.11	10.61
(49-51%)	(0.84)	(0.84)	(0.87)	(0.88)	(0.86)	(0.87)	(0.89)	(0.92)	(0.90)
Mean	27.08	24.24	29.72	29.89	28.66	28.53	22.22	20.83	25.67
	(0.75)	(0.71)	(0.74)	(0.72)	(0.73)	(0.73)	(0.76)	(0.83)	(0.78)
PM_{10}									
1st iteration	40.10	42.62	33.21	34.52	41.91	43.81	36.72	30.81	31.89
(97-3%)	(0.66)	(0.53)	(0.56)	(0.51)	(0.54)	(0.54)	(0.57)	(0.70)	(0.54)
2nd iteration	36.91	35.82	35.55	28.61	33.11	31.81	31.83	30.10	32.81
(93-7%)	(0.63)	(0.54)	(0.61)	(0.61)	(0.62)	(0.55)	(0.67)	(0.67)	(0.64)
3rd iteration	33.81	32.91	32.34	30.11	37.51	32.91	25.75	28.02	28.11
(88-12%)	(0.72)	(0.59)	(0.62)	(0.64)	(0.63)	(0.63)	(0.72)	(0.77)	(0.68)
4th iteration	30.61	31.91	38.12	32.07	30.16	30.11	22.56	24.31	26.01
(84-16%)	(0.71)	(0.62)	(0.77)	(0.69)	(0.74)	(0.63)	(0.75)	(0.88)	(0.84)
5th iteration	35.33	29.08	32.10	25.44	25.07	22.71	20.66	21.59	23.80
(79-21%)	(0.69)	(0.68)	(0.79)	(0.73)	(0.77)	(0.71)	(0.81)	(0.83)	(0.79)
6th iteration	29.61	23.81	25.13	17.02	19.01	17.10	19.10	18.50	19.62
(72-28%)	(0.79)	(0.74)	(0.80)	(0.78)	(0.83)	(0.79)	(0.87)	(0.85)	(0.83)
7th iteration	21.80	16.92	19.62	21.71	15.22	14.03	16.02	14.07	15.41
(66-34%)	(0.82)	(0.79)	(0.81)	(0.82)	(0.86)	(0.85)	(0.89)	(0.85)	(0.82)
8th iteration	17.91	13.71	14.90	16.92	16.75	12.88	13.05	11.66	11.88
(55-45%)	(0.83)	(0.83)	(0.84)	(0.84)	(0.88)	(0.86)	(0.80)	(0.89)	(0.87)
9th iteration	14.71	14.15	12.55	13.71	14.77	11.80	10.72	6.36	9.82
(49-51%)	(0.85)	(0.86)	(0.87)	(0.87)	(0.85)	(0.88)	(0.90)	(0.94)	(0.91)
Mean	28.97	26.77	27.05	24.45	25.94	24.12	21.82	20.60	22.15
	(0.74)	(0.68)	(0.74)	(0.72)	(0.74)	(0.71)	(0.77)	(0.82)	(0.76)

6.5 Computational Time of Alternate ML Models

For the ML models and during training stage, the computation time has been illustrated in Fig. 6.7. In the figure, the computational time values in seconds have been rounded up to the nearest integer. The computations were performed on a core i7, CPU 2 GHz and 8 GB RAM computer. The computation time for classical regression approach enhanced drastically with the increasing number of neurons and epochs. Despite its higher training time, the MLP has been the most efficient in terms of the lowest computationally time-intensive ML approach for the prediction of GHG emissions. DT and RF were the most rapid models in terms of the computation time (varied in the order of milliseconds. This performance in time execution was high due to the DT's functionality in terms of recursive partitioning. This feature ensured higher interpretability and lower computational burden for the DT. However, for RF, its higher computation time could be due to parallelization as it introduced additional overhead and coordination between multiple processes or threads. This potentially led to longer computation times. Also, for both DT and RF, the prediction accuracy was poor in comparison to other ML models. On the other hand, the GA model has been proven to be a good choice for CH_4-CO_2 predictions, both in terms of both the prediction accuracy (with $R^2 = 0.95$) and computational time. The slowest ML models were L2 (increased memory requirements and ridge penalty term), and L1 (increased model complexity due to regularization and non-differentiability). For SVR, the CPU time was acceptable. This is due to the choice of the kernel (linear being the fastest) and the optimization problem resolution as a quadratic program. Nevertheless, the advantage of these ML approaches exists in the fact that the expensive computational cost is only required in the training stage.

6.6 Forecasting Competence of Alternate ML models

The forecasting studies were addressed in this work using ARIMA models for CO_2, CH_4, and PMs. Accordingly, useful insights could be obtained. For the GHG emission and for landfill and incineration cases, the ARIMA configuration of (1,1,0) × (0,1,1) were the mostly acceptable except for the negative spikes at lag 12. This conveyed the inclusion of SMA term in the model (Table 6.3). For GHG emission studies with ARIMA configuration of (5,1,0) × (0,1,1), (3,1,0) × (0,1,1) and (3,1,0) × (0,1,1), the RMSE and R^2 values of (9.32 and 0.74); (8.19 and 0.86) and (5.67 and 0.89) were achieved respectively for

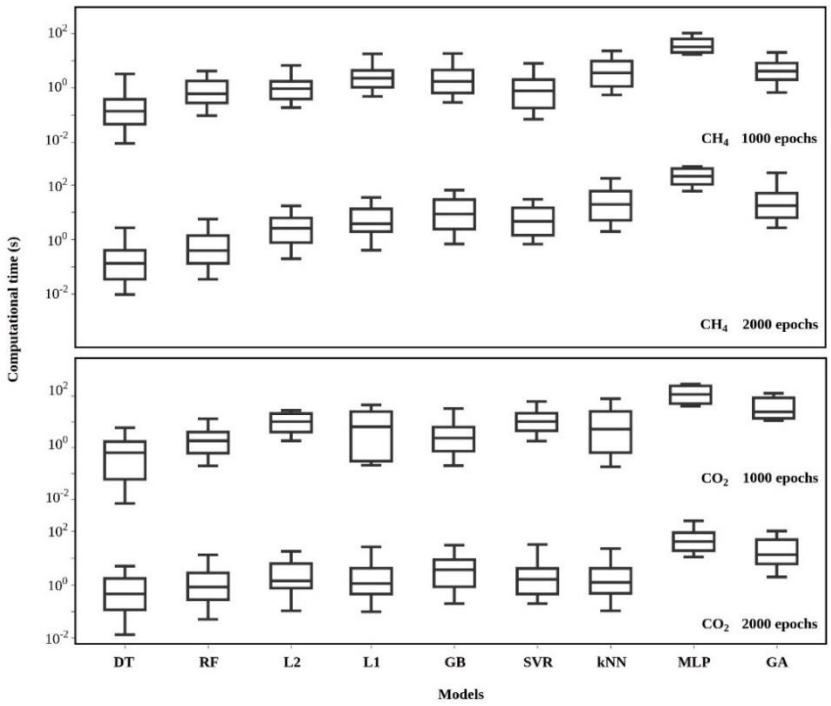


Fig. 6.7 Chart depicting the computational time of alternate ML models for the prediction of CH_4 and CO_2 emissions fluxes

landfill and (13.54 and 0.75); (10.31 and 0.79) and (8.17 and 0.86) respectively for the incineration case.

For CO_2 concentration, the ARIMA model successfully captured the underlying trends (2019-2027) and fluctuations in the historical data (Fig. 6.8 (a)). The model indicated a gradual increase in CO_2 levels with time (average ~2-2.5 mg m^{-2} day^{-1}). This is in apparent with the well-known phenomenon of anthropogenic CO_2 emissions contributing to the climate change. The forecasted values projected a continued rise in CO_2 concentration, and affirmed the persistence of this environmental challenge in the future. Regarding CH_4, the ARIMA model provided valuable insights into its temporal dynamics (Fig. 6.8 (b)). The ARIMA model effectively captured the periodic patterns in the historical CH_4 data (3-3.2 mg m^{-2} day^{-1}). The forecasted values confirmed fluctuations in CH_4 concentration. These are likely to be influenced by factors such as agricultural practices, waste management, and natural emissions. For PMs, the model yielded informative predictions about their levels in the atmosphere (~118-280 mg m^{-3}). The model effectively accounted for the temporal patterns and variations in the historical PM data

Table 6.3 Performance characteristics of alternate ARIMA configurations for the GHG emissions case.

Model	R^2	RMSE	MAE	MAPE	SSE
Landfill					
ARIMA (5,1,0) × (0,1,1)$_{12}$	0.74	9.32	45.66	28.77	3.21
ARIMA (3,1,0) × (0,1,1) $_{12}$	0.86	8.19	40.71	21.62	2.12
ARIMA (1,1,0) × (0,1,1) $_{12}$	0.89	5.67	34.91	15.43	0.65
Incineration					
ARIMA (5,1,0) × (0,1,1) $_{12}$	0.75	13.54	47.11	27.63	22.17
ARIMA (3,1,0) × (0,1,1) $_{12}$	0.79	10.31	43.62	24.19	14.92
ARIMA (1,1,0) × (0,1,1) $_{12}$	0.86	8.17	37.15	19.21	10.14

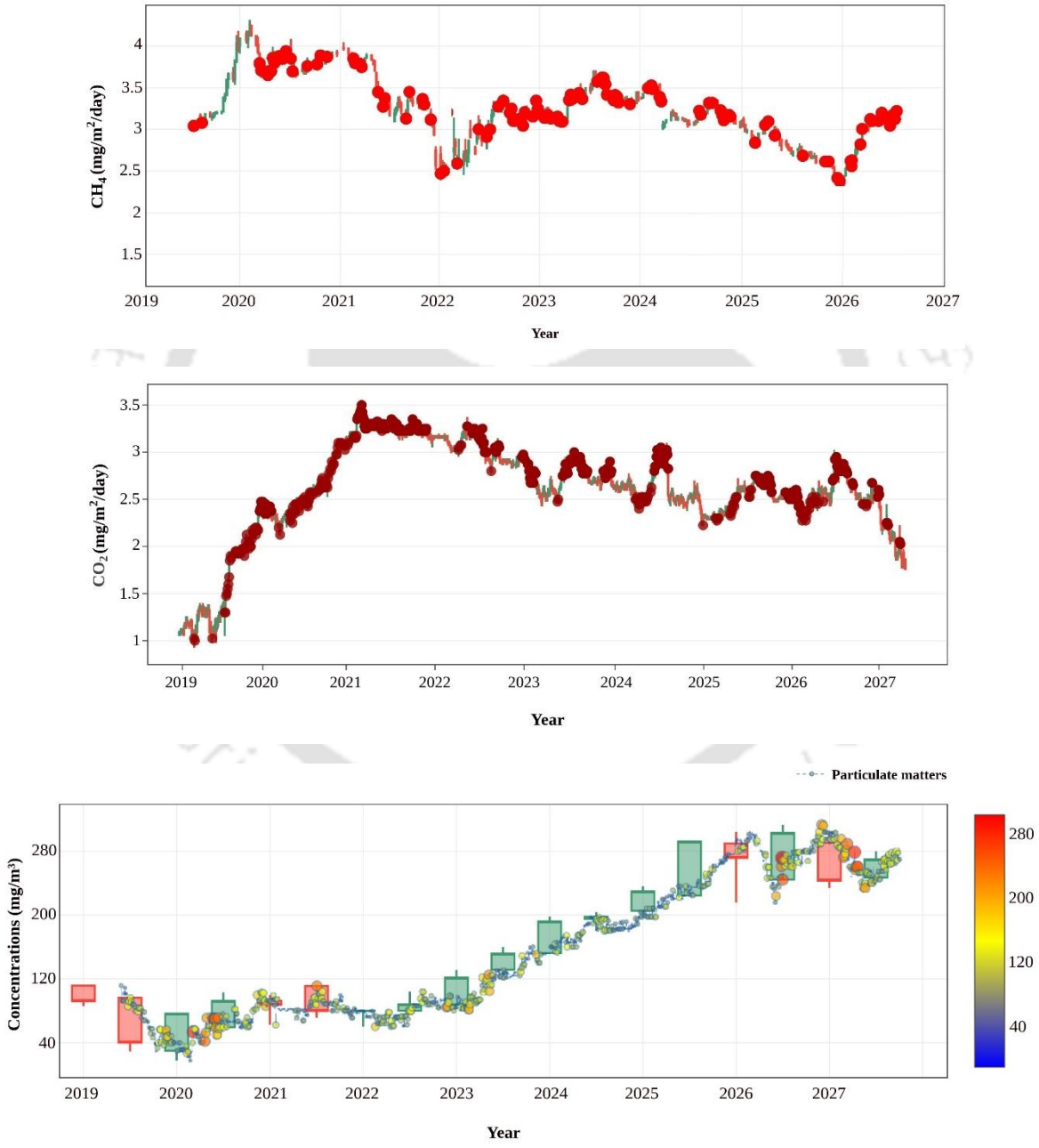


Fig. 6.8 Forecasted yearly trends of **(a)** CH_4 fluxes **(b)** CO_2 fluxes and **(c)** PM concentrations

(Fig. 6.8 (c)). The forecasted values suggested alterations in PM concentrations. This is due to factors such as meteorological conditions, emissions control measures, and regional air quality policies. These predictions clearly emphasize upon the importance of ongoing efforts to reduce PM pollution and mitigate its detrimental effects on human health and the environment.

Table 6.4 presents a comparison of the best findings of this work with those reported previously. Among these, Fallah et al. (2020) deployed ANN model for CH_4 fluxes prediction and reported a mean absolute percentage error (MAPE) of 3.03%. Also, Jayaraman et al. (2022) reported a MAPE of 2.418, RMSE of 2.433, MAE of 1.953, and an R^2 value of 0.271. Accordingly, the authors indicated the model's performance for CO_2 emissions prediction (632 $MtCO_2e$) and for input variables such as T, RH, Pr, and WS. However, the authors could have included other important meteorological parameters to further enhance the model robustness. In another study Abushammala et al. (2011), reported higher level of accuracy (R^2 of 0.937) for the model to effectively correlate and predict CH_4 emissions. Additionally, the mean squared error (MSE) was reported to be 0.0082, a value that as well validated very accurate model's performance for CH_4 emissions estimation. The model's accuracy was limited to the lower data size and could not delineate for the larger datasets. In this study, DT, RF, L2, L1 regularization, GB, SVR, *k*NN, MLP, and GA were utilized for the prediction of CH_4, CO_2 fluxes along with PMs. Comparatively, this work involved a data size of 12,492, a six-fold enhanced data size in comparison to the best case of Fallah et al. (2020). Also, all cases of CO_2, CH_4, PMs were not addressed in a comparative framework. Hence, with a impressive R^2, RMSE and IoA of 0.97, 0.32 and 0.94, the MLP provided best performance. Also, ARIMA was deployed for long-term forecasting.

The incineration plant in the Guwahati city has a daily capacity of 600 tons of MSW, while the landfill is designed to accommodate 1200 TPD for Guwahati city. These capacity values serve as essential parameters for the estimation and assessment of GHG emissions and particulate matters associated with waste management in the urban environment of Guwahati city. Along with these, the emissions from these sectors along with the standards mentioned by the Central Pollution Control Board (CPCB) have been mentioned in Table 6.4(a). The table conveys that the emissions of both GHG and PM of the city

have been lower than those set by the CPCB, India but comparable to the city-wide WTE data from the literature data (Karmakar et al., 2023).

Table 6.4(a). Literature comparison of incineration plant capacity and landfill capacity vis-a-vis PM and GHG emissions data of Guwahati city.

S. No	Case	Plant Capacity (TPD)	GHG emissions		PM emissions		References
			NO_x (mg m^{-3} day^{-1})	SO_2 (mg m^{-3} day^{-1})	$PM_{2.5}$ (mg m^{-3} day^{-1})	PM_{10} (mg m^{-3} day^{-1})	
1	Landfill (modelled emissions data)	1200	32	50	67	84	This work
2	Incineration (modelled emissions data)	600	63	68	89	102	
3	City wide Standard	-	80	80	60	100	Karmakar et al., 2023
4	City wide Standard	-	400	200		50	CPCB, New Delhi & Ministry of Environment, Forest and Climate Change (MoEFCC), New Delhi

Table 6.4(b) A summary of best findings of this work and prior art for the ML based prediction and forecasting of GHG and PMs emissions.

S. No.	Input parameter	Pollutants	Output	Data size	Model (s)	Prediction performance	References
1	T, RH, P, Pr, WS, WD, Population growth, GDDP, CH_4, CO_2, N_2O and SO_2 fluxes, $PM_{2.5}$/PM_{10}	CH_4, CO_2, N_2O SO_2, $PM_{2.5}$/PM_{10}	GHG emissions	12,492	DT, RF, L2, L1, GB, SVR, *k*NN, MLP, GA	Best model (MLP) R^2: 0.97; RMSE: 0.32; IoA: 0.94 MAE: 28.32; MAPE: 49.21; SSE: 0.21	This work
	Forecasting				ARIMA	R^2: 0.89 RMSE:5.67	
2	T, RH, Pr, WS	CO_2	CO_2 emissions ($MtCO_2e$)	632	GB, DT, RF, *k*NN, SVR, sARIMA,	MAPE:2.418 RMSE: 2.433 MAE: 1.953 R^2: 0.271	Jayaraman et al. (2022)
3	Mean T_{atm}, P_{max}, minimum & maximum RH, maximum WS	CH_4	Daily CH_4 generation rate	1883	ANN	MAPE 3.03%	Fallah et al. (2020)

4	Soil and T_{atm}, soil MC, CH_4 and O_2 concentration	CH_4	CH_4 emissions	121	ANN	R^2: 0.937 MSE: 0.0082	Abushammala et al. (2011)

6.7 Summary

In the Ph.D. thesis, measurement and analysis of major air pollutant concentrations during the year range of 1970-2018 was considered. The results of nine ML algorithms have been compared with six indices concerning their prediction accuracy for CH_4, CO_2, N_2O and SO_2 fluxes and PMs emissions. To determine the predictors' highest statistical significance and their interrelationship on the prediction of GHG emissions, several analyses were considered such as correlation, stepwise regression and NCA. Accordingly, T_{atm}, RH, P and Pr have been inferred to be the most influential parameters. For the stepwise regression (cubic) case of CH_4, CO_2, PMs, N_2O and SO_2, the R^2 and RMSE were 0.94 and 23.2; 0.96 and 9.5; 0.85 and 45.6; 0.78 and 31.4; 0.61 and 49.2 respectively.

Furthermore, the seasonal variation of $PM_{2.5}/PM_{10}$ was deciphered along with the extreme episodes. The PM concentration levels were 11 times above the daily limit value (60 mg m^{-3}) and the highest value was measured during the coldest period of the winters ($PM_{2.5}$ = 96.22; PM_{10} =138.73 mg m^{-3}). A clear seasonal pattern was identified for each pollutant. This was especially for the case of PM_{10}. For PM_{10}, the highest concentrations were in the winter and the lowest were during the rest of the year (Summer: $PM_{2.5}$ = 25.62; PM_{10} = 38.16 mg m^{-3}). Among all deployed algorithms, the MLP algorithm assured higher accuracy (with R^2 = 0.97, RMSE = 0.32, IoA = 0.94, MAE = 28.32, MAPE = 49.21 and SSE: 0.21). For forecasting, ARIMA configuration (1,1,0) × (0,1,1) was mostly acceptable but for the negative spikes at lag 12 with 5 iterations and with significant R^2 and RMSE values of 0.89 and 5.67 respectively.

The developed predictive model encourages the utility of data-driven decision-making processes. Relying upon empirical data and statistical analysis, the studied ML models facilitate evidence-based decision-making. Thereby, they lead to more informed and effective actions to address GHG. Also, the predictive model provided insights into the complex dynamics of CO_2 and CH_4 fluxes to better understand the GHG dynamics. Facilitating the identification and quantification of the key variables

and processes influencing fluxes, the modelling effort contributed a deeper realization of the trends and underlying factors of landfill and incinerators for the Guwahati city.

The accuracy of the models in this work can lead to better management and mitigation strategies for CO_2 and CH_4 emissions. Through the identification of factors that significantly influence fluxes, decision-makers can implement targeted measures to reduce emissions. These refer to optimization of land use practices, modification of agricultural techniques, or implementation of carbon capture and storage technologies. The model also allows for the assessment of the potential impacts of climate change on CO_2 and CH_4 fluxes. Through an incorporation of future climate scenarios, the reported modelling pedagogy can project alterations in fluxes for variant climate conditions. Accordingly, it can offer support to the schemes and strategies required for climate change scenarios and mitigation strategies. Also, the modelling efforts of this article can objectively define and set emissions reduction goals, effectiveness of existing policies and development of new regulations to curb CO_2 and CH_4 emissions for India and developing countries.

Chapter 7
Modelling Climate Sensitivity Analysis for the Machine Learning-based Prediction and Forecasting of Compost Production rate

Chapter 7

Application of ANN and Traditional ML Algorithms in Modelling Compost Production under Different Climatic Conditions

This chapter addresses the accurate prediction and optimization of compost production (CP) processes. To do so, complex relationships between climate variables and CP outcomes were considered using ML and hybrid algorithms. Section 7.1 orients the reader with respect to the relevant background and specific novelty in the chosen field of research. Following this, section 7.2 delineates upon the data attributes and correlation analysis for useful insights into competent for relationships between CP and climate variables, interdependencies and potential predictors. In section 7.3, the modelling and prediction performance findings for alternate ML algorithms has been presented. In the following section 7.4, the seasonal variations have been addressed. Section 7.5 elucidates on the long-term forecasting of the CP system with the ML and ARIMA algorithms. Section 7.6 delineates upon the findings of the case study for the Guwahati city. Following this, literature comparison has been addressed in section 7.7. Finally, section 7.8 summarizes the key findings of the fulfilled objective in the Ph.D. thesis[#].

Overview

Targeting valuable insights into the optimization of composting process for the realization of enhanced waste management strategy, the chapter considered diverse environmental conditions being experienced in variant climatic conditions. The data size consisted of approximately 864 sample records of the meteorological parameters and for the year range of 2010-2021. Model validation resulted in an RMSE of 0.757 and R^2 of 0.99 for GB model, and a correlation index of 0.68 between observed and predicted values as the best case in comparison to all other models. However, forecasted data for ten years affirmed the best performance of hybrid ARIMA-MLP with a standard error of 21.1152 and a CP annual yield of 74958 kg.

7.1 Introduction

In this chapter, MLP (a feedforward ANN), *k*NN and prominent tree-based ML models have been applied as predictive models for the modelling of CP rate. These ML methods can generalize the relationships between input and output variables (specific descriptors) through inductive inference, and can eventually contribute to informed decisions from the implicit relations in new cases. This will be based on the relationships learned from the empirical data. Also, possible scope can be explored for further investigations in the modelling efforts. The specific novelty of the addressed work in this chapter of the thesis are as follows:

[#] Published article: Singh, T., & Uppaluri, R. V. (2023). Application of ANN and traditional ML algorithms in modelling compost production under different climatic conditions. Neural Computing and Applications, 1-20.

(a) Application of MLP and traditional ML approaches for the prediction of CP rate for the Guwahati city with meteorological parameters, MERRA-2 data, waste data from CPCB and Open Government Data (OGD) Platform, India (2010-2021)

(b) critically study the influence of seasonal variations on the composting process performance for the Guwahati city

(c) forecasting of the CP rate for the Guwahati city with a time series model (ARIMA)

7.2 Data Characteristics and Correlation Analysis

The datasets being summarised in Table 7.1 refer to the scenario prior to the pre-processing stage. They constituted 180 compost data points. Also, 762 more data points were tied to MERRA-2 and CPCB data. After pre-processing, the compost data and MERRA-2 and CPCB data were downsized to 144. The downsizing of the datapoints was due to two main reasons. Firstly, data analysis requires quality datasets to solve important problems and resolve critical questions. If the data is incorrect or redundant and out of date or poorly formatted, it would not be able to fulfil its intended purpose. Henceforth, it is mandatory to "clean" and prepare data and thereby enhance the focus on the quality of the data. Secondly, downsizing of the datapoints facilitates maximization of the available information. Accordingly, it can ensure upon the creation of comprehensive and targeted results on a granular level. Thus, the approach improves the ability to transform data into actionable insights. In this study, OW data was constructed using MSWG rate and for a fixed choice of OW as 53 wt% in the overall MSWG of the Guwahati city (Singhal et al., 2022).

Table 7.1 Conditioned and integrated datasets summary for the training and testing studies of CP prediction under variant climatic conditions.

Parameters	***Data Sources***	***No. of Datapoints***
Dependent variable		
Compost Produced *(CP)*	OGD (2011-2021)	144
Independent variables		
Temperature *(T)*	MERRA-2 (2011-2021)	144
Relative Humidity *(RH)*	MERRA-2 (2011-2021)	144
Precipitation *(P)*	MERRA-2 (2011-2021)	144
Wind speed *(WS)*	MERRA-2 (2011-2021)	144
Organic Waste *(OW)* [Calculated]	CPCB (2011-2021)	144

The correlation analysis was primarily conducted to determine and rank the climatic parameters and waste factors. Higher correlation indices with respect to the dependent variables was sought. Accordingly, the parameters with weak or no correlation with one other were isolated. Such analysis would be helpful to obtain best relevant information through the modelling efforts. To do so, a correlation matrix that estimates the correlation index between each variable pair has been determined and analyzed.

Table 7.2 summarizes the correlation matrix among CP, climatic parameters and OW. The first column in the table confirms that among all parameters, the T possessed the highest correlation coefficient of 0.88. Fig. 7.1 illustrates this positive correlation between T and CP and thereby inferred that CP rate is significantly influenced by T. The positive trends of OW-CP (0.74) and P-CP (0.43) have been consistent with the inferences indicated in the relevant prior art (Thomas, 2004; Sun et al., 2011). Also, correlation coefficients for the pairs RH-CP (0.18), WS-CP (0.32), have all affirmed positive correlation. Further, RH-P (0.68), P-T (0.54), WS-T (0.30) parameter pairs have been observed to be significantly correlated with one another and RH-T (-0.42) and WS-RH (-0.22) have been negatively correlated to one another. Accordingly, to subdue the multi-collinearity issues, such model pairs have been avoided in due course of the modelling effort. Scattered plots of Population growth, LP, GDDP, HH counts and sizes with respect to MSW rate further illustrate such pairing effect (Fig. 7.1 (a-e)).

Table 7.2 Correlation matrix data of CP, climatic parameters and OW in the CP prediction case study.

	CP	***T***	***RH***	***P***	***WS***	***OW***
CP	1					
T	0.83	1				
RH	0.18	-0.42	1			
P	0.43	0.54	0.68	1		
WS	0.32	0.30	-0.22	0.32	1	
OW	0.74	0.13	0.17	0.11	0.14	1

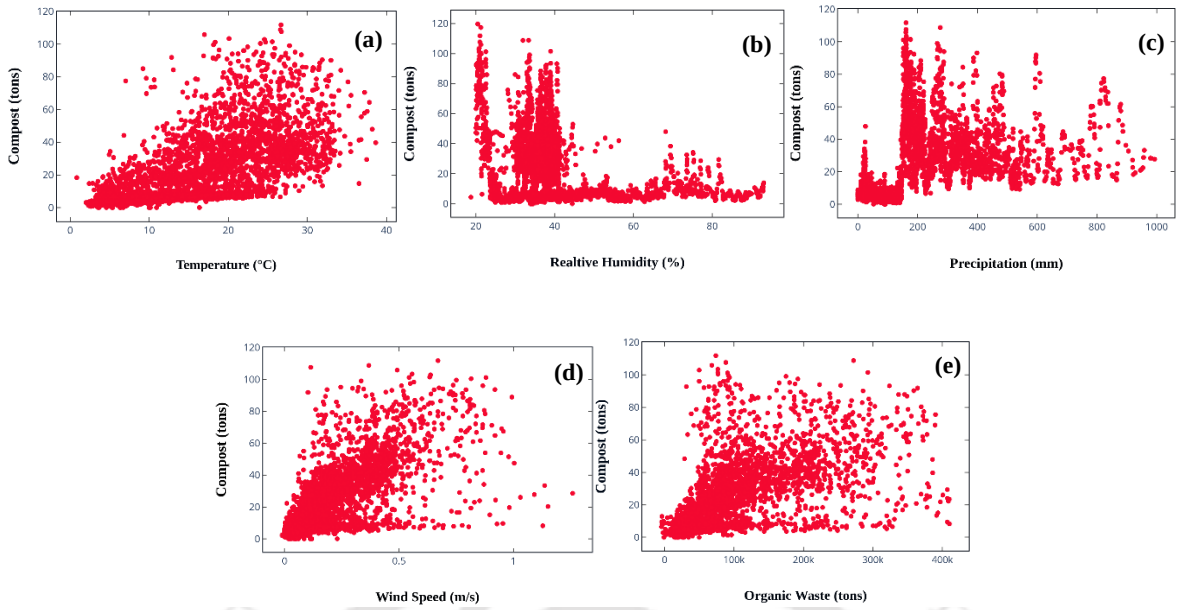


Fig. 7.1 Graphs depicting correlation results of compost generation rate with respect to various variables **(a)** T **(b)** RH **(c)** Pr **(d)** WS and **(e)** OW rate

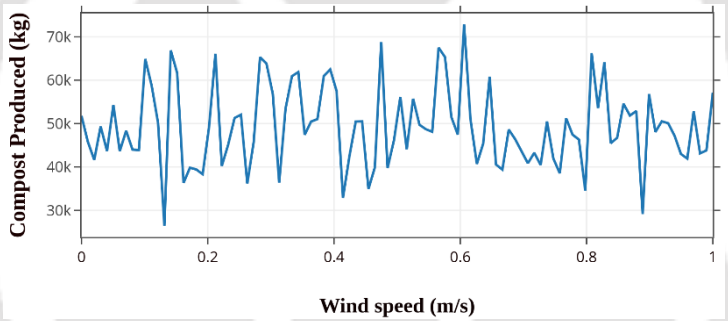


Fig. 7.2 CP rate alteration with WS (m s^{-1}) in the compost generation rate case study

RH and WS exhibited lesser positive influence on the CP. This is due to the reason that enhanced composting time and stronger winds significantly detriment temperature and vice versa for lower WS. Thus, the low significance in the correlation for the parameter is justified. Also, a maximum peak can be seen for a mature CP annual rate of 72,000 kg and moderate WS (0.6 m s^{-1}) followed by 68,000 kg annual rate of CP with lower WS (0.47 m s^{-1}). Similarly, the lowest (26,000 kg; 0.13 m s^{-1}) and second lowest peaks (29,000 kg, 0.88 m s^{-1}) do confirm that the CP enhancement does get affected by the moderate WS (Fig. 7.2).

7.3 Prediction Efficacy of ML Algorithms

7.3.1 k- Nearest Neighbour

The *k*NN functioned as an average model with train and test scores of 0.81 and 0.79 respectively. The *k*NN model was trained for *k*=2. The model's accuracy was reduced with increasing value of *k*. The RMSE, SSE and R^2 values for *k*NN were 59.56, 63.41 and 0.77, respectively. The model's prediction and actual values can be fitted with a line and can thereby explain that their relationship that does not get affected due to outliers/noise in the data, as illustrated in (Fig. 7.3a). However, certain limitations do exist for the the non-generalization of the datasets. In due course of the fitness, the *k*NN model aggregated the *k*-nearest values of the training dataset and could not explain the underlying structure of the data distribution and the relationship between the dependent and independent variables. Thus, the model was inaccurate due to the inclusion of testing new datasets from different *in-situ* measurements that account for a climatic data with respect to WS, RH and T range. Another important issue is that almost no training was involved with the *k*NN. Most operations occurred during the testing phase, which is computationally expensive computationally and slow. For this reason, even though *k*NN performed well, it was not suitable for the chosen problem and study.

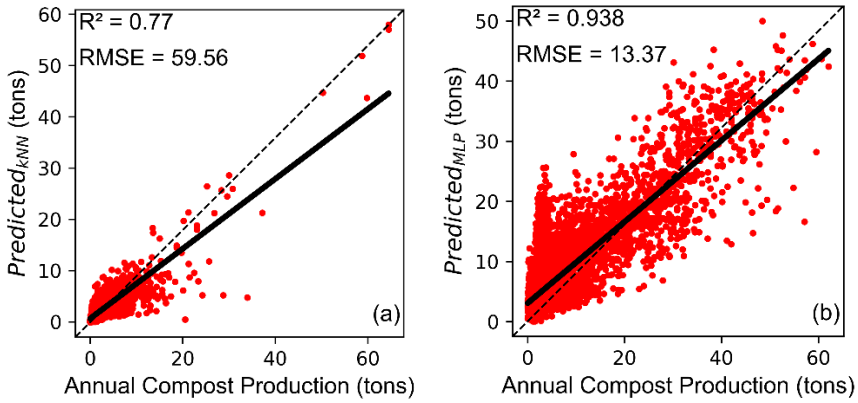
7.3.2 Multilayer Perceptron

The MLP model's train (0.87) and test (0.85) scores have been consistently higher. The MLP was trained with three hidden layers (size of five for each layer), *tanh* as activation function and with few other peripheral parameters. The RMSE (13.37) and SSE (24.29) values have been better but not the best. The R^2 (0.93) value has been significantly better than other studied models. The predicted data is still susceptible to outliers and this was confirmed by the scatter plot (Fig. 7.3(b)). The MLP can be extremely powerful as a competent regressively training model and can henceforth foster the production of more accurate models. However, for the case, due to computational issues, the model could not be trained with more hidden and deep layers.

7.3.3 Tree Models

The two tree-based models namely GB and RF produced quality results. Train scores of 0.98 and 0.78 and test scores 0.96 and 0.76 were obtained for GB and RF respectively. The two models have been trained with trees of maximum depth of 15 and 5 respectively. The RF and GB have been effective to

fix the overfitting issue. The scatter plots for RF and GB can be seen in Fig. 7.3(c) and (d) respectively. The data points for GB have been closer to the fit line in comparison to the RF. The RMSE (0.757), SSE (12.126) and R^2 (0.757) of GB have been better in comparison to the RMSE (29.1), SSE (38.37) and R^2 (0.823) of RF. This is probably due to the reason that the GB upon being trained for one tree at a time could correct the mistakes of the previous trees for each new tree. On the other hand, trees in the RF were constructed independently. An RF tree can choose its outputs in any order as each tree is independent. Thereafter, all of the individual predictions were combined into a single one, which is either the average value in regression issues or the majority class in classification problems. On the contrary, the order was set for the execution of GB trees. Henceforth, the sequence could not be altered. Hence, it can be concluded that the GB provided the best results for all parameters than RF. Moreover, at times, the RF repeated the predicted value for data points closer to one another. However, this was not the scenario for the GB model. A comparison of R^2 for the models has been presented in Fig. 7.4(a). The RMSE, MAE and index of agreement (IoA) for the models have been illustrated in Fig. 7.4 (b), (c), and (d) respectively. The ratio between the MSE and the potential error has been represented with the IoA. A perfect match indicates an IoA value of 1, whereas 0 refers to no agreement at all. The RMSE is a measurement of goodness of fit. Also, lower RMSE implies a lower error or better fit. The overall evaluation and comparison of the four ML algorithms confirm that GB is the best-suited ML model.



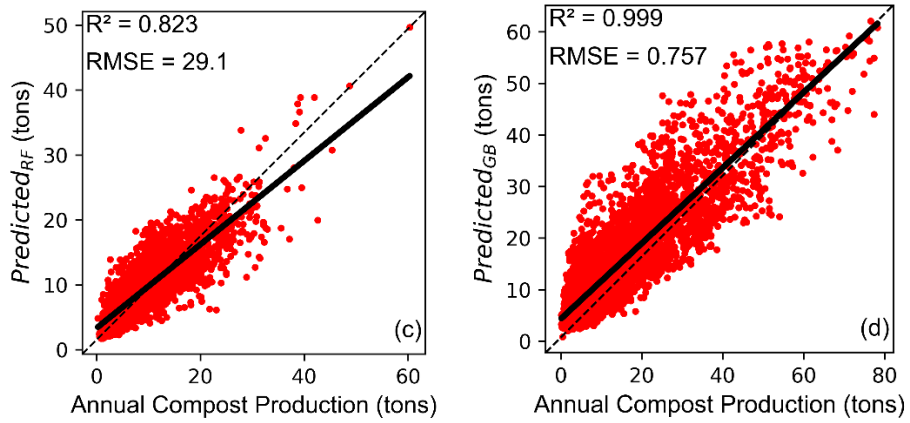


Fig. 7.3 Parity plots depicting prediction efficacy of **(a)** kNN **(b)** MLP **(c)** RF and **(d)** GB models in the climate variation-based CP rate prediction case study

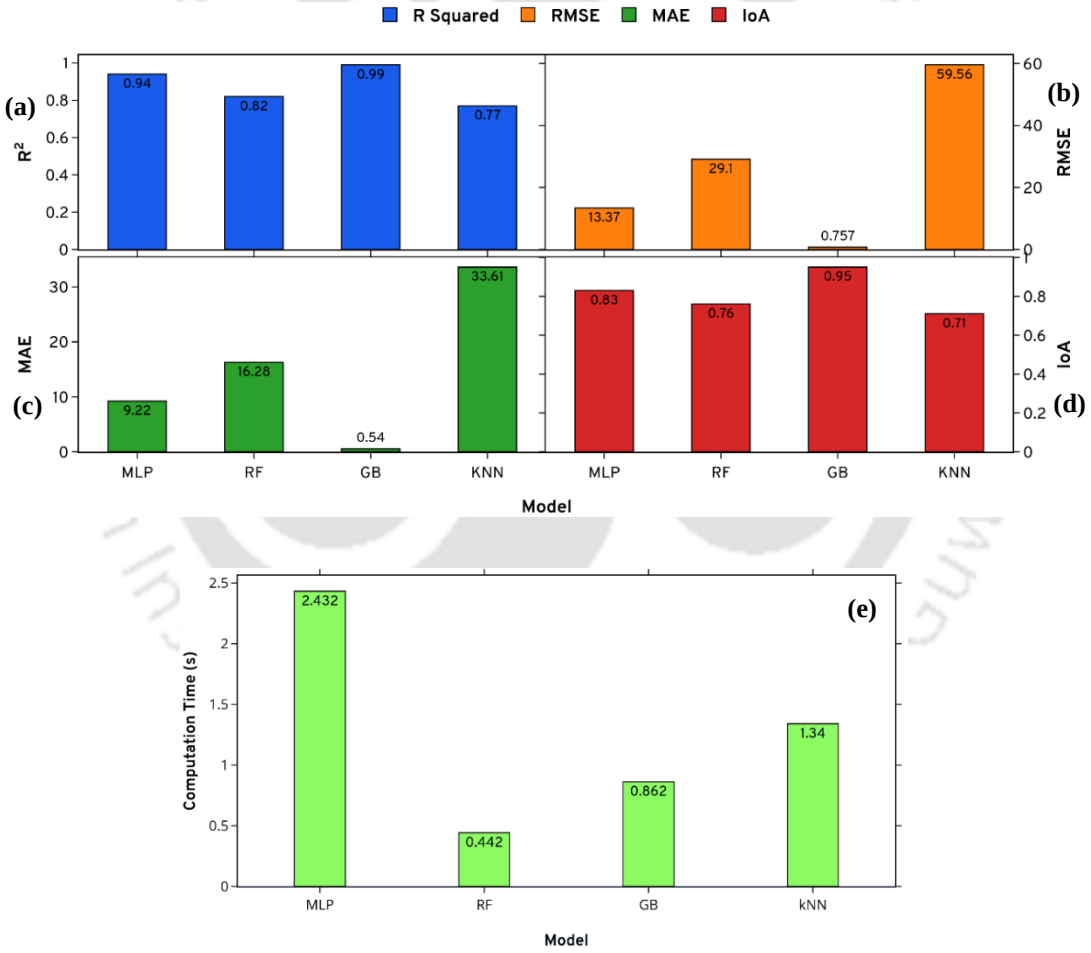


Fig. 7.4 (a) R^2, **(b)** RMSE, **(c)** MAE, **(d)** IoA and **(e)** computational time of alternate ML models in the climate variation-based CP rate prediction case study

Fig. 7.4(e) depicts a comparison of each model in terms of their computation time (sec). Comparatively, the RF, as a lesser complex model had lower computation time (HPO time of less than 0.44 s). However,

for GB, the HPO time has been about 0.862 s. Even higher complex models such as *k*NN and MLP demanded the maximum computation time of 2.432 and 1.34 s respectively.

7.4 Influence of Seasonal Variations

Seasonal parametric variations have been presented in Table 7.3. T critically influenced the composting process performance. Also, the released heat influenced process performance. In due course of CP, temperature alterations do follow a pattern (Khalil et al., 2011). The most favourable conditions for compost generation have been in the summer season (Apr-Jun) (Fig. 7.5 (a)). For the season, T_{max} reached a value above 30 ± 2°C and RH enhanced to a value above 80%. Henceforth, they prompted the realization of highest CP rates. On the contrary, the monsoon season (Jul-Sep) (Fig. 7.5 (b)) is challenging for the CP rate due to air moisture content enhancing composting time and eventually delaying the evaporation of moisture in the compost. However, during winter season (Dec-Feb) (Fig. 7.5 (c)), with a T_{max} of 20°C and RH below 17%, a marginal influence has been apparent on the CP rate. Thus, for the progressive months, CP rate gets affected with a negative pattern.

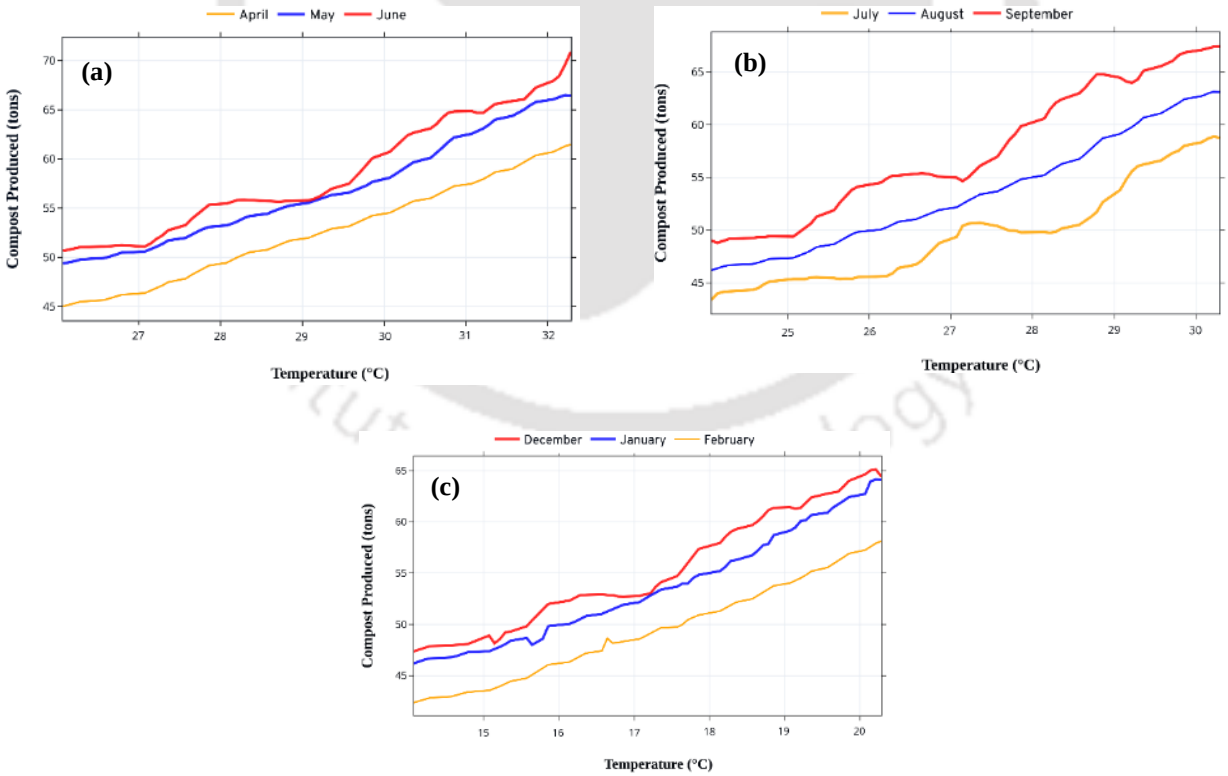


Fig. 7.5 Graphs depicting seasonal variation of CP rate of Guwahati city during **(a)** Summer, **(b)** Monsoon and **(c)** Winter seasons

Table 7.3 Seasonal parametric summary (T, RH, WS and P) in the CP rate prediction case study.

Parameters	***Seasons***		
	Summer	*Monsoon*	*Winter*
T (°C)	30.0 ± 5	30.0 ± 2	15.0 ± 5
RH (%)	76.0 ± 2	82.0 ± 4	69.0 ± 3
WS (m s^{-1})	1.3 ± 0.2	0.8 ± 0.3	0.6 ± 0.2
P (mm)	244 ± 4	549 ± 6	10 ± 0.18

7.5 ARIMA based Long-term Forecasting of CP Rate

In this study, ARIMA methodology was deployed for the forecasting of CP rate and with statistical models. The data until December, 2021 was used for the ARIMA modelling. Based on the mentioned dataset, the CP rate was forecasted with various algorithms. Among several attempted traditional ML algorithms, the ARIMA-MLP confirmed best accuracy. The ARIMA (model.predict_test.csv) data corresponds to the data obtained until December, 2021. The dataset was obtained by exporting the ml prediction dataframe into .csv format.

Also, noise reduction was achieved along with easy outlier removal. The strategy was based on the gathering of long-range correlations. In order to determine the stationarity of the forecasted trends, the autocorrelation result of the differential sequence was examined (Table 7.5). The autocorrelation generates a short-term correlation in the sequence. This was subsequently subjected to the white noise test. The sequence used to fit the ARIMA model was not classified as white noise. This was due to the result possessing a significance level of the test (0.05) and the p-value of the delayed statistics being lower than 0.05. The ARMA (p, q) models whose moving average delay order and autocorrelation delay order have been both less than or equal to 5.

Since the errors of the model were autocorrelated, the arising problem was fixed by adding one lag of the dependent variable to the prediction equation i.e., by regressing the first difference of Y on itself lagged by one period. For the parameter (p, d, q) configuration, the lowest value of standard error has been treated as the best configuration. From such configuration, ARIMA-MLP and ARIMA-GB have been set as the best configuration of ARIMA to forecast future values (Table 7.4). Therefore, the final ARIMA equation refers to ARIMA (1,1,0). This conveys which indicates that it has one autoregressive (AR) lag, one as the order of differencing to get the stationary series, and has no moving average (MA)

terms. Thus, the model corresponds to a first-order autoregressive model with one order of non-seasonal differencing and a constant term i.e., an ARIMA (1,1,0) model. Accordingly, the following prediction equation was obtained:

$$\overline{Y_t} = \alpha + Y_{t-1} + \theta(Y_{t-1} - Y_{t-2}) \quad \textbf{(7.1)}$$

where $\overline{Y}_t$ is the forecasted goal, α is the constant and θ_1 is the slope coefficient.

The ARIMA (1, 1, 0) model was used to predict the CP for the next ten years, and the predicted and the forecasted results for the year 2021 have been correlated for all models (Fig. 7.6(a-b)). The table conveyed that the ARIMA-MLP assured best result among all models. As being depicted in the graphs, future trends for Guwahati city have been forecasted for the year range of 2022-2031 and with the predicted past data. Table 7.6 compares the validation results for the forecasted CP rate with the actual CP rate (60000 kg) for the year 2021.

Table 7.4 Standard error rate values for alternate ARIMA configurations in the CP rate prediction case study.

ARIMA→	**Parameter(p,q,d) configuration**		
Models↓	(5,1,0)	(3,1,0)	(1,1,0)
MLP	99.49	17.97	2.66
GB	82.11	22.92	9.34
RF	91.53	38.18	18.50
***k*NN**	299.38	98.63	56.60

Table 7.5 Autocorrelation table findings at Lag value (K) of 17 in the CP rate prediction case study.

Lag	**Covariance**	**Correlation**	**Standard error**
0	83.2251	1.0000	0
1	46.0489	0.50443	0.15214
2	31.6231	0.40221	0.18522
3	27.4317	0.31483	0.20418
4	20.1193	0.21417	0.22676
5	24.7127	0.24448	0.23719
6	29.5949	0.31109	0.24531
7	22.3621	0.23915	0.25727
8	8.0165	0.11613	0.26010
9	5.2820	0.05206	0.26093
10	-0.8140	-.009138	0.26356
11	-3.8215	-.03112	0.27118
12	-5.7717	-.06249	0.27323
13	-7.0428	-.05982	0.27534
14	-14.6572	-.13824	0.27602
15	-25.3291	-.25910	0.28114
16	-20.4936	-.21427	0.28279

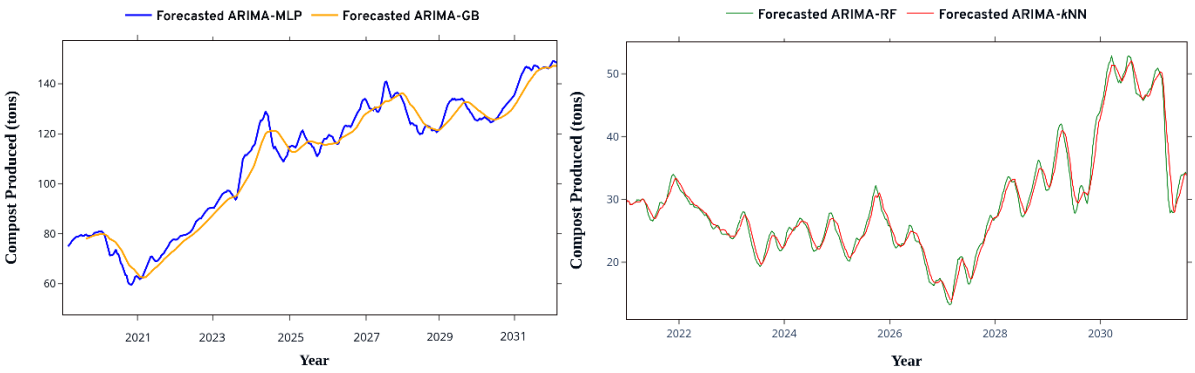


Fig. 7.6(a-b) Temporal series results for each model to forecast CP rate in the year 2022-2031

Table 7.6 Standard error values summary for alternate ML models in the CP rate forecasting case study.

Forecast for variable CP				
Models	***Forecasts***	***Standard Error***	***95% CL***	
MLP	74958	21.1152	74912	75002
GB	78009	26.4123	77993	78017
RF	29174	51.9438	29183	29201
*k*NN	30042	49.2242	30013	30079

Table 7.7 Organic fraction composition of Guwahati city.

Waste Collection Area	Vegetable/ Fruits/ Food Waste (%)	Paper (%)	Garden/ tree leaves (%)	Wood Scraps (%)
High Income Group	33-67 %	8-17 %	0.12-7 %	1-4 %
Middle Income Group	37-53 %	13-19 %	0.53-3 %	0.46-2.5 %
Low Income Group	23-44 %	7-16 %	0.78-17 %	0.34-5.2 %
Slum	18-41 %	4-21%	0.21-7 %	0.16-5 %
Commercial	21-46 %	14-24 %	0.06-13 %	0.54-3 %
Vegetable Markets	13-50 %	13-38%	1.16-14 %	1.3-14 %
Average	29-46 %	11-22%	2.67-8 %	0.45-4 %

7.6 Model Sensitivity with respect to OW Fraction

The organic fraction of MSW varies significantly and is region specific. On an average, MSW consists about 40-45% of organic fraction (CPCB, 2015). In another article (Kashyap and Borthakur, 2019), the organic fraction characteristic for Guwahati city was mentioned (Table 7.7). Thus, based on prior art, it can be analyzed that the organic fraction varies about 40-65% of the MSW for Guwahati city. Considering this, the performance of each ML model was evaluated with OW data being generated with a random number. Accordingly, the objective function has been appropriately modified as:

$$CP = f(T, RH, P, WS, OW_{[40-65\ \%]}) \quad \textbf{(7.2)}$$

Recently, a novel metaheuristic called Jaya was used to solve a variety of challenging optimization problems (Venkata Rao, 2016). Numerous metaheuristics use some type of stochastic optimization that involves the outcome as a complex function of the generated random variables (Bianchi et al., 2009). Metaheuristics are generally more effective than optimization algorithms, iterative techniques, or basic heuristics in combinatorial optimization. This is due to their ability to conduct searches effectively in a larger feasible space range (Blum and Roli, 2003). No algorithm-specific control parameters are to be set for the Jaya algorithm, as it simply needs common regulatory parameters. The population size and the maximum number of function evaluations (iterations or generations) are the often-set criteria. The idea behind such a population-based algorithm is that the search process should constantly lean toward the best design and shall steer away from the worst solution. At iteration *i*=0, a uniform distribution function is often used to generate a random population of people (possible solutions) in the chosen search domain. Then, from the original population, the best and worst answers are taken out and are used to update other solutions for the following iteration. At each iteration, the best and worst solutions must be updated.

The following function defines the simple movement strategy (i.e., updating equation) for a new solution:

$$x'_{m,n} = x_{m,n} + r_{m,1}\left(x_{m,best} - \left|x_{m,n}\right|\right) - r_{m,2}\left(x_{m,worst} - \left|x_{m,n}\right|\right) \quad \textbf{(7.3)}$$

$$x_{m+1,n} = \begin{cases} x'_{m,n}, & if\ f(x'_{m,n}) > f(x_{m,n}) \\ x'_{m,n}, & if\ f(x'_{m,n}) \leq f(x_{m,n}) \end{cases} \quad \textbf{(7.4)}$$

In the above expressions, $x_{m,n}$ corresponds to the for the nth candidate solution at the mth iteration; $r_{m,1}$ and $r_{m,2}$ are two independent random values being produced from a uniform distribution *U* in the range [0,1] to impart a stochastic exploration to the algorithm in the search space. The $x_{m,best}$ and $x_{m,worst}$ nomenclature respectively refer the best and worst candidate solutions in the entire pool.

The term $r_{m,1}\left(x_{m,best}-\left|x_{m,n}\right|\right)$ denotes a solution's propensity to become more similar to the best solution. In order to assist a candidate solution to avoid the worst possible solution, the term $r_{m,2}\left(x_{m,worst}-\left|x_{m,n}\right|\right)$ has been created. The $x_{m,n}$ will be retained in the subsequent iteration if it offers a superior objective function value than the $x_{m+1,n}$ being computed with equation 7.4

After simulation, the accuracy of MLP, GB RF and *k*NN varied significantly. The predictive performance of MLP outperformed with respect to the GB. This is due to the reason that the MLP works well with unstructured data points and it has better self-learning capabilities. The forecasting results also convey the superior performance of the MLP-ARIMA model. This is due to the effectiveness of as hybrid models in overcoming limitations of the forecasting accuracy of single models.

The seasonal evaluation of summer, winter and monsoon seasons results in the CP rate alteration by approximately ~30%, ~20% and ~15% respectively. This is similar to the findings reported in previous investigations. Hence, the variation does not indicate any influence on the mentioned randomization. Fig. 7.7(a-d) illustrate accuracy and loss graphs to demonstrate the suggested model performance. In these plots, the y-axes denote the accuracy and loss and the x-axis refers to the sample size. The x-axis has been represented as percentage or the number of training cycles in the entire dataset. A closer examination of accuracy graph reveals that for a few samples in the beginning, the curve increases steeply for each model (80%, 88%, 93% and 95% accuracies for *k*NN, RF, GB and MLP respectively). Also, with increasing percentage samples, both accuracy curves depict an upward trend. In actuality, the first through third sample sizes corroborate to the dramatic expansion scenario. Thereafter, the growth rate steadily reduced and reached a stagnant profile. Similarly, loss measures refer to the model error and hence the poor performance status of the model. Thus, the figures thereby convey that the loss (error) eventually became lower and better model performance was achieved. Despite indicating the existence of few peaks and valleys in the longer time frame, the reduction of loss in due course of time affirms that the model learnt effectively and henceforth performed well.

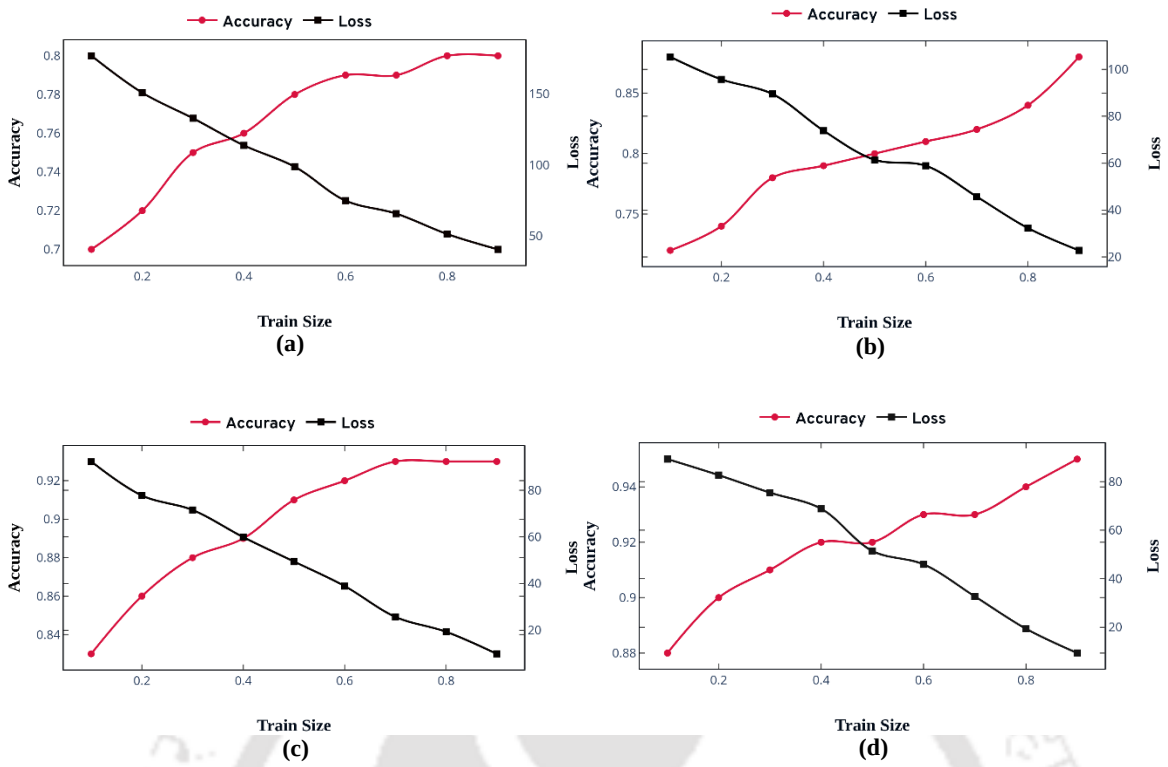


Fig. 7.7 Accuracy and loss graphs for **(a)** *k*NN **(b)** GB **(c)** RF and **(d)** MLP models in the CP rate prediction case study

Moreover, a larger dimension of the training data enhanced ML performance and upto a certain extent. Beyond this, it may compromise upon the performance of the ML algorithm. Fig. 7.7 (b) (for GB) and 7.7 (d) (for MLP) depict that the performance accuracy enhanced with iteration values of the training size. Accordingly, the figures infer that both GB and MLP models are less biased. However, for *k*NN (Fig. 7.7 (a)) and RF (Fig. 7.7(c)), the pertinent plateau effect phenomenon affirmed that the effectiveness measures reduced with time. The existence of plateau affirmed that the learning rate reduced with time. Thus, these findings conveyed that the inaccuracy altered with respect to the training iterations and lessened than the threshold value of θ. Henceforth, these models can be concluded to be susceptible with overfitness issues and with increasing size of the iterations.

Table 7.8 A summary of performance indices of alternate ML models for the CP rate forecasting case study.

	Performance Indices			
Models	***RMSE***	***R^2***	***IOA***	***Computation Time***
kNN	40.52	0.80	0.75	1.51
RF	22.8	0.88	0.82	0.48
GB	9.85	0.92	0.86	0.913
MLP	9.32	0.95	0.87	3.125

Table 7.9 A summary of best findings of this work and prior art in the CP rate prediction and forecasting case studies.

Application	Authors	Approach	Data Size	Input parameters	Performance Evaluation	Inference
Compost Maturity	This work	MLP (feedforward ANN)	864 (Historical data: 2010-2021)	T, RH, Pr, OW and WS	RMSE: 13.37, R^2: 0.93	Experimental parameters such as moisture content, pH etc., can be included to check the performances using ML models
		GB			RMSE: 0.75, R^2: 0.99	
		RF			RMSE: 29.1, R^2: 0.82	
		*k*NN			RMSE: 59.56, R^2: 0.77	
					Model accuracy: 93%	
	Kujawa et al. (2014)	ANN	1536	Colour and texture features of the compost	Classification error: 1.56%	Lack of comparison experiments and high complexity
	Gao et al. (2007)	WNN	500	T, MC, VS, value of faecal bacteria & GI	MSE = 0.06 and a testing accuracy of 87.5%	The model had limited data

Table 7.8 presents the performance indices of each model for the case study. The analysis indicates that the MLP performed better than the all-other considered ML models (RMSE value of 9.32, coefficient of determination, (R^2) of 0.9 and IoA of 0.87). However, with the increasing complexity of the model, the computation time for the MLP was found to be high (3.125 s). Table 7.9 summarises the findings in comparison with the best available prior art. The high-end ML algorithms yielded highly accurate findings and these are similar to the previously reported datasets (Kujawa et al., 2014; Boniecki et al., 2013). Lower-complexity ML models can also outperform higher-complexity models, but with a certain concern towards their accuracy. Lesser historical data records, other approaches and procedures being linked with hyperparameter optimization, and differences in the training to testing split ratio have all been liable for the algorithms' poor performance. As a result, good performance of the ML algorithms can be achieved towards the effective prediction of the CP rate by considering these factors in the modelling effort for the mitigation of overfitness problems.

7.7 Summary

MSWG data at the regional level is essential for the affirming of a successful SWM planning phase. However, many regions and particularly those that are developing do not have trustworthy data. To predict CP in Guwahati, the conducted study deployed publicly available data and adequate data analytics. Five parameters were used (T, RH, WS, P and OW) as independent factors. With best R^2 for T (0.83) and OW (0.74), the data analysis infers that these factors may explain the majority of the variability in CP rate. However, RH (0.18) does not show much significance in CP. This study also contributes to the efficacy of alternate estimating methods for CP rate determination. *k*NN, MLP and tree-based models, RF, and GB, were chosen and compared based on their strengths and weaknesses as popular and powerful ML models to represent CP data. In comparison, the GB produced promising outcomes in CP prediction with R^2 and RMSE values of 0.99 and 0.757, respectively. Following the GB, the MLP performed well with good R^2 (0.938) and RMSE values (13.37). However, with the deployment of randomization technique, MLP outperformed in the prediction phase and assured an RMSE, R^2 and IoA values of 9.32, 0.95 and 0.87 respectively. The overall model accuracy was found to be 90%. Seasonal analysis also confirms that the favourable conditions for compost generation have been in the summer season (Apr-Jun) with T_{max} value of about 30 ± 2°C and RH value above 80%. For the forecasting of CP, ARIMA was deployed. Accordingly, ARIMA-MLP model performed better in comparison to *k*NN, GB and RF models. This method for CP forecasting can be utilized as a model for other locations that dealt with similar development and environmental challenges.

An accurate prediction of CP rate is important for sustainable and efficient solid waste management. With this as a primary objective, the widely used ML models have been proven for their efficacy. Despite affirming thematic prior art, the chapter for the first time assessed the performance of *k*NN, RF GB and MLP for CP rate prediction. Also, ARIMA approaches have been deployed for long-term forecasting of the CP rate. In the chapter, the comprehensive major findings are as follows. Firstly, the HPO fine-tuned tree-based ML models had higher test scores and confirmed superior model predictive accuracy values. Secondly, hyper-tuned RF exhibited linear prediction and the GB model performance improved with the enhancement of the learning speed. Thus, the suggested methodology is generic in nature and can be applied suitability for any city and for much complex input datasets through an

appropriate modification of influential parameters. Also, the accuracy graph-based analysis confirmed that for a given set of samples, each model curve grew quickly during the initial stages, and the *k*NN, RF, GB and MLP had accuracies of 80%, 88%, 93% and 95% respectively. Finally, the quest for generalized application of ML algorithms for CP rate prediction has been complimented to target MLP and GB for its prediction speed and accuracy to handle complex datasets with greater ease. The research confirmed that the estimation of MSWG rate is very crucial for the subsequent system planning of MSW management from both short- and long-term perspectives. Using historical data in different cities, the presented generic approach can be applied for CP rate prediction in any city in the world. To do so, influencing parameters need to be investigated through careful introspection and analysis of the complex data. Best findings of this work can foster the Guwahati metropolitan region to initiate and develop an integrated decentralized community-based SWM approach. Accordingly, enhanced recycling and composting practices (bio-waste) can be commissioned for the subsequent realization of a circular bio-economy.

Chapter 8
Prediction and Forecasting of Biogas Production Rate with ML, ARIMA and MA Approaches

Chapter 8

Biogas Generation Prediction from Meteorological Parameters and Organic Waste using ML Approaches

The chapter highlights the importance of accurate biogas yield prediction and forecasting for optimal renewable energy production. Accordingly, it addresses the hybrid modelling effort that systematically accommodates simultaneous lab scale investigations and can be envisaged with the MLP algorithms for their promising performance. After summarizing the relevant introductory information and objectives and scope of the study in section 8.1, section 8.2 discusses the findings of the steps taken after pre-processing of the dataset. Also, the section as well explores the correlation between meteorological parameters, OW, and BP. Thereafter, section 8.3 and 8.4 respectively discusses the ML models used for BP rate prediction and ARIMA and MA models applications for long-term forecasting of BP rate. The influence of seasonal variations on BP rate and the utilization of ARIMA models to capture and analyze these variations have been investigated in section 8.5. Following this, section 8.6 summarizes the findings in comparison with the best available prior art. Finally, section 8.7 presents the key findings of the study and provides a concise summary of the research on biogas generation prediction and long-term forecasting[#].

Overview

The chapter summarizes the findings associated to climatic conditions-based BP rate variation of meteorological and OW with respect to alterations in parameters. The data size consisted of 728 sample points records of the meteorological parameters, and refers to OW and biogas yield data for the year range of 2015-2021. The model validation studies resulted in an RMSE of 2.76 and R^2 of 0.94 for the MLP model. Additionally, the ARIMA-MLP model resulted in the best performance for the forecasted data (with a standard error of 5.1152 and a BP yield of ~90 m^3). Thereby, the findings of this thesis work convey the limitations of the broader application of the empirical approaches and the feasibility of ML algorithms as a potential reconstruction technique for the development of robust and accurate region-specific BP rate models. Henceforth, they can serve as much needed tools for integrated circular agricultural system development and a sustainable global future.

8.1 Introduction

This chapter targets the criticality of climatic parameters for BP prediction with ML techniques. Climatic and other suitable meteorological conditions, along with other physical, biological, and chemical parameters, profoundly contribute to the BP. SVR, *k*NN, tree-based ML models, MLP (a feedforward ANN) were employed as predictive models. The targeted investigation aims to evaluate

[#] Published article: Singh, T., & Uppaluri, R. V. (2023). Feed-forward ANN and traditional machine learning-based prediction of biogas generation rate from meteorological and organic waste parameters. *Journal of Super Computing. 1-34.*

the critical influence of meteorological variables namely T, P, RH, Pr, and OW on BP. Through inductive inference, the ML techniques have been extended to investigate the correlations between input and output parameters and eventually assist in the informed decision making based on the implicit correlations learnt from the empirical findings. Thereafter, using performance metrices (R^2, RMSE, MAE, IoA), the best fit models, computation time and the accuracy and loss graphs were identified. Further it shall be noted that useful insights have been gained with respect to seasonal patterns for the identification of microbial activity, gas production rates, and overall biogas yield. Also, the utility of statistical analysis models such as ARIMA and MA (comparative analysis) were targeted for long-term forecasting of the BP rate. The findings contributed to the development of strategies for optimal BP rate generation and accordingly realize enhanced sustainability of energy generation in Guwahati city.

8.2 Dataset Processing and Correlation Results

The datasets summarised in Table 8.1 involved 162 biogas data points during pre-processing. Also, 728 more data points corresponded to MERRA-2 and CPCB data. After pre-processing, the biogas data, MERRA-2 and CPCB data together downsized to 118. In this study, OW data was assumed to be 53 wt% of the overall MSWG for the Guwahati city (Singhal et al., 2022).

The correlation analysis was conducted to determine prominent meteorological indicators and waste factors according to their higher correlation indices and their rank. Subsequently, weak or non-correlated parameters were isolated. Such an approach will be beneficial for the concise treatment of the best information for modelling efforts. The correlation index between each pair of data was computed with a correlation matrix.

Table 8.1 A summary of conditioned and combined datasets for training and testing investigations in the biogas rate prediction case study.

Parameters	***Data Sources***	***No. of Datapoints***
Dependent variable		
Biogas Produced *(BP)*	OGD (2015-2021)	118
Independent variables		
Temperature *(T)*	MERRA-2 (2015-2021)	118
Pressure *(P)*	MERRA-2 (2015-2021)	118
Relative Humidity *(RH)*	MERRA-2 (2015-2021)	118
Precipitation *(Pr)*	MERRA-2 (2015-2021)	118
Organic Waste *(OW)* [Calculated]	CPCB (2015-2021)	118

Fig. 8.1 summarizes the heat map corresponding to the Pearson correlations heat map for the properties among BP, climatic parameters and OW. Among all parameters, the T had the highest correlation coefficient (0.81) and is apparent in the first column of the heatmap. Fig. 8.2 illustrates this positive correlation between T and BP and thereby inferred that the BP rate is significantly influenced by T. The upward trends in OW-BP (0.65) and RH-BP (0.43) are consistent with the conclusions drawn from the pertinent prior art (Droke, 2001; Genedy and Ogejo, 2021). Also, correlation coefficients for the pairs RH-BP (0.26), P-BP (0.17) and Pr-BP (0.091), have all affirmed positive but weak correlation. Accordingly, targeting the subduing of the associated multi-collinearity issues, these model pairs were omitted during the modelling effort.

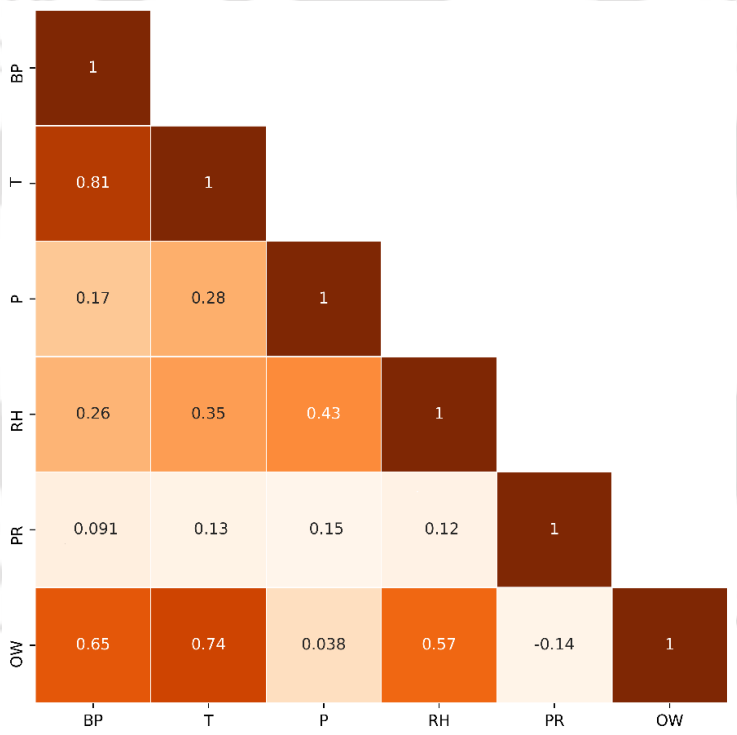
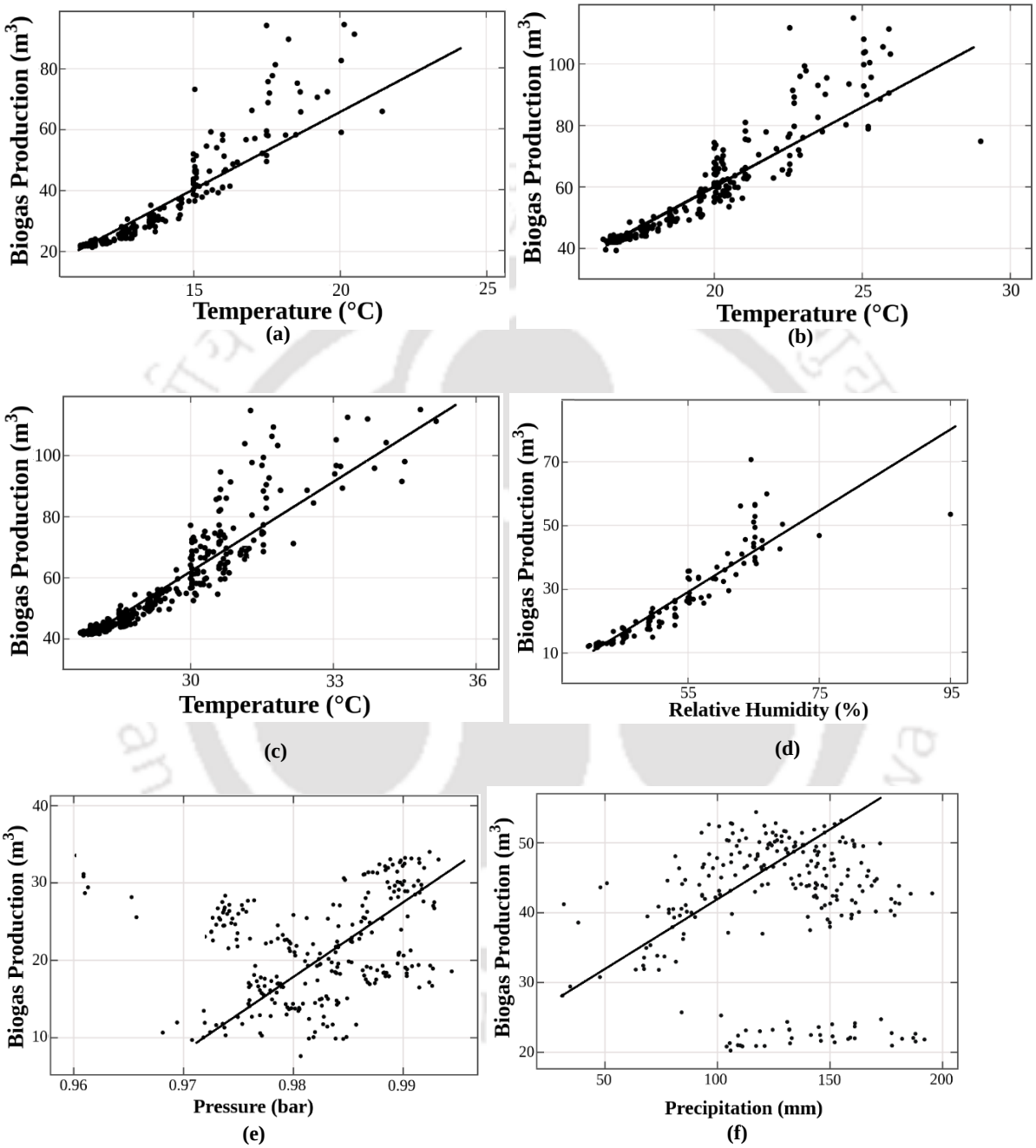


Fig. 8.1 Heatmap representing Pearson correlation analysis findings in the biogas rate prediction case study

Scattered plots of T (showing the seasonal aspects), P, RH, Pr and OW with respect to BP rate further illustrate such pairing effect (Fig. 8.2 (a-g)). BP prediction is influenced by temperature, a critical factor in anaerobic digestion. Mesophilic conditions (15-25°C) promote moderate microbial activity, ensuring robust and consistent biogas yields. In the temperature range of 20-30°C, performance efficiency gets enhanced due to accelerated microbial activity. Transitioning to the thermophilic phase (30-36°C) further boosts microbial activity. This translates into notably higher BP and enhanced methane content.

Careful management is crucial at the upper temperature limit for the prevention of inhibition. Fig. 8.2 (a), (b) and (c) demonstrate distinct trends across these temperature ranges, highlighting the temperature-dependent nature of BP rate.



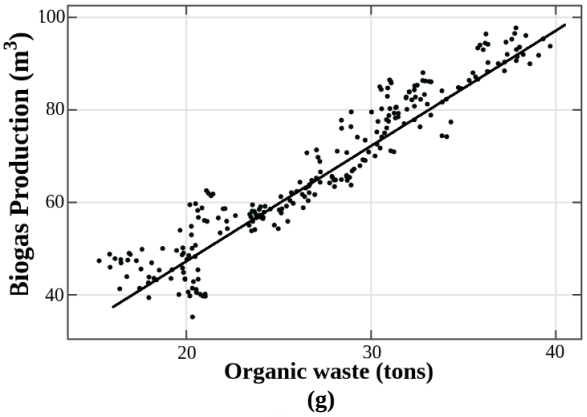


Fig. 8.2 (a-g) Correlation analysis findings depicting sensitive dependence of BP rate on various independent variables

Fig. 8.2 (d - f) depict the correlation analysis of the BP rate with respect to relative humidity (RH), pressure, precipitation and OW respectively. Among these, the interplay of RH, atmospheric pressure, and precipitation significantly and sensitively influences BP rate in AD systems (Wiśniewska et al., 2020). The optimal RH is crucial to sustain microbial activity. It directly influences enzymatic processes and ensures efficient breakage of the organic matter. Deviations in the optimal humidity range can impede microbial function, and thereby affect overall BP rate efficiency. Atmospheric pressure variations influence gas solubility, particularly for methane and carbon dioxide and thereby sensitively influences gas production kinetics in anaerobic digesters. Precipitation events introduce complexities by altering feedstock characteristics, and thereby necessitate upon the careful management strategies for optimal microbial digestion in BP (Ellacuriaga et al., 2021).

8.3 Modelling and Predictive Performances

8.3.1 Support Vector Regression

The SVR model was trained with *pr* and ε parameter values of 1 and 0.2 respectively. A train score value of 0.77 and a test score value of 0.74 were confirmed by the SVR. As a result, the model resulted in appropriate RMSE (103.92), SSE (124.41), and R^2 (0.68) values. Thus, despite affirming acceptable R^2 value, the model was unable to provide an accurate prediction (Fig. 8.3 (a)). Such under performance of the algorithm is due to large data samples and more independent variables.

8.3.2 *k*- Nearest Neighbour

With respective train (0.80) and test (0.79) scores, the *k*NN can be inferred to be an average model. The *k*NN was trained for *k*=2. However, the accuracy of the model decreased with increasing *k*. Also, the model affirmed average RMSE (53.26), SSE (61.11) and R^2 (0.76) values. A straight line can be fitted to the predicted model and actual values to infer that their correlation was not influenced by outliers/noise in the data (Fig. 8.3 (b)). However, certain limitations do exist for the utilization of the prominent limitation in terms of the non-generalization for other datasets. In due course of the fitness, the aggregated *k*-nearest values of the training dataset was unable to explain the correlation between the dependent and independent parameters and also underlying pattern of the data distribution. Thus, the *k*NN model will become inaccurate with the inclusion of new testing datasets from variegated *in-situ* measurements with respect to climatic data and altered combinations of P, RH and Pr. Another important issue is that almost no training occurred for the *k*NN. Most operations occurred during the computationally expensive testing phase. For this reason, even though *k*NN performed well, it was not suitable for the chosen problem and case study.

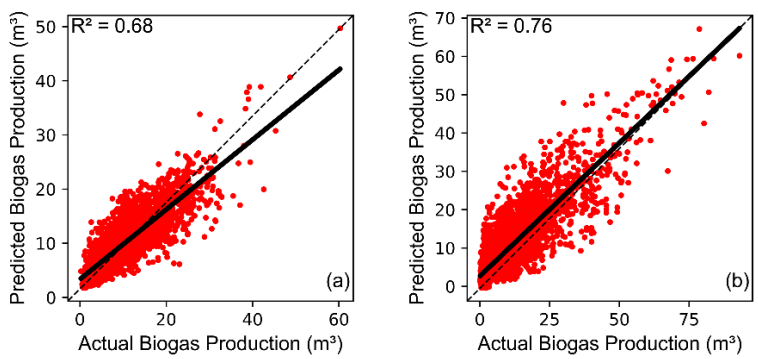
8.3.3 Multilayer Perceptron

The MLP resulted in consistently better train and test scores of 0.87 and 0.84 respectively. Three hidden layers (each with a size of five), the tanh activation function, and a few arbitrary parameters were used to train the MLP. The RMSE (2.76) and SSE (18.43) values have been better but not the best. The R^2 (0.94) value has been significantly better than that obtained for other considered models. The predicted data has been susceptible to outliers. This can be confirmed with the scatter plot (Fig. 8.3 (c)). As a regressively training model, the MLP can be extremely powerful, and can accordingly assure the production of even more accurate models. However, due to computational competency, the model was unable to get trained with additional hidden and deep layers.

8.3.4 Tree-based Models

Both GB and RF produced quality results in terms of the respective test (0.85 and 0.73) and train scores (0.88 and 0.77). The maximum tree depth for the RF and GB was 15 and 5 respectively and were effective to resolve the overfitting problem. Fig. 8.3 (d) and (e) respectively depict the scatter charts for the tree models. Compared to the RF, the data points for GB are closer to the fit line. The RMSE (3.23), SSE (20.52) and R^2 (0.92) of GB have been better in comparison to the RMSE (22.14), SSE (48.92) and

R^2 (0.83) of RF. This is probably due to the reason that the GB upon being trained for one tree at a time, could correct the mistakes of the previous trees. Hence, a new tree with better data could be generated. On the other hand, the trees in the RF have been constructed independently. An RF tree can choose its outputs in any order as each tree is independent. Thereafter, the individual predictions have been combined into a single prediction, which is either the average value in regression issues or the majority class in classification problems. On the contrary, the run for the GB trees is henceforth set and alteration of the sequence is not possible. Hence, it can be concluded that the GB provided the best results for all parameters than RF. Moreover, at times, the RF repeated the predicted data points that exist closer to one another. However, this was not the case for the GB model. A comparative summary of R^2, RMSE, MAE, IOA and computation time for the models have been presented in Table 8.2. The overall evaluation affirmed that the MLP is the best-suited ML model among the five considered ML algorithms. Table 8.2 also reveals a comparative summary of each model in terms of their computation time (min). Due to lesser complexity of the algorithms, the *k*NN and RF had a PO time lesser than 0.50 s (0.0083 min). However, for the GB, the PO time has been about 1.43 min. Models with even higher complexity such as SVR and MLP demanded a maximum computation time of 2.673 and 3.781 min respectively.



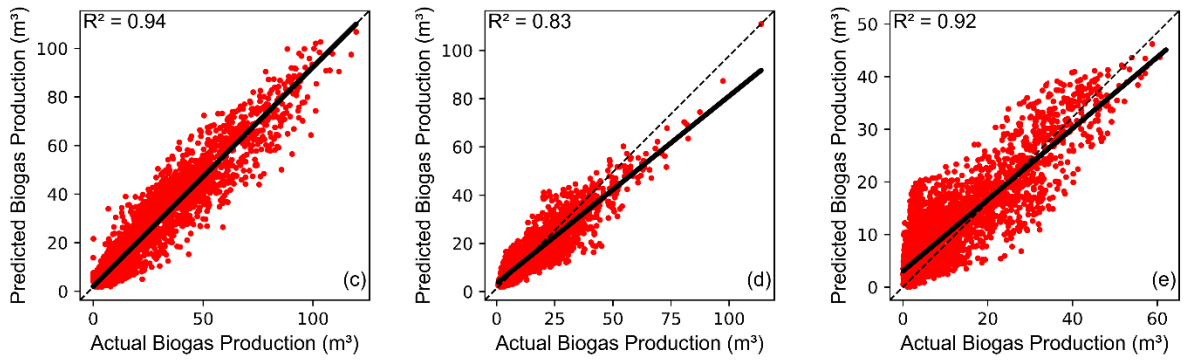


Fig. 8.3 Parity plots depicting BP rate prediction efficacy of **(a)** SVR **(b)** *k*NN **(c)** MLP **(d)** RF and **(e)** GB models

Table 8.2 (a) R^2, **(b)** RMSE, **(c)** MAE, **(d)** IoA and **(e)** computational time of alternate ML models in the BP rate prediction case study.

Models	R^2	RMSE	MAE	IoA	Computation time (min)
SVR	0.68	103.92	93.22	0.73	2.673
kNN	0.76	61.11	42.81	0.78	0.00683
RF	0.83	22.14	15.86	0.81	0.00816
GB	0.92	3.23	2.61	0.89	1.43
MLP	0.94	2.76	1.93	0.84	3.781

The accuracy and loss graphs of MLP and GB, as well as their performance, have been shown in Fig. 8.4 (a-d). In these plots, the y-axis denotes the accuracies and losses for both models and the x-axis represent the training epoch (sample size). The x-axis is plotted as a percentage or the overall dataset's training cycle count. When accuracy graph is examined more closely, it becomes clear that for some samples, the initial curve profiles were steep and accounted for 92% and 94% accuracies for GB and MLP respectively. Also, with increasing percentage sample size, both accuracy curves depict an upward trend. In actuality, it can be seen that the there is a significant growth scenario during the first through third sample sizes. Thereafter, the growth rate steadily reduced and reached a stagnant profile. Similarly, loss measures refer to the model error and hence poor performance status of the model. Thereby, the figures convey that the loss (error) is eventually becoming lower and hence better performance has been assured with increasing epoch. Despite indicating the existence of few peaks and valleys in the longer time frame, the reduction of loss in due course of time affirms that both GB and MLP models learnt effectively and performed well.

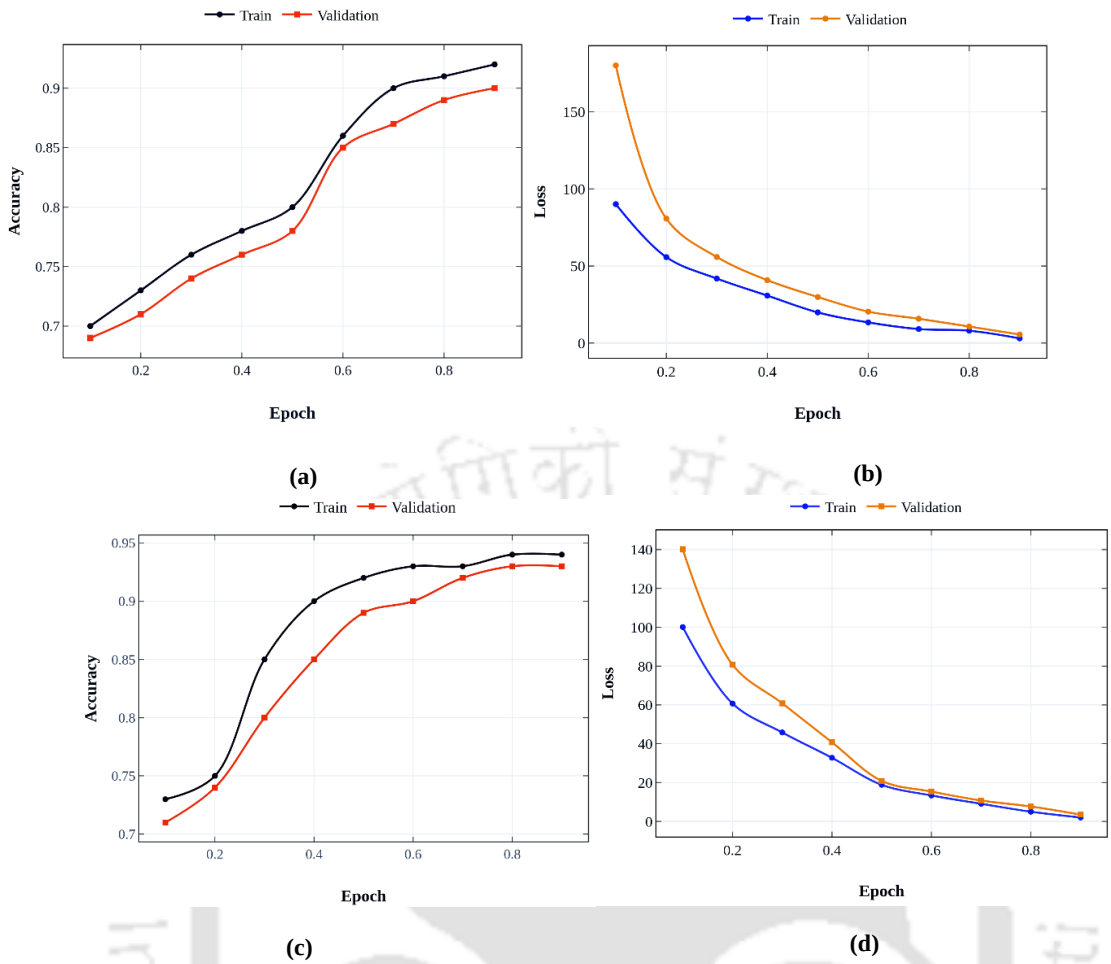


Fig. 8.4 Accuracy and loss curves for **(a-b)** GB algorithm and **(c-d)** MLP models in the BP rate prediction case study

8.4 Long-term Forecasting with ARIMA and MA Models

In this study, ARIMA methodology was deployed for the forecasting of BP with statistical models. Noise reduction was targeted along with easy removal of the outliers. The strategy is based on the gathering of long-range correlations. The correlation outcome of the differential sequence has been analysed in order to establish the stationarity of the predicted trends (Table 8.3). The autocorrelation generates a short-term correlation in the sequence, and was eventually subjected to the white noise test. The sequence used to fit the ARIMA model could not be classified as white noise due to the result possessing a 0.05 significance level and lagged p-value statistics being less than 0.0001. The ARMA (p, q) models with a moving average delay order and autocorrelation delay order have been both less than or equal to 7. To forecast the BP, the ARIMA (1, 1, 0) model was employed for the next ten years, and the predicted and the forecasted results for the year 2021 have been correlated for each model (Fig.

8.5(a)). Accordingly, it can be observed that the ARIMA-MLP confirmed better result followed by ARIMA-GB. As being depicted in the graphs, the future BP rate trends for Guwahati city have been forecasted for the year range of 2022-2028 and with the predicted past data. The validation results of forecasted BP with the observed (actual, 70 m^3) BP data for the year 2021 have been summarized in Table 8.4.

Table 8.3 Autocorrelation table findings for a lag value (K) of 17 in the BP rate forecasting case study.

Lag	Covariance	Correlation	Standard error
0	62.2817	1.0000	0
1	46.0912	0.62102	0.15762
2	33.1341	0.51792	0.18022
3	25.5102	0.34281	0.20441
4	20.6629	0.21991	0.22077
5	18.1361	0.23903	0.23831
6	14.9211	0.27061	0.24471
7	11.1359	0.24518	0.25802
8	9.3653	0.19213	0.26994
9	3.2972	0.06718	0.26468
10	-0.3123	-.009213	0.26321
11	-1.1562	-.03241	0.27175
12	-3.7770	-.06118	0.27331
13	-6.4863	-.05882	0.27540
14	-10.3019	-.12516	0.27752
15	-22.3291	-.25792	0.28116
16	-26.4936	-.21552	0.28872

Table 8.4 Data summary elucidating upon the forecasting efficacy of alternate ML models in the BP rate case study.

	Forecast for variable BP			
Models	***Forecasts***	***Standard Error***	***95% Confidence Limits***	
MLP	71319	5.1152	71217	71992
GB	72009	9.4123	71988	72561
RF	72174	10.9438	72011	73190
***k*NN**	68042	15.2242	67830	68452
SVR	57332	21.8331	56109	57901

Another forecasting approach namely MA methodology was also considered for the forecasting of BP with statistical models. Such a strategy was considered to achieve noise reduction and easy outlier removal. For the forecasting and analysis of time-series data, the MA technique was used as the statistical tool. For this, the strategy has been based on the gathering of long-range correlations (Chynoweth et al., 1994; Wang et al., 2021). The forecasted BP rate for Guwahati city and in the year range of 2022-2028 for considered models has been shown in Fig. 8.5 (b). Three MA techniques, EMA,

SMA, and WMA, were considered to achieve significant insights and useful evaluation of the BP rate patterns with the studied models. In the year range of 2022-2028, the BP for EMA-MLP and EMA-GB have been forecasted to be 64913-62576 m^3 and 58334-59713 m^3, respectively. The RF (59711-58452 m^3), SVR (43662-53902 m^3) and *k*NN (33770-43612 m^3), performed with least forecasted accuracy among all models. For all models, the EMA was found to be the best predictor of the BP rate, followed by the WMA, and the SMA. Further, along with the correlation coefficient, the model performances were accurately estimated with the MAE and RMSE tests.

The model errors have been summarised in Table 8.5(a) for the year 2021 and in conjunction with the available data (70 m^3 BP rate). In comparison to the rest of the models, the MLP and GB exhibited lowest forecasting errors (7.2-11.6 %) and (16.7-17.6 %) respectively. *k*NN (51.7-56.5 %) and SVR (37.6-43.0 %) indicated higher error percentage and the RF confirmed nearly identical error numbers with respect to GB. Among all, the MLP was the best to predict MSWG rates. This has been also confirmed with its R^2 values.

During forecasting studies, according to their rank, the meteorological and OW parameters were first integrated one by one into the model. However, parameters such as T, OW and RH have been found to significantly enhance model performance. Table 8.5(b) highlights the models' predictive performance (in terms of RMSE and R^2) for their optimal performance with respect to the three MA techniques. The RMSE and R^2 values for the best performing MLP and GB models were 0.90, 3.76; 0.88, 6.93 and 0.86, 7.942; 0.86, 11.5312 for EMA and ARIMA techniques respectively. The trends were comparable to those realized in the training scenarios.

Table 8.5(a) Forecasting error summary for alternate ML models in the BP rate case study.

Approach	*MLP forecasted (m^3)*	*MLP model error (%)*	*GB forecasted (m^3)*	*GB model error (%)*	*RF forecasted (m^3)*	*RF model error (%)*	*SVR forecasted (m^3)*	*SVR model error (%)*	*kNN forecasted (m^3)*	*kNN model error (%)*
EMA	64900	7.2	58300	16.7	59700	14.7	43660	37.6	33770	51.7
WMA	62720	10.4	58620	16.2	57840	17.3	41760	40.3	32970	52.9
SMA	61840	11.6	57650	17.6	55430	20.8	39850	43.0	30450	56.5

Table 8.5(b) RMSE and R^2 values summary for the BP rate forecasting case study.

Method	MLP		GB		RF		SVR		*k*NN	
	R^2	*RMSE*	R^2	*RMSE*	R^2	*RMSE*	R^2	*RMSE*	R^2	*RMSE*
ARIMA	0.90	3.76	0.86	7.94	0.81	25.80	0.72	55.97	0.64	82.97
EMA	0.88	6.93	0.86	11.53	0.78	36.99	0.70	72.80	0.61	90.33
SMA	0.87	10.81	0.85	18.62	0.77	52.77	0.67	88.44	0.60	101.41
WMA	0.83	13.20	0.80	24.66	0.74	78.54	0.66	90.67	0.58	121.82

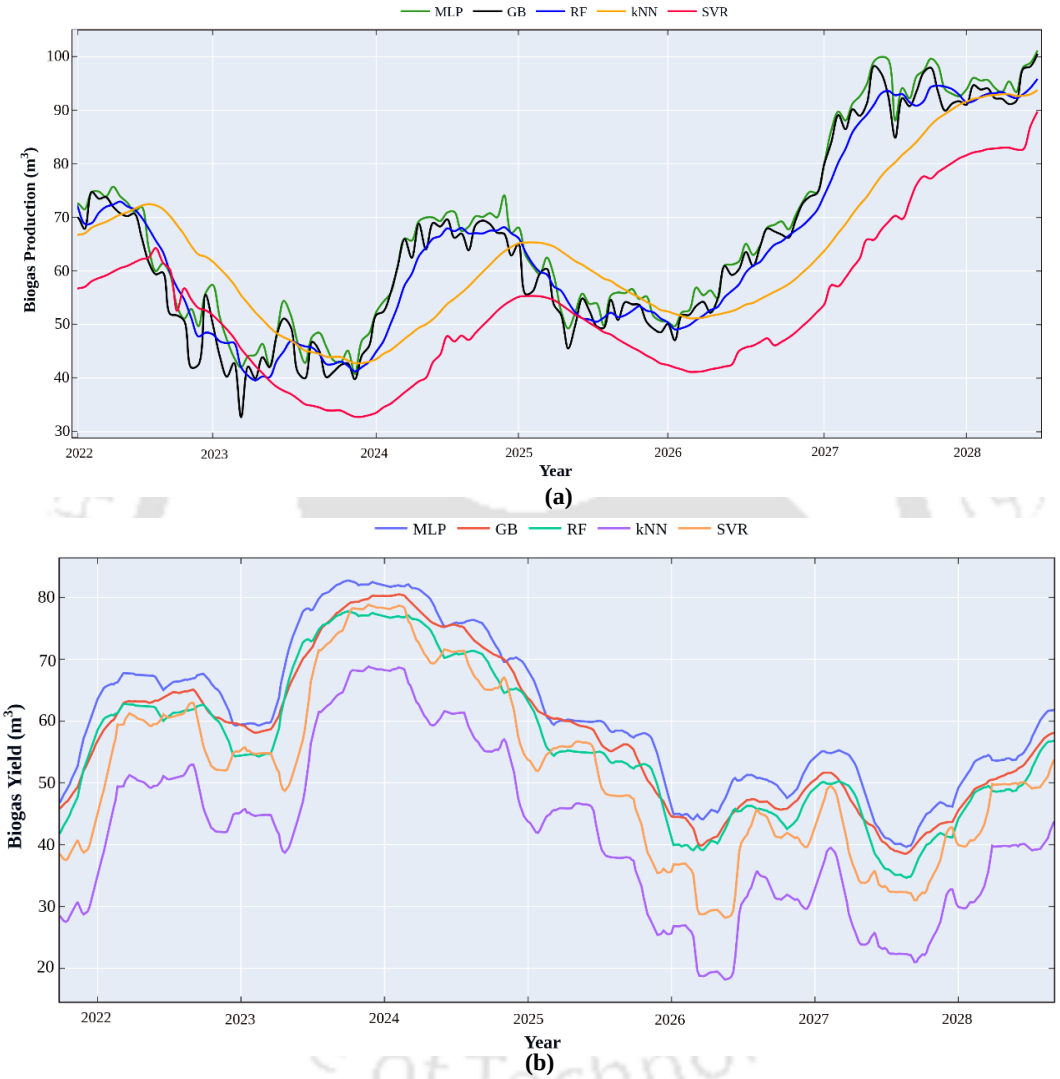


Fig. 8.5 Graphs depicting forecasted **(a)** BP rate and **(b)** Biogas yield rate trends in the year range of 2021-2030

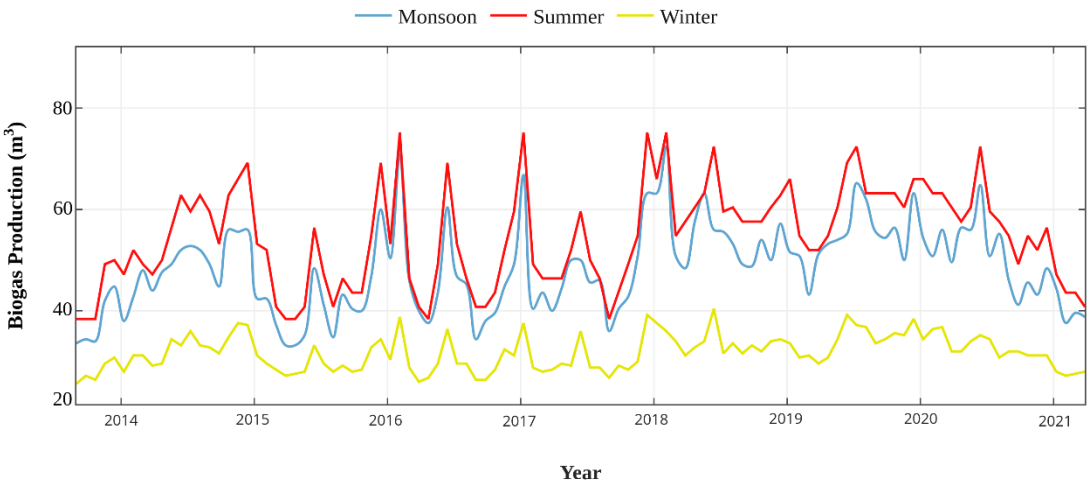


Fig. 8.6 BP rate variation for Guwahati city in the 2012-2022 year range for Guwahati city during Monsoon, Summer and Winter seasons.

Table 8.6 Seasonal parameters (T, RH, P and Pr) summary in the BP rate case study.

Parameters	***Seasons***		
	Summer	*Monsoon*	*Winter*
T (°C)	30.0 ± 3	25.0 ± 5	20.0 ± 5
RH (%)	76.0 ± 2	82.0 ± 4	69.0 ± 3
P (bar)	0.96 ± 0.02	0.68 ± 0.04	0.43 ± 0.02
Pr (mm)	244 ± 4	549 ± 6	10 ± 0.18

8.5 ARIMA based Seasonal Variation of the BP Rate

T critically influenced BP rate and the released heat as well as influenced the process performance. The literature also corroborated with the fact that the T alteration confirmed to the patterning of the BP rate (De Clercq et al., 2019). The most favourable conditions for biogas generation have been in the summer season (Apr-Jun) that witnessed a maximum T (30 ± 3°C) and higher RH (80%). Thereby, the highest BP rates can be achieved. On the contrary, the monsoon season (Jul-Sep) is challenging for the BP due to higher MC of air enhancing biogas generation time. However, during winter season (Dec-Feb), with a T (max) of 20°C and RH below 17%, a marginal influence has been apparent for the BP rate. Thus, with the progressive year, the BP rate has been forecasted (Fig. 8.6) and the seasonal parametric variations have been presented in Table 8.6.

8.6 Generic Nature of the Modelling Framework

Table 8.7 summarizes the findings in comparison with the best available prior art. The high-end ML algorithms yielded highly accurate findings and these are similar to the previously reported findings (Nkuna et al., 2021; Clarke et al, 2004; Blasius et al., 2020). Lower-complexity ML models can also

outperform higher-complexity models, but with a certain concern towards their accuracy. Lesser historical data records, other approaches and procedures linked with hyper parameter optimization (HPO), and differences in the training to testing split ratio have all been liable for the algorithms' poor performance. As a result, good performance of the ML algorithms can be achieved towards the effective prediction of the BP rate by considering all these factors in modelling efforts and incorporating the mentioned techniques for the mitigation of overfitness problems.

Table 8.7 A summary of the best findings of this work and prior art in the BP rate prediction and forecasting case studies.

Application	Authors	Approach	Data Size	Input parameters	Performance Evaluation	Inference
Biogas Prediction	This study	MLP (feedforward ANN) GB RF *k*NN SVR	708 (Historical data: 2015-2021)	T, RH, P, Pr and OW	RMSE: 2.76, R^2: 0.94 RMSE: 3.23, R^2: 0.92 RMSE: 22.14, R^2: 0.83 RMSE: 61.11, R^2: 0.76 RMSE: 103.92, R^2: 0.68 Model accuracy: 90%	T and OW affirmed the influencing parameters. MLP performed the best with accuracy of 94%. ARIMA-MLP affirmed as best suited for forecasting.
	Cinar et al. (2021)	*k*NN, xGBoost, RF, SVR, LR	242	Optimal temperature	SVM performed better with R^2 (0.93) and testing accuracy of 76 %	The model had limited data
	Nkuna et al. (2021)	GB, RF, NN	25,516 rows of data (1990-2020)	RH and Pr	RF outperformed with RMSE (0.286), MAE (0.092) and R^2 (0.99)	Lack of comparison experiments and high complexity
	Clarke et al. (2004)	GB, RF, SVR, Ridge, Lasso, *k*NN	1398	OW	For the best performance in making predictions, a tree-based regression model was chosen with RMSE (247.31) and R^2 (0.72)	The models lacked hyper tuning and hence may result in biasedness.

8.7 Summary

The capability of predictive analytical techniques, including ML, to anticipate the best operating conditions for AD was discussed in the chapter. Due to higher complexities in experimentations, and varying degree of accuracies of biological systems, numerous ML algorithms with accuracies were

studied in the prior art. As a result, for each system ad environment, it is ideal to develop and build a separate model. This research also created a framework for the development of an ML model, data collection and evaluation of the most competent ML model. If necessary, the generic methodology highlighted in the chapter can be revised and customized for other systems. Historical datasets were collected and trained in order to build an ML model and were applied in different regression and classification models. With an RMSE value of 2.76, the MLP had the best accuracy among the models. This was followed by GB (RMSE of 3.23). Henceforth, the model exhibited notable accuracy for generated biogas prediction. The level of nonlinearity in the models can be used to explain the lesser accuracy of other models. Since the models were developed with historical data under several environmental factors, it was simpler to maintain a consistent combination of data and associated parameters. The operational parameters of an industrial-scale biogas plant might even illustrate daily base variations in a real-world application. The system's dynamic behaviour outcomes were more infrequent during model upgradation phase. The best tuning parameters were obtained with the Python programming environment's meta-estimator Grid-search function after reaching the highest accuracy with the MLP model. Furthermore, a comparison of the forecasting techniques affirmed that ARIMA-MLP resulted in the best forecasting accuracy with good RMSE (3.76) and R^2 (0.90) values.

Further, the data analysis suggests that T (0.81) and OW (0.65) may account for the majority of the variation in BP as they have the best R^2. However, RH (0.26), P (0.17) and Pr (0.091) did not exhibit much significance in BP. This study also contributes to the efficacy of alternate estimating methods for BP rate determination. In comparison, the MLP produced promising outcomes in BP prediction followed by GB. The performance matrices of ARIMA-GB and EMA-GB affirmed RMSE values of 7.94 and 11.53, respectively and the overall model accuracy was found to be 84%. Seasonal analysis also confirms that the favourable conditions for biogas generation have been in the summer season (Apr-Jun with T_{max} value of about 30 ± 3 °C and RH value above 80%). The forecasting approach can be utilized as a model for other locations with similar developmental and environmental challenges.

Chapter 9
AD Calculator and Machine learning Algorithms based Optimality of Biogas Production Rate in Indian Biogas Plants

Chapter 9

Optimizing Biogas Production: A Novel Hybrid Approach Using Anaerobic Digestion Calculator and Machine Learning Techniques on Indian Biogas Plant

The chapter addresses BP optimization methodologies for Indian biogas plants and with ML algorithms. A novel hybrid approach that integrates AD calculator, ML and hybrid ML techniques has been delineated for its effectiveness to enhance performance efficiency. Section 9.1 provides a comprehensive overview of the research objectives, scope, and the significance of the study. It establishes the context for the subsequent sections by highlighting the need to enhance BP efficiency and maximize the economic and environmental benefits of biogas plants in India. The model's ability to consider temperature (T) and RT for the estimation of the relative volatile solids (RVS) along with prior data trends has been discussed in section 9.2. Further, the subsections 9.2.1 and 9.2.2 elucidate upon a comparative analysis of existing models for biogas generation, drawing insights from relevant case studies in the literature and accordingly address the implementation of the AD calculator and ML techniques for the daily biogas generation rate estimation in the Sonapur farm of the Guwahati city. Section 9.3 addresses the comprehensive comparisons of operational revenues and costs in terms of the NPV and PBP. Thereafter, sub-sections 9.3.1 and 9.3.2 respectively assess upon the efficacy of alternate operating modes in terms of the predictive ability for optimal revenues and costs in biogas plant operations, and forecasting of maximum biogas yield with ML algorithms. Thereafter, section 9.4 presents the significant findings of the study in terms of the optimal BP process plant performance[#].

Overview

The findings associated to the proposed AD calculator and ML-based predictive models as a comprehensive approach for biogas yield estimation, optimization, and economic analysis has been summarized in the chapter. The hybrid ARIMA-MLP model was used with a dataset of 350 samples being collected for one year time frame for the Sonapur AD plant, Guwahati. The results depict that the ARIMA-MLP model outperformed the basic MLP model and SVR models and thereby demonstrated its robust predictive potential. Accordingly, the findings contribute to the efficient, sustainable, and profitable operation of biogas plants, and thereby support India's transition towards a bio-based smart economy. Thereby, the environmental and public health concerns of the with waste management can be effectively addressed and realized.

[#]Published article: Singh, T., & Uppaluri, R. V. (2023). Optimizing biogas production: A novel hybrid approach using anaerobic digestion calculator and machine learning techniques on Indian biogas plant. Clean Technologies and Environmental Policy.

9.1 Introduction

Emphasising on a two-phase design approach, the chapter in the first phase proposes an AD calculator for the evaluation of biogas yield and prediction. The second phase involved historical data-based studies for yield estimation in terms of feedstock characteristics, RT and T. The AD calculator has the potential to serve as a valuable resource for operators seeking to manage their AD units on a daily basis. It facilitates the monitoring of seasonal variations influencing on BP rates, analyzes and optimizes parameters related to domestic biogas plants, and ascertains their effective management under diverse climatic conditions. The optimization process can help to maximize the long-term profitability of the AD unit. The specific novelties of the addressed work in this chapter are (a) a robust model for the prediction of BP rate, and (b) an economic model for the evaluation of the unit's operational revenue and costs. Small-scale AD units usually have limited resources to measure the key parameters such as T, pH, and organic loading rate in real-time. Such a limitation can make it difficult to optimize the performance of the AD unit and may result in suboptimal BP. Also, it is financially not feasible to install additional devices/laboratory equipment. Thereby, restrictions exist to develop and apply sophisticated process models. Based on RT, T, and OW volume, the model proposed in this chapter computes the biogas yield, which represents the quantity of biogas generated per unit mass of feed. In summary, the calculator is tailored to suit the needs of AD units that are commonly employed in India. The location of the site can significantly impact energy prices, product values, and government incentives. To instil confidence in the AD calculator, the prediction performance has been compared with literature data. Thereafter, the tool has been used to evaluate a 250kW AD unit deployed in an Indian farm biogas plant.

9.2 Evaluation of Biogas Generation Model

The BP model necessitates input data in terms of feed flow rates, operating temperature, and specific digester specifications, including orientation (vertical or horizontal) and feed materials. This is generally available for most AD units, and ensures that the model is readily applicable to various sites. Fig. 9.1(a) and (b) demonstrate that the model has the capability to consider T and RT for RVS estimation. The observed trend of enhanced RVS with RT, and up to upper limit of potential biogas yield, aligns with considered prior art data trends (Angelidaki and Ellegaard, 2003;

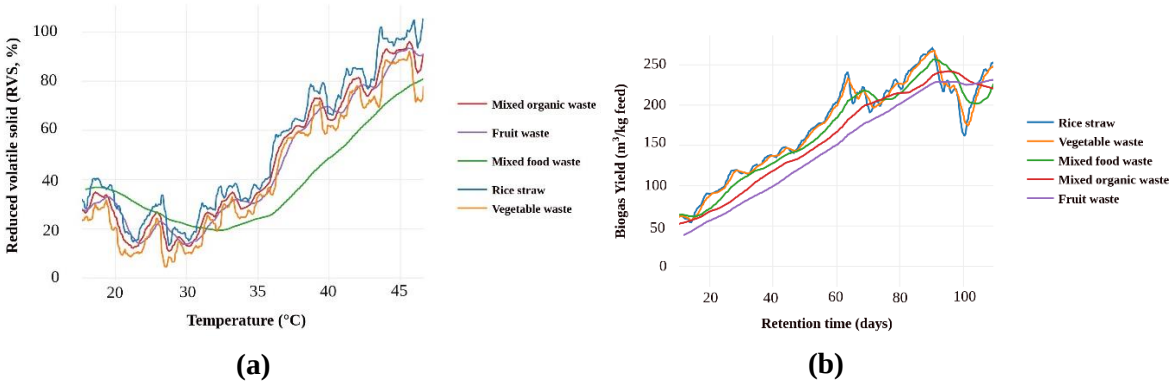
Chen et al., 2021). On-site model calibration necessitates experimental measurements of potential biogas yield for various feed materials, (as depicted in Fig. 9.1(b)) and for alternate feedstocks.

9.2.1 Comparative Analysis of Design Model with Literature Reported Data

A comparative analysis of models involved case studies of AD units up to 3 MW and comparison of the estimated variables from the proposed model with the base empirical models of the studies. The comparison of the proposed model with base empirical models was limited to the evaluation of its ability to provide reasonably accurate preliminary steady-state estimations. This is due to non-calibration status of the proposed model. The results (presented in Table 9.1) convey that the preliminary BP estimations of the proposed model were within ±10% of the reported values in several case studies. The BP estimations generated by the proposed model for the three types of studied digesters (Deenbandhu, Floating drum, and Balloon) were comparable to those reported in the case studies. This affirmed the reliability of the proposed model. The findings suggest that the proposed model is comparable to the case study data, and could be a useful tool for the BP estimation in AD units.

9.2.2 Estimating the Daily Biogas Generation at Sonapur farm AD Unit (Kamrup)

The performance of the proposed model was evaluated with twelve months of daily sample data (360 samples) from the Sonapur AD unit and, for day-to-day biogas yield estimation. Through a comparison of the expected BP with the measured biogas generation, the AD operators could establish performance benchmarks. Fig. 9.2 depicts the feedstocks profiling for a 12-month period, and thereby affirmed significant variation in the mass flow rate. Parity plots were used to compare the BP rate from literature data with that of the proposed model. The uncalibrated model's maximum prediction error for the Balloon digester was found to be 268.8 (Table 9.2). In the comparison study, Deenbandhu digester affirmed better prediction (R^2: 0.94, RMSE: 161.4 and SSE: 1.04) followed by floating drum design (R^2: 0.93, RMSE: 187.5 and SSE: 1.42). Fig. 9.3 depicts the application of the uncalibrated BP model and its performance for the prediction of the BP rate.



(a) **(b)**

Fig. 9.1 (a) Graph depicting Temperature and feedstock type influence on % RVS **(b)** Graph depicting RT and feedstock type influence on biogas yields

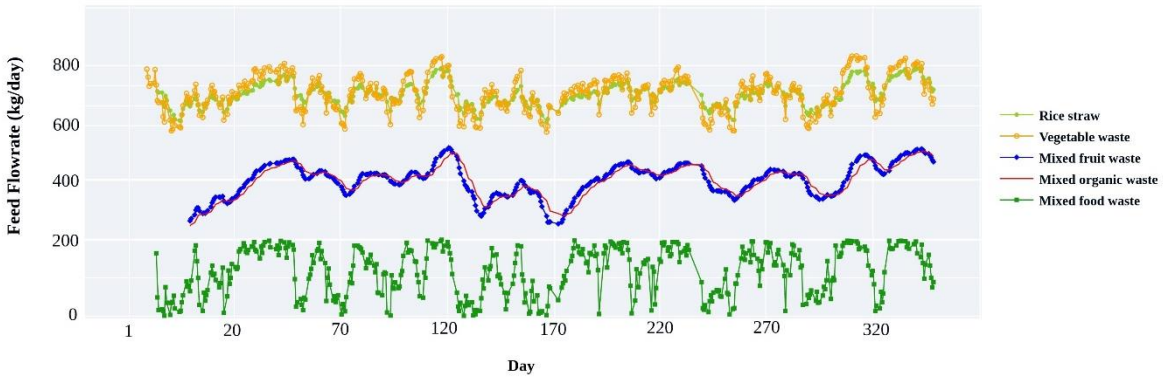


Fig. 9.2 Measured feed flowrates of the Sonapur AD unit for various feedstock cases

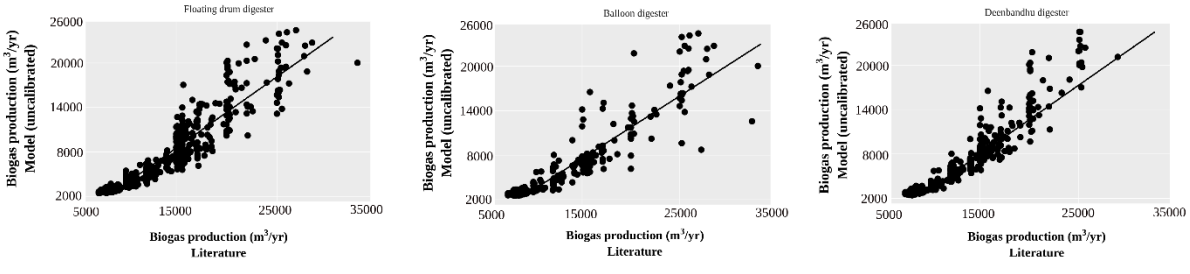


Fig. 9.3 Parity plots for the uncalibrated BP rate model with respect to literature data

Table 9.1 A summary of uncalibrated AD calculator based yield data in the prior art.

Case studies	Location	Feedstock and daily load	T & RT	Yield	Volumes of the reported and estimated study (m^3)	References
499 kW AD unit	UK	Grass silage; 32 t d^{-1}	35 °C 50 days	BGY (m^3 yr^{-1})	Reported: 5370; Calibrated model: [1] 6219 (+16%); [2] 6714 (+27%); [3] 4328 (+19%)	WRAP, 2014
Schmack Biogas	UK	Grass silage: 11 t d^{-1} Maize silage: 12 t d^{-1}	35 °C 55 days	BGY (m^3 yr^{-1})	Reported: 4767; Calibrated model: [1] 4931 (+3%); [2] 5699 (+20%); [3] 4547 (-5%)	The Andersons Centre, 2010
Municipal Corporation	Vijaywada, AP	Vegetable market wastes, Mixed OW; 2 t d^{-1}	35 °C 50 days	BGY (m^3 yr^{-1})	Reported: 8800; Calibrated model: [1] 9600 (+9%); [2] 9200 (+5%); [3] 9000 (+2.2%)	Chanakya et al., 2009
Vensa Biotek Ltd.	Samalkot, A.P.	Rice silage, grass silage; 30 t d^{-1}	35 °C 55 days	BGY (m^3 yr^{-1})	Reported: 24000; Calibrated model: [1] 21000 (-13%); [2] 27000 (+13%); [3] 23000 (-4%)	Jain and Sharma, 2011
SELCO International Ltd. Pvt. Ltd.	Hyderabad	Food waste: Potatoes, Waste from Sugar Cane, Groundnut Shells and Castor & Cotton; 60 t d^{-1}	53 °C 12 days	BGY (m^3 yr^{-1})	Reported: 36000; Calibrated model: [1] 39000 (+8%); [2] 41000 (+14%); [3] 37000 (+3%)	Jain and Sharma, 2011
Mappilai Sago Factory	Salem, T.N.	Potato (tapioca) waste; 1050 kg d^{-1}	39 °C 45 days	BGY (m^3 yr^{-1})	Reported: 9000; Calibrated model: [1] 9500 (+6%); [2] 9200 (+2%); [3] 8400 (-7%)	Rajendran et al., 2011
Samagra Agro	Kanpur, U.P.	Fruit & Vegetable Market waste; 2000 kg d^{-1}	38 °C 26 days	BGY (m^3 yr^{-1})	Reported: 12000; Calibrated model: [1] 15000 (+25%); [2] 13000 (+8%); [3] 11000 (-8%)	Jain and Sharma, 2011
Tirupati Starch and Chemicals Ltd.	Dhar, M.P.	Maize Processing Effluent; 43 t d^{-1}	40 °C 28 days	BGY (m^3 yr^{-1})	Reported: 25000; Calibrated model: [1] 27000 (+8%); [2] 28000 (+12%); [3] 21000 (-16%)	MNRE, 2017
Solapur Bio-energy Systems Pvt. Ltd.	Solapur, Maharashtra	Thermophilic AD of OW through DRYADTM Technology; 62 t d^{-1}	37.1 °C 27 days	BGY (m^3 yr^{-1})	Reported: 38000; Calibrated model: [1] 42000 (+11%); [2] 40000 (+5%); [3] 35000 (-8%)	Kulkarni and Ghanegaonkar, 2019
ARTI, Pune	Pilot plant at ARTI office,Pune	Kitchen waste, 1 – 2 kg d^{-1}	32 °C 10 days	BGY (m^3 yr^{-1})	Reported: 88; Calibrated model: [1] 94 (+7%); [2] 95 (+8%); [3] 90 (+2%)	Patil and Gadiwad, 2015

ARTI, Pune	Home of ARTI employee	Kitchen waste, 1.5 kg d^{-1}	33 °C 11 days	BGY (m^3 yr^{-1})	Reported: 73; Calibrated model: [1] 82 (+12%); [2] 79 (+8%); [3] 77 (+5%)	Patil and Gadiwad, 2015
BARC, Mumbai	BARC Campus, Mumbai	Segregated household waste and food waste from canteens; 1 – 1.2 t d^{-1}	30 °C 7 days	BGY (m^3 yr^{-1})	Reported: 9700; Calibrated model: [1] 9800 (+1%); [2] 10000 (+3%); [3] 9300 (-4%)	Voegeli and Zurbrügg, 2008
BARC, Mumbai	Hiranandani, Thane	Segregated OW from HH; 1.5 – 3 t d^{-1}	31 °C 12 days	BGY (m^3 yr^{-1})	Reported: 11000; Calibrated model: [1] 14000 (+27%); [2] 12000 (+9%); [3] 9900 (-1%)	Voegeli and Zurbrügg, 2008
BIOTECH, Trivandrum	PMG Office (Postal Live Insurance), Trivandrum	Kitchen waste; 30 kg d^{-1}	32 °C 20 days	BGY (m^3 yr^{-1})	Reported: 1460; Calibrated model: [1] 1700 (+16%); [2] 1900 (+30%); [3] 1600 (+10%)	Jha et al., 2016
BIOTECH, Trivandrum	Kottarakara, Grama Panchayath, Kollam District	Market waste (vegetables and fish); Initial phase planned for 250 kg d^{-1}	33 °C 24 days	BGY (m^3 yr^{-1})	Reported: 2500; Calibrated model: [1] 3000 (+20%); [2] 2900 (+16%); [3] 2400 (-4%)	Jha et al., 2016
BIOTECH, Trivandrum	Sreekariyam, Grama Panchayath, Trivandrum District	Market waste (vegetables and fish); 100 kg d^{-1}	33 °C 20 days	BGY (m^3 yr^{-1})	Reported: 1800; Calibrated model: [1] 2300 (+28%); [2] 2500 (+39%); [3] 2100 (+14%)	Jha et al., 2016
MAILHEM, Pune	Magarpatta City, Pune	Mixed household waste (sorted at plant); 2 t d^{-1}	33 °C 34 days	BGY (m^3 yr^{-1})	Reported: 11000; Calibrated model: [1] 10000 (-9%); [2] 13000 (+18%); [3] 7000 (-4%)	Voegeli and Zurbrügg, 2008

Abbreviations: Proposed model: [1] = Deenbandhu digester [2] = Floating drum and [3] = Balloon digester; BGY: Biogas Yield

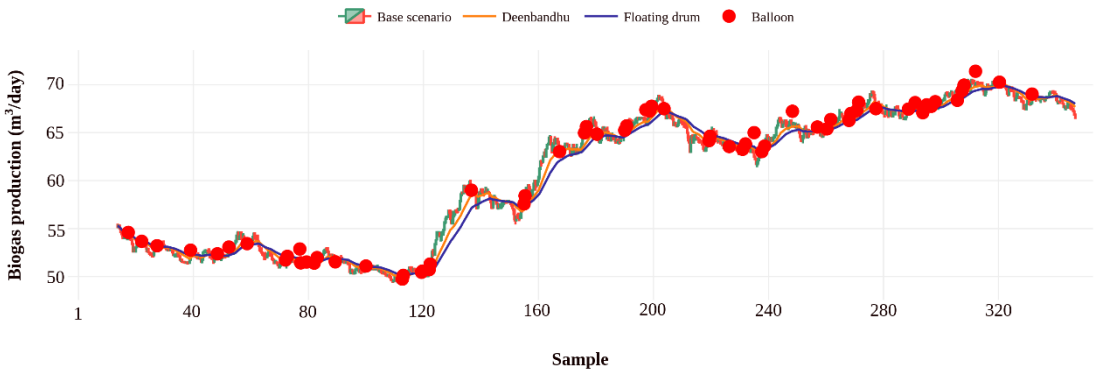


Fig. 9.4 Graph depicting comparative trends of measured and calibrated model-based BP rate

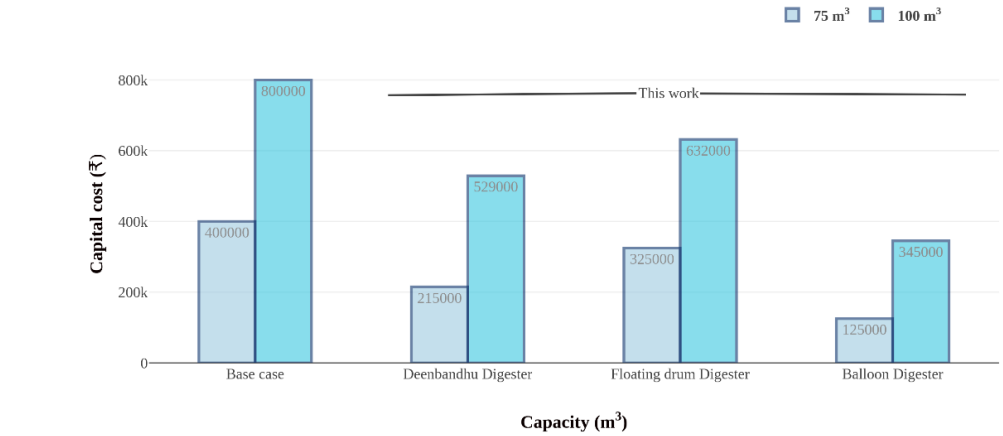
Fig. 9.4 suggested that the floating drum and Deenbandhu digesters exhibited higher biogas yield in comparison with other digesters. The proposed model generally underestimated BP in comparison to the actual production. This highlights the significance of model calibration for daily BP estimations. The estimated BP for the balloon digester had notable discrepancies that were not observed in the measured values. These discrepancies were due to significant variations in the feed flow rates. Simple calibration of the model could improve the evaluation of potential biogas yield. The dataset consisting of 365 days of samples was used for the calibration the potential biogas yields of the feed materials. Subsequently, the calibrated model was deployed for estimation of the BP rate with the same dataset.

9.3 Associated Cost and Revenues

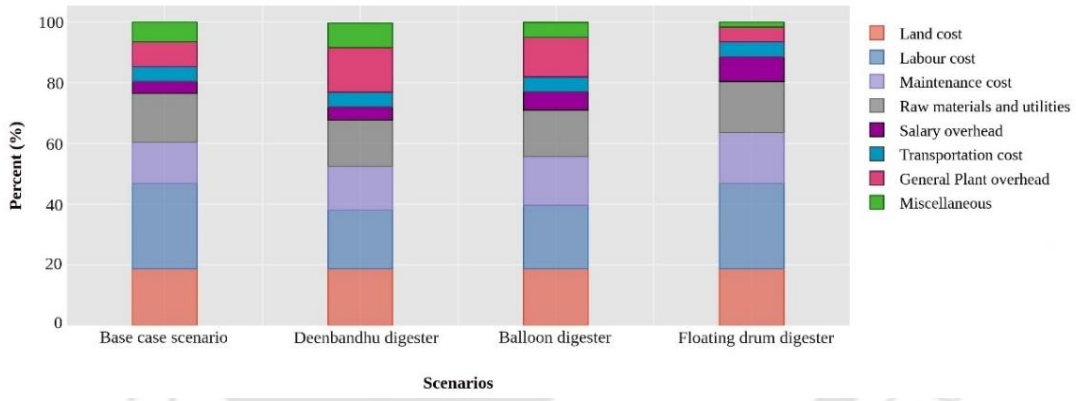
The financial data of AD unit operations is not widely available due to concerns from owners. This resulted in limited reports and evaluations. The operational revenue and cost comparison is therefore restricted. In 2021, a floating dome AD unit was reported to operate at 35°C with a 250 kW capacity. It was fed with 1 ton day^{-1} bio-waste and 3 tons day^{-1} rice silage for 90 days RT. The farm benefited from reduced electricity grid dependency, and with 75% of revenue being derived from government incentives. Nevertheless, this situation also emphasizes the financial difficulties encountered by AD technology due to the gradual reduction of incentives. Fig. 9.5 (a) compares the CAPEX to capacity (m^3) data for the present AD unit and proposed digesters. Among all, the Deenbandhu digester was feasible. Despite this, the CAPEX of the AD unit remained high in comparison to typical biogas installations. Fig. 9.5(b) depicts AD unit plant expenses that fluctuated with raw materials, utilities, and overhead salary. However, other expenses remained relatively unchanged during plant upgradation.

Table 9.2 Performance indices of the digesters.

Digesters	R^2	RMSE	SSE	MAE
Floating drum	0.93	187.5	1.42	1.52
Fixed dome (Deenbandhu)	0.94	161.4	1.04	1.64
Balloon digester	0.89	268.8	4.53	1.81



(a)



(b)

The direct sale of biogas from smaller digesters (60 m^3 day^{-1}) may not be commercially viable as the Indian market for the biogas is currently dominated by large-scale producers. This is obviously due to the economies of scale. For small scale systems, the production volume is relatively low and thereby makes it difficult to justify the costs of infrastructure and logistics for purification, transportation, and marketing. The economic evaluation case study was conducted for the biogas prices of 0.25, 0.50, and 0.75 \$ m^{-3} and with the base price of 0.25 \$ m^{-3} that aligns with the current biogas selling price index. Notably, the chosen base price of 0.25 \$ m^{-3} was derived by considering that the information that 240 cubic meters of the biogas equates to 109 cylinders of compressed biogas (CBG) (this data statement deduced from the work of Prof. V. K. Vijay of IIT Delhi and from his webpage

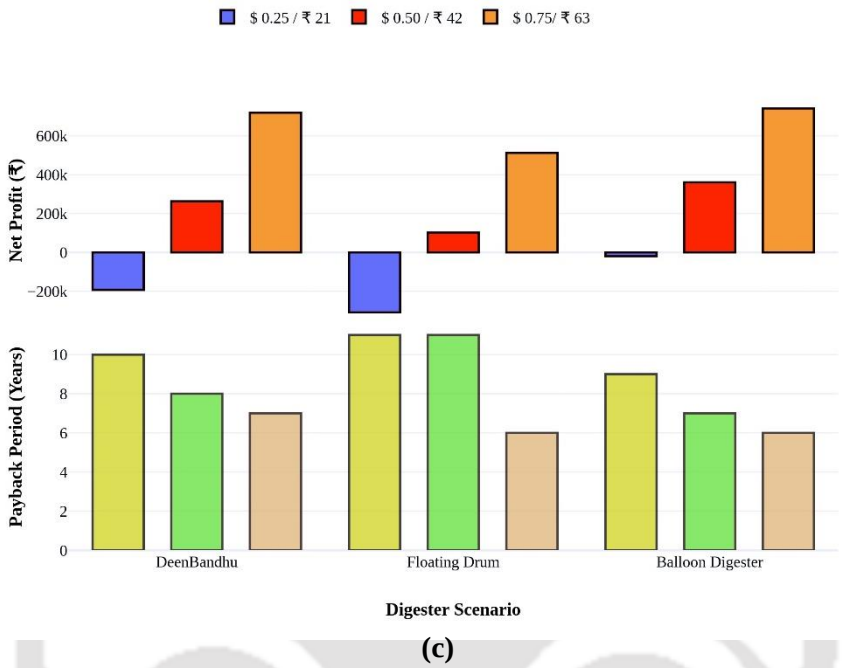
https://web.iitd.ac.in/~vkvijay/files/biogas%20entrepreneurship.pdf) at a cost of $46 per kg (as per the 2018 IOCL White Paper - https://iocl.com/download/White_Paper_EOI_1.pdf). Such an evaluation affirms that a biogas price (excluding biomethane) of ₹ 21.8 m^{-3} or $0.25 m^{-3}.

The selling price of the biogas at 0.25 $ m^{-3} translated into a negative net profit of ₹ -254,398 ($ -3,051 USD, as in Fig. 9.5 (c)). However, positive net profits of ₹2,62,920 ($3,154) and ₹ 7,19,445 ($8,629) have been achieved for the hypothetical biogas price alteration to 0.50 and 0.75 $ m^{-3} respectively (Fig. 9.5 (c)). This underscores the necessity for upgradation in the digester plant. Therefore, the biogas process plants despite undergoing optimization through machine learning strategies can at the best operate for integration in community projects being driven with the philanthropic cause or with corporate social responsibility framework. Thus, to realize the mentioned higher prices of the biogas in the thesis, price subsidy measures will have to be implemented in the waste management policies. Otherwise, sustainability of the biogas plants is not feasible.

Fig. 9.5(c) illustrates the net profit and digester scenarios for each biogas digester, specifically Deenbandhu, Floating Drum, and Balloon digesters, across a range of selling prices from $0.25 to $0.75. Upon scrutiny of the financial data, it becomes evident that at a selling price of $0.25 (₹21), the net profit consistently exhibits negative values, indicating substantial losses of $ -2321 (₹ -193,428), $ -3708 (₹ -309,182), and $ -236 (₹ -19,677) for Deenbandhu, Floating Drum, and Balloon digesters, respectively. However, a noteworthy trend emerges as the cash inflow gradually transitions to positive values as the selling price is increased to $0.25(₹ 42) with a profit range of $3136 (₹ 262876), $ 1225 (₹ 101723) and $ 4344 (₹ 360744). Notably, at an average selling price of $0.50 (₹42), a significant shift in the financial trajectory is observed for the biogas scenarios, signifying enhanced profitability. This underscores the sensitivity of these digesters to variations in selling prices, with the $0.50 point serving as a pivotal juncture for achieving positive cash inflows.

Furthermore, the alteration in payback periods with changes in selling prices is depicted. As the selling price varies from $0.25 to $0.75, the payback period undergoes changes from 9.2 to 6.4 years for Balloon, 11.6 to 6.1 years for Floating Drum, and 10.4 to 7.7 years for Deenbandhu digester scenarios. The base price considered for the payback period is $0.25. Consequently, with higher payback period

values, it becomes apparent that biogas technology may necessitate a price subsidy to significantly reduce the payback period to approximately 4 years. In scenarios where the selling price of biogas is even lower than $0.25, the payback period values would escalate beyond 11-13 years, rendering it economically impractical.



(c)

Fig. 9.5 (a) Capital cost of alternate biogas digester **(b)** Cost contribution of various entities for plant upgradation in the design-based biogas digester case study **(c)** Variation of biogas prices with net profit and payback period for alternate biogas designs

The PBP in the biogas scenario was found to vary between 2 and 8 years and was dependent upon the type of chosen digester, (Fig. 9.5 (c)). The cost reduction enabled a reduction in the minimum biogas price. For a 3-year PBP, the minimum price was evaluated as ₹ 80 m^{-3} ($ 0.97 USD) against ₹ 79 m^{-3} ($ 0.96 USD) for the base case. The study conveyed that the PBP for biogas plants in the given scenarios was about 3.5 years (average value). Thus, the project is more attractive in terms of the biogas price. However, the necessity of bio-fertiliser should be as well highlighted.

9.3.1 Competence of Alternate Operating Models and Predictive Analysis

The objective of the proposed AD calculator was to evaluate the long-term revenue potential of the AD unit, and thereby aid plant owners to optimize their operations for maximum profitability. The calculator

estimates costs, as well as financial break-down for a particular configuration. Accordingly, the owners could make informed decisions on the process operability.

For comparison purpose, a farm-based AD unit was utilized to process a mixture of OW (kitchen waste + vegetable waste) and rice silage for BP. The generated biogas was used on-site for the production of 20 kW of electricity, and the excess was exported to the gas grid. Additionally, revenue was generated through the sale of bio-fertilizers. The owner had the opportunity to explore different scenarios by adjusting the feed rate or RT. Accordingly, it assured an assessment of the profit margins and associated sensitivities. Although there was a limitation of available waste on the farm, the owner could still vary the amount of feed fed to the AD unit. Henceforth, the following cases can be considered:

a. *Case 1: Operating the AD unit as a baseline scenario*
b. *Case 2: Feed flow is reduced while RT is enhanced*
c. *Case 3: Enhance the feed flow and reduce the RT*
d. *Case 4: Operate at the base RT, but with half of the feed flow rate*

Fig. 9.6 (a-c) depict a comparative analysis of the base case and the proposed scenarios based on feed flow and RT variation. The RT critically influenced and altered biogas generation/upgradation. A reduction in the RT enabled greater feed throughput (Fig. 9.6 (a)). With decreased feed flow and increased RT, the average BP will lead to process inhibition due to increased gas production (Fig. 9.6 (b)). A reduced feed throughput resulted in lower BP rate in comparison to the previous cases. Hence, it led to process inhibition (Fig. 9.6 (c)). Table 9.3 presents the revenue estimates for each evaluated scenario. Case 3 emerged as the most profitable option. However, it is also important to compare Case 1 and 2, as a higher yield obtained by leaving the material in the digester to digest for a longer time may not offset the loss in material fed to the digester. The table also highlights a conflict of interest between sustainable practices and economic profitability. While increasing biogas yield and production of more biogas from feed material may seem economically sound, it may lead to the release of CH_4/CO_2 into the atmosphere, and accordingly contribute to higher GHG emissions. Conversely, processing more material and accepting the loss in yields is more profitable from an economic standpoint (at least as per the evaluated cases).

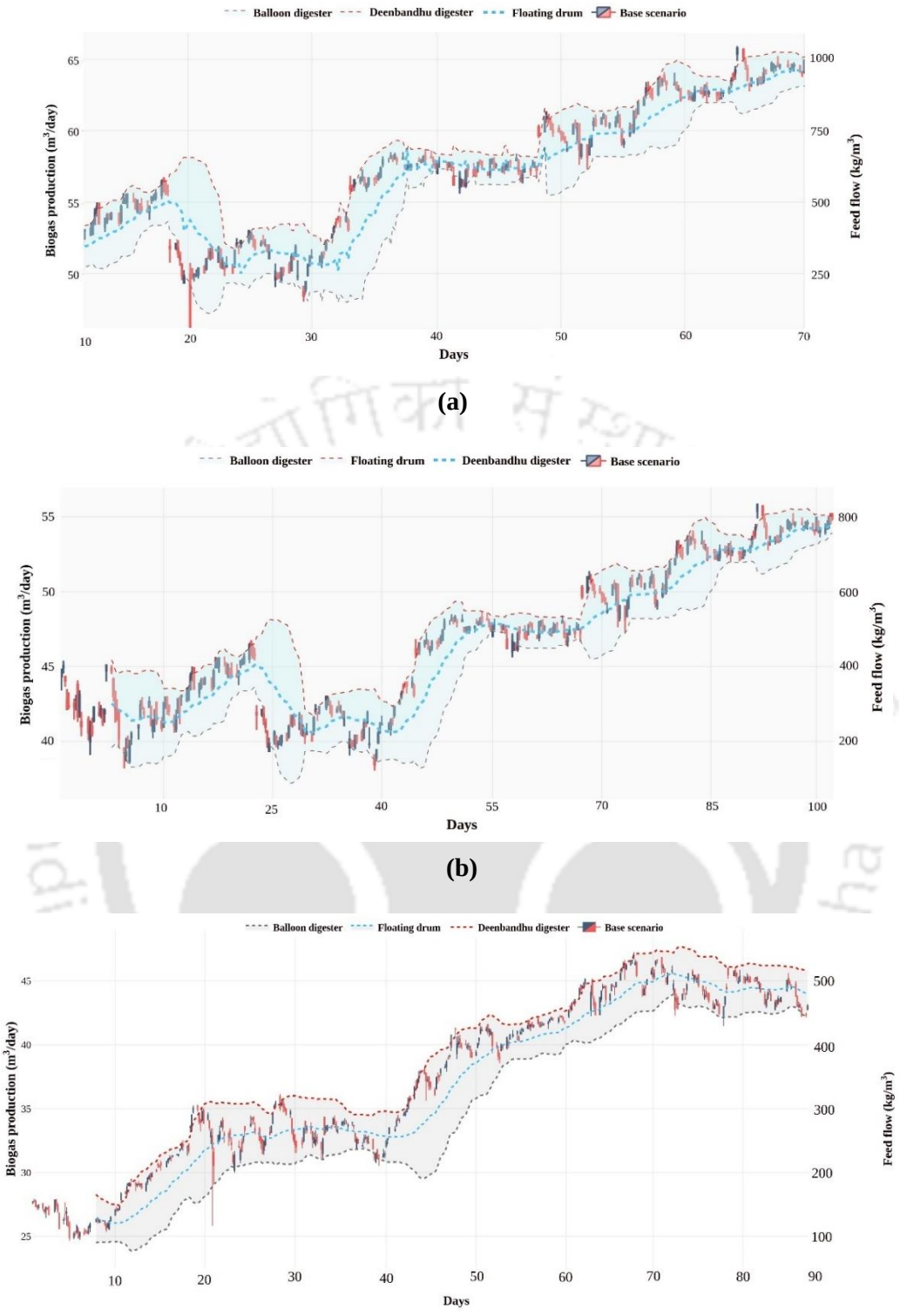


Fig. 9.6 Comparison of three hypothetical case scenarios with **(a)** reduced feed flow and enhanced RT **(b)** enhanced feed flow and reduced RT **(c)** reduced (halved) feed flow rate and base case RT

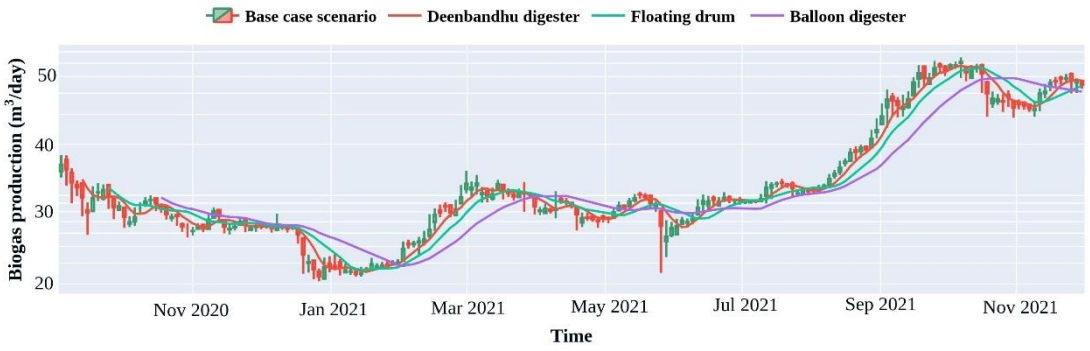


Fig. 9.7 Time dependent variation of BP (m^3 day^{-1}) rate (a year) for the proposed digesters

Fig. 9.7 shows the variation of daily BP rate (m^3 day^{-1}) and for altered psychrophilic conditions. A comparison of the base case scenario with the calibrated model has been addressed. During the psychrophilic AD process, the T range has varied for the base case from 22-35°C, (reported in Table 2.8 of the sub-section 2.8.5.4 in the thesis). The daily T average was determined as the mean of the maximum and minimum temperatures of the day. According to a prior report, BP is positively correlated with temperature, as higher temperatures support the activity of methanogenic bacteria (Nsair, et al., 2020). Kamrup is a city with colder nights than days, and the microbes are mostly active during the daytime. As a result, the BP is higher during the daytime and in the summer months (June - August, as shown in Fig. 9.7). However, during the winter season, the T drops to about 10°C during the night. This translated into a reduced methanogenic bacterial activity. Consequently, total BP decreased as being observed in the thesis outcome. Under psychrophilic conditions, T fluctuations can significantly affect bacterial activity thereby lead to decreased biogas yield.

9.3.2 ML-based Prediction and Forecasting Efficacy for BP Rate Evaluation

To assess upon the predictive ability of ML models for biogas yield, the performance of the calibrated models was evaluated against a 350-sample dataset of measured biogas yield from the AD unit in the test dataset (as shown in Fig. 9.8). Metrics such as R^2 and RMSE were utilized to examine the accuracy and precision of the models. This thesis deployed three different models namely hybrid (ARIMA-MLP), MLP, and SVR algorithms. The SVR model was trained with a penalty parameter (C) value of 1 and ε of 0.2, and enabled a train score of 0.75 and a test score of 0.70. In comparison, the MLP model consistently exhibited higher train (0.87) and test (0.84) scores. The MLP was trained with three hidden

layers (each with a size of five), *tanh* as an activation function, and with few other parameters. For the forecasting, ARIMA has been incorporated for the short-term forecasting (4 years). For such evaluation, it was assumed that the PBP of the plant was 4 years.

The performance of the hybrid ARIMA-MLP model for the test dataset was promising (R^2 of 0.96). This is a positive indicator for a model trained on real-world industrial data. An alternative model using ANN and SVR was developed for comparison purposes. The ANN is widely used to predict the performance of AD facilities. The MLP was used as the baseline model for comparison with SVR and the hybrid model. The results conveyed that the ARIMA-MLP model (R^2 = 0.96, RMSE = 5.71) outperformed the MLP (R^2 = 0.87, RMSE = 23.62) and SVR (R^2 = 0.57, RMSE = 465.31) models. In the test dataset, it was observed that the relative RMSE was approximately 15% for the ARIMA-MLP model and 19% for the MLP model. For the forecasting purpose, ARIMA was deployed for a 4 years PBP. Accordingly, the model forecasted a good prediction performance (R^2 = 0.76, RMSE =573) of the biogas yield (6.3×10^7 - 7.2×10^7 Joule year^{-1}).

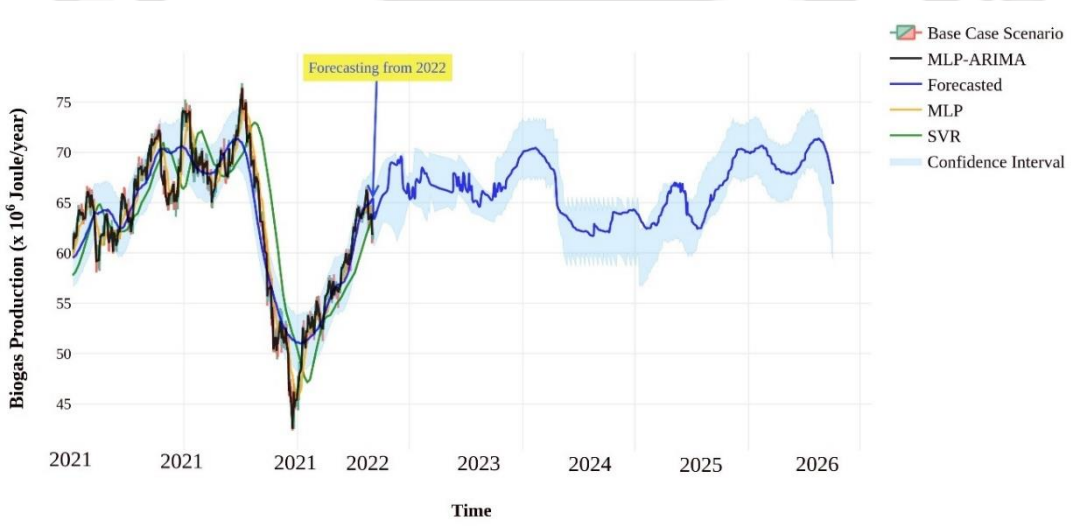


Fig. 9.8 ML based predicted and forecasted BP rate variation with time

For the purpose, firstly, autocorrelation was deployed to generate short-term correlations in the sequence. Thereafter, white noise test was conducted. The ARIMA model did not meet the white noise criteria. Hence, ARMA (p, q) models were selected with a maximum delay order of 5. To predict the BP rate for the next 5 years (PBP), the ARIMA (1, 1, 0) model was deployed. The predicted and

forecasted results for each model were correlated for the year 2021. Accordingly, ARIMA-MLP exhibited the best results followed by the MLP. The future biogas generation rate trends were then forecasted for the year range of 2022-2026 and with the predicted past data. The validation results of the forecasted BP in comparison to observed BP data for the year 2021 (61×10^6 Joule $year^{-1}$) have been presented in Table 9.4.

Table 9.4 also provides data associated to the net profits of the Deenbandhu, Floating drum and Balloon digesters. For the base case scenario with 0.25 \$ m^{-3} selling biogas price and feed input of 1500 kg day^{-1}, the net profit is ₹ -389174 (\$ -4663.75 USD) for a produced biogas of 53 m^3 day^{-1}. However, corresponding net profit values for the digesters affirmed consistently negative values. Thereby, the values convey substantial losses of -2321 \$ (- 193,428 ₹), -3708 \$ (- 309,182 ₹), and - 236 \$ (-19,677 ₹) for Deenbandhu, Floating Drum, and Balloon digesters, respectively. Thus, the biogas technology also requires price subsidy to significantly increase the net profit and simultaneously reduce payback period (years) hence reduce the payback period (years).

Table 9.3 Data summary of three alternate hypothetical case scenarios for the three digesters.

	Deenbandhu digester				**Floating drum digester**			**Balloon digester**		
	Case 1 (Base scenario)	Case 2	Case 3	Case 4	Case 2	Case 3	Case 4	Case 2	Case 3	Case 4
Feed material (kg day^{-1})	Mixed OW: 1000 Rice silage: 500	Mixed OW: 700 Rice silage: 300	Mixed OW: 800 Rice silage: 400	Mixed OW: 500 Rice silage: 250	Mixed OW: 700 Rice silage: 300	Mixed OW: 800 Rice silage: 400	Mixed OW: 500 Rice silage: 250	Mixed OW: 700 Rice silage: 300	Mixed OW: 800 Rice silage: 400	Mixed OW: 500 Rice silage: 250
RT (days)	90	100	70	90	100	70	90	100	70	90
BP (m^3 day^{-1})	53	52	60	55	48	54	51	43	50	44
Net profit (₹ annum^{-1})	₹ -389174 ($ -4663.75 USD)	₹ -254398 ($ $ -3051 USD)	₹ -193429 ($ -2321 USD)	₹ -231551 ($ -2777 USD)	₹ -354789 ($ -4255 USD)	₹ -309180 ($ -3708 USD)	₹ -331943 ($ -3981 USD)	₹ -72876 ($ -874 USD)	₹ -19677 ($ -236 USD)	₹ -65286 ($ -783 USD)
LPG cylinders	-	387	627	231	189	268	134	208	464	109

Table 9.4 Prediction and validation results for the forecasted BP rate and for 2021 year.

Prediction				
Models	**Prediction (Joule year^{-1})**	**Standard error**	**95% Confidence interval**	
ARIMA-MLP	60×10^6	0.35	60134476	60734348
MLP	61×10^6	10.54	61990488	62004561
SVR	56×10^6	21.62	56871293	57332175
Forecast				
	Forecasts (Joule year^{-1})	**Standard error**	**95% Confidence interval**	
ARIMA-MLP	59×10^6	3.67	59873213	60342219

Table 9.5 A comparison of ML and ANN based AD process investigations based on the best findings of this work and prior art data.

Feed	**Input parameter**	**Prediction parameter**	**Data size**	**Model**	**Prediction performance**		**Training: Testing ratio**	**Authors**
					R^2	*RMSE*		
Mixed OW, rice silage, vegetable waste	RT, Feed flow, feed characteristics	BP Forecasting	350	ARIMA-MLP: MLP: SVR: ARIMA-MLP	0.89 0.82 0.57 0.76	5.83 140.27 854.31 573	70:30	This work
Food waste	pH, RT, T	Biogas generation	205	ANN (Back propagation)	0.8456	-	80:20	Neto et al., 2021
Corn Straw+ grass silage	Substrate characteristics	Biogas flow rate	287	MLP	0.97	11.07	70:30	Beltramo and Hitzmann, 2019

The proposed method that utilized ANN models for the prediction and/or optimizing BP has been compared to those conveyed in various studies in Table 9.5. From the thesis work, it can be inferred that the R^2 and RMSE values of the MLP did not convey better performance in comparison to those inferred by Beltramo and Hitzmann, 2019. According to Guo et al. (2021), one possible explanation for such limitations of neural networks in providing a mechanistic understanding of the AD process is that they cannot justify their reasoning process or provide a basis for the reasoning. Even though neural networks can estimate process changes using empirical data, they may not fully capture the complex nature of these changes. However, the authors suggest that ANN can still be used for mechanistic understanding of the AD process, provided insights exist for the underlying nature of the process alterations. Thus, to resolve this issue, a hybrid ML model has been proposed in this work to model AD

process data for the prediction and forecasting of the biogas yield. The performance resulted in decent metrics (R^2 = 0.89 and RMSE = 5.83).

9.3.3 Biomanure Value

Considering and quantifying the financial value of biomanure, the AD process can be further explored for the manure's utility as a soil fertiliser. In line with India's waste policy directives, AD can be used to manage animal manure and agricultural waste. However, the cost recovery price for biomanure is rarely higher than its market value due to its limited marketability, associated market costs, and perceived value by customers (Edwards et al., 2015). The proposed AD calculator values biomanure based on the quantity of fertiliser it replaces for the farm. As a result, the valuation of biomanure is dependent on the site-specific issues. If the Indian market for biomanure develops in the future, it could lead to better estimates and improved AD process economy.

9.4 Significant Findings and Summary

The utility of AD technology is crucial to manage OW and for renewable energy production. Precisely, AD technology is considered to be a pragmatic approach for enhanced BP rate through the utilization of existing infrastructure and addressal of challenges related to substrate properties and system stability in single-substrate AD processes. Effective optimization and management of various incoming waste streams are key factors for a successful operation of the AD processes. This work employed various techniques such as modelling, simulation and cost evaluations of the AD processes and upgrading scenarios. The results were consistent with experimental, literature, and spreadsheet data. Ensuring model interpretability is important, and this is especially relevant for decision-making purposes. The expertise and experience of the researchers in bioprocesses and AD facility operation were utilized to guide the model development and interpretation. This ascertains scientific insights into interesting and operationally relevant outcomes. The investigation of the influence of input variables such as feed flow and RT on biogas yield was conducted, and important scores were used to determine the variables having greatest impact.

The AD calculator proposed in the thesis work was capable to estimate BP rates for both short-term and long-term operations in AD units. In comparison with the results, the model was validated with case studies of other AD units, and the model was found to predict values within a ±10% range of the reported values. Among the proposed digesters, the Deenbandhu digester exhibited the most promising outcome with a higher yield of approximately 60 m^3 day^{-1}, a favourable biogas price, and a net profit of ₹ 225110 (2724.05 USD). The calculator assists the AD owners to determine the most profitable operational approach, and thereby support the technology's sustainable use.

This work also contributes to the enhanced research in the exploration of the ML models' application for BP prediction and improved understanding of the AD process. ML models are highly effective to capture the behaviour of complex systems that are difficult to model with traditional methods. The ARIMA-MLP model was used with a dataset of 350 data samples from the Sonapur AD plant, (collected for a period of one year). The results show that the ARIMA-MLP model outperformed the basic MLP model and SVR. This demonstrated its robust predictive potential. Enhanced predictive models for AD performance can result in increased operational efficiency, improved understanding of waste inputs and operating conditions, and a more sustainable OW valorization industry.

Chapter 10
GIS-based Decision Model and Techno-Economic Calculator for Compost Facilities

Chapter 10

Development of a GIS-based Decision Model and Techno-economic Calculator for Compost Facilities: A case Study for Guwahati City

In this chapter, the development of a GIS-based decision model and a tailored techno-economic calculator for compost facilities has been addressed for the Guwahati city. Accordingly, the practical application of the model has been demonstrated. The conducted research aims to enhance the decision-making process and economic viability of compost facility in the Guwahati city. Section 10.1 summarizes the pertinent introductory details and delineates upon the objectives of the study. Thereafter, section 10.2 addresses findings from GIS spatial analysis to infer upon suitable locations for compost facilities in Guwahati city. Section 10.3 discusses the model's accuracy, reliability, and performance efficacy. Thereby, it provides insights into the model's effectiveness for compost facility planning. Section 10.4 compares the developed compost model with similar models used in other case studies reported in the prior art. Also, seasonal study has been conducted with the calibrated compost model, and for scale variant compost systems such as high scale compost facility (HSCF), medium scale compost facility (MSCF), and yard waste compost facility (YWCF) systems. These have been duly addressed in section 10.5 of the thesis. Subsequently, sections 10.6, 10.7 and 10.8 elucidate on the economic perspectives and associated financial implications of the operated compost facilities. This is as per the developed decision model, in terms of operational revenues and costs (section 10.6) followed by sensitivity analysis (section 10.7) and subsequent findings of the ML model applications to predict and forecast CP rate (section 10.8). Finally, section 10.9 elucidates with a concise summary, and emphasizes upon the significance of the developed decision model to envision potential avenues for future research in the field of compost facility planning[#].

Overview

The sub-objectives being addressed in the chapter refer to site selection using ArcGIS and multi-criteria overlay analyses for a compost processing facility tailored to the needs of a specific municipal body; a techno-economic calculator for the assessment of small to large scale composting facilities and a framework to accurately and efficiently predict and forecast CP volumes for the improved planning and management of composting facilities in the Guwahati city. The techno-economic calculator evaluated costs and capacity for the upgraded scenarios. Accordingly, the HSCF facility has been evaluated for a total cost of $120,000 USD ($105,000 USD of the base case). The performance metrices R^2 and RMSE for MLP, GA and SVR algorithms were 0.87 and 34.73; 0.82 and 103.22 and 0.76 and 231.51

[#] Article communicated: Singh, T., & Uppaluri, R. V. (2023). Development of a GIS-based decision model and techno-economic calculator for compost facilities: A case study for Guwahati city.

respectively. The PBP of the plant was assumed to be 7 years. The hybrid ARIMA-MLP model performed the best for the testing dataset and with an R^2 of 0.94 (0.84 for ARIMA-GB). The most popular ANN (MLP) was deployed as the baseline case in the conjunction with the hybrid GA and SVR models. Accordingly, the ARIMA-MLP (R^2 = 0.94, RMSE = 23.22) outperformed ARIMA-GA (R^2 = 0.84, RMSE = 81.43) and ARIMA-SVR (R^2 = 0.67, RMSE = 383.94) models. In summary, the thesis findings highlighted the effectiveness and practicality of the developed GIS-based decision model and techno-economic calculator for scale variant compost facilities planning in the Guwahati city.

10.1 Introduction

For the Guwahati city, the chapter elaborates upon the findings of different alternate schemes such as landfill disposal, waste conversion facilities, and establishment of new landfills. Accordingly, optimal waste management practices have been highlighted. The previous chapter presented a novel hybrid approach that integrates an AD calculator with the ML algorithm for the optimization and enhanced efficiency of the BP systems in the Indian biogas plants. However, the thesis work also aims to establish a comprehensive decision-making model that can aid municipalities in realizing well-informed choices in the field of the disposal and reuse of solid waste. Considering these issues, the specific novelty of the addressed work in this chapter are as follows:

(a) to devise a decision-making model that incorporates economic, and environmental parameters for the identification of the optimal waste disposal technology for the Guwahati city

(b) to determine the most suitable location for a compost processing facility being tailored to the needs of a specific municipal body in Guwahati, India

(c) to develop a techno-economic calculator for the assessment of small to large scale composting facilities (HSCF, MSCF and YWCF)

(d) to develop a ML algorithm and hybrid model (ARIMA-ML) for the prediction and forecasting of CP rate in the Guwahati city. Accordingly, improved planning and management of composting facilities in the Guwahati city can be realized.

10.2 GIS-based Spatial Analysis for Site Selection

Due to various jurisdictions in the Kamrup district of Assam, India, Guwahati (gateway of North-east India) is of prime attention. In Kamrup, municipalities emphasis is upon various important aspects such as waste management perspectives at source, collection services, waste diversion from landfill, reuse, recycling and composting of diverted waste, and recovery and generation of energy from residual waste. The basis for such co-ordinated policy research is due to the associated jurisdictions in terms of the optimal use and disposal of MSW. Kamrup district has 12 towns, 7 municipalities, and other districts (Singh and Uppaluri, 2022). The district faces challenges related to inadequate waste management infrastructure including non-existent proper landfill sites. Also, the existing landfills are often poorly managed, and could lead to environmental pollution and health hazards. The Kamrup district administration (GMC) has been reportedly working to improve waste management practices, such as the commissioning of new waste treatment plants and exploration of alternative solutions for waste disposal. Envisioning that the landfill capacity of 1 lakh tonnes of waste in 1980 would increase significantly in the subsequent decades, the authorities of Kamrup district wish to divert as much waste as the Guwahati city produces for the composting process. However, the compost plant has been non-functional since 2020.

In 2021, Kamrup had a population of approximately 1,710,000 (Census of India, 2021). Currently the district generates approximately 3 million tonnes of waste per year (Singh and Uppaluri, 2022) and has only one landfill site (Boragoan). Also, former sites were closed and converted into transfer stations. Fig. 10.1(a) shows the LULC for the Kamrup district by considering protected region, railways, buildings, land use, urban settlements, waterways and roads parameters for 2019 (Data Source: Geopackage). The district area covered by the built-up land increased considerably from 225.6 km^2 in 2000 to 453.2 km^2 in 2018 and for Guwahati, the area enhanced from 25.8 km^2 (2000) to 123.3 km^2 (2018). Consequently, the surrounding rural areas, comprising of cropland, wetland, and forest land, experienced a decreasing trend in the land cover during the same period. Precisely, cropland reduced from 4329.6 km^2 (2000) to 3235.9 km^2 (2018) in the district and from 216.5 km^2 (2000) to 210.3 km^2 (2018) in the Guwahati city. Similarly, the wetland and forest area for Kamrup altered from 3612 km^2

to 4215 km^2 and from 581.5 km^2 to 467 km^2 respectively in the year range of 2000-2018. Incidentally, for the Guwahati city, these reduced from 81.9 km^2 to 78.3 km^2 and 71.4 km^2 to 63.2 km^2 respectively. Grassland, which is also present in the nearby vicinity, also varied marginally from 21.7 km^2 (2000) to 25.3 km^2 (2018). Such alterations have been attributed to the rapid urbanization of the district, which is contributed primarily by Guwahati city. Subsequently, vegetation and cropland areas transformed into densely populated areas. Such a trend has been also supported in a recent study that infers that the corresponding reduction in the land covers of Kamrup district were accompanied with an increase in the built-up area (Nath et al., 2021). Additionally, bare land, which is present in small amounts, also affirmed a reduced trend from 4023.7 km^2 (2000) - 3223.5 km^2 (2018) and 49.2 km^2 (2000) - 26.9 km^2 (2018) in Kamrup district and Guwahati respectively. The roadways, waterways and railways also contributed significantly to the LULC variation during the mentioned period. Kamrup has six existing transfer stations and waste is currently transported from these stations to the landfill site near Boragaon. Building a waste conversion facility to treat the waste from both city and part of the neighbouring city's waste could enable the district to realize a sustainable waste management system.

Considering the above-mentioned parameters using ArcGIS modelling system, a multi-criteria overlay analyses was deployed to identify potential locations for a waste facility plant (compost) in Kamrup district. Fig. 10.1(b) depicts the final constraint map of the overlay analysis. For the case, weights were assigned for each criterion and were in terms of the relative importance in the decision-making process (0-1). In this study, EPAs, transportation lines and urban settlements were assigned higher weights than the available land. The data was thereafter normalized for each criterion to a common scale for the realization of a composite map. For this, each criterion was overlayed to depict areas that met the desired criteria for a composting facility in the district. Based on the multi-criteria analysis conducted for the Kamrup metropolitan region, the three most suitable locations for the waste conversion facility are Panikhaiti (26.2006° N, 91.8757° E), Jorabat (26.0989° N, 91.8623° E), and Sonapur (26.1172° N, 91.9802° E). These locations were identified based on a variety of factors, including proximity to waste generation sources, transportation infrastructure, and compatibility with zoning regulations and land use patterns. Further, site assessments and community engagement shall be determined for the final

location and thereby ensure that they meet all necessary regulatory requirements and are acceptable to the local community. Further, for Kamrup (rural), the most suitable locations for the waste conversion facility are found to be Chhaygaon (26.0481° N, 91.3867° E), Boko (25.9778° N, 91.2356° E), Palashbari (25.2794° N, 89.3535° E), Nagarbera (26.1001° N, 90.9797° E), and Kamalpur (24.1969° N, 91.8331° E).

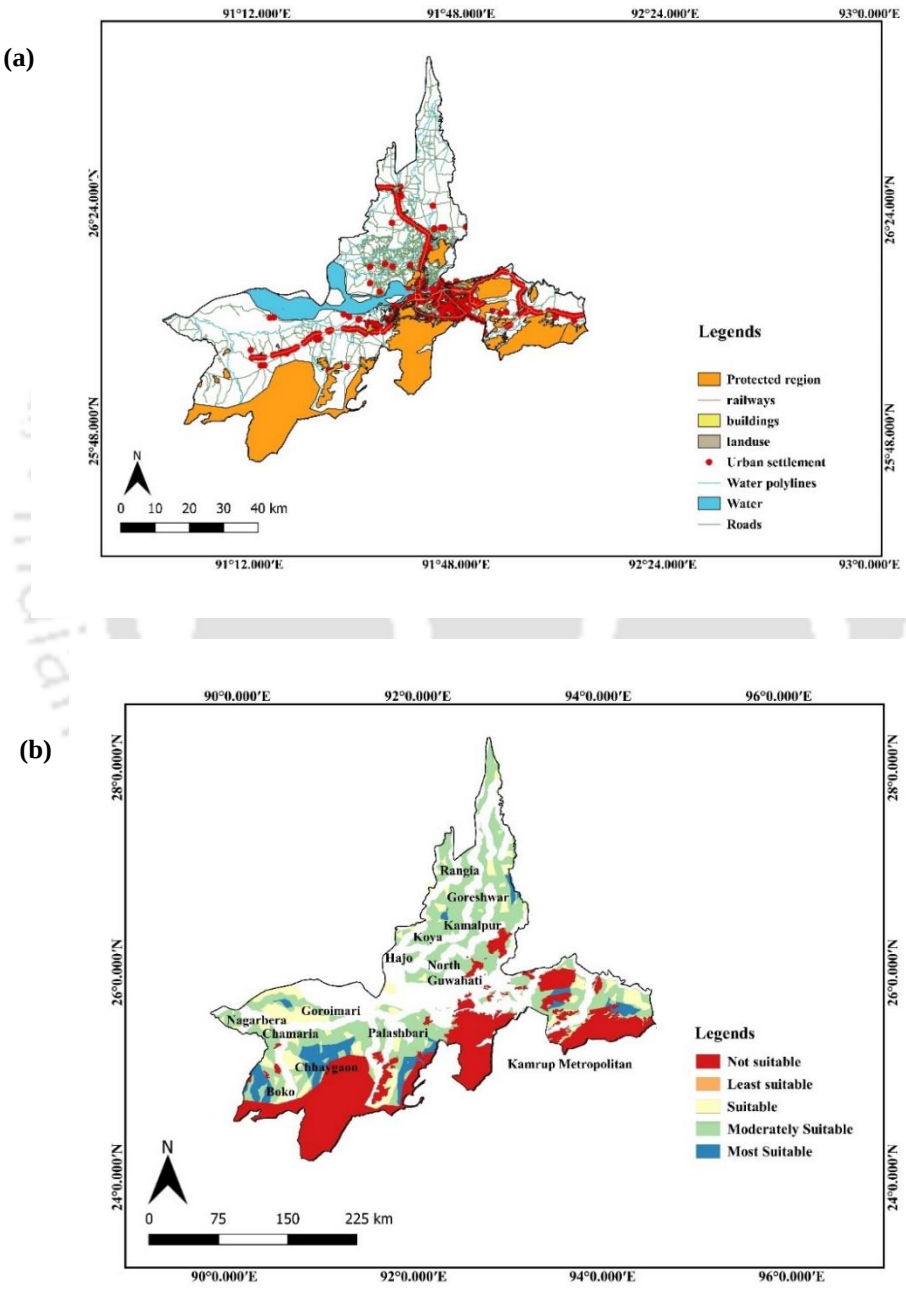


Fig. 10.1(a) LULC mapping of Guwahati city using preference analysis and **(b)** final constraint map of Guwahati city based on the multi criteria analysis of waste conversion facility for Kamrup district

10.3 Compost Model Assessment

The proposed CP model requires input information with respect to the feedstocks (which determine the RT and potential compost yield), and relevant information with respect to the equipment and technology and the econometric estimations. Thus, the model should be readily usable for most sites. Three years sample data (200 samples from 2016-2019) from the GMC compost unit was used to evaluate the performance of the proposed model for CP estimation. For an RT of 120 days, the plant owners can benefit from the performance benchmark of the CP system and with the seasonal variation in the T. Seasonal changes in T that occurred in the windrows during the composting period (120 days) for the GMC compost plant have been shown in Fig. 10.2. For the case, the ambient temperature was about 34±3 °C in the day and 25±2 °C in the night for summer and autumn cases, and was about 22±2 °C and 15±3 °C in the night for winter and spring cases respectively. The T of the windrows gradually increased from 47, 49, 36, 38 °C and reached a maximum of 60, 61, 55, 49 °C after 120 days of composting time during autumn, summer, winter and spring seasons respectively. Thus, the T alterations during the compost process followed a typical pattern that exists in many composting systems (Khalil, 1996). The temperature peak during the composting of MSW at Andhra Pradesh (India) compost plant (Kurmana and Srinivas, 2021) usually exceeds 60 °C. Well-aerated composts often attain temperatures of 55-65 °C and may even reach 70 °C (Stutzenberger et al., 1970).

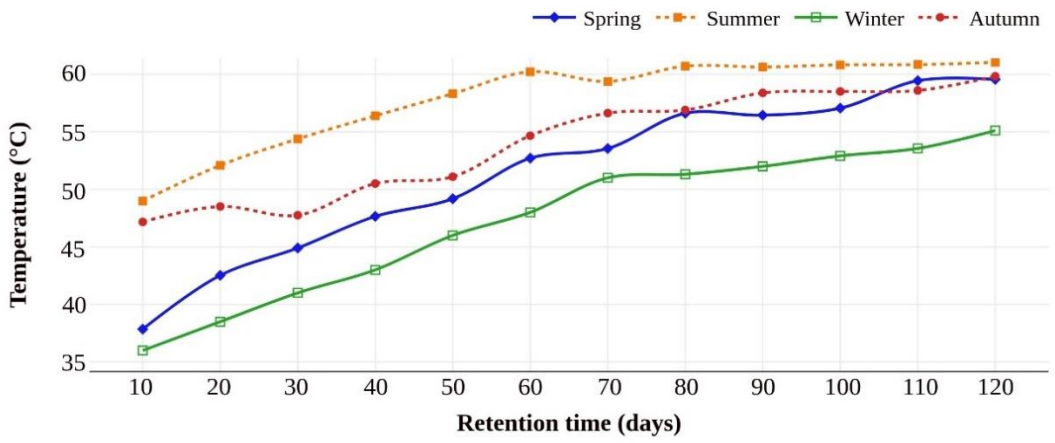


Fig. 10.2 Seasonal variation of the RT of the composing process (windrow composting, GMC plant)

10.4 Competence of Compost Models based on Literature Case Studies

Reported case studies of the compost units were used to compare the model estimations of the prior art based empirical models (Composting Calculator, The Energy and Resources Institute (TERI); Compost Production Cost Calculator by the International Fertilizer Development Center (IFDC); Composting Financial Model by the Agro-Enterprise Center (AEC); Compost Calculator, IIT Delhi) and outcomes few case studies. Thus, the proposed model was not calibrated. Hence, the comparison confined to the model's ability to assess upon fairly accurate preliminary steady-state estimates. These have been summarised in Table 10.1. Among several case studies, the preliminary CP estimations were within ±20% of the reported cases. In the considered cases, the CP was estimated with the HSCF, MSCF and YSCF designs and is comparable with those presented in the prior art and relevant case studies. To assure confidence levels, the CP profiles were compared with the prior art data and with the parity plots (Fig. 10.2). Based on the prior art case study comparison, the uncalibrated model had a maximum prediction error of up to 187.5 for MSCF. The effectiveness of the uncalibrated compost model in predicting CP rate was evaluated and depicted in Fig. 10.3. The performance metrices for HSCF (R^2: 0.95, RMSE: 121.5 and SSE: 1.01), MSCF (R^2: 0.92, RMSE: 187.5 and SSE: 1.63) and YWCF (R^2: 0.93, RMSE: 162.1 and SSE: 1.32) ensure that good performance of the developed calculator is possible after calibration.

A visual observation of the base case scenario and the calibrated model has been depicted in Fig. 10.3(a). In general, the proposed models very well predicted the CP data for various samples. This further emphasised upon the importance of model calibration during its application at the plant site for effective CP volume estimation. Further, there could be notable gaps in the estimated CP that did not appear in the measured values. These gaps could be due to sharp alterations in the feed stock quantity. Simple model calibration could be carried out to get better estimates of the compost yield potentials. Three year data samples (200) were used to calibrate the potential yields, and the calibrated model was used to estimate CP across the same data set (Fig. 10.3(b)). The variation of the CP rate for each of the facilities after calibration has been depicted in Fig. 10.3(c). The HSCF utilizes low-moisture feedstock and has a longer RT ranging from 20-120 days. This resulted in a high-quality compost product due to

complete decomposition of organic material. However, longer RT led to lower CP rates in comparison to other composting methods. The MSCF typically utilizes a mix of feedstock with a higher MC in comparison to an HSCF. The RT in an MSCF is typically shorter than that of the HSCF, and alters from 20-100 days. Such shorter RT allows faster decomposition and higher CP rates than the HSCF. However, the quality of the compost may not be as high as in the HSCF. This is due to lesser complete decomposition of the organic material in the YWCF system that typically uses only yard waste as feedstock. Incidentally, this feedstock has a lower MC in comparison to food waste. Depending upon the process needs, the RT in a YWCF can vary. Since yard waste composting facilities typically only accept yard waste, the compost produced is also limited to only yard waste-based compost.

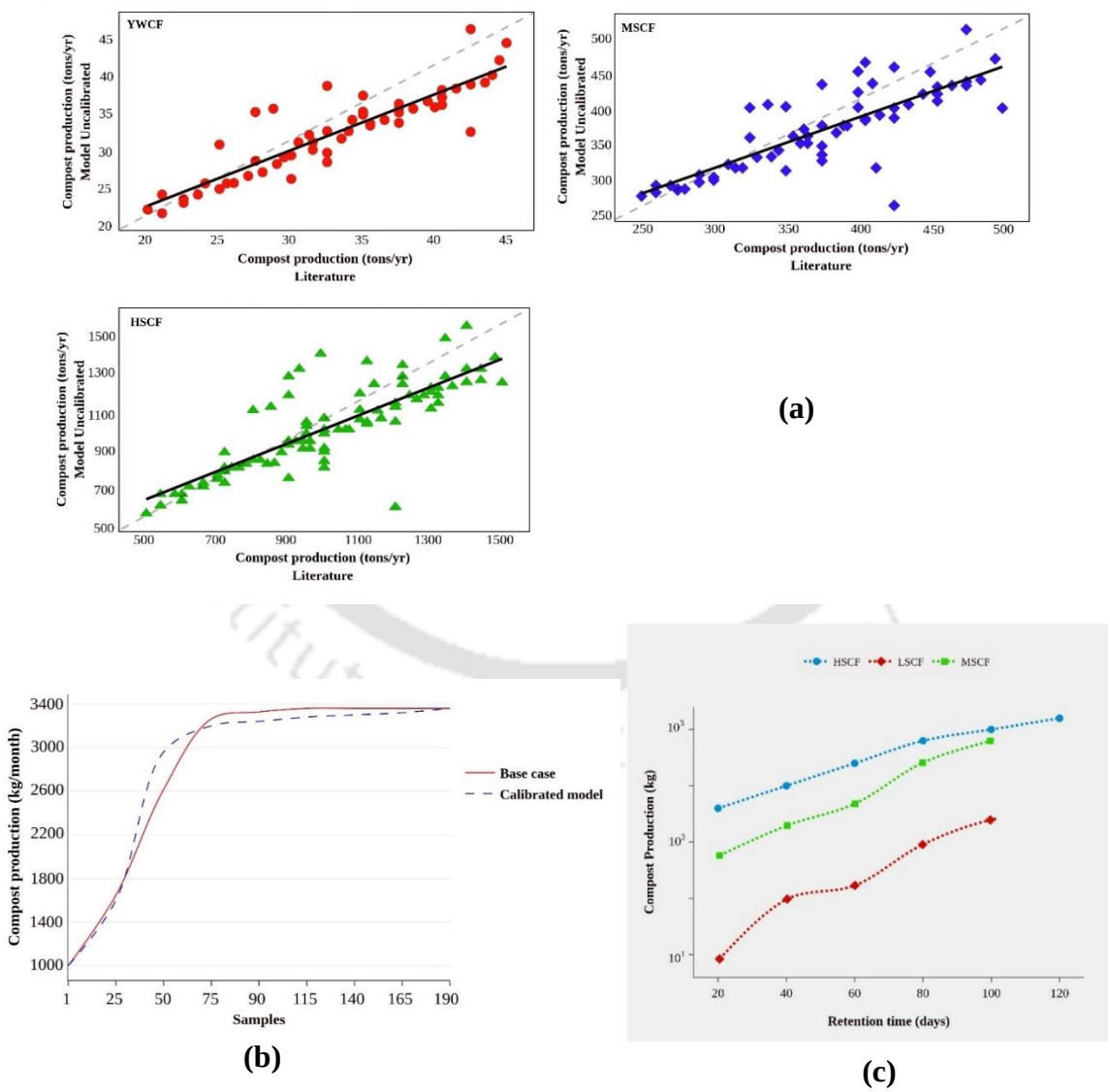


Fig. 10.3(a) Calibrated model parity plots for the CP rate case study **(b)** calibrated model performance with respect to sample size **(c)** CP rate variation for alternate facility after calibration

10.5 Calibration Model based Seasonal Study

Seasonal variations in feedstock availability can affect the performance of composting processes in large-scale (HSCF), medium-scale (MSCF), and yard-scale (YWCF) operations. During spring and summer seasons, an abundance of nitrogen-rich feedstock, such as food waste and grass clippings, confirms greater MC due to higher water content. Accordingly, bulking agents may be required to achieve the optimal moisture range of 50-70% (Liang et al., 2002). The T may reach 60-70°C due to higher microbial activity. During autumn and winter, the compost pile may require nitrogen-rich material to maintain the optimal carbon-to-nitrogen ratio of 25-30:1, and additional insulation may be required to maintain the T range of 55-65°C in HSCF operations, 55-65°C in MSCF and 50-75°C in YWCF operations (Mahapatra et al., 2022). Accordingly, the reduced MC may decrease, and water may have to be added or the pile be covered to prevent moisture loss. In terms of feedstock availability, nitrogen-rich materials may decrease, and supplementation with manure or other nitrogen sources may be necessary (Wang et al. 2022). Overall, the seasonal variations in composting across different scales depend upon feedstock availability, MC, and T, and operators need to adjust their composting practices to maintain optimal conditions for the composting process in the entire year (Fig. 10.4 (a-c)).

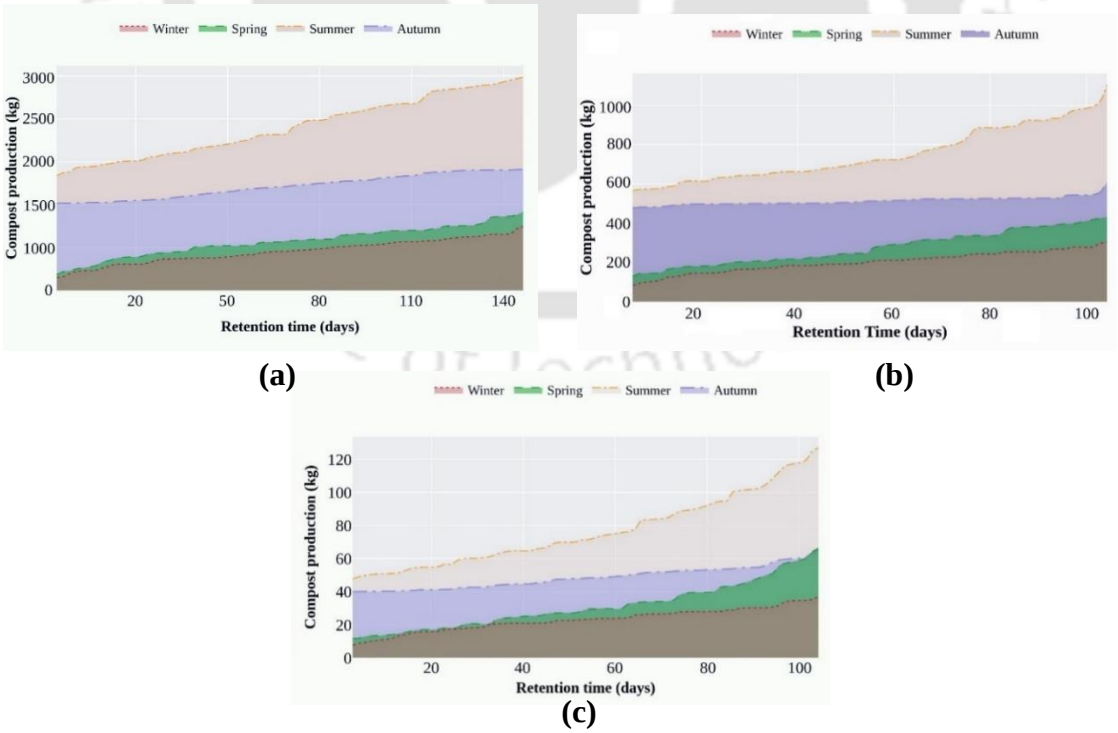


Fig. 10.4 CP rate seasonal variation for **(a)** HSCF (feed: 10 tons) **(b)** MSCF (feed: 5 tons) and **(c)** YWCF (feed: 1 tons)

Table 10.1 Prior art based comparative summary of compost calculator performance data.

Studies	Location	Feedstock and load	RT	Output	Reported and estimated study report (kg)	References
Compost Production Cost Calculator	Chennai Corporation, Tamil Nadu	MSW, Agricultural waste: 600 kg	120 days	CP (tons yr^{-1})	Reported model [1]: 100, [2]: 88, [3]: 75; Proposed model: [1] 150 (+10%), [2] 120 (+12%), [3] 90 (+19%)	International Fertilizer Development Center (IFDC)
Composting Calculator	Indian Institute of Science Education and Research (IISER), Pune	Food waste: 200 kg	155 days	CP (tons yr^{-1})	Reported model [1]: 50, [2]: 60, [3]: 55; Proposed model: [1] 70 (+16%), [2] 60 (+7%), [3] 80 (+15%)	TERI
Composting Financial Model	Coimbatore Municipality, Tamil Nadu	MSW, Mixed OW; Sewage sludge: 500 kg	90 days	CP (tons yr^{-1})	Reported model [1]: 120, [2]: 180, [3]: 150; Proposed model: [1] 170 (-12%), [2] 190 (+14%), [3] 200 (+15%)	Agro-Enterprise Center (AEC)
Compost Calculator	IIT Delhi	Food waste, yard waste silage: 150 kg	80 days	CP (tons yr^{-1})	Reported model [1]: 50, [2]: 35, [3]: 50; Proposed model: [1] 60 (+20%), [2] 78 (+13%), [3] 55 (-10%)	Dadhich et al., 2021
Kharagpur Municipality	Kharagpur, West Bengal	Food waste, MSW: 60 kg	120 days	CP (tons yr^{-1})	Reported model [1]: 10, [2]: 20, [3]: 11; Proposed model: [1] 15 (+8%), [2] 22 (+12%), [3] 30 (+18%)	Kumar and Goel, 2009
CHF India Foundation	Shivajinagar, Pune	Food and vegetable waste: 700 kg	90 days	CP (tons yr^{-1})	Reported model [1]: 130, [2]: 130, [3]: 150; Proposed model: [1] 180 (+12%), [2] 150 (+15%), [3] 160 (+10%)	Ambastha and Aich, 2020
Okhla Compost Plant	Delhi	Sewage sludge, Agricultural waste, MSW: 2000 kg	140 days	CP (tons yr^{-1})	Reported model [1]: 300, [2]: 340, [3]: 250; Proposed model: [1] 500 (+13%), [2] 480 (+9%), [3] 280 (-5%)	Mandal et al., 2014
AC Tech	Anna University, Chennai,	Food waste: 50 kg	80 days	CP (tons yr^{-1})	Reported model [1]: 8, [2]: 5, [3]: 10; Proposed model: [1] 12 (-16%), [2] 15 (-11%), [3] 20 (+7%)	Jayaprakash & Jagadeesan, 2018
Waste Ventures India	Hyderabad	MSW, sewage sludge, 1000 kg	120 days	CP (tons yr^{-1})	Reported model [1]: 100, [2]: 120, [3]: 75; Proposed model: [1] 210 (+15%), [2] 190 (+11%), [3] 140 (-9%)	Swaminathan, 2018

Abbreviations: Reported base case [1]: Compost Production Cost Calculator, (IFDC) [2]: Composting Calculator (TERI) [3]: Compost Calculator, IIT Delhi; Proposed model: [1]: HSCF [2]: MSCF [3]: YSCF

During summer season, the HSCF (with a feed quantity of 10 tons), MSCF (5 tons feed capacity) and YWCF (1 ton feed capacity) are expected to generate approximately 1800-3000 kg, 590-1500 kg and 50-130 kg of compost respectively. During autumn season, CP volume was 1500-1800 kg, 500-590 kg, 40-60 kg of compost for HSCF, MSCF and YWCF respectively. Similarly, in the spring season, the maximum CP volume reached up to 1400, 400 and ~70 and in the winter, it reached to about 1200, ~220 and ~40 respectively for HSCF, MSCF and YWCF composting systems.

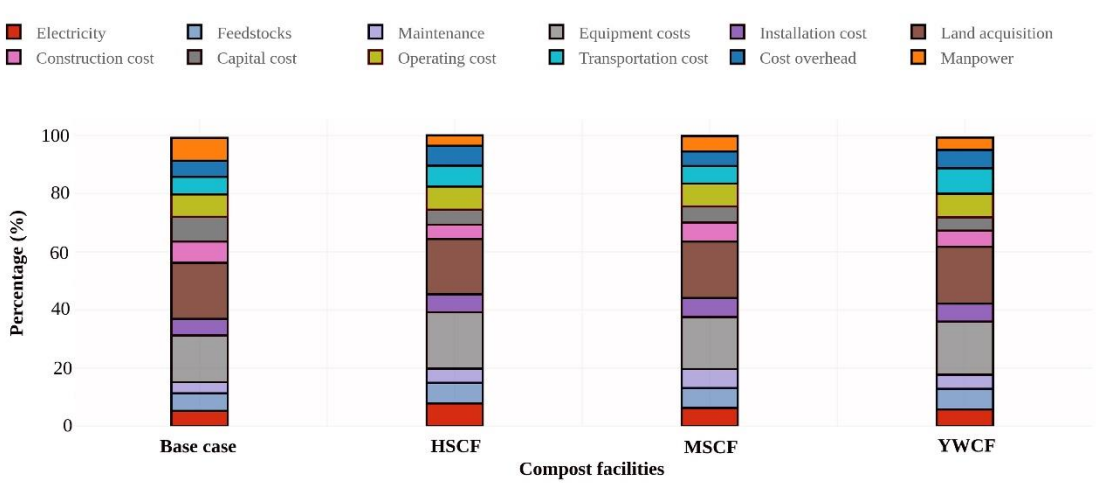
10.6 Operational Revenues and Costs

Fig. 10.5 (a) represents the expenses of the compost plants that witness fluctuations in the raw materials and utilities, equipment and overhead costs. All other expenses such as land acquisition, operating cost remained relatively unchanged in the plant despite its upgradation. The percentage difference in electricity costs for the YWCF system would be lower in comparison to a HSCF and MSCF system. This is due to the smaller size and lower capacity of the YWCF system. However, the actual difference will depend on factors such as the type of equipment used and the amount of electricity required. With respect to the feedstock costs, the percentage difference in feedstock costs for the YWCF would be higher in comparison to the HSCF due to the smaller volume of processed feedstocks. This may also result in the fluctuations in feedstock availability and quality. With respect to maintenance and overhead costs, the percentage difference for the YSCF would be higher in comparison to a larger scale facility due to the limited availability of resources. However, the actual difference will depend on factors such as management efficiency and production volume. Also, the HSCF may have more complex electrical and mechanical systems. Accordingly, it will have higher equipment and maintenance costs in comparison to MSCF and YSCF. However, a larger facility will benefit from economies of scale in terms of feedstock procurement and transportation costs.

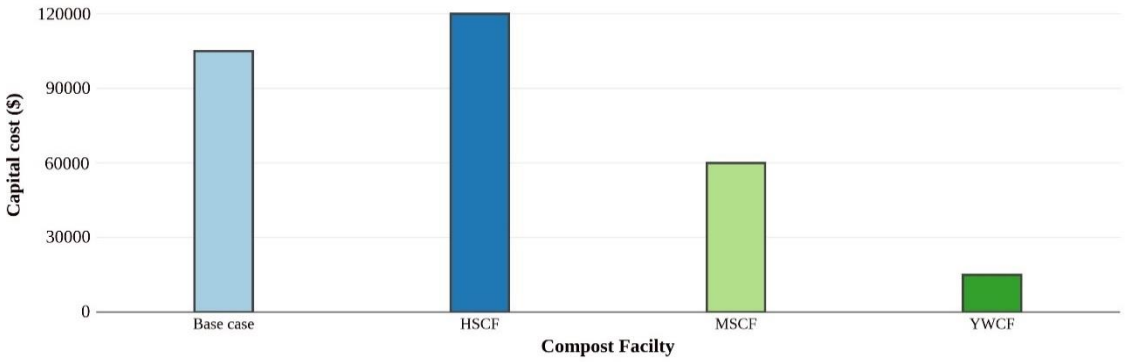
In general, many compost plant owners reserve themselves to publish the financial data of their processes. Not many financial reports are available for the compost plant units. The available few reports do not even provide a detailed account of the conducted evaluations. Thus, a comparative assessment of the operational revenue and costs is limited. Further, it shall be noted that the government incentives can play a major role in supporting the development and operation of compost facilities. For

such facilities, majority revenue is generated from CP sale to customers (farmers, landscapers, and garden centers). Government incentives can include subsidies or grants for facility construction or operation, tax credits, or other financial incentives. These incentives can help to offset the costs related to the development and operation of a compost facility. However, they are not the typical primary revenue source for the facility. Additionally, the availability and incentives extent can vary widely due to location and government specific policies. Hence, it's important to carefully evaluate the financial viability of a compost facility in terms of on its revenue potential from compost sales and other sources. Thus, the financial challenges can be better visualized for the compost technology. With government incentives being reduced with time and for increased overall capacity (tariff digression), alternate ways shall be explored to better manage and improve the process performance.

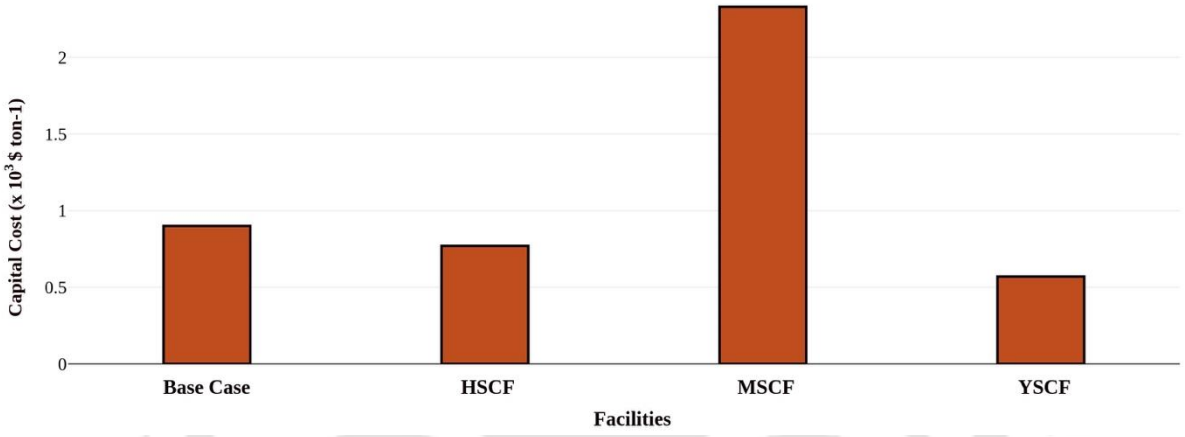
A comparison of capital cost and the capacity measured in tons for the base case unit and the proposed compost facilities have been shown in Fig. 10.5 (b). The calculator evaluated costs and capacity for the upgraded scenarios. Accordingly, the HSCF possessed a total cost of $120,000 USD in comparison to the $105,000 USD of the base case. The capital cost of the HSCF is high in comparison to the base case



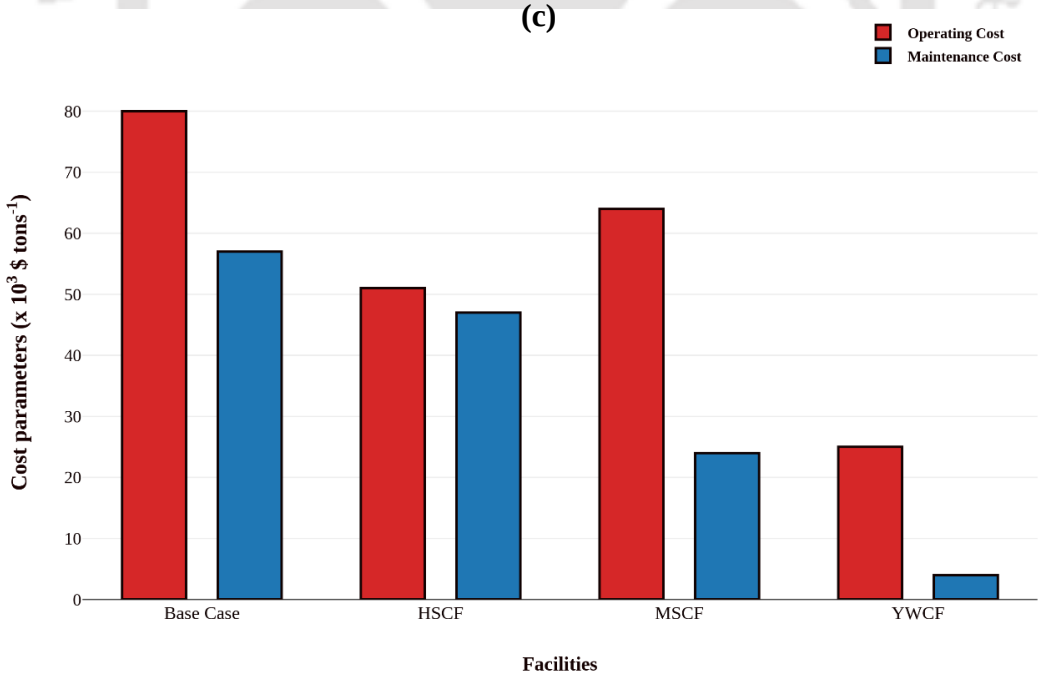
(a)



(b)



(c)



(d)

Fig. 10.5 (a) Cost contribution for the upgradation of the base case scenario to various composting scale units **(b)** Bar chart depicting capital cost of alternate composting systems **(c, d)** Operating and Maintenance Cost for base case scenario to various composting scale units

installations, and especially for the upgraded case (indicated in Fig. 10.5 (a)). This is due to the construction and installation costs. Henceforth, despite higher costs, it is a feasible solution due to the production of high-quality composts. Table 10.2 summarizes the capital, operating and maintenance costs data for all the three facilities. Certain assumptions were relevant for the cases in which accurate data were not available.

CAPEX refers to the total capital (money) required to acquire, develop, and set up a physical asset or infrastructure for a specific project or business (Fig. 10.5 (c) depicts the data for the present case). It is a one-time expenditure incurred at the beginning of a project or during the acquisition of a significant asset. CAPEX is typically associated with long-term investments and play a crucial role in determining the financial feasibility and viability of a project. Also, it's important to note that the OPEX and maintenance cost for composting facilities can vary widely based on factors such as location, technology, regulatory requirements, and specific facility design (Fig. 10.5 (d)).

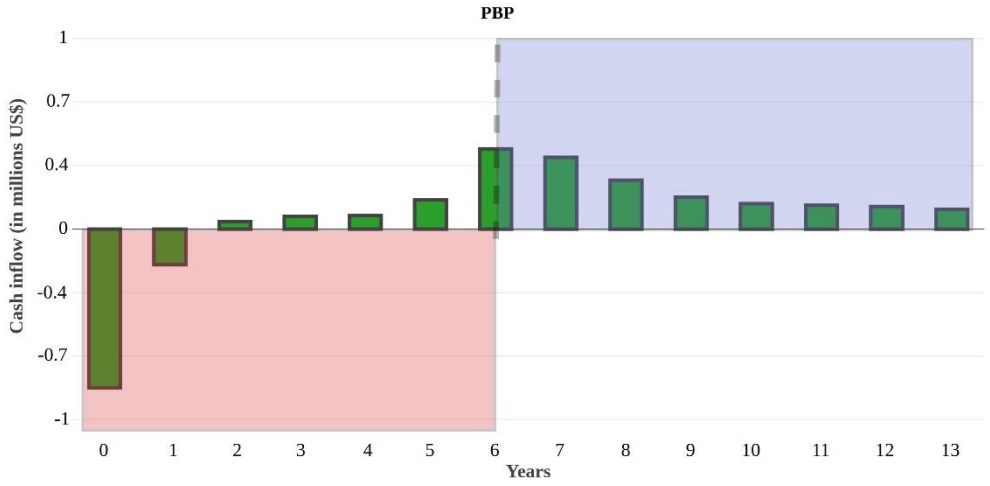
Table 10.2 Cost estimates of for all three types of MSW Compost Facilities.

Item/ Equipment	**Cost ($)**			**References**
	HSCF	***MSCF***	***YSCF***	
Paving	\$10000 $acre^{-1}$	\$3000 $acre^{-1}$	\$500 $acre^{-1}$	Parthan et al., 2012
Grading	\$850 $acre^{-1}$	\$500 $acre^{-1}$	\$280 $acre^{-1}$	Parthan et al., 2012
Land acquisition		\$20,000 $acre^{-1}$		Based on the Land value (Assam)
Construction cost	\$40 ft^{-1}			Parthan et al., 2012
Capital cost:				
Windrow turner	\$25000	\$10000	-	Komilis and Ham, 2004
Front-end loader	-	-	\$3000	Tchobanoglous et al., 1993
Odour-control system	\$60	\$50	-	Komilis and Ham, 2004
Screens	\$34000	\$14000	\$7500	Tchobanoglous et al., 1993
Hammermill	\$60000	\$35000	-	Tchobanoglous et al., 1993
Tub grinder	-	-	\$4500	Tchobanoglous et al., 1993
Maintenance	\$47000 $year^{-1}$	\$24000 $year^{-1}$	\$4000 $year^{-1}$	Mandal et al., 2014
Electricity	\$0.12 kWh^{-1}			Dadhich et al., 2017
Overhead cost	\$10 h^{-1} equipment operator (Overhead taken as 50% of manpower cost)			Assumed

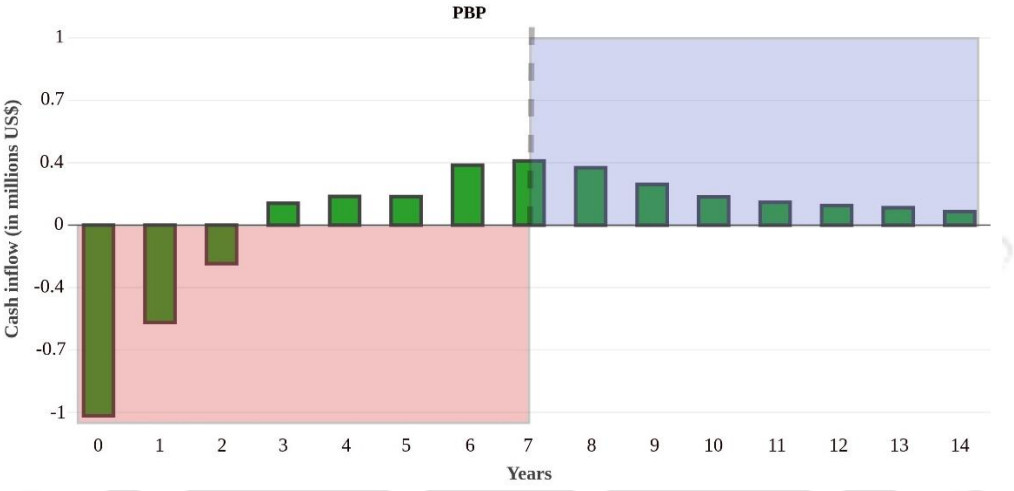
The compost generated from the composting plants is difficult to sell. It may not even serve as an organic fertilizer due to the public perception and lack of on field studies of such compost for fertilizer applications. However, given that the modest application could be that of an organic fertilizer, the selling price of the compost was fixed as Rs. 1500 per metric ton, a value being fixed by the Government

of India for organic fertilizer as per the Press Information Bureau (Press Information Bureau dated 30th June 2023, Internet website link: https://pib.gov.in/PressReleasePage.aspx?PRID=1936466). Thus, the equivalent bas price value of price has been chosen as $ 18 per metric ton for the economic evaluation. Nonetheless, since cost of compost may vary due to several factors including inflation, uncertainty in the vendors etc., the general range of compost price has been varied from 10 – 25 $ ton^{-1}.

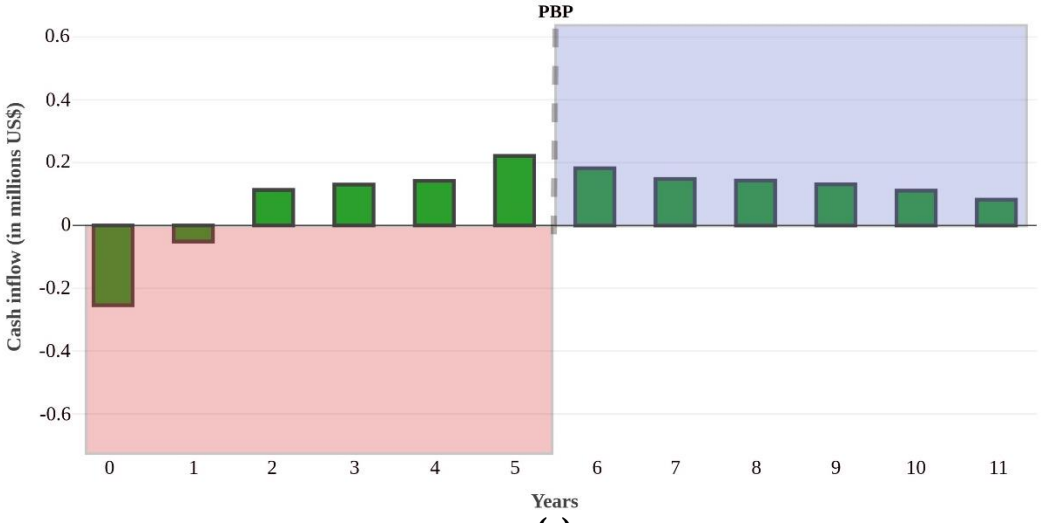
The profitability indices for the process were first based on an assumed selling price of \$10 $tons^{-1}$. For such an assumption, the profit margin of \$1.83 t^{-1} was achieved for the base production conditions. Fig. 10.5 (c, d, e) depict the PBP for HSCF, MSCF and YWCF respectively and Fig. 10.5 (f) depicts the cash flow of the investment for a 14-year period. The initial total investment was \$1.82 million, \$1.51 million and \$1.26 million for HSCF, MSCF and YWCF respectively. The analysis affirms a payback period of 5.5 - 7 years (Fig. 10.5(c, d)), and 13-year net present value (NPV) of approximately \$0.581 million at an internal rate of return (IRR) of 16.2% for HSCF and \$0.401 million at IRR of 13.3% for the MSCF. These figures convey a strong potential for the breakeven and full recovery of the initial investment in the fourth year after full operation commences, and increased profitability with prolonged plant life. To account for the uncertainty of the plant life, for the HSCF, the respective NPV and IRR for the operation after 7 years were approximately \$0.38 million and 14.02% and were \$0.22 million and 12.15 % respectively after 10 years. For the MSCF, these were \$0.39 million and 11.53% and \$0.26 million and 10.04% respectively. Accordingly, a very strong maintenance routine must be ensured to realize maximum plant life. Thus, reduced production downtime must be an essential factor in the maximization of the profitability of the process. Also, for the YWCF, Fig. 10.5(e) depicts a payback period of 5.5 years for YWCF with \$0.23 million NPV at IRR of 9.6%. After 6 years, the NPV will be \$0.18 million with an IRR of 8.4%. These will reach the values of \$0.11 million and 7.1% after 10 years.



(c)



(d)



(e)

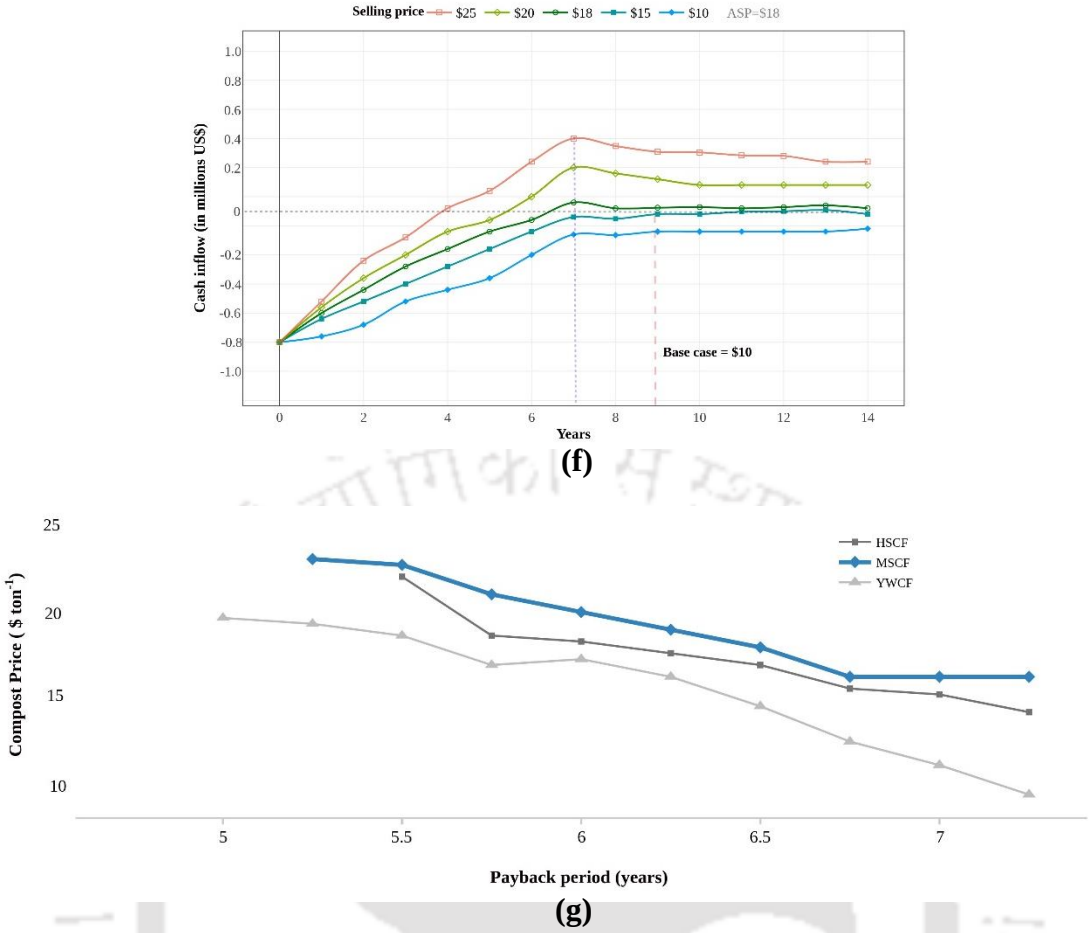
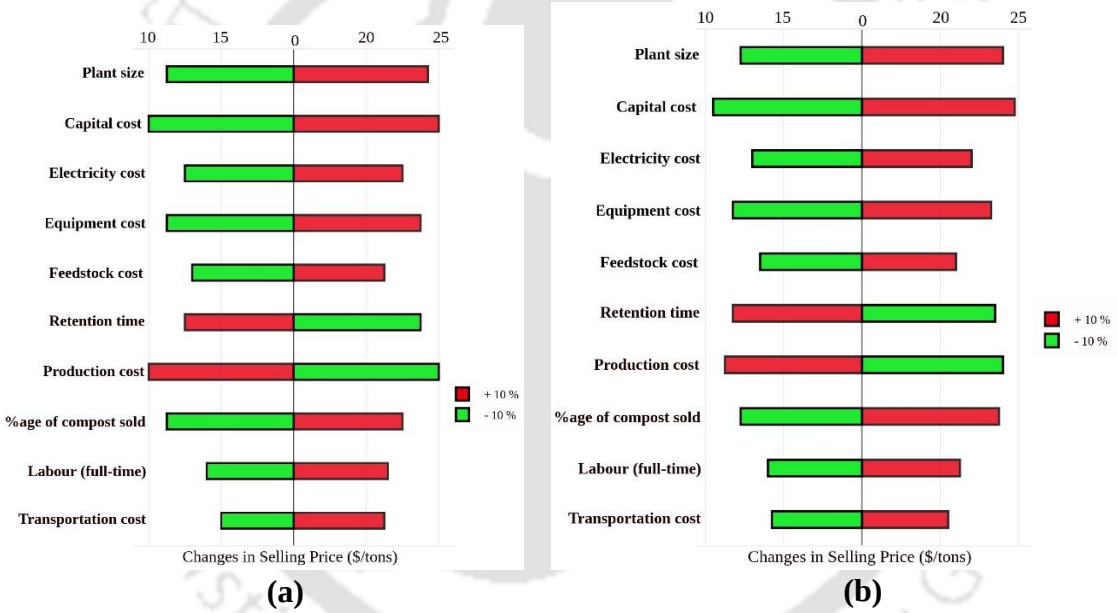


Fig. 10.5 Payback period for **(c)** HSCF, **(d)** MSCF **(e)** YWCF systems and for $ 10 tons^{-1} selling price and **(f)** Cash flow trends for various selling prices scenarios, average selling price (ASP) ($ 18 tons^{-1}) **(g)** Variation of compost Prices with PBP (Years)

Fig. 10.5 (f) depicts the cash inflows for various cases of the selling prices of the compost product. An examination of the financial data reveals that across a selling price spectrum ranging from $10 ton^{-1} to $25 ton^{-1}, the cash inflow for HSCF, MSCF, and YWCF compost facilities consistently registers as negative, indicating substantial losses (HSCF: -$0.8 million, MSCF: -$1.1 million, YWCF: -$0.28 million). However, a noteworthy trend emerges as cash inflow gradually transitions to positive values within an average timeframe of 2-3 years, suggesting a prospective turnaround in financial performance for these compost facilities. At an average selling price of $18 ton^{-1}, a notable shift in the financial trajectory is observed for the compost facilities. This change implies enhanced profitability and underscores the sensitivity of these compost facilities to variations in selling prices, with the $18 ton^{-1} point serving as a pivotal juncture for positive cash inflows.

Fig. 10.5 (g) depicts the payback period alteration with compost price ($ ton^{-1}). For a variation in compost price from 10 – 25 $ ton^{-1}, the payback period altered from 5.5-2.6, 7.1 – 5.5 and 6.3 – 4.6 years respectively for YWCF, MSCF and HSCF compost systems. The payback period for a base price of 18 $ ton^{-1} were 4.1, 6.2 and 5.3 years for YWCF, MSCF and HSCF compost systems. Thus, with greater payback period values, the composting technology may also require price subsidy to significantly reduce the payback period to about 3 years. If the compost selling price is bound to be even lower than 10 $ ton^{-1}, the payback period values will enhance beyond 5 – 7 years, which is an impractical scenario from an economics perspective.

10.7 Sensitivity analysis of the costs

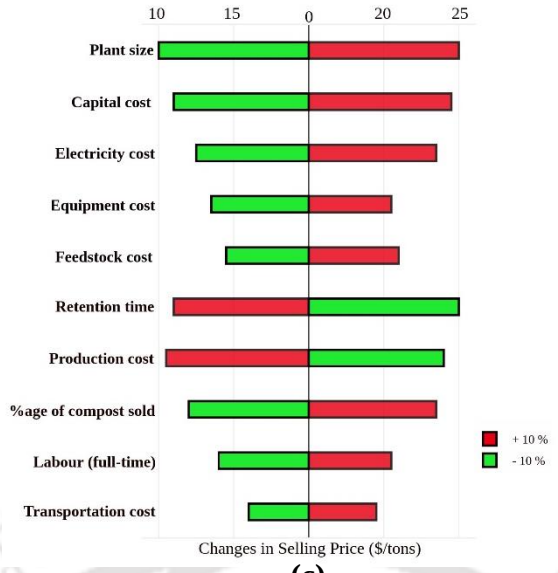


Since the techno-economic model was developed with few basic assumptions, sensitivity analysis was conducted to evaluate the sensitive influence of various entities on the production cost of the compost (Fig. 10.6 (a-c)). Thereby, opportunities for further improvement can be sought with respect to equipment, electricity, capital and transportation costs, RT, compost price and percentage of compost sold and labour charges (full-time). For a reduction in plant size from 10,000 tons yr^{-1} to 50 tons yr^{-1}, the profit reduced by $9 tons^{-1} for HSCF and MSCF cases (Fig. 10.6(a, b)) and by $5.8 tons^{-1} for the YWCF case (Fig. 10.6(c)). The estimated base capital cost was $1,20,000 for HSCF, $60,000 for MSCF and was $15,000 for the YWCF. The sensitivity analysis of ±10% range implied a low estimate of

$109,800 and a high estimate of $130,800 (HSCF), $54,900 and $65,400 (MSCF) and lastly $14,250 and $15,750 (YWCF). Similar analysis was obtained for other variables for each facility.

Reduced RT translates into number of composters, and thereby increases the profit. However, it will involve a trade-off with the equipment scheduling for the full utilization of the equipment. The enhanced equipment cost of the HSCF case by 10 % will increase the capital cost and reduce the overall profitability. Accordingly, net income gets reduced and breakeven period gets enhanced. On the other hand, if the equipment cost is reduced by 10 %, the capital cost of the plant will decrease. This leads to lower breakeven point and higher net income. Similar observations can be seen for MSCF (±10 %) and YWCF (±6 %). The ±10% range implied an alternation in plant size of $ 11.11 and $24.62 respective changes in the selling price per tons for HSCF. Similar estimations were conducted in terms of MSCF and YWCF.

Additionally, the composting was sensitive with respect to the compost price that accounted for the majority revenues. Overall, the sensitivity analysis depicted that the profitability of the process, being determined with minimum selling price (MSP) and NPV, was mostly influenced with the alteration in production and percentage of compost sold. Profitability has also been significantly sensitive with respect to feedstock costs. Thus, the important measures for the maximization of process profitability are (a) reducing the number and labour (charges) required for the operations, (b) reducing the costs associated with the feedstock by sourcing feedstock at lower prices, and (c) ensuring that the plant is located very close to the feedstock sources and thereby reduce the transportation costs. The analysis also indicated that proper maintenance of the equipment is required for the assurance of maximum productivity. Proper conditioning of the feedstock at its optimal constitution as well holds the key to optimize the process profitability. For the increased or reduced electricity cost by ±10 % (HSCF and MSCF) and ±6 % (YWCF), the plant costs increase and reduce respectively and thereby influence overall profitability and net income.



(c)

Fig. 10.6 Sensitivity analysis of the composting process profitability for **(a)** HSCF **(b)** MSCF and **(c)**YWCF systems

Overall, the results of the sensitivity analysis conveyed that plant size, capital cost and feedstock cost have a significant effect on each of the compost facility. For HSCF, a 10% increase in capital cost had the most significant influence on the financial performance of the compost facility. This resulted in a decrease in NPV by 8%. Similarly, a 10% increase in equipment cost resulted in a decrease in NPV by 6.5%. The influence of a 10% increase in other costs such as percentage of compost sold, labour costs, and transportation cost was found to be relatively small, with a reduced NPV ranging from 2% to 5%. On the other hand, a 10% reduction in the production cost resulted in a reduced NPV by 7.3%, while a 10% reduction in electricity cost had a minimal influence on the financial performance of the compost facility. This resulted in an increased NPV by about 2%.

Seasonal alterations foster sensitive dependence of the area requirements on temperature, precipitation and evaporation. Thereby, the composting area requirements increase by 20% in dew and cold seasons, by 10% in spring due to rains, by 10% due to precipitation. However, they reduce by 10% in the summer due to the heat (Rastogi et al., 2019). Precipitation influences area alterations by about 5-10% and due to the drainage adjustments (Zbinden et al., 2015; Arvanitoyannis et al., 2008). Thus, seasonal adjustments require alteration in few composting process parameters so as to facilitate optimal compost process performance in the entire year. Accordingly, the thesis considered the following as optimal conditions for the windrow composting system: 1.5 m windrow height; 2.5 m^2 windrow base area; 3

times per week turning frequency; 1.2 turning area factor. Temperature adjustments range from 5-10%, and henceforth favour insulation or heat dissipation. To determine the area needed for windrow composting in each month, volume per meter, turning frequency, and month-specific adjustments for temperature, precipitation, and evaporation are required. Thereafter, the obtained area per ton value is multiplied with the total amount of processed waste (say 50 TPD) to obtain the overall area required for the composting process. The area per ton is evaluated using the expression:

$$\textbf{Area per ton} = \frac{\textbf{Volume per meter}}{(\textbf{Turning frequency} \times \textbf{Turning area factor} \times \textbf{Month factor} \times \textbf{Temperature factor} \times \textbf{Precipitation factor})}$$

Thereby, the area requirements in every month have been determined and are presented in Fig. 10.7. As conveyed, for a waste processing capacity of 50 tons per day, the area requirements enhanced steadily from January till July (from 41.5 to 64.5 m^2) and thereafter declined upto the month of September. For the months October to December, the area requirements remained fairly constant.

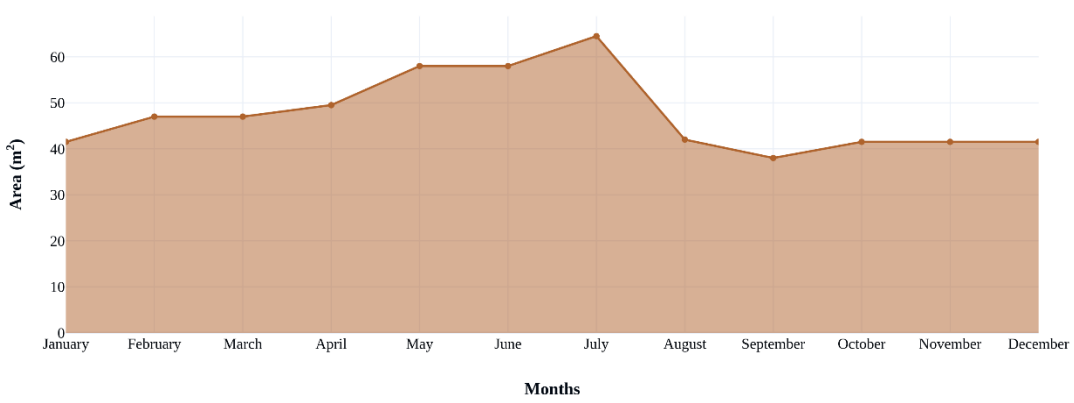


Fig. 10.7 Area requirement for composting with seasonal variations

10.8 ML based Prediction and Forecasting of CP Rate

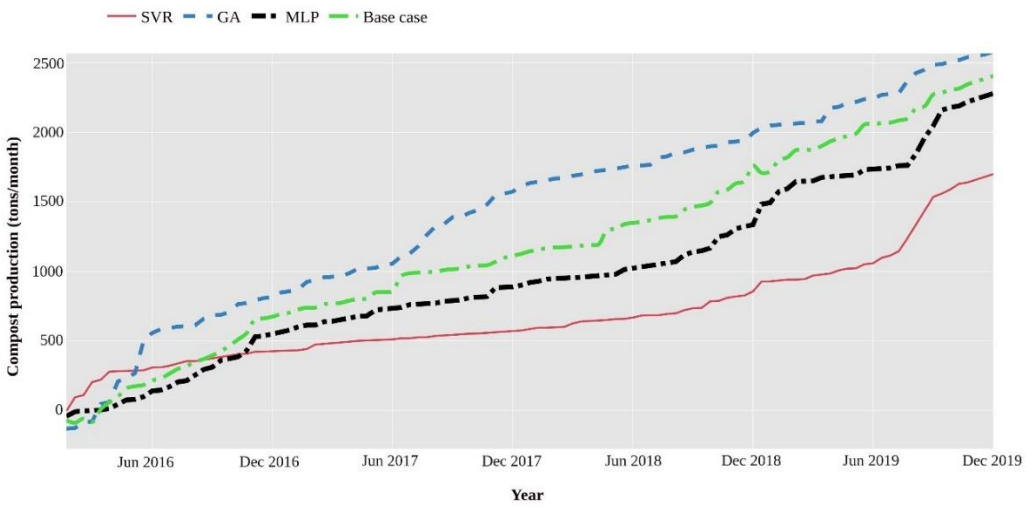
To address the competence of ML models for compost yield prediction, the ML model performance was assessed through a comparison of the calibrated model data with the compost yield achieved with the test dataset (200-sample dataset as measured data of the GMC unit) (Fig. 10.8). R^2 and RMSE were used as the metrics to examine the precision and accuracy of the model. In this study, hybrid MLP, SVR and GA models have been implemented. The SVR model was trained with an ε of 0.2 and a penalty parameter (C) value of 1. The SVR affirmed a train and test score of 0.76 and 0.72 respectively. For the MLP, the train (0.84) and test (0.82) scores have been consistently higher. The MLP was trained with

three hidden layers (each with a size of five), *tanh* as activation function and with a few other peripheral parameters. For the GA, chromosomes have been produced by randomly selecting location points from a pre-processed dataset and with a random selection method. Each chromosome had 20 bits size (each bit being a gene). The size of the primary population (100 chromosomes) did not alter during the algorithmic progression. The population-based fitness measure was used to choose the best performing chromosomes. As a result, the best chromosome is the one that has a repetitive mean value in multiple generations. Accordingly, the mean and 100 best chromosomes from 100 generations were together considered with the fitness value. For the case of fitness value becoming saturated, chromosomes were altered to boost fitness value. The best chromosomes have been chosen in terms of the fitness measure. The performance metrices R^2 and RMSE, for MLP, GA and SVR were 0.87 and 34.73; 0.82 and 103.22 and 0.76 and 231.51 respectively.

For the conducted evaluations, it was assumed that the payback period of the plant was 7 years. The hybrid ARIMA-MLP model performed well for the test dataset and with an R^2 of 0.94 (0.84 for ARIMA-GB). This represents a good predictive capacity for a model trained on real-world data. For the sake of comparison, an alternative model with the ANN model and SVR was also developed. ANN has been the most commonly deployed technique to predict the performance of compost facilities (Lin et al., 2016; Abdi et al., 2023). For the case, the most popular ANN (MLP) was deployed as the base scenario for comparison with both GA and SVR hybrid models. Accordingly, the ARIMA-MLP (R^2 = 0.94, RMSE = 23.22) outperformed ARIMA-GA (R^2 = 0.84, RMSE = 81.43) and ARIMA-SVR (R^2 = 0.67, RMSE = 383.94) models. The computation time was approximately 6 mins and was better than the computational time for ARIMA-GA and ARIMA-SVR algorithms. For forecasting purposes, the ARIMA was deployed with a 7 years payback period (2018-2024). The model forecasted well for all three facilities in comparison with the base case scenario (Fig. 10.9(a-c)).

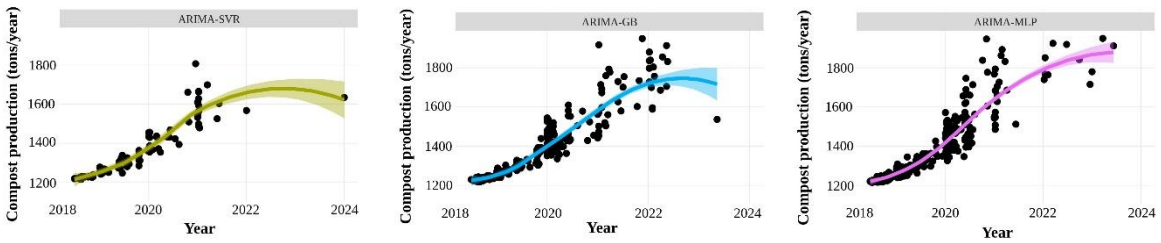
The autocorrelation feature generated a short-term correlation in the sequence. Henceforth, it was eventually subjected to the white noise test. The sequence used to fit the ARIMA model cannot be classified as white noise due to the result possessing a 0.05 significance level and lagged p-value being less than 0.05. To forecast the compost yield, the ARIMA (1, 1, 0) model was employed for the next 10

years forecasted period. Thus, the predicted and the forecasted results for the year 2021 have been correlated for each model. Accordingly, it can be observed that the ARIMA-MLP assured better result and thereafter, it was followed by MLP. As being depicted in the graphs, the future CP volume trends have been forecasted in the year range of 2020-2029 and using the predicted past data. The validation of forecasted CP with the observed value (2300 tons yr^{-1}) of CP data for the year 2018-19 have been summarized in Table 10.3. From Fig. 10.8(d), it can be inferred that while the HSCF can return more profits and revenues, it affirmed longer PBP due to the maintenance of the upgraded equipment and other capital and operating expenses. All these had a detrimental influence on the profitability. The alterations in the equipment and production costs of the compost could have a significant influence on the net income of the HSCF plant. Henceforth, it is important to carefully consider such factors and optimize the plant design and operation to minimize costs and maximize profitability.

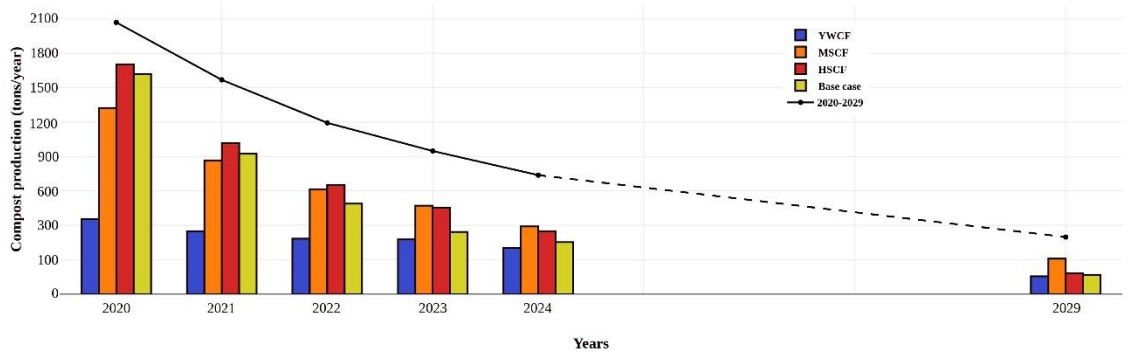


(a)

Fig. 10.8 ML algorithms-based CP rate prediction in the year range of 2016-2019



(a) **(b)** **(c)**



(d)

Fig. 10.9 Forecasted CP rates for **(a)** HSCF **(b)** MSCF and **(c)** YWCF and, **(d)** Long-term forecasting of CP rate for alternate composting systems in the year range of 2020-2029

Table 10.3 Predicted and forecasted CP rate and error summary values for 2018-2019 year range.

	Prediction			
Models	**Prediction (tons yr^{-1})**	**Standard error**	**95% Confidence interval**	
MLP	1300-2250	1.93	1270	2370
GA	2000-2600	19.27	1910	2765
SVR	800-2000	89.32	720	2180
	Forecast			
	Forecasts (tons yr^{-1})	**Standard error**	**95% Confidence interval**	
ARIMA-MLP	1750	5.22	1620	1860
ARIMA-GA	1600	34.18	1530	1820
ARIMA-SVR	680	101.43	570	890

Table 10.4 (a) A subjective summary of alternate techno-economic calculators for the composting case study.

	Composting Calculator by TERI	Compost Production Cost Calculator (IFDC)	Composting Financial Model (AEC)	Compost Calculator designed by IIT Delhi	This Work
Purpose	Helps to estimate the amount of compostable waste generated, compost quality, and nutrient content of compost.	The optimal blend of feedstocks based on C:N ratio have been calculated	The financial feasibility of a composting operation has been estimated	The amount of materials needed for a compost pile has been calculated	This work helps to estimate the compost generated from small scale (yard) facility, medium scale (MSCF) to large scale (HSCF) from OW
Input variables	Quantity and type of waste being composted MC, pH, and T of the compost, nitrogen, phosphorus, and potassium content of the compost.	Amount and type of compostable waste, compost quality parameters, nutrient content	Feedstock type and quantity, equipment and labour costs, utilities	Capital expenses, operating expenses, revenue, production volume	RT, T, MC, equipment and operational costs (electricity, capital, operating, maintenance, transportation, labour), compost price, odour control system, area acquisition, and cost overhead
Output variables	Amount of compost generated, nutrient content	Estimated amount of compostable waste generated, estimated compost quality, estimated nutrient content of compost	Total cost of CP, net profit	NPV, IRR, payback period, cash flow	CP, cash inflow, payback period, IRR, NPV, profit analysis, project profitability and other performance metrics
Inferences	It does not provide information on other factors such as odor control or disease suppression.	Limited to estimating the amount of compostable waste generated and nutrient content of compost	Does not account for all factors that may affect the cost of CP	Does not estimate all factors that may affect the financial viability of a composting operation	The nutrient content, C:N, pH factor have not been estimated in this calculator.

Table 10.4 (b) Summary of the best performing ML models for various studied composting processes

Feed	Input parameter	Prediction parameter	Data size	Model	Prediction performance		Training: Testing ratio	Authors
					R^2	*RMSE*		
Food waste, paper waste, yard waste	RT, T, feedstock quantity	CP Forecasting	200	MLP: GA: SVR: ARIMA-MLP ARIMA-GA ARIMA-SVR	0.87 0.82 0.76 0.94 0.84 0.67	34.73 103.22 231.51 23.22 81.43 383.94	70:30	This work
Food waste, yard waste, sewage sludge, and paper mill sludge	Feedstock quantity, MC, and C/N ratio	Compost yield	302	Random Forest (RF)	0.92	2.63	60:40	Wan et al. (2022)
Vegetable waste, Straw, Livestock manure	Images of different composting stage	Compost maturity	9403	Deep learning: CNN and SVR	-	-	80:20 Accuracy: 99.5%	Xue et al. (2019)
Maize straw waste	Colour and texture features of the compost	Compost maturity	1536	Artificial Neural Network (ANN)	0.73	83.88 Classification error: 1.56%	70:30	Kujawa et al. (2014)
Food waste	T, MC, VS, GI	CP rate	300	Wavelet Neural Network	0.82	4.81	80:20	Gao et al. (2007)

Table 10.4(a) and (b) respectively present an overview of the techno-economic calculators being reported in various studies and in this work. A striking feature of the techno-economic calculator of this work has been in terms of scalability and consideration of several technical and economic factors. The work did not delineate upon the nutrient content, C:N ratio and pH. Such studies need to be addressed in the near future as data consolidation remains a stiffer challenge for such objectives. Table 10.4(b) infers that the R^2 for ANN was not better in comparison to that inferred by Kujawa et al. (2014). For this, the reason could be that the neural networks could not justify or provide the basis for their reasoning process. Since ANN estimate the pertinent process alterations with empirical data, an insight into the understanding of the nature behind the alteration complexity cannot be explored at times to obtain

mechanistic insights into the compost process (Guo et al., 2021). The accuracy of the SVR model however resulted to be low in conjunction with the CNN (Xue et al., 2019) as SVR could have suffered sensitivity to the choice of hyperparameters. Henceforth, the resultant model has been more time-consuming and involved significant computational resources. Thus, to resolve this issue, a hybrid ML model has been proposed in this work to model compost process data for the forecasting of the compost yield. The performance metrics resulted in decent metrics (R^2 = 0.94 and RMSE = 23.22). In summary, the MLP performance has been comparable with that of Wan et al. (2022) that considered similar type of parameters and output variables.

10.9 Implications of the Study

The implementation of resource recovery and waste recycling is very crucial for the realization of a circular economy concept. For the quicker transition from a linear to a circular economy, Waste-to-Compost (WTC) technologies can ensure effective management of a country's solid waste. Addressing sustainable development, the case study for the Guwahati city demonstrated the strength of the techno-economics of composting technologies and their relevance for the effective handling of huge quantities of waste in the Asia-pacific region.

The case study in this Ph.D. work focussed on Assam, India, and especially the Kamrup region. This is due to it being one of the most populous provinces and the economic center of north-east India. The district has a population of approximately 2 million, and an annual waste generation rate of 2.1 kg person^{-1} day^{-1}, the total amount of organic waste generated annually would be 673,750 tons. If a composting facility is to be set up to manage this waste, the appropriate scale of the facility would depend on several factors, such as the available space, budget, and other environmental constraints.

Since 2020, Kamrup lacks waste conversion facility, a feature that ensures sustainable environment through the effective waste management and promotion of environmental sustainability. Through the commissioning of such waste conversion facilities, municipalities and states can manage their waste more efficiently and accordingly reduce their dependence on fossil fuels, and promote the circular economy. Furthermore, these facilities can provide economic and social benefits by creating jobs,

reducing landfill costs, and providing a source of renewable energy. Utilizing, multi-criteria analysis via ArcGIS, the study examined the potential of waste conversion facilities and their capacities based on segregated and mixed waste along with the consideration of environmental costs and benefits.

The study also concludes that Kamrup has the potential to implement WTC projects that provide adequate financial and economic returns. Being environmentally attractive, such measures can support a circular economy with the greater utility of the proposed techno-economic calculator for the estimation of CP volumes with relevant combination of process parameters. Techno-economic calculators can serve composting operators as a very useful tool to evaluate the economic feasibility and environmental benefits of their operations. This, in turn, can help to optimize the resources utilization, waste minimization and product value maximization. Through the promotion of resource efficiency and waste reduction, the composting can contribute to a more circular and sustainable economy. It also depicts the viability of a centralized compost facility in the capacity range of 1000 to 10,000 tons $year^{-1}$. Through the implementation of such development projects, waste management issues in the region can be largely eliminated along with the assurance of economic and social benefits.

Also, in this study, the utility of ML as predictive models for compost prediction has been proven for accurate prediction of CP. Thereby, the methodology can assist decision-makers in the identification of the optimal conditions for the composting system. Thus, it can help identify potential issues and resolve them prior to their evolution as significant problems. Several recommendations have been provided by the models, and these can be followed by the compost producers for the production of high-quality compost and accordingly realize increased revenue and customer satisfaction. Also, since ML predictive models for compost prediction can provide valuable insights, they can be also used to make informed decisions, optimize resources, and improve the efficiency and profitability of composting operations.

10.10 Findings and Summary

After conducting a thorough TEA of composting, it can be concluded that composting is a viable and sustainable waste management option. In this case study, three MSW composting plants i.e., HSCF, MSCF and YWCF were assumed with a processing capacity of 25 tons, 15 tons and 3-5 tons day^{-1} respectively. From the analysis, it can be concluded that for the setting up or upgrading of a HSCF

(similar to that of the base case), the total capital investment required is around $0.5 million. This includes the cost of land, construction, equipment, and other miscellaneous expenses. The operating cost for the plant has been estimated to be $600,000 $year^{-1}$. This includes labour, maintenance, and other operational expenses.

The projected revenue (annually) from the sale of compost has been about $900,000. This revenue is based on the assumption that the compost is sold at $40 ton^{-1}, which is a conservative estimate based on the current market price. From these figures, the NPV of the composting plant for 8-year period has been $0.581 million at an IRR of 18.2 % and a PBP of approximately 3.5 years. These values convey that the composting plant is a financially feasible investment, and has a relatively short PBP and a positive NPV.

The analyses demonstrated that HSCF compost has strong potentials for profitability with an estimated production cost of $51 t^{-1}, and market value ranging between $80 and $150 t^{-1} of the product. The sensitivity analysis exhibited that the RT, feedstock, and equipment are the most significant factors to influence production cost. Measures to minimize these costs are, therefore, key to maximize the process profitability.

In conclusion, the TEA approach in this work demonstrated that composting is a viable and sustainable waste management option with both financial and environmental benefits. The composting plant in this case study has a relatively short payback period and a positive NPV. This affirmed that it is a financially feasible investment. Moreover, composting is an environmentally friendly solution that can significantly reduce GHG emissions and promote soil health. Overall, the study not only established potential benefits related to mitigation of environmental issues but also ensured higher potentials for enhanced profitability of the composting plants in Guwahati, India.

Chapter 11
Conclusions and Future Work

Chapter 11

Conclusions and Future Work

In the chapter, section 11.1 provides a brief summary of the key findings achieved from the conducted investigations in the PhD thesis. These have been presented in different sub-sections and each of these met the objectives of the PhD thesis. Furthermore, section 11.2 outlines a comprehensive summary of potential avenues for future research in the field of MSWM. Accordingly, scientific advancement and development for an effective and pragmatic waste management strategy can be realized.

11.1 Conclusions

Firstly, the PhD. thesis addressed detailed investigations that targeted the quantification, and characterization of MSW in the Guwahati city. Thereafter, the application of ML techniques and time series models enabled an accurate prediction and forecasting of the total MSWG rate, OW and recyclables waste generation patterns. All these assists in the efficient waste management strategies for the city. Also, ML techniques have been deployed to model and simulate the performance of composting and biogas processes under altered climatic conditions. Thereby, they enabled efficient planning and targeted optimized compost and biogas operations for the mitigation of potential environmental impacts. Finally, an AD calculator being deployed along with ML techniques allowed optimization of BP rates and the integration of GIS data with a decision model and a techno-economic calculator served as a comprehensive toolset for the evaluation and performance optimization of compost facilities. All these can lead to enhanced operational efficiency, increased energy generation and cost-effective operations in biogas and compost plants respectively. The following sub-sections delineate upon the most relevant objective and subjective findings of eight major objectives being set and realized in the PhD. thesis.

11.1.1 Socio-demographic Groups-based Characterization, Quantification and Seasonal Classification of MSW Rate

- The waste disposal classification analysis conveyed that 51.62% of the solid waste was OW, 24.22% of it was package waste (papers and plastics), 8.13% of it was burnable waste (clothes, diaper, rug, bag, shoes, slippers, pillow, carpet), 4.62% of it was park and garden waste, 3.54% of it was metals and glass waste, and 7.87% of it was miscellaneous.

- For all three socioeconomic (poor, medium and high income) categories, a daily average of 0.201 kg of garbage was produced per person. The findings depict that more income leads to enhanced solid waste generation. The per capita generation rates were 0.18, 0.23, 0.17, and 0.516 kg capita^{-1} day^{-1} for A1 (middle income), A2 (high income), A3 (poor income) and commercial area, respectively.
- In comparison with the best available prior art (Suthar and Singh, 2015; Kamran et al., 2015), the waste generation patterns in different socioeconomic groups and commercial areas, have been comprehensive and improved due to the methodology that involved a survey, correlation analysis, and statistical analysis. The work also examined various factors, and seasonal variations. The findings revealed a strong correlation between the number of HH and population growth with waste generation. Additionally, the thesis inferred a reduced waste generation from August to January, followed by a gradual increase from the month of June.

11.1.2 Prediction and Forecasting of Total MSWG Rate

- Firstly, population growth, HH counts, LP, GDDP indicated strong positive correlation index. Secondly, among all algorithms, the RF and GB with response train and test scores of 0.90 and 0.94 respectively performed well and among these two, the GB had a model accuracy of 97%. Thirdly, critical novelty referred to the overfitting issues of the DT structures being mitigated with the grid search 10- fold cross validation (a hyper tuning approach) and forecasting using the MA based approach. Thereby, this work uniquely contributed to overcome overfitness issues of DT to model non-linear data and lower learning speed with the HPO for the RF model.
- The performance of the GB model (r = 0.94, R^2 = 0.99, RMSE = 3.01, MAE = 2.86, and IoA = 0.94) was superior to that of the RF model (r = 0.90, R^2 = 0.98, RMSE = 83.21, MAE = 74.84, and IoA = 0.72) and the DT model (r = 0.82, R^2 = 0.97, RMSE = 325.82, MAE = 302.20, and IoA = 0.45). Additionally, the R^2 and RMSE values for EMA-GB and SMA-GB were 0.998, 2.11 and 0.994 3.83, respectively, and henceforth affirmed their statistical significance. Conversely, the RMSEs for EMA-DT and WMA-DT were 4.22 and 11.45, respectively, and

indicated notable difference among the models. Consequently, the GB model demonstrated the best performance for the prediction of the MSWG rate in the Guwahati city.

- In this work, the ML methods have been effective and outperformed with respect to those reported in the prior art (Kannangara et al., 2018; Dissanayaka and Vasanthapriyan, 2019; Johnson et al., 2017). Moreover, the ML models simultaneously minimized prediction uncertainty and enhanced prediction vigor. Besides these, the ML models can provide insight into decision-making for stakeholders, and open pathways for the systematic improvisation of pragmatic solid waste environmental pragmatic policies.

11.1.3 ML and Time Series Models-based Prediction and Forecasting of OW and Recyclables Waste Rate

- For the realization of the set objective, ML and hybrid ML models were applied to predict and forecast organic and recyclables generation rate respectively in the Kamrup metropolitan region. Using socioeconomic and demographic data, the developed models were able to predict municipal-level recyclables and OW solid waste generation rates. The modelling was based on an integrated dataset consisting of socioeconomic and demographic information from the Indian census, annual solid waste generation (organic) data for the HH level (in the Kamrup metropolitan under GMC), and data obtained after elimination of the diversion of plastics and paper.
- For effective prediction, the MLP, GA, SVR, GB, and *k*NN algorithms were deployed. Compared to all other approaches, the neural network approach affirmed best performance. Using the test data, it generated MSWG models with an overall 92% accuracy rate. However, plastics and paper diversion model performed less well (error rate of 18.21% in the sample data). The MLP algorithm affirmed train and test scores of 0.85 and 0.82, respectively, and performed well with a model accuracy of 95%.
- For both GB and MLP models, overfitting issues in both the GB and MLP models were addressed through grid-based search, tenfold cross-validation. Thereafter, and the forecasting was addressed with a MA-based approach. The thesis ensured contributions in terms of the overfitting problems associated to the modelling of nonlinear data and enhanced the learning

speed with the HPO. Overall, the MLP model exhibited superior performance for OW (R^2 = 0.92, RMSE = 10.43) and recyclables (plastics and papers) (R^2 = 0.92, RMSE = 18.21) in comparison to other considered algorithms. The hybrid ARIMA-MLP models exhibited statistically significant distinctions with a 95% confidence interval, and with standard errors of 5.67 and 10.93 for OW and recyclables generation rate respectively.

11.1.4 Prediction and Forecasting of GHG and PMs Emissions

- The objective involved the evaluation of GHG emissions-based air pollution from solid waste sent to landfill and incinerators. They conducted measurements and analyzed the concentrations of major air pollutants from 1970 to 2018. Available data refers to the conducted measurements-based concentrations of major air pollutants in the year range of 2018. Accordingly, the study compared the results obtained for nine ML algorithms and six indices (CH_4, CO_2, N_2O, SO_2 fluxes, and PMs emissions).
- To determine the most significant predictors and their interrelationship for GHG emissions prediction, various analyses such as correlation, stepwise regression, and NCA were performed. The researchers found that atmospheric T_{atm}, RH, P, and Pr were the most influential parameters. The stepwise regression (cubic) yielded R^2 and RMSE values of 0.94 and 23.2 for CH_4; 0.96 and 9.5 for CO_2; 0.85 and 45.6 for PMs; 0.78 and 31.4 for N_2O and 0.61 and 49.2 for SO_2 respectively.
- Additionally, the thesis examined the seasonal variation of $PM_{2.5}/PM_{10}$ and identified extreme episodes. The concentrations of PM exceeded the daily limit value (60 µg m^{-3}) by about 11 times. The highest levels were observed during the coldest period of winter ($PM_{2.5}$ = 96.22 µg m^{-3}; PM_{10} = 138.73 µg m^{-3}). Each pollutant displayed a clear seasonal pattern. Particularly for PM_{10}, higher concentrations in winter have been achieved in comparison to other seasons of the year (summer: $PM_{2.5}$ = 25.62 µg m^{-3}; PM_{10} = 38.16 µg m^{-3}).
- In general, the MLP algorithm exhibited best performance accuracy in comparison to other algorithms, (R^2 = 0.97, RMSE = 0.32, IoA = 0.94, MAE = 28.32, MAPE = 49.21, and SSE = 0.21). For forecasting purpose, an ARIMA configuration of (1,1,0) × (0,1,1) was found to be

acceptable, but it had issues with negative spikes at lag 12, (required 5 iterations). Nonetheless, it yielded significant R^2 and RMSE values of 0.89 and 5.67, respectively. Also, the developed model can even effectively predict HCl, HF, Dioxin concentrations, and NO_2 emissions. However, the same has not been addressed in the thesis due to the non-availability of hourly monitored datasets from EDGAR database. This conveys upon the need for robust data sources for comprehensive environmental predictions in the Guwahati city.

11.1.5 CP Rate Prediction with Alternate ML Models

- The availability of reliable regional-level data is crucial for effective SWM planning. However, many developing regions do not have trustworthy data. The thesis aims to predict the CP rate of Guwahati city by utilizing publicly available data and appropriate data analytics. Five parameters, namely T, RH, WS, Pr, and OW, were considered as independent factors. The analysis revealed that T (with an R^2 of 0.83) and OW (with an R^2 of 0.74) have significant explanatory power for the variability in CP, and RH (with an R^2 of 0.18) did not convey much significance.
- To predict CP, popular ML models such as *k*NN, MLP, RF, and GB were compared based on their strengths and weaknesses. Among these models, GB produced promising results (R^2 of 0.99 and RMSE of 0.757) followed by MLP (R^2 of 0.938 and RMSE of 13.37). However, upon the deployment of a randomization technique, the MLP outperformed other models (RMSE of 9.32, R^2 of 0.95, and IoA of 0.87, resulting in an overall model accuracy of 90%). Also, the high-end ML algorithms yielded highly accurate findings and these are similar to the previously reported datasets (Kujawa et al., 2014; Boniecki et al., 2013).
- Seasonal analysis indicated that the summer season (April to June) exhibited favourable conditions for compost generation, (characterized with a T_{max} of about 30 ± 2°C and RH above 80%). For long-term CP forecasting, an autoregressive integrated moving average (ARIMA) model was deployed, and the ARIMA-MLP model demonstrated better performance in comparison to *k*NN, GB, and RF models.

- Accurate prediction of CP is crucial for sustainable and efficient SWM. The thesis demonstrated the effectiveness of various ML models for CP rate prediction and provided novel insights through the evaluation of *k*NN, RF, GB, and MLP algorithms. The research also highlighted the applicability of ARIMA models for long-term forecasting. The findings suggest that the proposed methodology can be adopted for other cities having similar development and environmental challenges.
- The PhD. thesis demonstrated the importance of ML models for CP rate prediction and presented significant findings. The optimization of tree-based ML models through parameter fine-tuning resulted in higher test scores and improved predictive accuracy. While hyper-tuned RF exhibited linear prediction, the GB model's performance improved with enhanced learning speed.
- The proposed methodology is generic in nature and can be applied to predict CP rates for any city and for complex input datasets and appropriate modifications of influential parameters. Accuracy analysis indicated that each model curve grew quickly during the initial stages. Accordingly, *k*NN, RF, GB, and MLP achieved accuracies of 80%, 88%, 93%, and 95%, respectively. MLP and GB have been recommended for their prediction speed and accuracy in handling complex datasets. Ultimately, these findings can support the development of an integrated decentralized community-based SWM approach, enhanced recycling, and composting practices to foster a circular bio-economy in the Guwahati metropolitan region.

11.1.6 Prediction and Forecasting of BP Rate

- The thesis deployed appropriate data analytics and publicly available data to predict BP rate for the Guwahati city. Five parameters namely T, P, RH, Pr and OW were used as independent factors. The data analysis suggests that T (0.81) and OW (0.65) account for the majority of the variation in BP (since they have the best R^2). Further, RH (0.26), P (0.17) and Pr (0.091) does not exhibit much significance in the BP rate.
- The thesis work also contributed to the efficacy of alternate estimation of BP rates. In comparison, the MLP followed by GB produced promising outcomes for BP prediction. Also,

for the forecasting of BP, ARIMA and MA were deployed and thereby both ARIMA-MLP and MA-MLP models performed well. The performance indices of ARIMA-GB and EMA-GB affirmed RMSE values of 7.94 and 11.53, respectively and the overall model accuracy was found to be 84%.

- Seasonal analysis also confirmed that the favourable conditions for biogas generation have been in the summer season (Apr-Jun with T_{max} value of about 30 ± 3 °C and RH value above 80%). The forecasting approach can be utilized as a model for other locations that dealt with similar developmental and environmental challenges. A hybrid modelling effort that systematically accommodates simultaneous lab scale investigations with the modelling efforts can be envisaged with the MLP due to its promising performance.

11.1.7 Optimizing BP Rate with AD Calculator and Application of ML Models for Biogas Plant Efficacy

- For the objectives, various techniques such as techno-economic calculations, modelling, and simulation to assess the AD process and upgradation scenarios were deployed. The results obtained were consistent with the literature data.
- The investigation focused on the evaluation of the influence of input variables such as feed flow and RT on biogas yield. Relevant prominent scores were deployed to identify the variables with the greatest impact. The developed AD calculator in this PhD. work has been proven to be effective for BP rate determination for both short-term and long-term AD operations. The model's accuracy was validated through a comparison of the predictions reported for other AD units (with results within a ±10% range).
- The direct sale of biogas from smaller digesters (60 m^3 day^{-1}) faces commercial challenges in India due to dominance by large-scale producers and limited production volume. Economic evaluation indicates that a base biogas price of \$0.25 m^{-3} results in a negative net profit, emphasizing the need for upgradation in digester plants. Higher prices, achievable at \$0.50 and \$0.75 m^{-3}, require price subsidies in waste management policies to ensure sustainability. Optimal operation may be viable in community projects driven by philanthropy or corporate social responsibility.

- The experimental data encompassed 350 samples, and three models, namely ARIMA-MLP, MLP, and SVR, were compared. Among these, the ARIMA-MLP model achieved the highest R^2 value of 0.89, followed by MLP (with an R^2 of 0.82), and SVR (with the lowest R^2 of 0.57). The corresponding RMSE values were 5.83, 140.27, and 854.31, respectively. The training to testing data ratio was set at 70:30. However, it can be inferred that even though R^2 and RMSE for MLP have not conveyed better performance in comparison to those inferred by Beltramo and Hitzmann, 2019, the hybrid modelling effort outperformed the MLP with better model accuracy.
- Further, the management of supernatant and digestate is beyond the scope of the PhD. thesis. However, it can be conceptualized that mature eco-friendly technologies would be available at an affordable price for the digestate and supernatant. Their cost implications with minimal public interest to purchase processed byproducts along with even modest operating and capital costs for the sub-processes do convey that the biogas process plants would have even stringent economic viability scenarios. Thus, sustainable ecology and environment as the primary cause shall be considered for the adequate implementation of smaller and even larger biogas plants with the existing biogas price index value. Otherwise, biogas price subsidy shall be provided by the Government of India to further enhance sustainability of the BP plants.

11.1.8 GIS-based Decision Model and Techno-economic Evaluation of Compost Facilities

- For the objective, the proposed case study considered three alternate MSW composting plants namely HSCF, MSCF, and YWCF. Their respective daily processing capacities were 25 tons, 15 tons, and 3-5 tons. Based on the conducted analysis, it can be concluded that the setting up or upgradation of a composting plant similar to the HSCF plant described in this case study would require a total capital investment of approximately $0.5 million. This investment includes expenses such as land acquisition, construction, equipment, and miscellaneous costs. The annual operating cost for the plant was estimated to be $600,000 (covering labor, maintenance, and other operational expenses).

- The projected revenue from compost sales, was estimated to reach an annual value of $900,000. Such a projection assumed a conservative selling price of $40 ton^{-1} and was based on the current market price. Considering these figures, the NPV of the composting plant for an 8-year period was determined as $0.581 million, and with an IRR of 18.2%. The PBP for the investment is approximately 6 years. These results indicate that the composting plant is a financially viable investment. This is due to relatively short PBP and a positive NPV.
- In general, the OPEX of a higher-capacity facility is often lower on a per-unit basis in comparison with the medium or small (yard) facility in the composting industry. This is due to factors such as economies of scale, efficiency gains, bulk handling, technology adoption, and regulatory compliance. Also, numerous small-scale digesters have been successfully implemented in various solid waste industries. They provide a promising solution for waste management. In the context of HSCF, MSCF, and YWCF systems implementation in India, the thesis focussed on proposed capacities of 500 TPD, 100 TPD, and 50 TPD, respectively. For each of these systems, the operational expenditures (OPEX) have been evaluated and were $119,000 per ton, $129,000 per ton, and $99,000 per ton for HSCF, MSCF, and YWCF, respectively.
- Moreover, it is essential to consider that while larger facilities may benefit from economies of scale, there are other factors that can influence operational costs. These include the specific technologies used, regional variations in costs, and regulatory environments. Henceforth, the operational cost of a higher-capacity composting facility is often lower per unit of output in comparison to medium or small facilities. This is due to economies of scale and operational efficiencies. However, the actual cost structure can vary based on several factors, and each facility's unique circumstances should be considered for the comprehensive analysis.

11.2 Future Work

The following areas of work have been identified for consideration in the near future to further encourage pragmatic and customized subjective research in the vast field of sustainable MSWM system in the targeted developing countries.

- Future studies need to target more sophisticated and complex survey-based data along with a wider variety of demographic and socio-economic data of MSWG to supplement and verify ML algorithm's competence in terms of updation, retraining and CPU time marginalization.
- The consumption and waste generation patterns of major socio-economic categories can be considered as future studies together with comparable socio-economic aspects that are as well important for solid waste characterization and quantification.
- The proposed tree-based model could be also strengthened by choosing the most critical input elements from the dataset and with dimensionality-reduction techniques such as PCA and NCA.
- Variables related to socio-economic factors, land-use patterns, and policy interventions should be also integrated with the GHG prediction model to improve its predictive competence and comprehensive insights into potential factors shall be sought for their contribution to GHG emissions and PMs and their pragmatic mitigation strategies.
- More historical data on meteorological parameters such as WD and solar radiation should be targeted as future studies for CP prediction. Such research could also target alternative optimization algorithms for the effective training of the ML models for better CP rate prediction.
- The ML can enable real-time monitoring and control systems for biogas plants. By integrating meteorological data, OW characteristics, and process parameters, the ML algorithms can predict BP in real-time. Such information can be used to adjust process variables, such as T, pH, and feedstock composition and thereby achieve BP rate maximization.
- Enhanced predictive model performance for AD systems can result in increased operational efficiency. This can lead to the feedstock selection for higher yields, betterment in risk assessment and mitigation driven with the prediction of potential issues, cost reduction based on process scale up and optimization and identification of optimal conditions, design optimization and improved understanding of waste inputs and operating conditions for a better sustainable OW valorization industry.

- In the near future, investigations targeting the effectiveness of pH, nutrient content and C:N ratio in the composting process shall be studied. This will explore and ascertain the possibility to reduce treatment costs, reduce the cost of feedstock acquisition and accordingly maximize the process profitability.

References

Abbas, M. A., Iqbal, M., Tauqeer, H. M., Turan, V., & Farhad, M. (2022). Microcontaminants in wastewater. In Environmental micropollutants (pp. 315-329). Elsevier.

Abbasi, M., & El Hanandeh, A. (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. Waste management, 56, 13-22.

Abdi, R., Shahgholi, G., Sharabiani, V. R., Fanaei, A. R., & Szymanek, M. (2023). Prediction compost criteria of organic wastes with Biochar additive in in-vessel composting machine using ANFIS and ANN methods. Energy Reports, 9, 1684-1695

Abdoli, M. A., Falahnezhad, M., & Behboudian, S. (2011). Multivariate econometric approach for solid waste generation modeling: Impact of climate factors. Environmental Engineering Science, 28(9), 627-633.

Abdoulaye, T., & Sanders, J. H. (2005). Stages and determinants of fertilizer use in semiarid African agriculture: the Niger experience. Agricultural Economics, 32(2), 167–179

Abraham, B., & Ledolter, J. (1986). Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. International Statistical Review/Revue Internationale de Statistique, 51-66.

Abushammala, M. F. M., Noor Ezlin Ahmad Basri, Basri, H., Ahmed Hussein El-Shafie, & Kadhum, A. A. H. (2010). Regional landfills methane emission inventory in Malaysia. Waste Management & Research, 29(8), 863–873. https://doi.org/10.1177/0734242x10382064

Abushammala, M. F., Basri, N. E. A., Elfithri, R., Younes, M. K., & Irwan, D. (2014). Modeling of methane oxidation in landfill cover soil using an artificial neural network. Journal of the Air & Waste Management Association, 64(2), 150-159.

Adjuik, T. A., & Davis, S. C. (2022). Machine Learning Approach to Simulate Soil CO2 Fluxes under Cropping Systems. Agronomy, 12(1), 197. https://doi.org/10.3390/agronomy12010197

Afon, A. O. (2007). Informal sector initiative in the primary sub-system of urban solid waste management in Lagos, Nigeria. Habitat International, 31(2), 193-204.

Al-Khatib, I. A., Monou, M., Zahra, A. S. F. A., Shaheen, H. Q., & Kassinos, D. (2010). Solid waste characterization, quantification and management practices in developing countries. A case study: Nablus district–Palestine. Journal of environmental management, 91(5), 1131-1138.

Alter, H. (1983). Handbook of materials recovery from solid wastes.

Ambastha, R., & Aich, S. (2019). Decentralized Community-Led Solid Waste Management. Sustainable Waste Management: Policies and Case Studies, 77–89

Anderson, R. C., Hilborn, D., & Weersink, A. (2013). An economic and functional tool for assessing the financial feasibility of farm-based anaerobic digesters. Renewable energy, 51, 85-92.

Angelidaki I, Ellegaard L (2003) Codigestion of Manure and Organic Wastes in Centralized Biogas Plants: Status and Future Trends. Applied Biochemistry and Biotechnology 109:95–106. https://doi.org/10.1385/abab:109:1-3:95

Angelidaki, I., Ellegaard, L., & Ahring, B. K. (1999). A comprehensive model of anaerobic bioconversion of complex substrates to biogas. Biotechnology and bioengineering, 63(3), 363-372.

Antanasijević, D., Pocajt, V., Popović, I., Redžić, N., & Ristić, M. (2013). The forecasting of municipal waste generation using artificial neural networks and sustainability indicators. Sustainability science, 8, 37-46.

Aragonés-Beltrán, P., Pastor-Ferrando, J. P., García-García, F., & Pascual-Agulló, A. (2010). An analytic network process approach for siting a municipal solid waste plant in the metropolitan area of Valencia (Spain). Journal of environmental management, 91(5), 1071-1086.

Arvanitoyannis, I. S., Kassaveti, A., & Ladas, D. (2008). Food waste treatment methodologies. Waste management for the food industries, 345.

Ayeleru, O. O., Okonta, F. N., & Ntuli, F. (2021). Cost benefit analysis of a municipal solid waste recycling facility in Soweto, South Africa. Waste Management, 134, 263-269.

Banar, M., & Özkan, A. (2008). Characterization of the municipal solid waste in Eskisehir City, Turkey. Environmental engineering science, 25(8), 1213-1220.

Bandara, N. J., Hettiaratchi, J. P. A., Wirasinghe, S. C., & Pilapiiya, S. (2007). Relation of waste generation and composition to socio-economic factors: a case study. Environmental monitoring and assessment, 135, 31-39.

Bao, Z., Lu, W., Chi, B,, Yuan, H., & Hao, J. (2019). Procurement innovation for a circular economy of construction and demolition waste: Lessons learnt from Suzhou, China. Waste Manag 99:12–21.

Barr, S., Ford, N. J., & Gilg, A. W. (2003). Attitudes towards recycling household waste in Exeter, Devon: quantitative and qualitative approaches. Local Environment, 8(4), 407-421.

Bautista, J., & Pereira, J. (2006). Modeling the problem of locating collection areas for urban waste management. An application to the metropolitan area of Barcelona. Omega, 34(6), 617-629.

Beigl, P., Lebersorger, S., & Salhofer, S. (2008). Modelling municipal solid waste generation: A review. Waste management, 28(1), 200-214.

Beigl, P., Wassermann, G., Schneider, F., & Salhofer, S. (2004). Forecasting municipal solid waste generation in major European cities.

Beltramo, T., & Hitzmann, B. (2019). Evaluation of the linear and non-linear prediction models optimized with metaheuristics: Application to anaerobic digestion processes. Engineering in Agriculture, Environment and Food, 12(4), 397-403.

Beltramo, T., Klocke, M., & Hitzmann, B. (2019). Prediction of the biogas production using GA and ACO input features selection method for ANN model. Information Processing in Agriculture, 6(3), 349-356.

Benítez, S. O., Lozano-Olvera, G., Morelos, R. A., & de Vega, C. A. (2008). Mathematical modeling to predict residential solid waste generation. Waste Management, 28, S7-S13.

Bianchi, L., Dorigo, M., Gambardella, L. M., & Gutjahr, W. J. (2009) A survey on metaheuristics for stochastic combinatorial optimization. Natural Computing, 8(2):239–287

Bishop, C. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: springer.

Blasius, J. P., Contrera R. C., Maintinguer, S. I., & Alves de Castro, M. C. A. (2020). Effects of temperature, proportion and organic loading rate on the performance of anaerobic digestion of food waste. Biotechnology Reports 27:e00503

Blum, C., & Roli, A. (2003) Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. ACM Computing Surveys, 35(3):268–308

Boniecki, P., Dach, J., Mueller, W., Koszela, K., Przybyl, J., Pilarski, K., & Olszewski, T. (2013). Neural prediction of heat loss in the pig manure composting process. Applied Thermal Engineering, 58(1-2), 650-655.

Bonk, F., Bastidas-Oyanedel, J. R., & Schmidt, J. E. (2015). Converting the organic fraction of solid waste from the city of Abu Dhabi to valuable products via dark fermentation–Economic and energy assessment. Waste Management, 40, 82-91.

Box, G. E., & Jenkins, G. M., (1976). Time series analysis: forecasting and control. John Wiley & Sons, Hoboken, New Jersey.

Breiman, L. (2017). Classification and regression trees. Routledge.

Buenrostro, O., Bocco, G., & Bernache, G. (2001). Urban solid waste generation and disposal in Mexico: a case study. Waste Management & Research 19:169–176.

Buenrostro, O., Bocco, G., & Cram, S. (2001). Classification of sources of municipal solid wastes in developing countries. Resources, Conservation and Recycling, 32(1), 29-41.

Bunsan, S., Chen, W. Y., Chen, H. W., Chuang, Y. H., & Grisdanurak, N. (2013). Modeling the dioxin emission of a municipal solid waste incinerator using neural networks. Chemosphere, 92(3), 258-264.

Cao, L. J., Keerthi, S. S., Ong, C.-J., et al (2006). Parallel Sequential Minimal Optimization for the Training of Support Vector Machines. IEEE Transactions on Neural Networks 17:1039–1049.

Carrere, H., Antonopoulou, G., Affes, R., Passos, F., Battimelli, A., Lyberatos, G., & Ferrer, I. (2016). Review of feedstock pretreatment strategies for improved anaerobic digestion: From lab-scale research to full-scale application. Bioresource technology, 199, 386-397.

Castano, J. M., Martin, J. F., & Ciotola, R. (2014). Performance of a small-scale, variable temperature fixed dome digester in a temperate climate. Energies, 7(9), 5701-5716.

Central Pollution Control Board, CPCB, Annual Report (2014). cpcb.nic.in/annual-report.php

Chakrabarti, S. (2019). Treatment of urban solid waste: engineering and integrated management. The Energy And Resources Institute

Chanakya HN, Sharma I, Ramachandra TV (2009) Micro-scale anaerobic digestion of point source components of organic fraction of municipal solid waste. Waste Management 29:1306–1312. https://doi.org/10.1016/j.wasman.2008.09.014

Chanakya, H. N., & Sreesha, M. (2012). Anaerobic retting of banana and arecanut wastes in a plug flow digester for recovery of fiber, biogas and compost. Energy for sustainable development, 16(2), 231-235.

Chang, A. Y., Parrales, M. E., Jimenez, J., Sobieszczyk, M. E., Hammer, S. M., Copenhaver, D. J., & Kulkarni, R. P. (2009). Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. International journal of health geographics, 8(1), 1-11.

Chelani, A. B., Rao, C. C., Phadke, K. M., & Hasan, M. Z. (2002). Formation of an air quality index in India. International Journal of Environmental Studies, 59(3), 331-342.

Chen G, Wu W, Xu J, Wang Z (2021) An anaerobic dynamic membrane bioreactor for enhancing sludge digestion: Impact of solids retention time on digestion efficacy. Bioresource Technology 329:124864.

Chen Q. L., Li H., Zhou X. Y., et al (2017). An underappreciated hotspot of antibiotic resistance: The groundwater near the municipal solid waste landfill. Science of The Total Environment, 609:966–973.

Chen, M., Huang Y, Liu H, et al. (2019). Impact of different nitrogen source on the compost quality and greenhouse gas emissions during composting of garden waste. Process Safety and Environmental Protection 124:326–335.

Chen, X., Geng, Y., & Fujita, T. (2010). An overview of municipal solid waste management in China. Waste management, 30(4), 716-724.

Chiueh, P. T., Lo, S. L., & Chang, C. L. (2008). A GIS-based system for allocating municipal solid waste incinerator compensatory fund. Waste Management, 28(12), 2690-2701.

Chuang Y.H., Huang W. J., Nguyen K. L. P., et al (2019). Redundancy analysis for characterizing the groundwater quality in coastal industrial areas. Environmental Forensics, 20:77–91.

Chung, S. S. (2010). Projecting municipal solid waste: The case of Hong Kong SAR. Resources, Conservation and Recycling, 54(11), 759-768.

Chynoweth, D. P., Svoronos, S. A., & Lyberatos, G., et al (1994). Real-time expert system control of anaerobic digestion. Water Science and Technology 30:21–29

Cinar, S., Cinar, S. O., Wieczorek, N., et al (2021). Integration of Artificial Intelligence into Biogas Plant Operation. Processes 9:85

Clarke, S. M., Griebsch, J. H., & Simpson, T. W. (2004). Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses. Journal of Mechanical Design 127:1077–1087

Cohen, B. (2006). Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. Technology in society, 28(1-2), 63-80.

Converti, A., Oliveira, R. P. S., Torres, B. R., Lodi, A., & Zilli, M. (2009). Biogas production and valorization by means of a two-step biological process. Bioresource technology, 100(23), 5771-5776.

Dadhich, S. K., Yadav, G. K., Yadav, K., Kumawat, C., & Munalia, M. K. (2021). Recycling of Crop Residues for Sustainable Soil Health Management: A Review. International Journal of Plant & Soil Science, 66–75

Das, K., Keener, H. M., (1997). Moisture effect on compaction and permeability in composts. J Environ Eng 123(3):275–281.

Daskalopoulos, E., Badr, O., & Probert, S. D. (1998). Municipal solid waste: a prediction methodology for the generation rate and composition in the European Union countries and the United States of America. Resources, conservation and recycling, 24(2), 155-166.

De Clercq, D., Jalota, D., Shang, R., Ni, K., Zhang, Z., Khan, A., ... & Yuan, K. (2019). Machine learning powered software for accurate prediction of biogas production: A case study on industrial-scale Chinese production data. Journal of cleaner production, 218, 390-399.

Debrah, J. K., Vidal, D. G., & Dinis, M. A. P. (2021). Raising awareness on solid waste management through formal education for sustainability: A developing countries evidence review. Recycling, 6(1), 6.

Dhakal, S., Koirala, M., Sharma, E., & Subedi, N. R. (2010). Effect of land use change on soil organic carbon stock in Balkhu Khola watershed southwestern part of Kathmandu valley, central Nepal. World Academy of Science Engineering and Technology, 66.

Diaz, L. F., De Bertoldi, M., & Bidlingmaier, W. (2007). Compost science and technology. Elsevier, Amsterdam; Boston, Ma

Diaz, M. J., Madejón, E., López, F., et al. (2002). Optimization of the rate vinasse/grape marc for co-composting process. Process Biochemistry 37:1143–1150.

Dijkema, G. P. J., Reuter, M. A., & Verhoef, E. V. (2000). A new paradigm for waste management. Waste management, 20(8), 633-638.

Dijkstra, F. A., Prior, S. A., Runion, G. B., Torbert, H. A., Tian, H., Lu, C., & Venterea, R. T. (2012). Effects of elevated carbon dioxide and increased temperature on methane and nitrous oxide fluxes: evidence from field experiments. Frontiers in Ecology and the Environment, 10(10), 520–527. https://doi.org/10.1890/120059

Ding, S., Huang, W., Xu, W., Wu, Y., Zhao, Y., Fang, P., ... & Lou, L. (2022). Improving kitchen waste composting maturity by optimizing the processing parameters based on machine learning model. Bioresource Technology, 360, 127606.

Dissanayaka, D. M. S. H., & Vasanthapriyan, S. (2019, December). Forecast municipal solid waste generation in Sri Lanka. In 2019 International Conference on Advancements in Computing (ICAC) (pp. 210-215). IEEE.

Draper, N. R., & Smith, H. (1998). Applied regression analysis (Vol. 326). John Wiley & Sons.

Droke, C. (2001). Moving averages simplified / by Clif Droke. Marketplace Books, Columbia, Md

Dunteman, G. H. (2008). Principal components analysis (No. 69). Sage.

Dyson, B., & Chang, N. B. (2005). Forecasting municipal solid waste generation in a fast-growing urban region with system dynamics modeling. Waste management, 25(7), 669-679.

Edwards J, Othman M, Burn S (2015) A review of policy drivers and barriers for the use of anaerobic digestion in Europe, the United States and Australia. Renewable and Sustainable Energy Reviews 52:815–828.

Ellacuriaga, M., García-Cascallana, J., & Gómez, X. (2021). Biogas production from organic wastes: Integrating concepts of circular economy. Fuels, 2(2), 144-167.

Emery, A., Davies, A., Griffiths, A., & Williams, K. (2007). Environmental and economic modelling: A case study of municipal solid waste management scenarios in Wales. Resources, Conservation and Recycling, 49(3), 244-263.

Eskandari, M., Homaee, M., & Mahmodi, S. (2012). An integrated multi criteria approach for landfill siting in a conflicting environmental, economical and socio-cultural area. Waste Management 32:1528–1538.

Fadhullah, W., Imran, N. I. N., Ismail, S. N. S., Jaafar, M. H., & Abdullah, H. (2022). Household solid waste management practices and perceptions among residents in the East Coast of Malaysia. BMC public health, 22(1), 1-20.

Fallah, B., Ng, K. T. W., Vu, H. L., & Torabi, F. (2020). Application of a multi-stage neural network approach for time-series landfill gas modeling with missing data imputation. Waste Management, 116, 66-78.

Ferronato, N., & Torretta, V. (2019). Waste mismanagement in developing countries: A review of global issues. International journal of environmental research and public health, 16(6), 1060.

Filmer, D., & Pritchett L. H. (2001). Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India. Demography 38:115.

Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge university press.

Freund, R. J., Wilson, W. J., & Sa, P. (2006). Regression analysis. Elsevier.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232.

Fulford, D. (2015). Small-scale rural biogas programmes: A handbook. Practical Action Publishing.

Gao MJ, Tian JW, Jiang W, & Li, K. (2007). Research of sludge compost maturity degree modeling method based on wavelet neural network for sewage treatment. In: Li K, Fei M, Irwin GW, et al. (eds) Bio-Inspired Computational Intelligence and Applications. LSMS 2007. Lecture Notes in Computer Science, vol. 4688. Berlin, Germany: Springer, pp.608–618.

Gao, M., Li, B., Yu, A., Liang, F., Yang, L., & Sun, Y. (2010). The effect of aeration rate on forced-aeration composting of chicken manure and sawdust. Bioresource Technology, 101(6), 1899-1903.

Gardner, M., & Dorling, S. (2000). Artificial neural network-derived trends in daily maximum surface ozone concentrations. Journal of the Air & Waste Management Association, 51(8), 1202-1210.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., ... & Zhao, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). Journal of climate, 30(14), 5419-5454.

Genedy, R. A., & Ogejo, J. A. (2021). Using machine learning techniques to predict liquid dairy manure temperature during storage. Computers and Electronics in Agriculture 187:106234

Ghanbari, F., Kamalan, H., & Sarraf, A. (2021). An evolutionary machine learning approach for municipal solid waste generation estimation utilizing socioeconomic components. Arabian Journal of Geosciences, 14(2), 92.

Ghinea, C., Drăgoi, E. N., Comăniţă, E. D., Gavrilescu, M., Câmpean, T., Curteanu, S. I. L. V. I. A., & Gavrilescu, M. (2016). Forecasting municipal solid waste generation using prognostic tools and regression analysis. Journal of environmental management, 182, 80-93.

Gomez, G., Meneses, M., Ballinas, L., & Castells, F. (2008). Characterization of urban solid waste in Chihuahua, Mexico. Waste Management, 28(12), 2465-2471.

Gómez, G., Meneses, M., Ballinas, L., & Castells, F. (2009). Seasonal characterization of municipal solid waste (MSW) in the city of Chihuahua, Mexico. Waste management, 29(7), 2018-2024.

Goswami, U., & Sarma, H. P. (2008). Study of the impact of municipal solid waste dumping on soil quality in Guwahati city. Pollution research, 27(2), 327-330.

Granger, C. W., & Ramanathan, R. (1984). Improved methods of combining forecasts. Journal of forecasting, 3(2), 197-204.

Grazhdani, D. (2016). Assessing the variables affecting on the rate of solid waste generation and recycling: An empirical analysis in Prespa Park. Waste Management, 48, 3-13.

Guerrero, L. A., Maas, G., & Hogland, W. (2013). Solid waste management challenges for cities in developing countries. Waste management, 33(1), 220-232.

Guo, H., Wu, S., Tian, Y., Zhang, J., & Liu, H. (2021). Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. Bioresource Technology, 319, 124114

Gutiérrez, M. C., Serrano, A., Siles, J. A., Chica, A. F., & Martín, M. A. (2017). Centralized management of sewage sludge and agro-industrial waste through co-composting. Journal of environmental management, 196, 387-393.

Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. Feature extraction: foundations and applications, 1-25.

Hall, S. G. (1998). Temperature feedback and control via aeration rate regulation in biological composting systems. Ph.D. Dissertation, Cornell University, Ithaca, USA.

Hamelers, H. V. M. (2004). Modeling composting kinetics: a review of approaches. Rev Environ Sci Biotechnol 3(4):331–342.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning.

Hatik, C., & Gatina, J. C. (2017). Waste production classification and analysis: a PCA-induced methodology. Energy Procedia, 136, 488-494.

Haug, R. T. (1993). Compost engineering. In: The library of compost engineering software.

Hensen, A., Skiba, U., & Famulari, D. (2013). Low cost and state of the art methods to measure nitrous oxide emissions. Environmental Research Letters, 8(2), 025022.

Hering, J. G. (2012). An end to waste?. Science, 337(6095), 623-623.

Hira M., Yadav S., Morthekai P., et al (2018). Mobile Phones—An asset or a liability: A study based on characterization and assessment of metals in waste mobile phone components using leaching tests. Journal of Hazardous Materials, 342:29–40.

Hockett, D., Lober, D. J., & Pilgrim, K. (1995). Determinants of per capita municipal solid waste generation in the Southeastern United States. Journal of Environmental Management, 45(3), 205-218.

Hristovski, K, Olson, L., Hild, N., et al (2007). The municipal solid waste system and solid waste characterization at the municipality of Veles, Macedonia. Waste Management 27:1680–1689.

Huang, J., & Koroteev, D. D. (2021). Artificial intelligence for planning of energy and waste management. Sustainable Energy Technologies and Assessments, 47, 101426.

Huang W. J., Lin Y. H., Chen W. Y., et al (2015). Causal relationships among biological toxicity, geochemical conditions and derived DBPs in groundwater. Journal of Hazardous Materials, 283:24–34.

Ihara, I., Yano, K., Andriamanohiarisoamanana, F. J., Yoshida, G., Yuge, T., Yuge, T., ... & Umetsu, K. (2020). Field testing of a small-scale anaerobic digester with liquid dairy manure and other organic wastes at an urban dairy farm. Journal of Material Cycles and Waste Management, 22, 1382-1389.

IOCL. White Paper on Expression of Interest (EOI). Retrieved from https://iocl.com/download/White_Paper_EOI_1.pdf

Jain, S., & Sharma, M. P. (2011). Power generation from MSW of Haridwar city: A feasibility study. Renewable and Sustainable Energy Reviews, 15(1), 69-90.

Jayapriya J., & Hema J. (2020). Sustainable Waste Management in Higher Education Institutions—A Case Study in AC Tech, Anna University, Chennai, India. Green Engineering for Campus Sustainability, 163–172

Jayaraman, V., Parthasarathy, S., & Arun Raj Lakshminarayanan. (2022). Forecasting the Emission of Greenhouse Gases from the Waste using SARIMA Model. International Conference on Trends in Electronics and Informatics (ICOEI).

Jha, B., Jaiswal, K. K., Ramaswamy, A. P. (2016). Impact of poly-aluminium chloride on foam suppression in a chicken waste-based biogas plant: A case study at KEPCO Kerala. International Journal of Environmental Sciences, 6(6), 934-940. International Journal of Environmental Sciences 6:934–940.

Jha, A. K., Sharma, C., Singh, N., Ramesh, R., Purvaja, R., & Gupta, P. K. (2008). Greenhouse gas emissions from municipal solid waste management in Indian mega-cities: A case study of Chennai landfill sites. Chemosphere, 71(4), 750–758.

Jha, A. K., Singh, S. K., Singh, G. P., & Gupta, P. K. (2011). Sustainable municipal solid waste management in low income group of cities: a review. Tropical Ecology, 52(1), 123-131.

Jiang, T., Schuchardt, F., Li, G., Guo, R., & Zhao, Y. (2011). Effect of C/N ratio, aeration rate and moisture content on ammonia and greenhouse gas emission during the composting. Journal of Environmental Sciences, 23(10), 1754-1760.

Johnson, N. E., Ianiuk, O., Cazap, D., Liu, L., Starobin, D., Dobler, G., & Ghandehari, M. (2017). Patterns of waste generation: A gradient boosting model for short-term waste prediction in New York City. Waste management, 62, 3-11.

Jones, P., & Salter, A. (2013). Modelling the economics of farm-based anaerobic digestion in a UK whole-farm context. Energy Policy, 62, 215-225.

Kabyanga, M., Balana, B. B., Mugisha, J., Walekhwa, P. N., Smith, J., & Glenk, K. (2018). Economic potential of flexible balloon biogas digester among smallholder farmers: A case study from Uganda. Renewable Energy, 120, 392-400.

Kaiser, J. (1996). Modelling composting as a microbial ecosystem: a simulation approach. Ecological modelling, 91(1-3), 25-37.

Kalia A., & Singh S. (2004). Development of a Biogas Plant. Energy Sources 26, 707–714.

Kalia, V. C., Sonakya, V., & Raizada, N. (2000). Anaerobic digestion of banana stem waste. Bioresource Technology, 73(2), 191-193.

Kamran, A., Chaudhry, M. N., & Batool, S. A. (2015). Effects of socio-economic status and seasonal variation on municipal solid waste composition: a baseline study for future planning and development. Environmental Sciences Europe, 27, 1-8.

Kamran, A., Chaudhry, M. N., & Batool, S. A. (2015). Effects of socio-economic status and seasonal variation on municipal solid waste composition: a baseline study for future planning and development. Environmental Sciences Europe, 27, 1-8.

Kannangara, M., Dua, R., Ahmadi, L., & Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. Waste management, 74, 3-15.

Karacan, C. Ö. (2008). Modeling and prediction of ventilation methane emissions of US longwall mines using supervised artificial neural networks. International Journal of Coal Geology, 73(3-4), 371-387.

Karimi, N., Richter, A., & Ng, K.T. W. (2020). Siting and ranking municipal landfill sites in regional scale using nighttime satellite imagery. Journal of Environmental Management 256:109942.

Kashyap, A., & Borthakur, R. (2019) Identification of Types and Source-Specific Characterization and Quantification Study of Solid Waste in Guwahati City, Assam, India. In Waste Management and Resource Efficiency (pp. 373-384). Springer, Singapore

Kashyap, A., Kalita, J., Kalita, S., & Mazumdar, K. (2010). Present scenario of solid waste with special reference to plastic and other non-biodegradable solid waste and its management for the sustainable urban poor development in Guwahati city, Assam, India. management, 1, 1.

Katsamaki, A., Willems, S., & Diamadopoulos, E. (1998). Time series analysis of municipal solid waste generation rates. Journal of Environmental Engineering, 124(2), 178-183.

Khalil, A. I. (1996). Composting and the utilization of organic wastes: an environmental study [Ph.D. Thesis, Institute of Graduate Studies & Research, University of Alexandra, Alexandria, Egypt.]

Khalil, A. I., Hassouna, M. S., El-Ashqar, H. M. A., & Fawzi, M. (2011). Changes in physical, chemical and microbial parameters during the composting of municipal sewage sludge. World Journal of Microbiology and Biotechnology, 27(10):2359-2369

Khalil, M., Iqbal, M., Turan, V., Tauqeer, H. M., Farhad, M., Ahmed, A., & Yasin, S. (2022). Household chemicals and their impact. In Environmental micropollutants (pp. 201-232). Elsevier.

Khan, M. M. U. H., Jain, S., Vaezi, M., & Kumar, A. (2016). Development of a decision model for the techno-economic assessment of municipal solid waste utilization pathways. Waste management, 48, 548-564.

Kirkegaard, R. H., McIlroy, S. J., Kristensen, J. M., Nierychlo, M., Karst, S. M., Dueholm, M. S., ... & Nielsen, P. H. (2017). The impact of immigration on microbial community composition in full-scale anaerobic digesters. Scientific reports, 7(1), 9343.

Kishimoto, M., Preechaphan, C., Yoshida, T., & Taguchi, H. (1987). Simulation of an aerobic composting of activated sludge using a statistical procedure. MIRCEN journal of applied microbiology and biotechnology, 3, 113-123.

Kolekar, K. A., Hazra, T., & Chakrabarty, S. N. (2016). A review on prediction of municipal solid waste generation models. Procedia Environmental Sciences, 35, 238-244.

Kollikkathara, N., Feng, H., & Stern, E. (2009). A purview of waste management evolution: Special emphasis on USA. Waste management, 29(2), 974-985.

Komilis, D. P., & Ham, R. K. (2004). Life-Cycle Inventory of Municipal Solid Waste and Yard Waste Windrow Composting in the United States. Journal of Environmental Engineering, 130(11), 1390–1400

Kong, E. J., Bahner M. A., & Turner S. L. (1996). Addendum to Assessment of Styrene Emission Controls for FRP/C and Boat Building Industries. United States Environmental Protection Agency, Research and Development, National Risk Management Research Laboratory

Kontokosta, C. E., Hong, B., Johnson, N. E., & Starobin, D. (2018). Using machine learning and small area estimation to predict building-level municipal solid waste generation in cities. Computers, Environment and Urban Systems, 70, 151-162.

Kroese, D. P., Botev, Z., Taimre, T., & Vaisman, R. (2019). Data science and machine learning: mathematical and statistical methods. CRC Press.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (Vol. 26, p. 13). New York: Springer.

Kujawa, S., Mazurkiewicz, J., & Czekała, W. (2020). Using convolutional neural networks to classify the maturity of compost based on sewage sludge and rapeseed straw. Journal of Cleaner Production, 258, 120814.

Kujawa, S., Nowakowski, K., Tomczak, R. J., Dach, J., Boniecki, P., Weres, J., ... & Carmona, P. C. R. (2014). Neural image analysis for maturity classification of sewage sludge composted with maize straw. Computers and Electronics in Agriculture, 109, 302-310.

Kulkarni, M. B., Ghanegaonkar, P. M. (2019). Methane enrichment of biogas produced from floral waste: A potential energy source for rural India. Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 41:2757–2768.

Kumar, A., & Samadder S. R. (2017). A review on technological options of waste to energy for effective management of municipal solid waste. Waste Management 69:407–422.

Kumar, A., Mandal, B., & Sharma, A. (2015). Advancement in biogas digester. Energy sustainability through green energy, 351-382.

Kumar, A., Samadder, S. R., Kumar, N., & Singh, C. (2018). Estimation of the generation rate of different types of plastic wastes and possible revenue recovery from informal recycling. Waste Management, 79, 781-790.

Kumar, K. N., & Goel, S. (2009). Characterization of Municipal Solid Waste (MSW) and a proposed management plan for Kharagpur, West Bengal, India. Resources, Conservation and Recycling, 53(3), 166–174

Kurmana, A., & Srinivas, N. (2021). Quality assessment of compost from Centralized windrow composter (CWC) and Source segregate automatic vessel composter (AVC) at Hyderabad city in India. Journal of Applied and Natural Science, 13(2), 450–454

Lansana, F. M. (1993). A comparative analysis of curbside recycling behavior in urban and suburban communities. The Professional Geographer, 45(2), 169-179.

Legendre P., & Legendre L. (2012). Numerical ecology. Elsevier

Li, C., He, P., Peng, W., Lü, F., Du, R., & Zhang, H. (2022). Exploring available input variables for machine learning models to predict biogas production in industrial-scale biogas plants treating food waste. Journal of Cleaner Production, 380, 135074.

Li, H., Sanchez, R., Qin, S. J., Kavak, H. I., Webster, I. A., Tsotsis, T. T., & Sahimi, M. (2011). Computer simulation of gas generation and transport in landfills. V: Use of artificial neural network and the genetic algorithm for short-and long-term forecasting and planning. Chemical engineering science, 66(12), 2646-2659.

Li, Y., Jin, Y., Li, H., Borrion, A., Yu, Z., & Li, J. (2018). Kinetic studies on organic degradation and its impacts on improving methane production during anaerobic digestion of food waste. Applied energy, 213, 136-147.

Li, Y., Manandhar, A., Li, G., & Shah, A. (2018). Life cycle assessment of integrated solid state anaerobic digestion and composting for on-farm organic residues treatment. Waste Management, 76, 294-305.

Liang, C., Das, K. C., & McClendon, R. W. (2003). The influence of temperature and moisture contents regimes on the aerobic microbial activity of a biosolids composting blend. Bioresource technology, 86(2), 131-137.

Lin, C., Wei, C. C., & Tsai, C. C. (2016). Prediction of influential operational compost parameters for monitoring composting process. Environmental Engineering Science, 33(7), 494-506

Lin, L., Shah, A., Keener, H., & Li, Y. (2019). Techno-economic analyses of solid-state anaerobic digestion and composting of yard trimmings. Waste management, 85, 405-416.

Lin, Y. P., Huang, G. H., Lu, H. W., & He, L. (2008). Modeling of substrate degradation and oxygen consumption in waste composting processes. Waste Management, 28(8), 1375-1385.

Liu, S., Wang, D., Li, H., et al (2017). The Ecological Security Pattern and Its Constraint on Urban Expansion of a Black Soil Farming Area in Northeast China. ISPRS International Journal of Geo-Information 6:263.

Lohani, S. P., & Havukainen, J. (2018). Anaerobic digestion: factors affecting anaerobic digestion process. Waste Bioremediation, 343-359.

Lu, W., Lou, J., Webster, C., Xue, F., Bao, Z., & Chi, B. (2021). Estimating construction waste generation in the Greater Bay Area, China using machine learning. Waste Manag 134:78–88.

MacArthur, E. Foundation. (2013). Towards the circular economy. J of Ind Ecol 23–44.

Magazzino, C., Mele, M., & Schneider, N. (2020). The relationship between municipal solid waste and greenhouse gas emissions: Evidence from Switzerland. Waste Management, 113.

Maharashtra Energy Development Agency Annual Report (2012). mahaurja.com/meda/en/programmes/rporec/rpo/meda_reports

Mahapatra, S., Ali, M. H., & Samal, K. (2022). Assessment of compost maturity-stability indices and recent development of composting bin. Energy Nexus, 6, 100062.

Mahini, A. S., & Gholamalifard, M. (2006). Siting MSW landfills with a weighted linear combination methodology in a GIS environment. International Journal of Environmental Science & Technology, 3, 435-445.

Mandal, I., & Pal, S. (2020). COVID-19 pandemic persuaded lockdown effects on environment over stone quarrying and crushing areas. Science of the Total Environment, 732, 139281.

Mandal, P., Chaturvedi, M. K., Bassin, J. K., Vaidya, A. N., & Gupta, R. K. (2014). Qualitative assessment of municipal solid waste compost by indexing method. International Journal of Recycling of Organic Waste in Agriculture, 3(4), 133–139

Mao, C., Feng, Y., Wang, X., & Ren, G. (2015). Review on research achievements of biogas from anaerobic digestion. Renewable and sustainable energy reviews, 45, 540-555.

Mason, I, G., Milke, M. W. (2005). Physical modelling of the composting environment: a review. Part 1: reactor systems. Waste Manag 25(5):481–500.

Mavrotas, G., Gakis, N., Skoulaxinou, S., Katsouros, V., & Georgopoulou, E. (2015). Municipal solid waste management and energy production: Consideration of external cost through multi-objective optimization and its effect on waste-to-energy solutions. Renewable and Sustainable Energy Reviews, 51, 1205-1222.

McBain, M. C., Warland, J. S., McBride, R. A., & Wagner-Riddle, C. (2005). Micrometeorological measurements of N2O and CH4 emissions from a municipal solid waste landfill. Waste Management & Research, 23(5), 409–419.

McDougall, F. (2001). Life cycle inventory tools: supporting the development of sustainable solid waste management systems. Corporate environmental strategy, 8(2), 142-147.

Mckenzie, S. K., Carter, K. N., Blakely, T., & Ivory, V. (2011). Effects of childhood socioeconomic position on subjective health and health behaviours in adulthood: how much is mediated by adult socioeconomic position?. BMC Public Health, 11, 1-9. Abu Qdais HA, Hamoda MF, Newham J (1997) Analysis of Residential Solid Waste at Generation Sites. Waste Management & Research 15:395–405.

Mehrdad, S. M., Abbasi, M., Yeganeh, B., & Kamalan, H. (2021). Prediction of methane emission from landfills using machine learning models. Environmental Progress & Sustainable Energy, 40(4).

Meza, J. K. S., Yepes, D. O., Rodrigo-Ilarri, J., & Cassiraga, E. (2019). Predictive analysis of urban waste generation for the city of Bogotá, Colombia, through the implementation of decision trees-based machine learning, support vector machines and artificial neural networks. Heliyon, 5(11), e02810.

Miezah, K., Obiri-Danso, K., Kádár, Z., Fei-Baffoe, B., & Mensah, M. Y. (2015). Municipal solid waste characterization and quantification as a measure towards effective waste management in Ghana. Waste management, 46, 15-27.

Minghua, Z., Xiumin, F., Rovetta, A., Qichang, H., Vicentini, F., Bingkai, L., ... & Yi, L. (2009). Municipal solid waste management in Pudong new area, China. Waste management, 29(3), 1227-1233.

Ministry of Environment and Forests (MoEF) Report, (2015). cag.gov.in/uploads/download_audit_report/2015/Union_Compliance_Scientific_Departmen_Report_30_2015_chap_7.pdf

Ministry of Housing and Urban Affairs (MOHUA) Annual Report (2016). mohua.gov.in/publication/annual-reports.php

Ministry of New and Renewable Energy (MNRE) Annual Report (2018). mnre.gov.in/knowledge-center/publication

Ministry of New and Renewable Energy Annual Report (2017-18). Ministry of New and Renewable Energy (MNRE)

Mir, I. S., Cheema, P. P. S., & Singh, S. P. (2021). Implementation analysis of solid waste management in Ludhiana city of Punjab. Environmental Challenges, 2, 100023

Mmereki D, Baldwin A, Li B, Liu M (2015) Healthcare waste management in Botswana: storage, collection, treatment and disposal system. Journal of Material Cycles and Waste Management 19:351–365.

Moeinaddini, M., Khorasani, N., Danehkar, A., & Darvishsefat, A. A. (2010). Siting MSW landfill using weighted linear combination and analytical hierarchy process (AHP) methodology in GIS environment (case study: Karaj). Waste management, 30(5), 912-920.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). Foundations of machine learning. MIT press.

Molina-Gómez, N. I., Díaz-Arévalo, J. L., & López-Jiménez, P. A. (2021). Air quality and urban sustainable development: the application of machine learning tools. International Journal of Environmental Science and Technology, 18(4), 1029-1046.

Molugaram, K., Rao, G. S., Shah, A., Davergave, N. (2017). Statistical techniques for transportation engineering. Butterworth-Heinemann.

Moncks, P. C. S., Corrêa, É. K., Guidoni, L. L. C., Moncks, R. B., Corrêa, L. B., Lucia Jr, T., ... & Marques, F. S. (2022). Moisture content monitoring in industrial-scale composting systems using low-cost sensor-based machine learning techniques. Bioresource Technology, 359, 127456.

Namlis, K. G., & Komilis, D. (2019). Influence of four socioeconomic indices and the impact of economic crisis on solid waste generation in Europe. Waste management, 89, 190-200.

Nandi, R., Saha, C. K., Sarker, S., Huda, M. S., & Alam, M. M. (2020). Optimization of reactor temperature for continuous anaerobic digestion of cow manure: Bangladesh perspective. Sustainability, 12(21), 8772.

Nas, B., Cay, T., Iscan, F., & Berktay, A. (2010). Selection of MSW landfill site for Konya, Turkey using GIS and multi-criteria evaluation. Environmental monitoring and assessment, 160, 491-500.

National Solid Waste Association of India (NSWAI), Report (2017). https://www.nswai.org/knowledge_reports.php

Navarro-Esbrı, J., Diamadopoulos, E., & Ginestar, D. (2002). Time series analysis and forecasting techniques for municipal solid waste management. Resources, conservation and Recycling, 35(3), 201-214.

Neto, J. G., Ozorio, L. V., de Abreu, T. C. C., Dos Santos, B. F., & Pradelle, F. (2021). Modeling of biogas production from food, fruits and vegetables wastes using artificial neural network (ANN). Fuel, 285, 119081.

Ngai EWT, Poon JKL, Chan YHC (2007) Empirical examination of the adoption of WebCT using TAM. Computers & Education 48:250–267.

Nguyen, K. L. P., Chuang, Y. H., Chen, H. W., & Chang, C. C. (2020). Impacts of socioeconomic changes on municipal solid waste characteristics in Taiwan. Resources, Conservation and Recycling, 161, 104931.

Nguyen, X. C., Nguyen, T. T. H., La, D. D., Kumar, G., Rene, E. R., Nguyen, D. D., … Nguyen, V. K. (2021). Development of machine learning - based models to forecast solid waste generation in residential areas: A case study from Vietnam. Resour, Conserv and Recycl 167(2020): 105381.

Ngwabie, N. M., Wirlen, Y. L., Yinda, G. S., & VanderZaag, A. C. (2019). Quantifying greenhouse gas emissions from municipal solid waste dumpsites in Cameroon. Waste Management, 87, 947–953.

Nisbet, R., Elder, J., & Miner, G. D. (2009). Handbook of statistical analysis and data mining applications. Academic press.

Nkuna, R., Roopnarain, A., Rashama, C., & Adeleke, R. (2021). Insights into organic loading rates of anaerobic digestion for biogas production: a review. Critical Reviews in Biotechnology 42:487–507

Noori, R., Abdoli, M. A., Ghazizade, M. J., & Samieifard, R. (2009). Comparison of neural network and principal component-regression analysis to predict the solid waste generation in Tehran. Iranian Journal of Public Health, 38(1), 74-84.

Noori R., Abdoli M. A., Ghasrodashti A. A., Jalili Ghazizade M. (2009). Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: A case study of Mashhad. Environmental Progress & Sustainable Energy, 28:249–258.

Nsair A, Onen Cinar S, Alassali A, et al (2020) Operational Parameters of Biogas Plants: A Review and Evaluation Study. Energies 13:3761.

Ogwueleka, T. C. (2013). Survey of household waste composition and quantities in Abuja, Nigeria. Resources, Conservation and Recycling, 77, 52-60.

Ojeda-Benítez, S., Armijo-de Vega, C., & Marquez-Montenegro, M. Y. (2008). Household solid waste characterization by family socioeconomic profile as unit of analysis. Resources, Conservation and Recycling, 52(7), 992-999.

Ozkaya, B., Demir, A., & Bilgili, M. S. (2007). Neural network prediction model for the methane fraction in biogas from field-scale landfill bioreactors. Environmental Modelling & Software, 22(6), 815-822.

Pao, H. T., & Chih, Y. Y. (2006). Comparison of TSCS regression and neural network models for panel data forecasting: debt policy. Neural Computing & Applications, 15, 117-123.

Parthan, S. R., Milke, M. W., Wilson, D. C., & Cocks, J. H. (2012). Cost estimation for solid waste management in industrialising regions – Precedents, problems and prospects. Waste Management, 32(3), 584–594

Patil RB, Gadiwad A (2015) Renewable energy and sustainable development: a study of Appropriate Rural Technology Institute (ARTI), Pune, India. Journal of Environmental Research and Development 9:1264–1270

Phuong, N. T. L., Yabar, H., & Mizunoya, T. (2021). Characterization and analysis of household solid waste composition to identify the optimal waste management method: A case study in Hanoi city, Vietnam. Earth, 2(04).

Pradhan, P. K., Mohanty, C. R., Swar, A. K., & Mohapatra, P. (2012). Urban solid waste management of Guwahati city in north-east India. Journal of Urban and Environmental Engineering, 6(2), 67-73.

Qu, X. Y., Li, Z. S., Xie, X. Y., Sui, Y. M., Yang, L., & Chen, Y. (2009). Survey of composition and generation rate of household wastes in Beijing, China. Waste Management, 29(10), 2618-2624.

Quinlan, J. R. (1999). Simplifying decision trees. International Journal of Human-Computer Studies, 51(2), 497-510.

Radojević, D., Antanasijević, D., Perić-Grujić, A., Ristić, M., & Pocajt, V. (2019). The significance of periodic parameters for ANN modeling of daily SO2 and NOx concentrations: A case study of Belgrade, Serbia. Atmospheric Pollution Research, 10(2), 621-628.

Rafey, A., & Siddiqui, F. Z. (2021). A review of plastic waste management in India–challenges and opportunities. International Journal of Environmental Analytical Chemistry, 1-17.

Rahimi, V., & Shafiei, M. (2019). Techno-economic assessment of a biorefinery based on low-impact energy crops: a step towards commercial production of biodiesel, biogas, and heat. Energy conversion and management, 183, 698-707.

Rahman, M. A., Saha, C. K., Feng, L., Møller, H. B., & Alam, M. M. (2019). Anaerobic digestion of agro-industrial wastes of Bangladesh: Influence of total solids content. Engineering in Agriculture, Environment and Food, 12(4), 484-493.

Rajendran R, Soora M, Kandasamy S, et al (2011) Biogas plants efficiency in purifying Indian sago factory waste water with wide C/N ratios: strategies for process water reuse. International Journal of Sustainable Engineering 4:

Rao, M. S., Singh, S. P., Singh, A. K., & Sodha, M. S. (2000). Bioenergy conversion studies of the organic fraction of MSW: assessment of ultimate bioenergy production potential of municipal garbage. Applied energy, 66(1), 75-87.

Rao CR (1964) The Use and Interpretation of Principal Component Analysis in Applied Research. Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 26:329–358

Rastogi, M., Nandal, M., & Nain, L. (2019). Seasonal variation induced stability of municipal solid waste compost: an enzyme kinetics study. SN Applied Sciences, 1, 1-16.

Rathod, T., Hudnurkar, M., & Ambekar, S. (2020). Use of Machine Learning in Predicting the Generation of Solid Waste. PalArch's J of Arch of Egypt/Egyptology 17(6): 4323-4335.

Rawat, S., & Daverey, A. (2018). Characterization of household solid waste and current status of municipal waste management in Rishikesh, Uttarakhand. Environmental Engineering Research, 23(3), 323-329.

Rimaitytė, I., Ruzgas, T., Denafas, G., Račys, V., & Martuzevicius, D. (2012). Application and evaluation of forecasting methods for municipal solid waste generation in an eastern-European city. Waste Management & Research, 30(1), 89-98.

Risberg, K., Sun, L., Levén, L., Horn, S. J., & Schnürer, A. (2013). Biogas production from wheat straw and manure–impact of pretreatment and process operating parameters. Bioresource Technology, 149, 232-237.

Roweis, S., Hinton, G., & Salakhutdinov, R. (2004). Neighbourhood component analysis. Adv. Neural Inf. Process. Syst.(NIPS), 17(4).

Ruiz, L. A. L, Ramón X. R., & Domingo SG (2020). The circular economy in the construction and demolition waste sector–A review and an integrative model approach. J Clean Prod 248:119238.

Rumelhart DE, Mcclelland JL (1986) Parallel distributed processing: explorations in the microstructure of cognition: Volume 1: foundations. Mit Press, Cambridge, Mass.

Sabbir, A. Y. B., Saha, C. K., Nandi, R., Zaman, M. F. U., Alam, M. M., & Sarker, S. (2021). Effects of seasonal temperature variation on slurry temperature and biogas composition of a commercial fixed-dome anaerobic digester used in Bangladesh. Sustainability, 13(19), 11096.

Sahni, S., & Aulakh, R. S. (2014). Planning for Low Carbon Cities in India. Environment and Urbanization Asia, 5(1), 17–34

Samarasinghe, K., & Wijayatunga, P. D. (2022). Techno-economic feasibility and environmental sustainability of waste-to-energy in a circular economy: Sri Lanka case study. Energy for Sustainable Development, 68, 308-317.

Sammut, C., & Webb, G. I. (Eds.). (2011). Encyclopedia of machine learning. Springer Science & Business Media.

Sankoh, F. P., Yan, X., & Conteh, A. M. H. (2012). A situational assessment of socioeconomic factors affecting solid waste generation and composition in Freetown, Sierra Leone. Journal of Environmental Protection, 2012.

Scozzari, A. (2008). Non-invasive methods applied to the case of Municipal Solid Waste landfills (MSW): analysis of long-term data. Advances in Geosciences, 19, 33-38.

Seadon, J. K. (2010). Sustainable waste management systems. Journal of cleaner production, 18(16-17), 1639-1651.

Şener, B., Süzen, M. L., & Doyuran, V. (2006). Landfill site selection by using geographic information systems. Environmental geology, 49, 376-388.

Shailendra, K., Mishra, B. P., Khardiwar, M. S., Patel, S. K., Yaduvanshi, B. K., & Solanki, B. P. (2016). Biogas plants in Chhattisgarh (India): A case study. Current World Environment, 11(2), 599.

Sharholy, M., Ahmad, K., Mahmood, G., & Trivedi, R. C. (2008). Municipal solid waste management in Indian cities–A review. Waste management, 28(2), 459-467.

Sharma, A., Gupta, A. K., & Ganguly, R. (2018). Impact of open dumping of municipal solid waste on soil properties in mountainous region. Journal of Rock Mechanics and Geotechnical Engineering, 10(4), 725-739.

Shi, T., & Horvath, S. (2006). Unsupervised learning with random forest predictors. Journal of Computational and Graphical Statistics, 15(1), 118-138.

Singh, T., & Uppaluri, R. V. S. (2022). Machine learning tool-based prediction and forecasting of municipal solid waste generation rate: a case study in Guwahati, Assam, India. International Journal of Environmental Science and Technology.

Singh, B., Szamosi, Z., & Siménfalvi, Z. (2020). Impact of mixing intensity and duration on biogas production in an anaerobic digester: a review. Critical reviews in biotechnology, 40(4), 508-521.

Singh, C., Tiwari, S., Gupta, V. K., & Singh, J. S. (2018). The effect of rice husk biochar on soil nutrient status, microbial biomass and paddy productivity of nutrient poor agriculture soils. CATENA 171:485–493.

Singh, R. P., Tyagi, V. V., Allen, T., Ibrahim, M. H., & Kothari, R. (2011). An overview for exploring the possibilities of energy generation from municipal solid waste (MSW) in Indian scenario. Renewable and Sustainable Energy Reviews, 15(9), 4797-4808.

Singh, S., Sharma, S. K., & Akbar, M. A. (2022). Developing zero carbon emission pavements with geopolymer concrete: A comprehensive review. Transportation Research Part D: Transport and Environment, 110, 103436.

Singhal, A., Gupta, A. K., Dubey, B., & Ghangrekar, M. M. (2022) Seasonal characterization of municipal solid waste for selecting feasible waste treatment technology for Guwahati city, India. J of the Air and Waste Manag Asso, 72(2):147–160

Sjöström, M., & Östblom, G. (2010). Decoupling waste generation from economic growth—A CGE analysis of the Swedish case. Ecological Economics, 69(7), 1545-1552.

Sole-Mauri, F., Illa, J., Magrí, A., Prenafeta-Boldú, F. X., Flotats, X. (2007). An integrated biochemical and physical model for the composting process. Biores Technol 98(17):3278–3293.

Soni U, Roy A, Verma A, Jain V (2019). Forecasting municipal solid waste generation using artificial intelligence models—a case study in India. SN Applied Sciences 1:

Spokas, K., Bogner, J., & Chanton, J. (2011). A process-based inventory model for landfill CH4 emissions inclusive of seasonal soil microclimate and CH4 oxidation. Journal of Geophysical Research: Biogeosciences, 116(G4).

Sposob, M., Moon, H. S., Lee, D., Kim, T. H., & Yun, Y. M. (2020). Comprehensive analysis of the microbial communities and operational parameters of two full-scale anaerobic digestion plants treating food waste in South Korea: seasonal variation and effect of ammonia. Journal of Hazardous Materials, 398, 122975.

Stutzenberger, F. J., Kaufman, A. J., & Lossin, R. D. (1970). Cellulolytic activity in municipal solid waste composting. Canadian Journal of Microbiology, 16(7), 553–560

Sujauddin, M., Huda, S. M., & Hoque, A. R. (2008). Household solid waste characteristics and management in Chittagong, Bangladesh. Waste management, 28(9), 1688-1695.

Sultana, A., & Kumar, A. (2012). Optimal siting and size of bioenergy facilities using geographic information system. Applied Energy, 94, 192-201.

Sultana, A., & Kumar, A. (2012). Optimal siting and size of bioenergy facilities using geographic information system. Applied Energy, 94, 192-201.

Sumathi, V. R., Natesan, U., & Sarkar, C. (2008) GIS-based approach for optimized siting of municipal solid waste landfill. Waste Management 28:2146–2160.

Sun, N., & Chungpaibulpatana, S. (2017). Development of an appropriate model for forecasting municipal solid waste generation in Bangkok. Energy Procedia, 138, 907-912.

Sun, W., Huang, G. H., Zeng, G., Qin, X., & Yu, H. (2011) Quantitative effects of composting state variables on C/N ratio through GA-aided multivariate analysis. Science of the Total Environment, 409(7):1243–1254

Supriyadi, S., Kriwoken, L. K., & Birley, I. (2000). Solid waste management solutions for Semarang, Indonesia. Waste Management & Research, 18(6), 557-566.

Suthar, S., & Singh, P. (2015). Household solid waste generation and composition in different family size and socio-economic groups: A case study. Sustainable Cities and Society, 14, 56-63.

Sutthasil, N., Chiemchaisri, C., Chiemchaisri, W., Ishigaki, T., Ochiai, S., & Yamada, M. (2020). Greenhouse gas emission from windrow pile for mechanical biological treatment of municipal solid wastes in tropical climate. Journal of Material Cycles and Waste Management, 22(2), 383–395.

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. Handbook of statistics, 24, 303-329.

Swaminathan, M. (2018). How can India's waste problem see a systemic change. Economic and Political Weekly, 21

Talaiekhozani, A., Nematzadeh, S., Eskandari, Z., Dehkordi, A. A., & Rezania, S. (2018). Gaseous emissions of landfill and modeling of their dispersion in the atmosphere of Shahrekord, Iran. Urban climate, 24, 852-862.

Tavares, G., Zsigraiová, Z., & Semiao, V. (2011). Multi-criteria GIS-based siting of an incineration plant for municipal solid waste. Waste management, 31(9-10), 1960-1972.

Tchobanoglous, G., & Kreith, F. (1994). Handbook of solid waste management. McGraw-Hill Education.

Tchobanoglous, G., Theisen, H., Vigil, S., & George, T. (1993). Integrated Solid Waste Management: Engineering Principles and Management Issues. McGraw-Hill Science/Engineering/Math

Thanh, N. P., Matsui, Y., & Fujiwara, T. (2010). Household solid waste generation and characteristic in a Mekong Delta city, Vietnam. Journal of Environmental Management, 91(11), 2307-2321.

Thomas, G. V. (2004). vermicomposting of coconut leaves by Eudrilus sp . in coastal tract of Kerala. Journal of Plantation Crops, 32(JANUARY):486–490

Thompson, T. M., Young, B. R., & Baroutian, S. (2019). Advances in the pretreatment of brown macroalgae for biogas production. Fuel Processing Technology, 195, 106151.

Tokida, T., Fumoto, T., Cheng, W. S., Matsunami, T., Adachi, M., Katayanagi, N., … Hasegawa, T. (2010). Effects of free-air CO2 enrichment (FACE) and soil warming on CH4 emission from a rice paddy field: impact assessment and stoichiometric evaluation. 7, 7(9), 2639–2653.

Tomei, M. C., Braguglia, C. M., Cento, G., & Mininni, G. (2009). Modeling of anaerobic digestion of sludge. Critical Reviews in Environmental Science and Technology, 39(12), 1003-1051.

Towler, G., & Sinnott, R. (2008). Chemical engineering design: principles, practice and economics of plant and process design. Butterworth-Heinemann.

Tripathi, D. K., Kumar, M., & Biswas, V. (2022). Spatial Modelling for Municipal Solid Waste Management Using Remote Sensing and Geographic Information System. In Remote Sensing and Geographic Information Systems for Policy Decision Support (pp. 483-497). Singapore: Springer Nature Singapore.

Trisakti, B., Manalu, V., Taslim, I., & Turmuzi, M. (2015). Acidogenesis of palm oil mill effluent to produce biogas: effect of hydraulic retention time and pH. Procedia-Social and Behavioral Sciences, 195, 2466-2474.

U.S. Energy Information Administration (EIA) Report (2013). www.osti.gov/servlets/purl/1081575

Ugwu, S. N., & Enweremadu, C. C. (2019). Biodegradability and kinetic studies on biomethane production from okra (Abelmoschus esculentus) waste. South African Journal of Science, 115(7-8), 1-5.

United Nations Development Programme (UNDP) Annual Report, (2008). www.undp.org/publications/undp-annual-report-2008

Unnikrishnan, S., & Singh, A. (2010). Energy recovery in solid waste management through CDM in India and other countries. Resources, Conservation and Recycling, 54(10), 630-640.

Van de Klundert, A. and Anschutz, J. (2001). Integrated Sustainable Waste Management-the Concept, Tools for Decision-Makers, Experiences from the Urban Waste Expertise Programme, 1995-2001

van den Wollenberg A. L. (1977) Redundancy analysis an alternative for canonical correlation analysis. Psychometrika 42:207–219

Venkata Rao, R. (2016) Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems. International Journal of Industrial Engineering Computations, 7(1):19–34

Vicente, P., & Reis, E. (2008). Factors influencing households' participation in recycling. Waste Management & Research, 26(2), 140-146.

Vijay, V. K. Biogas Entrepreneurship. Retrieved from https://web.iitd.ac.in/~vkvijay/files/biogas%20entrepreneurship.pdf

Vögeli, Y., & Zurbrügg, C. (2008). Biogas in cities-a new trend. Sandec News, 9(2008), 8-9.

Wan, X., Li, J., Xie, L., Wei, Z., Wu, J., Tong, Y. W., ... & Zhang, J. (2022). Machine learning framework for intelligent prediction of compost maturity towards automation of food waste composting system. Bioresource Technology, 365, 128107

Wang, F., Cheng, Z., Reisner, A., & Liu, Y. (2018). Compliance with household solid waste management in rural villages in developing countries. Journal of Cleaner Production, 202, 293-298.

Wang, H., Qin, Y., Xin, L., Zhao, C., Ma, Z., Hu, J., & Wu, W. (2023). Preliminary techno-economic analysis of three typical decentralized composting technologies treating rural kitchen waste: a case study in China. Frontiers of Environmental Science & Engineering, 17(4), 47.

Wang, R., Ren, T., Feng, L., Wang, T., & Wang, T. (2023). Prediction of the Moisture Content in Corn Straw Compost Based on Their Dielectric Properties. Applied Sciences, 13(2), 917.

Wang, Y. Q., Zhang, X. Y., & Draxler, R. R. (2009). TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data. Environmental Modelling & Software, 24(8), 938-939.

Wang, Y., Huntington, T., & Scown, C. D. (2021). Tree-based automated machine learning to predict biogas production for anaerobic co-digestion of organic waste. ACS Sustainable Chemistry & Engineering, 9(38), 12990-13000.

Wang, Y., Huntington, T., & Scown, C. D. (2021). Tree-Based Automated Machine Learning to Predict Biogas Production for Anaerobic Co-digestion of Organic Waste. ACS Sustainable Chemistry & Engineering 9:12990–13000

Wang, Y., Lü, F., Kang, X., Xu, X., Chen, W., Chai, H., ... & He, P. (2022). Odor characteristics and health risks during food waste bioconversion by housefly (Musca domestica L.) larvae. Journal of Cleaner Production, 376, 134343.

Wannasek, L., Ortner, M., Amon, B., & Amon, T. (2017). Sorghum, a sustainable feedstock for biogas production? Impact of climate, variety and harvesting time on maturity and biomass yield. Biomass and Bioenergy, 106, 137-145.

Wilkinson, L. (1979). Tests of significance in stepwise regression. Psychological bulletin, 86(1), 168.

Wiśniewska, M., Kulig, A., & Lelicińska-Serafin, K. (2021). Odour nuisance at municipal waste biogas plants and the effect of feedstock modification on the circular economy—a review. Energies, 14(20), 6470.

Woon, K. S., & Lo, I. M. (2016). A proposed framework of food waste collection and recycling for renewable biogas fuel production in Hong Kong. Waste Management, 47, 3-10.

World Bank Report (2019). pubdocs.worldbank.org/en/195181530913257957/2019-WDR-PPT.pdf

WRAP, 2014 A survey of the UK Anaerobic Digestion industry in 2013. Prepared by LRS Consultancy, Banbury, UK 2014, UK

Wu, A., Lovett, D., McEwan, M., Cecelja, F., & Chen, T. (2016). A spreadsheet calculator for estimating biogas production and economic measures for UK-based farm-fed anaerobic digesters. Bioresource technology, 220, 479-489.

Wu, S., Ni, P., Li, J., Sun, H., Wang, Y., Luo, H., ... & Dong, R. (2016). Integrated approach to sustain biogas production in anaerobic digestion of chicken manure under recycled utilization of liquid digestate: Dynamics of ammonium accumulation and mitigation control. Bioresource technology, 205, 75-81.

Xu, J., Shi, Y., Xie, Y., & Zhao, S. (2019). A BIM-Based construction and demolition waste information management system for greenhouse gas quantification and reduction. Journal of Cleaner Production, 229, 308-324.

Xu, M., Yang, M., Sun, H., Meng, J., Li, Y., Gao, M., ... & Wu, C. (2022). Role of multistage inoculation on the co-composting of food waste and biogas residue. Bioresource Technology, 361, 127681.

Xue, W., Hu, X., Wei, Z., Mei, X., Chen, X., & Xu, Y. (2019). A fast and easy method for predicting agricultural waste compost maturity by image-based deep learning. Bioresource Technology, 290, 121761

Yang, X. S. (2019). Introduction to algorithms for data mining and machine learning. Academic press.

Yoshizaki, T., Shirai, Y., Hassan, M. A., Baharuddin, A. S., Abdullah, N. M. R., Sulaiman, A., & Busu, Z. (2012). Economic analysis of biogas and compost projects in a palm oil mill with clean development mechanism in Malaysia. Environment, Development and Sustainability, 14, 1065-1079.

You X, Wu D, Wei H, et al (2018). Fluoroquinolones and β-lactam antibiotics and antibiotic resistance genes in autumn leachates of seven major municipal solid waste landfills in China. Environment International 113:162–169.

Zade, M. J. G., & Noori R. E. (2008). Prediction of municipal solid waste generation by use of artificial neural network: A case study of Mashhad. 13-22.

Zaki, M. J., & Meira Jr, W. (2020). Data mining and machine learning: Fundamental concepts and algorithms. Cambridge University Press.

Zawadzka, A., Krzystek, L., Stolarek, P., & Ledakowicz, S. (2010). Biodrying of Organic Fraction of Municipal Solid Wastes. Drying Technology, 28(10), 1220–1226. https://doi.org/10.1080/07373937.2010.483034

Zhang, C., Dong, H., Geng, Y., Song, X., Zhang, T., & Zhuang, M. (2022). Carbon neutrality prediction of municipal solid waste treatment sector under the shared socioeconomic pathways. Resources, Conservation and Recycling, 186, 106528. https://doi.org/10.1016/j.resconrec.2022.106528

Zhang, H., Duan, H., Andric, J. M., Song, M., & Yang, B. (2018). Characterization of household food waste and strategies for its reduction: A Shenzhen City case study. Waste Management, 78, 426-433.

Zhang, H.-H., He, P.-J., & Shao, L.-M. (2008). N2O emissions from municipal solid waste landfills with selected infertile cover soils and leachate subsurface irrigation. Environmental Pollution, 156(3), 959–965. https://doi.org/10.1016/j.envpol.2008.05.008

Zhang, J., Li, N., Dai, X., Tao, W., Jenkinson, I. R., & Li, Z. (2018). Enhanced dewaterability of sludge during anaerobic digestion with thermal hydrolysis pretreatment: new insights through structure evolution. Water Research, 131, 177-185.

Zhang, Y., Li, L., Ren, Z., Yu, Y., Li, Y., Pan, J., ... & Han, Y. (2022). Plant-scale biogas production prediction based on multiple hybrid machine learning technique. Bioresource Technology, 363, 127899.

Zhang, Y., Lu, W., Wing-Yan, Tam, V., & Feng, Y. (2018). From urban metabolism to industrial ecosystem metabolism: a study of construction in Shanghai from 2004 to 2014. J Clean Prod 202:428–438.

Zhou, H., Long, Y., Meng, A., Li, Q., & Zhang, Y. (2015). Classification of municipal solid waste components for thermal conversion in waste-to-energy research. Fuel, 145, 151-157.

Zhu, L., & Rahman, K. A. (2020). Impact of purchasing power parity and consumption expenditure rise on urban solid waste generation in China. International Journal of Asian Social Science, 10(9), 458-470.

Zupančič, G. D., & Grilc, V. (2012). Anaerobic treatment and biogas production from organic waste. Management of organic waste, 2.

Zurbrügg, C., Drescher S., Patel, A., & Sharatchandra H. C. (2004). Decentralised composting of urban waste – an overview of community and private initiatives in Indian cities. Waste Management 24:655–662.

Appendix A: Survey Questionnaires and Stakeholders' Information

Survey Questionnaire (Households)

This interview is made to you to undertake a study for the partial fulfilment of the requirements for the PhD degree. The goal of this study is to gather preliminary information to assess the importance of solid waste management for sustainable development in Guwahati city. The indirect benefit of the study is to improve understanding of the impact of solid waste on the environment and municipal solid waste management practices in Guwahati city. Your response will help policy makers to formulate an informed policy for a sustainable integrated solid waste management (ISWM). The interview will take a few minutes and the answers will be completely confidential and strictly for academic purpose. Thus, please answer the questions honestly and as truthfully as you can.

- **Participation in the study is completely voluntary.**

- **The questionnaire is to be answered by the owner of the house.**

A.Household Details

1. Name of the respondent: ……

2. Are you the head of the household? Yes/No

3. Name of the head of the house hold: …….

4. Total members of the Household ……

   Male……. Female…… Children (6-14) ……

   Kids (1- 5 years) ……. Infants (<1 year) …….

5. Education of the HH head……: Highest education among the members of the HH…….

   0. Illiterate; 1. Primary; 2. Middle; 3. Matric; 5. BA/BSc/BCS; 6. MA/MSc; 7. Above

6. Total number of HH who are employed…….

7. Employment status of Head of Household Head

   1. Unemployed 2. Street Vendor/Small Informal Business

   3. Government Employee 4. Own Business

   5 Private Employee 6. Other

8. Average Monthly Household Income of all earning members

1. Less than ₹10,000 2. ₹ 10,001-20,000

3. ₹ 20,001-50,000 4. ₹ 50,001-100,000

5. Greater than ₹100,000

9. How long do you spend on your smartphone/TV on a weekday/ weekend?

1. Four hours 2. Five hours

3. Six-seven hours 4. More

10. Do you think that media has raised your awareness about water, sanitation and solid waste management?

Yes 1 No 0 (if yes cont. to Q.11)

11. What type of mass media component was more effective in generating your awareness?

1. Radio 2. Television 3. Newspaper 4. Social media

B. Home Composting Questionnaire

Please tick the answers that are relevant to you

1. Do you own a composting bin?

Yes ☐

No ☐

2. How many composting bins do you have?

One ☐

Two ☐

More ☐

None ☐

3. If you **do not compost**, would you consider starting composting at home?

Yes ☐

No ☐

If no, please state why

……...……

……

4. Which type of composting will you prefer? ☐

Self-composting

Community-based composting ☐

Outsourced composting ☐

5. Which type of property do you live in? ☐

Detached house

Semi-detached house ☐

Terrace house ☐

Bungalow ☐

Cottage ☐

Flat ☐

6. Do you own or rent your house?

Own ☐ Rent ☐

7. Do you take waste for disposal?

If yes, please specify which wastes and how often ……..……..

How many bags approximately?

One ☐

Two ☐

3-5 ☐

6-8 ☐

8. Does a maid take care of waste management at your home?

Yes ☐

No ☐

If so, then what is the policy of your Household?

……

……

……

9. What is the compostable waste generation in a day?

1 kg ☐ 4-5 kg ☐

2-3 kg ☐ More ☐

10. Where will you perform the composting for self-composting?

Backyard composting ☐

Terrace composting ☐

11. Where will you perform the composting for community-based composting?

Backyard composting ☐

Terrace composting ☐

Community composting ☐

12. Where will you perform the composting for outsourced composting?

Backyard composting ☐

Terrace composting ☐

Community composting ☐

13. How would you compost?

Compost Bin ☐

Compost heap ☐

Other ☐

If other, please specify ……

1. Introduction

This questionnaire is designed to facilitate the assessment of the current situation of solid waste management service in an urban area. To enable an accurate assessment, it is important that all information requested in the questionnaire should be provided as completely and accurately as possible.

2. General Information

2.1 Name and address of authority responsible for solid waste management

……

……

……

2.2 Area of jurisdiction

Urban area ……. (sq. km)

Rural area ……. (sq. km)

Total area ……. (sq. km)

2.3 Population

	1991	2001	2011
Urban Population			
Rural Population			
Total Population			

3. Planning and Development

3.1 Physical characteristics of solid waste

If data on waste characteristics are available, please complete the following table:

Component	% By Weight
Paper	
Plastic and rubber	
Organic or vegetables	
Glass and ceramic	
Ferrous metal	
Aluminium	
Wood	
Textile	
Garden waste	
Others	
Total	

3.2 Physical characteristics of solid waste

(1) Does the Department have a storage bin standardization policy If so, please briefly outline the policy.

……

……

……

……

……

……

……

(2) Type of storage bin used (please tick appropriate space)

Type of Containers		Residential Area				Commercial Area			
		A	**F**	**S**	**N**	**A**	**F**	**S**	**N**
Individual Containers	Metal bin								
	Plastic bin								
	Plastic bag								
	Oil drum								
	Others								
Communal Containers	Metal bin								
	Plastic bin								
	Oil drum								
	Roll-on-roll-off								
	Concrete bin								
	Others								

A= Almost exclusively used

F= Frequently used

S= Sometimes used

N= Never used

3.3 Collection

(1) Collection service coverage for domestic premises for the year 20.....

	% of Total Population	Frequency of Collection
Urban Population		
Rural Population		

(2) Collection service coverage for commercial/ trade premises for the year 20.....

	% of Premises	Frequency of Collection
Collected by the Municipality		
Collected by the Municipal contractor		
Collected by owner's contractor		

(3) Amount of waste collected (by both the Department and contractors) last year i.e. 20.....

Waste Type	**Estimated Recycling Rate (%)**	**Amount Collected**			
		Measure	**Estimate**	**Measured**	**Estimated**
Domestic, institutional, commercial and trade waste					
Industrial waste					
Street/park waste					
Drain waste					
Bulky waste					
Total					

3.4 Disposal

Items	**Disposal Site**
Name of the site	
Total area (acre/hectare)	
Year when disposal started	
Amount of waste deposited daily (tons/day)	

Distance from collection area to the site (km)	
Disposal method (*)	
Existence of animals on site	Yes/No
Existence of waste pickers or scavengers on site	Yes/No
Existence of open burning on site	Yes/No

*Note: For disposal method, please specify as follow:

O = Open dumping

C = Controlled tipping (with occasional soil cover)

S = Sanitary landfill (with daily cover)

D = Dumping into water body (river etc.)

4. Operation

4.1 Contractual services

Service Component	Proportion of contractual service (last 3 years)			Number of contractors in last 3 years		
	20….	20….	20….	20….	20….	20….
Collection and transport						
Street sweeping						
Grass cutting						
Landfill operation						
Vehicle maintenance						
Others						

4.2 Vehicles and equipment

(1) General information

Is there any policy to standardize the vehicles and equipment used by the Municipality? If so, please outline how this policy is being implemented.

……

……

……

……

(1) Equipment for prints collection (i.e., collection of solid waste front households to communal bin or depot for subsequent collection by collection vehicles)

Equipment Type	Number	Average capacity (m^3)
Wheel barrows (1 wheel)		
Push carts (2-4 wheels)		
Others		

(2) Vehicles

Vehicle type	**No.**	**Ave. Cap. m^3**	**No. of vehicle by condition (See note below)**			**No. of vehicle by age (year)**			
			G	**F**	**B**	**> 10**	**S—10**	**2—5**	**<2**
Compactor vehicles									
Tipping truck with sliding cover									
Open truck with tipping mechanism									
Open truck without tipping mechanism									
Open truck with crane									
Tilt-frame or hoist truck handling bid metal bin									
Nightsoil tanker									

Vacuum truck									
Tractor									
Vehicle for administration									
Others									

Note: G - Good condition. F = Fair condition, B = Bad condition

(4) Machinery used in landfill, including machinery owned by both the municipal corporation and contractors

Machinery type	**No.**	**No. of machinery bv condition**			**No. of machinery by age (years)**			
		G	**F**	**B**	**>10**	**5-10**	**2-5**	**<2**
Bulldozers								
Bucketloaders								
Backhoes								
Compactors								
Tractors								
Others								

List of NGOs engaged in waste collection under Guwahati Municipal Corporation

Sl No	**Ward No**	**N G O Name**	**Area covered**	**No. of Allotted House Holds**
1	1	Udayan Social Welfare	Jalukbari, Sundarbari, Lankeshwar	7152
2	2	Yuva Prerona NGO	Pandu	6158

3	3	Naba Pragati Development Society	Adabari	5110
4	4	Shyamkanu Samajik Kalyankami Anusthan	Maligaon, Borabazar	5192
5	5	Puberun Social Welfare	Maligaon Railway Colony	7235
6	6	Sibanga NGO	Tetelia, Dhirenpara	6540
7	7	Udayan Social Welfare	Gorchuk, Kotabari	8021
8	8	Nabadeep Social Welfare Society	Kumarpara, Santipur	6175
9	9	Akashi NGO	Paltanbazar	6543
10	10	Jiban Sathi Welfare Society	Fancybazar	6609
11	11	Subhakangsha	Kharghuli, Uzanbazar	5885
12	12	Bahnisikha	Chandmari	5910
13	13	Nabarup	Goswami Service, Rajgarh road	6628
14	14	Suraj	Ulubari, Rehabari	6140
15	15	Asthitya	Ambari, Lamb road, Guwahati Club	7604
16	16	Bonde	Kalapahar, Cycle factory	7920
17	17	New Evergreen NGO	Lalganesh	5907
18	18	Pragati Sangha	Bhangagarh, Birubari	5976
19	19	Suprabhat Welfare Society	Nabagraha	6751
20	20	Gitanjali	Anil Nagar, Zoo Road	6079
21	21	Ma Kamakhya	Nabin nagar	7048
22	22	Pragjyotish Samaj Kalyan Society	Noonmati	7959
23	23	Samannaya Gosthi	Jyotinagar	7270
24	24	Amar Prayash	Narengi	8642
25	25	Uttaran Social Welfare Society	Hengarabari	8839
26	26	Uttaran Social Welfare Society	Hatigaon, Downtown Hospital	8198
27	27	Bharati Yuba Shakti NGO	Kahelipara	7550
28	28	Enajori	Beltola	6992
29	29	Orion Society	Sijubari, Mazhar road	6550
30	30	Pragjyotish Samaj Kalyan Society	Basistha Chariali	7703
31	31	Lokakalyan Samajik Santha	Khanapara, Panjabari	7806
		Total		**214092**

Lorry

To,

The Commissioner,

Guwahati Municipal Corporation, Guwahati

Subject: Kind request and consent towards providing **unit-level data of Guwahati city** for my PhD thesis related dissertation on the dynamics and impact of solid waste generation in Guwahati city.

Respected Sir,

I am Tinka Singh, and currently a research scholar from the Centre for the Environment, IIT Guwahati. Presently, I am working on **solid waste management in smart cities** and am researching exclusively towards Capital cities and towns of Northeast India. As a part of my PhD thesis work, I am collecting solid waste management-related data for smart cities. Thereafter, I intend to use the collected data to gain useful insights into the dynamics and impact of solid waste generation in Guwahati city.

Thereby, on behalf of myself and my supervisor Prof. Ramagopal V. S. Uppaluri, we heartily express our gratitude for considering our plea and providing us the valuable data related to unit-level data related to solid waste management parameters in Guwahati city. Kindly do indicate to us the protocols that need to be followed along with privacy regulations.

Looking forward to your kind consent and needful sharing of requested data at the earliest. Thanking you.

Yours sincerely,

Tinka Singh

Tinka Singh

Research Scholar, IIT Guwahati

Forwarded, Recommended & Consented

Ramagopal

11/11/21

Prof. **Ramagopal Uppaluri**
Department of Chemical Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, INDIA

JC (II)
Pl help.

12/11/21

9085063663
UTAL J.B

01. provide the required data.

12/11/21

To,

The Member Secretary

Pollution Control Board Assam, Guwahati

Subject: Kind request and consent towards providing **solid waste data of Guwahati city** for my PhD thesis related dissertation on the dynamics and impact of solid waste generation in Guwahati city.

Respected Sir,

I am Tinka Singh, and currently a research scholar from the Centre for the Environment, IIT Guwahati. Presently, I am working on **solid waste management in smart cities** and am researching exclusively towards Capital cities and towns of Northeast India. As a part of my PhD thesis work, I am collecting solid waste management-related data for smart cities. Thereafter, I intend to use the collected data to gain useful insights into the dynamics and impact of solid waste generation in Guwahati city.

Thereby, on behalf of myself and my supervisor Prof. Ramagopal V. S. Uppaluri, we heartily express our gratitude for considering our plea and providing us the valuable data related to unit-level data related to solid waste management parameters in Guwahati city. Kindly do indicate to us the protocols that need to be followed along with privacy regulations.

Looking forward to your kind consent and needful sharing of requested data at the earliest. Thanking you.

Yours sincerely,

Tinka Singh 08-11-2021

Tinka Singh

Research Scholar, IIT Guwahati

Consented, Recommended and forwarded

Ramagopal

8|11|21

Prof. Ramagopal Uppaluri
Department of Chemical Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, INDIA

To,

The Additional Director of Economics & Statistics,

Directorate of Economics & Statistics, Guwahati

Subject: Kind request and consent towards providing **unit level data of Guwahati city** for my PhD thesis related dissertation on the dynamics and impact of solid waste generation in Guwahati city.

Respected Sir,

I am Tinka Singh, and currently a research scholar from the Centre for the Environment, IIT Guwahati. Presently, I am working on **solid waste management in smart cities** and am researching exclusively towards Capital cities and towns of Northeast India. As a part of my PhD thesis work, I am collecting solid waste management-related data for smart cities. Thereafter, I intend to use the collected data to gain useful insights into the dynamics and impact of solid waste generation in Guwahati city.

Thereby, on behalf of myself and my supervisor Prof. Ramagopal V. S. Uppaluri, we heartily express our gratitude for considering our plea and providing us the valuable data related to unit-level data related to solid waste management parameters in Guwahati city. Kindly do indicate to us the protocols that need to be followed along with privacy regulations.

Looking forward to your kind consent and needful sharing of requested data at the earliest. Thanking you.

Yours sincerely,

Tinka Singh
08-11-2021
Tinka Singh

Research Scholar, IIT Guwahati

Consented, Recommended and forwarded

Ramagopal
8/11/21

Prof. Ramagopal Uppaluri
Department of Chemical Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, INDIA

OFFICE OF THE GUWAHATI MUNICIPAL CORPORATION
BHANGAGARH, GUWAHATI-5

No. GER/MSWM (P)/10/2023 Pt-I/27 dated 21/09/2023

To

Tinka Singh,
Research Scholar,
Indian Institute of Technology (Guwahati)

Sub: "Characterization of Municipal Solid Waste Generation and Seasonal Classification
for Various Socio-demographic Groups in Guwahati City"

Ref : Your communication via mail dated 19th Sept' 2023

Dear Madam,

With reference to you communication, we would like to acknowledge the summery of your study on the subject mentioned above and appreciate your endeavour in this important sector of Municipal Solid Waste Management. Although we did not get the opportunity to go through the complete study-report, from the submitted summery, it is understood that the selected area of study is of utmost importance for the urban managers for tailoring their waste-management strategies. Characterisation of MSW generation with seasonal variations and on the basis of lower, middle and high economic groups of generators is necessary for taking up viable project initiatives for up-gradation of the existing MSWM infrastructures.

Best wishes for your future endeavours.

Executive Engineer
Municipal Solid Waste Management Project,
Guwahati Municipal Corporation
Dispur Super-market, Guwahati-05

Memo no. GER/MSWM (P)/10/2023 Pt-I/27 dated 21/09/2023

Copy to-

1. The Commissioner, Guwahati Municipal Corporation for information.
2. Additional Commissioner, GMC for information
3. Joint Commissioner, GMC for information.
4. The Additional Chief Engineer, GMC for information.
5. File

Executive Engineer
Municipal Solid Waste Management Project,
Guwahati Municipal Corporation

Appendix B: Summarized Model Parameters Values for Biogas Yield Estimations

For the AD calculations, Table B1 presents average values and multiplication factors used to estimate the total plant volume based on the rated daily gas production of different types of plants. Table B2 summarizes the volume calculations for the considered digesters. Additionally, Table B3 provides the summarized parameter values and references for these parameters, which are used in the model. Lastly, Table B4 includes literature-reported values for potential biogas yield.

Table B1 Proportions and multiplication factors for estimating digester and gas storage volume from total plant volume and rated daily gas production.

Plant types	***Volume estimations of the digester and gas storage***			***Multiplication factors for estimating plant volume from rated daily gas production***		
	Digester volume	***Gas storage volume***	***Total volume***	***Digester volume***	***Gas storage volume***	***Total volume***
Fixed dome plant	80%	20%	100%	2.4	0.6	3.0
Floating drum plant	70%	30%	100%	1.4	0.6	2.0
Balloon/bag digester	75%	25%	100%	1.8	0.6	2.4

Table B2 Total plant volume calculations for different plant types.

Plant Types	**Plant volume calculations**
Fixed dome plant (Deenbandhu design)	$V_P = \frac{2}{3}\pi r^3 + \frac{1}{6}\pi k(3r^2 + k^2)$
Floating drum plant	$V_P = V_d + V_g$
Balloon digester	$V_P = \pi\left(\frac{D}{2}\right)^2 L$

Table B3 Fixed values for parameters to calculate biogas yield.

RT of the feedstock (in days)	Temperature (°C)					
	16-18	***19-21***	***22-24***	***25-27***	***28-30***	***31-33***
6-10	5.41	7.98	10.83	13.59	15.91	18.33
11-15	4.73	6.79	8.99	11.09	12.88	14.74
16-20	4.21	5.90	7.68	9.37	10.82	12.32
21-25	3.79	5.22	6.70	8.11	9.33	10.59
26-30	3.44	4.69	5.95	7.15	8.20	9.28
31-35	3.16	4.25	5.35	6.39	7.32	8.26
36-40	2.91	3.88	4.86	5.78	6.60	7.44
41-45	2.71	3.58	4.45	5.27	6.02	6.77
46-50	2.53	3.32	4.10	4.85	5.53	6.21
51-55	2.37	3.09	3.81	4.49	5.11	5.74
56-60	2.23	2.89	3.55	4.18	4.75	5.33
61-65	2.10	2.72	3.33	3.91	4.44	4.98
66-70	1.99	2.57	3.13	3.67	4.17	4.67
71-75	1.89	2.43	2.95	3.46	3.93	4.40
76-80	1.80	2.30	2.80	3.27	3.71	4.15
81-85	1.72	2.19	2.66	3.10	3.52	3.94
86-90	1.65	2.09	2.53	2.95	3.34	3.74
91-95	1.58	2.00	2.41	2.81	3.19	3.56
96-100	1.52	1.92	2.31	2.69	3.04	3.40

*Sources: IRENA, 2016

Table B4 Rates of biogas yield from different fermentation materials and their primary chemical constituents.

Feed material	VS (%)	Biogas yield m^3 BG/kg VS	Density kg m^{-3}	References
Rice straw	0.36	0.578	380	Rajendran et al., 2012
Wheat straw	0.39	0.516	177	Fulford, 2015; SEAI (2012)
Grass	0.51	0.656	485	Seadi et al., 2008
Corn stalk	0.43	0.611	127	Fulford, 2015; SEAI (2012)
Fruit waste	0.14	0.626	621	Rajendran et al., 2012
Vegetable waste	0.16	0.654	540	SEAI (2012)
Mixed food waste	0.08	0.720	500	Seadi et al., 2008
Mixed organic waste	0.26	0.550	502	Seadi et al., 2008

Appendix C: Equations for the Design of Composting Facilities

Tipping Area

$$Area_{tipping} = \frac{Storage_{tipping} \times vol_1}{Height_{tipping}} \quad \textbf{(A-1)}$$

Where $Area_{tipping}$ is the tipping floor area (m^2), $Storage_{tipping}$: time requirements for storage and equipment downtime; default value 2 days, $Height_{tipping}$ is the height of wastes in tipping floor; default value 2 m.

The calculation of the shredding/screening area takes into account the number of operational screens and hammermills in the staging area, as well as their individual footprint areas. An allocated space will be designated specifically for storing the windrow turners and front-end loaders.

Buffer zone area

The determination of the buffer distance relies on local regulations, which assign varying buffer distances for different areas such as inhabited areas, water bodies, protected area, agricultural land, etc. Typically, buffer distances can range from 30 m to 500 m. The calculation of the buffer area takes into consideration the buffer zone distance, the facility's length-to-width ratio, and the total area of the facility and is calculated as follows:

$$Width_{facility} = \frac{Area_{staging} + Compost_{padding} + Curring\ pad + Office\ area + Area_{reject}}{Road_{width}} \quad \textbf{(A-2)}$$

$$Lenght_{facility} = \frac{length}{width} \times \left(Width_{facility} - Width_{road}\right) + Width_{road} \quad \textbf{(A-3)}$$

$$Area_{facility} = Area_{staging} + Compost_{padding} + Curring\ pad + Area_{rejected} + Area_{office} + Area_{staging} + Area_{buffer} \quad \textbf{(A-4)}$$

where $Area_{staging}$ is the total staging area (m^2), $Curring\ pad$ is the total required curing stage area (m^2), $Area_{curing}$: area required for curing of wastes in HSCF (m^2), $Area_{office}$: total area of office space (m^2), $Office_{coeff}$: office space coefficient; default of 20 m^2/employee, $Area_{rejected}$ is the area required for temporary storage of

rejected wastes (m^2), $Width_{facility}$: facility width (m), ratio of facility length to facility width; default of 2, $Lenght_{facility}$: facility length (m), $Width_{road}$: road width; default 5 m, $Area_{buffer}$ is the buffer zone area (m^2), $Distance_{buffer}$: buffer zone distance; varies with local legislation and is a function of the location of the composting facility and adjacent sites (e.g., rivers, lakes, wells, EPAs etc.); since odour control systems are used for the MSCF and HSCF, a default of 153 m will be used; a 60 m buffer distance will be used for the YWCF and $Area_{facility}$ is m^2 (Karmakar et al., 2007; Komilis and Ham, 2004).

Grading cost

The grading cost for the facility does not include the buffer area, as it will not be subject to grading. Here is the breakdown of the facility grading cost:

$$Cost_{grad} = grad\ unit\ cost \times (Area_{fac} - Area_{buffer}) \quad \textbf{(A-5)}$$

Paving cost

The paving requirements for different areas of the facility will vary. Specifically, the roads, staging area, rejects area, and curing pad will be paved with 4 inches of asphalt and 8 inches of gravel. On the other hand, the composting pad will be paved with 2 inches of asphalt and 8 inches of gravel. The office area will have a direct cost coefficient (in $/ft2) assigned to it, which encompasses both the paving and the building. It's important to note that the buffer zone will not undergo any paving.

$$Cost_{pav} = pav\ unit\ cost \times (Area_{stag} + Area_{road} + curring\ pad + Area_{rej}) \quad \textbf{(A-6)}$$

$$Cost_{fence} = unit\ cost\ of\ the\ fence \times perimeter\ of\ the\ facility \times 3 \quad \textbf{(A-7)}$$

Building cost

The cost associated with the steel fencing can vary significantly based on whether the composting and curing pads are covered. In the MSCF, the composting pad, including side walls, will be covered. In the HSCF, both the composting and curing pads will be covered. Additionally, in both types of MSW compost facilities, the staging area will also be covered. Furthermore, all three composting facilities will include a building dedicated to office space.

$$Cost_{build} = office\ unit\ cost \times Area_{off} \times 9 + building\ unit\ cost \times \left(Area_{stag} + compost\ pad\right) \times 9$$ **(A-8)**

Equipment cost (Source: Tchobanoglous et al. (1993))

$$Cost_{turner} = number\ of\ turner \times turner\ unit\ cost$$ **(A-9)**

$$Cost_{hammer} = number\ of\ hammer \times hammer\ unit\ cost$$ **(A-10)**

$$Cost_{grinder} = number\ of\ grinder \times grinder\ unit\ cost$$ **(A-11)**

$$Cost_{trommel} = number\ of\ trommel \times trommel\ unit\ cost$$ **(A-12)**

$$Cost_{FEL} = number\ of\ FEL \times FEL\ unit\ cost$$ **(A-13)**

Overhead cost

The calculation of overhead costs involves determining a fraction of labour salaries. Overhead costs encompass various expenses such as overtime, office supplies, insurance, social security, vacation, sick leave, and other services. Overhead is calculated by the following equation:

$$Cost_{overhead} = 40\% \times Cost_{labour}$$ **(A-14)**

Windrow turner maintenance cost

By excluding labor and diesel fuel costs, the average hourly maintenance costs are anticipated to fall within the range of $20 to $50 per operational hour. A default value of $25 per hour will be employed for calculation purposes.

Also, this cost encompasses expenses related to flail replacement, hydraulic filter replacement, other replacement parts, and routine maintenance. The subsequent equations are utilized to estimate the annual maintenance costs for windrow turners.

$$Turner_{maint} = \frac{average\ hourly\ cost\ for\ windrow\ turner\ operation \times hours\ of\ operation\ of\ turner\ per\ week}{mass_1 \times operating\ days\ per\ week}$$ **(A-15)**

Hammermill maintenance cost

Diaz et al. (2007) conducted a study to estimate the maintenance costs associated with hammermills. They compared the maintenance costs for three options: performing hammer build-up once per week, daily hammer buildup, and the "wear and scrap" option.

$$Hammer_{maint} = hammer\ unit\ maintenance \times \frac{mass_3}{mass_1} \quad \textbf{(A-16)}$$

Utilities cost

The primary utility cost that needs to be considered is electricity. Electricity is utilized by specific equipment and is consumed during the operation of the odour-control system and buildings.

$$Cost_{util} = \frac{Elect_{hp} \times 0.746 \times unit\ cost\ of\ electricity}{mass_1 / operating\ hours} \quad \textbf{(A-17)}$$

Appendix D: Redundancy Analysis and Standardized Scores

Table D1 Eigenvalues and Percentage of Inertia for unconstrained RDA.

	Eigenvalue	Percentage of Inertia	Cumulative Inertia
RDA1	90.34115	52.12%	52.12%
RDA2	34.9286	20.15%	72.27%
RDA3	13.29827	7.67%	79.95%
RDA4	7.52388	4.34%	84.29%
RDA5	3.98102	2.30%	86.58%
RDA6	2.502	1.44%	88.03%
RDA7	0.72288	0.42%	88.44%
RDA8	0.30543	0.18%	88.62%

Table D2 Standardized scores for PC axes.

PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7	PC 8	Location
58.81%	**22.74%**	**8.66%**	**4.90%**	**2.59%**	**1.63%**	**0.47%**	**0.20%**	
			Standardized Scores					***Scores Labels***
-1.49785	1.06045	-0.55837	-0.23705	-2.90141	6.7332	4.2566	12.34898	**w19**
-1.50183	0.41232	-2.89626	-0.99788	-3.24413	-3.81352	0.02263	-5.68271	**w21**
0.2177	1.68813	2.32797	-4.9239	0.91532	-1.11834	-3.76734	-8.60793	**w24**
1.79285	2.22937	-1.25149	-0.71835	1.68263	-3.14347	-2.19006	2.92782	**w22**
0.46309	0.54241	0.98924	1.43315	-0.13851	3.30025	7.67789	13.01891	**w28**
1.92077	3.61526	-0.8144	3.31068	0.95598	-1.11775	-4.68722	-13.67013	**w30**
-2.45759	-1.06879	1.11482	2.06777	-0.47117	-1.48579	-6.63573	-6.02726	**w4**
-2.68953	-1.66277	-2.85808	-0.2161	4.13755	2.2413	2.5223	2.95055	**w15**
-1.77439	-0.98004	4.29688	0.62388	0.16985	-2.13893	3.61722	4.56081	**w16**

RESEARCH OUTPUT

Published articles in International refereed Journals

1. Singh, T., & Uppaluri, R. V. S. (2022). Machine learning tool-based prediction and forecasting of municipal solid waste generation rate: a case study in Guwahati, Assam, India. International Journal of Environmental Science and Technology, 1-24. **(IF= 3.5)**
2. Singh, T., & Uppaluri, R. V. (2023). Application of ANN and traditional ML algorithms in modelling compost production under different climatic conditions. Neural Computing and Applications, 1-20. **(IF=6.0)**
3. Singh, T., & Uppaluri, R. V. S. (2023) Optimizing biogas production: A novel hybrid approach using anaerobic digestion calculator and machine learning techniques on Indian biogas plant. Clean Technologies and Environmental Policy. **(IF=4.7)**
4. Singh, T., & Uppaluri, R. V. (2023). Feed-forward ANN and traditional machine learning-based prediction of biogas generation rate from meteorological and organic waste parameters. The Journal of Supercomputing, 1-34. **(IF=3.3)**
5. Singh, T., Naik, A., & Uppaluri, R. V. S. (2023) Characterization of municipal solid waste and seasonal classification for various socio-demographic groups in Guwahati city. Material Cycle and Waste Management. (In Press) **(IF=3.3)**

Book Chapter

1. Singh, T., & Uppaluri, R. V. S. (2023) Prediction and forecasting municipal solid waste generation of Northeastern cities by retrofitting ML models of Guwahati, India in: R. Bhattacharjee, D. R. Neog, K. R. Mopuri, S. K. Vipparthi (Eds.) Artificial Intelligence and Data Science based R&D interventions: Proceedings of NERC 2022 Springer Nature Singapore, Singapore, 2023

Communicated/Under preparation/Collaborative work

1. Singh, T., & Uppaluri, R. V. S. (2023) ML-Based Prediction and Forecasting of GHG Emissions and Particulate Matters from MSW Landfill and Incineration in Guwahati city *(Communicated)*
2. Singh, T., & Uppaluri, R. V. S. (2023) Enhancing the Efficacy of Solid Waste Generation Prediction through Supervised ML and Time Series Modelling: A Case Study in Kamrup, Guwahati *(Communicated)*
3. Singh, T., Prasad, A., & Uppaluri, R. V. S. (2023) Machine learning for crop yield prediction and forecasting: A meteorological data-driven approach for Kamrup subdivision *(Communicated)*
4. Singh, T., & Uppaluri, R. V. S. (2023) Development of a GIS-based decision model and techno-economic calculator for compost facilities: A case study for Guwahati city *(Communicated)*