Deep Learning Models for Cardiac Abnormality Detection from ECG Signals: An Interpretability Perspective

Thesis submitted in partial fulfilment of the requirements for the award of the degree of

Doctor of Philosophy

in

Computer Science and Engineering

by

Deepankar Nankani

Under the supervision of

Dr. Rashmi Dutta Baruah

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati

Guwahati - 781039 Assam India

October, 2022

Dedicated to

My Beloved Family

Acknowledgements

I would like to take this opportunity to express my sincere thanks and heartfelt gratitude to each and every individual who directly or indirectly have helped me to achieve the goal to successfully complete my thesis.

I would like to express my sincere and heartfelt gratitude to my supervisor *Dr. Rashmi Dutta Baruah* for nurturing me to grow as a researcher. I am always indebted for her time and patience during my course of doctoral research. I am wholeheartedly thankful for her consistent support, and positive guidance that helped me to complete this work at the right time with the right perspective.

I am highly grateful to *Prof. Pradip K Das*, the chairperson of my doctoral committee, for his invaluable advice, and support throughout my PhD. I would also like to thank my other doctoral committee members - *Dr. Vijaya V. Saradhi* and *Dr. Prithwijit Guha* for their insightful comments and suggestions which made me improve the quality and clarity of my thesis work. I am also thankful to my thesis examiners: *Prof Aurobinda Routray* and *Dr. Sadasivan Puthusserypady*. I am also grateful to the DPPC Chairman: *Prof. Ashish Anand*. I am thankful to the Viva Voce committee members: *Prof S. B. Nair* and *Dr. Debanga Raj Neog*.

I would like to thank the head of the Department of Computer Science and Engineering during my PhD at IIT Guwahati for their support in availing me the departmental facilities and resources including the travel support for the conferences. I am obliged to all the faculty members, Bhriguraj Borah and Nanu Alan Kachari, Other faculty members, Staff, Security personnel for their constant help and support.

I am thankful to MHRD, ONGC for providing financial assistantship during the Ph.D. program.

I would like to express my heartfelt thanks to Anik, Mehul, Nikunj, and Nikita for their kind co-operation in the course of my work. It was a great experience to work with them. I would also like to thank all my colleagues and friends during my journey as a Ph.D. scholar. Special thanks to Pallabi Saikia for always being there for me and for creating a warm atmosphere of support and encouragement.

Finally, I would like to thank my family for their unconditional love, support, care, warmth, encouragement and sacrifices all these years. They never doubted my intentions and wholeheartedly supported me in all my endeavours and there are no words to express my love for them.

I would also like to thank Mohit Kumar, P. Bhagath, Sathisha Basavaraju, Swarup Behera, Sameer Hansda, Divya Kulkarni, Akshay Parakh, Sonia, Tushar, Suraj, Komal, Karnish, Vanshali, Sandip, Prachurya, Hemraj, Nilotpal, Maithilee, Dipojjwal, Ujjwal, Arijit, Palash, Aparajita, Madhurima, Anasua, Bhim Singh, Gaurav Talukdar, Mohit Mishra.

I am also thankful to the anonymous reviewers for making me a better researcher.

October 29, 2022 Deepankar Nankani

Declaration

I certify that

- The work contained in this thesis is original and has been done by myself and under the general supervision of my supervisor.
- The work reported herein has not been submitted to any other Institute for any degree or diploma.
- Whenever I have used materials (concepts, ideas, text, expressions, data, graphs, diagrams, theoretical analysis, results, etc.) from other sources, I have given due credit by citing them in the text of the thesis and giving their details in the references. Elaborate sentences used verbatim from published work have been clearly identified and quoted.
- I also affirm that no part of this thesis can be considered plagiarism to the best of my knowledge and understanding and take complete responsibility if any complaint arises.
- I am fully aware that my thesis supervisor is not in a position to check for any possible instance of plagiarism within this submitted work.

October 29, 2022 Deepankar Nankani

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati
Guwahati - 781039 Assam India

Dr. Rashmi Dutta Baruah
Assistant Professor
Email : r.duttabaruah@iitg.ac.in
Phone : +91-361-2583253

Certificate

This is to certify that this thesis entitled "**Deep Learning Models for Cardiac Abnormality Detection from ECG Signals: An Interpretability Perspective**" submitted by **Deepankar Nankani**, in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy, to the Indian Institute of Technology Guwahati, Assam, India, is a record of the bonafide research work carried out by him under my guidance and supervision at the Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, Assam, India. To the best of my knowledge, no part of the work reported in this thesis has been presented for the award of any degree at any other institution.

Date: October 29, 2022
Place: Guwahati

Dr. Rashmi Dutta Baruah
(Thesis Supervisor)

ABSTRACT

Cardiovascular diseases (CVDs) are unpredictable, with life-threatening consequences leading to death. Subjects with CVD have been rapidly rising, particularly within urban communities. Detecting CVD before its occurrence using an Electrocardiogram (ECG) signal helps in risk stratification, better medical assistance, and patient treatment. Deep learning models (DLM) have been widely used for detecting cardiac abnormalities. Although DLMs outperform conventional machine learning models, they lack interpretability or explanation of the model predictions. According to General Data Protection Regulation, the importance of reasons behind a diagnosis predicted by the models is also essential. In the medical domain, explanations are also crucial, and the lack of interpretability limits the real-world deployment of DLMs.

The thesis investigates the interpretability perspective of deep learning models for detecting cardiac abnormalities from ECG signals. The recordings get contaminated with low-frequency noise baseline wander (BW) noise during the ECG signal acquisition, making classification challenging. BW is removed using Variational Mode Decomposition (VMD), a signal decomposition technique, and its efficacy is compared with other signal decomposition and filtering techniques. The clean single and multichannel ECG is used for detecting irregular heartbeats and rhythmic cardiac abnormalities using DLM. The predicted diagnosis is later explained to stakeholders using interpretability techniques.

The first contribution aims at the detection and classification of irregular heartbeats such as supraventricular ectopic beat (SVEB), ventricular ectopic beat (VEB), and normal beats (N) from a single-lead ECG signal. Initially, the R-peaks are extracted from the ECG signal using fractal-based mathematical morphological operators. The peaks are used for segmenting heartbeats. The rare occurrence of SVEB and VEB stemming from complex biological, physiological systems causes scarcity of irregular heartbeats and limits the development of a generalized classifier. Therefore, SVEB, VEB, and N beats are synthesized using a deep convolution conditional generative adversarial network. The heartbeats are classified using Penalty Induced Prototype-based eXplainable Residual Neural Network (PIPxResNet) that addresses the black-box nature of deep neural networks by providing explanations. PIPxResNet encodes the temporal variations of heartbeats through a pretrained Residual Neural Network (ResNet) following the concept of task transfer learning. The algorithm then extracts prototypes that are most representative of the training dataset that explain

model predictions to general physicians. PIPxResNet has been tested on four public datasets acquired from different geographical locations.

Life-threatening arrhythmias such as atrial fibrillation (AF) and ventricular tachyarrhythmia are caused due to prolonged occurrence of SVEB and VEB. In addition, these cardiac arrhythmias occur rhythmically with varying episode lengths in the nonstationary and nonlinear ECG signals making the classification challenging. The smaller segments might miss an abnormal rhythm, and larger segments might not focus on the relevant regions causing the cardiac abnormality. Therefore, an analysis of optimum signal length is also performed in the subsequent two contributions.

In the second contribution, ventricular tachyarrhythmias such as ventricular tachycardia and ventricular fibrillation are detected using ResNet. The ResNet models were tested on five datasets acquired from different geographical locations. The predictions are interpreted using three model agnostic gradient backpropagation-based posthoc interpretability techniques such as guided backpropagation (GBP), gradient class activation map (Grad CAM), and guided Grad CAM. The techniques are validated through sanity checks by performing a data permutation and model weight randomization test. Although the techniques passed the sanity checks, Grad-CAM highlighted the most critical signal timestamps. GBP and Guided Grad CAM worked as peak detectors that help detect heartbeats.

The third contribution aims at detecting and interpreting AF from single-channel ECG using convolution neural network (CNN), attentive CNN (ACNN), ResNet, and transformer neural network (TNN). The models are verified using PhysioNet Computing in Cardiology Challenge 2017 database. ResNet achieves good performance but hinders interpretability. The interpretability aspect is incorporated through attention mechanism in ACNN and TNN models that provide explanations by highlighting clinically meaningful characteristic ECG waves responsible for/against the diagnosis. TNN highlights more meaningful timestamps compared to ACNN.

The single-channel ECG provides lower resolution and often misses information required to detect several cardiac abnormalities. Moreover, the cardiac abnormality interpreted by one cardiologist may differ when interpreted by another cardiologist, leading to multiple possible cardiac abnormalities for a multichannel ECG (MECG) recording. Therefore, the fourth contribution uses the multi (12,6,4,3,2) channel ECG that provides better resolution for classifying thirty cardiac abnormalities. Initially, a ResNet is used for single-label 12-lead ECG classification and compared with re-

current and attention-based models. Convolution-based ResNet outperformed other models in terms of performance and computation time. Next, demographic and heartbeat features are fused with static CNN to perform multilabel MECG classification. Multiple labels improved the performance drastically, but adding heartbeat features increased training and inference time multifold, and hence heartbeat features were excluded. Lastly, demographic feature fused channel-specific dynamically built CNN (CSD-CNN) is used that eliminates manual effort and provides less trainable parameters for reduced lead ECG, making the model less prone to overfitting and memory-efficient. The models are verified on six publicly available standard datasets. In addition, the CSD-CNN model introduces an interpretability mechanism that highlights the crucial leads and relevant signal timestamps responsible for cardiac pathology prediction.

The proposed methods would be useful for automated pre-screening of subjects with CVD and potentially could be incorporated into the clinical decision support system. The visual explanations improve model assessment, increase transparency, enhance trust in the model, and increase the likelihood of model acceptance by the medical practitioners.

Contents

List of Figures

List of Algorithms

List of Tables

List of Notations

Notations	Terms
μ	Mean
σ	Variance
$G(x)$	Generator
$D(x)$	Discriminator
ν	Stiffness
P_E	Encoded prototypes
P_A	Actual prototypes
N	Number of prototypes
ρ_{min}	Minimum density
ρ_{max}	Maximum density
S	Supporting prototypes
Y_r	Reliability Score
$\theta_{w,u}$	Convolution Filter Weights

List of Abbreviations

Terms	Abbreviations
1-D	One Dimensional
2-D	Two Dimensional
ANN	Artificial Neural Network
AM	Attention Mechanism
AUC	Area under the Region Operating Characteristic curve
AAMI	Association for the Advancement of Medical Instrumentation
Acc	Accuracy
ADASYN	Adaptive Synthetic Sampling
AED	Automatic External Defibrillator
AF	Atrial Fibrillation
AFL	Atrial flutter
AHA	American Heart Association
AM	Attention Mechanism
ANN	Artificial Neural Network
BBB	Bundle branch block
Brady	Bradycardia
BN	Batch Normalisation
CAM	Class Activation Map
CC	Correlation Coefficient
CNN	Convolutional Neural Network
CART	Classification And Regression Tree
CLBBB	Complete left bundle branch block
CRBBB	Complete right bundle branch block
CVD	Cardiovascular Disease
CNN	Convolutional Neural Network
CPR	Cardiopulmonary Resuscitation

CUDB	Creighton University Ventricular Tachyarrhythmia Database
DL	Deep Learning
DT	Decision Tree
DWT	Discrete Wavelet Transform
DNN	Deep Neural Network
DO	Dropout
ECG	Electrocardiogram
F1	F1-score
FC	Fully Connected
FLM	Filter Leakage measure
FN	False Negative
FP	False Positive
GA	Genetic Algorithm
GBP	Guided Backpropagation
HRV	Heart Rate Variability
IAVB	1st degree AV block
IRBBB	Incomplete right bundle branch block
IVADB	Ideo Ventricular Arrhythmia Database
LAD	Left axis deviation
LAnFB	Left anterior fascicular block
LBBB	Left Bundle Branch Block
LPR	Prolonged PR interval
LQRSV	Low QRS voltages
LQT	Prolonged QT interval
LSTM	Long-Short Term Memory
ML	Machine Learning
MIT-BIH	Massachusetts Institute of Technology-Boston's Beth Israel Hospital
MITDB	MIT-BIH Arrhythmia Database
NSIVCB	Nonspecific intraventricular conduction
NSRDB	MIT-BIH Normal Sinus Rhythm Data Base
NSR	Normal Sinus Rhythm
PAC	Premature atrial contraction
PAFDB	Paroxysmal Atrial Fibrillation Data Base
PPI	Peak-to-Peak Interval

Pr	Precision
PR	Pacing rhythm
PRWP	Poor R wave Progression
PVC	Premature ventricular contractions
PSD	Power Spectral Density
QAb	Q wave abnormal
RAD	Right axis deviation
RBBB	Right Bundle Branch Block
ReLU	Rectified Linear Unit
ResNet	Residual Neural Netowork
RMSE	Root Mean Square Error
ROC	Receiver Operating Characteristic
RNN	Recurrent Neural Network
RRI	R-to-R Peak Interval
SA	Sinus arrhythmia
SB	Sinus bradycardia
SCD	Sudden Cardiac Death
SMOTE	Synthetic Minority Oversampling Technique
STach	Sinus tachycardia
SVM	Support Vector Machine
SVM-SMOTE	Borderline SMOTE using Support Vector Machine
SVPB	Supraventricular premature beats
TAb	T wave abnormal
TInv	T wave inversion
TN	True Negative
TNN	Transformer Neural Network
TP	True Positive
VF	Ventricular Fibrillation
VFDB	MIT-BIH Malignant Ventricular Ectopy Database
VMD	Variational Mode Decomposition
VPB	Ventricular premature beats
VT	Ventricular Tachycardia
WT	Wavelet Transform
ZCR	Zero-Crossing Rate

Citation to Published Work

Chapter 3 is based on the following papers:

- **Deepankar Nankani** and Rashmi Dutta Baruah. "Effective Removal of Baseline Wander from ECG Signals: A Comparative Study." In International Conference on Machine Learning, Image Processing, Network Security and Data Sciences (MIND), pp. 310-324. Springer, Singapore, 2020.
- **Deepankar Nankani**, Bhagath Parabattina, Rashmi Dutta Baruah and Pradip K. Das, "R-Peak Detection from ECG Signals Using Fractal Based Mathematical Morphological Operators", In IEEE Region 10 Conference (TENCON) 2021.
- **Deepankar Nankani** and Rashmi Dutta Baruah. "Investigating Deep Convolution Conditional GANs for Electrocardiogram Generation." In IEEE International Joint Conference on Neural Networks (IJCNN), pp. 1-8., 2020.

Chapter 4 is based on the following papers:

- **Deepankar Nankani** and Rashmi Dutta Baruah, "Improved Diagnostic Performance of Arrhythmia Classification Using Conditional GAN Augmented Heartbeats", In Springer Book entitled "Generative Adversarial Learning: Architectures and Applications" (In Press).
- **Deepankar Nankani** and Rashmi Dutta Baruah, "PIPxResNet: Penalty Induced Prototype-Based eXplainable Residual Neural Network for Heartbeat Classification", In Neural Computing and Applications, Springer Journal. (Under Review Since August 2021).

Chapter 5 is based on the following papers:

- **Deepankar Nankani** and Rashmi Dutta Baruah, "Ventricular Arrhythmia Classification and Interpretation Using Residual Neural Network with Guided Backpropagation", In IEEE Region 10 Conference (TENCON) 2021 (In Press).

- **Deepankar Nankani** and Rashmi Dutta Baruah, "Posthoc Interpretability Techniques for Explaining Ventricular Tachyarrhythmia Prediction using Residual Neural Network", In Applied Intelligence, Springer Journal. (Under Review Since November 2021).

Chapter 6 is based on the following papers:

- **Deepankar Nankani** and Rashmi Dutta Baruah. "An End-to-End framework for automatic detection of Atrial Fibrillation using Deep Residual Learning." In IEEE Region 10 Conference (TENCON), pp. 690-695. IEEE, 2019.
- **Deepankar Nankani** and Rashmi Dutta Baruah. "Classification and Interpretation of Atrial Fibrillation Using Transformer Neural Network", In IEEE International Joint Conference on Neural Networks, 2022. (Under Review)

Chapter 7 is based on the following papers:

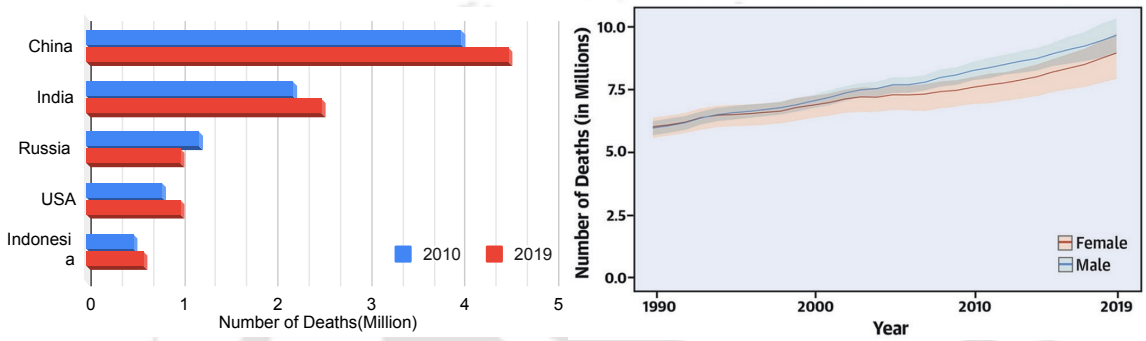
- **Deepankar Nankani**, Pallabi Saikia and Rashmi Dutta Baruah, "Automatic Concurrent Arrhythmia Classification using Residual Neural Networks", In IEEE Computing in Cardiology (CinC), 2020, pp. 1-4.
- **Deepankar Nankani** and Rashmi Dutta Baruah, "Incorporating Demographic and Heartbeat Features with Multichannel ECG for Cardiac Abnormality Detection using Parallel CNN and GAP Network", In IEEE Computing in Cardiology (CinC), 2021, pp. 1-4.
- **Deepankar Nankani** and Rashmi Dutta Baruah, "Feature Fused Multichannel ECG Classification using Channel Specific Dynamic CNN for Detecting and Interpreting Cardiac Abnormalities", In Special Issue: Classification of Multilead ECGs, Physiological Measurement Journal, Institute of Physics and Engineering in Medicine. (Under Review Since December 2021).

1

Introduction

Cardiovascular diseases (CVDs) are the most widespread disease with unpredictable and life-threatening consequences throughout the world [1,9]. The total cases of CVD nearly doubled from 271 to 523 million from 1990 to 2019, and the number of deaths steadily increased from 12.1 million in 1990, reaching 18.6 million in 2019 as illustrated in Figure 1.1. The number of subjects with CVD has been rapidly rising, particularly within the urban communities [10]. Detecting CVD through long-term non-invasive ECG signal helps in risk stratification, better medical assistance, and patient treatment [11]. The ECG signal measures the skin surface measurement of the heart's electrical activity from different angles through different leads and describes the morbid status of the cardiovascular system. An abnormality in ECG or deviation of electrical impulses from normal sequence leads to a host of cardiac abnormalities ranging from arrhythmias to coronary heart disease to electrolyte imbalance. The number of leads varies from single lead to standard twelve-lead configuration depending on the resolution of ECG required by physicians. In recent years, miniaturized ECG recorder have enabled recording of the single-lead ECG signals in urgent situations or remote locations as twelve lead devices suffer from limited accessibility. Moreover, the limited accessibility of twelve lead devices motivates the use of reduced lead devices that are cheap and easily accessible but capture limited useful information [12]. The length of the ECG signal varies from a few seconds to several days, depending on the disease the cardiologist aims to detect. The amount of cardiac data depends upon the sampling frequency, the number of leads and recording time, etc. Ideally, the sampling rate during ECG acquisition ranges from 100 Hz to 2 kHz, thereby generating a vast amount of data for further interpretation. The population growth has also led to massive amounts of ECG that require examination by

skilled physicians or cardiologists. Cardiologists manually interpret the ECG signal through visual inspection, but this process is expensive in terms of human resources and prone to error. A skilled cardiologist uses single or multilead ECG for screening and diagnosing the source of cardiac abnormalities [11]. The intervention of experienced medical personnel is an exhaustive process that needs to be automated using automatic detection algorithms [13].



(a) Country wise CVD Deaths. (b) Gender wise CVD Deaths.

Figure 1.1: Number of Cardiovascular Deaths from 1990 to 2019 [1].

The detection algorithms are embedded in a Computer-Aided Disease Diagnosis System (CADDS) that effectively detect cardiac abnormalities from the ECG signal and work as clinical decision support system to assist physicians in making accurate and informed decisions. Since quality is an essential constraint for CADDS, many automatic quality-controlled classification methods have been attempted based on well-designed evaluation metrics for guaranteeing correct predictions. However, the practical deployment of developed models is limited as the models lack explanations/justifications behind the prediction. Explanations must be designed in such a manner that they fit healthcare provider needs. Unfortunately, due to the lack of interpretability evaluation metrics focused on medical applications, most claims regarding model interpretability lack rigorous evaluations utilizing real end-users [14, 15]. CADDS encompasses different modules that perform noise removal, heartbeat detection, and disease classification. The noises present in ECG ranges from low to high frequencies that corrupt the signal and reduce the classification performance. For fine-grained ECG classification, the clean ECG signal is used to segment heartbeats to perform heartbeat classification. The beat classification is also necessary because they are a precursor or indication of life-threatening arrhythmias that may arise in the future. Therefore, beat detection is essential before classification. A good detector aims at minimizing capturing of extra beats that may take more storage

space, computing power and precisely capture irregular and regular beats without missing critical beats. Detected beats are classified into normal beats and abnormal beats. As ECG is a non-stationary signal, irregular rhythms may arise at any time. Therefore, coarse-grained ECG classification is performed in which single-channel and multi-channel ECG rhythms are examined for detecting cardiac abnormalities.

Many automated methods have been proposed in the literature for both fine and coarse-grained ECG classification. These methods exploit the short-term correlation between adjacent timestamps of beats, the considerable similarity between adjacent ECG beats (long-term correlation), and the correlation between the intra-lead ECG beats. The automated methods can be categorized into three broad categories: (i) Handcrafted feature extraction and classification with traditional machine learning (ML) approaches; (ii) Deep learning-based approaches for end-to-end ECG classification; and (iii) Hybrid approaches incorporate handcrafted features and uses raw ECG signal with deep learning models [16]. The classifier performance is highly dependent upon the selection of relevant handcrafted features [17]. The features could be temporal, morphological, statistical, frequency domain features, entropy features, etc. In recent years, deep learning-based approaches have shown promise as they automatically extract, select, and achieve good performance for ECG classification. However, the deep learning models (DLM) suffer from the following limitations.

1. The requirement of a considerable amount of labeled and realistic-looking ECG is challenging to obtain in practice. The challenge in synthesizing ECG occurs because of the complex biological and physiological systems generating the electrical impulses.
2. The developed classifiers are not generalized as they are trained on dataset acquired from a couple of geographical locations and perform poorly on unseen dataset from different geographical locations.
3. In the medical domain, the explanations behind the diagnosis are essential. The black-box nature of DLM lacks interpretability or the ability to explain model predictions, inhibiting their real-world deployment.

Several interpretability approaches have been proposed in the literature that aim at providing explanations or interpretations of the model to stakeholders. Interpretability in DLM allows failure identification and establishes trust for providing better decisions in the future. Interpretability also explains where the network focuses on the input, activated neurons, and the decision-making process. The model interpretation could be weights assigned to features responsible for a specific prediction.

Depending on the end-user, different explanations about the model may be required. Recently, there has been increased attention to the lack of user-specific explanations, despite the acknowledgment that user goals, expertise, and time constraints are central in defining explanation needs [14,15,18–23]. A good and useful explanation is left to the judgment of expert model developers, whose knowledge and backgrounds are generally not representative of end-user expertise [21,22,24]. Developers are primarily concerned with the statistical and modeling challenges of generating an explanation. The display of the explanation often receives less attention and is rarely influenced by end-user needs or insights from the literature [15, 21, 24, 25]. Moreover, it is unclear from the current evaluation studies how end-users interpret and utilize explanations designed by modeling experts, which often require some level of understanding of ML models to interpret accurately, leading to a lack of usability for end-users.

The explanation design methodology influences the perceptions mentioned earlier for CADDS that fit physician requirements. For instance, whether a physician can comprehend the information provided by CADDS affects whether physician knowledge can be integrated and alter physician perceptions about the effort required to use the system to make better and informed decisions. In addition, interpretability should not increase the time and cognitive resources as it may reduce the likelihood of system deployment. Researchers in the ML communities have proposed frameworks for providing explanations based on ML models [25–30]. The explanations are focused on the task to be accomplished and why the explanations are required [26,28,30]. These are essential elements in understanding the context of an explanation yet receive little attention about where or when users require explanations. Thus, the current interpretability frameworks fail to consider the end-user needs when utilizing a predictive model in clinical practice. Moreover, limited approaches have been proposed in the past that provides interpretations or explanation for an ECG classification system.

1.1 Motivation

The above discussion shows that an automatic interpretable ECG signal classification system requires the following components to be readily deployed in the healthcare industry: (i) Training data that encompasses diverse ECG characteristics accounting for different patterns of cardiac abnormalities to develop a generalized classifier; (ii) Classifier capable of learning correlations between input-output from the data; (iii) Classifier should provide explanations or justifications of their predictions. This research work attempts to present an interpretability perspective of deep learning

models for cardiac abnormality detection from ECG signals by considering the design philosophy of the mentioned components.

The training data is crucial in developing a generalized classifier that achieves good performance when tested for new subjects with CVD. An alternative for diverse training data is transferring pre-trained models, but prior ECG classification models are not open-sourced, limiting the transfer learning approach. The data could also be collected from multiple geographic locations or generated for broader data distribution of cardiac abnormalities.

The second component focuses on building a generalized model with sufficient trainable parameters to extract and learn the information from the training dataset and achieve good performance during the inference on unseen test data. This generalized classifier could, in practice, be deployed into CADDS.

The third and most crucial component is generating explanations or reasons for model predictions to circumvent the black box problem of DLM. However, classification performance and explainability are the trade-offs that must be considered simultaneously. A highly explainable classifier might not achieve good performance, and a black box DLM might not provide explanations of its predictions. Thus, this work is oriented towards developing a good classifier that achieves good performance and model interpretability.

1.2 Challenges

There are various challenges in developing a model based on deep learning approaches, particularly models that are interpretable in nature. The challenges are provided as follows:

1. Noisy Datasets: The low-frequency Baseline Wander, dirty or loose electrodes, might deteriorate the acquired signal quality during the data acquisition, making ECG signal classification challenging.
2. Detecting Heartbeats: Heartbeats extracted from a single-lead ECG signal indicate the occurrence of arrhythmia that might turn fatal if left untreated. But detecting the beats is challenging as ECG is a non-stationary signal, and the signal varies from patient to patient. Heartbeat detection or finding QRS complexes from ECG is necessary as its a precursor to finding heart rate, which is a crucial parameter to analyze many heart diseases, such as arrhythmias. It also serves as a reference for detecting other characteristic waves and parameter measurements.

3. Synthesising Heartbeats: Irregular beats such as supraventricular ectopic beat (SVEB) and ventricular ectopic beat (VEB) need to be detected and classified as prolonged occurrence of SVEB, and VEB may lead to atrial and ventricular fibrillation, which are life-threatening arrhythmias. The classification methods can achieve good performance, but they require huge amount of labeled, diverse, and realistic-looking heartbeats, that are difficult to acquire in practice. The challenge in synthesizing ECG occurs because the biological and physiological systems that generate these beats are highly complex. Publicly available datasets often might miss out on relevant information, thereby failing to satisfy specific criteria concerning a study, leading to synthesis of heartbeats for improving the classification performance.
4. Explaining and Classifying Heartbeats: The data generated by the increasing population makes it difficult for cardiologists to analyze heartbeats (Normal, SVEB, and VEB) manually. In addition, the human eye is poorly suited to detect the morphological and temporal variation in a patient's heartbeats, imposing the need to develop an automated diagnostic system. The traditional methods perform well on the training data but fail on test sets of new subjects with CVD as the testing is limited to one or two datasets. The recent deep learning models have achieved good performance, but their black-box nature limits real-world deployment. In the medical domain, the explanation behind diagnosis is also necessary.
5. Classification and Interpretation of Single-Channel ECG Rhythms: CVD indicators may appear in random episodes on the nonstationary and nonlinear ECG signals. The CVD episodes present in single-channel ECG might be of different lengths making the segment classification necessary due to their fatal nature. Defining the optimum episode length is crucial as small segments might miss an abnormal rhythm. Large segments might not focus on a specific part of the segment responsible for classifying the rhythm as arrhythmic. Diseases such as Atrial and Ventricular Fibrillation occur in an episodic fashion. They are related to rapid irregular contractions of the heart's muscle fibers which are generated due to untreated SVEB and VEB beats. In addition, the explanation of the classification is also required to support the model decisions.
6. Multichannel Multilabel ECG Classification and Interpretation: The cardiac abnormalities on single-channel ECG often persist in lower resolution but provide better resolution on the multi (12,6,4,3,2) channel ECG. Moreover, the limited accessibility of standard 12-Lead ECG demonstrates a need for smaller,

lower-cost, and easier-to-use devices. Subsets of the standard twelve leads can be comparable to the full set of leads in limited contexts. Still, there is limited evidence that reduced-lead ECGs can capture the wide range of diagnostic information captured by 12-lead ECGs. The cardiac abnormality interpreted by one cardiologist may differ when interpreted by another cardiologist, leading to multiple cardiac abnormalities for a single recording (Note: The abnormality might be a sub-category of the same class). The classification studies performed in the past are trained, tested, or developed in single, small, or relatively homogeneous datasets. In addition, most algorithms focus on identifying small numbers of cardiac arrhythmias that do not represent the complexity and difficulty of ECG interpretation.

1.3 Contributions

This research work mainly focuses on investigating the perspective of interpretability in deep learning models for detecting cardiac abnormalities from ECG signals and explaining the predictions to the stakeholders. The thesis contributes in the following directions. Initially, the low-frequency baseline wander is removed from the ECG signal for effective analysis, followed by detecting heartbeats (QRS complex) from the ECG signal and synthesizing irregular heartbeats to circumvent the imbalanced heartbeat distribution. In the first contribution, a penalty-induced prototype-based explainable ResNet is proposed for heartbeat classification that addresses the black-box nature of deep neural networks and provides explanations in the form of candidate heartbeats of the corresponding class. In the second contribution, gradient backpropagation-based posthoc interpretability techniques are investigated to explain ventricular tachyarrhythmia prediction using ResNet. The techniques are validated through sanity checks, thereby providing reliable diagnosis explanations to clinicians. In the third contribution, an attentive transformer neural network is proposed to classify and interpret varied length ECG segments into atrial fibrillation, normal sinus rhythm, noisy rhythms, and other rhythms. In the fourth and last contribution, multichannel multilabel ECG classification and interpretation is performed using a channel-specific dynamic CNN with demographic features, eliminating manual effort with less trainable parameters for reduced lead ECG configuration, making the model less prone to overfitting and memory-efficient in real-time processing. The contributions are summarised and represented in Figure 1.2. The detailed contributions are described below.

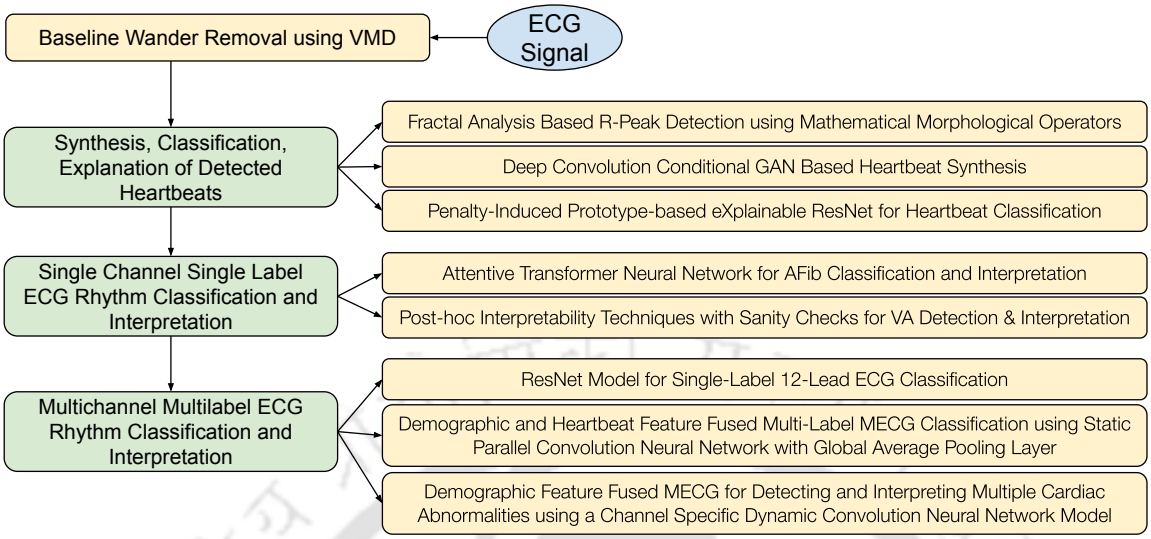


Figure 1.2: Thesis Contributions.

1. Baseline Wander (BW) Removal using Variational Mode Decomposition (VMD): The ECG signal recordings get contaminated with BW during signal acquisition, making classification challenging. BW is removed using VMD in this work, and the results are shown for normal sinus rhythm and ventricular tachycardia. Firstly, the ECG signal is decomposed into variational modes using VMD, and the noisy modes are removed, and the remaining variational modes are combined to obtain the clean signal. The performance of VMD is estimated using percentage root mean square difference, pearson Correlation, maximum absolute error, and the signal decomposition time. In addition, mean filtering, mean median filtering, empirical mode decomposition, ensemble empirical mode decomposition, and complete ensemble empirical mode decomposition with adaptive noise are compared with VMD.

2. Heartbeat Detection: The R-peaks indicate heartbeats in the ECG signal. A simple, reliable, and intuitive algorithm is proposed for real-time R-peak detection using the pattern recognition property of fractals. Mathematical morphological operators such as erosion and dilation are implemented using dynamic programming with memoization to calculate the area under the ECG signal, and peaks are extracted after resampling and thresholding the area curve. The algorithm achieved higher sensitivity and predictivity for non-arrhythmic records compared to arrhythmic records. The predictivity was high compared to sensitivity; thus, fewer extra beats were detected with more missed beats. A similar pattern was observed in the detection error rate. Fractals' fast and low compu-

tational complexity properties make an effective real-time QRS detector.

3. Heartbeat Synthesis: A Deep Convolution Conditional Generative Adversarial Network is proposed for synthesizing different beat classes, namely, supraventricular ectopic beats (SVEB), ventricular ectopic beats (VEB), and normal beats (N) as recommended by the Association for the Advancement of Medical Instrumentation [7] to circumvent the imbalanced heartbeat distribution. The generated beats are evaluated quantitatively through five metrics, namely, Frechet Distance, Dynamic Time Warping, Maximum Mean Discrepancy, Root Mean Square Error, and Time Warp Edit Distance. Qualitatively, the beats are visually analysed by plotting against the original beats.

4. Heartbeat Classification and Explanation: A Penalty Induced Prototype-based eXplainable Residual Neural Network (PIPxResNet) is proposed to classify N, SVEB, and VEB. The model addresses the black-box nature of deep neural networks. PIPxResNet encodes the temporal variations of heartbeats by employing pre-trained ResNet following the concept of task transfer learning. The algorithm extracts prototypes that are most representative of the training dataset that explain model predictions to general physicians, making them clinically relevant. The prototypes of a particular class having a close resemblance to other class prototypes are penalized, and their contribution towards the corresponding class is reduced. PIPxResNet is adopted from xDNN [31]. PIPxResNet takes into account the following conditions: (i) Beats and prototypes of class should be in close vicinity; and (ii) Other class prototypes should be farther away from corresponding class prototypes. PIPxResNet uses the beats and generates encoded and actual prototypes associated with that beat class and the actual prototypes are used for explaining the predictions. PIPxResNet is tested on four publicly available standard datasets such as Massachusetts Institute of Technology-Beth Israel Hospital Arrhythmia Database [32], MIT-BIH Supraventricular Arrhythmia Database [33], St. Petersburg INCART 12-lead Arrhythmia Database [34], and China Physiological Signal Challenge 2020 [35]. The data is collected from multiple geographical locations and is diverse enough to account for a real-world scenario.

   CVD indicators may appear in random episodes on the nonstationary and nonlinear ECG signals making the segment classification necessary. In the subsequent two contributions, diseases such as Atrial and Ventricular Tachyarrhythmia are covered that occur in an episodic fashion. They are related to rapid

irregular contractions of the heart's muscle fibers which are generated due to untreated SVEB and VEB beats.

5. Posthoc Interpretability Techniques for Explaining Ventricular Tachyarrhythmia Prediction using ResNet: Ventricular Tachyarrhythmias such as Ventricular Tachycardia and Ventricular Fibrillation along with Normal Sinus Rhythm are classified using ResNet and the predictions are interpreted using gradient backpropagation-based posthoc interpretability techniques such as Guided Backpropagation (GBP) [36], Gradient Class Activation Map (Grad CAM) [37], and Guided Grad CAM [37] that highlight the signal timestamps responsible for a particular diagnosis. The models are tested on five datasets acquired from different geographical locations, which include Massachusetts Institute of Technology [38], Malignant Ventricular Arrhythmia Database [39], Creighton University Ventricular Tachyarrhythmia Database [40], Ideo Ventricular Arrhythmia Database [41], and American Heart Association Database [34]. Quantitatively, ResNet layers and ECG signal length were varied, and the effect of Augmentation techniques, namely, Synthetic minority oversampling technique (SMOTE) [42], Borderline SMOTE using Support Vector Machine [43], and Adaptive synthetic sampling approach [44] was also analyzed for achieving optimum performance. Qualitatively, the saliency maps generated by posthoc interpretability techniques are used for interpretation, and the techniques are validated using sanity checks; namely, weight randomization and training data permutation [45]. Grad CAM performed better than other interpretability techniques and precisely highlighted relevant signal timestamps responsible for/against the diagnosis. GBP and Guided Grad CAM worked as peak detectors that usually aid in detecting heartbeats for most cases. The sanity checks validate interpretability techniques by perturbing the respective saliency maps, providing reliable diagnosis explanations to clinicians.

6. Attentive Transformer Neural Network for Atrial Fibrillation Classification and Interpretation: An end-to-end framework is developed for detecting Atrial Fibrillation from varied length ECG segments using CNN, ResNet, Attentive Convolution Neural Network, and Transformer Neural Network. The framework is verified using PhysioNet Computing in Cardiology Challenge 2017 database. The effect of ECG length is also analyzed. ResNet achieves good performance but is not interpretable. The interpretability aspect is incorporated in DLM through Attentive Convolution Neural Network and Transformer Neural Net-

work models that provide explanations through feature maps highlighting the clinically meaningful characteristic ECG waves responsible for and against the diagnosis.

7. Multichannel Multilabel ECG Classification and Interpretation: The cardiac abnormalities often provide lower resolution on single-channel ECG but provide better resolution on multi (12,6,4,3,2) channel ECG. Moreover, the cardiac abnormality interpreted by one cardiologist may differ when interpreted by another cardiologist, leading to multiple possible cardiac abnormalities for a single ECG recording. This problem is tackled in three stages. Initially, 12-lead ECG classification is performed with a single label through convolution, recurrent, and attention-based models. Convolution-based ResNet outperformed other models in terms of performance and time complexity and are preferred in later stages. Next, demographic and heartbeat features extracted from modified limb lead II are fused with static CNN to perform multilabel classification. Multiple labels improved the performance drastically, but adding heartbeat features did not improve the performance significantly. The training and testing time of feature-based models increased due to the extraction of heartbeat features. Therefore, heartbeat features were excluded in the next stage. Lastly, a demographic feature fused channel-specific dynamically built CNN is used that eliminates manual effort and provides less trainable parameters for reduced lead ECG, making the model less prone to overfitting and memory-efficient in real-time processing. The methods are verified on six publicly available standard datasets [8, 34, 46]. In addition, the model also introduces an interpretability mechanism that highlights the important leads and relevant signal timestamps responsible for cardiac pathology prediction, which builds trust in the model.

1.4 Thesis Outline

This thesis comprises of seven chapters and the organisation of the chapters in the thesis is as follows:

1. **Chapter 1** provides the introduction, motivation to pursue the problem, various underlying research challenges, contributions made to tackle the challenges, and a brief outline of the thesis.
2. **Chapter 2** provides the background about the cardiovascular system, including electrophysiology behind heartbeat genesis, history of clinical ECG, and

its acquisition. It provides an overview of an ideal ECG classification framework, including denoising, heartbeat segmentation, synthesis, cardiac abnormality classification, and interpreting diagnosis. Lastly, the chapter covers the interpretability landscape, including a brief literature review of interpretability and explainability approaches such as implicit and model agnostic posthoc approaches. Overall, the chapter presents a systematic review of the application of deep neural networks and the perspective of interpretability in these models.

3. **Chapter 3** describes the preprocessing steps applied to single-lead ECG signals. Firstly, Low-Frequency Baseline Wander is removed using Variational Mode Decomposition and compared with variants of Empirical Mode Decomposition and filtering techniques. Secondly, a fractal-based approach is described for R-peak Detection from ECG signals for segmenting heartbeats. Lastly, a deep convolution conditional generative adversarial network is discussed to synthesize minority class or irregular heartbeats.
4. **Chapter 4** describes a framework for heartbeat classification and explanation of predictions using a penalty-induced prototype-based explainable ResNet. The method classifies heartbeats into N, SVEB, and VEB and provides explanations through visual prototypes supporting the model diagnosis.
5. **Chapter 5** investigates three gradient backpropagation-based posthoc interpretability techniques for explaining Ventricular Tachyarrhythmia prediction using ResNet. The interpretability techniques include GBP, GradCAM, and Guided GradCAM, highlighting the signal timestamps responsible for Ventricular Tachyarrhythmia. The validity of these techniques is verified using sanity checks, namely, weight randomization and training data permutation.
6. **Chapter 6** proposes a framework that classifies single-lead ECG segments into atrial fibrillation, normal sinus rhythm, noisy rhythms, and other rhythms using CNN, ResNet, attentive CNN, and transformer neural network. The transformer neural network provides interpretability by highlighting the essential characteristic waves responsible for and against the clinical features responsible for the diagnosis.
7. **Chapter 7** performs multichannel multilabel ECG classification and interpretation. Different lead configurations of ECG are used to detect cardiac abnormalities. Convolution, recurrent, and attention-based models are examined. The effect of fusing demographic and heartbeat features with raw MECG is also analyzed. A channel-specific dynamically built CNN is also proposed that introduces an interpretability mechanism that highlights the essential leads and

relevant signal timestamps responsible for the corresponding prediction.

8. **Chapter 8** is the final chapter of the thesis, which highlights the conclusions with a discussion on new avenues for future research in this field.

2

Background

This chapter provides the background about the cardiovascular system, individual components of an ideal ECG classification framework, and a brief overview of interpretability techniques proposed in the literature. Section 2.1 provides an overview of the cardiovascular system, which includes electrophysiology behind heartbeat genesis, history of clinical ECG, and ECG acquisition. Section 2.2 provides an overview of an ideal ECG classification framework which includes signal denoising, heartbeat segmentation, heartbeat synthesis, cardiac abnormality classification, and diagnosis interpretation. Section 2.3 provides a literature review on interpretability, including the definition of interpretability, its relationship to explainability, approaches for achieving interpretability, and taxonomy of interpretability approaches such as implicit and posthoc approaches, and the available frameworks for generating explanations in the ECG domain. Section 2.4 summarises the chapter.

2.1 The Cardiovascular System

The cardiovascular system is composed of the heart and blood vessels which facilitate blood circulation throughout the body. Mechanical events of the cardiac cycle are initiated and synchronized by electrical events [47–49]. An Electrocardiogram (ECG) is an essential tool for assessing the working of a cardiovascular system for diagnosing several cardiac abnormalities [49, 50]. Therefore, this section presents a brief outline of the cardiovascular system, which includes details about the heart, ECG, and its acquisition.

The heart is an autonomous muscular organ about the size of a fist, located just behind and slightly left of the middle of the chest. The heart pumps blood through

the network of arteries and veins [47]. The cross-sectional view of the heart is depicted in Figure 2.1. The heart has four chambers. The upper two chambers are the atria that fill with the blood returning to the heart from the body and lungs. The lower two chambers are called the ventricles that send blood to the body and lungs. First, the de-oxygenated blood is received by the right atrium through large veins called superior vena cava and inferior vena cava [51]. Then, the blood is transferred from the right atrium (RA) to the right ventricle (RV) through the tricuspid valve. The function of the tricuspid valve is to prevent the backflow of blood from RV to RA. The right ventricle pumps the blood to the lungs for oxygenation. The left atrium collects the oxygen-rich blood. The blood is then passed to the left ventricle through the mitral valve. The left ventricle (the strongest chamber) pumps the oxygenated blood to the aorta through the aortic valve. Then, the aorta transfers the oxygenated blood to the whole body [47].

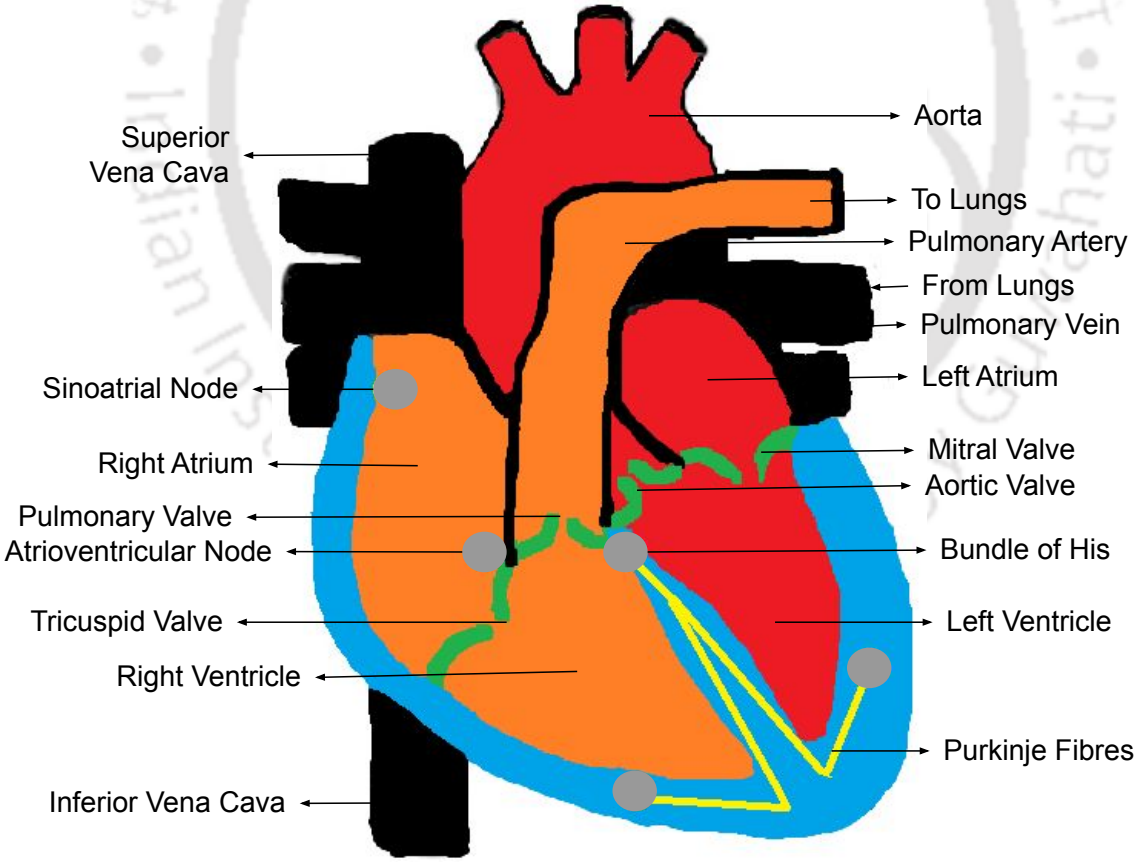


Figure 2.1: The Cross Sectional View of the Heart.

2.1.1 Electrophysiology behind Heartbeat Genesis

The entire heart activity is coordinated by an electrical conduction path or system [52]. This electrical conduction path is established using four specialized cardiac myocytes or pacemaker cells, namely, sinoatrial (SA) node, atrioventricular (AV) node, the bundle of HIS, and Purkinje fibers [51]. Firstly, the left and right atrium are depolarized due to the firing of the SA node. The firing rate of the SA node varies from 60 to 100 beats per minute. Then, the firing of the AV node transfers the heart's electrical activity from the atria to the ventricles. The AV node is also called the gatekeeper of the heart. The AV node has a firing rate varying from 40 to 60 beats per minute. The depolarization of septum and ventricles are due to the firing of HIS bundle and Purkinje fibers [52]. The intrinsic peaks of these two pacemaker cells vary from 25 to 45 beats per minute. The aforementioned electrical conduction of the heart is measured through an ECG signal that captures the electrical activity of the heart from different angles through multiple electrodes placed at different body parts. Figure 2.2 represents normal conduction of the heart. Starting from phase 1 to 6. The atria contracts causing the P-wave followed by ventricular contraction that generates QRS-complex and ending by ventricular relaxation, causing T-wave. This electrical signal follows a pattern [53]: atrial depolarization (P wave), ventricular depolarization (QRS complex wave), and ventricular repolarization (T wave). Atrial repolarization is masked by ventricular depolarization as both coincide, and therefore atrial repolarization is not evident in the ECG signal. Here, the Q wave represents the activation of the anteroseptal region of the ventricular myocardium, the R wave represents the depolarization of the ventricular myocardium, and the S wave is the activation of the posterior basal portion of the ventricles [54]. Each ECG cycle contains several time-varying characteristic waves (P, Q, R, S, T). These waves are examined through their amplitudes, intervals, and shapes. The alteration in characteristic waves are the signs and symptoms of cardiac conditions [55, 56].

2.1.2 A Brief History of Clinical Electrocardiogram

The timeline of important landmarks in the development of ECG highlighting the efforts of researchers is illustrated in Figure 2.3. In 1786, Luigi Galvani recorded electrical activity from dissected skeletal muscles at the University of Bologna, Italy [3]. In 1842, Carlo Matteucci demonstrated that electrical current accompanies every heartbeat in a frog at the University of Pisa [57]. The first ECG was recorded with a mercury capillary electrometer by Augustus Waller in 1887 at St. Mary's Hospital,

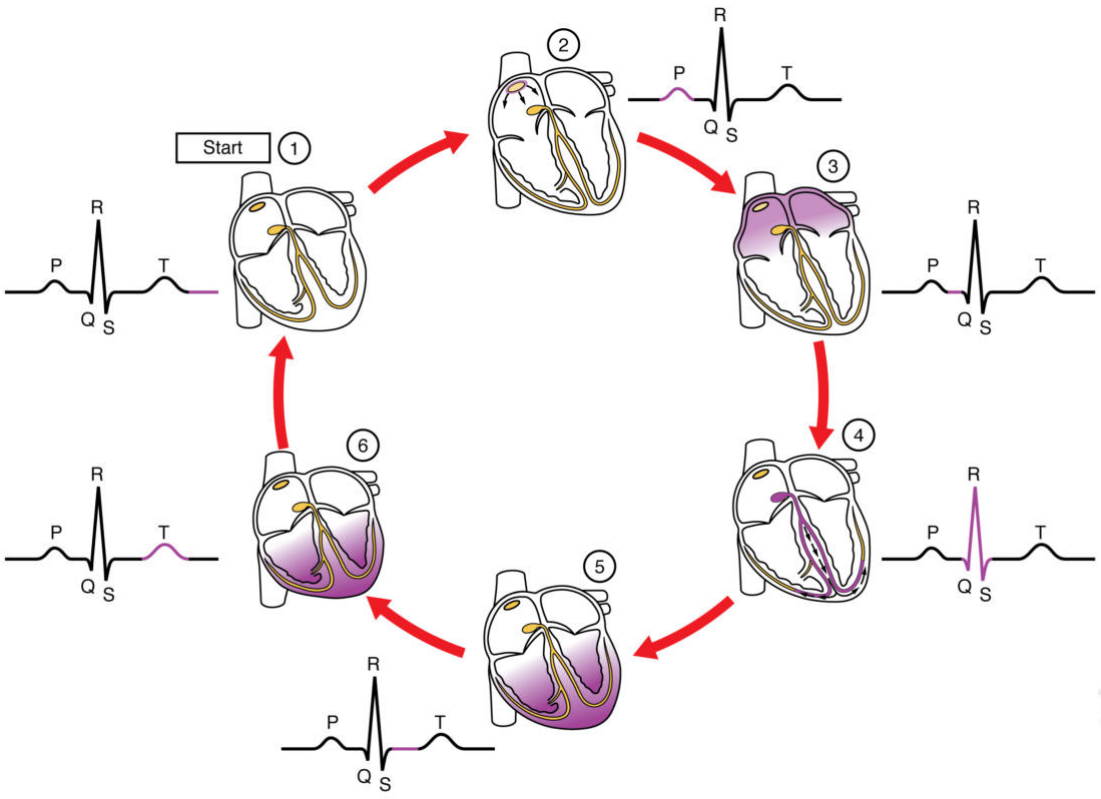


Figure 2.2: Conduction System of Heart [2].

London. He demonstrated that electrical activity preceded ventricular contraction, but the tracings were poor and exhibited only two distorted deflections [58]. In 1891, William Bayliss and Edward Starling improved triphasic cardiac electrical activity using capillary electrometer [59, 60].

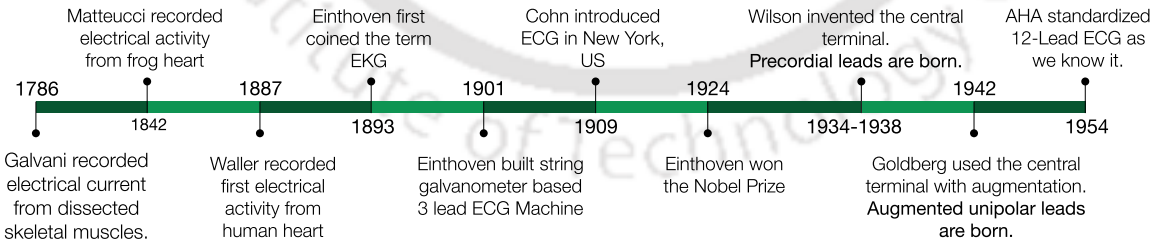


Figure 2.3: Timeline of important landmarks responsible for the ECG development.

In 1893, Willem Einthoven refined the capillary electrometer by adjusting inertia resulting in the PQRST deflections and coined the term 'electrocardiogram' at the Dutch Medical Meeting [61, 62]. In 1901, he successfully developed a new string galvanometer with very high sensitivity and recorded his electrocardiograph [60, 63] as illustrated in Figure 2.4. The positive leads were placed on the left arm and leg to

produce positive deflections on the electrocardiogram tracing as the normal electrical activation of the heart was noted to be from the upper right quadrant to the lower left quadrant [64].

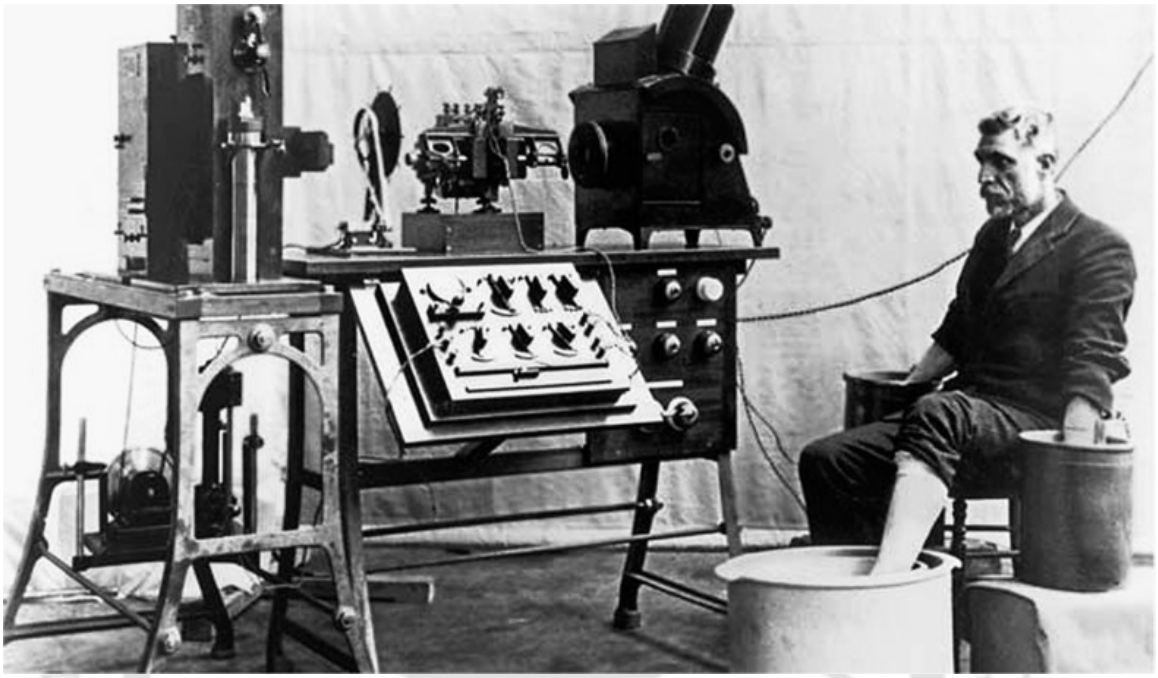


Figure 2.4: String galvanometer illustrating the machine with the patient rinsing his limbs in the cylindrical electrodes filled with electrolyte solution. [3].

Alfred Cohn introduced the ECG machine at Mt. Sinai Hospital, New York, the United States, in 1909 [60]. In 1909, Thomas Lewis observed irregular heartbeat as a result of atrial fibrillation using the ECG at University College Hospital, London [65]. In 1924, Willem Einthoven was awarded the Nobel Prize in physiology and medicine for the invention of the electrocardiograph. By 1930, the importance of ECG was recognized for differentiating between cardiac and non-cardiac chest pain. While 3-lead ECG was satisfactory, researchers realized that some regions of the heart appeared silent, leading to less detection rate of myocardial infarction [66]. In 1934, Frank N. Wilson developed the 'central terminal' concept by connecting the three limb electrodes, a central negative lead reflecting 'ground' or reference terminal. These 'unipolar' leads contrasted with the 'bipolar' leads that measure the potential difference between two sites on the body surface. In 1938, the American Heart Association (AHA) and the Cardiac Society of Great Britain published their recommendation for recording the exploring lead from six sites named V1 through V6 across the precordium. Thus, the chest leads/ Precordial leads/ frank leads were born that represented the horizontal plane of the heart.

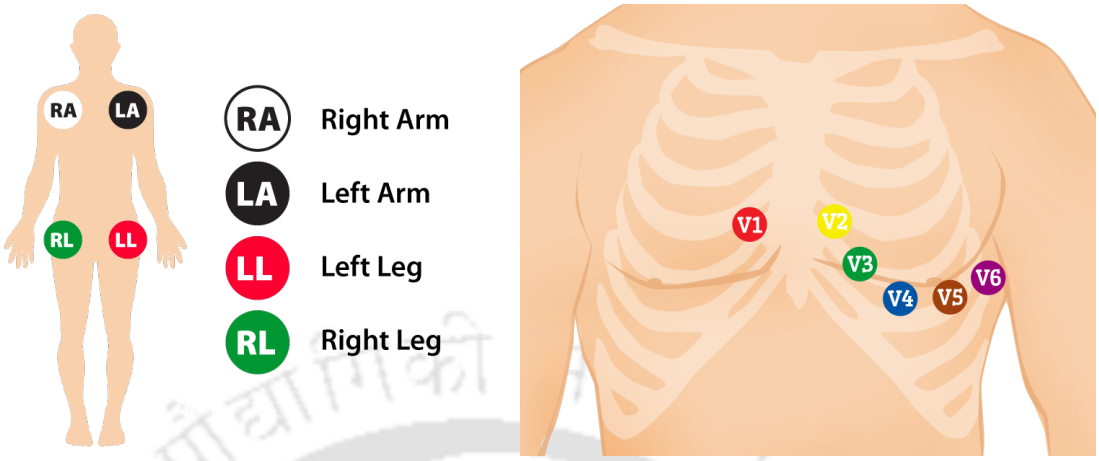
In 1942, Emanuel Goldberger used Wilson's central terminal, constructed unipolar leads with the central (zero) terminal and connected to additional positive unipolar leads on each of the left and right arms and the left leg [67] to provide more detailed coverage of the frontal plane. To increase the signal amplitude of unipolar leads, Goldberger designed a method to augment these signals resulting in the augmented unipolar limb leads a-VL, a-VR, and a-VF. The invention of the unipolar leads concluded the major advancement toward the 12-lead ECG. In 1954, the AHA published their recommendation for standardization of 12-lead ECG [68]. The 12-Lead ECG remains a standard diagnostic tool that provides a complete picture of the heart's electrical activity by recording information through 12 different perspectives.

2.1.3 Multichannel Electrocardiogram and its Acquisition

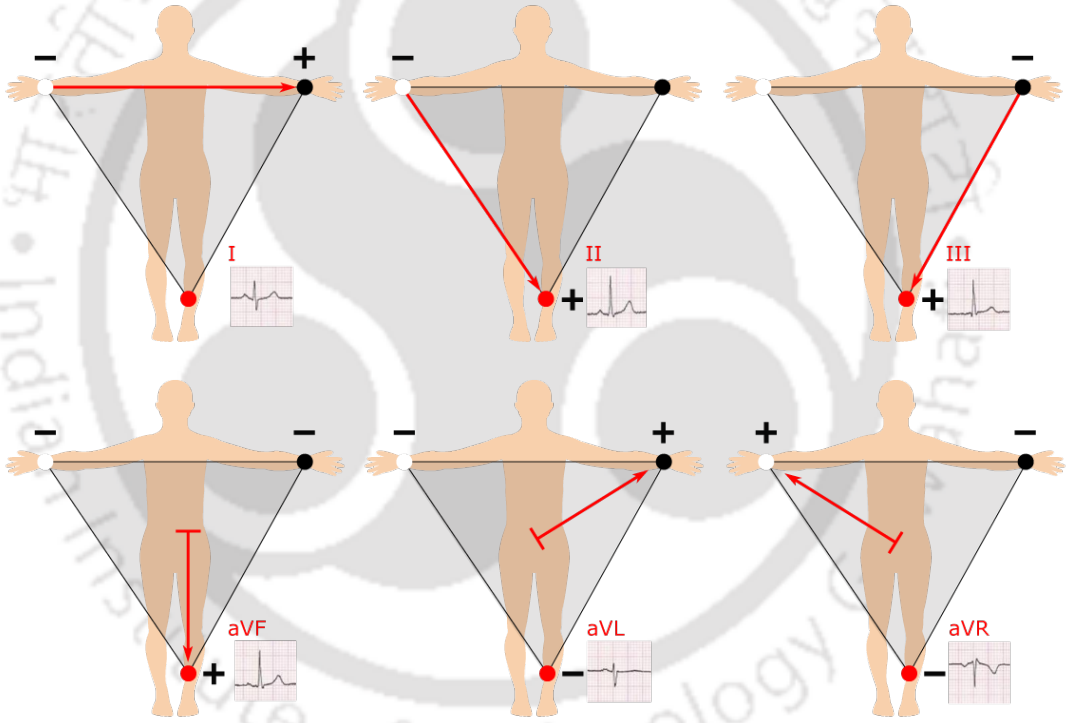
The standard 12-lead ECG is a non-invasive yet most valuable diagnostic tool that records the heart's electrical activity. A lead is a glimpse of the heart's electrical activity from a particular angle. In 12-lead ECG, ten electrodes provide 12 perspectives of the heart's activity using different angles through two electrical planes - vertical and horizontal planes. These 12 views are collected by placing electrodes through patches on the chest (precordial), wrists, and ankles. These electrodes are connected to a machine that registers the heart's electrical activity. The 12-lead ECG consists of bipolar leads (I, II, III), unipolar limb leads (aVR, aVL, aVF), and precordial leads (V1, V2, V3, V4, V5, V6).

The bipolar leads and the unipolar limb leads are derived using the electrodes that are placed on the left leg (LL), the right leg (RL), the left arm (LA), and the right arm (RA), respectively [69] to provide information about the heart's vertical plane. The placement of limb lead electrodes and their respective recording method are illustrated in Figure 2.5a and Figure 2.5c. The following lead configurations are produced: Lead I, Lead II, Lead III, Augmented Vector Right (aVR), Augmented Vector Left (aVL), Augmented Vector Foot (aVF). Leads I, II, and III require a negative and positive electrode (bipolarity) for monitoring. Lead I, II, and III are developed by Einthoven, and aVR, aVL, and aVF are developed by Goldberger [4]. The leads represent the Einthoven's Triangle formed by leads placed at RA, LA, LL.

The precordial leads are obtained from six standardized positions on the surface of the chest or thorax to provide information about the heart's Horizontal Plane. The placement of chest lead electrodes is illustrated in Figure 2.5b. It is important to correctly place V1 and V2 because the remaining chest leads are placed following



(a) Limb (Extremity) Electrodes Placement (b) Chest (Precordial) Electrodes Placement.



(c) Six frontal lead recording using just 4 limb electrodes.

Figure 2.5: Standard 12-Lead ECG Placement on Human Body [4].

these leads. Leads V1 to V6 are placed on the chest. The lead V1 and the lead V2 are positioned at the fourth intercostal space just to the right sternum and left sternum, respectively [69]. The lead V4 is placed on the fifth inter-coastal space at the midclavicular line. The lead V3 is positioned in between the lead V2 and lead V4. The lead V5 and lead V6 are placed at the anterior auxiliary and midaxillary lines. V4-V6 should line up horizontally along the fifth intercostal space [70]. The

12-lead ECG of normal sinus rhythm is illustrated in Figure 2.6. The variation in morphological features of multilead ECG such as amplitude, duration, and shape of local characteristic waves are indicators of life-threatening cardiac ailments [71]. The cardiologist examines the aforementioned components of ECG for diagnosing cardiac abnormalities [52].

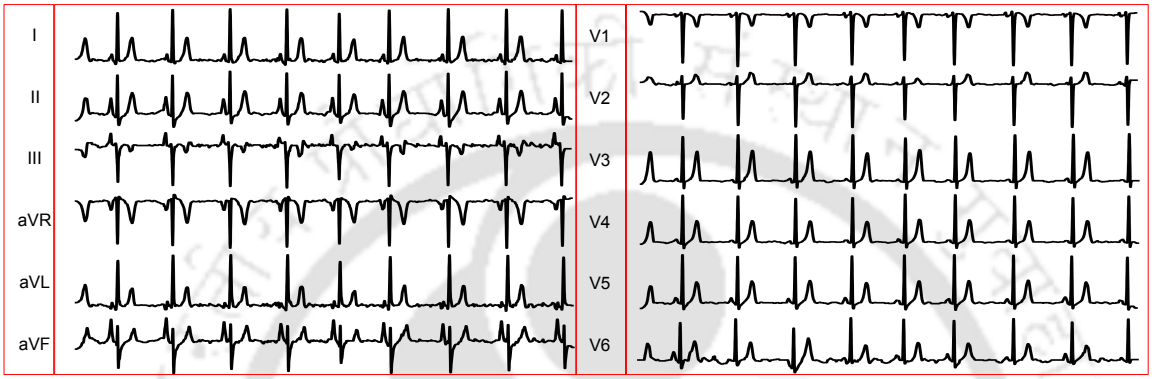


Figure 2.6: Standard 12-Lead ECG Waveform depicting Normal Sinus Rhythm.

2.2 Overview of a Typical Electrocardiogram Classification Framework

The cardiologist examines standard 12-lead ECG for detecting life-threatening cardiac abnormalities [52]. However, the large amount of data generated during continuous recording of 12-lead ECG makes it difficult for the cardiologist to manually inspect the ECG for diagnosing cardiac ailments. In addition, the human eye is poorly suited to detect the morphological and temporal variation in ECG signals. Therefore, the need for a Computer-Aided Disease Diagnostic System (CADDS) is urgent to assist the cardiologist in diagnosing the cardiac abnormalities [72]. Figure 2.7 illustrates CADDS which consists of ECG preprocessing stage, diagnostic information extraction stage, classification stage, and decision explanation stage [73]. The preprocessing stage involves ECG filtering, segmentation, and augmentation (depends on data distribution). The classification stage performs feature extraction using signal processing techniques, and the extracted diagnostic information is classified into one or more cardiac abnormalities. In the last stage, the diagnosis are explained to the physician using interpretability techniques.

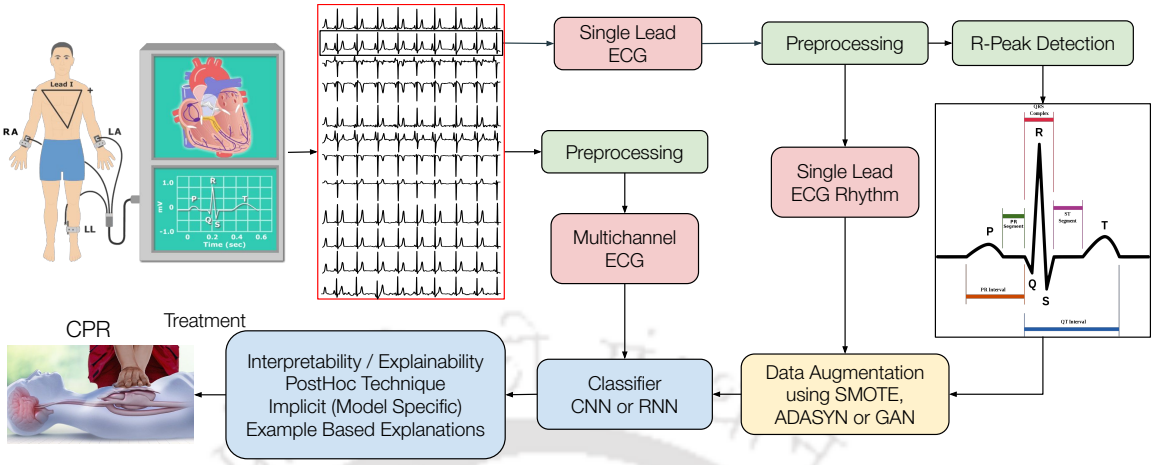


Figure 2.7: Computer Aided Disease Diagnostic System.

2.2.1 Denoising ECG Signal

Low and high-frequency noises distort the ECG signal during acquisition and make the subsequent processing challenging [74]. For effective analysis, the signal has to be denoised from embedded noises [75]. The interfering noise could be originating from different sources such as Power line interference (PLI) [76, 77], Baseline Wander (BW) [76, 78], Muscle artefacts (MA) [79], and Channel Noise (CN) [80]. Table 2.1 describe the characteristics of different noise sources present in ECG signal [75]. The noises change the amplitude of the ECG signal, degrade the PQRST morphology, and finally hinder the doctors in analyzing the signal and increase the chances of wrong diagnosis [81, 82]. The noises could be removed using various filtering techniques, wavelet transform and signal decomposition techniques. Chatterjee et al. [75] provided a review of several noise removal techniques in ECG signals. The preprocessing stage is followed by heartbeat segmentation from a single lead ECG.

Table 2.1: Characteristics of different noise sources present in ECG signal.

Type	Causes	Changes	Frequency	Effects
BW [76, 78]	Body Movement, Respiration, Bad Electrode Contact, Skin-Electrode Impedance	Depending on skin impedance electrode-electrolyte property and subject movement.	Between 0.05 and 1 Hz.	Distorts ST-segment, other LF components.
PLI [76, 77]	Inductive and capacitive couplings of 50/60 Hz during acquisition.	50% of the peak-to-peak ECG signal amplitude.	Narrowband centered at 50 60Hz with bandwidth<1Hz	Distortion in duration and amplitude.
MA [79]	Muscle activity arising from eyes, muscle, neck movement, swallowing.	10% of the peak-to-peak ECG signal amplitude.	Bandwidth ranges between 20 and 1000 Hz	Alter shapes of local waves of ECG signal
CN [80]	Transmission Channel	–	–	–

2.2.2 Heartbeat Segmentation from Single Lead ECG

Heartbeats extracted from ECG signals might indicate the occurrence of irregular or arrhythmic beats that are a precursor to determining several heart diseases such as arrhythmias, making heartbeat segmentation necessary [13]. The detected beats serve as a reference for detecting other characteristic waves and extracting heart rate features, ST segments to detect heart rhythm abnormalities or abnormal heartbeats. An ideal heartbeat constitutes a P-wave, QRS-complex, and T-wave. The QRS complex is the most striking waveform that plays a significant role in the detection of arrhythmias and irregular rhythms [83]. Widely used Holter monitors include a QRS detector that records ECG when arrhythmia appears and these recorded segments are further interpreted by cardiologists. These devices require accurate QRS detectors, which minimizes false negative and false positive rates so that no unnecessary transmission occurs, and the device does not need massive memory to store ECG segments [84]. QRS detection is difficult because of the physiological variability, and various noises present in the ECG signal [85]. QRS detection has been researched in the past three decades and a detailed discussion is provided in Section 3.2. The detected beats are provided to a classifier that performs heartbeat classification.

2.2.3 Synthesising ECG Signals

An ECG classifier achieves state-of-the-art performance using a massive amount of labeled, diverse, and realistic-looking ECGs [86]. Acquisition of high-quality normal and arrhythmic class ECGs is difficult. Often, the access to personal ECGs is restricted due to privacy concerns, hindering the development of a generalized classifier [87]. In addition, the public databases sometimes fail to satisfy a particular criterion concerning a study as they might miss out on relevant information, necessitating ECG synthesis [88]. Data imbalance could be mitigated through augmentation, improving model performance [89]. ECG synthesis is challenging because the biological and physiological systems generating these signals are complex. Synthesis could be performed using the traditional augmentation techniques such as Synthetic minority oversampling technique (SMOTE) [42], Borderline SMOTE using Support Vector Machine [43], and Adaptive synthetic (ADASYN) sampling approach [44]. Recently developed Generative models [90] could also be employed to synthesize class-specific ECG signals [87, 91]. The augmented signals mitigate the problem of data imbalance and provide sufficient data to develop a generalized classifier that could, in practice, be deployed in real-world scenarios.

2.2.4 Classification using Deep Learning Models

An ideal classifier should correctly perform disease diagnosis of a new patient without their prior information [92]. It should be able to cope with the inter-patient variability as similar cardiac abnormalities might have different waveforms for different subjects with CVD, and different cardiac diseases may have similar waveform resemblance [93]. Cardiac abnormalities such as Atrial and Ventricular Fibrillation appear rhythmically with varying episode lengths, where smaller segments might miss an abnormal rhythm, and larger segments might not localize the segment responsible for arrhythmic rhythm, making segment classification necessary [94, 95]. The rhythmic cardiac abnormalities can be prevented through early-stage heartbeat classification as irregular heartbeats are precursors to life-threatening arrhythmias. Although single lead devices are portable, they suffer from limited resolution. Therefore, multichannel ECG is used, which provides better resolution and captures a large variety of cardiac abnormalities [8, 46]. The extracted heartbeats, single-lead ECG segments, and multichannel ECG segments could be classified using traditional machine learning (ML) classifiers with handcrafted features or deep learning models (DLM). Handcrafted features include time and frequency domain features, morphological features, statistical features, entropy features that extract information from ECG [96]. These features are classified using traditional ML classifiers; namely, support vector machine, naive bayes, boosting classifier, and linear classifiers [96]. In contrast, DLM perform end-to-end ECG classification and achieve state-of-the-art performance [97–99]. DLM extracts complicated and abstract features from the raw ECG signals. Figure 2.8 summarizes different approaches for detecting cardiac abnormalities from ECG signals. The rule-based methods do not have ML modules and rely on handcrafted features. Transforming medical knowledge into hard-coded rules for computers is more challenging. Therefore, the ML methods and DLM are widely used for ECG classification.

2.2.5 Interpreting Model Diagnosis

Researchers have demonstrated commendable performance of DLM, yet they lack practical deployment due to lack of interpretability, or black-box nature, which inhibits the model from explaining its predictions [100–102]. According to General Data Protection Regulation [103], the reasons behind the model's predictions are as important as the prediction. Especially in the medical domain, the explanation behind diagnosis is also necessary to support the model decisions. The problem of the black

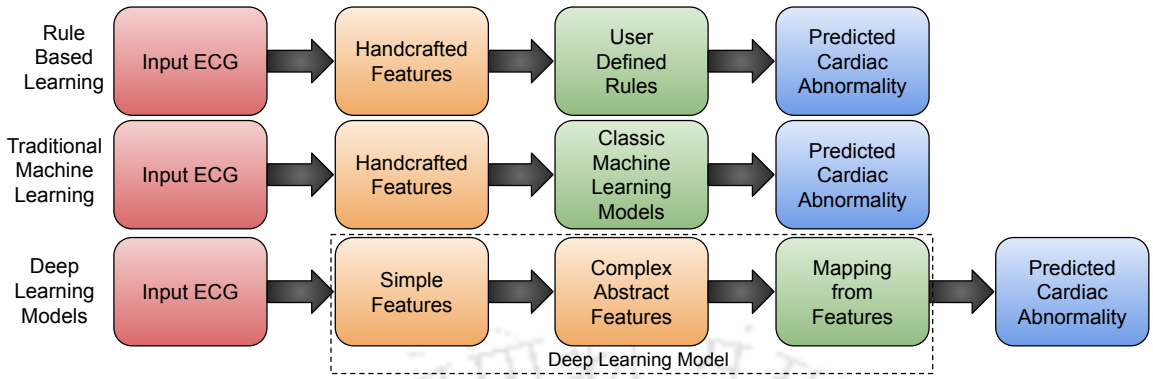


Figure 2.8: ECG Classification Approaches

box could be mitigated by providing explanations or model interpretability to establish trust in stakeholders. Several interpretability techniques have been proposed in the literature to provide insights into the model predictions, which are discussed in Section 2.3.

2.3 Interpretability Landscape

Interpretability can be defined using the following definitions: (i) Degree to which an individual can comprehend the cause of a prediction [21]; (ii) Degree to which an individual can consistently predict the result [104]. As the interpretability of a machine learning system increases, its comprehension becomes more straightforward about the predictions. Interpretability in DLM allows failure identification, establishes trust in stakeholders, and generates rules to enable better decisions in the future [105]. Although there has been significant progress in this direction [106–108], the interpretability methods developed in the past are mainly oriented towards computer vision applications rather than time-series data such as ECG, which has received relatively scarce attention. Few interpretability methods have been published for time series classification [109, 110] and ECG in particular [111]. The demand for interpretability has arisen when the goals of real-world deployment require a system to satisfy evaluation criteria that are hard to formalize or quantify [14, 29, 112]. Interpretability is a multi-faceted concept closely related to explanation. Thus, when users demand interpretability, they often seek some explanation about a model or a system to assist them in evaluating whether specific criteria are satisfied. This close relationship between the concepts of interpretability and explanation often results in the term "explainability" being used interchangeably with "interpretability" [14,112–114]. The following sections describe the differences between interpretability and explainability,

the evaluation criteria to be considered when choosing an interpretability technique, the taxonomy of interpretability, and highlight the perspective of interpretability in cardiology.

2.3.1 Defining Interpretability and Explainability

The methods proposed in the literature have used the term interpretability and explainability interchangeably and also lack a concrete definition for the term interpretability [23, 29, 112, 113, 115–117]. Interpretability generally takes the form of an explanation, but explanations are qualitative, which are widely debated within the ML community [21, 29, 118, 119]. Miller et al. [21] defined explainability as an answer to a "why-question". Interpretability is achieved when the system provides satisfactory explanations for all user questions. However, the questions and explanations depend on the user system relationship [120] and a single explanation cannot satisfy all users. The user expertise, time constraints, and goals are also important to obtain explanations that answer a question [19, 20, 29, 114, 118, 120]. The difficulty of developing adequate explanations for varied users while still providing an accurate description of the underlying ML system operations has been noticed by researchers [19, 20, 113, 118, 120, 121]. Gilpin et al. [118] suggests that explanations should be comprehensive, meaning it is human-understandable and takes into account the cognition, knowledge, and biases of end-user. Explanations should be complete, meaning that they should describe the system operations accurately. However, the trade-off between comprehensive and completeness is that an explanation with completeness may generate technical explanations, resulting in low comprehensiveness. Thus balancing this trade-off heavily relies on the user [112, 113, 122].

2.3.2 Interpretability Criteria

Auditing and debugging of DLM can be made easier if they are interpretable. The ability to interpret proves valuable for research and development and even after deployment in low-risk environments, such as item recommendations in online shopping websites such as Amazon or Flipkart or song recommendations in Gaana or Shazam. Interpretation for an erroneous prediction helps to understand the cause of the error, thereby delivering directives for fixing the system. Considering an example from [123], where the model tries to differentiate between a husky versus wolf and misclassified some huskies as wolves. Using interpretability, the cause was the snow in the image and the classifier learned snow as a feature for classifying images as a wolf in the

training dataset, but the same is not applicable in real-world use. Therefore, the following traits ensure that the model correctly explains its decisions:

1. Fairness and Bias Reduction: Ensuring that predictions are unbiased and do not implicitly or explicitly discriminate against underrepresented groups [15, 29, 112, 116, 124].
2. Privacy: Ensuring that sensitive information in the data is protected [29, 124].
3. Causality: Only causal relationships are selected [29, 125].
4. Trust: It is easier for humans to trust a system that explains its decisions compared to a black box [125]. An explainable system is trustworthy as compared to black-box model.
5. Adherence to ethical principles: Algorithm decisions or suggestions conform to ethical standards [112, 124].
6. Accountable and Liable: Assigning responsibility of a suggestion or decision to an algorithm [15, 120, 124, 126, 127].
7. Transferable, Reliable, and Robust: Algorithms exhibit a certain level of precision or accuracy when deployed in new environments such that minor changes in the input do not cause drastic changes in predictions [20, 29, 112].
8. Informative: Provide new information or insights from data to increase user knowledge base to improve decision-making [15, 112].
9. Justification: Model justifies existing domain knowledge [113, 128].

2.3.3 Taxonomy of Interpretability Methods

Interpretability techniques provide model interpretations that take the form of explanations. Hence, in this thesis, the broader class "interpretability" describes the interpretability techniques. Researchers differentiate between the interpretations based on whether they describe the model behavior [19, 112, 115, 118, 121, 126, 129–131]. Interpretability encompasses different aspects such as training algorithm, hyperparameter settings, weight representation, mathematical relationships between inputs and outputs [112, 115]. Figure 2.9 provides the taxonomy of interpretability techniques that can be categorized through the following aspects: (i) Generating interpretations; (ii) Scope; (iii) Level of Evaluation; and (iv) Nature of the technique.

Generating Interpretations: Interpretation methods can be differentiated according to the interpretations they produce. The interpretability techniques can generate interpretations in the following ways:

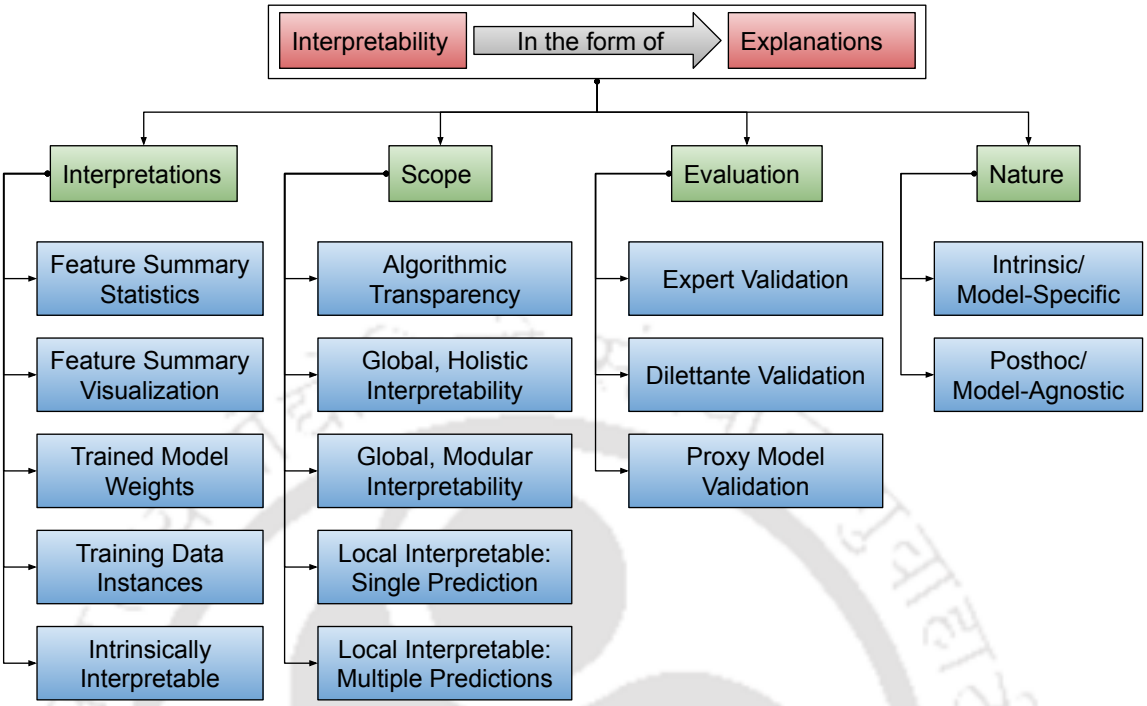


Figure 2.9: Taxonomy of interpretability techniques.

1. *Feature Summary Statistics*: A statistical summary for input features denotes individual feature importance, or pairwise feature correlation [125].
2. *Feature Summary Visualization*: The statistical summary could be visualized to provide better insight into the feature contribution responsible for the prediction. For instance, partial dependence plots [132], shapley Additive explanations [133], local interpretable model-agnostic explanations [123].
3. *Trained Model Weights*: The intrinsically interpretable models fall in this category. A few examples are linear model weights, the features and thresholds at nodes of decision trees, visualization of convolution filter maps in CNN [125].
4. *Training Data Instances*: The techniques producing data instances (either training data samples or a combination of training data samples) to make a model interpretable. For example, an explainable deep neural network that generates explainable prototypes to make the model interpretable [31, 134]. The method is suitable for computer vision and the natural language processing domain but provides less valuable insight for data with large numerical and categorical features.
5. *Intrinsically Interpretable*: Black box models can be interpreted through approximation using a piece-wise linear or an interpretable model. The interpretable model could be interpreted either through learned weights or sum-

marised feature statistics [125].

Scope of Technique: The interpretations provided by the technique could be local or global or somewhere in between local and global [19, 20, 29, 114, 120, 125, 135–137]. The global interpretations provide an understanding of the logic that the model follows and a generalized reasoning leading towards different outcomes [114]. Global interpretations provide insight on how a model is working, where the insight includes training information, functional level performance such as F1 score, and failure modes, i.e., where the model fails to predict correctly [137]. Although the interpretations have completeness, they suffer from a lack of comprehensiveness. On the other hand, local or instance level interpretations are specific model outputs that are comprehensive, but lack completeness [114,125,137]. Global interpretations describe model processes, and local interpretations describe model behavior. Hall et al. [135] suggested that best interpretations are a combination of global and local interpretations. Following are the major scopes defined in the literature:

1. *Algorithmic Transparency*: The scope highlights the transparency of the model's learning procedure using the training data and relationships learned during training. In CNN, the convolution filter visualization explains the type of detectors learned but not how individual predictions are made. Linear models are transparent, but gradient backpropagation of DLM is less transparent [112].
2. *Global, Holistic Model Interpretability*: If the overall model decision-making is comprehensive through a holistic view of input features, trained model weights, and model architecture, the model is interpretable at a Global level [23]. It helps to understand the distribution of target outcomes based on the features. However, it is difficult to achieve as a large number of model parameters make it difficult for humans to comprehend, as a larger dimension space cannot be understood by humans [112].
3. *Global Modular Interpretability*: Interpreting a large number of parameters could be performed by understanding the individual weights or a group of weights at a modular level to achieve global model interpretability. However, models might not be interpretable at a parameter level. For instance, the split and leaf nodes are helpful for decision trees, the weights are beneficial for linear models [112].
4. *Locally Interpretable for Single Prediction*: The models can be interpreted locally for the individual prediction that may depend on a few features. Local explanations could be more accurate than global explanations [112]. This the-

sis focuses on interpreting individual predictions for model-agnostic posthoc methods and model-specific implicit methods.

5. *Locally Interpretable for Multiple Predictions*: A group of predictions could also be explained with global and modular methods or instance-based explanations. Here, the group of instances is treated as if they represent the complete dataset followed by individual explanation methods for each instance followed by aggregation of explanations for the group [112].

Evaluating Interpretability Techniques: Although measuring interpretability is an active research field, a real consensus about interpretability in machine learning and evaluation metric is still lacking. However, there are a few evaluation approaches proposed in the literature [29, 125]:

1. *Expert Validation*: The technique embedded in a product performing a real task is validated by experts. For instance, an ECG classification system will generate heatmaps highlighting relevant signal timestamps for a diagnosis, and the cardiologist will evaluate the model.
2. *Dilettante Validation*: The technique is validated by dilettante/non-experts. The advantage is that no domain expert is required and could be validated by several users. For instance, users can choose the best explanation from various explanations.
3. *Proxy Model Validation*: The technique adopts a proxy function/model that is already validated by non-experts. Suppose users understand decision trees, then tree depth can be provided as a proxy for explanation where short trees get a high explainability score.

Nature of Interpretability Techniques: The techniques can be applied to a model after or during the model building. The nature of the techniques becomes intrinsic if they are part of the model architecture and posthoc if they are applied later to a pretrained model. The intrinsic and posthoc techniques are the main focus of the thesis and are discussed in detail in Section 2.3.4.

2.3.4 Implicit vs Posthoc Techniques

The nature of interpretation techniques can be categorized into implicit or posthoc interpretations [125]. The implicit or transparent or white-box interpretation provide insight to the detailed model internal operations, such as why the model may fail, and also aid during model debugging, monitoring, and model improvement [112, 115,

118, 121, 126, 129, 131]. On the other hand, the techniques focusing on model outputs and highlighting changes in inputs are known as posthoc interpretations as they provide reasons for output and not the information regarding the model's internal working [115, 121, 126, 127, 130, 131, 131]. The posthoc interpretations also provide insights into why a model may fail or generate undesired effects [129]. However, the interpretations change with the posthoc techniques, but they can be adopted for pretrained models, and retraining can be omitted. Thus, these interpretations have better comprehensiveness with less completeness [131]. Model interpretability has been classified into various categories in the literature [14, 20, 112, 114, 118, 125, 129, 135, 138] but the general consensus is that it can be categorized into implicit and posthoc approaches [14, 15, 112, 114, 117, 120, 125, 129, 138–140]. A general description and various sub-categories of implicit and posthoc techniques are provided.

Implicit Approaches: The intrinsic or implicit interpretability methods are considered interpretable due to their simple structure, which restricts the complexity of ML models [125]. The implicit approaches are generally model-specific as they are embedded in model architecture and describe model behavior at local and global levels, where global-level interpretation is more common [14, 114, 125]. Decision trees and linear models are considered intrinsically interpretable due to their simple structure. The interpretation of regression weights in a linear model is an example of model-specific interpretation. Implicit approaches can be sub-divided into transparent, hybrid, and approaches generating prototypes as explanations [14, 118].

1. *Transparent Approaches*: Family of models whose internal mechanisms are comprehensive [14, 112]. These include decision trees, Naive Bayes, logistic, and linear regression models. These models themselves serve as an explanation and are also termed as an intrinsically interpretable model, transparent model, white-box model [20, 22, 114, 120, 141]. These models generate explanations that have completeness, but comprehensiveness depends on model complexity, which leads to inferior performance, such as in the case of the linear regression model.
2. *Explainable Prototypes*: The approaches are applied in a cascaded fashion to a black-box model that provides prototypes as explanations for improving transparency of internal mechanisms of the model [118]. The trade-off between completeness and comprehensiveness varies with the approach. For instance, in [31, 134], the authors combined clustering and neural networks for classification and generation of explainable prototypes.
3. *Hybrid Approaches*: Approaches that combine transparent and black-box models, thereby sacrificing comprehensiveness but achieving better performance [14].

In [14], authors combined logistic regression and support vector machine, and the resultant model was transparent and provided explanations. The attention-based models produce heatmaps that highlight relevant signal timestamps for ECG classification [142–149]. However, these models are inapplicable to previously developed pretrained models.

Posthoc Approaches: Posthoc approaches are applied to already built trained models and are also applicable to intrinsically interpretable models [20, 112, 114, 125, 129]. The posthoc approaches are generally model-agnostic or model-independent. They can be applied to a pretrained intrinsically interpretable model to analyze the input-output relationship to describe the model's internal mechanism [14,114,125,135, 139, 140]. The approaches work at global and local levels, where local-level interpretations are more common. The generated interpretations have low completeness but high comprehensiveness and are more generalized. The approaches allow customization for varying user needs, thereby enabling comparisons and model switching in an already deployed system [140]. The posthoc methods approximate the reasoning rather than providing a cause-effect relationship for a pretrained classifier [31]. They are deprived of model weights or architectural information. Moreover, their explanations change with classifier, leading to multiple conflicting yet convincing explanations for a prediction [31]. The posthoc approaches generate visualizations through partial dependence plots, rule-extraction, feature influence methods such as sensitivity analysis [114, 125]. For instance, feature ranking is calculated for decision trees.

Posthoc techniques can be categorized into Perturbation and Backpropagation based methods. The perturbation-based methods perturb the input signal with noise and observe the changes in predicted probability. The perturbation of most contributing signal timestamps reduces the likelihood with the maximum amount [150]. However, multiple feed-forward runs are required for a given input, making the technique computationally expensive. The method is unstable to surprise artifacts as sudden perturbations to a normal class signal might result in a diseased class prediction, making the technique unsuitable. The backpropagation-based posthoc techniques are computationally cheaper and measure the contribution of input signal timestamps through backpropagation-based methods. The backpropagation-based posthoc techniques can be categorized into relevance score backpropagation-based techniques and gradient backpropagation-based techniques.

1. *Relevance score backpropagation methods*: These techniques calculate the relevance of each input signal timestamp by backpropagating the probability score

instead of the gradient. In the literature, two relevance score-based techniques are proposed, namely, Layerwise Relevance Propagation (LRP) [151] and Deep Learning Important FeaTures (DeepLIFT) [152]. LRP propagates relevance scores and redistributes as per the proportion of the activation of previous layers. Since the redistribution is based on activation scores, the technique does not suffer difficulties with non-linear activation layers. DeepLIFT explains the difference in the output prediction and prediction on a baseline reference signal instead of directly explaining the output prediction [152].

2. *Gradient backpropagation methods*: These techniques backpropagate the output gradient to the input layer to describe the influence of input timestamp on the predicted class. The methods describe the local behavior of the output at specific input timestamps. The method requires a single forward and backward pass instead of multiple forward passes for the perturbation-based methods, making it computationally inexpensive and free of artifacts. The Gradient backpropagation-based techniques include Guided Backpropagation (GBP) [36], Gradient Class Activation Map (Grad CAM) [37], and their hybrid Guided Grad CAM [37].

This thesis focuses on gradient-backpropagation-based posthoc techniques for ECG interpretation, including GBP, Grad CAM, Guided Grad CAM, and implicit approaches, including attention mechanism and explainable prototypes.

2.3.5 Model Interpretability in Cardiology

The stakes are high in cardiology, motivating the use of the DLM to assist physicians in achieving various goals [19]. DLM analyzes patient data and derives insights to aid clinical decision-making [19, 22, 153, 154]. Effective use of DLM requires medical personnel to assimilate knowledge from the model with proper reasoning and combine prior domain knowledge to make informed decisions [100,114,153,154]. The black-box nature of DLM limits the clinical deployment of models that undermines their usefulness to clinicians [155]. Lack of interpretability in DLM can reduce trust, especially in cardiology, which deals with life and death issues [156]. Despite the wide application of DLM in cardiology [157–159], cardiologists are cautious of blindly trusting the prediction of a DLM that does not provide insight into the decision-making process due to their black-box nature [160]. Medical personnel is legally and ethically responsible for decisions based on DLM, making it unlikely to adopt these models without proper justification to account for errors or data bias [19, 22, 100, 113, 124, 128, 154].

Model interpretability is needed to derive actionable insights, verify model outputs, defend care decisions, generate new knowledge that could help advance the field [19,154,161], provide a critical understanding of the model's strengths and weaknesses [162]. The newly discovered knowledge can assist experts in causal reasoning and hypothesis generation in identifying explanatory variables to study [161, 163] potentially leading to new therapies and interventions that improve prediction performance with low costs [154, 163, 164]. Since medical practitioners are already overwhelmed by vast amounts of data, explanations must be delivered in a manner that does not exacerbate this problem [19, 124]. In addition, the user should be satisfied with the provided information [19, 124]. If a CADDS predicts the cardiac abnormality without indicating appropriate reasons, the predictions may cause a delay in treatment and waste cardiologists' time for false positives [165]. Thus, much of the literature motivates the need for model interpretability in cardiology by claiming that it is integral to a DLM's usability, acceptability, and trustworthiness. [19, 22, 100, 113, 124, 161, 166]. The objective is to provide better and more effective care for a patient [19].

Researchers have published several articles supporting interpretability in DLM [167]. However, the research focuses more on creating intelligible explanations that explain the model's decision-making process through a transparent design that more closely mirrors evidence-based clinical reasoning or by identifying the important features that describe how it predicts specific predictions [168,169]. The techniques target transparency [169] to increase clinician trust in black-box models [168]. Krause et al. [170] performed a human evaluation of a custom visual explanation approach, but the users were data scientists/analysts and not healthcare providers. Lundberg et al. [171] evaluated whether explanations improved anesthesiologists' ability to predict hypoxemia risk during surgery without considering the system satisfaction achieved by anesthesiologists regarding explanations.

2.4 Summary

This chapter presents the background of the cardiovascular system, ECG classification, and interpretation. The cardiovascular system includes electrophysiology behind heartbeat genesis, history of clinical ECG, its acquisition. Then individual components of an ideal ECG classification framework are discussed, including signal denoising, heartbeat segmentation, heartbeat synthesis, cardiac abnormality classification, and interpreting model predictions. Lastly, the interpretability landscape

is explained, which defines interpretability, the difference between interpretability and explainability, the criteria for developing an interpretable model, the taxonomy of interpretability methods proposed in the literature, and the perspective of interpretability in cardiology.

3

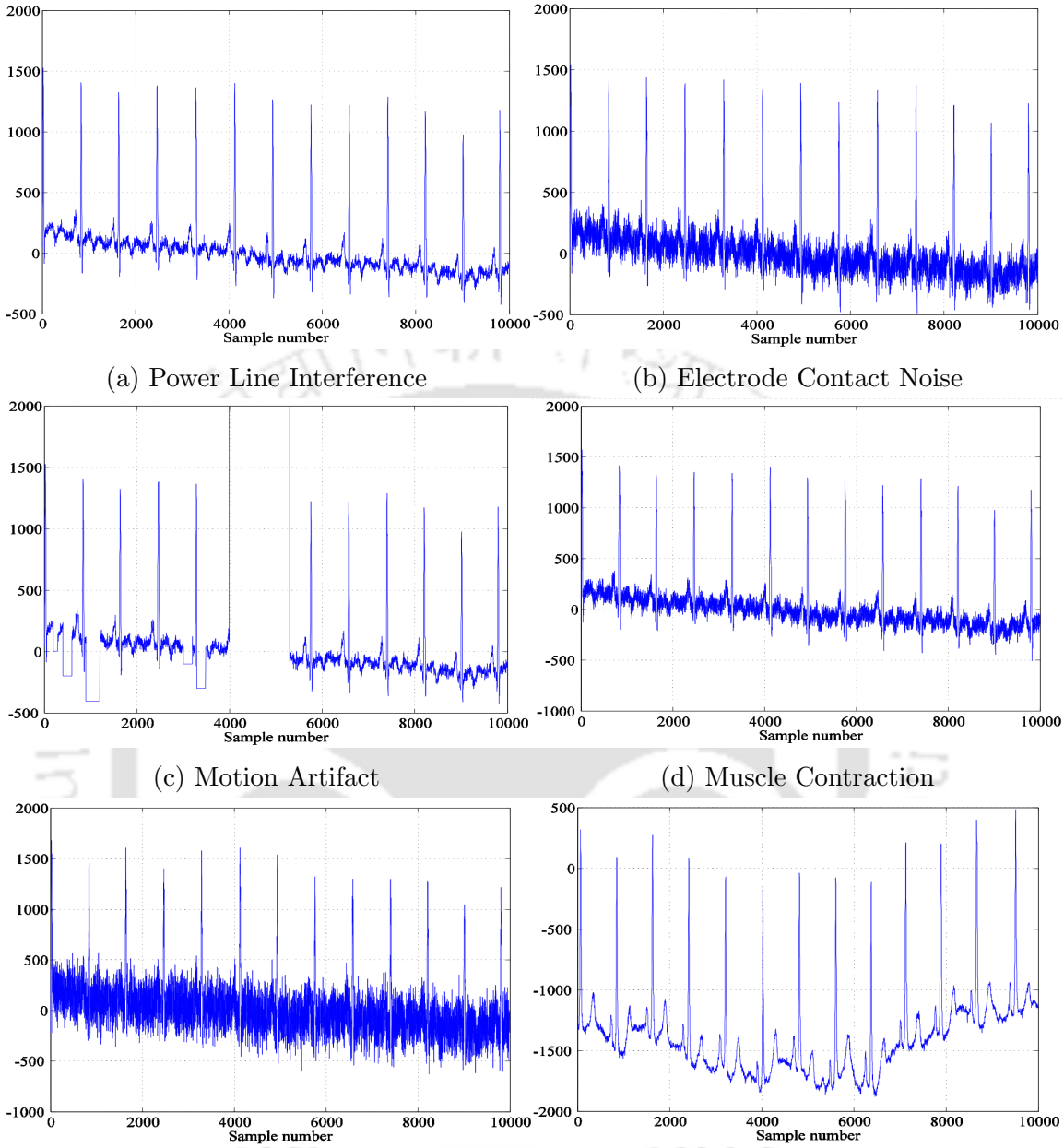
Preprocessing Electrocardiogram Signal

The preprocessing of ECG involves noise filtering, segmentation to extract heartbeats, and synthesis of heartbeats to alleviate data imbalance. ECG signal is contaminated with low and high frequency noises that deteriorate the signal quality during the data acquisition, making further processing challenging. Section 3.1 describes different types of noises corrupting ECG and provides an empirical study for baseline wander removal techniques from ECG signal. The heartbeats extracted from clean ECG signal indicate the occurrence of arrhythmia that might turn fatal if left untreated. Detecting the beats is challenging as ECG is a non-stationary signal, and the signal amplitude varies in both intra and inter patient ECG. Heartbeat detection or finding R-peaks from ECG is necessary as its a precursor to finding heart rate, which is a crucial parameter to analyze many heart diseases, such as arrhythmias. R-peaks also serves as a reference for the detection of other characteristic waves and parameter measurements. Section 3.2 describes a fractal based approach to detect R-peaks from ECG signals. The classifiers can achieve good performance using labeled, diverse, and realistic-looking heartbeats, which are difficult to acquire in practice. The challenge in synthesizing beats occurs because of the complex biological and physiological systems generating these beats. Publicly available datasets often might miss out on relevant information, thereby failing to satisfy specific criteria concerning a study. This necessitates synthesising heartbeats to improve the classification performance. Section 3.3 describes a heartbeat synthesis approach that uses generative models to synthesise class specific heartbeats.

3.1 Baseline Wander Removal to Enhance ECG Signal Quality

ECG signals are usually non-stationary, containing low and high frequency noises that distort the ECG signal. The presence of noise and artefacts corrupt the signal and make the subsequent signal processing analysis less accurate. For effective ECG analysis, it has to be decluttered from embedded noises. The noises originating from different sources are as follows:

- Power line interference (PLI) originates from power sources such as wires with interfering frequency of 50 Hz. A typical PLI corrupted signal is shown in Figure 3.1a. A number of filters have been proposed for removing PLI [172].
- Electrode contact noise (ECN) originates due to disconnected measurement system from the subject which maybe caused due to movements and vibration. An ECN corrupted signal with 60 Hz frequency [173] is shown in Figure 3.1b.
- Motion artefacts are transient baseline changes occurring due to variations in the electrode-skin impedance. It can be caused due to movement or vibrations of the patient. A typical motion artefact is shown in Figure 3.1c.
- Muscle contractions cause artefact of millivolt-level as shown in Figure 3.1d.
- Electrosurgical noise is caused by the electrosurgical instrument that is used to record ECG and operates on high voltages [174], see Figure 3.1e. During electrosurgical instrument activation, extraordinarily large transient voltages (100 - 400 V) are generated ubiquitously over the patient's skin surface with frequencies between 100 KHz and 1 MHz.
- Baseline Wander (BW) occurs due to variations in electrode-skin impedance [175]. BW can be represented as a sinusoidal component at the frequency of respiration added to the ECG signal, see Figure 3.1f. Low-frequency noise caused due to BW ranges between 0.5 ± 0.5 hertz frequency and is caused due to respiration or motion of the subject, dirty leads and improper skin contact of electrode. BW hinders the doctors in analyzing the ST segment as both of them have a similar frequency spectrum [81,82]. BW introduces a gradual increase in the amplitude of ECG signal, thereby degrading the PQRST morphology. This section focuses on removing BW for effective ECG analysis.



(a) Power Line Interference (b) Electrode Contact Noise

(c) Motion Artifact (d) Muscle Contraction

(e) Electrosurgical Noise (f) Baseline Wander

Figure 3.1: Types of noises corrupting ECG signals.

3.1.1 Brief Overview of Baseline Wander Removing Techniques

In the past, researchers have removed baseline wander from ECG signal using cubic spline curve fitting method [176], non-linear filter banks [177], median filters [178], mean median filters [179], adaptive filters [180] and combination of wavelet and adaptive filters [181]. Lifting-based discrete wavelet transform [182], statistical techniques

like independent component analysis [183] have also been used to remove artefacts from ECG. Filtered residue [184], independent component analysis [183, 185] have also been used for BW removal from ECG singals. BW removal from ECG signal has also been performed using empirical mode decomposition (EMD) and its variants [186–191]. EMD itself is unable to remove BW as it distorts the QRS complex and attenuates R-peak. So, different techniques were employed in addition to EMD including mathematical morphology [187], adaptive filter [188], and wavelet transform [189]. Ensemble EMD was also used to remove noise [190]. Complete ensemble EMD with adaptive noise and wavelet threshold [191] was also used to remove BW. In most of the aforementioned techniques, filtering, wavelet transform and EMD based methods are prevalent for BW removal. The techniques based on EMD and its variants provide comparatively better results but require a high execution time. EMD performs signal decomposition into high and low frequency components that are commonly known as Intrinsic Mode Function (IMF). High frequency component denotes the QRS complex and high frequency noise such as the interference from power sources. Low frequency components are P, T waves, ST segments and BW. Direct removal of higher order IMF ruptures the ST segment morphology. Hence, EMD is used in tandem with different techniques. The problem with wavelet transform is the requirement of the P, T wave morphology that is difficult to obtain and also the methods fail in the presence of other noises.

3.1.2 Investigated Techniques

Several signal decomposition and filtering techniques are explored for detecting baseline wander from ECG signals. The signal decomposition techniques include Empirical Mode Decomposition, Ensemble Empirical Mode Decomposition, Complete Ensemble Empirical Mode Decomposition with Adaptive noise, and Variational Mode Decomposition. The filtering techniques include median filtering and mean median filtering. A brief description of these techniques is as follows.

Empirical Mode Decomposition (EMD) [192] is a data-driven technique that decomposes a non stationary signal (generated from non linear systems) in narrowband monocomponent signals also called as intrinsic mode functions (IMF). IMFs are zero mean amplitude modulated frequency modulated (AMFM) components. However, it is not guaranteed that an IMF consists of a single oscillatory mode, and neither be narrowband signal nor meaningful due to its limitations. The IMF should satisfy the following stopping criteria: (i) Number of extrema (local maxima and local

minima) and zero-crossings must differ at most by one; (ii) At any timestamp, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is near zero [192]. Algorithm 1 calculates EMD of signal $y(t)$.

Algorithm 1 Empirical Mode Decomposition

Input: ECG signal ($y(t)$)
Output: IMF_{EMD} ($z(t)$)

1. Determine all local maxima $y_{max}(t)$ and local minima $y_{min}(t)$ for $y(t)$.
2. Interpolate $y_{max}(t)$ and $y_{min}(t)$ using cubic spline.
3. Calculate mean $m(t)$: $m(t) = (y_{max}(t) + y_{min}(t))/2$.
4. Calculate $d(t)$: $d(t) = y(t) - m(t)$.
5. Check if $d(t)$ is an IMF using the stopping criteria, if it satisfies the criteria then goto step 6 or else goto step 1.
6. The above procedure is called sifting. After obtaining the first IMF, subtract it from $y(t)$ and obtain the remaining signal. Perform sifting on the obtained signal until the residue persists any meaningful frequency information.
7. The final decomposed signal can be obtained as a sum of IMF's $d_n(t)$ and a residue $r_n(t)$ as provided in Equation 3.1.

$$z(t) = \sum_{n=0}^{N} d_n(t) + r_n(t) \tag{3.1}$$

The IMFs obtained using EMD suffers from oscillation with multiple frequencies in a single mode or single frequency in multiple modes. This problem is commonly known as "mode mixing". Adding multiple realizations of a specific amount of noise removes mode mixing by utilizing the dyadic filter bank behaviour of EMD [193]. This phenomenon was termed as ensemble EMD (EEMD) [194].

EEMD [194] decomposes the original signal for multiple ensembles of noise and produces the modes by averaging. Algorithm 2 calculates EEMD of signal $y(t)$.

Algorithm 2 Ensemble Empirical Mode Decomposition

Input: ECG signal ($y(t)$)
Output: IMF_{EEMD} (D_k)

1. Generate a new input by adding multiple noise realizations of $N(\mu = 0, \sigma = 1)$.
2. Decompose each new input using EMD and obtain the IMF d_k^n.
3. Assign d_k as the k^{th} IMF obtained from $y(t)$ by averaging the corresponding IMFs as given in Equation 3.2.

$$D_k = \frac{1}{I} \sum_{i=1}^{I} {d_k}^i \tag{3.2}$$

Each pair of signal and noise is individually decomposed and their residue ${r_k}^i =$

${r_k}^{i-1} - {d_k}^i$ is obtained thereby eliminating the estimation of local means.

EEMD alleviates mode mixing problem but introduces the problem of residual noise that corresponds to the difference between reconstructed and original signal. Another problem is that the averaging of IMFs is difficult due to the fact that varying number of IMFs are generated by EEMD. This led to the development of Complete Ensemble EMD using Adaptive Noise (CEEMDAN) [195].

CEEMDAN [195] not only achieves negligible reconstruction error but also solved the problem of varying number of modes for different noise realizations. The basic intuition of CEEMDAN comes from the fact that it utilises all final modes generated by multiple noise realization of signal for the calculation of the next mode. This estimates the local mean of modes in an efficient and sequential manner for each noise realization. Suppose $E_k(.)$ generates k^{th} IMF via EMD. Then CEEMDAN of signal $y(t)$ is calculated using Algorithm 3.

Algorithm 3 Complete Ensemble EMD with Adaptive Noise

Input: ECG signal $(y(t))$

Output: $IMF_{CEEMDAN}$ (x)

1. For every $j = \{1, \ldots, J\}$, decompose each $y^{(j)} = y + \beta_0 w^{(j)}$ using EMD until the first CEEMDAN mode is obtained. Then compute $\overline{d_1} = \frac{1}{J}\sum_{j=1}^{J} d_1^{(j)}$.
2. Calculate first residue using $r_1 = y - \overline{d_1}$.
3. Generate first mode of $r_1 + \beta_1 E_1(w^{(j)})$ by EMD, where $j = \{1, \ldots, J\}$ and calculate second CEEMDAN mode as $\overline{d_2} = \frac{1}{J}\sum_{j=1}^{J} E_1(r_1 + \beta_1 E_1(w^{(j)}))$.
4. For $k = \{1, \ldots, K\}$ calculate the k^{th} residue as $r_k = r_{(k-1)} - \overline{d_k}$.
5. Calculate first mode of $r_k + \beta_k E_k(w^{(j)})$ by EMD, where $j = \{1, \ldots, J\}$ and calculate the $(k+1)^{th}$ CEEMDAN mode as $\overline{d_{(k+1)}} = \frac{1}{J}\sum_{j=1}^{J} E_1(r_k + \beta_k E_k(w^{(j)}))$.
6. Goto step 4 for the calculation of next mode k.
7. Iterate steps 4 to 6 until the residue satisfies IMF conditions or it has less than 3 local extremum points. The last residue satisfies: $r_K = y - \sum_{k=1}^{K} \overline{d_k}$, where K is the number of IMFs. Therefore, the overall signal can be represented by Equation 3.3.

$$x = \sum_{k=1}^{K} \overline{d_k} + r_K \tag{3.3}$$

Modes extracted using CEEMDAN provide exact reconstruction of the original signal. Final number of IMFs is solely determined by the data and the stopping criterion. However, CEEMDAN also suffers from residual noise as the signal informa-

tion appears in higher order IMF as compared to EEMD and some "spurious" lower order modes [195]. Theoretical and mathematical literature still lacks in finding out the number of ensembles and the amplitude of noise to be added in order to boost performance.

Variational Mode Decomposition (VMD) [196] is a data adaptive technique that generates the variational modes from multicomponent signal $y(t)$ in an entirely non recursive and concurrent fashion. The variational modes (u_k) are quasi orthogonal and bandlimited around center frequency (ω_k) that are capable to reproduce the input signal. VMD comprises of a strong mathematical framework. It uses the concepts of Wiener filtering, Fourier transform, Hilbert transform, analytic signal and the frequency shifting through harmonic mixing. Algorithm 4 describes signal decomposition through VMD.

Algorithm 4 Variational Mode Decomposition

Input: ECG signal $(y(t))$

Output: IMF_{VMD} $(y_k(t))$

1. For each mode, the analytical signal is computed using the hilbert transform to acquire a unilateral frequency spectrum.
2. The spectrum of the obtained mode is mixed with an exponential that shifts it to an estimated center frequency.
3. The bandwidth of the mode is estimated through the squared norm of the demodulated signal.
4. Perform above steps until convergence and calculate $y_k(t)$ using Equation 3.4.

$$\min_{\{y_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * y_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} s.t. \sum_{k=1}^{K} y_k(t) = y(t) \tag{3.4}$$

where, δ is the dirac distribution, t is the time, K is the number of modes and $*$ is the convolution operator.

Mean-Median Filtering (MMF) [197] utilizes the convex combination of the sample median and sample mean of signal $y(t)$ as provided in Equation 3.5.

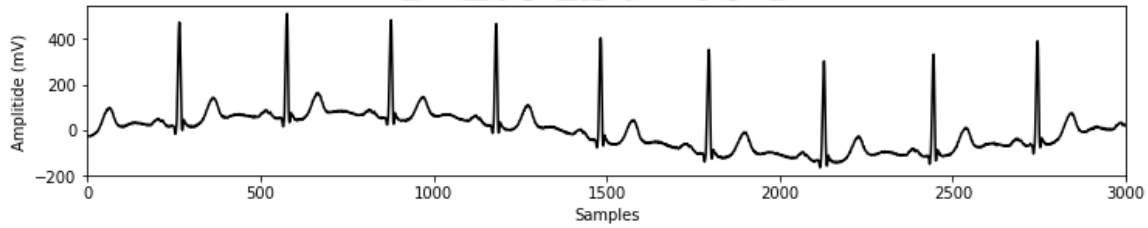
$$MMF = (1-\alpha) * mean(y(t)) + \alpha * median(y(t)) \tag{3.5}$$

where, $\alpha \in [0, 1]$ is the 'contamination factor'.

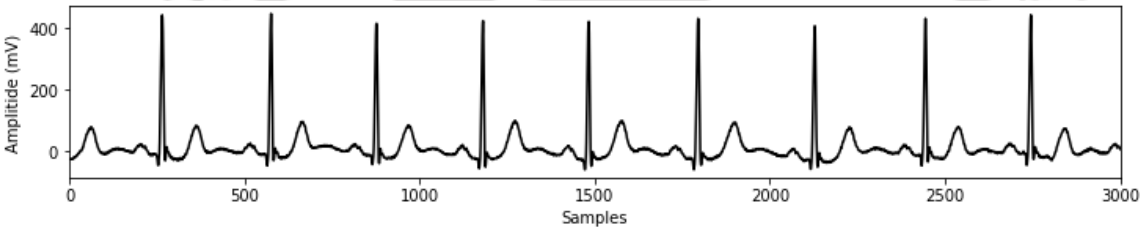
3.1.3 Dataset Description

Massachusetts Institute of Technology-Beth Israel Hospital (MIT-BIH) database [32, 38] is used for experimental purposes. The Database is freely available at Phys-

ioNet [34] since 1999. It consists of 23 normal and 25 arrhythmic dual-channel ECG recordings with sampling frequency of 360 Hz. The two channels are Modified limb lead II (MLII), MLII lead signal is used for experimental purposes. Figure 3.2a represents a normal sinus rhythm from record 103 that is contaminated with BW and Figure 3.2b represents the clean normal sinus rhythm (NSR). Similarly, Figure 3.3a represents a segment of ventricular tachycardia (VT) from record 205 that is contaminated with BW and Figure 3.3b represents the clean segment of VT.

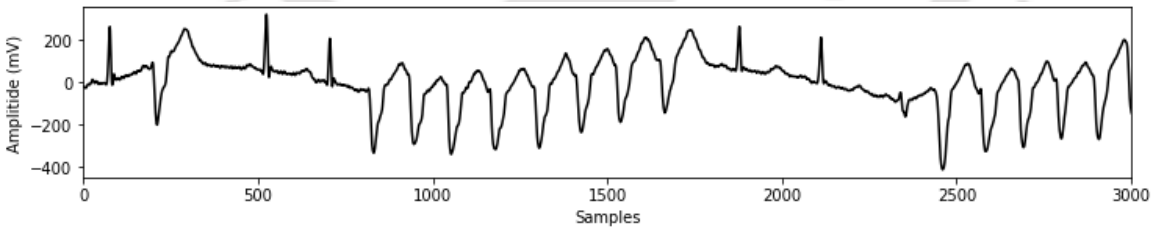


(a) NSR contaminated with Baseline Wander

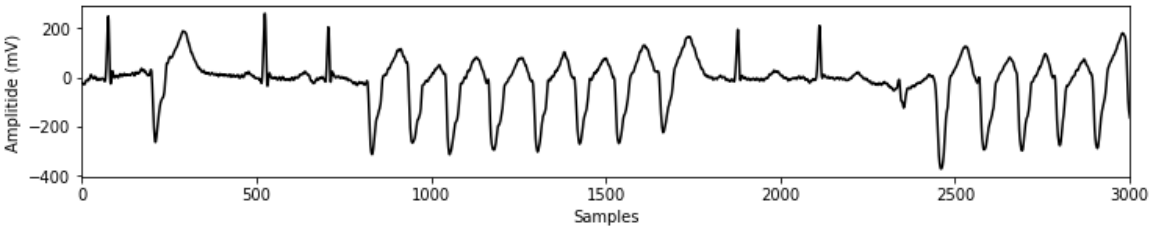


(b) Clean Normal Sinus Rhythm.

Figure 3.2: Normal Sinus Rhythm from record 103 of MIT-BIH Dataset.



(a) VT contaminated with Baseline Wander



(b) Clean Ventricular Tachycardia Segment

Figure 3.3: Ventricular Tachycardia Segment from record 205 of MIT-BIH Dataset.

3.1.4 Methodology

A detailed analysis is presented for estimation of BW using Variational Mode Decomposition. Different parameters of VMD are varied, namely, the bandwidth constraint and number of modes for better decomposition of the input signal into a clean signal and BW. A comparison between different variants of EMD, filtering techniques, namely median, mean median filtering, and Variational Mode Decomposition, is also performed for an effective estimation of BW. For testing the robustness of the techniques, artificially generated noise of frequency around 0.4 ± 0.4 Hz was added to the original signal. These frequencies correspond to the BW frequency. The noisy signal was provided to different BW estimation techniques. The techniques include median filter [177], MMF [197], Blanco EMD [186], combination of MMF with EMD [198], EEMD with fixed cut off frequency [194], and VMD [196]. The estimated baseline is subtracted from noisy signal producing the denoised signal as illustrated in Figure 3.4.

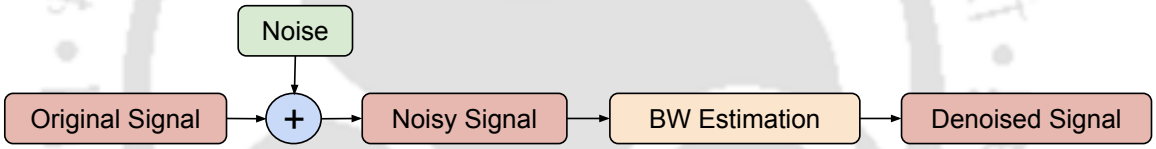


Figure 3.4: Baseline wander estimation workflow.

3.1.5 Results and Discussion

For comparing the technique efficacy, three evaluation metrics have been employed: Percentage root mean square difference (PRMSD), Maximum Absolute Error (MAE), and Pearson Correlation (PC) as provided in Equation 3.6, 3.7, and 3.8, respectively. In addition to the metrics, time taken by each technique for BW removal was also measured. Here, $x[n]$ represents the signal contaminated with baseline wander, $\tilde{x}[n]$ represents the clean signal and N represents number of samples in the signal. $x[n]$ and $\tilde{x}[n]$ are similar length signal.

The NSR segment is decomposed into variational modes/components using VMD and then original signal is reconstructed from variational modes after removal of noisy component. The difference between the original and reconstructed signal is illustrated with the help of Figure 3.5. The number of modes/components varied from 2 to 15 and center frequencies varied from 1000 to 60000. The PRMSD and MAE are maximum when number of modes is less, and bandwidth constraint is very high. As variational

modes increases, the bandwidth constraint should also be increased in order to obtain less error while reconstructing the original signal.

$$\text{PRMSD} = \sqrt{\frac{\sum_{n=1}^{N}\left[x(n)-\tilde{x}(n)\right]^2}{\sum_{n=1}^{N}\left[x(n)\right]^2}} \times 100\% \tag{3.6}$$

$$\text{MAE} = \max_{n=1}^{N}\{|x(n)-\tilde{x}(n)|\} \tag{3.7}$$

$$\text{PC} = \frac{N\sum_{n=1}^{N}x(n)\tilde{x}(n) - \left(\sum_{n=1}^{N}x(n)\sum_{n=1}^{N}\tilde{x}(n)\right)}{\sqrt{\left[N\sum_{n=1}^{N}x(n)^2 - \left(\sum_{n=1}^{N}x(n)\right)^2\right]\left[N\sum_{n=1}^{N}\tilde{x}(n)^2 - \left(\sum_{n=1}^{N}\tilde{x}(n)\right)^2\right]}} \tag{3.8}$$

As specified by [196], both over-binning and under-binning have advantages and disadvantages. During under-binning (less number of variational modes), mode sharing occurs between the neighbouring frequency for small center pulsation and high-frequency variational modes are discarded, as these modes are considered as noise for large pulsation. During over-binning (higher number of variational modes), larger values of pulsation allows a low-frequency band in the decomposed modes providing very compact band in frequency spectrum but with increased execution time for mode extraction. After the signal decomposition using VMD, the baseline wander was mostly present in the 1^{st} component. A similar pattern can be observed for correlation where the PC increases as K and ω increase together. In the case of low K and high ω, the correlation becomes insignificant. The memory consumption also increases by 50 mega bytes for each additional variational mode. The time for mode extraction via VMD increases exponentially with each new mode. Hence, for higher number of modes, the execution time limits the real world use. Therefore, it can be inferred that there exists a relation between the variational modes and bandwidth constraint such that if either of them increases then the other has to increase in order to produce consistent modes with least reconstruction error in least square sense. It is also clear that larger values of variational modes and bandwidth constraint produce modes with compact frequency spectrum when compared to smaller values, but the execution time and RAM requirement also increases. Moreover, the variational modes extracted by VMD for the corresponding signal precisely captures their center frequencies. The trend and mid frequency bands of the obtained modes

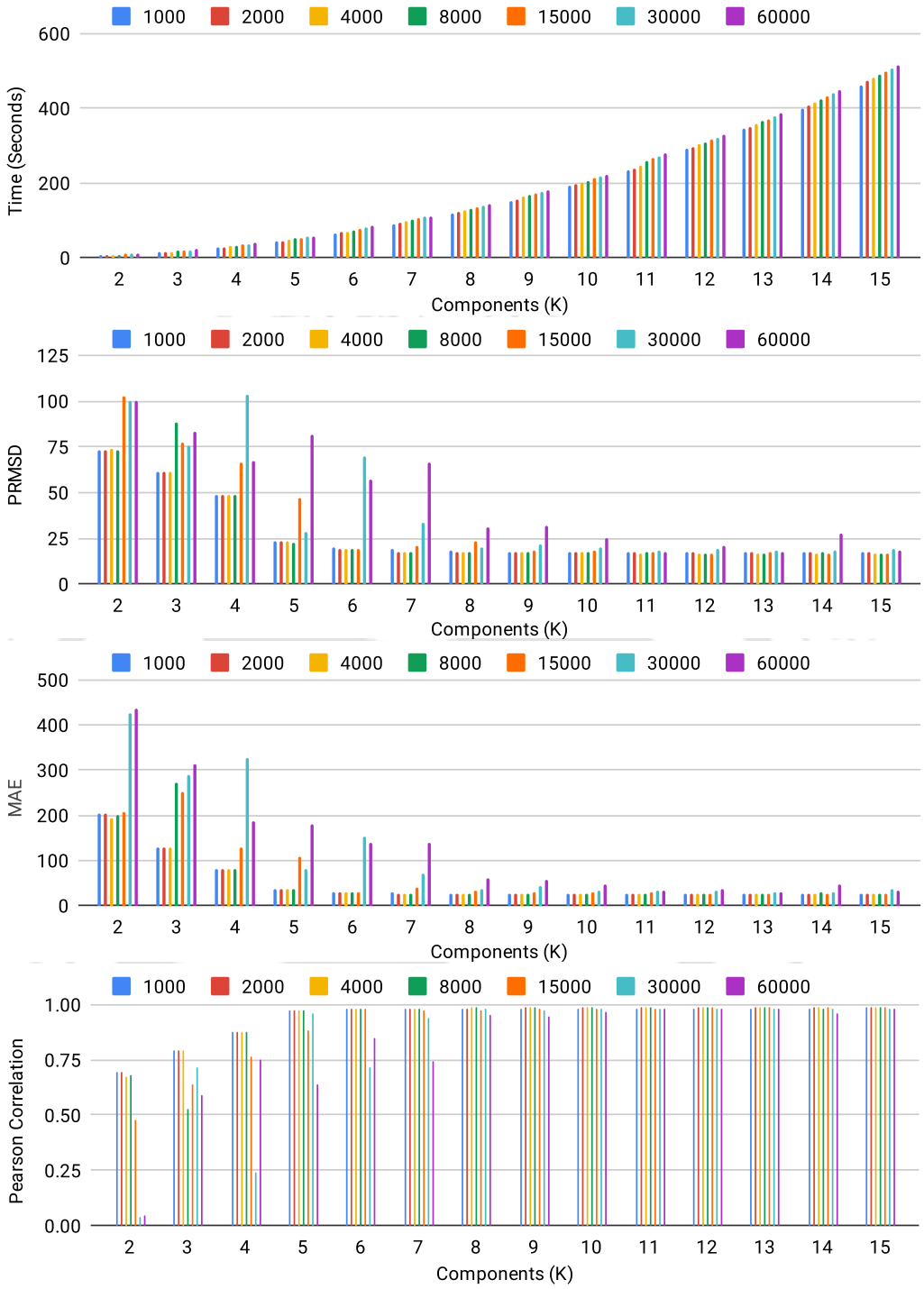


Figure 3.5: Application of VMD on NSR where the variational modes vary from 2 to 15 and center frequencies vary from 1000 to 60000.

consists of less spurious oscillations when compared to EMD. In addition to the above characteristics, no additional spectral and temporal feature estimates are required for discriminating the BW components from the ECG. As a precise value for number of

variational modes and bandwidth constraint was difficult to determine, $\omega = 8000$ and $K = 8$ was chosen for achieving least reconstruction error.

Comparison with Other Techniques: Comparison with median filter, mean median filter and EMD along with its other variants for BW removal in NSR and VT is also performed. Median filters [177] were employed in a cascading fashion where the output of first filter was provided as input to second filter and a step like waveform is obtained as the resultant BW. The window length for the filters was kept at 251 and 601 for first and second filter, respectively. Thus providing a high value of correlation between obtained and BW present in the signal. Mean median filter (MMF) [197] were also applied in a similar fashion as the median filters with similar window length with $\omega = 0.6$. The mean median filters produce a very smooth baseline because of the presence of mean filter. The mean filter overestimates BW because of the presence of QRS complex and the median filter produces trimmed mean that in turn leads to severe wave distortion. Hence, MMF not only preserves the outline of BW but also avoids step like waveform as generated by the traditional median filter. However, the drawback is that the discontinuity is still present in the obtained baseline at the signal endpoints. Blanco's EMD. [186] method was chosen where EMD is employed for signal decomposition to obtain IMFs with multiband filtering for BW estimation. The EMD algorithm produces high frequencies in lower order IMFs and low frequencies in higher order IMFs. So, the BW is present in higher order IMFs (except the residual mode due to less number of extrema). However, it is worth mentioning that generated baseline has phase difference compared to original baseline. Hence, if the two baselines are aligned together, they produce a very high correlation. BW obtained through MMF resulted in discontinuities at the starting and ending point of the baseline. Hence, the fourth experiment combines MMF and EMD [198], where EMD smoothens the baseline obtained from MMF. Two mean median filters with window length of 250 and 600 were used that produced the BW. The obtained BW was decomposed using EMD and noisy IMFs were removed using statistical methods.

According to the results, BW was present up to the last 6 IMF with $L = 0.05$. These values were obtained in contrast to the PRMSD and Pearson correlation which turn out to be around 0.85 and 61.37. It can be observed from the Figure 3.6 that due to the shifted baseline, the performance metrics deteriorated. We performed two more variations to the [186] approach by employing EEMD and CEEMDAN in place of EMD that helped in better estimation of baseline wander. However, the time required by CEEMDAN was very high making it unreasonable for real-time

applications. Hence, the results of CEEMDAN are not included in this work. The performance for all the techniques for NSR are provided in Figure 3.6. The best PC was obtained for VMD at 0.98 followed by median filter, and EEMD with fix cut off frequency. Median Filter correlation constantly reduced from 0.97 to 0.83 as the artificially induced noise was increased. Except for Blanco EMD method, other techniques did not produce much change in MAE when the noise was increased. VMD produced least MAE among all at 27%. Median filter, EEMD_Fixcut, and Blanco EMD produced MAE in an increasing fashion as the noise was increased. For PRMSD, median filter and Blanco EMD produced an increase in error as the noise increased. VMD again provided the least error irrespective of the noise. The time taken by decomposition techniques namely EMD, EEMD and VMD were higher than other techniques. Median filter, MMF and MMF-EMD took the least time at around 0.1, 0.6, and 3 seconds, respectively. VMD took around 5 seconds.

Results on VT for all techniques for all evaluation metrics are provided in Figure 3.7. The best PC was obtained for VMD at 0.97 followed by EEMD_Fixcut, and median filter. The PC values for MMF and MMF with EMD were better than the ones obtained for NSR. Blanco EMD method performed similar to MMF for high noise frequencies. MAE values kept varying for all the techniques at different noise frequencies. However, VMD provided less error at most frequencies and MMF, MMF-EMD and Blanco EMD method provided highest error. VMD, EEMD fix cut achieved low PRMSD ranging between 20% to 25%. PRMSD kept increasing for MMF, MMF-EMD, and Blanco EMD method producing the highest PRMSD values. Median filter, MMF and MMF-EMD took the least time, whereas decomposition took relatively higher execution time.

The higher the complexity of the present baseline, the more execution time the algorithm took to decompose the signal. Hence, as the noise increased the time to decompose also increased. Results for CEEMDAN are not included as its execution time exceeds by a huge margin as compared to other approaches. The comparison depicted that VMD estimates better baseline as compared to other techniques in terms of PC, PRMSD, and MAE. However, the time required to decompose the signal is relatively higher than the filtering techniques. The preprocessing stage is followed by segmentation of heartbeats from single lead ECG signal.

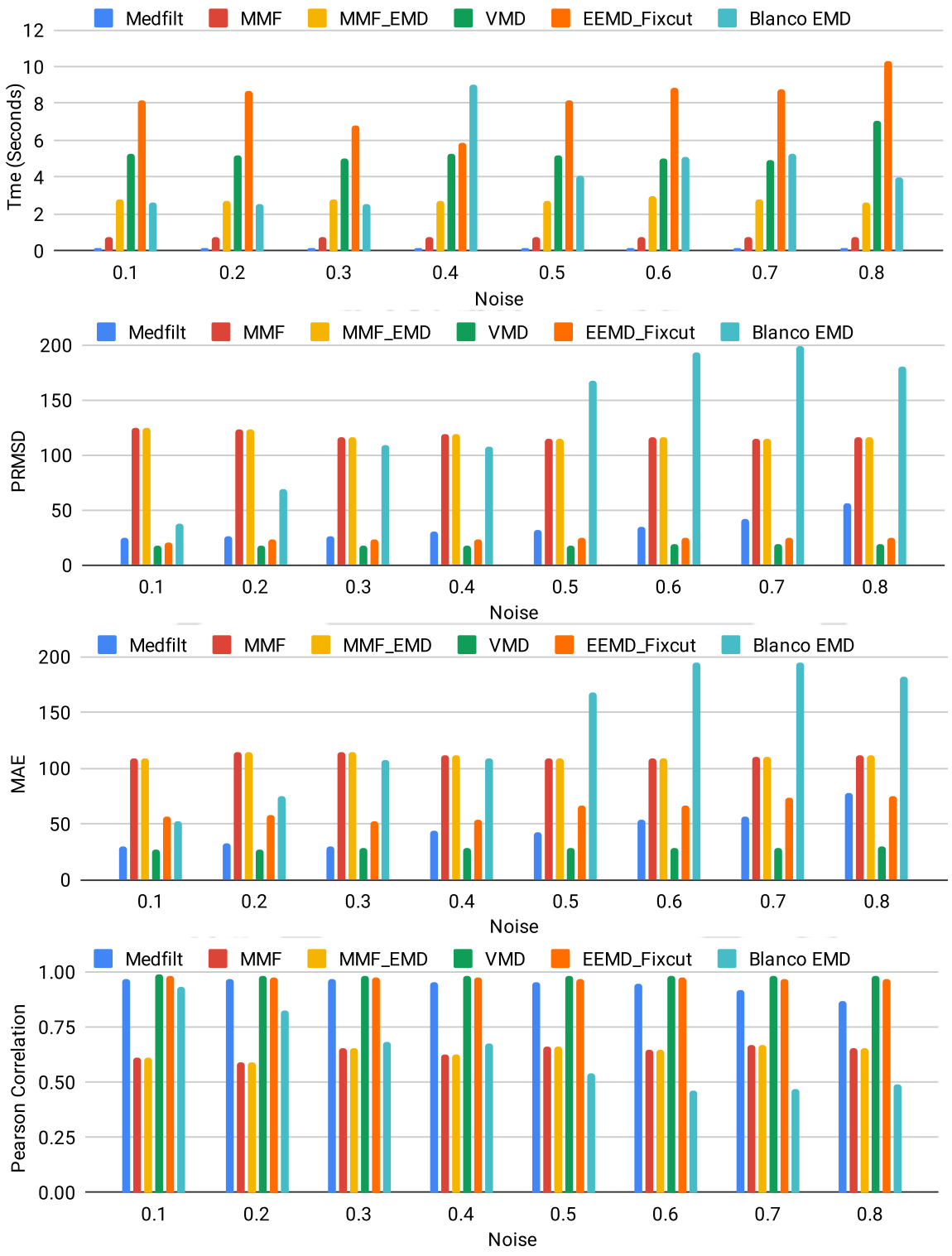


Figure 3.6: Comparison between the techniques for BW removal from NSR.

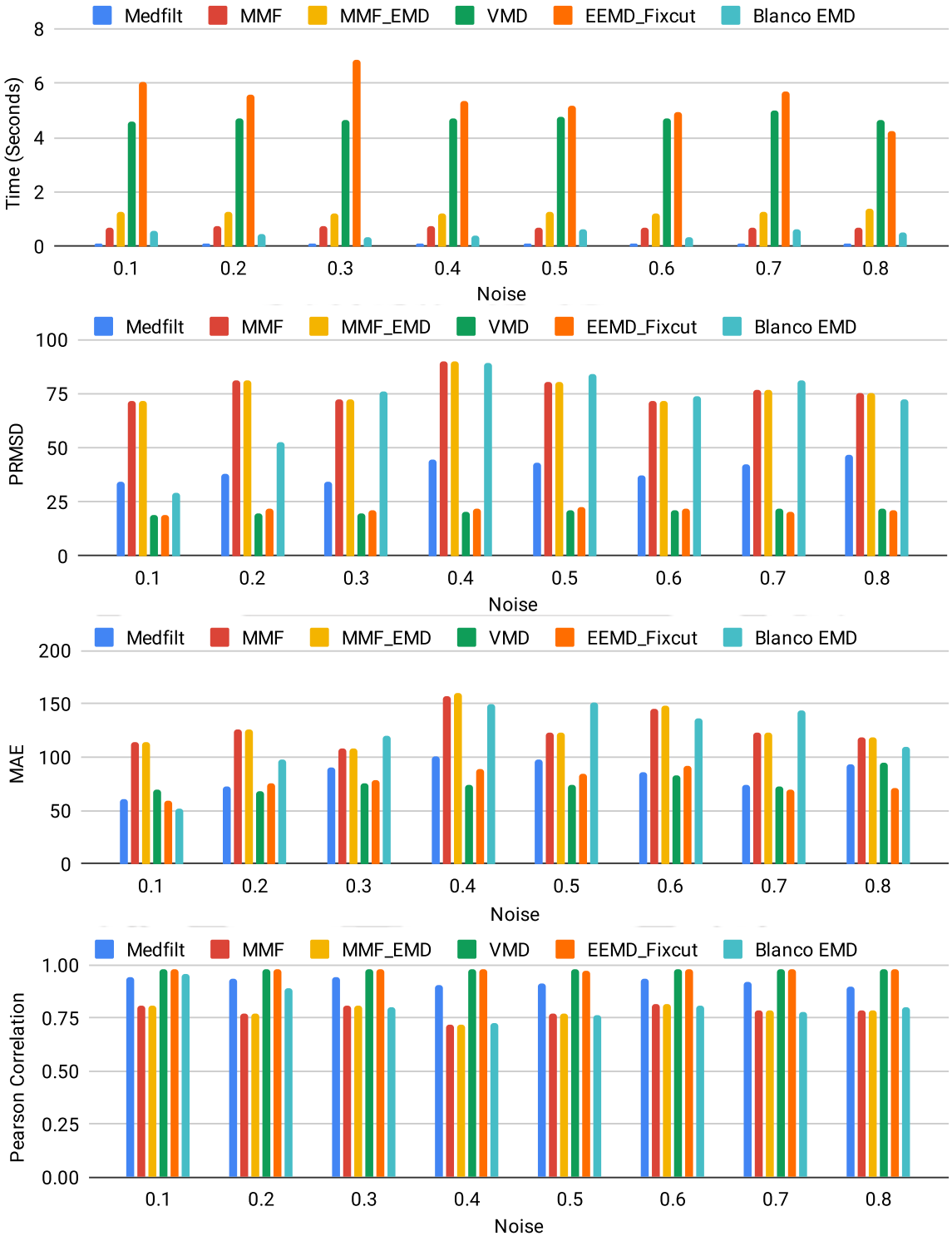


Figure 3.7: Comparison between the techniques for BW removal from VT Segment.

3.2 R-Peak Detection from Single Lead ECG

Heartbeats extracted from a single-lead ECG signal might indicate the occurrence of irregular or arrhythmic beats that might turn fatal if left untreated. But detecting

the beats is challenging as ECG is a non-stationary signal, and the signal varies from patient to patient and irregular beats may occur sporadically. Heartbeat detection or finding QRS complexes from ECG is necessary as its a precursor to finding heart rate, which is a crucial parameter to analyze many heart diseases, such as arrhythmias. It also serves as a reference for the detection of other characteristic waves and parameter measurements. Figure 3.8 provides an illustration of regular and irregular heartbeats extracted from single-lead ECG. An ideal heartbeat constitutes a P-wave, QRS-complex, and T-wave representing atrial depolarization, ventricular depolarization, and ventricular repolarization, respectively as described in Figure 3.9. According to the medical definition [83], R-peak is positive, Q-peak and S-peak are negative. Accurate QRS detection enables individual beat examination for acquiring information about heart rate variability, ST segments, heart rhythm abnormalities, and effective fine grained heartbeat classification. The QRS detection is challenging due to the dynamic behaviour of heart that produces non-stationary ECG signals and the presence of high and low frequency noise in ECG signal that gets added during signal acquisition. The types of noise and preprocessing needed is provided in Section 3.1. Efficient QRS detectors avoid storing unnecessary segments and capture essential irregular segments for clinical examination by physicians.

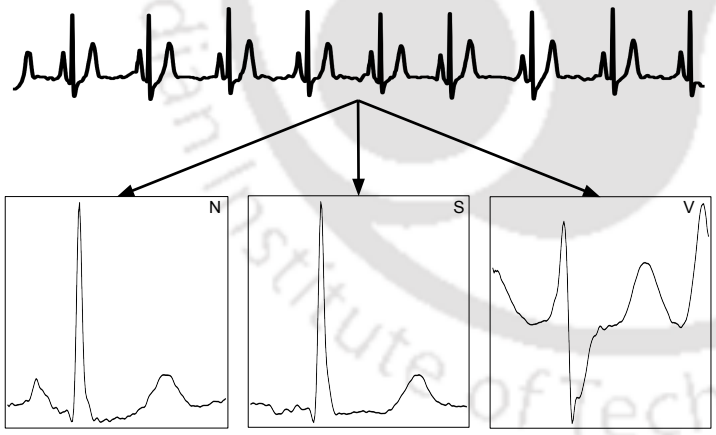


Figure 3.8: Heartbeats Extracted from ECG.

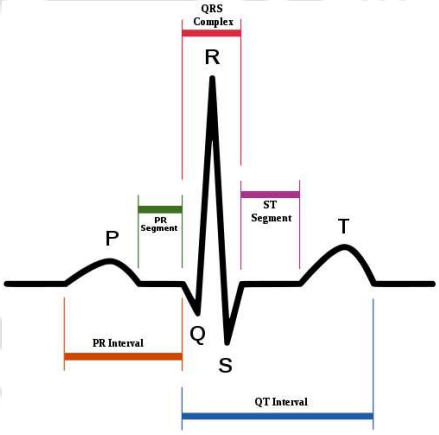


Figure 3.9: Normal Beat.

3.2.1 Brief Overview of Beat Detection Techniques

The first robust QRS detection technique analyzed positions and magnitudes of sharp waves [85]. However, the technique was not sensitive to the change of the integral information of QRS, leading to mistakes in the cases of low-amplitude QRS, sudden change of amplitude and high P-waves and T-waves. Li et al. [199] employed dyadic

wavelet transform (WT) to decompose signals into elementary building blocks that further provide sharp variation points of ECG. WT does not require any predefined cutoff frequency for detection but are non adaptive and put ECG under microscopic examination, making the method computationally expensive, restricting real world deployment. Xu et al. [200] used slope vector waveform where variable stage differentiation extracts slope vectors and non-linear amplification improves signal-to-noise ratio. Suarez et al. [201] employed template matching (TM) algorithm that calculates similarity between the predefined template and input signal using cross correlation. However, TM can not be adapted to nonstationary patterns, requires predefined templates and similarity metric. As patient's ECG might change over time, TM requires reconfiguration, involving regular human supervision. Training TM requires data which might be unavailable for new subjects with CVD. Yeh et al. [202] employed difference operation method that splits ECG signal into positive and negative halves and performs R peak detection, but the fixed thresholding makes the method, person and disease dependent. Pal et al. [203] used EMD to decompose ECG signal into high and low frequency components followed by nonlinear transformation to enhance QRS complex. However, the extracted components were sensitive to noise requiring effective filtering and threshold selection. Kim et al. [204] utilised the dual criteria of amplitude and QRS complex duration using finite impulse response filter, differentiation, and thresholding. He et al. [205] used K-nearest neighbours algorithm with model parameters adjusted through Particle Swarm Optimization. Malik et al. [206] utilised changes in heart rate and local amplitudes to detect R-Peak. Chen et al. [207] used exponential transform to narrow the magnitude difference of R-Peak and proportional-derivative to control the threshold to detect R-Peak. The aforementioned methods focus on strengthening the QRS complex and weaken noise. The filtering methods are simple and require less computational power but the selected frequency range limits precise R-Peak detection. The wavelet-based techniques require selection of mother wavelet apriori, making the algorithm unsuitable for real world.

3.2.2 Proposed Fractal Based R-Peak Detection

A simple, reliable, and intuitive algorithm for real-time R-peak detection is proposed using the geometry of fractals that meets the clinical needs. Fractals are described as an object with a never ending self similar structure that are created by repeating a simple process over and over as in recursion [208]. Fractal geometry provide meth-

ods for analyzing irregular structures existing in nature. It helps to understand the amount of similarity between two objects. The fact that ECG signal is produced by a nonlinear system, generating small or large variations, motivated the use of fractals for QRS detection. The QRS complex structure repeats periodically in ECG making it suitable for detection using fractals. The fast and low computational complexity properties of fractals make an effective real-time QRS detector. The approach consists of a Preprocessing stage, Fractal Analysis, and a Post Processing stage as described in Figure 3.10. During preprocessing, ECG is cleaned from noise followed by calculation of area under the curve using Mathematical Morphology (MM) operators such as erosion and dilation. The operators are implemented using dynamic programming with memoization that helps in achieving accurate results in a shorter duration. During post processing, the final peaks are extracted after resampling and thresholding the area curve to produce the peaks of ECG signal.

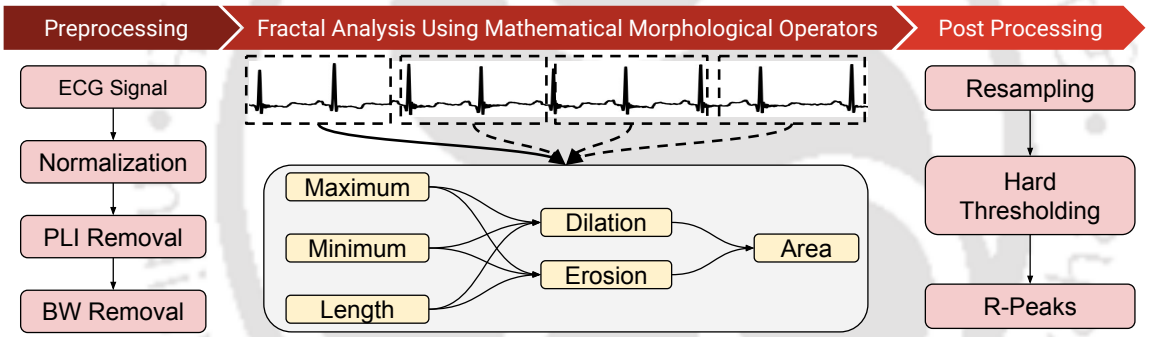
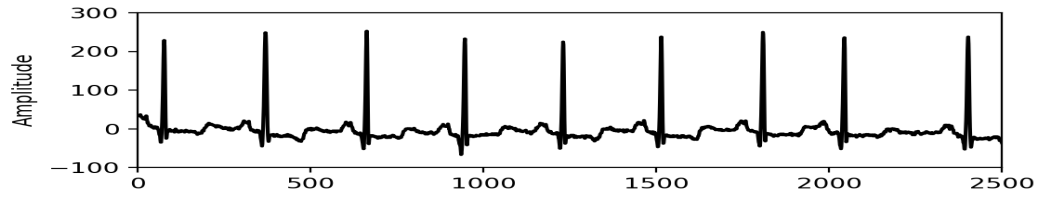


Figure 3.10: Different stages of the proposed approach.

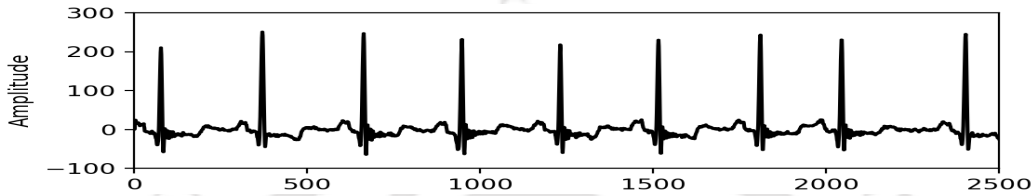
The validity of the application of MM principles for signal analysis depends on the following two properties: 1) Partial ordering and 2) Each subset should have a maximum and minimum. To check the above mentioned properties in the context of ECG signal, consider a signal $x[n]$ where $x[n] \subset \mathbb{R}$. Any set that is a subset of $\mathbb{R}$ satisfies the property, partial ordering $\preceq$ by the relation $\leq$ for any pair of elements $x_i, x_j \ \forall i, j, \in \mathbb{N}$. The second property is satisfied in $x[n]$ because of the existence of peaks and valleys in the signal. They are the local maxima and local minima of any subset of elements in $x[n]$. Therefore, MM operators are applicable for R-peak detection from ECG signals.

Data Description and Preprocessing: MLII lead from recordings of MIT-BIH database described in Section 3.1.3 are used for R-Peak detection. The original ECG signal is shown in Figure 3.11a. PLI is removed using a notch filter and BW is removed using a mean median filter with an initial window of 70 and the final window

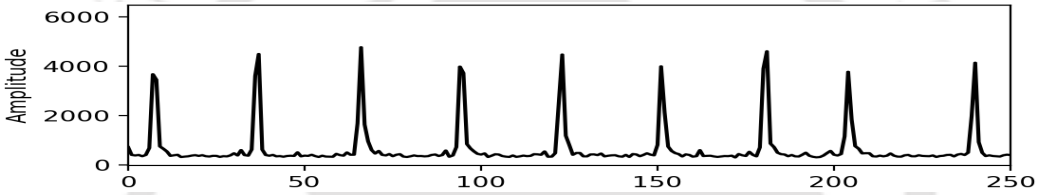
of 250. The clean signal is illustrated in Figure 3.11b.



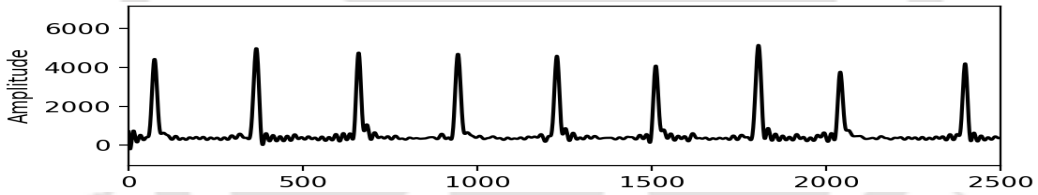
(a) Normalised Signal



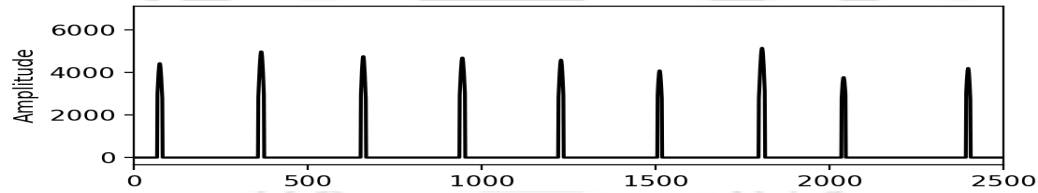
(b) Clean Signal



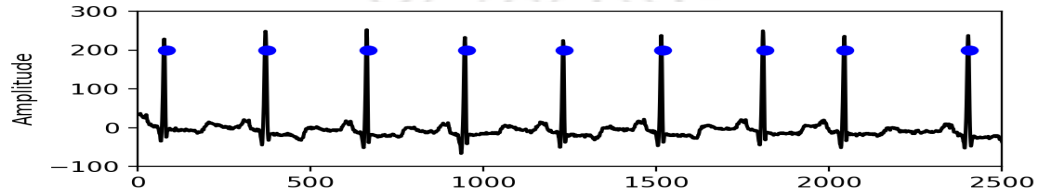
(c) Area obtained through Fractal Analysis



(d) Resampled Area



(e) Hard Thresholding



(f) Signal overlapped with Obtained Peaks

Figure 3.11: R-Peak Detection for Record 100 from MIT-BIH.

Fractal Analysis Using MM Operators: The MM operators, dilation and erosion are given in Equations 3.9 and 3.10 respectively. The area under a curve is

calculated as the difference between dilation and erosion and is given in Equation 3.11. The key idea of morphological analysis is the understanding of features of the regions in the signal by filling those regions by different operators. The Steps to calculate area are described in Algorithm 5. The recursive nature of Algorithm 5 was further optimised using the approach of dynamic programming with memoization that reduced the time complexity of the algorithm and made it computationally efficient. For a discrete-time finite-length ECG signal $S[n]$, $n(= 0, 1, ..., N)$, covers are used at discrete scales $\epsilon = (1, 2, ..., signal\_length/2)$ then the structuring element $G[n]$ (at scale $\epsilon = 1$) is restricted to three sample support with only two shapes: a triangle, defined by $G_t[-1] = 0$, $G_t[1] = 0$ and $G_t[0] = h \geq 0$, or a rectangle, defined by $G_r[-1] = G_r[1] = G_r[0] = h \geq 0$. The height h matches the amplitude range of the signal $S[n]$ and controls finesse or coarseness of multi-scale area measurements [209]. The final area is calculated using the difference between dilation and erosion. Figure 3.11c depicts the area obtained after applying MM operators.

$$\begin{aligned} S \oplus G[n] &= \max_{-1 \leq k \leq 1} \{S[n+k] + G[k]\}, \ \epsilon = 1 \\ S \oplus G_{\epsilon+1} &= (S \oplus G_\epsilon) \oplus G, \ \epsilon \geq 2 \end{aligned} \tag{3.9}$$

$$\begin{aligned} S \ominus G[n] &= \min_{-1 \leq k \leq 1} \{S[n+k] - G[k]\}, \ \epsilon = 1 \\ S \ominus G_{\epsilon+1} &= (S \ominus G_\epsilon) \ominus G, \ \epsilon \geq 2 \end{aligned} \tag{3.10}$$

$$\mathbb{A} = \sum_{n=0}^{N} \Big[(S \oplus G_\varepsilon)[n] - (S \ominus G_\varepsilon)[n] \Big] \tag{3.11}$$

Post-Processing: After obtaining the area through fractal analysis in Figure 3.11c, the length of area is reduced with a factor of frame size. So for accurate QRS detection we resampled the signal to original signal length. Figure 3.11d depicts the resampled area curve. The frame size is varied to achieve best results. After resampling, thresholding was performed to reduce false detections. However, it leads to some missing beats. Figure 3.11e depicts the thresholded signal. The thresholding was also varied to obtain best results. Thresholding also led to both false and missed beats. Figure 3.11f depicts the overlapping of obtained peaks and original signal. For the comparison of obtained beats with the annotation provided in the data set by cardiologist, a window of size 60 samples was employed. If the extracted markers are not in this range then the beat is assumed to be missed by our algorithm and if more

Algorithm 5 Area Calculation using MM operators

Input: ECG signal S[n], FS: Frame size
Output: Area
$n_{frames} \leftarrow length_{signal}/FS$
$k \leftarrow 0$
$Area \leftarrow \emptyset$
for $k \leftarrow 0$ **to** n_{frames} **do**
  $seg \leftarrow signal[FS * k : FS * (k + 1)]$
  $h \leftarrow max(seg) - min(seg)$
  $h \leftarrow h/len(seg)$
  $area_k \leftarrow dilation(seg, h) - erosion(seg, h)$
  $k \leftarrow k + 1$
  $Area \leftarrow Area \cup area_k$
end

than one marker is found out by our algorithm then it is assumed to be an extra beat.

3.2.3 Results and Discussion

The performance of proposed R-peak detection method is measured using Sensitivity, Predictivity, and Detection Error Rate (DER) as provided in Equation 3.12. Here, TP represents a real peak that is identified correctly, FN when a real peak is not detected, FP when a false peak detected as a real peak. So, if the algorithm misses more number of beats, the sensitivity reduces and if more number of extra beats are detected then the predictivity is reduced. Similarly, if the algorithm finds actual beats with minimum extra and missed beats, the DER reduces. The frame size was varied from a window of 10 timestamps to 150 timestamps and the thresholding was varied from 1 to 9 times of mean area curve. Fixed hard thresholding was performed [84] for detecting final R-Peaks.

$$\begin{aligned} Sensitivity &= \frac{TP}{TP + FN} \\ Predictivity &= \frac{TP}{TP + FP} \\ DER &= \frac{FP + FN}{TP} \end{aligned} \tag{3.12}$$

Detection using Triangular Structuring Element: The sensitivity and predictivity achieved with varying frame size and thresholding using triangular structuring element for area calculation is provided using the radar plot in Figure 3.12. An

increase in frame size, the detection performance increased till frame size 120 and gradually deteriorated with increasing frame size with lower threshold. It can also be observed that with increasing frame size, an increase in threshold also provides better performance. Increasing the hard thresholding after five times of mean area signal degraded the model performance. The performance remains consistent when both thresholding and frame size are increased together. The maximum sensitivity and predictivity were not obtained at a particular threshold and frame size. Maximum sensitivity was obtained using area mean as threshold and frame size of 10 timestamps. Maximum predictivity was obtained using mean area as threshold and frame size of 120 timestamps. Figure 3.13 depicts the Logarithmic Column Chart for change in detection error rate with varying frame size and thresholding using triangular structuring element. The minimum DER was achieved for mean thresholding and a frame size of 90 timestamps. Increasing thresholding with less frame size also enabled the algorithm to detect significant peaks thereby minimizing DER. The algorithm achieved a sensitivity of 95.78%, predictivity of 97.53%, and DER of 8.44% using triangular structuring element with frame size of 90 timestamps and mean thresholding. In the figures, only those combinations of threshold and frame size are depicted where the algorithm was able to detect significant amount of peaks in the MIT-BIH records. Those combinations are excluded where the algorithm completely missed out on R-peaks.

Detection using Rectangular Structuring Element: The sensitivity and predictivity achieved with varying framesize and thresholding using rectangular structuring element for area calculation is provided using the radar plot in Figure 3.14. The performance is quite similar, if not better than triangular structuring element. Increase framesize increases sensitivity and predictivity till framesize 120 and gradually deteriorates later. Increasing framesize with threshold provides consistent performance. As explained earlier, maximum sensitivity and predictivity were obtained using different combinations. Maximum sensitivity was obtained using area mean as threshold and framesize of 10 timestamps. Maximum predictivity was obtained using twice the mean area as threshold and framesize of 20 timestamps. Figure 3.15 depicts the Logarithmic Column Chart for change in detection error rate with varying framesize and thresholding using rectangular structuring element. The minimum DER was achieved for mean thresholding and a framesize of 90 timestamps. Increasing thresholding with less framesize also enabled the algorithm to detect significant peaks thereby minimizing DER. The algorithm achieved a sensitivity of 94.95%, predictivity of 97.70%, and DER of 8.21% with framesize of 90 timestamps and mean

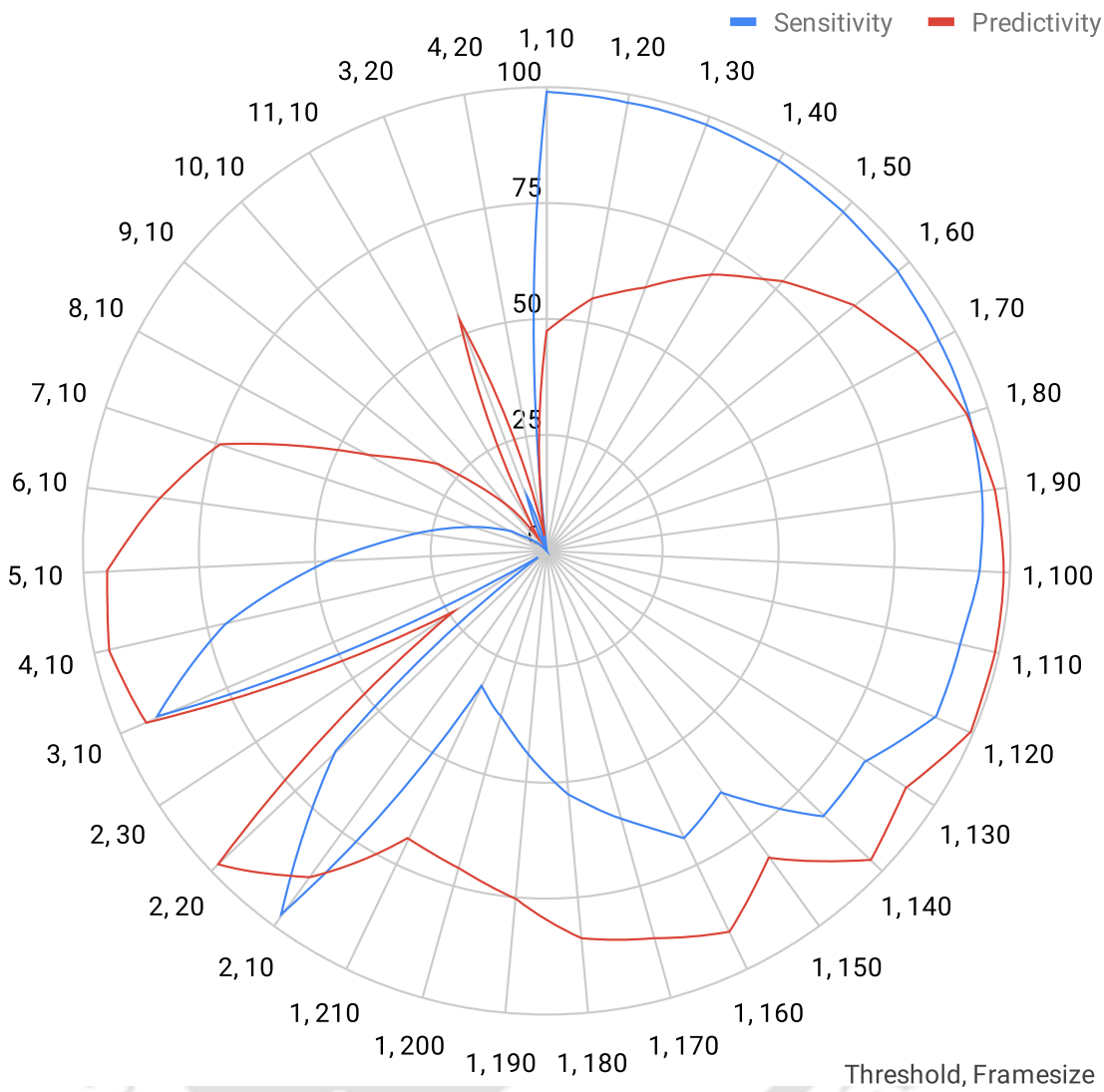


Figure 3.12: Radar Plot of Sensitivity and Predictivity with varying threshold and frame size using Triangular Structuring Element.

thresholding. In the figures, only those combinations of threshold and framesize are depicted where the algorithm was able to detect significant amount of peaks in the MIT-BIH records.

Detailed MIT-BIH Evaluation using Rectangular Element: Area calculation using rectangular element achieved better performance than triangular element. Detailed results for MIT-BIH records are provided in Figure 3.16. The highest number of beats detected in each record was obtained using mean of area as thresholding and framesize of 90 timestamps. Figure 3.16a depicts the change in predictivity, Figure 3.16b depicts the change in sensitivity, and Figure 3.16c depicts the change in detection error rate using the mentioned parameters. It is evident from the figures

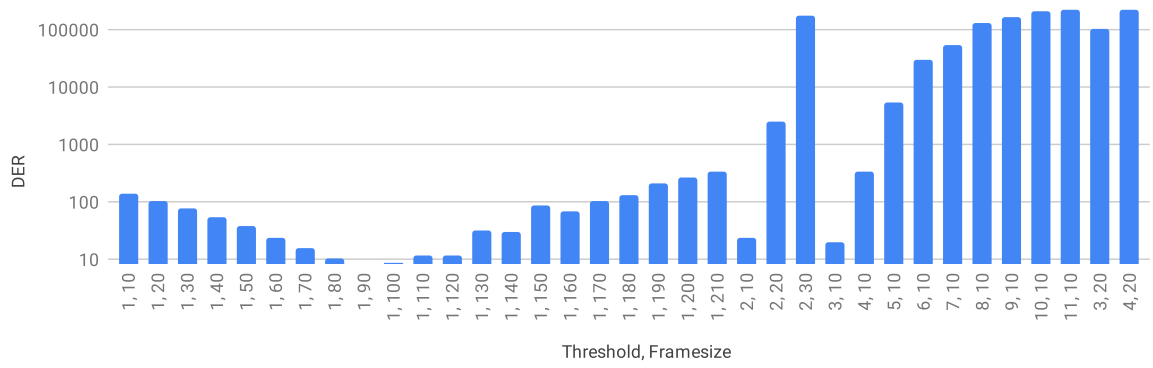


Figure 3.13: Logarithmic Column Chart of DER with varying threshold and framesize for Triangular Structuring Element.

that the proposed algorithm was able to achieve higher sensitivity and predictivity for non-arrhythmic records (100-124) in comparison to arrhythmic records (200-234). The predictivity is high in most of the records as compared to sensitivity, meaning that less extra beats are detected but more beats are missed by the algorithm. A similar pattern was observed in DER, less DER was observed in non-arrhythmic records as compared to arrhythmic records, meaning that higher number of actual and extra beats were detected and fewer beats were missed. The few key areas or patterns we discovered that caused the algorithm to under preform were broad duration and small amplitude QRS complexes. The reason behind low accuracy and high DER were reversed QRS polarity and low-amplitude QRS complexes as the area amplitude was insufficient for detection. The false positives increased due to reversed QRS polarity and high amplitude P and T wave. To counteract this issue, we implemented a new function that checked for multiple peaks in a window and choose the later peak as the former peak occurred probably due to the P wave. This lead to reduced sensitivity but high predictivity as many beats were missed. This issue was tackled by reducing the framesize from 200 to 60 timestamps. FN occurred due to high magnitude thresholding in the algorithm. Fixed hard thresholding was performed [84] for detecting final R-Peaks.

Comparison with Existing Methods: The proposed fractal based QRS detection approach is compared with state of the art methods in Table 3.1. Since the performance of the algorithms depend strongly on the selection of the ECGs for the evaluation purposes, PT method [85], KNN-PSO method [205] and the DOM [202] are re-implemented for comparing the proposed approach. Often the previous approaches exclude extremely noisy data before peak detection. The proposed approach achieves Sensitivity, Positive predictivity and detection error rate of 95.82%, 95.15%

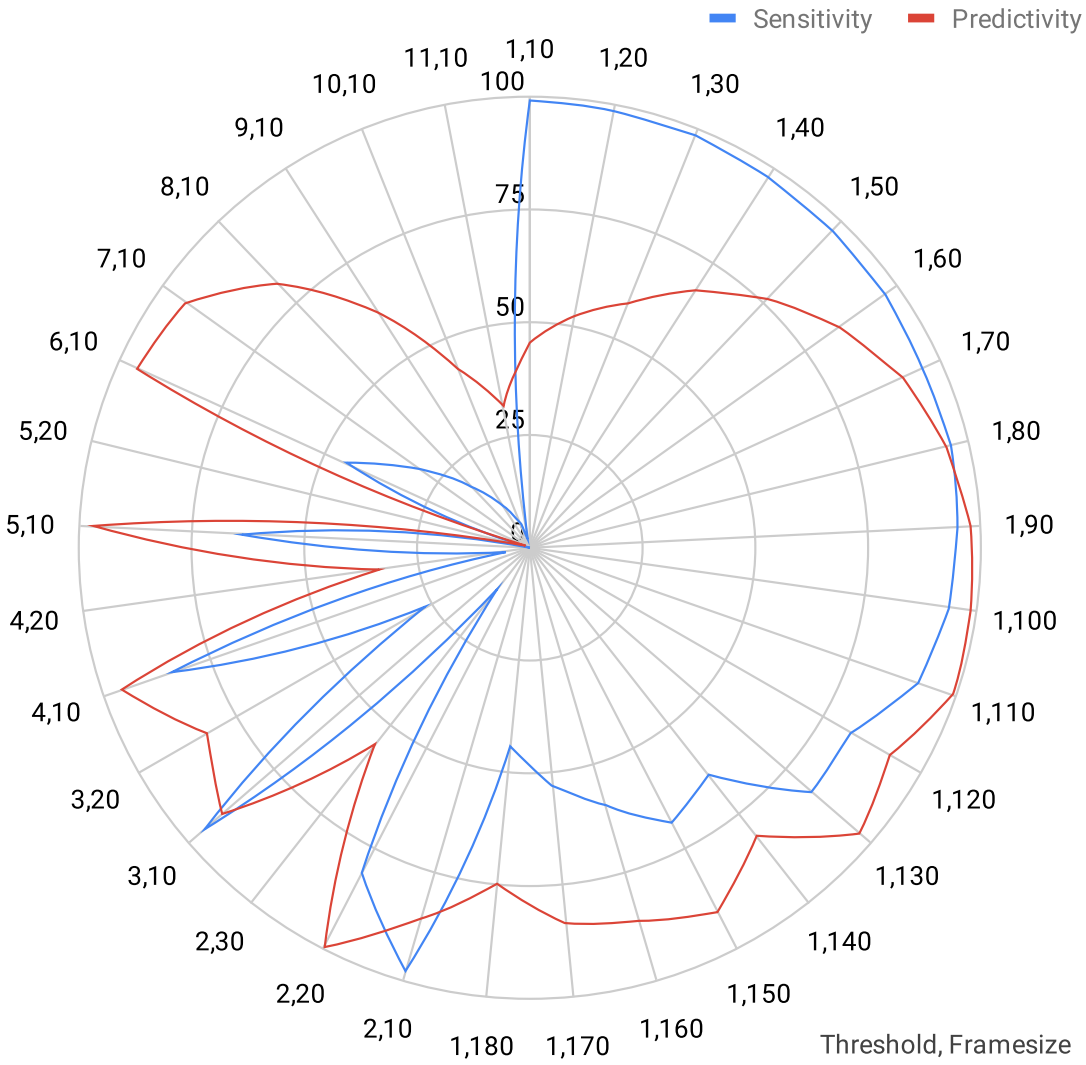


Figure 3.14: Radar Plot of Sensitivity and Predictivity with varying threshold and framesize using Rectangle Structuring Element.

and 10.16%. All judgements of correctness are based upon the annotations in the database.

Table 3.1: Comparison with Existing Methods.

Method	Sensitivity (%)	Predictivity (%)	DER
PT [85]	96.08	97.89	9.43
KNN-PSO [205]	96.26	98.33	5.97
DOM [202]	96.81	97.95	5.78
Fractal_Rect	99.03	99.41	8.21
Fractal_Tri	95.78	94.99	10.43

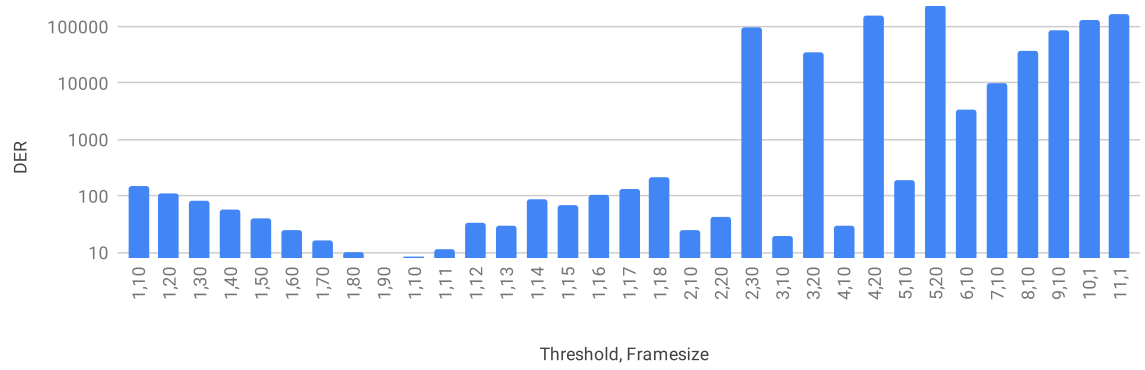
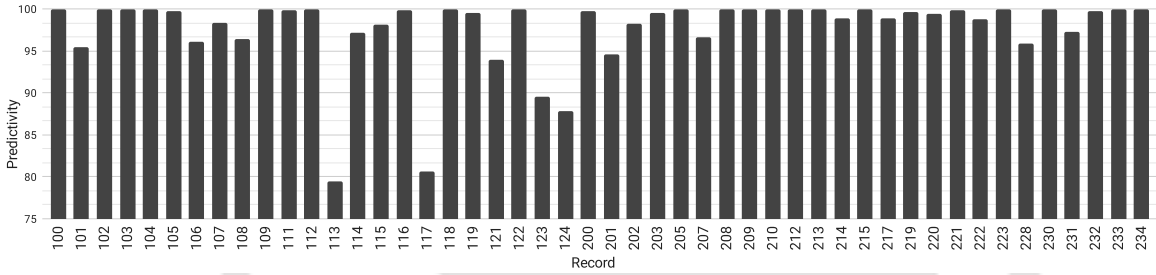
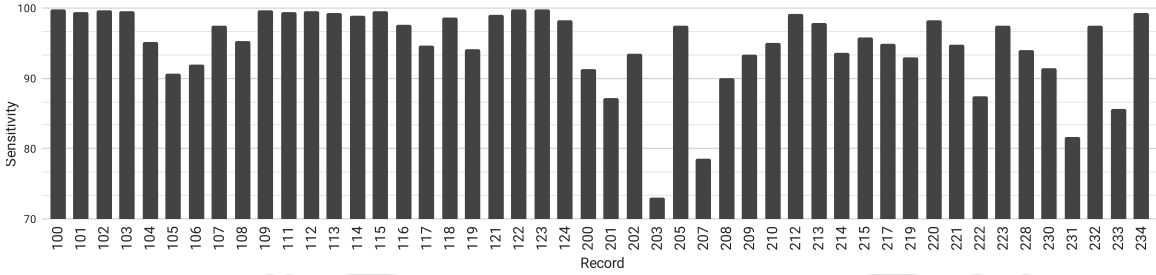


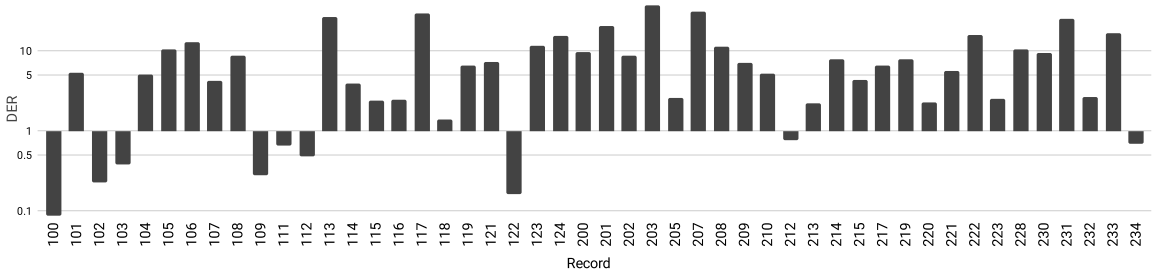
Figure 3.15: Logarithmic Column Chart of DER with varying threshold and framesize for Rectangular Structuring Element.



(a) Predictivity for all records in MIT-BIH database.



(b) Sensitivity for all records in MIT-BIH database.



(c) Logarithmic Column chart of DER for all records in MIT-BIH database.

Figure 3.16: Results using Rectangular Structuring Element with mean thresholding and a framesize of 90 timestamps.

The proposed R-peak detection algorithm uses erosion and dilation operators

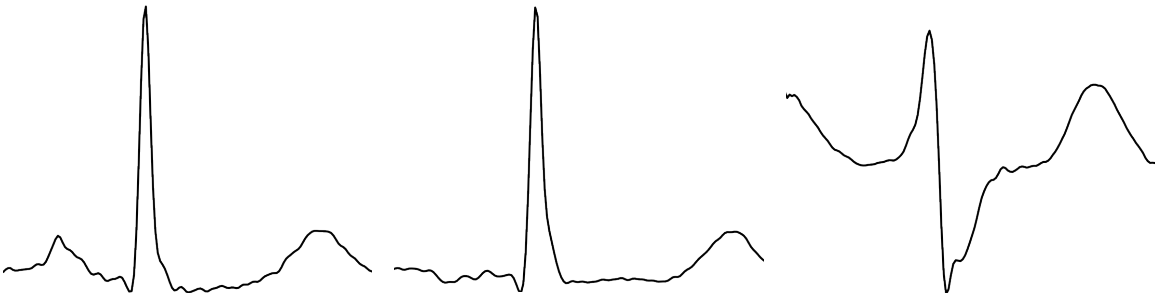
for area calculation. The area obtained using either rectangle or triangle structuring elements is further resampled and thresholded to detect peaks. The method achieved a sensitivity of 95.82%, predictivity of 95.15%, and DER of 10.16%. The algorithm was tested against all the records available in the dataset. The detected beats are provided to a classifier that performs fine grained heartbeat classification.

3.3 Heartbeat Synthesis using Generative Models

A generalized classifier can achieve state of the art performance using huge amount of labeled, diverse, and realistic-looking regular and irregular heartbeats, which is difficult to acquire in practice. Privacy concerns restrict the access to personal ECGs, hindering the development of a generalized classifier. The challenge in synthesizing heartbeats occurs because the biological and physiological systems generating these beats are highly complex. Publicly available datasets sometimes fail to satisfy a certain criterion concerning a study as they might miss out on relevant information, making heartbeat synthesis crucial. This necessitates the synthesis of regular (Normal beat) and irregular beats: supraventricular ectopic beat (SVEB) and ventricular ectopic beat (VEB) as recommended by the Association for the Advancement of Medical Instrumentation (AAMI) [7]. The ectopic beats are felt as palpitations caused due to an extra or skipped heartbeat making the patient feel the heart lurch or an extra strong beat occurring momentarily. They usually occur in succession like ectopic-normal-ectopic or continuation of several ectopic beats within a small duration. Ectopic beats are more sinister when arising from ventricles than those arising from atria. SVEB originates from the upper chambers of heart or atria. These are also called atrial premature beats or premature atrial contraction and may lead to atrial fibrillation. VEB originates from the lower chambers of heart or ventricles. These are also called Premature ventricular contractions (PVC) and may lead to ventricular tachycardia and ventricular fibrillation. A Normal Beat (N), SVEB, and VEB is illustrated in Figure 3.17. In a normal beat, all the characteristic waves P, QRS, T are visible. In SVEB, the P wave is usually missing, and in VEB, the QRS complex is wider than the normal beat with a discordant ST segment.

3.3.1 Brief Overview of ECG Synthesis Techniques

In the past, ECG has been generated synthetically using predefined synthesizers, interpolation, and extrapolation, but these methods generate limited variation in syn-



(a) Normal Beat (b) SVEB (c) VEB

Figure 3.17: Illustration of regular and irregular beats.

thesized beats. The augmentation techniques could be categorised into traditional methods that generates ECGs using heart rate features [210], 3-D vectorcardiogram formulation and non-linear models [211]. However, these techniques depend on predefined functions making it difficult to incorporate sufficient information about the basic ECG characteristics.

ECG can also be generated with the help of deep learning based generative models such as Generative Adversarial Networks (GAN) models that learn the training data distribution and then generates ECG segments of a cardiac abnormality [91, 212, 213]. Brophy et al. [212] used Wasserstein GANs (WGANs) [214] with gradient penalty to synthetically generate sinusoidal waves, PPG signals, and ECG signals by feeding rasterized images of these signals to GANs and generated new images. However, their approach was limited by the resolution of the generated image. Haradal et al. [213] employed a recurrent neural network (RNN) based long short term memory (LSTM) model for ECG and EEG generation. Zhu et al. [215] employed two layers of bidirectional LSTM (BiLSTM) for generator followed by a convolution neural network (CNN) based discriminator to generate ECG signals. They compared the developed model with recurrent neural network autoencoder and recurrent neural network variational autoencoder. Zhou et al. [216] detected anomalous rhythms in ECG by proposing Beat GAN that generated new beats by extracting the latent input noise from actual ECG rhythms, making the model input dependent. Delaney et al. [217] generated synthetic Sine waves and normal ECG using two BiLSTM followed by CNN along with a mini-batch discrimination layer. Hazra et al. [218] proposed SynSigGAN using bidirectional grid LSTM in generator and CNN in discriminator and generated ECG, EEG, Electromyogram (EMG), and PPG signals.

Researchers have also encoded conditional dependencies in GANs to generate class specific signals. Wang et al. [219] employed an auxiliary classifier GAN (AC-

GAN) to generate normal beat, left bundle branch block (LBBB), right bundle branch block (RBBB), premature atrial contraction (PAC), and premature ventricular contraction (PVC). Ye et al. [220] proposed RPSeqGAN by training a sequence generative adversarial network (SeqGAN) using the concept of policy gradient from reinforcement learning. Wulan et al. [221] proposed three GAN models, namely, WaveNet-based model, SpectroGAN model, and WaveletGAN model for the generation of normal, LBBB, and RBBB. Their first model preprocesses the input using μ-law companding transformation followed by convolutional layers with dilation; the second and third model used short-term Fourier transform (STFT) and stationary wavelet transform (SWT) to obtain suitable input for GAN. Golany et al. [87] proposed a personalized GAN (PGAN) using an LSTM model that synthesizes minority class beats, including SVEB and VEB, and incorporated two loss functions, mean square loss and binary cross-entropy loss. Golany et al. [88] in their second work incorporated ordinary differential equations representing the heart dynamics into the optimization process of GAN.

Aforementioned methods have generated synthetic ECG signals. However, the drawbacks include: (i) Only a few methods follow the AAMI [7] recommended beat generation, thereby posing a problem of lack in minority class beats that needs to be tackled; (ii) Use of latent input extracted from actual ECG beats, making the generative model biased; and (iii) Lack of evaluation metrics for generated signal verification. In this work, AAMI recommended N, SVEB, and VEB are generated by encoding class related dependencies in the Deep Convolution Conditional Generative Adversarial Network (DCCGAN).

3.3.2 Proposed Deep Convolution Conditional GAN

The heartbeat generation framework generates three classes of beats using a convolution based conditional GAN. GANs consist of a generative (probabilistic) model that learns the input probability distribution and a discriminative model that discriminates between the real and generated samples. Generator $G(z)$ maps latent input to the probability distribution of real data P_r. Here, z represents the latent input. Discriminator $D(x)$ classifies the input sample as real or fake, here x represents the input beats provided to the discriminator. Mathematically, GAN training resembles a two-player minimax game, where the generated data distribution is brought as close as possible to the real data distribution, and the discriminator aims to be better at differentiating between the real and generated beats. GAN optimisation is performed

using the loss function defined in Equation 3.13. Here, z is randomly sampled from synthetically generated data distribution (P_z) and x is randomly sampled from real data distribution (P_r). The former term in Equation 3.13, $\log[D(x)]$ predicts that the data is real and the latter term $\log[1 - D(G(z))]$ predicts that synthetically generated data is fake. The generator and discriminator training alternates by keeping the training of other model constant and alternately maximizing and minimizing $V(G, D)$ until synthetically generated data is indistinguishable from real data.

$$\min_G \max_D V(D, G) = E_{x \sim P_r(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \tag{3.13}$$

Broadly categorising, GANs can be classified into: Vanilla GAN [90], Conditional GAN [222], and Deep Convolutional GAN (DCGAN) [223]. Vanilla GAN is the basic GAN built using the multi-layer perceptron (MLP) that optimizes the Equation 3.13 using stochastic gradient descent. The Conditional GAN encodes conditional dependencies in generator and discriminator along with the conventional input allowing for the generation of specific beats. Deep Convolutional GAN (DCGAN) is the most popular and successful GAN that replaces traditional MLP with convolution layers (with strides) without max-pooling layers, making GAN training fast and stable.

Conditional GAN: A class encoded convolution GAN is used to generate class-specific heartbeats (N, SVEB, and VEB). The class information (c) is incorporated with conventional input transforming $G(z)$ to $G(z|c)$ and $D(x)$ to $D(x|c)$ leads to the deep convolution conditional generative adversarial network (DCCGAN). Figure 3.18 illustrates the architecture of the proposed DCCGAN. The architecture details of the generator and discriminator are illustrated in Figure 3.19.

Modified DCCGAN Loss Function: The proposed DCCGAN takes two inputs, a class label (c) and the real beats corresponding to that class ($x|c$) or the generated beats corresponding to that class ($G(z|c)$). This modifies the basic GAN loss function as described in Equation 3.13 to DCCGAN loss function in Equation 3.14. Here, $P_r(x)$ and $P_z(z)$ are the real and generated beat distributions, respectively.

$$\min_G \max_D V(D, G) = E_{x \sim P_r(x)}[\log D(x|c)] + E_{z \sim P_z(z)}[\log(1 - D(G(z|c)))] \tag{3.14}$$

Generator Model: The generator model intakes a latent input vector (z) and class label (c) and generates a fake heartbeat ($G(z_n)$), where $z = \{z_n\}_{n=1}^{N}$, N is the length of noise vector, $G(z_n) \in \mathbb{R}^k$, and k is similar to the dimension of a real heart-

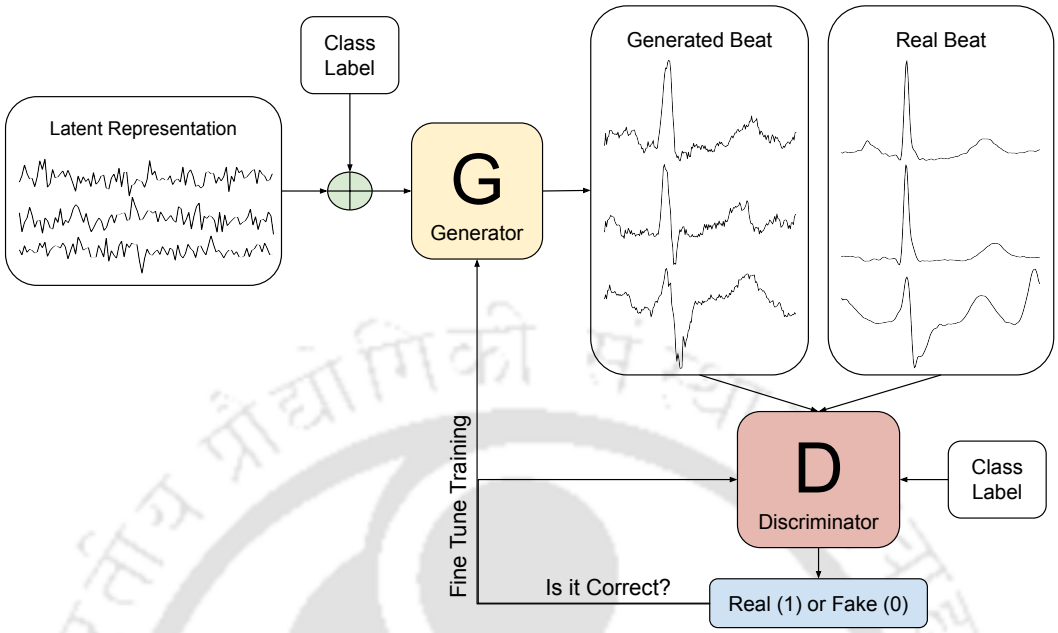


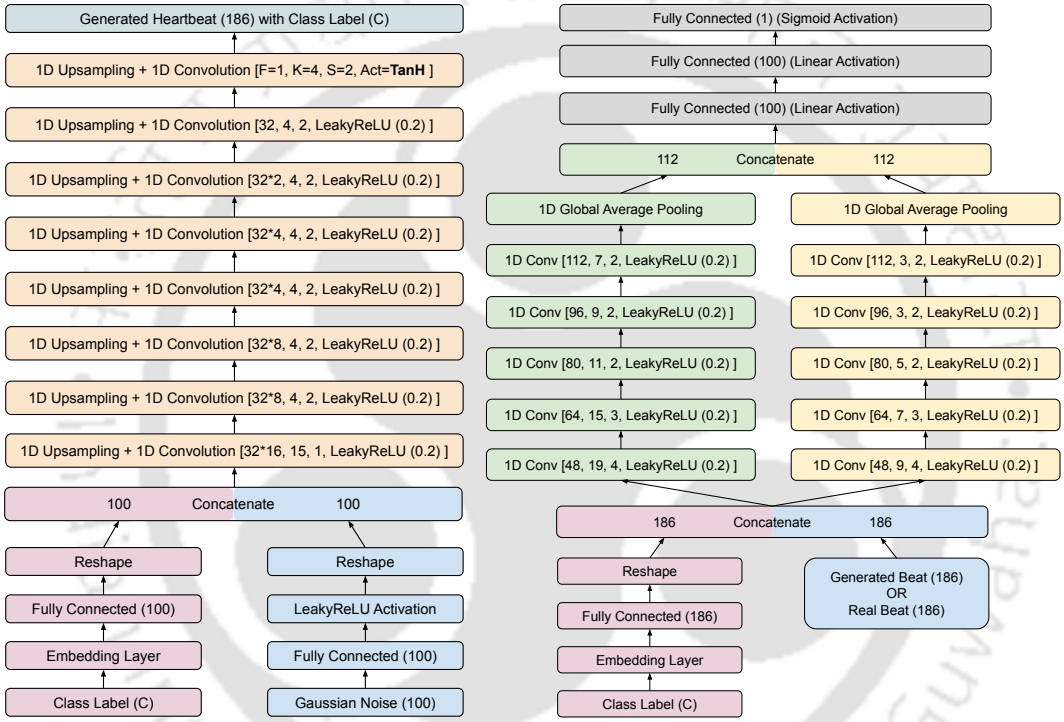
Figure 3.18: Basic Architecture of Conditional GAN.

beat. The latent input is randomly sampled from a gaussian noise distribution. The generator aims to approximate the underlying real heartbeat distribution given the corresponding class label $P(hb|c)$ and produces fake beats to deceive the discriminator model. The modified generator loss is depicted by Equation 3.15. It tries to minimize the log probability predicted by the discriminator model for synthetically generated beats, thereby encouraging the generation of those beats with less probability of being fake.

$$L(G) = \texttt{minimize}[\log(1 - D(G(z|c)))] \tag{3.15}$$

Figure 3.19a depicts the architecture of the generator. Generator model intakes a latent input (z) of length 100 sampled from a gaussian (normal) noise distribution ($\mathcal{N}(\mu = 0, \sigma = 1)$) [91, 224] and a class label (c), where $c \in \{0, 1, 2\}$. Here, $(0, 1, 2)$ corresponds to Normal beat, SVEB, and VEB. The label is incorporated by employing an embedding layer followed by a fully connected layer with 100 neurons, similar to input shape, and later reshaped to match the dimensions of gaussian input noise. The Gaussian noise is provided to a fully connected layer with 100 neurons followed by LeakyReLU activation function with a negative slope coefficient of 0.2 followed by a reshaping layer for concatenation with the encoded class information. The input noise and class label are then concatenated and provided to a 1-Dimensional (1D) Upsampling layer that increases the dimension of input by a factor of 2 followed by a 1D Conv layer with 32×16 filters of size 15 and stride 1 with LeakyReLU

activation function with a negative slope coefficient of 0.2. Six more Upsampling and Conv layers are added in a cascaded fashion for making the generator model as illustrated in Figure 3.19a. The combination of the 1D Conv layer and 1D Upsampling layer imitates a 1D Deconvolution layer. In penultimate layer, activation function is changed from LeakyReLU to hyperbolic tangent function, and generated heartbeat with 186 dimensions, in practice, should resemble a realistic-looking heartbeat of class (c).



(a) Generator Architecture (b) Discriminator Architecture

Figure 3.19: Illustration of Generator and Discriminator Architecture.

Discriminator Model: The discriminator model intakes real heartbeats (x), and synthetically generated heartbeats $(G(z|c))$ along with their corresponding class label (c) and classifies them as either real or fake. Here, x is represented as $x_n \in \mathbb{R}^T$, $G(z)$ is represented as $G(z_n) \in \mathbb{R}^T$, T corresponds to the beat dimension. The modified discriminator loss is described using Equation 3.16. The loss function aims at maximizing the log probability of real beats and inverse log probability of synthetically generated beats.

$$L(D) = \texttt{maximize}[\log D(x|c) + \log(1 - D(G(z|c)))] \tag{3.16}$$

Figure 3.19b depicts the architecture of the discriminator. The discriminator model intakes the heartbeat (both real and fake) (x) of 186 length and a class label (c), where $c \in \{0, 1, 2\}$. Here, $(0, 1, 2)$ corresponds to Normal, SVEB, and VEB class of beats. The class label is incorporated in a similar fashion as described in the generator model. The input beats and reshaped embedded class label are further concatenated, and a parallel convolution neural network is employed. The main idea behind applying such convolution layers in parallel is that the beats of 186 dimension consist of both local and global patterns. The global patterns are extracted using the large kernels embedded in the left part of PCNN, and the local patterns are extracted using the small kernels embedded in the right branch of PCNN. The number of filters, stride, and activation function in each adjacent convolution layer are kept similar, and only the kernel size is varied. For instance, the first convolution layer in the left branch after input concatenation encompasses a 1-Dimensional convolution layer with 48 filters of size 19 and stride 4 with LeakyReLU activation function with a negative slope coefficient of 0.2. Four more convolution layers are added in a cascaded fashion in both branches, followed by 1-Dimensional global average pooling (GAP) layer [225].

GAP calculates the spatial average of filters, making it robust to spatial translations present in input ECG beats. The advantages of GAP over the combination of flatten and fully connected layer are (i) less prone to overfitting; (ii) no dependency on external regularization as it behaves like a structural regularizer; (iii) no trainable parameters similar to max-pooling layer. The reduced parameters lead to significantly faster training and reduced model size, making it suitable for mobile devices. The GAP layer reduces the last layer dimensions from $(2, 112)$ to (112). These outputs are further concatenated and provided as input to two fully connected layers with 100 neurons, each with a linear activation function followed by a single neuron layer with the sigmoid activation function that generates a value (v), where $v \in (0, 1)$. (v) is thresholded and is classified as fake if $v \in (0, 0.5)$ and real if $v \in [0.5, 1)$.

DCCGAN Training: The training alternates between generator and discriminator. During the discriminator training, n samples are randomly sampled from real heartbeats and n synthetic heartbeats are generated using the generator network for the three corresponding classes. For the real heartbeats, the discriminator is provided with labels in the range $[0.8, 1]$, and for synthetic heartbeats, the discriminator is provided with labels in the range $[0, 0.2]$. It can be observed that the labels are modified from 0 and 1 to an interval using the concept of soft labels allowing a faster convergence in the discriminator loss during initial batches of training.

After training the discriminator, the generator model is trained where the dis-

criminator model weights are frozen so that they do not get affected during generator training. During generator training, $2 \times n$ samples are randomly sampled from the Gaussian noise distribution and provided to the generator in addition to the class information allowing the generator to generate fake heartbeat samples. These fake and real heartbeats are provided as input to the discriminator with the corresponding class labels. Again soft labels are provided as input to the discriminator, i.e., labels in the range $[0.8, 1]$ for real beats and labels in the range $[0, 0.2]$ for synthetic beats. The discriminator performs a forward propagation and compares the predicted label with soft labels and backpropagates the error to the generator network. During the backpropagation, the discriminator weights are not changed, and the generator weights are modified, thereby training the generator. Therefore, the generator network is trained in an adversarial fashion using the discriminator.

This training sequence is followed for training both the networks alternatively for several batches, and their losses are recorded. The training is not stopped following any early stopping criteria as this remains an open problem in the research area of GANs and because the GAN training is unstable. However, the generator models were saved after every second batch. The training aims to achieve Nash Equilibrium by approximating generator probability distribution $P_z(z)$ to the real heartbeat distribution. Adam optimizer [226] is employed during the training with $LR = 0.0002$ as suggested in [223] along with batch-wise training. More filters with large sizes are preferred for generator and discriminator as larger filters cover more signal timestamps and account for more information present in the beat and previous filter. It also allows for maintaining smoothness in the information present in the filters.

Moreover, DCCGAN took several hundred batches of training before generating any meaningful beats for the corresponding class. In initial batches, random noise signals were generated, and later, realistic-looking beats were generated. For this reason, early stopping criteria was not used as sometimes the generator model was unable to generate meaningful beats after several hundred batches of training. For some models, the discriminator loss approached 0, and the generated beats resembled random noise-looking signals. Therefore, the training was restarted after modification of certain parameters such as the kernel size or adding or removing certain layers.

3.3.3 Dataset Description and Preprocessing

Four publicly available standard datasets are adopted to evaluate the performance of proposed methodology. The datasets include Massachusetts Institute of Technology-

Beth Israel Hospital Arrhythmia Database (MITARDB) [32], MIT-BIH Supraventricular Arrhythmia Database (MITSVDB) [33], St. Petersburg INCART 12-lead Arrhythmia Database [34], and China Physiological Signal Challenge (CPSC) 2020 [35] data. The data is chosen such that it covers multiple demographics and is diverse enough to account for any real-world scenario. The dataset includes normal and clinically significant arrhythmic beats. The modified limb lead II signal from each record is used for this work. The annotations provided in the dataset categorize the beats into 15 different classes (f, /, N, L, e, R, j, Q, A, J, a, V, S, E, F), each denoting a different type of beat. Table 3.2 describe the major and minor division provided by AAMI [7] into N, SVEB, VEB, Fusion beats (F), and Unknown beats (Q). This chapter focuses on N, SVEB, and VEB class beats.

Table 3.2: AAMI [7] recommended beat labels and classes

Group	Symbol	Class
Any heartbeat not categorized as SVEB or VEB. Normal Beat (N)	N	Normal Beat
	L	Left Bundle Branch Block Beat
	R	Right bundle branch block beat
	e	Atrial escape beat
	j	Nodal (junction) escape beat
Supraventricular ectopic beat (SVEB)	A	Atrial premature beat
	a	Aberrated atrial premature beat
	J	Nodal (junction) premature beat
	S	Supraventricular premature beat
Ventricular ectopic beat (VEB)	V	Premature ventricular contraction
	E	Ventricular escape beat
Fusion beat (F)	F	Fusion of ventricular and normal beat
Unknown Beat (Q)	P	Paced beat
	f	Fusion of paced and normal beat
	U	Unclassified beat

Preprocessing: The records are decluttered from high and low frequency noises. The Power Line Interference (PLI), i.e., high-frequency noise (50 or 60 Hz) is removed using a notch filter, and Baseline Wander (BW), i.e., low-frequency noise, is removed using a mean median filter [227, 228]. The beats are extracted from the clean signal using the annotations provided in the dataset from a window of 234 ms before and 486 ms after the R-peak. Since the datasets are sampled at different sampling frequencies, the extracted beats are resampled to the lowest sampling rate of 257 Hz or 186 samples. The beats are represented as $hb(i) = \{v_1, ..., v_{186}\}$, where v_t represent normalised value at time t.

The extracted beats are split into the train (DS1) and test (DS2) sets. DS1 and DS2 distribution for MIT ARDB [229] and MIT SVDB [230] datasets are shown in Table 3.3. CPSC and INCART data were split in half, and the first half (DS1) was used for training and the other half (DS2) for testing as no split strategy was provided in the literature. DS1 is used for training, and DS2 for testing/inferencing.

Table 3.3: Data distribution for MIT ARDB and MIT SVDB.

	MIT ARDB recordings
DS1	101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, and 230
DS2	100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, and 234
	MITSVDB recordings
DS1	802, 804, 805, 808, 810, 812, 841, 843, 845, 849, 866, 871, 873, 876, 877, 886, and 890
DS2	800, 803, 806, 807, 809, 811, 840, 842, 844, 846, 850, 867, 872, 874, 875, 878, and 879

3.3.4 Evaluation Metrics

DCCGAN performance is measured by the quality and diversity of the generated signals obtained from the generator model. Beats generated from the generator and randomly sampled from the real dataset are evaluated on five metrics, namely, Frechet Distance (FD), Dynamic Time Warping (DTW), Maximum Mean Discrepancy (MMD), Root Mean Square Error (RMSE), and Time Warp Edit Distance (TWED). A lower score of these metrics depicts similarity in real data distribution (P_r) and generated data distribution (P_g). Details of each evaluation metric are provided as follows.

Frechet distance [231] is adopted to measure the difference between real data distribution (P_r) and generated data distribution (P_g). Assuming the distributions to be multivariate gaussian distributions with mean (μ_r) and covariance matrix (σ_r) for real data distribution and mean (μ_g) and covariance matrix (σ_g) for generated data distribution, the distance can be calculated using Equation 3.17. Tr refers to the sum of elements along the diagonal of square matrix.

$$d^2((\mu_r, \sigma_r), (\mu_g, \sigma_g)) = ||\mu_r - \mu_g||^2 + Tr(\sigma_r + \sigma_g - 2 \times \sqrt{\sigma_r \times \sigma_g}) \tag{3.17}$$

Maximum Mean Discrepancy [232] tests the similarity between the distri-

bution P_r and P_g. n samples are independently and identically drawn from the two distributions. Smaller MMD corresponds to high similarity of the two distributions. Gaussian MMD is adopted that is calculated using Equation 3.18. Here, x and y denotes samples drawn from P_r and P_g respectively, $\kappa(x, x') = \sum_{j=1}^{k} e^{-\alpha_j \|x-x'\|^2}$ with bandwidth α denotes pairwise distance between the joint data.

$$MMD^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \kappa(x_i, x_j) - \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \kappa(x_i, y_j) + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \kappa(y_i, y_j) \tag{3.18}$$

Root Mean Square Error quantifies the distance between generated signal x and real signal y. It can be calculated using Equation 3.19. Here, n represents the number of samples in each signal (beats from both distributions are of equal length).

$$RMSE(x, y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2 \ / \ n} \tag{3.19}$$

Dynamic Time Warping Since the generated heartbeats might suffer from the condition of bradycardia or tachycardia, meaning slow or fast pace that regular rhythms, DTW is introduced to estimate dissimilarity between the generated and real heartbeats [233]. DTW aligns the two beats and calculates the distance between generated (x) and real (y) heartbeats with length (N). An optimised version of DTW, namely, Fast DTW, approximates DTW and reduces the computational complexity to $O(N)$ time [234]. Fast DTW is calculated using Equation 3.20. Here, $f(x_i, y_j) = (x_i - y_j)^2$.

$$D_{i,j} = f(x_i, y_j) + min\{D_{i,j-1}, D_{i-1,j}, D_{i-1,j-1}\} \tag{3.20}$$

Time Warp Edit Distance It measures the minimum cost sequence of "edit operations" required to transform the generated beat to the real beat. Edit operations are defined using a graphical editing process and produce a dynamic programming recursive algorithm called TWED [235]. It highlights a parameter that drives the stiffness (ν) of the elastic measure. TWED between real beat x and generated beat y of length N is calculated using Equation 3.21. Here, λ represents a mismatch penalty.

$$g(x_i, y_j) = \min\{g(x_i, y_j) + \delta(x_i, y_j), g(x_{i-1}, y_j) + \ \delta(x_i), g(x_i, y_{j-1}) + \delta(y_j)\} \tag{3.21}$$

$$
\begin{aligned}
Here,\ g(x_0, y_0) &= 0 \\
\delta(x_i, y_j) &= cost(x_i, y_j) + cost(x_{i-1}, y_{j-1}) + 2\nu.|i-j| \\
\delta(x_i) &= cost(x_i, x_{i-1}) + \lambda + \nu \\
\delta(y_j) &= cost(y_j, y_{j-1}) + \lambda + \nu \\
g(x_i, y_0) &= g(x_0, y_j) = \infty \\
cost(x_i, y_j) &= |x_i - y_j|^2
\end{aligned}
$$

The dissimilarity value $g(x_N, y_N)$ between P_r and P_g is defined as the TWED metric. Twenty four possible constant combinations for $\lambda = [0, 0.2, 0.4, 0.6, 0.8, 1]$ and $\nu = [0.001, 0.01, 0.1, 1]$ were computed to obtain a series of TWED values. Minimum of these values is used as the final value.

3.3.5 Results and Discussion

DCCGAN performance is evaluated both quantitatively and qualitatively. The real beats of N, SVEB, and VEB classes are randomly sampled from the training data, where equal beats of each class are extracted so that model has enough information about each class of beats. The fake beats are generated from the generator using the Gaussian noise as latent input and class labels. An equal number of fake beats of each class are generated so that discriminator does not get biased during the training. The DCCGAN is trained for 800 batches, and the generator loss, discriminator loss, and evaluation metrics for all classes are recorded. The models are saved after every second epoch during training as the training is unstable, and any intermediate model could generate the best quality of beats. Figure 3.20 depicts the generator and discriminator loss during DCCGAN training. It can be observed that multiple local minima are obtained by generator loss at batches 255, 525, 649, 663, and 757. The beats generated by the generator model at these batches depict a perfect resemblance with the real beats present in the training dataset. So, a detailed analysis of the generator models saved at these particular batches is performed.

Quantitative Evaluation: Five evaluation metrics are monitored during the training of the DCCGAN model, as mentioned above. The TWED metric is computationally expensive, and therefore, only 200 real and generated samples of each class are compared against each other after every two batches. The average of each evaluation metric is provided in Figure 3.21. The individual evaluation metric for Normal, SVEB, and VEB class is provided in Figure 3.22, 3.23, and 3.24. Overall and class respective metrics display similar pattern to training loss curves.

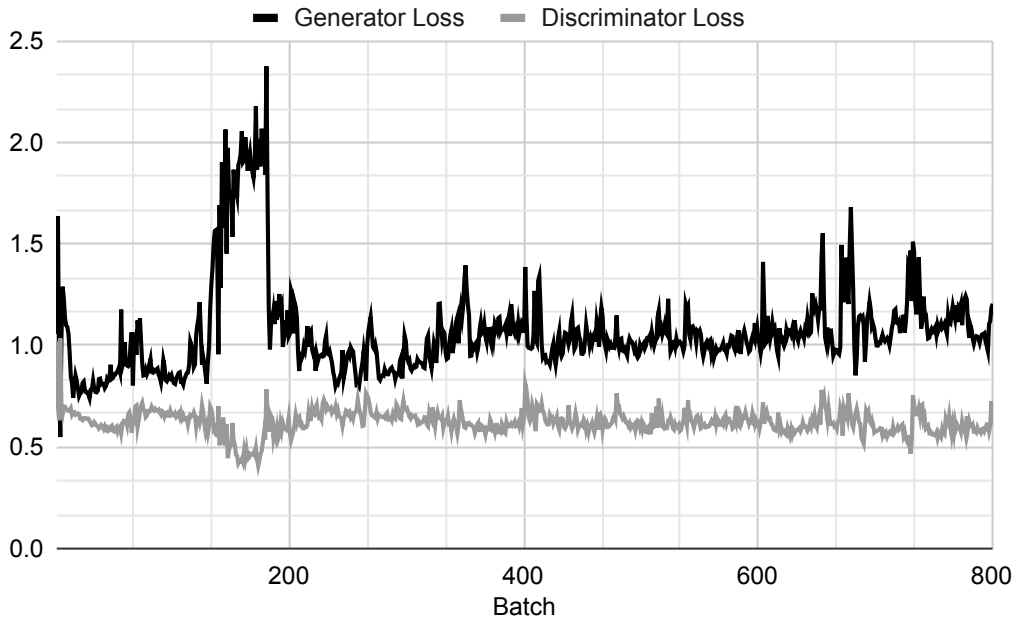


Figure 3.20: Generator and Discriminator Loss during training.

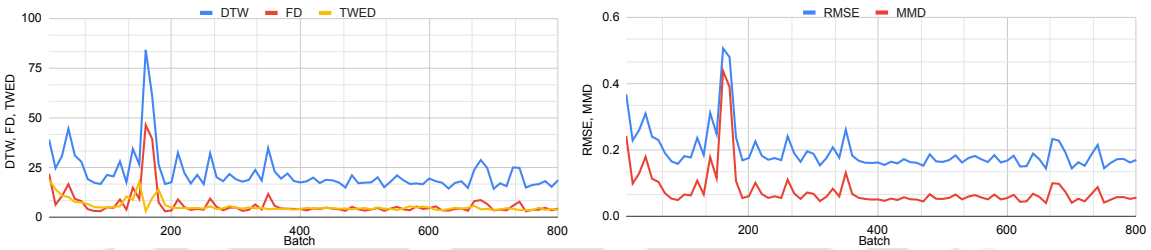


Figure 3.21: Overall Evaluation Metrics during DCCGAN training.

The normal class evaluation metrics in Figure 3.22 depicted a larger variation than the irregular beat evaluation metrics. The reason behind this might be the random undersampling performed for the normal class during data preparation. The Normal class beats were reduced as they were around 80% of the total beats, whereas SVEB and VEB contributed nearly 20% of total beats. The reduction in Normal beats might have reduced the variation present in the patient beats, thereby increasing the error in evaluation metrics.

SVEB and VEB class evaluation metrics in Figure 3.23 and Figure 3.24 depicted comparatively less error than Normal beats. The reason might not be because the model produced a good variation in irregular beats but because SVEB and VEB might have less variation in irregularities in the actual training dataset. Therefore, the generated beats could easily adapt to the variation present in SVEB and VEB class beats.

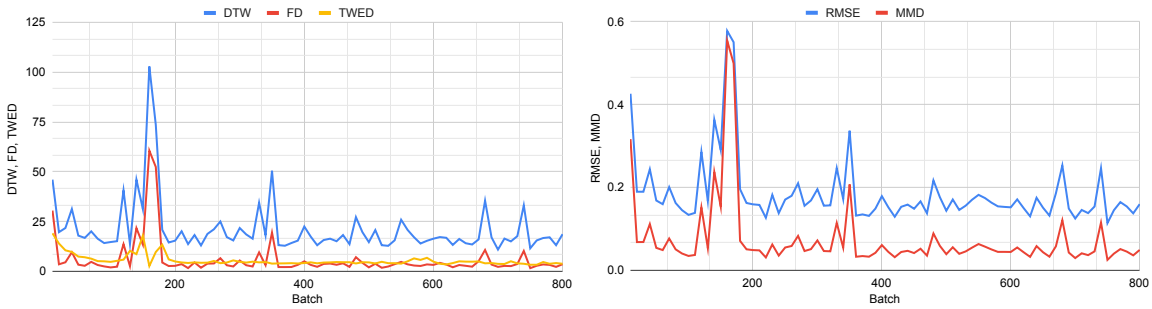


Figure 3.22: Evaluation Metrics for Normal Class during DCCGAN training.

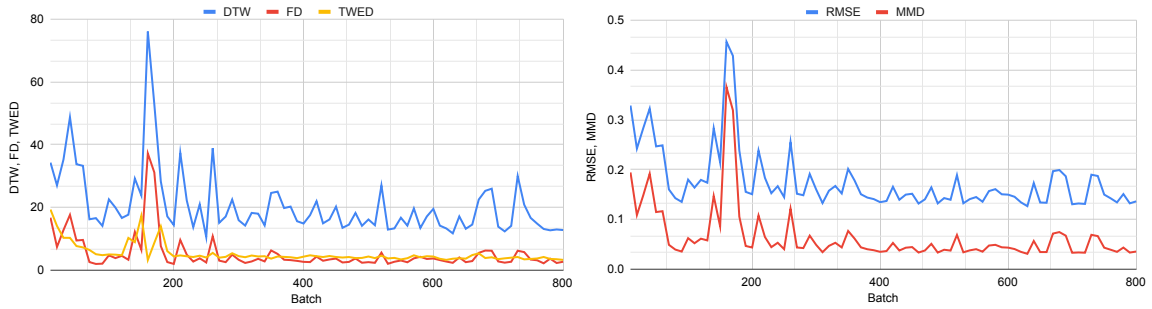


Figure 3.23: Evaluation Metrics for SVEB class during DCCGAN training.

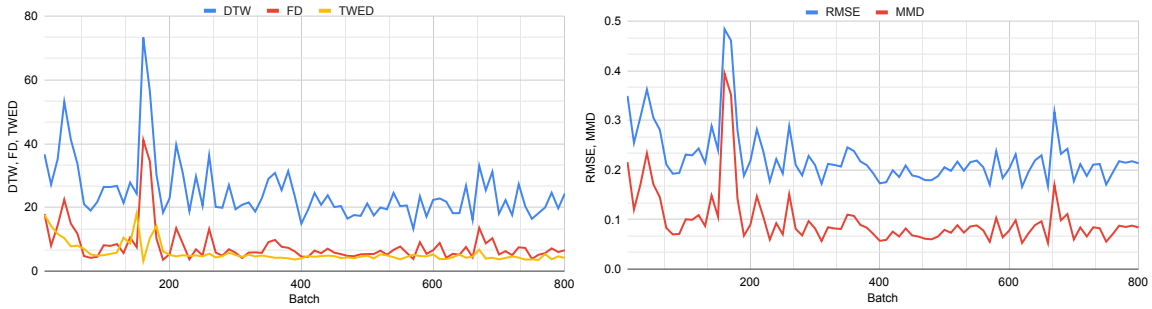


Figure 3.24: Evaluation Metrics for VEB class during DCCGAN training.

Similarly, less error in evaluation metric and less generator loss are observed in Figure 3.20, 3.21, 3.22, 3.23, and 3.24 for batches 255, 525, 649, 663, and 757, respectively. Therefore, the models saved at these batches are used to generate data for heartbeat classification. The model with lowest error in evaluation metrics is used to depict the generated beat quality.

Qualitative Evaluation: Out of the selected batches, the least error in evaluation metrics is obtained for batch 649. The model saved at batch 649 is used to generate beats of Normal, SVEB, and VEB class. Figure 3.25 illustrates original beats from DS1 and GAN Augmented beats from generator model saved at batch 649. Figure 3.25 (a), (b), (c) describe normal, SVEB, and VEB class beats from

training set of DS1. Figure 3.25 (d)-(f) describe the generated Normal Beat, Figure 3.25 (g)-(i) describe the generated SVEB, and Figure 3.25 (j)-(l) describe the generated VEB. The high error in evaluation metrics of normal beats might be due to irregularity or slight problem in generation of R-wave of QRS complex, i.e., the ventricular depolarisation of the heart. The other characteristic waves, such as P-wave and T-wave, resemble the real normal beats. The generated SVEB depicts a clear absence of P-wave and narrow QRS complex resembling real SVEB. This might be the reason for less error in evaluation metrics of SVEB. The generated VEB depicts an abnormal QRS complex with prolonged duration and elevated ST segment with a dominant S wave. It also shows an inverted or retrograde P-wave. This might be the reason for less error in the evaluation metrics of VEB.

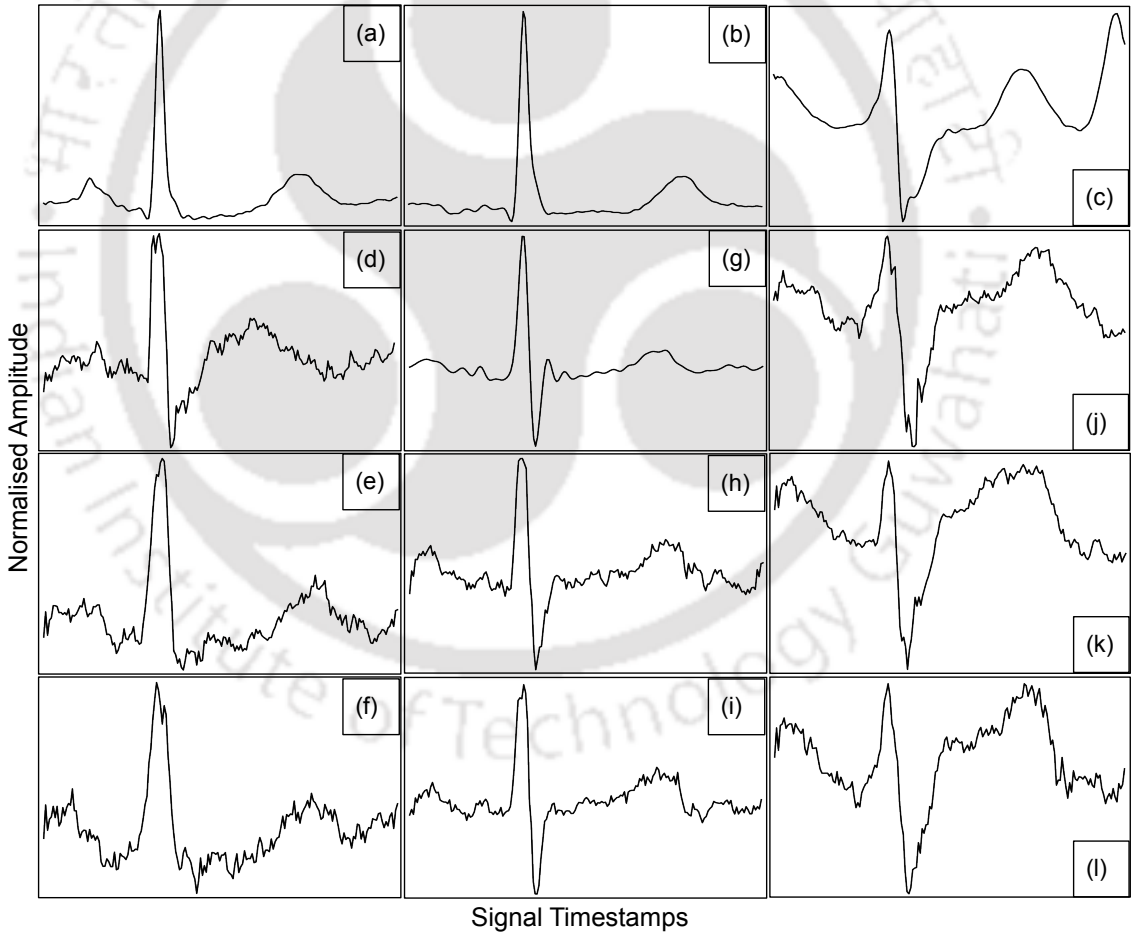


Figure 3.25: Illustration of Original beats from DS1 and GAN Augmented beats from generator model saved at batch 649. **(a)** Normal Beat, **(b)** SVEB, **(c)** VEB; **(d)-(f)** Generated Normal Beat, **(g)-(i)** Generated SVEB, **(j)-(l)** Generated VEB.

Figure 3.26 illustrates the incorrect beat generation by the generator model from

the models saved at different batches. Figure 3.26 (a-b) describe incorrectly generated normal beat , Figure 3.26 (c-d) describe incorrectly generated SVEB , Figure 3.26 (e-f) describe incorrectly generated VEB. The beats of all three classes do not resemble the actual beats of the respective class. Moreover, the generated beats are contaminated with high frequency noises.

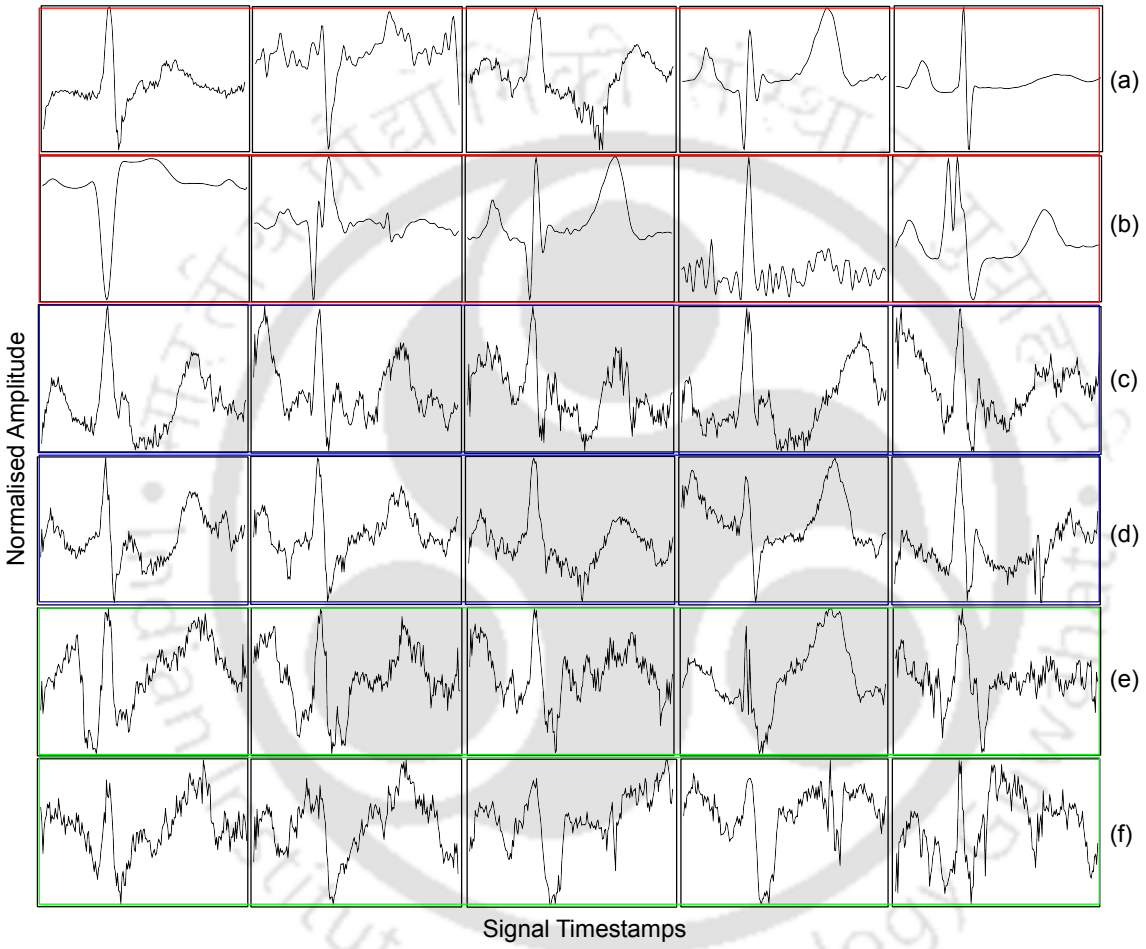


Figure 3.26: Illustration of Incorrect Beat Generation by the Generator Model. **(a-b)** Normal Beat, **(c-d)** SVEB, and **(e-f)** VEB.

The DCCGAN model generated different classes of ECG beats, including N, SVEB, and VEB recommended by AAMI. In DCCGAN, binary cross-entropy loss is preferred for discriminator and Gaussian noise as input for generator. Soft labels are preferred for faster convergence of the discriminator model. The results are presented both quantitatively and qualitatively. Quantitatively, the generated beats are evaluated using five quantitative metrics, and qualitatively, the beats generated by the generator model are plotted against the real/original beats to illustrate the sim-

ilarity in real and synthetically generated beats. The training curves of DCCGAN depicted stable training, and the change in evaluation metrics followed a similar pattern. Generated synthetic heartbeats resemble real beats as they encompass essential characteristics present in beats and follow the intricate structure present in the different classes of beats.

3.4 Summary

This chapter discusses preprocessing techniques involved in filtering the noisy ECG signals, segmentation to extract heartbeats, and heartbeat synthesis to improve minority class beat distribution. An empirical study is performed to remove BW from ECG signals using signal decomposition techniques such as EMD, EEMD, CEEMDAN, VMD and filtering techniques such as mean and mean median filtering techniques. The higher the complexity of the present baseline, the more execution time the algorithm took to decompose the signal. As the noise increased, the signal decomposition time also increased. CEEMDAN execution time exceeds by a huge margin as compared to other approaches. The comparison depicted that VMD estimates better baseline as compared to other techniques in terms of PC, PRMSD, and MAE. However, the time required to decompose the signal is relatively higher than the filtering techniques. The clean ECG signal is used to extract heartbeats using a fractal based approach that incorporates erosion and dilation, mathematical morphological operators which are implemented using dynamic programming with memoization. The method achieved an overall sensitivity of 95.82%, predictivity of 95.15%, and DER of 10.16% when tested against all the records available in the MIT-BIH dataset. Lastly, a DCCGAN model is employed to generate different classes of heartbeats, including N, SVEB, and VEB as recommended by AAMI. The generated heartbeats are verified quantitatively using five evaluation metrics and qualitatively by plotting against original beats. Generated heartbeats resemble real beats as they encompass essential characteristics and follow the intricate structure present in the different classes of original beats. The beats are provided to a classifier for heartbeat classification.

4

Penalty Induced Prototype-Based eXplainable ResNet for Heartbeat Classification

An abnormality in heart rhythm or deviation of electrical impulses from normal sequence leads to irregular heartbeats or cardiac arrhythmias. The heartbeats considered for this work are non life threatening heartbeats such as normal beat (N), supraventricular ectopic beat (SVEB), and ventricular ectopic beat (VEB) as these are the major categories of beats recommended by the Association for Advancement of Medical Instrumentation (AAMI) [7]. This chapter focuses on classifier design and explanation generation aspect of CAD as it highly influences the classifier performance [17]. To achieve state-of-the-art performance, the developed models require huge amount of labeled, diverse, and realistic-looking heartbeats. Therefore, heartbeat synthesis of regular and irregular heartbeats is performed as described in Section 3.3. The beats are classified by deep learning models [16]. Although deep learning models achieve good performance for heartbeat classification, their black box nature inhibits real world deployment [236]. A Penalty Induced Prototype based eXplainable Residual Neural Network (PIPxResNet) is proposed for heartbeat classification that provides explanations along with improved diagnostic performance. PIPxResNet is adopted from explainable deep neural network [31] due to its non-iterative and non-parametric nature that eliminates hyperparameter tuning. PIPxResNet employs pretrained residual neural network that extracts features from decluttered heartbeats and minimizes resource consumption (eliminating the need for GPU, reduced training and inference time) following the concept of task transfer learning. The features similar to other class features are penalised and their contribution towards the corresponding class is reduced. The selected encoded features represent prototypes. The

prototypes are combined to make rules that explain model predictions and enhance trust in predictions. In addition, the encoded and actual prototypes are utilised to generate a reliability score that depends on nearest prototype support and correlation between the actual prototype and heartbeat. PIPxResNet is verified on four publicly available standard databases.

The chapter organization is as follows. Section 4.1 provides the literature review of the previous heartbeat classification methods. Section 4.2 describes the proposed PIPxResNet technique. Section 4.3 describes experimental setup and results. Section 4.4 summarizes the chapter.

4.1 Literature Review

Heartbeat classification has been performed using a combination of feature extraction, feature selection, and classification methods. Several time domain, frequency domain, and morphological features have been proposed in the literature. The most commonly used features include RR interval (RRI), defined as the interval between two adjacent R-peaks, that achieves the best performance [229]. Other features include, Morphological features such as position, amplitude, and interval features [237], wavelet based features [17, 238–244], discrete cosine transform [245], Hilbert transform [246], Kalman filter-based features [247], stockwell transform (ST) [248], higher order statistical (HOS) features [249], Hermite polynomial based features (HBF) [250], mixture of features [251] and complex heartbeat representations such as mean connectivity degree, maximum connectivity degree, joint entropy, joint energy, mean joint degree extracted from temporal vectorcardiogram (TVCG) [252]. Redundant features are removed through dimensionality reduction techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA) [243]. Feature selection has been performed through particle swarm optimization (PSO) [252, 253], genetic algorithm (GA) [254], and bacteria foraging optimisation (BFO) [248] have been employed to reduce feature redundancy.

The classification of extracted features has been performed using weighted conditional random fields (CRF) [237], support vector machines (SVM) [93, 96, 245, 248, 252, 255], mixture-of-experts [256, 257], and K-nearest neighbors [240]. Clustering methods [250, 258] such as maximum margin clustering with immune evolution [259], fuzzy-entropy-based clustering [260], K-means clustering [260] have also been employed for classification. The handcrafted features sometimes fail to capture essential information present in data and limit the learning capability of machine learning

models and reduce the performance.

Therefore, end to end Deep Neural Networks have been applied that have achieved exceptional performance for heartbeat classification. These include recurrent neural networks (RNN) [239, 261] based long short term memory (LSTM) [239], fuzzy neural networks [241], convolution neural networks (CNN) [262–265], multiscale wavelet CNN [244], feedforward neural network (FFNN) [245], probabilistic neural network (PNN) [245], multilayer perceptron (MLP) [251], modular neural networks [256], hybrid neural network [266], deep belief networks [267], artificial neural network (ANN) [93], learning vector quantization NN [268], and radial basis function network [242, 269].

The developed classifiers lack explainability, leading to reduced confidence in model diagnosis. The intrinsic interpretation methods create interpretable models and explain the prediction using data points from training dataset. The selected data points or prototypes represent data in a condensed format that encompasses basic characteristics of the respective class [31]. Several such methods have been developed in the past including prototype with autoencoders [265], deep sequence model consisting of sequence encoder and prototype layer for classification and reasoning [270], and explainable deep neural network (xDNN) [31].

The methods developed in the past have achieved excellent performance but the limitations are: (i) Only a few methods follow the ANSI/AAMI EC57:2008 [7] recommended beat classes; (ii) Extraction and selection of handcrafted features; (iii) existing models perform inconsistently when tested on other datasets as their testing was limited to a single or couple of databases; (iv) the black box nature of neural networks limits real-world deployment. The aforementioned issues are addressed by performing heartbeat classification and prediction explanation through PIPxResNet.

4.2 Penalty Induced Prototype-Based eXplainable Residual Neural Network

The proposed approach is adopted from explainable deep neural network (xDNN) [31]. xDNN [31] is a non parametric approach that uses recursive search procedures requiring no hyperparameter tuning, eliminating human intervention. It addresses Deep Learning bottlenecks by offering an explainable architecture with minimal computational resources. The xDNN algorithm extracts prototypes for each class that efficiently represent training data in a condensed format. xDNN architecture encom-

passes the following layers: Feature, Density, Typicality, Prototype, and MegaClouds layer. The Feature layer extracts features from data using either conventional feature extraction methods or pretrained NN. The Density layer uses a distance metric for calculating mutual proximity of data instances using preceding layer features. Instances closer to global mean persist higher density values thereby indicating strong influence with neighboring samples. The Typicality layer empirically derives probability distribution function of density layer. The core of xDNN classifier is the Prototype layer which is responsible to provide an explainable model using prototypes, representing local peaks of data density. The prototypes are combined together using logical OR to form linguistic logical *if-then* rules. Lastly, the MegaClouds layer merges the neighbouring prototypes having the same class label facilitating human interpretability.

The Penalty Induced Prototype-based eXplainable Residual Neural Network (PIPxResNet) inspired from xDNN is proposed that addresses the black-box nature of deep neural networks. PIPxResNet encodes the temporal variations of heartbeats by employing pretrained residual neural network following the concept of task transfer learning. The algorithm extracts prototypes that are most representative of the training dataset that explain model predictions to general physicians, making them clinically relevant. The prototypes of a particular class having a close resemblance to other class prototypes are penalized and their contribution towards the corresponding class is reduced. PIPxResNet incorporates the information of other class prototypes in addition to corresponding class prototypes whereas xDNN [31] uses only single class data during prototype extraction. PIPxResNet follows these conditions: (i) Beats and prototypes of class C should be present in close vicinity; and (ii) Other class prototypes should be farther away from corresponding class prototypes. PIPxResNet initializes the following for each class: encoded prototypes (P_e), actual prototypes (P_A), mean (μ), variance (σ), number of prototypes (N) associated with that class, minimum density (ρ_{min}), maximum density (ρ_{max}) of that class, and number of beats contributing towards the corresponding prototype (S).

Algorithm 6 extracts the prototypes using the original heartbeat (X), 3-D encoded beat (X_e) as input and produces actual prototypes (P), encoded prototypes (P_e), and support (S) as output. The encoded beats are extracted from the pretrained neural network using the concept of transfer learning. The first sample of each class initialises the meta-parameters by creating the first prototype from *line* 2 to 5. In *line* 6 the `Density` function is calculated using Equation 4.1 that determines mutual proximity between the beat, μ, and σ of the corresponding class, similar to xDNN [31]. Heartbeats closer to the mean have higher density values. Therefore, the

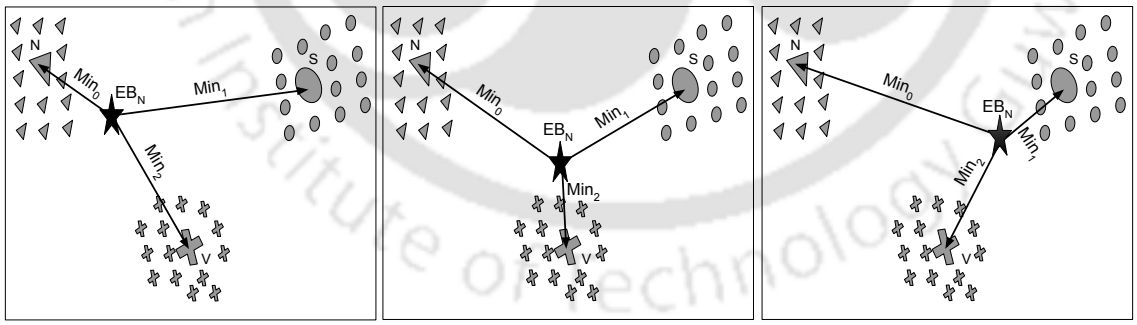
value of the data density indicates how strongly a particular heartbeat is influenced by other principal beats in data space due to their mutual proximity. The density function utilises an Euclidean distance-based density function to find similarity between existing prototypes and the encoded beat.

Algorithm 6 Prototype Extraction For Each Class

Input: Original Beats (X) and Encoded Beats (X_e)
Output: Actual Prototypes(P), Encoded Prototypes(P_e), Support(S)
for $i := 0$ *to* *Number of Original Beats* **do**
  if $i = 0$ **then**
    // CREATE FIRST PROTOTYPE
    $P_e \leftarrow X_e[i]$ $P \leftarrow X[i]$
    $\mu \leftarrow X_e[i]$ $\sigma \leftarrow \| X_e[i] \|^2$
    $N + 1$ $S \leftarrow 1$
    $\rho_{min} = \infty$ $\rho_{max} = -\infty$
    $d = \texttt{Density}(X_e[i], \mu, \sigma)$
  else
    $Penalty = \texttt{Penalise Beat}(P, X_e[i])$
    $E = Penalty \times X_e[i]$
    $O = Penalty \times X[i]$
    $d = \texttt{Density}(E, \mu, \sigma)$
    if $d \geq \rho_{max}$ *or* $d \leq \rho_{min}$ **then**
      // CREATE A NEW PROTOTYPE
      $P_e \leftarrow E$ $P \leftarrow O$
      $S \leftarrow 1$ $N + 1$
    else
      // UPDATE THE NEAREST PROTOTYPE
      $j^* = \texttt{Nearest\_Prot\_Index}(E, P_e)$
      $P_e[j^*] = \texttt{Update\_Prot}(S, j^*, P_e, E)$
      $P[j^*] = \texttt{Update\_Prot}(S, j^*, P, O)$
      $S[j^*] = S[j^*] + 1$
      $d = \texttt{Density}(P_e[j^*], \mu, \sigma)$
    end
    $\mu \leftarrow \frac{i \times \mu + E}{i+1}$
    $\sigma \leftarrow \frac{i \times \sigma + \|E\|^2}{i+1}$
  end
  $\rho_{min} = \texttt{Minimum}(\rho_{min}, d)$
  $\rho_{max} = \texttt{Maximum}(\rho_{max}, d)$
end

$$D = \frac{1}{1+ \parallel X_e - \mu \parallel^2 - \sigma + \parallel \mu \parallel^2} \tag{4.1}$$

The subsequent beats either create a new prototype or update an existing prototype. In both cases, the density of the encoded beat (EB) is compared with the minimum and maximum density of that class. Algorithm 7 uses all class prototypes (P) and EB (X_e) as input and calculates $Penalty$ by penalising encoded and actual beats that are near to other class prototypes and distant from their respective class prototypes. Here, P_0 represents the class to which EB belongs, whereas P_1 and P_2 denotes other class prototypes. Min_0 represent the distance between EB and nearest prototype of corresponding class, Min_1 and Min_2 denotes the distance between the EB and nearest prototype of other classes. The distance `MinDist` is calculated using Equation 4.2. Assuming that each class has only one prototype with several beats supporting the prototype, three possible scenarios are illustrated in Figure 4.1, where, the triangle represents the class in which EB belongs, and circle and plus represent the other class prototypes. In scenario 1, X_e lies closer to its own class prototypes (triangles) making Min_0 less than Min_1 and Min_2, using the beat without penalty. In scenario 2, X_e lies closer to second class prototype (plus) making Min_2 less than Min_0 and Min_1, penalising the beat with $Min_2/(Min_0 + Min_2)$. In scenario 3, X_e lies closer to third class prototype (circle) making Min_1 less than Min_0 and Min_2, penalising the beat with $Min_1/(Min_0 + Min_1)$.



(a) EB close to actual class. (b) EB closer to second class. (c) EB closer to third class.

Figure 4.1: Illustration of possible penalty scenarios. Star: New normal class encoded heartbeat; Triangle: Existing normal class encoded heartbeats; Circle and Plus: Encoded heartbeats of other classes.

$$Min = \min_{j=1,2,\ldots,P} \parallel P_j - X_e \parallel^2 \tag{4.2}$$

After penalising, if data density of penalise beat lies between ρ_{min} and ρ_{max}, the

Algorithm 7 Penalise Beat

Input: All Class Prototypes (P) and Encoded Beat (X_e)
Output: Penalty
$Penalty = 1$
$P_0 \leftarrow P[0]$ $P_1 \leftarrow P[1]$ $P_2 \leftarrow P[2]$
$Min_1 = \infty$ $Min_2 = \infty$
$Min_0 = \texttt{MinDist}(P_0, X_e)$
`if` $P_1 \neq \phi$ `then` $Min_1 = \texttt{MinDist}(P_1, X_e)$
`if` $P_2 \neq \phi$ `then` $Min_2 = \texttt{MinDist}(P_2, X_e)$
if $Min_1 < Min_2$ **then**
  if $Min_1 < Min_0$ **then** $Penalty = \frac{Min_1}{Min_0 + Min_1}$
else
  if $Min_2 < Min_0$ **then** $Penalty = \frac{Min_2}{Min_0 + Min_2}$
end

existing prototype is updated otherwise a new prototype is created. For updating the nearest existing prototype, the nearest prototype location is determined using `Nearest_Prot_Index` provided in Equation 4.3 that resembles EB.

$$j^* \leftarrow \underset{k=1,2,...,len(P)}{\text{argmin}} \parallel E - P_k \parallel^2 \tag{4.3}$$

Variable j^* updates the number of beats supporting prototypes (S), encoded (P_e) and actual (P) prototype using `Update_Prot` given in Equation 4.4. The data density is updated as updating the prototype might have changed the data density of the class. After updating the prototypes, μ, σ of the class and data density range (ρ_{min} and ρ_{max}) are updated.

$$P[j^*] \leftarrow \frac{S[j^*] \times P[j^*] + X_e}{S[j^*] + 1} \tag{4.4}$$

Inferencing: The inferencing is performed using original test beats (X), encoded test beats (X_e), extracted actual prototypes (P), extracted encoded prototypes (P_e), and support (S) as inputs. Algorithm produces the following: (i) Model Diagnosis (Y_d); (ii) Prototype-based Explanation (Y_e); and (iii) Reliability score (Y_r) supporting model diagnosis. Algorithm 8 searches for the nearest prototype from extracted prototypes in a linear fashion. For each class, the density is calculated by `Test` function provided in Equation 4.5 using the EB, prototypes, μ, and σ. The prototype nearest to EB is provided as explanation (Y_e), corresponding prototype class as diagnosis (Y_d) and the `Reliability Score` (Y_r) is calculated by Equation 4.6 using actual beat (X), actual prototype (P^{jk}) and normalised support (S^{jk}) associated with corresponding

prototype. The testing is performed in $O(N \times K)$, where N and K represent number of encoded beats and all class prototypes, respectively.

Algorithm 8 Prototype Based Diagnosis, Explanation and Reliability

Input: X, X_e, P, P_e, and S
Output: Y_d, Y_e, and Y_r
$D = -\infty$
for $j := 0$ *to Number of Classes* **do**
  for $k := 0$ *to Corresponding Class Prototypes* **do**
    $T = \texttt{Test}(X_e, P_e^{jk}, \mu^j, \sigma^j)$
    if $T > D$ **then**
      $D \leftarrow T$
      $Y_d \leftarrow j$
      $Y_e \leftarrow P_e^{jk}$
      $Y_r \leftarrow \texttt{Reliability}(X, P^{jk}, S^{jk})$
    end
  end
end

$$T = \frac{1}{1 + \parallel X - P \parallel^2 - \sigma + \parallel \mu \parallel^2} \tag{4.5}$$

$$Y_r = S^{jk} + \frac{\sum (X \cdot P^{jk})}{\sqrt{\parallel X \parallel^2 \cdot \parallel P^{jk} \parallel^2}} \tag{4.6}$$

4.3 Experimental Setup and Results

The framework for heartbeat classification and explanation is described in Figure 4.2. The single lead ECG signal is preprocessed to remove noise. The obtained clean data is split into train and test sets followed by augmentation of training data using DCCGAN. The synthesised heartbeats are used to train neural network models which are later used for feature extraction from heartbeats. The prototypes are extracted through PIPxResNet from the encodings generated from pretrained models. The extracted prototypes are later used for inferencing and reasoning on test data.

4.3.1 Data Description and Preprocessing

The PIPxResNet is evaluated on Massachusetts Institute of Technology Beth Israel Hospital (MIT-BIH) Arrhythmia Database (ARDB) [32], MIT-BIH Supraventricular Arrhythmia Database (SVDB) [33], St. Petersburg INCART twelve lead Arrhythmia

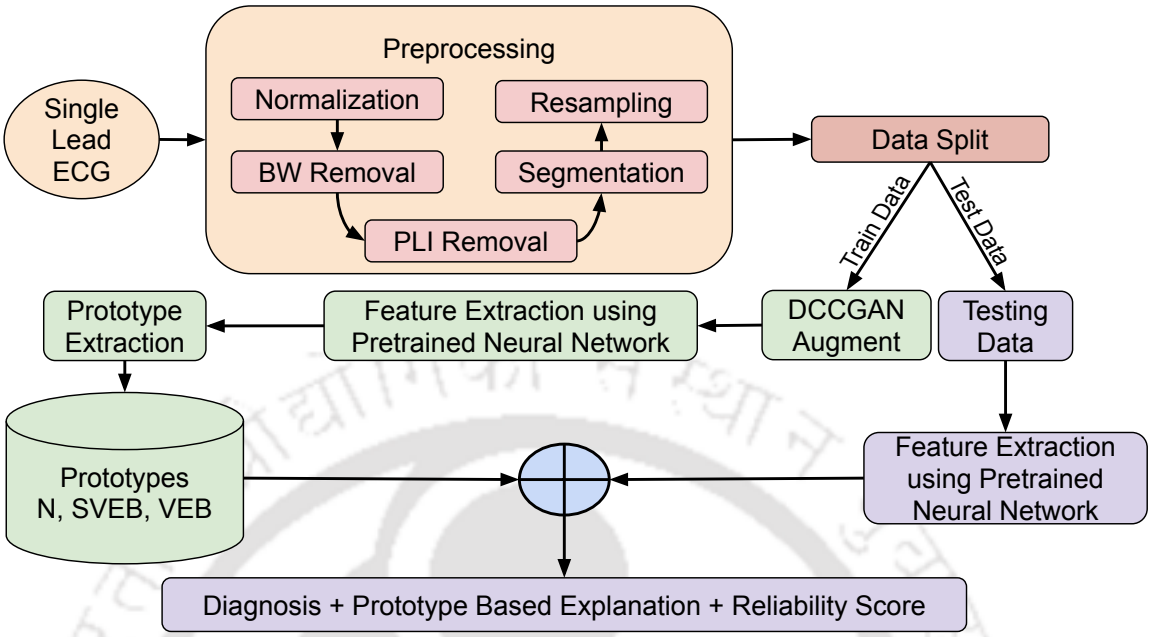


Figure 4.2: Heartbeat classification and explanation.

Database [34] available on PhysioNet website [34] and China Physiological Signal Challenge (CPSC) 2020 [35]. The data covers multiple demographics and its diversity accounts for real world scenarios. The detailed description and preprocessing is provided in Section 3.3.3 which includes Power Line Interference, Baseline Wander removal followed by N, SVEB, VEB beat extraction using the annotations provided in the database and resampled to the lowest sampling rate of 257 Hz or 186 samples. The beats were split into training and testing datasets. Distribution strategy of MIT-BIH ARDB [229] and MIT-BIH SVDB [230] was provided in literature whereas CPSC and INCART data were randomly split as splitting strategy was unavailable in the literature. The Deep convolution conditional generative adversarial network (DCCGAN) augmentation is described in Section 3.3.

4.3.2 Evaluation Metrics

The classification performance is evaluated quantitatively using Accuracy (Acc), Sensitivity (Se), Specificity (Sp), Precision (Pr), and F1 Score (F1) as described in Equation 4.7. Here, TP, FN, TN, and FP represent True Positive, False Negative, True Negative, and False Positive. Sensitivity measures the fraction of SVEB, VEB beats correctly classified as SVEB, VEB. Specificity measures the fraction of N beats that are correctly identified as N. Accuracy corresponds to the fraction of rhythms that are correctly classified, irrespective of the beats. Qualitatively, the method is validated

through prototype-based explanation and reliability scores (RS) using Equation 4.6.

$$\begin{aligned} Accuracy\ (Acc) &= \frac{TP+TN}{TP+FP+FN+TN} \\ Sensitivity\ (Se) &= \frac{TP}{TP+FN} \\ Specificity\ (Sp) &= \frac{TN}{TN+FP} \\ Precision\ (Pr) &= \frac{TP}{TP+FP} \\ F1Score\ (F1) &= 2 \times \frac{Pr \cdot Se}{Pr+Se} \end{aligned} \tag{4.7}$$

4.3.3 Feature extraction using Neural Networks

Convolution neural networks have produced benchmark results in domains like Image classification [271] and Sentence classification [272]. The basic idea of CNN was proposed long back in 1980s by Kunihiko Fukushima [273]. CNNs have also produced benchmark results for ECG classification [244, 263]. CNNs have the ability to process multidimensional data and are immutable to local spectral and spatial variations which makes them translational and shift invariant. The model encompasses four 1-D convolution layers, each followed by batch normalization (BN) [274] and Maxpooling [275] layer. The description of the individual layers is as follows.

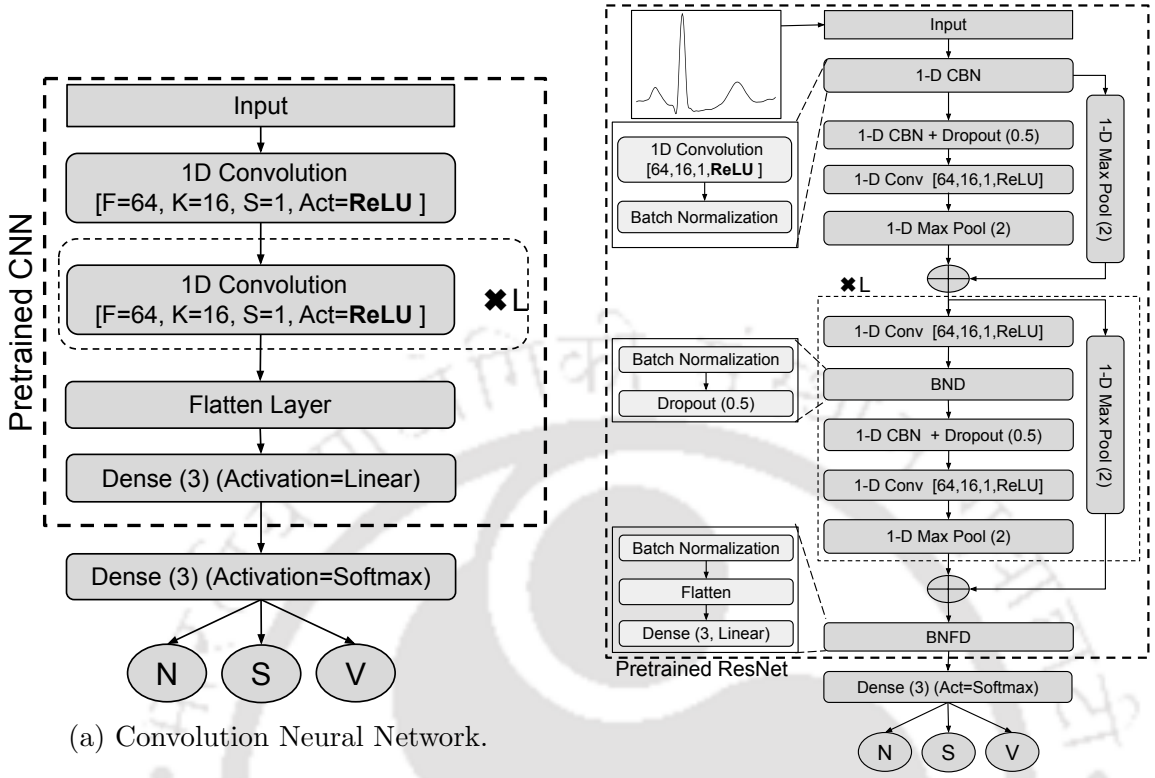
1. Input: Takes raw signal or handcrafted features as input.
2. 1-Dimensional Convolution: Performs computations over the input and filters. It works as both feature extractor and selector. The convolution layer has the following main parameters: Number of filters (F), Kernel Size (K), and Stride (S).
3. Batch Normalization (BN) [274]: A regularizer that accounts for the co-variance shift in the incoming inputs from the previous layer.
4. Dropout [276]: A regularizer that shuts off neurons/filters of a particular layer with probability p.
5. Max-pooling [275]: A sub-sampling layer that reduces the number of parameters and computational burden of the NN thereby reducing overfitting. It tries to capture only the most salient elements of the previous layer. It has a parameter (P) that pools or calculates the maximum of P values from the kernel.
6. Flatten: Reshapes the values of the previous layer.
7. Fully-connected (FC): Connects all the neurons of the previous layer with the current layer.

Increasing the size and number of layers does not always optimize the neural network, it instead deteriorates its performance due to overfitting and suffers from problems such as the notorious vanishing and saturation of the gradient. To tackle this, Residual Neural Networks (ResNet) was invented that introduced the concept of "identity shortcut connection", that controls the amount of information allowed to flow across the shortcut [277]. It solves the problem of vanishing and exploding gradients, by reusing activations from the previous layers. The shortcut connection simplifies the network by using a few layers and less backpropagation thereby reducing the model size and execution time. Without these identity mapping, the model explores more feature space and leaves the original hypothesis and leads to performance deterioration and necessitates the use of extra training data.

CNN (C) and ResNet (R) are used for feature extraction. Figure 4.3a and Figure 4.3b illustrates the architecture of CNN and ResNet. The models were trained with L layers, where $L \in [1, 9]$. Here, a ResNet with 5 layers is represented as R5. The layers are repeated depending on the depth of CNN and ResNet. The trained model is transferred from first layer upto the penultimate layer following the transfer learning concept.

4.3.4 Pretrained Neural Network Performance

The neural network models were trained for 50 epochs while monitoring the validation loss with early stopping criteria. The models were trained with following configurations: (1) Original Data (OD); (2) WWT Augmented Data (WWTOD); (3) ADASYN Augmented Data (ADAOD); (4) Models trained with class weights for normal, SVEB, and VEB class as 1.30, 5.85, and 5.63 using original data (CWOD). Weights depend on number of training instances for each class; DCCGAN synthesised data as described in Section 3.3 at batch (5) 255; (6) 525; (7) 649; and (8) 757. The performances of these 8 configurations on test data is provided in Figure 4.4. The specificity was always better than the accuracy. The F1 score for all models except CWOD was is in range $[0.8, 0.85]$ but in CWOD it deteriorated. The performance of conventional augmentation techniques were less as compared to the DCCGAN augmented data. The CNN models underperformed in comparison to ResNet models. Overall the model performance improved with increasing the layers of ResNet, but the same trend was not observed when the CNN layers were increased. The CNN performance initially improved but later got deteriorated. The comparison of best performing model is provided in Figure 4.5a. Following are the details of model with



(a) Convolution Neural Network.

(b) Residual Neural Network.

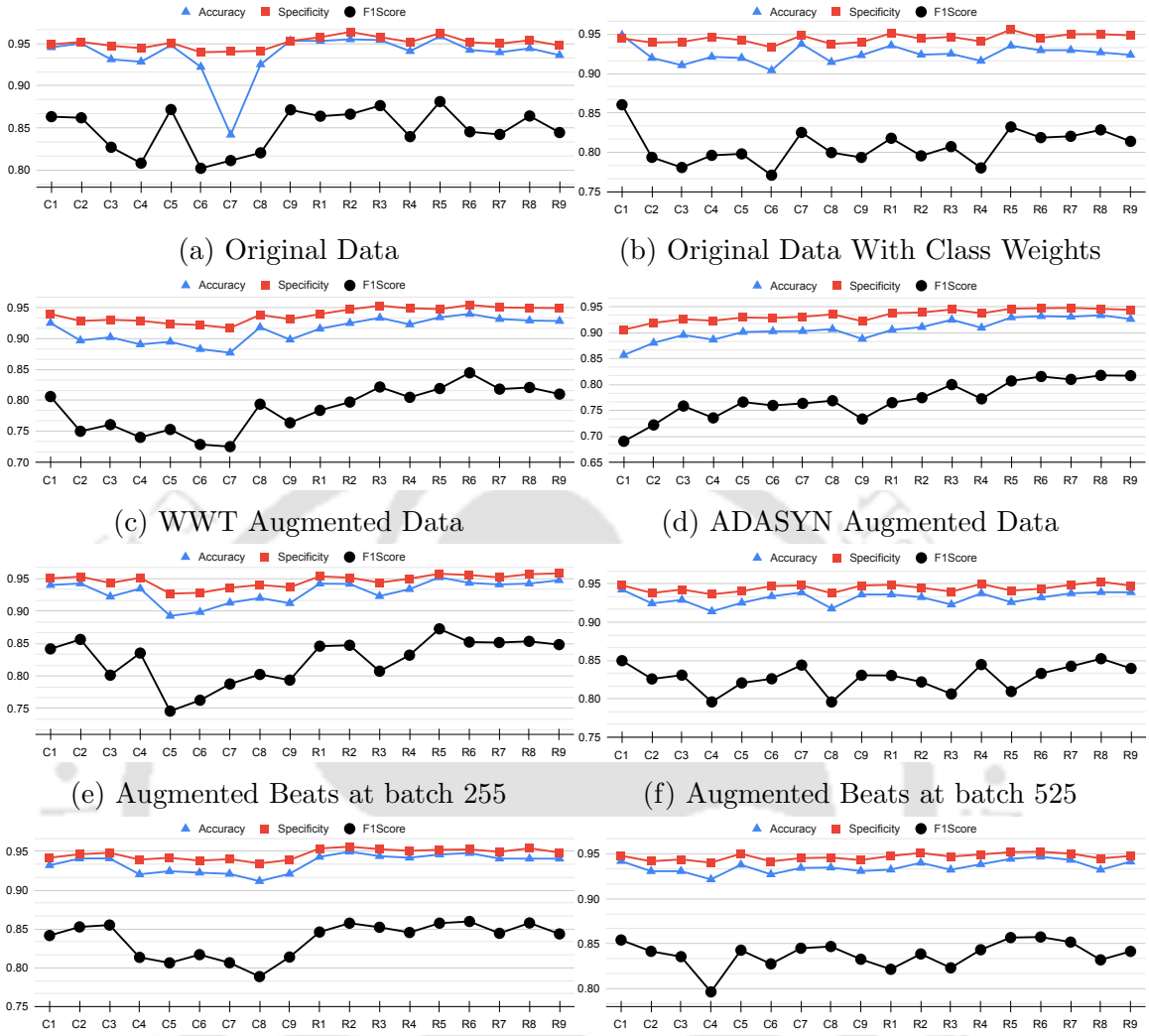
Figure 4.3: Neural Network Architecture.

respective data configuration: R5 on OD, R5 on CWOD, R6 on WWTOD, R3 on ADAOD, R5 on 255^{th} batch, R8 on 525^{th} batch, R6 on 649^{th} batch, R6 on 757^{th} batch of DCCGAN augmented data.

4.3.5 Evaluation of Prototype-Based Techniques

The prototype based techniques are evaluated in three stages: (i) Prototype extraction is performed on raw heartbeats; (ii) eXplainable ResNet; (iii) PIPxResNet.

Prototype Extraction from Heartbeats: The extraction of prototypes from raw heartbeats without using pretrained neural network as feature extractors deteriorated the performance. However, the explanations and reliability scores provided evidence for misclassified beats, thereby allowing better model assessment. Figure 4.5b provides the performance over original data and augmented data. Since, no neural network is employed, CWOD is skipped. The Pr and Se is consistently low for all the models. The training time was around 100 seconds for full training set and testing time (Te_Ti) was around 300 seconds for testing set. The time ($\approx$ seconds) is significantly less than training neural networks ($\approx$ day). The testing time



(a) Original Data

(b) Original Data With Class Weights

(c) WWT Augmented Data

(d) ADASYN Augmented Data

(e) Augmented Beats at batch 255

(f) Augmented Beats at batch 525

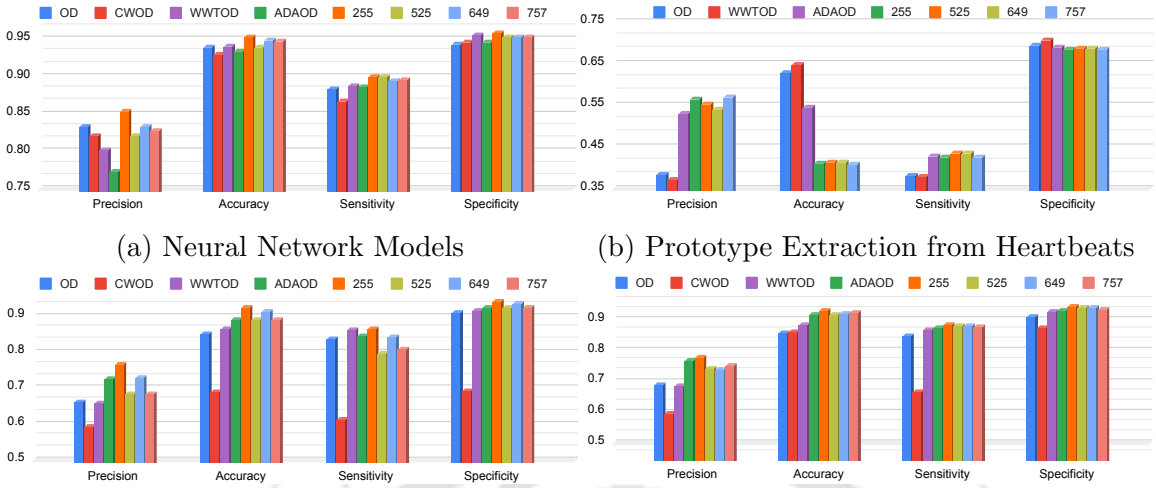
(g) Augmented Beats at batch 649

(h) Augmented Beats at batch 757

Figure 4.4: Neural Network Model Performance. The x-axis represent models and y-axis represent the performance achieved for corresponding metric.

is directly proportional to the number of prototypes extracted during training phase. Table 4.1 shows the number of extracted prototypes from each configuration without using pretrained neural networks. DCCGAN generated very less prototypes for all batches as compared to other configurations. The reason might be the high resemblance of synthesised beats which were not generated of so better quality with other augmentation techniques. Majority beats (N) generate more prototypes compared to minority beats (SVEB and VEB). For individual class performance, the method produces highest Pr, for Normal class beats, highest Se for SVEB, and highest Acc and Sp for VEB.

eXplainable ResNet (xResNet): Pretrained Neural Networks were used to



(a) Neural Network Models

(b) Prototype Extraction from Heartbeats

(c) eXplainable ResNet

(d) Proposed Method (PIPxResNet)

Figure 4.5: Best performing models from each configuration.

Table 4.1: Number of individual class prototypes extracted from heartbeats without using neural networks.

Class	DCCGAN	OD	WWTOD	ADAOD
N	14	66	66	66
S	12	28	17	30
V	10	23	33	25

extract features and ResNet [31] was used for heartbeat classification and explanation without penalising the beats. Figure 4.5c provides the performance of all configurations. Model trained using DCCGAN augmented beats on batch 255 performed better than ADAOD and WWTOD. Table 4.2 shows the number of extracted prototypes from each configuration using xResNet. The prototypes of OD, WWTOD, and ADAOD reduced as compared to the prototypes extracted from heartbeats without using neural networks as described in Table 4.1. The number of individual class prototypes follow the probability distribution of corresponding class beats for DCCGAN synthesised beats, i.e., N prototypes are highest followed by SVEB followed by VEB. Other configurations also produce highest number of N prototypes followed by irregular beat class. CWOD configuration produced many prototypes of Normal class which might be due to the less penalising of missclassified Normal class beats during network pretraining. The results improved significantly over prototype-only method and reduced the prototypes resulting in reduced testing time. The training and testing times reduced due to the reduction of the data dimension from 186 to 3. Moreover, whenever the data dimension reduces, the method tends to perform

better and faster. Model performed better on CPSC and INCART data with good explanation maybe because same patient beats (at different period) were present in the train and test data.

Table 4.2: Number of extracted prototype using xResNet

Class	255	525	649	757	OD	CWOD	WWTOD	ADAOD
N	21	15	16	14	8	45	15	17
S	10	13	15	14	10	9	8	12
V	8	11	9	12	16	16	11	9
All	39	39	40	40	34	70	34	38

PIPxResNet: A data driven penalty is imposed on encoded and actual beats if they are near to other class prototypes as compared to their corresponding class prototypes. Suppose, encoded N beat resembles with SVEB then it contributes less during prototype updation to reduce missclassification while inferencing. Table 4.3 shows the number of extracted prototypes from each configuration using the proposed PIPxResNet. The number of prototypes reduced for 649^{th} batch of DCCGAN, OD, CWOD, and WWTOD and increased in other configurations as compared to xResNet, increasing the performance. Introducing the penalty helps in extraction of better prototypes that are more representative of the training dataset rather than influencing the number of prototypes. Figure 4.5d provides the performance of all configurations and shows that imposing penalty increases the model performance in general. The testing time also reduced by 10 times making the model suitable for mobile devices as provided in Figure 4.6. The Pr and Sp improved as penalising reduces the contribution of ambiguous beats, thereby reducing the undesired overfitting effects, and provides better model generalization. Normal class prototypes reduced without major change in performance. SVEB prototypes increased with an increase in Pr, Acc, and Sp. VEB prototypes nearly doubled and reduced Pr but increased Acc and Se.

Table 4.3: Number of individual class prototypes extracted using PIPxResNet.

Class	255	525	649	757	OD	CWOD	WWTOD	ADAOD
N	21	18	16	23	5	22	7	13
S	8	11	8	10	7	23	11	14
V	14	15	13	14	11	16	13	15
All	43	44	37	47	23	61	31	42

Comparing xResNet and PIPxResNet: We compared the best performing xResNet with PIPxResNet and calculated the percentage increase in performance for

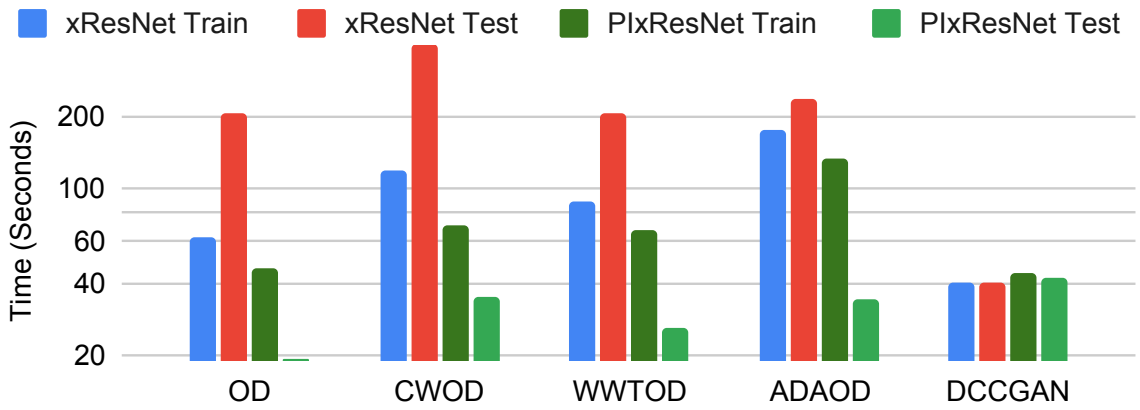


Figure 4.6: Training and Testing time taken by xResNet and PIPxResNet for various data configurations.

all evaluation metrics as described in Table 4.4. Inducing the penalty in xResNet algorithm generated better quality prototypes and had less influence in the number of prototypes, improving the prediction performance of the classifier. The *Pr* for DCCGAN augmented data at 757 batch improved by 9.68%, *Acc* for CWOD improved by 24.68%, *Se* for DCCGAN augmented data at 525 batch improved by 10.69%, and *Sp* for CWOD improved by 26.35%. The best performing PIPxResNet was at DCCGAN augmented data at 255 batch, which is further compared for individual classes with best performing ResNet.

Table 4.4: Comparison between the best performing xResNet and PIPxResNet.

	xResNet				**PIPxResNet**				**Change (%)**			
	Pr	**Acc**	**Se**	**Sp**	**Pr**	**Acc**	**Se**	**Sp**	**Pr**	**Acc**	**Se**	**Sp**
OD	0.66	0.85	0.83	0.91	0.69	0.86	0.84	0.91	4.20	0.65	1.12	0.06
CWOD	0.59	0.69	0.61	0.69	0.59	0.86	0.66	0.87	0.15	24.68	8.52	26.35
WWTOD	0.66	0.86	0.86	0.92	0.68	0.88	0.87	0.93	4.25	2.40	0.50	1.10
ADAOD	0.72	0.89	0.84	0.92	0.77	0.91	0.87	0.93	5.71	2.80	3.35	0.42
255	0.76	0.92	0.86	0.94	0.78	0.93	0.88	0.94	1.68	0.68	2.12	0.25
525	0.68	0.89	0.80	0.92	0.74	0.92	0.88	0.94	8.69	2.87	10.69	1.66
649	0.73	0.91	0.84	0.94	0.74	0.92	0.88	0.94	0.87	0.80	4.38	0.34
757	0.68	0.89	0.81	0.92	0.75	0.92	0.88	0.93	9.69	3.75	8.66	1.09

4.3.6 Prototype Interpretation

The 3-D encoding of extracted prototypes for DCCGAN batch 525 provided in Table 4.3 is illustrated in Figure 4.7. The encoding is obtained from penultimate layer of pretrained neural network. The prototypes represented in the 3D space display a

clear separation between the three classes. The extracted prototypes for DCCGAN batch 525 representing the 186 dimensional heartbeat described in Table 4.3 are illustrated in Figure 4.8 to provide an explanation to the general physician. According to the medical definition [278], ECG signal captures P-wave, QRS complex, and T-wave, where R-peak is positive, Q-peak and S-peak are negative. An abnormality in these waves deteriorates ECG and leads to cardiovascular problems. In Normal beats, all the characteristic waves P, QRS, T are visible. In SVEB, the P wave is usually missing, and in VEB, the QRS complex is wider than the normal beat with a discordant ST segment. The ectopic beats are felt as palpitations caused due to an extra or skipped heartbeat making the patient feel the heart lurch or an extra strong beat occurring momentarily. SVEB originates from the upper chambers or atria and are also called atrial premature beats or premature atrial contraction and may lead to atrial fibrillation. VEB originates from the lower chambers or ventricles and are also called Premature ventricular contractions and may lead to ventricular tachycardia and fibrillation. Both atrial and ventricular fibrillation are life threatening.

Figure 4.8a represent the normal class prototypes with the presence of a P-wave, QRS-complex, and T-wave, capturing the characteristics of normal beats correctly and therefore provide good explanation of normal beats. Figure 4.8b represent the prototypes of SVEB. It is characterized by an abnormal P-wave (either missing or negative P-wave) and a very short duration of the QRS-complex, mainly a missing P-wave and narrow QRS-complex describe SVEB as compared to N. The SVEB prototypes also represent the ideal SVEB correctly and thus provide a good explanation. Figure 4.8c represent the prototypes of VEB. They are categorized by abnormal QRS-complex lasting longer than normal QRS-complexes. The identified prototypes match the beats of subjects. These prototypes can further verify the knowledge learned by the model and point out if the model misses on some important information. Since prototypes represent ideal candidate beats of the class, they might provide additional insight and increase the knowledge base of the physicians. If the prototypes are unable to represent the corresponding class correctly then more data could be added to further improve the model performance. The reliability score quantifies the diagnosis significance. A high score represents a reliable diagnosis, i.e., more beats supporting the prototype with a good pearson correlation between test beat and actual prototype and vice versa.

Additional Experiments: Few additional experiments were also performed: (1) Medoid based testing; (2) Correlation based testing; (3) Voting during testing; and (4) Use of Cosine similarity in place of euclidean distance. In medoid based test-

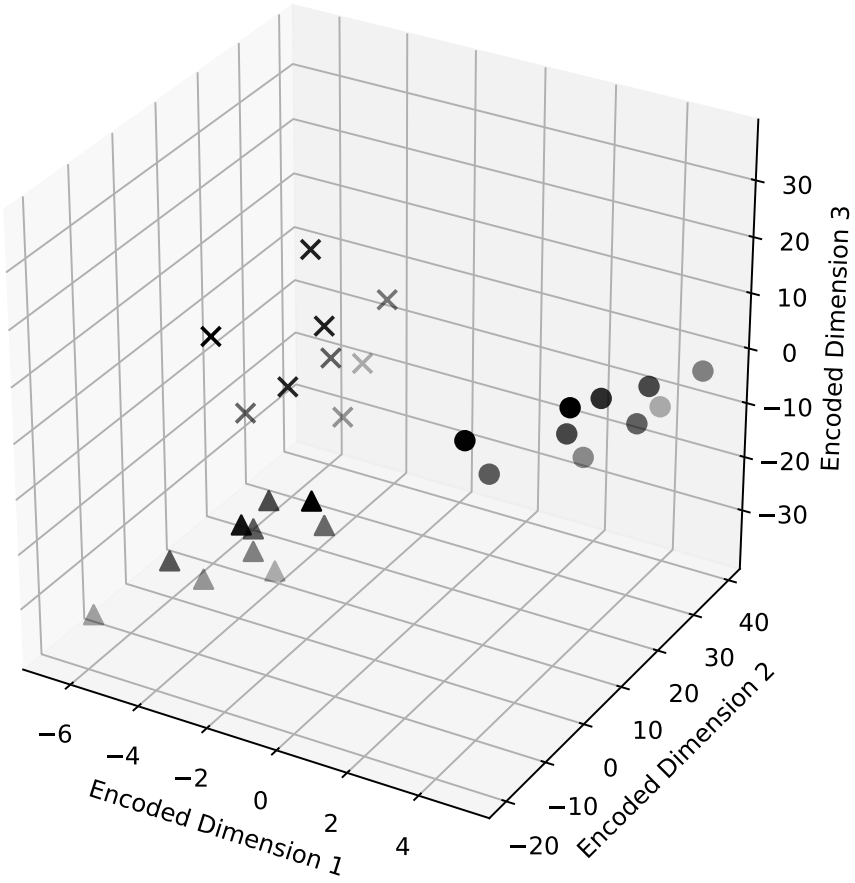
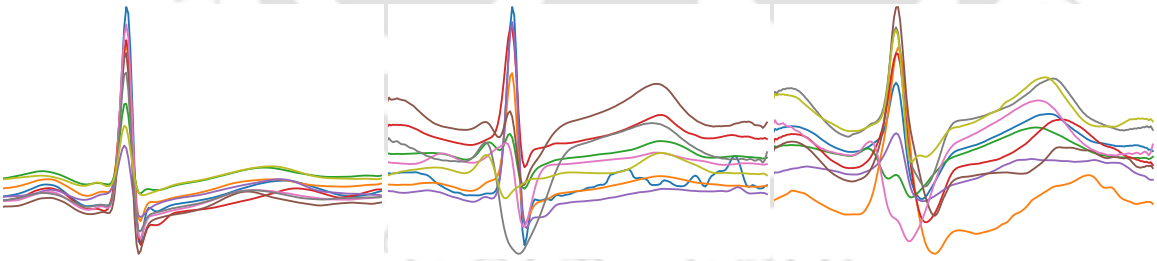


Figure 4.7: Encoded Prototypes. **Circle**: Normal, **Triangle**: SVEB, **Cross**: VEB.



(a) Normal (b) SVEB (c) VEB

Figure 4.8: Actual Prototypes of N, SVEB, and VEB Class.

ing, encoded prototypes are replaced with nearest EB in training data. $O(P \times N)$ time is consumed for replacing the beats with medoids, where P and N represent the number of prototypes and training beats. The results degraded as several prototypes were a linear combination of the respective class beats and after swapping, the new prototypes were unable to capture corresponding class characteristics. Table 4.5 depict the results of medoid based testing. In correlation based testing, correlation

between actual beats and actual prototypes was calculated and performance degraded as correlation performs a linear one to one mapping whereas beat deformity might occur at any timestamp and may not resemble with any existing prototype. Table 4.5 depict the results of correlation based testing. Testing without voting resembles 1-Nearest Neighbor. Following this, a K-Nearest Neighbor inspired voting was performed, where 3 nearest prototype votes were considered for classification. The performance degraded as described in Table 4.5. The performance declined during voting because majority votes might have been from other classes than corresponding class votes. The euclidean distance based metric in density function does not fully capture the high dimensional data distribution. To remedy this, cosine distance was exploited but the number of prototypes increased from 40 to ≈ 3000, N to 1144, S to 1036, V to 68 and testing time also increased manifold, making the metric unsuitable for real time applications.

Table 4.5: Results of KNN based Voting, Medoid and Correlation Based Testing.

Testing	**KNN Voting**				**Medioid Testing**				**Correlation Testing**			
Set	**Pr**	**Acc**	**Se**	**Sp**	**Pr**	**Acc**	**Se**	**Sp**	**Pr**	**Acc**	**Se**	**Sp**
OD	0.68	0.83	0.80	0.89	0.36	0.63	0.46	0.72	0.56	0.79	0.71	0.85
CWOD	0.68	0.90	0.74	0.90	0.34	0.73	0.32	0.73	0.60	0.82	0.73	0.85
WWTOD	0.63	0.84	0.80	0.90	0.25	0.67	0.22	0.64	0.57	0.80	0.74	0.85
ADAOD	0.66	0.88	0.78	0.90	0.41	0.63	0.46	0.74	0.55	0.75	0.64	0.81
255	0.77	0.90	0.83	0.92	0.46	0.72	0.49	0.76	0.56	0.86	0.56	0.83
525	0.73	0.90	0.88	0.94	0.41	0.75	0.54	0.85	0.55	0.78	0.69	0.84
649	0.71	0.90	0.87	0.93	0.38	0.73	0.52	0.84	0.61	0.83	0.66	0.81
757	0.71	0.90	0.87	0.93	0.39	0.71	0.50	0.83	0.54	0.76	0.71	0.83

4.3.7 Comparison with Existing Methods

The comparison of PIPxResNet with previous techniques is provided in Table 4.6. The previous methods used RR interval [229], morphological features [237], wavelet based features [17,239,241,243,244], discrete cosine transform (DCT) [245], stockwell transform (ST) [248], higher order statistical (HOS) features [249], Hermite polynomial based features (HBF) [250], mixture of features [251], and complex heartbeat representations from temporal vectorcardiogram [252]. The extracted feature redundancy is handled through dimensionality reduction techniques such as principal component analysis (PCA), linear discriminant analysis (LDA), independent component analysis (ICA) [243]. Lastly, the classifiers with meta-heuristic techniques such as particle swarm optimization (PSO) [252, 253] and bacteria foraging optimi-

Table 4.6: Comparison of proposed method with previous methods on the basis of different evaluation metrics.

Paper	Features and Classifier	Data	Class	Se	Sp	Pr	Ac	Exp
[229]	Morphological features, RR intervals	MIT ARDB	S V	.76 .77	-	.38 .82	-	×
[17]	Morphological, WT, Temporal Features PCA, PSO, and FFANN	MIT ARDB	S V	.63 .84	.99 .98	.53 .87	.97 .98	×
[237]	RRI, Morphological features, HBF, HOS Weighted Conditional Random Fields	MIT ARDB	N S V	-	-	-	.76 .80 .86	×
[243]	PCA, LDA, ICA reduced DWT subbands. ANN, SVM, PNN	MIT ARDB	All	.99	.99	.99	.99	×
[244]	Stationary Wavelet Transform 1D CNN	MIT ARDB	N S V	.99 .87 .98	.97 .99 .99	.99 .93 .97	.99 .99 .94	×
[245]	PCA reduced DCT coefficients FFNN, SVM, PNN	MIT ARDB	All	.98	.99	-	.99	×
[248]	Stockwell Transform BFO with SVM Classifier	MIT ARDB INCART	S V	.74 .91	.98 .99	.73 .91	.97 .98	×
[252]	Temporal vectorcardiogram Mean, Maximum connectivity degree Joint degree entropy and energy Mean joint degree. PSO with SVM	MIT ARDB	N S V	.94 .62 .87	-	.98 .53 .60	.94	×
[249]	Temporal features Morphological features Satistical features	MIT ARDB	N S V	.89 .83 .86	-	.99 .33 .75	-	×
[251]	ST and Temporal features Mixture of features MLPNN classifier	MIT ARDB	N S V	.95 .78 .89	-	-	.94 .97 .96	×
[253]	Morphological and Temporal features PSO with SVM	MIT ARDB	All	-	-	-	.89	×
[279]	Power Spectral Density and DFT Deep genetic ensemble of classifiers	MIT ARDB	All	.94	.99	-	.99	×
[280]	Signal denoising using DWT QRS features optimized with cuckoo search SVM with feed-forward neural network	MIT ARDB	All	.95	-	.98	.98	×
[96]	Wavelet features, LBP, HOS, RRI Ensemble of SVM	MIT ARDB	N S V	.96 .78 .95	-	.98 .50 .94	-	×
[258]	Temporal and morphological features Clustering Based Classification	MIT ARDB	S V	.54 .43	.84 .82	.23 .16	-	×
[270]	Combination of Sequence Encoder Prototype layer and Dense layer	MIT ARDB	All	.92	-	.85	.97	✓
[281]	RRI and Morphological features Echo State Networks	MIT ARDB AHA	All	.82	.98	.85	.97	×
Our	DCCGAN for Augmentation ResNet for Feature extraction Prototype Based Explanation Reliability Score Supporting Diagnosis	MIT ARDB INCART MIT SVDB CPSC	N S V	.92 .77 **.94**	.91 **.96** **.94**	.84 **.74** **.76**	.91 **.94** .92	✓

sation (BFO) [248] are applied for effective heartbeat classification. Although a few methods [243–245, 279, 280] exhibit substantial performance but the evaluation is performed solely on MIT ARDB dataset or an additional dataset, making the model

biased towards the dataset. PIPxResNet is generalised as it has been tested on four public datasets acquired from different geographical regions. The deep learning models have fared better, but the black-box aspect limits real-world implementation. The PIPxResNet and [270] provide prototype based explanations with better performance. Fixed number of individual class prototypes hinders in capturing diversity for majority class beats and creates redundant prototypes for minority class beats with increased inference time [270]. The PIPxResNet prototype generation process is data-driven, requires no parameter adjustment, making it better suited for heartbeat classification.

4.4 Summary

A Penalty Induced Prototype based eXplainable Residual Neural Network is developed that addresses the black-box nature of deep neural networks. PIPxResNet extracts prototypes from heartbeats using pretrained residual neural network and minimizes resource consumption by eliminating GPU requirement. The method penalises prototypes of a class that resembles other class prototypes and reduces their contribution towards corresponding class. The performance is improved by DCC-GAN synthesised heartbeats. Quantitatively, PIPxResNet achieved state-of-the-art performance without compromising individual class performance when tested on four publicly available standard datasets. Qualitatively, the model simulates a clinical decision support system for general physicians by providing explanation through extracted prototypes. The method can also be adopted to domains that require explanations for classification purposes. The lower dimension data achieved better performance with significant reduction in training time. However, euclidean distance is less suitable for complex data, as more dimensions disrupt the Euclidean norm. The order in which encoded beats are provided to PIPxResNet might alter prototype creation and updation.

The supraventricular ectopic beat and ventricular ectopic beat may lead to atrial and ventricular fibrillation, respectively. Atrial fibrillation and ventricular fibrillation are life threatening cardiac abnormalities. The detection and diagnosis explanation of atrial and ventricular tachyarrhythmias is covered in Chapter 5 and 6, respectively.

5

Posthoc Interpretation Techniques for Explaining Ventricular Tachyarrhythmia Predictions using ResNet

Ventricular Tachyarrhythmias such as Ventricular tachycardia (VT) and Ventricular fibrillation (VF) accounted for 4.5% of the subjects with CVD and heart failure in India [282]. VT is a rapid and well-organized heart rhythm with 100 beats per minute. VF is a fast, disorganized, and erratic cardiac activity where multiple heart cells trigger together, leading the heart muscles to quiver and generate smaller amplitude waveform with fibrillation frequencies. The heart struggles to pump blood to the brain during these lethal arrhythmias, leading to unconsciousness and eventually death. Therefore, correct classification of VT and VF is of utmost importance [283]. In the past, deep learning models (DLM), have been employed to detect VT and VF through single channel Electrocardiogram (ECG) signal. DLM have achieved greater performance but they lack interpretability. Therefore, gradient backpropagation based posthoc techniques are explored in this chapter that highlight relevant signal timestamps. The physician might interpret the highlighted signal timestamps as an explanation even if the saliency methods work as peak detectors. For instance, the techniques might highlight R peaks that are responsible for detecting heartbeats from ECG signals. Therefore, sanity checks proposed by Adebayo et al. [45] are explored to verify whether the techniques highlight relevant signal timestamps responsible for the corresponding diagnosis.

The chapter organization is as follows. Section 5.1 provides the literature survey. Section 5.2 describes the methodology followed in the chapter. Section 5.3 describes experimental setup and results. Section 5.4 summarizes the chapter.

5.1 Literature Review

Studies conducted for ventricular tachyarrhythmia detection can be categorized into two broad categories: conventional feature-based and deep learning-based techniques [5, 94, 284–297]. Feature-based algorithms extract temporal, morphological, spectral, wavelet, and fractal features followed by a classifier for classification. Santos et al. [290] extracted Power Spectral Density (PSD) from ECG signals. Li et al. [94] extracted 14 features and selected optimal feature subsets using Genetic Algorithm (GA). The selected features were auxiliary count and Filter Leakage measure (FLM) followed by Support Vector Machine (SVM) classifier. Alonso et al. [295] extracted morphological and spectral features followed by SVM. Prabhakar et al. [291] used Zero-Crossing Rate (ZCR) and Peak-to-Peak Interval (PPI) features. Figuera et al. [285] extracted slope, waveform complexity, time-frequency, and time-domain features. Mohanty et al. [287] extracted FLM, spectral features, covariance measure, frequency, area, kurtosis, exponential algorithmic features, skewness, and few Nonlinear entropy features that include Shannon, Norm, Log, Threshold, and Sure entropy followed by SVM and Decision Tree (DT). Xu et al. [293] used adaptive variational mode decomposition (VMD) and extracted variational modes followed by Boosted Classification And Regression Tree (CART) classifier. Tripathy et al. [294] used Digital Taylor-Fourier transform to decompose ECG into various oscillatory modes and calculated magnitude and phase difference with SVM. Taye et al. [284] extracted Time domain Heart Rate Variability (HRV) features that include mean, standard deviation, root mean square differences of R-to-R Peak Interval (RRI); Frequency-domain HRV features include power in following bands (0–0.04 Hz), (0.04–0.15 Hz), (0.15–0.4 Hz), and ratio of low over high frequency; Mean and standard deviation of QRS signed area; Mean and standard deviation of R-peak amplitude. Mandala et al. [288] extracted mean and standard deviation of RRI, mean heart rate, standard deviation of RRI, mean QRS duration, mean and standard deviation for the amplitude of Q, R, and S wave. A brief comparison of these techniques in terms of dataset, features, classifier, and results is provided in Table 5.4. Amann et al. [298] implemented ten existing algorithms and concluded that the algorithms failed to achieve the claimed performance. Therefore, testing on multiple databases is essential for algorithm reliability.

The deep learning techniques exploit raw signals and perform end to end ECG classification. Recurrent Neural Networks (RNN) have proven to be effective and efficient for time dependent and variable length signals as they can integrate the

temporal relations present in signals. Picon et al. [286] used 1-D Convolution Neural Networks (CNN) and Long Short Term Memory (LSTM) that integrated the long and short time relations in the feature map. Fujita et al. [296] employed a six layer NN for detecting Normal, Atrial Fibrillation, Atrial Flutter, and Ventricular Fibrillation. Acharya et al. [5] detected shockable and non shockable VA using 11 layer 1-D CNN. In another work, Acharya et al. [297] performed diagnosis of congestive heart failure from ECG signals using 11-layer CNN. Nguyen et al. [292] used modified Variational Mode Decomposition (VMD) to decompose ECG into variational modes followed by CNN for feature extraction and Boosting classifier. Although DLM achieved cardiologist level performance, they lack interpretability. Researchers have developed interpretability methods to understand the reasoning behind model diagnosis. Only Strodthoff et al. [147] have used gradient based attribution method for detecting myocardial infarction, but did not verify the validity of extracted saliency maps. This leaves a research gap to develop and validate interpretability techniques that provide correct and reliable explanations.

The results presented in the literature are substantial but suffer with a few limitations. The handcrafted features are rhythm sensitive. The deep learning models achieve good performance but lack model interpretation. The implicit interpretability techniques are model specific and are inapplicable to pretrained models. The validity of interpretability techniques is not verified. The models developed in the past are trained and tested on limited data.

5.2 Methodology

The steps involved for achieving ventricular tachyarrhythmia classification are illustrated in Figure 5.1. The data from multiple geographical locations is acquired and preprocessed before feeding to the classifier. The inadequate VT and VF rhythms are synthesized using Synthetic minority oversampling technique [42] (SMOTE), Borderline SMOTE using Support Vector Machine [43], and Adaptive synthetic [44] augmentation technique. The convolution based residual neural networks are employed to detect VT, VF, and NSR from segmented ECG data. The effect of ResNet layers and ECG segment length is analysed to provide faster treatment in early stages. ResNet predictions are explained using three gradient backpropagation-based posthoc interpretability techniques, namely, Guided Backpropagation [36], Gradient Class Activation Map [37], and Guided Gradient Class Activation Map [37] to highlight relevant signal timestamps that contribute towards or against the diagnosis. The validity of

interpretability techniques is verified using sanity checks [45].

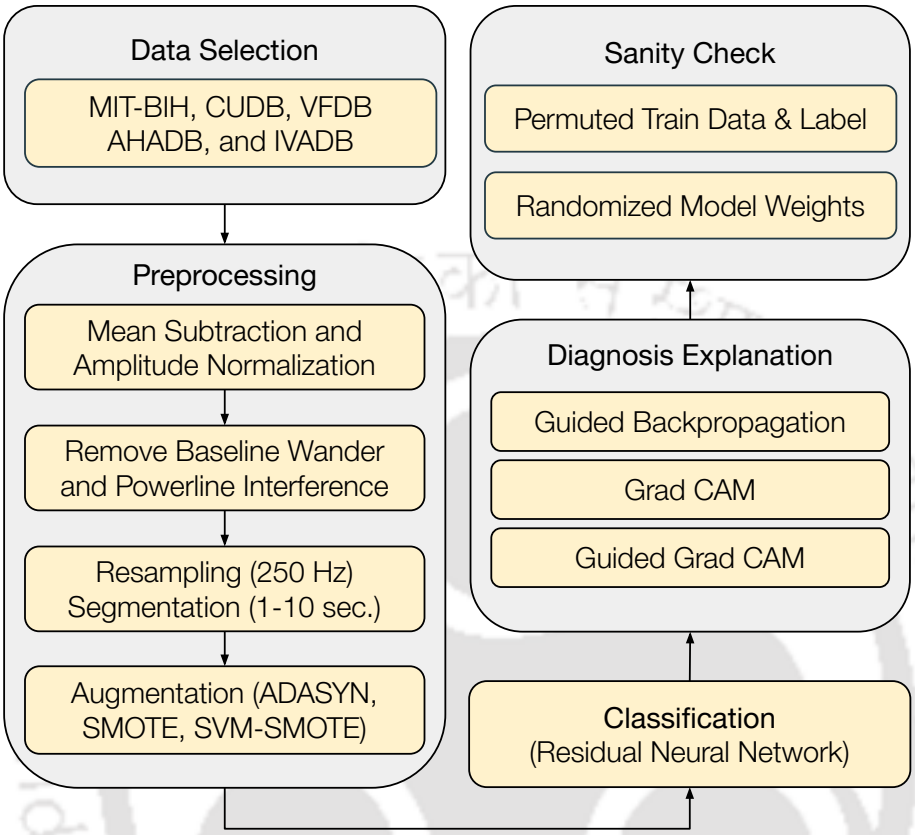


Figure 5.1: Block diagram of the proposed framework.

5.2.1 Data Description and Preprocessing

The databases include Creighton University Ventricular Tachyarrhythmia Database (CUDB) [40], Massachusetts Institute of Technology Database (MITDB) [38], Ideo Ventricular Arrhythmia Database (IVADB) [41], Malignant Ventricular Arrhythmia Database (VFDB) [39] and American Heart Association Database (AHADB) [34,299]. Table 5.1 describes detailed database description. Episodes of VT, VF, and NSR are extracted from the databases. Single channel recording from each record is considered to avoid redundancy in extracted signals. The mean was subtracted and the signals were min-max normalized.

The signals suffered from high and low frequency noises caused due to electromyogram induced noise or power line interference and baseline wander. Low frequency noises increase the amplitude of ECG that degrades the PQRST morphologies [82, 228]. Baseline wander, and power line interference was suppressed using a

Table 5.1: Detailed description of all databases.

Database	Channel	Recordings	Sampling
MITDB [38]	Double	47	360 Hz
VFDB [39]	Double	22	250 Hz
CUDB [40]	Single	24	250 Hz
IVADB [41]	Twelve	6	2 KHz
AHADB [34]	Double	10	250 Hz

high pass filter with a cutoff frequency of 1 Hz and a notch filter centered at 50 Hz, respectively [300].

Resampling and Segmentation: Cleaned signals were resampled to 250 Hz. Then overlapping segmentation was performed, ranging from 1 to 10 second segments [263]. The dataset was split into a 80 : 20 split for train and test dataset. The train dataset was further split into a 80 : 20 split for training and validation dataset.

Augmentation: Inadequate VT and VF rhythm bias the neural network towards NSR rhythms. This class imbalance was handled using augmentation techniques, namely, SMOTE, Borderline SMOTE using SVM, and ADASYN. SMOTE [42] generates new instances between existing samples using the K-Nearest Neighbors algorithm without considering majority class data. This leads to the generation of ambiguous instances if strong overlapping is present in the classes. ADASYN [44] generates synthetic samples that are inversely proportional to the density of existing samples. SVM SMOTE [43] replaces KNN with SVM and generates new instances in the close vicinity of each class's support vectors. It is important to note that the data augmentation is not applied on the test and validation dataset. The validation data is used to determine the optimum number of resnet layers for better performance and less computation and memory requirement. The results in Section 5.2.2 are performed only on the test dataset. The total number of segments for each second before and after data augmentation are provided in Table 5.2.

5.2.2 Residual Neural Network Classifier

The details of ResNet architecture are described in Section 4.3.3. Figure 5.2 illustrates a 1-layer ResNet. Consider $G(z)$ as a mapping of stacked layers with z denoting the identity connection, the residual block output is denoted by $G(z) + z$. The identity connections do not introduce computational burden. 1-D Convolution performs feature extraction using convolution filters (f) of size $W \times N$, where W is the filter width and N is the number of filters, pooling with k and stride of r generates $s(f)$

Table 5.2: Number of segments before and after data augmentation. NSR: Normal Sinus Rhythm; VT: Ventricular Tachycardia; VF: Ventricular Fibrillation.

Segment Length		Original			Augmented		
Samples	**Seconds**	**NSR**	**VT**	**VF**	**NSR**	**VT**	**VF**
250	1	39301	18043	16846	39301	39514	39056
500	2	19596	9160	8336	19596	19053	18791
750	3	13040	6071	5616	13040	13038	13595
1000	4	9818	4516	4210	9818	10196	9947
1250	5	7879	3608	3347	7879	8186	7981
1500	6	6519	3071	2771	6519	6488	6685
1750	7	5642	2540	2412	5642	5689	5589
2000	8	4920	2265	2083	4920	5083	5027
2250	9	4338	1998	1903	4338	4348	4327
2500	10	3897	1819	1698	3897	3981	3957

as represented in Equation 5.1.

$$s_{j,u}(f) = \left(\sum_{w=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} |f_{g(w,j,u)}|^p \right)^{1/p} \tag{5.1}$$

where, $g(w, j, u) = (r \cdot j + w, u)$ function maps s to f, p is the p-norm order. If $p \leftarrow \infty$, it resembles max pooling and if $r > k$, no overlapping pooling regions. Comparing the pooling operation defined in Equation 5.1 to convolutional layer c applied to feature map f given in Equation 5.2.

$$c_{j,u}(f) = \sigma \left(\sum_{w=-\lfloor k/2 \rfloor}^{\lfloor k/2 \rfloor} \sum_{u=1}^{N} \theta_{w,u} \cdot f_{g(w,j,u)} \right) \tag{5.2}$$

where, θ represent filter weights, $\sigma(\cdot)$ is Rectified Linear Unit (ReLU) activation function with $\sigma(x) = max(x, 0)$, and $u \in [1, N]$ is feature map index.

5.2.3 Posthoc Interpretability Techniques

The interpretation workflow is depicted in Figure 5.3. The input signal (S) is provided to the pretrained ResNet model (M) and the predicted targeted class score (C) before softmax activation function is extracted. The score (C) and last convolution layer filters (A^k) to visualize are further used for interpretation. Algorithm 9 explains the generation of saliency maps. First, a model is built with input similar to pretrained ResNet and output as the convolution layer. The ReLU activations are modified to

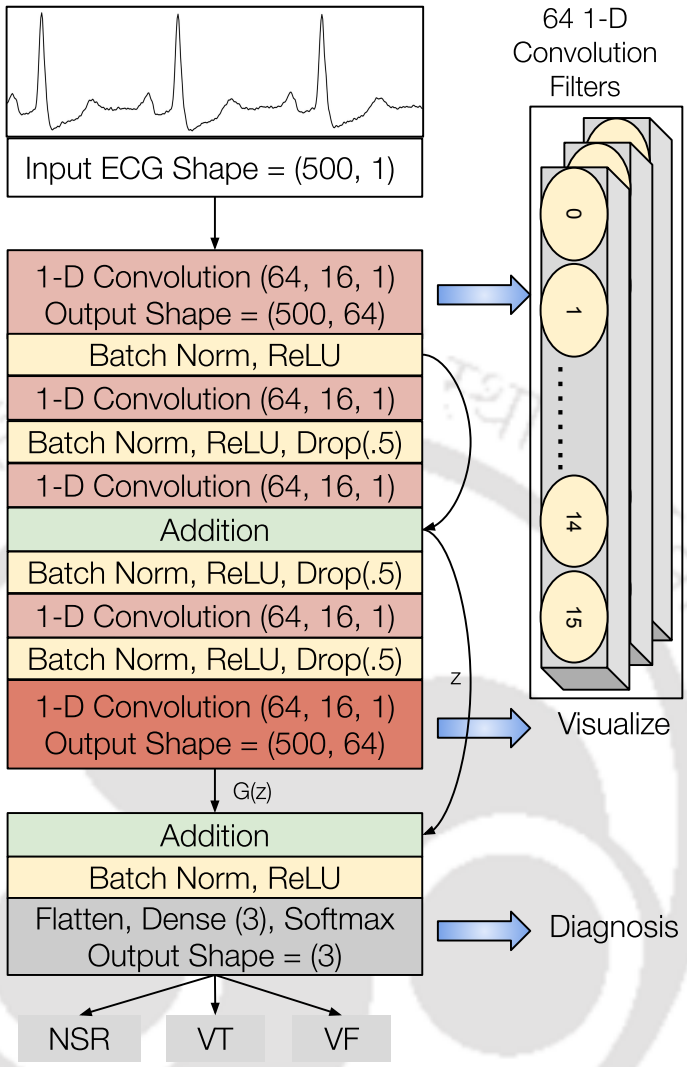


Figure 5.2: Single Layer Residual Neural Network. BN: Batch Normalization; ReLU: Rectified Linear Unit.

guided ReLU so that the gradient are overridden and only non negative gradients backpropagate. Finally, the clean ECG segment is forward propagated and gradients of the target class with respect to convolution feature maps are calculated. The score and gradient are used to calculate GBP, Grad CAM, and Guided Grad CAM saliency maps. The description and calculation of class discriminative localization maps for each interpretability technique is explained subsequently. Similar phenomena can be extended for other convolution layers.

Guided Backpropagation: The Guided Backpropagation (GBP) or guided saliency is the combination of vanilla backpropagation and DeconvNets [301], where only the positive gradients are backpropagated, and the negative gradients are clipped

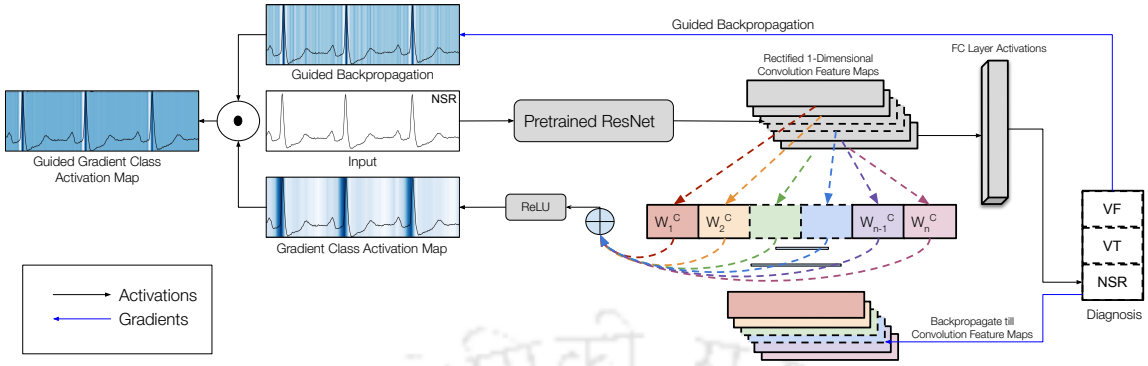


Figure 5.3: Proposed Interpretation Workflow generating Saliency Maps of GBP, Grad CAM, and Guided Grad CAM.

Algorithm 9 Posthoc Interpretability Maps

Input: Pretrained Model (M), Signal (S), and Convolution Layer to Visualize (A^k)
Output: GBP Maps (G_{GBP}), Grad CAM (G_{GC}), and Guided Grad CAM (G_{GGC})
// Build new model with input similar to M and output as A^k
$M_{new} = \text{Model}(M_{input}, A^k)$
    // Change ReLU to Guided ReLU in M_{new}
for L_{act} ***in*** M_{new} **do**
    if L_{act} ***== ReLU*** **then**
        // $x \leftarrow$ input of L_{act}
        // $dy \leftarrow$ gradient backpropagated to L_{act}
        $L_{act} = (dy > 0) \times (x > 0) \times (dy)$
    end
end
Target Class Score (Y^C) = Predict S using M_{new}
$G = \frac{\partial Y^C}{\partial A_i^k}$

// Calculating G_{GBP} by Min-Max Normalizing G
$G_{GBP} = \frac{G - Min(G)}{Max(G)}$

// Calculating G_{GC} by Multiplying A_k with Mean of A_k and applying ReLU activation
$G_{GC} = ReLU\left(\sum_k \sum_i GA^k\right)$

// Calculating G_{GGC} by Element-wise Multiplication of G_{GBP} and G_{GC}
$G_{GGC} = G_{GC} \cdot G_{GBP}$

to zero [36]. Figure 5.4 illustrates the forward and backward propagation during the training, where the input is passed through the ReLU activation function as described in Equation 5.3. Here, the input z_i^l represents the output of i^{th} neuron

of l^{th} layer. During the backward propagation, only those values are propagated that were positive during the forward pass and other values are clipped as described in Equation 5.4. The error signal E_i^{l+1} represents the error backpropagating from i^{th} neuron of $(l+1)^{th}$ layer. GBP modifies the backward propagation by passing only those values that were positive during the forward and backward propagation. The new error signal takes into account both z_i^l and E_i^{l+1} before passing the error to previous layer as described in Equation 5.5. Hence, the new gradient is guided by the input from the preceding layer and error from the succeeding layer. The backpropagated gradients highlights only those timestamps that highly influences the diagnosis, whereas, the conventional backpropagation does not mask negative entries during backward propagation. The intensity values of heatmap are divided by the maximum intensity value of heatmap to normalize the heatmap such that intensity values range between 0 and 1. The heatmaps are overlapped with input signal to obtain guided backpropagation heatmaps. GBP calculates the imputed version of gradient with respect to input, keeping the network weight matrix θ constant while using ReLU. Being a posthoc interpretable technique, GBP does not influence the decision-making capability of ResNet model.

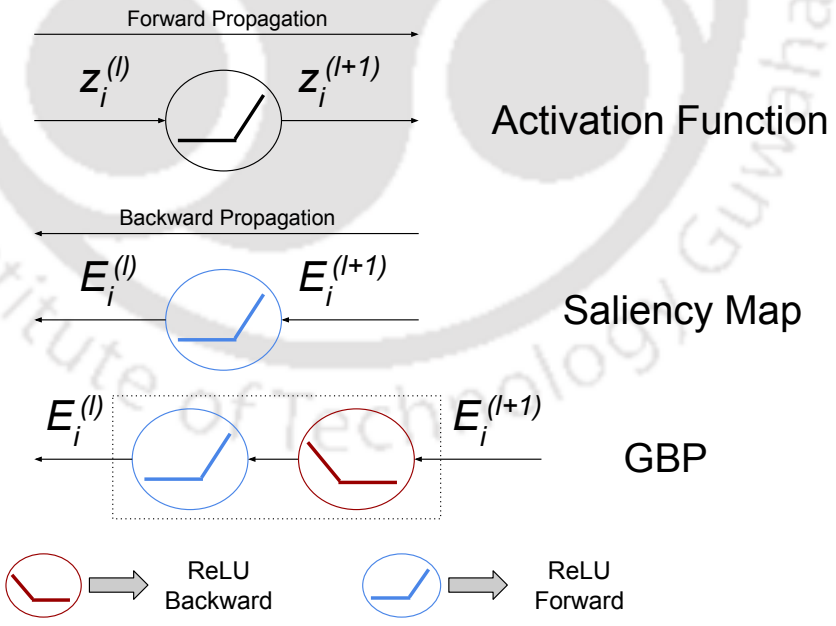


Figure 5.4: Illustration of GBP Method

$$z_i^{l+1} = ReLU(z_i^l) = max(z_i^l, 0) \tag{5.3}$$

$$E_i^l = E_i^{l+1} \; \forall \; (z_i^l > 0), \; where \; E_i^{l+1} = \frac{\delta z^{out}}{\delta z_i^{l+1}} \tag{5.4}$$

$$E_i^l = E_i^{l+1} \; \forall \; (z_i^l > 0) \; and \; (E_i^{l+1} > 0) \tag{5.5}$$

Gradient Class Activation Map: The Class Activation Mapping (CAM) technique indicates the discriminative regions used by a specific class of CNN to identify the predicted class [302]. CAM applies to CNN with Global Average Pooling layer, leading to inferior accuracies as new models needs to be trained [302]. Grad CAM improves over CAM shortcomings, does not require modifications to existing CNN models [37]. Grad CAM saliency is generated by performing weighted combination of A^k and backpropagated gradients followed by a ReLU as described in Equation 5.6. This results in generation of coarse heatmap of similar size as that of A^k. ReLU makes sure that features having positive influence on the target class are only highlighted otherwise more information is highlighted, leading to low localization performance.

$$G_{GC} = ReLU\left(\sum_k \sum_i \frac{\partial Y^C}{\partial A_i^k} A^k\right) \tag{5.6}$$

Guided Gradient Class Activation Map: Guided Grad CAM combines the best aspects of Grad CAM and GBP through element-wise multiplication and produces class-discriminative saliency maps for any arbitrary CNN architecture [37] as described in Equation 5.7.

$$G_{GGC} = G_{GC} \cdot G_{GBP} \tag{5.7}$$

5.2.4 Sanity Check Mechanism

Ideally, the saliency maps should correlate with model parameters/weights learned during training using the training data. Moreover, visual assessment of saliency maps is insufficient for understanding whether meaningful explanations are provided to stakeholders. Therefore, sanity checks proposed by Adebayo et al. [45] are employed to verify the validity of saliency maps by performing model parameter randomization and data permutation test. The sanity check mechanism is illustrated in Figure 5.5.

Model Weight Randomization Test: The weights of convolution layer filters are modified and the effect on saliency maps is observed. If the saliency maps gets changed, that means they depend on model's learned parameters, making the technique effective for diagnosis interpretation and if saliency maps remain unchanged, it

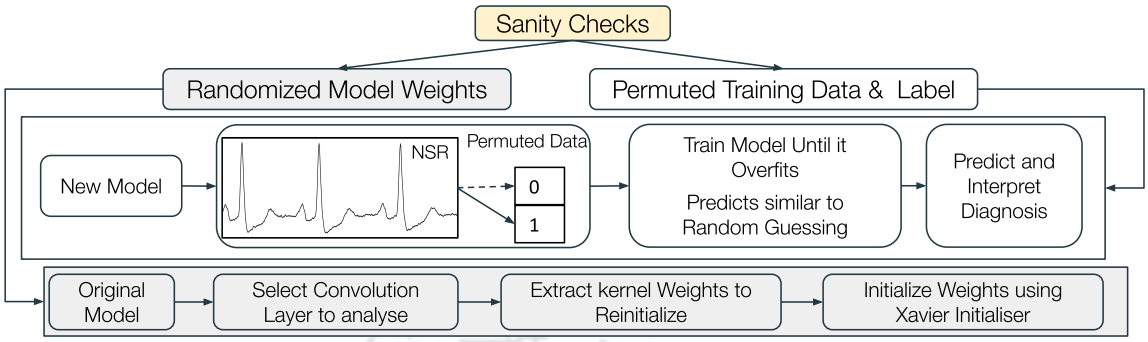


Figure 5.5: Sanity Check Mechanism. (1) Model Weight Randomization Test; (2) Training Data Permutation Test.

means that the saliency maps are unrelated to model parameters, the technique fails and will provide wrong diagnosis interpretation. Algorithm 10 describes the model parameter randomization test.

Algorithm 10 Model Weight Randomization

Input: Convolution Layer to Visualize (A^k)
Step 1: $AM^k \leftarrow$ Reinitialize A^k with HeNormal Initializer
Step 2: Update Model by replacing A^k with AM^k
Step 3: Predict Test data using Updated Model
Step 4: Generate saliency maps using Algorithm 9
if *Maps Remain Same as Earlier* **then**
| Maps depend on model weights, **Pass**
else
| Maps independent of model weights, **Fail**
end

Training Data Permutation Test: The test measures the sensitivity of interpretability techniques with respect to the relationship between training data and label combination. An explanation method insensitive to randomized data label combination cannot possibly provide correct explanation mechanisms present in the data generating process. Algorithm 11 describes the training data permutation test, where the labels corresponding to the ECG segments are permuted to destroy their relationship. The task is accomplish by training a new ResNet model until 95% accuracy is achieved for uncorrelated segment-label pairs (overfitting). For instance the NSR with label 1 is now trained with label 0. During testing, the model performs similar to a random guessing classifier.

Algorithm 11 Training Data Permutation

Input: Training Data (D) with Labels (L)
Step 1: Create Single Layer ResNet (M)
Step 2: Permute/Randomize D and L
Step 3: Train ResNet till 95% accuracy is achieved (Overfit)
Step 4: Predict test data using ResNet
Step 5: Generate saliency maps using Algorithm 9
if *Maps Remain Same as Earlier* **then**
| Maps depend on train data and label, **Pass**
else
| Maps independent of train data and label, **Fail**
end

5.3 Experiments and Results

For quantitative analysis, the model performance is calculated by varying the effect of ResNet layers, augmentation technique, and optimum ECG segment length required for accurate diagnosis. The model performance is measured using Accuracy (Acc), Sensitivity (Se), Specificity (Sp), and Precision (Pr) as described in Equation 4.3.2.

5.3.1 Quantitative Evaluation

ResNet Layer Effect: The model performance increases upto five ResNet layers but declined thereafter due to data scarcity for 2 second ECG segment as described in Figure 5.6. The performance achieved by 1-layer ResNet and 5-layer ResNet are almost similar. Therefore, 1-layer ResNet was preferred as it achieved an Acc of 0.97, Se of 0.94, Sp of 0.98, and Pr of 0.95. The 4 MB model took 5 minutes per epoch for training and 1 milli second inference time. It can be easily deployed in automatic external defibrillators and bedside monitors.

It can be seen in Figure 5.7 that in single layer model, the size of model is 3.5 MB which increases exponentially to 183.3 MB in 15 layer model. As the layers increase, the performance increases upto five resnet layers but declined thereafter due to data scarcity. The curse of dimensionality was observed in many models that deal with larger size ECG segments as the model performances drastically reduced when large ECG segments were provided as input to the model.

Effect of Augmentation Technique: The results obtained using original and augmented data for 2 second segment using 1-layer ResNet are described in Figure 5.8. The models trained on ADASYN augmented data as descried in Figure 5.8c, achieved best results as ADASYN considers the minority class density distribution rather than

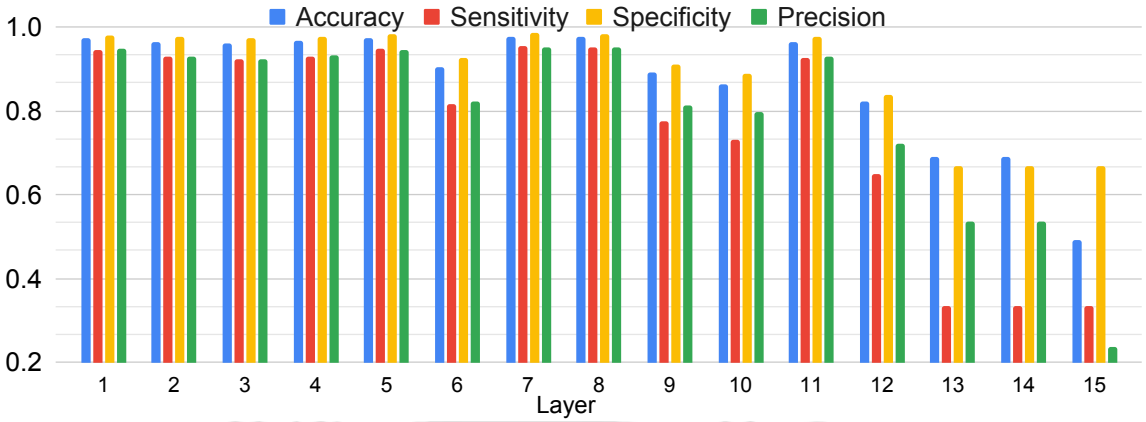


Figure 5.6: Model performance variation for increasing number of ResNet layers for two second ECG segment.

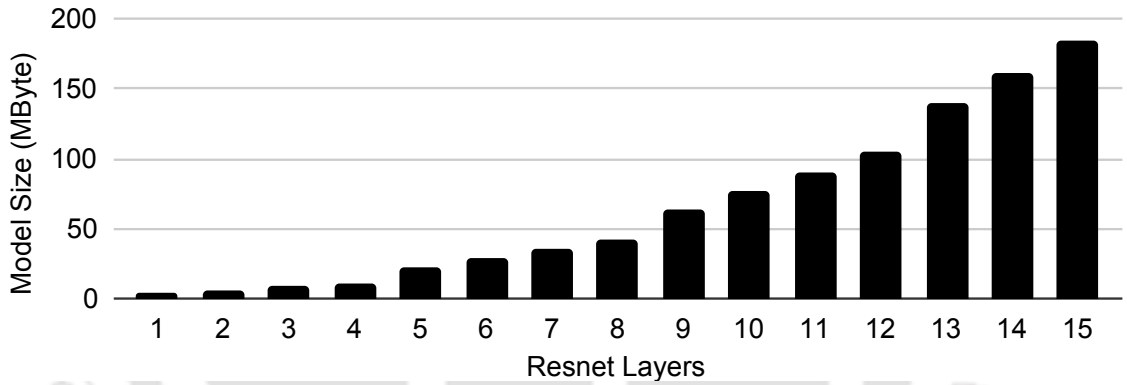
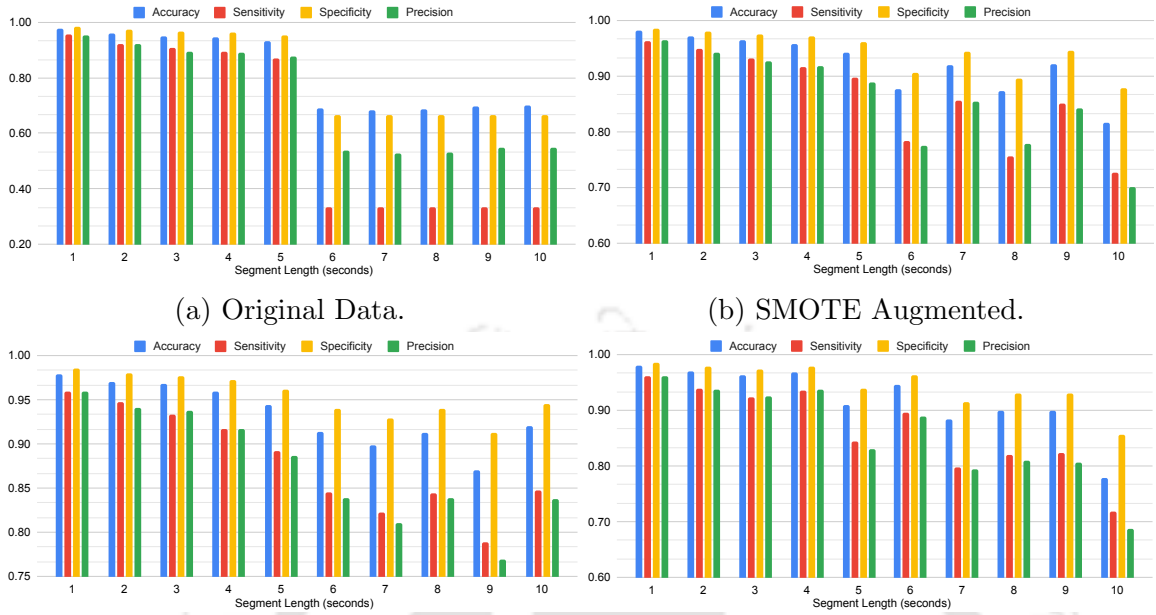


Figure 5.7: Model size variation with increasing number of ResNet layers for two second ECG segment.

assigning a uniform weight for all minority points while generating synthetic samples. Therefore, the newly generated instances are non linearly correlated with the original samples and persist high resemblance towards the minority classes. This helps model in generalizing towards the minority classes. A steep decline in model performance was observed without augmentation due to limited availability of training data. The performance increased when augmentation was applied as the number of training instances were sufficient. ADASYN outperformed other techniques for all segment lengths and was therefore preferred for further experimentation.

Varying Segment Length: The effect of segment length is described in Figure 5.8. It can be observed that, for all the segments, the model produces Sp that is better than Se, which is desirable as a lesser number of people would receive shocks and might suffer from cardiac arrest. The reason for lower Se was the limited number of training samples that capped the ResNet's generalization capability. The model performance



(a) Original Data.

(b) SMOTE Augmented.

(c) ADASYN Augmented.

(d) SVM-SMOTE Augmented.

Figure 5.8: Single layer ResNet model performance for original and augmented data for varying segment lengths.

also deteriorated with increasing signal length due to the decline in training instances causing model underfitting. The model performed better for ECG segments ranging between 1-4 second segments as shown in Table 5.3. The shorter duration segments perform relatively better than longer duration segments. Data augmentation enabled the model to generalize much better and provide a significant improvement over the baseline results. 1-layer ResNet achieved an Acc of 0.98, Se of 0.96, Sp of 0.98, Pr of 0.96 for 1 second segment and Acc of 0.97, Se of 0.95, Sp of 0.98, Pr of 0.94 for 2 second segment using ADASYN augmentation.

Model for Interpretation: Single layer ResNet developed for 2 second ECG with ADASYN augmentation is used for interpreting diagnosis. Figure 5.9 illustrates the confusion matrix for the prediction on test data (combined from test data of all sources). The model distinguishes between the NSR and VT-VF segments, but deteriorates while differentiating between VT and VF segments as 148 VF episodes were classified as VT and 174 VT episodes were classified as VF. Hence, we further examine the reasons behind the correct and incorrect diagnosis for 2 second segments.

Table 5.3: Single layer ResNet Performance for varying segment lengths on ADASYN augmented data.

Time	Class	Accuracy	Sensitivity	Specificity	Precision
1 sec.	NSR	0.99	0.99	0.99	0.99
	VT	0.97	0.94	0.98	0.95
	VF	0.97	0.95	0.98	0.94
2 sec.	NSR	0.98	0.98	0.99	0.99
	VT	0.97	0.92	0.98	0.93
	VF	0.96	0.94	0.97	0.90
3 sec.	NSR	0.99	0.99	0.98	0.99
	VT	0.95	0.92	0.96	0.89
	VF	0.96	0.88	0.98	0.93
4 sec.	NSR	0.99	0.99	0.99	0.99
	VT	0.94	0.86	0.97	0.90
	VF	0.95	0.90	0.96	0.86

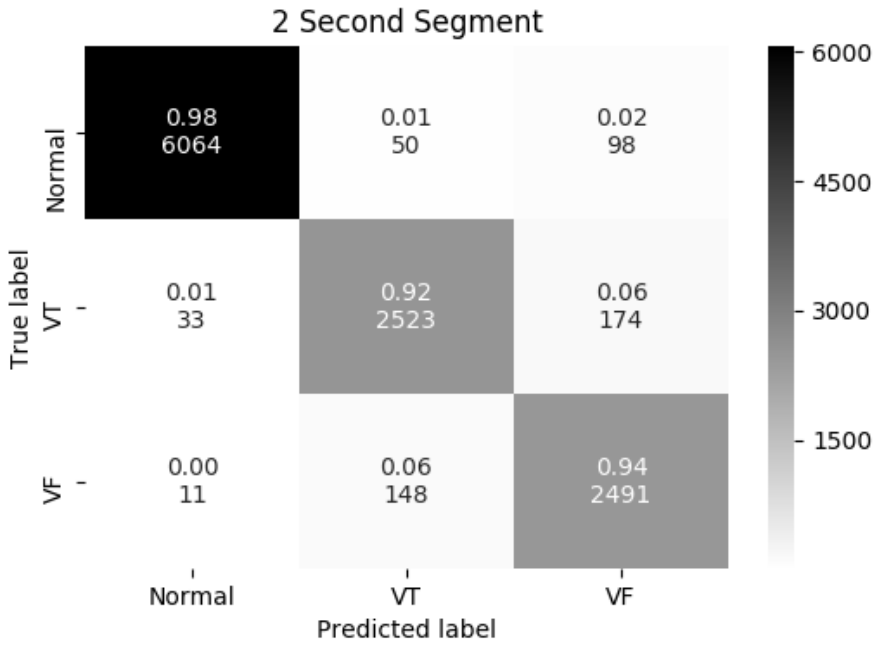


Figure 5.9: Confusion Matrix on test data for two second segment using single layer ResNet model.

5.3.2 Comparison with State of the Art Techniques

Table 5.4 compares the single layer ResNet classifier when tested on 1 second ECG segment with the techniques provided in the literature. Existing methods are verified on mostly two or three datasets, making them unreliable and underperform in real

world scenario. The proposed method has been verified on five datasets acquired from multiple geographic locations. The Normal Sinus Rhythm Database (NSRDB) was excluded as sufficient NSR were available for experimentation. ResNet was preferred as it had outperformed other CNN based models [303]. Moreover, to increase the model performance, augmentation techniques, namely, ADASYN, SMOTE, and SVM-SMOTE were employed for circumventing data imbalance and increasing performance.

Table 5.4: Performance comparison with state-of-the-art methods on the basis of different evaluation metrics.

Paper	Data	Method	Se	Sp	Acc
[5]	MITDB, CUDB, VFDB	11 layer 1D CNN	0.91	0.95	0.93
[295]	MITDB, CUDB, VFDB	Morphological and Spectral features SVM Classifier	0.92	0.97	–
[300]	AHA, CUDB, MITDB	Time-delay based algorithm	0.79	0.97	0.96
[94]	AHA, CUDB, VFDB	GA selected 14 TF features SVM Classifier	0.98	0.98	0.98
[287]	MITDB, CUDB, VFDB	Temporal, Spectral, Statistical features SVM and Decision Tree Classifier	0.98	0.99	0.99
[292]	CUDB, VFDB	CNN for feature extraction Boosting classifier	0.97	0.99	0.99
[286]	AHA, CUDB, VFDB	1-D CNN and LSTM Model	0.99	0.98	0.99
[304]	MITDB, CUDB, VFDB	Entropy features Random forest Classifier	0.96	0.97	0.97
[293]	MITDB, CUDB, VFDB	Adaptive-VMD Boosted CART classifier	0.97	0.98	0.98
[294]	CUDB, VFDB	Phase Difference features SVM Classifier with RBF Kernel	0.86	0.93	0.89
[284]	CUDB, PAFDB, NSRDB	11 HRV, 2 QRS, 2 R-peak features ANN Classifier	0.98	0.99	0.98
[288]	NSRDB, VFDB	5 time-domain, 7 QRS features SVM, DT, NB Classifier	0.89	0.91	–
[289]	CUDB, VFDB, PTBDB NSRDB	Box Count Features	–	–	0.94
[290]	CUDB, VFDB, MITDB NSRDB	Modified Power Spectral Density SVM Classifier	–	–	0.37
[291]	MITDB, CUDB, VFDB	ZCR and PPI	99.61	99.96	99.92
Ours	**AHA, CUDB, VFDB,** M**ITDB, IVADB**	**ResNet Classifier ADASYN Augmentation**	**0.96**	**0.98**	**0.98**

5.3.3 Qualitative Evaluation

The predictions are interpreted using three posthoc gradient backpropagation based interpretability techniques, namely, GBP, Grad CAM, and Guided Grad CAM that highlight relevant signal timestamps responsible for and against the prediction. The

analysis is performed for correct and incorrectly classified rhythms. In addition, the validity of aforementioned techniques is verified using Sanity checks that explains the reliability of interpretability techniques. The single layer ResNet performed better than other configurations. Hence, the diagnosis interpretation for 2 second ECG segment is explored using single layer ResNet. Firstly, the origination of NSR, VT, VF with their respective ECG waveforms is shown in Figure 5.11 followed by posthoc interpretation of the saliency maps of last convolution layer of correctly classified and miss-classified ECG segments. In addition, the interpretability technique efficacy is verified through sanity checks. The training data permutation test performs retraining of single layer ResNet using permuted training data and label combination until the model achieves 95% accuracy. Figure 5.10 illustrates the training curves of the model trained on permuted data. Table 5.5 describes the prediction on test data which resembles to random guesses. It can be concluded that convolution filters have learned something entirely different as compared to the previously trained models. The model weight randomization test is also performed to generate another class of models. The original model and the modified class of models are used to generate qualitative saliency maps for visualization. The colour coding of saliency maps represent the attribution score, where the blue colour influence the signal timestamps towards the model diagnosis and the white colour influence the signal timestamps against the model diagnosis.

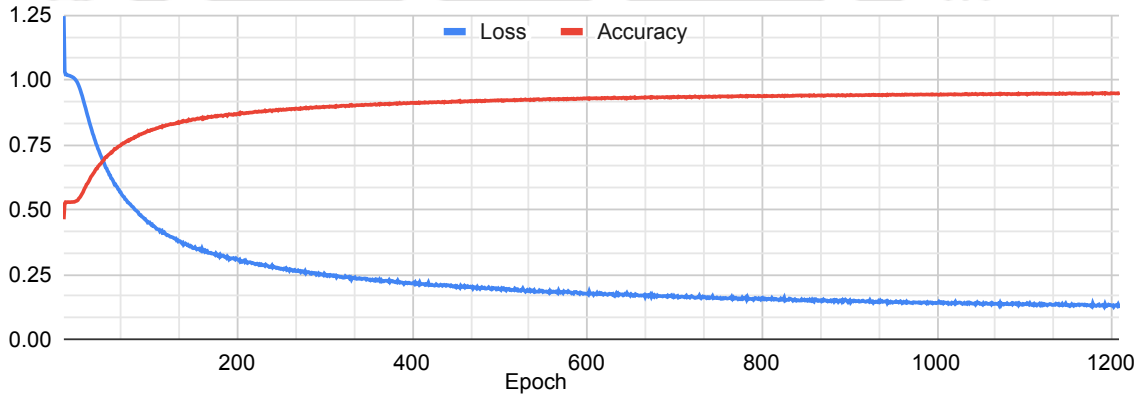


Figure 5.10: Model training curves for permuted data.

5.3.4 Ventricular Arrhythmia Origin

The origination of NSR, VT, and VF with their respective ECG waveforms is provided in Figure 5.11. During NSR, the impulse is generated from sinoatrial node and the waveform depicts narrow QRS complex with P-wave and T-wave as shown in Figure

Table 5.5: Prediction on test data using the model trained on permuted data and label combination.

Label	Precision	Accuracy	Sensitivity	Specificity
NSR	0.50	0.47	0.53	0.40
VT	0.24	0.65	0.24	0.77
VF	0.17	0.63	0.16	0.77
Average	0.31	0.58	0.31	0.65

5.11a. The reentrant monomorphic VT with left bundle branch block morphology generated from the right ventricle with wide QRS complexes is illustrated in Figure 5.11b. VT is a rapid and well-organized heart rhythm with a heart rate of more than 100 beats per minute (bpm). Sustained VT lasting more than 30 seconds or is symptomatic leads to polymorphic VT, which then degenerates to VF [305]. The coarse VF generated from various sites throughout the ventricle giving rise to multiple fast QRS complexes with uneven width and unclear R-peaks is illustrated in Figure 5.11c. VF is a fast, disorganized, and erratic cardiac activity where multiple heart cells trigger together, leading the heart muscles to quiver and generate smaller amplitude waveform with fibrillation frequencies.

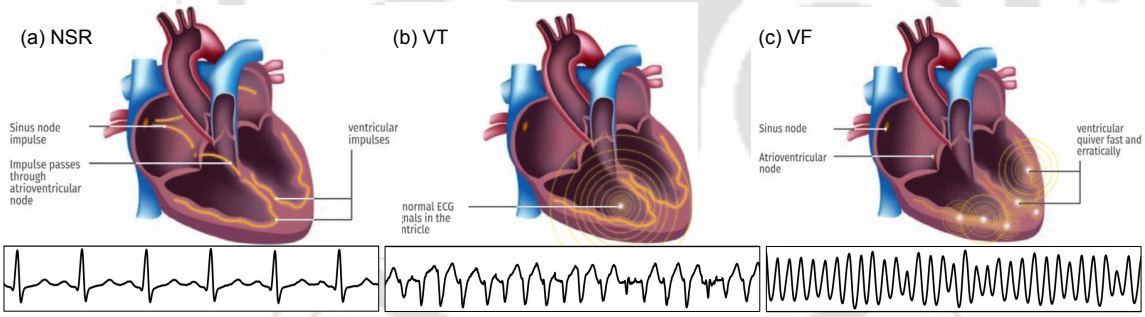


Figure 5.11: Illustration of (a) Normal Sinus Rhythm with impulse generated from Sinoatrial Node; (b) Reentrant monomorphic Ventricular Tachycardia with left bundle branch block morphology generated from the right ventricle; (c) Coarse Ventricular Fibrillation generated from various sites throughout the ventricle. (Edited from [5])

5.3.5 Correctly Diagnosed Rhythms

The saliency maps obtained from last convolution layer of ResNet model for correctly classified NSR, VT, and VF rhythms using GBP, Guided Grad CAM (GGC), and Grad CAM (GC) are illustrated in Figure 5.12, 5.13, and 5.14. The maps highlights essential areas responsible for/against diagnosis. Since, saliency maps were destroyed

during training data permutation test, only correct predictions from test data using overfit model are displayed. During correct classification of NSR in top row of Figure 5.12, 5.13, and 5.14, GGC and GBP find precise R-peaks whereas GC highlights the area from P-wave to R-peak. GC also highlights Q-wave with slight focus on T-wave. GBP work as R-peak detector and highlights areas of increasing slope in the waveform. During the sanity checks, the saliency maps were destroyed making the techniques suitable for interpretation. The middle row of Figure 5.13 and 5.14 describe correct classification of VT where GBP follows a similar behaviour as mentioned for NSR. It highlights increasing slope in the waveform. GGC also highlight slight R-peaks. All three techniques highlight wide QRS complex regions (seen in blue). The blue regions in VT are more as it is a faced paced rhythm. The saliency maps of GC were destroyed the most whereas the maps of other techniques retained some similarity with a focus on QRS complexes. Therefore, it can be inferred that they do not depend on train data as much as GC, making GC a better technique. Moreover, GBP again worked as a peak detector. The last row of Figure 5.12, 5.13, and 5.14 describe correct classification of VF where GC puts less focus on QRS complex but more on absent P-waves (basically before and after QRS complexes). GBP highlighted peaks present in ECG segment.

5.3.6 Incorrectly Diagnosed Rhythms

The saliency maps obtained from last convolution layer of ResNet model for miss-classified rhythms are illustrated in Figure 5.15. The miss classifications can be categorized into two broad categories. Firstly, the model is usually confused while classifying between VT and VF rhythms as sustained VT leads to VF. In later stages of VT the visibility of P-waves and T-waves start to diminish, making correct diagnosis difficult. Secondly, the fast heart rate around 150 bpm, confuses the model to classify between intermediate and shockable VT. In Figure 5.15a, the model predicted NSR as VT. The wide QRS complexes are highlighted in GC saliency maps, resulting the model to predict VT and not NSR. GGC and GBP highlight random signal timestamps. In Figure 5.15b, VT is classified as NSR. The GC maps highlight narrow QRS complexes and miss out evident P-waves, thereby predicting NSR. The GGC maps follows similar highlighted timestamps. GBP highlights only the increasing slopes. In Figure 5.15c coarse VF is predicted as NSR but GC focuses on narrow QRS complexes that are present in NSR. Saliency maps highlight TP segment rather than focusing on wide QRS complexes. GGC and GBP maps highlight random timestamps thereby

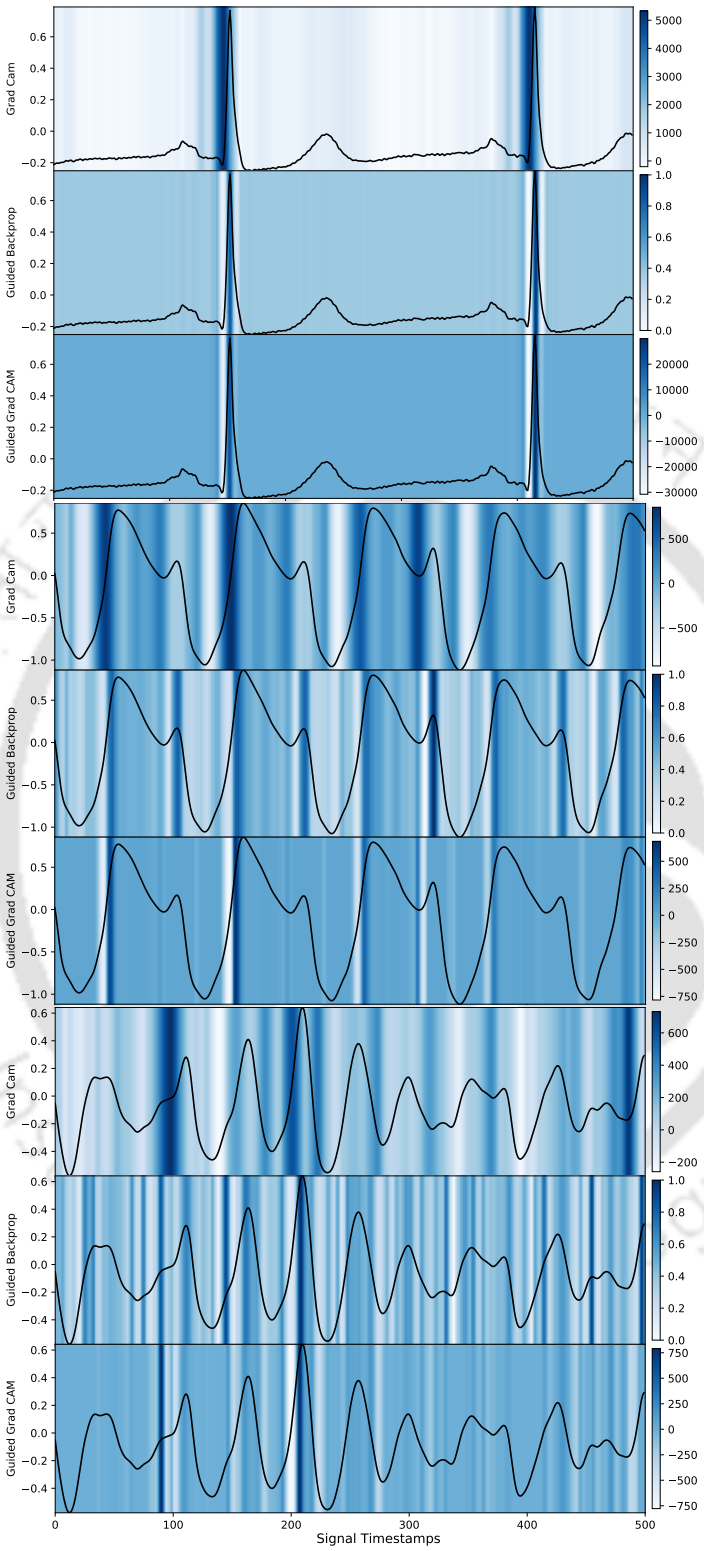


Figure 5.12: Original Saliency maps of correctly classified episodes. Top row: NSR, Middle Row: VT, Bottom Row: VF.

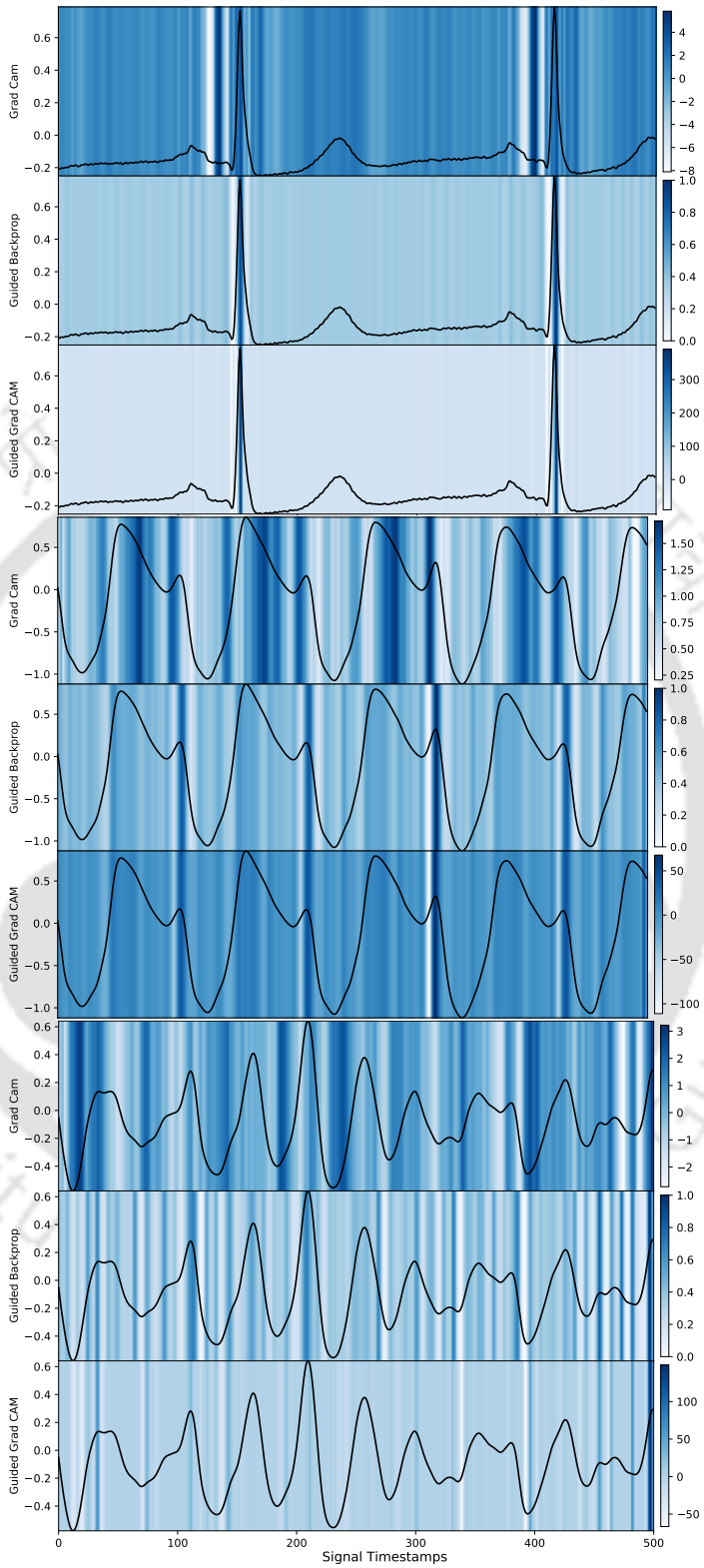


Figure 5.13: Saliency maps with Randomized Weights of correctly classified episodes. Top row: NSR, Middle Row: VT, Bottom Row: VF.

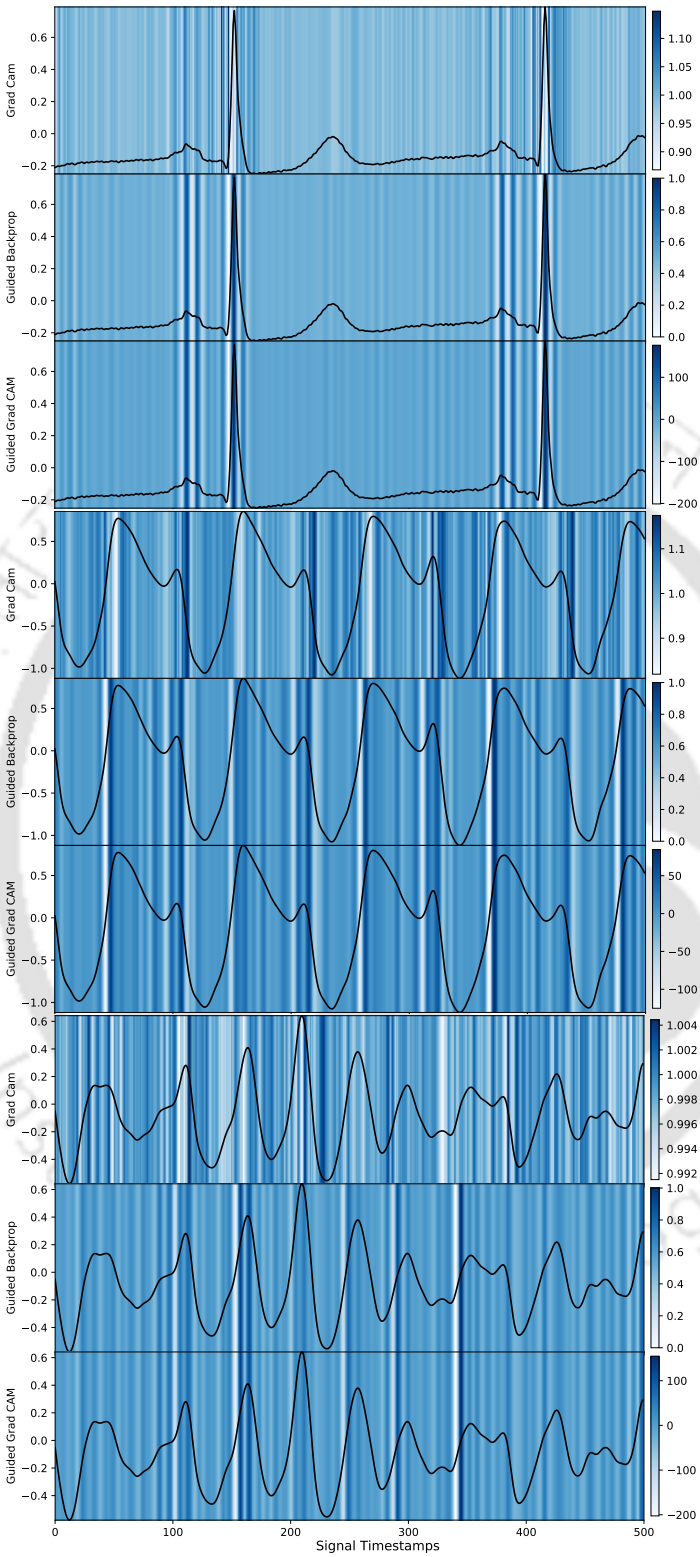


Figure 5.14: Saliency maps of Permuted Data-Label combination for correctly classified episodes. Top row: NSR, Middle Row: VT, Bottom Row: VF.

lacking the focus on specific morphologies. In Figure 5.15d NSR is predicted as fine VF. The amplitude of the signal is very low, ranging from $[0.9, 1.1]$. GGC shows clear demarcation on QRS. GGC and GBP focus on QRS complexes and two beats are present, causing a heart rate of 60 bpm. Low focus is provided to P and T-wave occurring in NSR and missed in the VF. In Figure 5.15e sustained VT is predicted as VF. GC maps highlight high number of wide QRS complexes. GBP maps highlight increasing slope regions. GGC maps focus on QRS complexes. The HR is 240 bpm as 8 beats are captured in 2 second ECG segment leading to classification of sustained VT as VF. Moreover, sustained VT leads to VF. Lastly, Figure 5.15f classifies VF as VT. The HR is 180 bpm. GGC focus mainly on R-peaks and work as peak detector and provide less focus on QRS complex. GC also provides slight focus on QRS complex. GC focuses on sides of QRS complex and classify VF as VT.

5.3.7 Interpretability Technique Comparison

The comparison of posthoc interpretability techniques is provided in Table 5.6. Although the techniques rely on trained model weights and relation between training data-label combination, GC deteriorated the most, whereas GBP and GGC mostly worked as peak detectors rather than providing reliable explanations. Therefore, GradCAM is the best technique followed by Guided GradCAM and Guided Backpropagation for generating saliency maps.

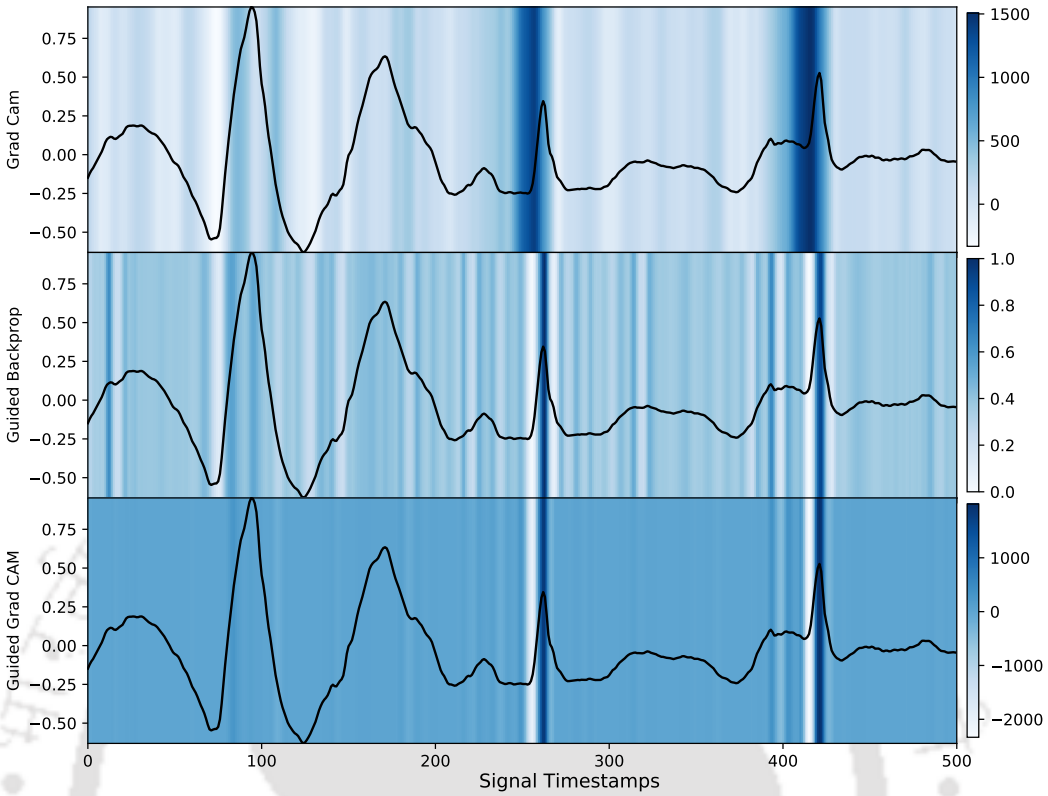
Table 5.6: Interpretability Technique Comparison

Sanity Checks	**GC**	**GGC**	**GBP**
Randomized Model Weights	✓	✓	✓
Permuted Train Data-Label	✓	✓	✓
Rank	1	2	3

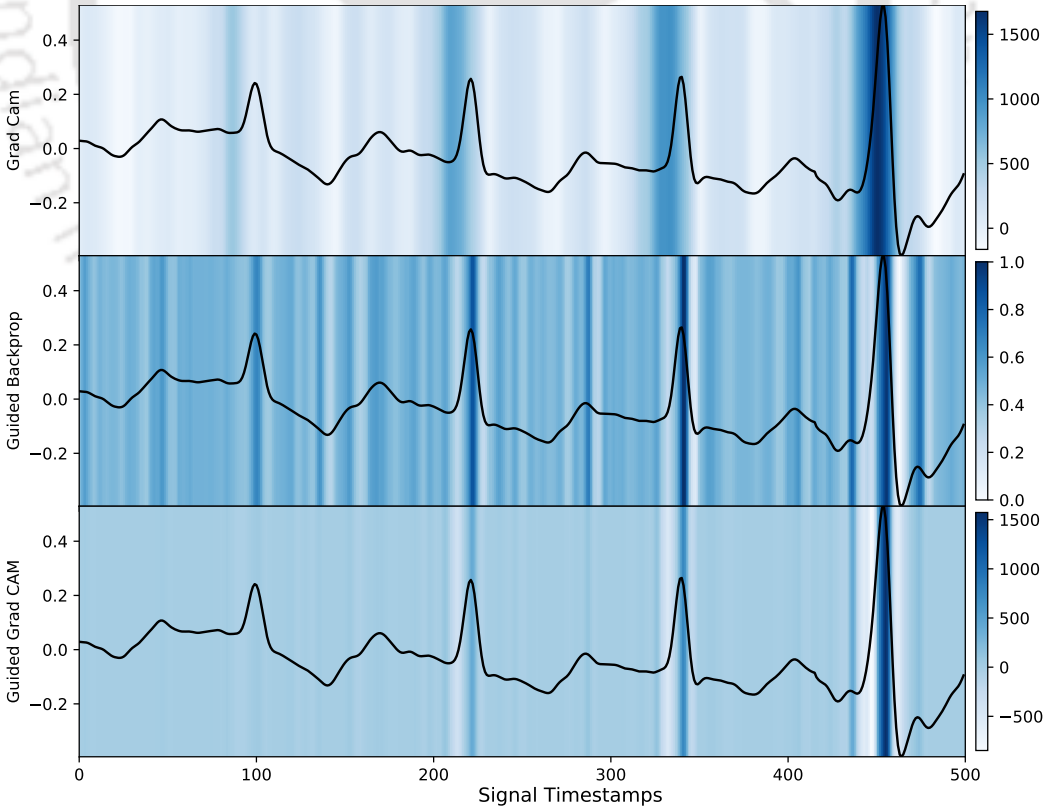
5.4 Summary

In this chapter, gradient backpropagation based posthoc interpretability techniques are investigated for explaining ResNet model predictions. The ResNet models detect VT, VF, and NSR from ECG signal. The effect of layers of the ResNet model, augmentation technique, and segment length were analysed and the optimum performance was obtained for 1-layer ResNet for 1 second segment using ADASYN augmentation technique. The models were tested on datasets acquired from mul-

tiple geographical locations. The interpretability techniques, Grad CAM highlighted relevant timestamps whereas GBP and Guided Grad CAM mostly worked as peak detectors which aid in heartbeat detection. The interpretability techniques were verified using sanity checks, namely, model weight randomization and training data label permutation test. All techniques passed the sanity checks and but Grad CAM saliency maps deteriorated the most, making it a better choice for signal interpretation. In conclusion, 1-layer ResNet with ADASYN augmentation and Grad CAM highlighted most relevant signal timestamps responsible for prediction. The next chapters highlight the importance of implicit interpretability techniques for ECG classification and interpretation.

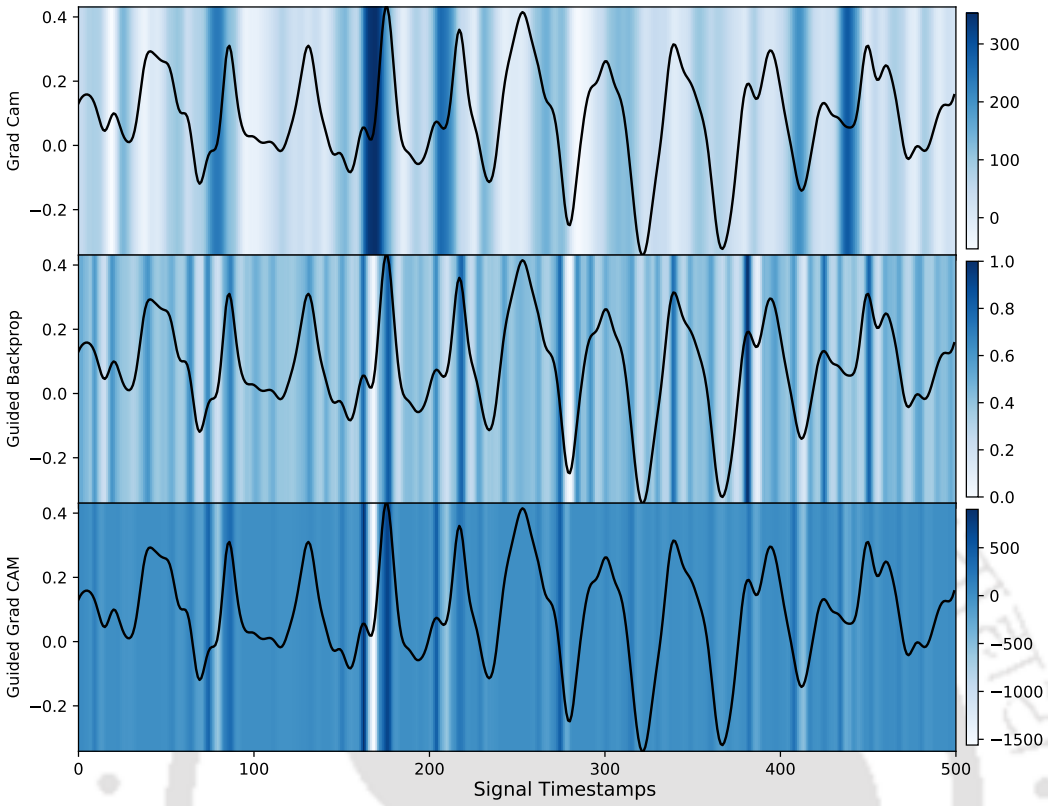


(a) NSR predicted as VT.

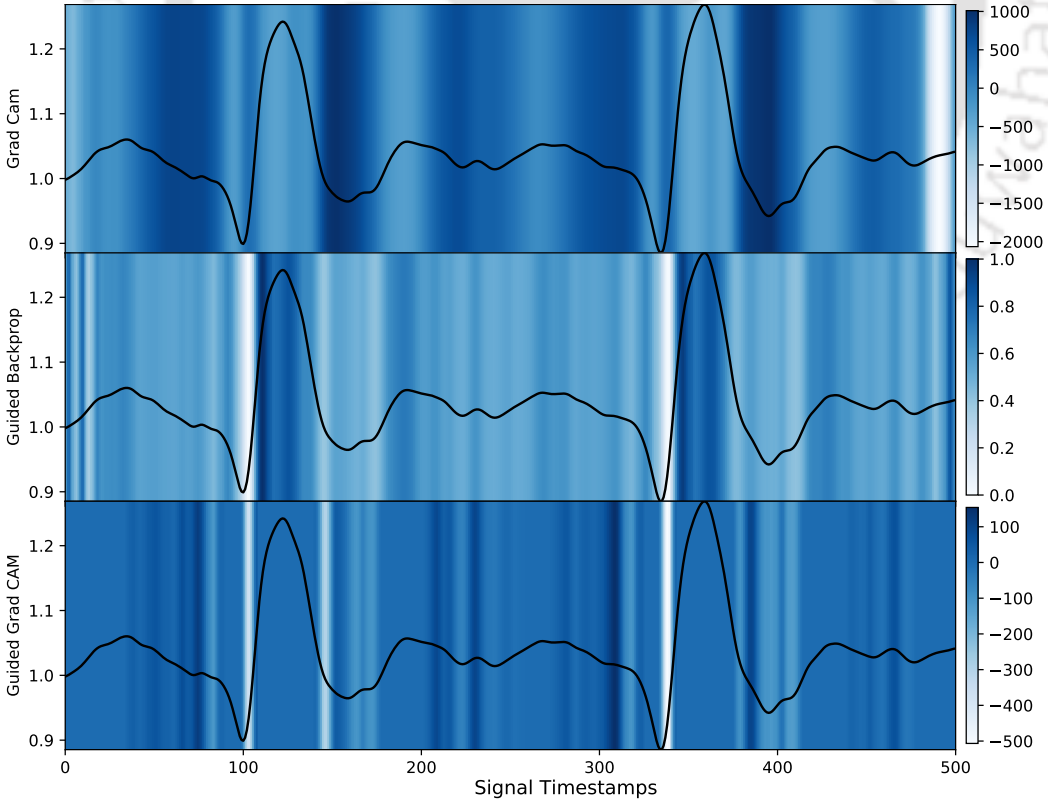


(b) VT predicted as NSR.

Figure 5.15: Saliency maps of incorrectly classified NSR, VT, and VF rhythms.

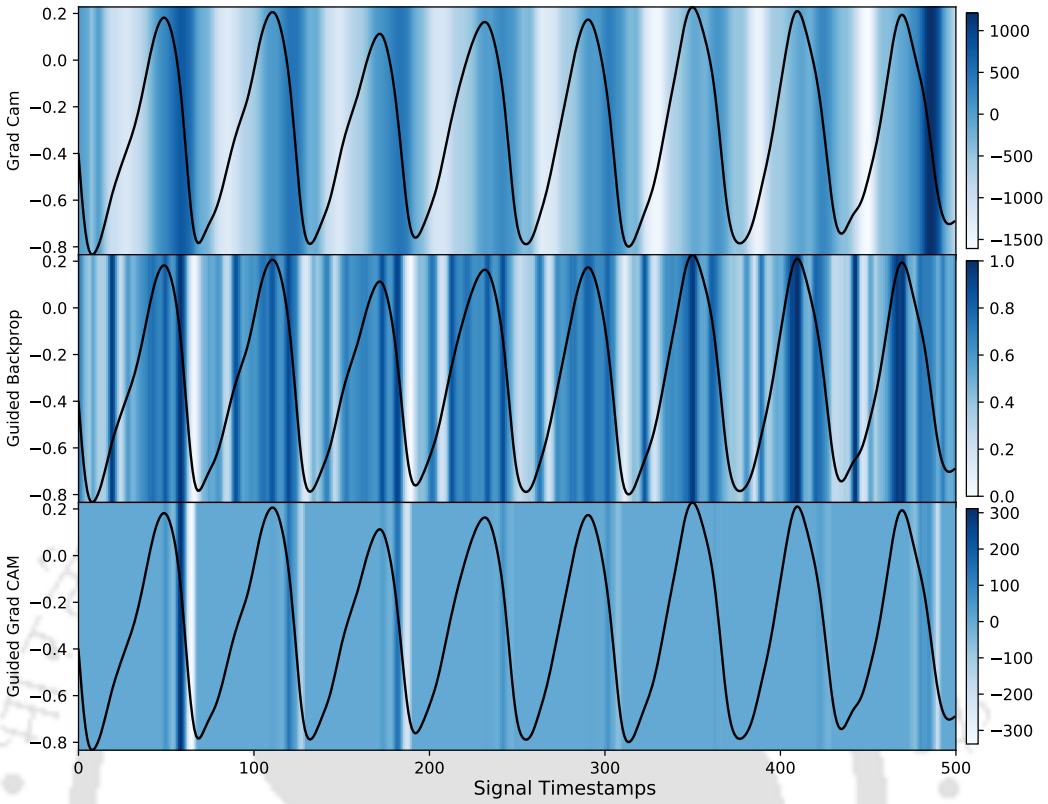


(c) Coarse VF predicted as NSR.

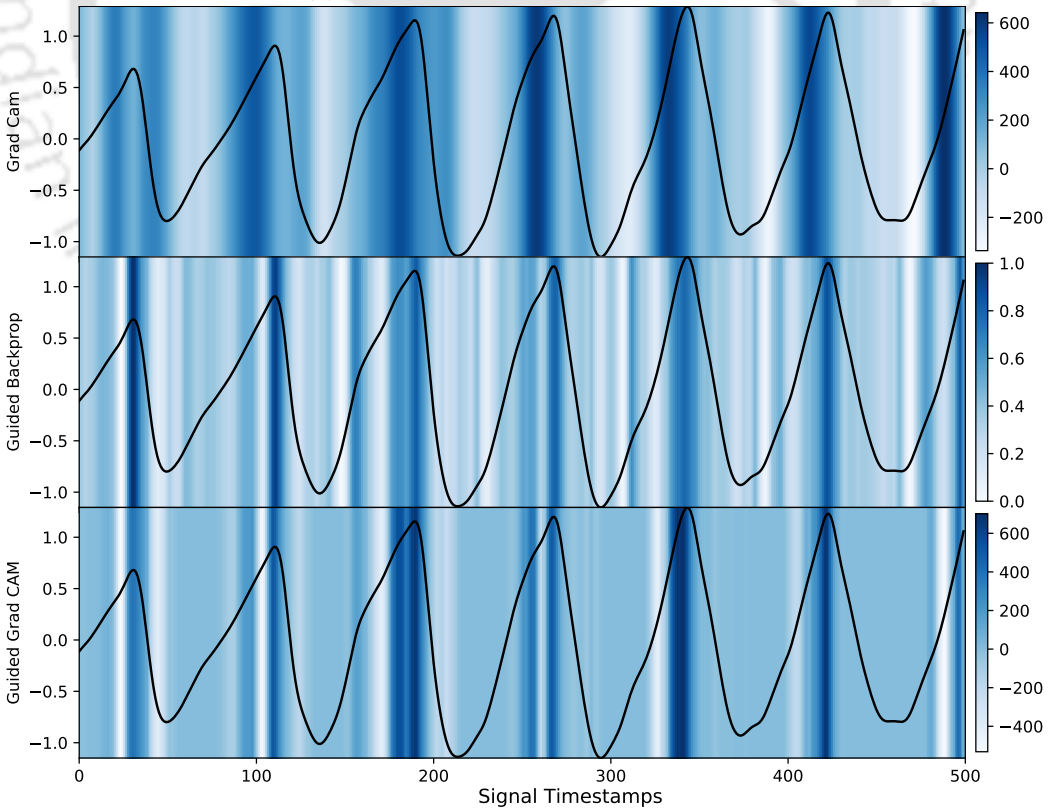


(d) NSR predicted as Fine VF.

Figure 5.15: Saliency maps of incorrectly classified NSR, VT, and VF rhythms.



(e) Sustained VT predicted as VF.



(f) VF predicted as VT.

Figure 5.15: Saliency maps of incorrectly classified NSR, VT, and VF rhythms.

6

Classification and Prediction Interpretation of Atrial Fibrillation using Attentive Transformer Neural Network

The World Health Organization reported that around six million people in the United States and around ninety million people worldwide suffer from Atrial Fibrillation (AF) and these numbers will become two-fold till 2050 [306, 307]. AF can be detected using the Electrocardiogram (ECG) signal clinically. AF occurs due to disorganized electrical activity at the upper left chamber of the heart (left atrium) making the atria quiver or fibrillate. This phenomenon introduces a chaotic activity between the QRS complexes during which the P-waves are absent and are replaced by f-waves. The absence of P-wave signifies no PR interval and presence of f-waves produces irregular RR intervals. The heart acts as a pump and sends blood throughout the body, and during AF, the blood may form a clot which is passed during blood circulation and causes stroke when the clot reaches the brain [308]. People having AF suffer from a higher risk of having a stroke and since it is sometimes asymptomatic, it is mostly diagnosed after the patient faces stroke or heart failure. The occurrence of AF episodes might be sustained or intermittent making its diagnosis challenging during routine clinical visits [309]. In this chapter, the interpretability aspect of deep learning models is investigated through an attention-based transformer neural network model. For comparison, baseline Convolution Neural Network, Residual Neural Network, Attentive Convolution Neural Network are also investigated. The importance of segment length is analyzed by varying ECG segment length from 1 second to 10 seconds. The models are trained, validated, and tested on PhysioNet Computing in Cardiology 2017 Challenge Database. Attentive Convolution Neural Network

and Transformer Neural Network provide explanations by highlighting clinically relevant features such as absent P-wave and Fibrillation wave, supporting the model predictions.

The chapter organization is as follows. Section 6.2 describes the methodology which includes dataset description, data preprocessing, investigated models, and the interpretation mechanism used to generate explanations. Section 6.3 describes the experiments and results followed by Section 6.4 which summarises the chapter.

6.1 Literature Review

Conventional AF detection is performed through visual inspection by nurses, physicians, and cardiologists. However, the increasing number of CVD subjects generating vast amounts of ECG increases workload of medical professionals, making manual ECG inspection time-consuming, tedious, and inefficient. This necessitates the development of automated AF detection for accurate and efficient diagnosis. Clinically, AF is detected by observing the absence of P waves [310] and irregular RR intervals [311] in the ECG. Algorithms using P wave morphology are prone to noise artefacts [312]. Clifford et al. [6] organised the physionet computing in cardiology challenge in 2017 that aimed towards AF classification using single lead ECG signals. The approaches could be categorised into the following three categories.

1. Feature extraction and classification using ML techniques [313–316]: The feature extraction methods extract features from ECG, followed by classifiers [313]. Several methods detect P, QRS, T waves and calculate heart rate variability (HRV) features, detect atrial activity by the presence of a P-wave and f-waves during TQ intervals [314], template-based wave morphology features [315]. Garcia et al. exploited the variability in ventricular and atrial activities reflected on ECG by extracting RR intervals and fibrillatory waves contained in the TQ intervals and estimated Coefficient of Sample Entropy and classified using multiclass support vector machine (SVM) [317]. Combination of HRV, RR interval, PR interval, P-wave and f-wave features have also been used for AF detection [318–320]. Datta et al. [316] also extracted RR intervals, HRV features, frequency domain features, statistical features and morphological features from detected PQRST points. The features were selected using Maximal Information Coefficient and minimum Redundancy Maximum Relevance and classified using adaptive boosting, an ensemble learning approach.

2. Classification using Deep Learning Approaches [263,321,322]: The Neural Network approaches are able to achieve better performance for end-to-end AF detection using Convolutional neural networks (CNN) Recurrent Neural Network (RNN), and Residual Neural Network (ResNet) [263]. Tran et al. [322] combined fully connected layer and residual blocks to detect AF. The hybrid methods use an R-peak detector for heartbeat extraction from single lead ECG followed by CNN [321]. Hong et al. [323] extracted features from centerwave and classified using an ensemble deep neural network classifier. Zihlmann et al. [324] employed a convolutional recurrent neural network and used log spectrogram of ECG signal as input to the neural network. Recently, Murat et al. [325] published a review of deep learning-based AF Detection that cohesively describes the recent advancement of neural network models for AF detection.
3. Hybrid methods: These methods extract features in tandem with deep learning models. Hong et al. [323] used a combination of expert features, including statistical features, medical features, center wave frequency features extracted using the dynamic time warping technique, and features extracted using a deep neural network for AF detection. Although significant performance is achieved by deep learning and hybrid models, they lack interpretability which inhibits their real-world usage.

The shortcomings of these methods is that they use a fixed set of handcrafted features that might fail in practical scenarios. This happens due to the drastic change in ECG rhythm either from AF to normal or normal to AF as they suffer from high variations. Such features might not even be sufficient to perfectly characterize an ECG signal. Although significant performance is achieved by deep learning models, they lack interpretability.

Researchers have developed interpretable techniques such as Attention mechanism [142], Transformer Neural Network (TNN) [143], shapley additive explanations (SHAP) [326] and class activation map (CAM) [327] to mitigate the black box nature of deep learning models and gain better insight into the relationships learned by these models. Duan et al. [144] employed attention mechanism followed by deep neural network for arrhythmia classification. Mousavi et al. [145] employed a bidirectional recurrent neural network followed by three attention mechanism levels, wave level for computing wave weights, beat level for computing beat weights, and window (multiple heartbeats) level to compute window weights, responsible for AF detection. TNN adopts attention mechanism and captures long-distance signal dependencies and identify relevant signal timestamps [143]. Yan et al. [146] fused transformer

encoder with RR interval features for heartbeat classification whereas Natarajan et al. [149] used transformer encoder and fully connected layer for multichannel ECG classification. Rouhi et al. [326] employed SHAP to rank important features based on relevance. Vijayranjan et al. [327] used global average pooling and softmax activation to visualize CAM. Yao et al. [148] employed an attention based time incremental convolution neural network that fused spatial and temporal information present in ECG.

6.2 Methodology

The methodology is described in Figure 6.1. Initially, the dataset is preprocessed and provided to deep learning models for classification. The models investigated for AF detection include Baseline Convolution Neural Network, Residual Neural Network, Attentive Convolution Neural Network, and Transformer Neural Network. The trained models are later used to perform inference and generate interpretations for model predictions.

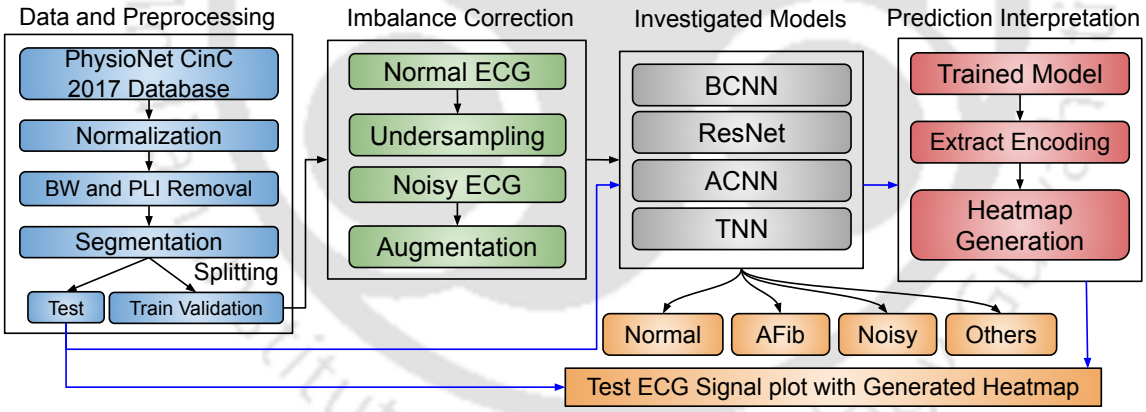


Figure 6.1: Overall methodology encompassing data preprocessing followed by modelling and interpretation mechanism.

6.2.1 Dataset Description and Preprocessing

The PhysioNet/CinC Challenge 2017 Database [6] is used for experimentation that consists of 8528 single lead annotated ECG recordings that were acquired and band-pass filtered by an AliveCor device: a portable ECG monitor. The database consists of four classes (recordings): AF (771), normal (5154), other (2557) and noisy (46) rhythms of lengths between 9 and 60 seconds (sec.) sampled at 300Hz. An instance

of each class is depicted in Figure 6.2. The normal class rhythms contain the P-wave, QRS complex and T-wave morphology. In AF, the absence of P-wave with irregular RR interval is reflected. The noisy rhythm overshadows the ECG waveform due to the presence of low and high frequency of noises. The other rhythms include all those that do not belong to other categories.

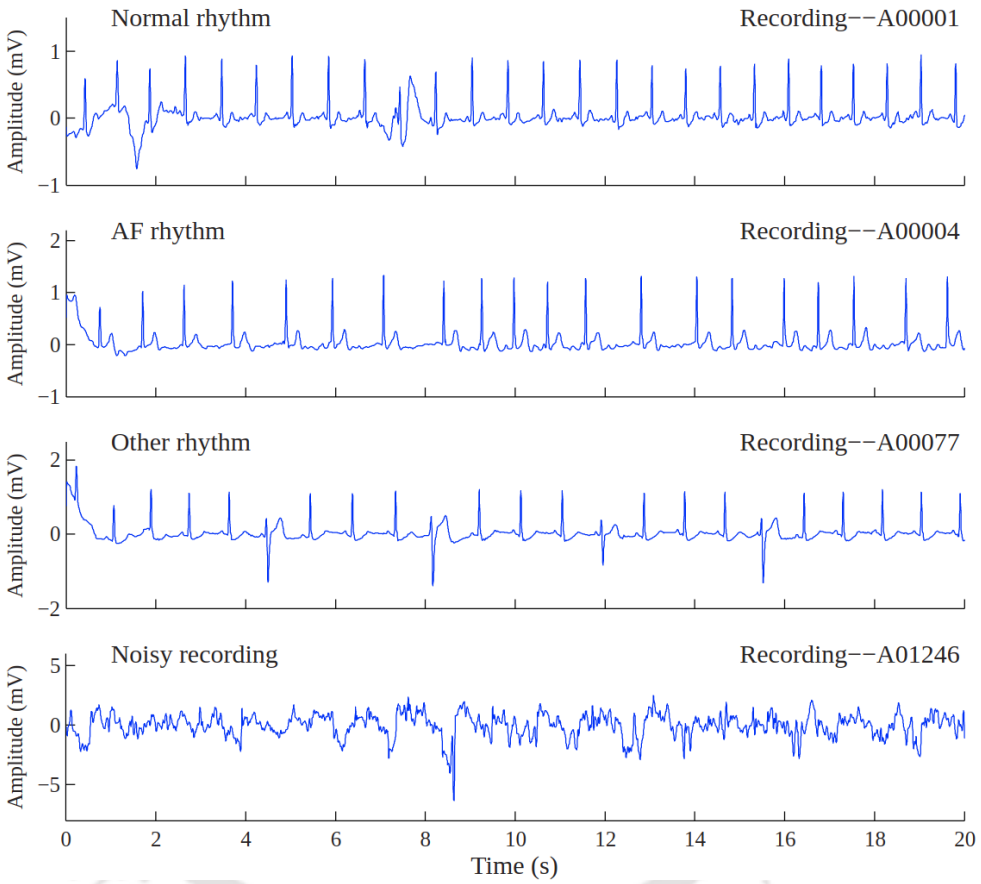


Figure 6.2: Illustration of different rhythms [6].

Preprocessing: The raw ECG recordings are normalized by subtracting the mean and dividing by maximum amplitude of the record. The records are contaminated with high-frequency noise or power line interference and low-frequency noise or baseline wander (BW) caused due to respiration, the motion of the subject and dirty or loose electrode. BW introduces a gradual increase ECG amplitude and hinders the analyses of ST segment morphology. Therefore, a bandpass Butterworth filter with the frequency range of $[5, 45]$ Hz was employed to eliminate these noises. Then segmentation is performed to extract equal length segments ranging from length of 1 second to 10 seconds. The segment length plays a major role in evaluating the performance of the model and it has not been much explored in the literature as per the

best of our knowledge. After the segment division, one-hot encoding was performed before feeding the data to the neural network model.

Imbalance Correction: The data of the noise class was removed prior to modelling in the literature, but we decided not just to keep it but rather augmented it to make the model more robust towards random perturbations. Our objective of developing an end-to-end framework relies negligibly on the signal processing methods and majorly on deep learning methods. Since, the minority class data (Noise class) is significantly less as compared to majority class data (Normal Class), the classifier becomes biased towards the majority class in producing results. Possible solutions to the problem of data imbalance in the literature consists of: (i) Collection of minority class data; (ii) Resampling: either under-sampling (loss of important information) or over-sampling (reduces decision boundary of the minority class data) or both to equate the number of samples in each class; (iii) Synthetically generating minority class samples. The noisy rhythms were augmented using the adaptive synthetic technique (ADASYN) [44] to make the model robust towards random perturbations. ADASYN does not introduce a bias towards noisy rhythms and increases the decision boundary. Random undersampling was also performed over the normal class data to reduce it in half of its original size so that the number of instances in each class are almost equal. The final number of segments in each rhythm class are described in Table 6.1. The extracted segments are split into training, validation, and test sets, classified using the deep learning models.

Table 6.1: Final number of segments in each rhythm class.

Length	**Original**				**Final**			
(Sec.)	**Other**	**AF**	**Noisy**	**Normal**	**Other**	**AF**	**Noisy**	**Normal**
1	82822	49033	6801	108702	82822	49033	34005	81498
2	41411	24516	3400	54349	41411	24516	17000	40749
3	27607	16343	2267	36234	27607	16343	11335	27166
4	20705	12257	1700	27174	20705	12257	8500	20374
5	16564	9805	1360	21739	16564	9805	6800	16299
6	13803	8171	1133	18115	13803	8171	5665	13583
7	11831	7003	971	15526	11831	7003	4855	11642
8	10352	6128	850	13587	10352	6128	4250	10187
9	9202	5447	755	12075	9202	5447	3775	9055
10	8282	4902	680	10869	8282	4902	3400	8149

6.2.2 Baseline Convolution Neural Network

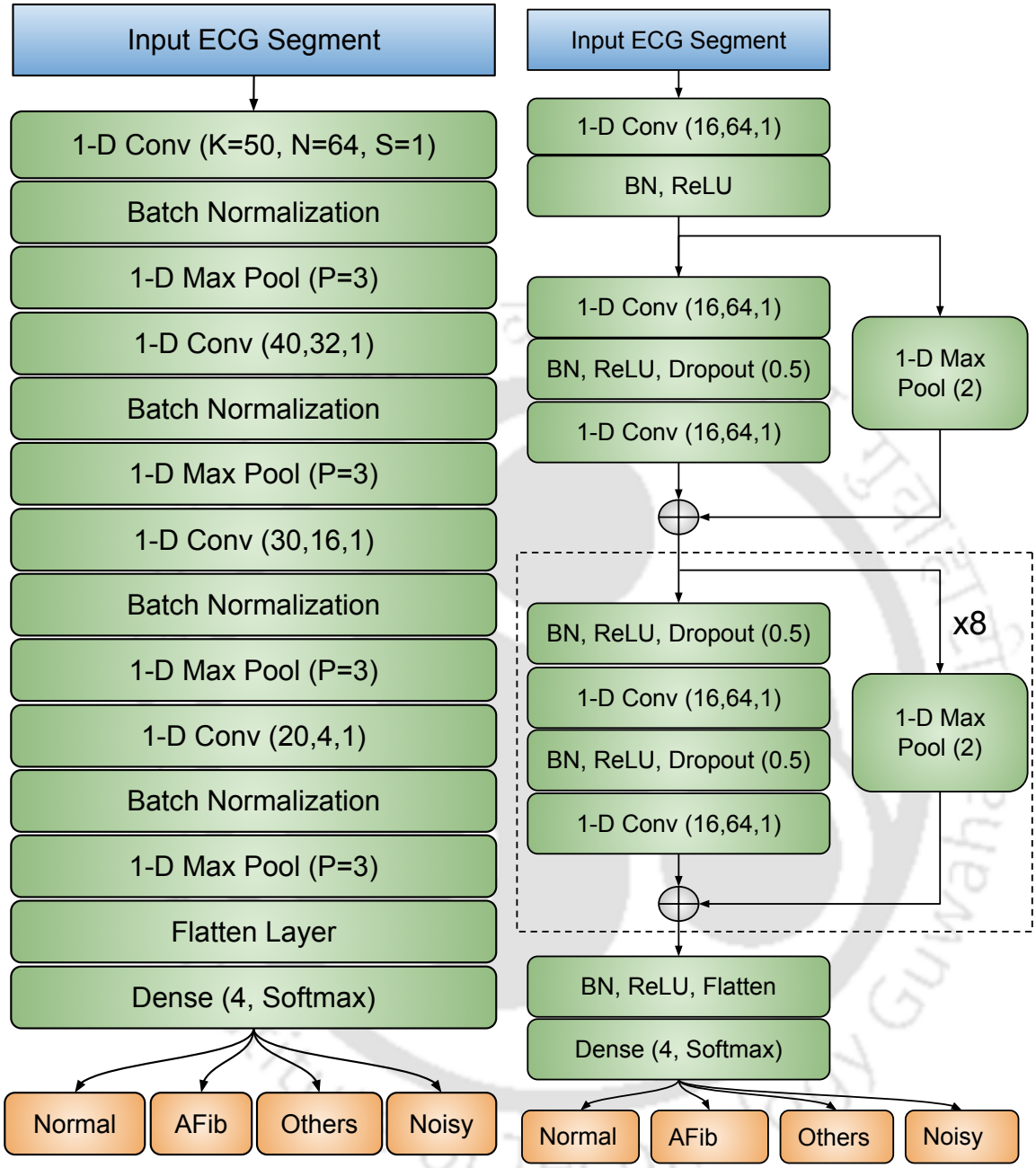
The basic details of a typical Baseline CNN (BCNN) are described in Section 4.3.3. The BCNN architecture is illustrated in Figure 6.3a. The model encompasses four 1-D convolution layers, each followed by batch normalization (BN) [274] and Maxpooling [275] layer. The convolution layer has the following parameters: Number of filters (N), Kernel Size (K), and Stride (S).

6.2.3 Residual Neural Network

Increasing the size and number of layers does not always optimize the neural network, it instead deteriorates its performance due to overfitting and suffers from problems such as the notorious vanishing and saturation of the gradient. To tackle this, Residual Neural Networks (ResNet) was invented that introduced the concept of "identity shortcut connection", that controls the amount of information allowed to flow across the shortcut [277]. It solves the problem of vanishing and exploding gradients, by reusing activations from the previous layers. The shortcut connection simplifies the network by using a few layers and less backpropagation thereby reducing the model size and execution time. Without these identity mapping, the model explores more feature space and leaves the original hypothesis and leads to performance deterioration and necessitates the use of extra training data for the model to recover. ResNet was implemented using 8 layers as depicted in Figure 6.3b.

6.2.4 Attentive Convolution Neural Network

The Attentive Convolution Neural Network (ACNN) employs an additional self-attention block after the stack of convolution, batch normalization, and ReLU activation layers. The ACNN architecture is provided in Figure 6.4. The attention approach is adapted from Luong-Style Attention, i.e., multiplicative or dot-product attention [328]. The intuition is to capture relative importance or attend the signal encoding globally. In the absence of a mechanism that highlights the salient information across the entirety of the encoding, important information may not be effectively focused. The attention mechanism addresses this, especially for larger ECG segments. The ECG segment provided as input generates encoding of length similar to input ECG as the convolution filter stride is unity. The last convolution filter produces encoding of size (S) for F filters, represented as $C^{F}_{S\times 1}$, which is provided to attention block that calculates self-attention through the following equations.



(a) Baseline CNN (b) 8-layer ResNet

Figure 6.3: Illustration of the Model Architectures.

1. Alignment score ($AS^F_{S\times S}$) is calculated using the dot product of CNN filter outputs ($C^F_{S\times 1}$) using Equation 6.1. The scores indicate how well the encodings align with the current output.

$$AS^F_{S\times S} = C^F_{S\times 1} \cdot C^F_{1\times S} \tag{6.1}$$

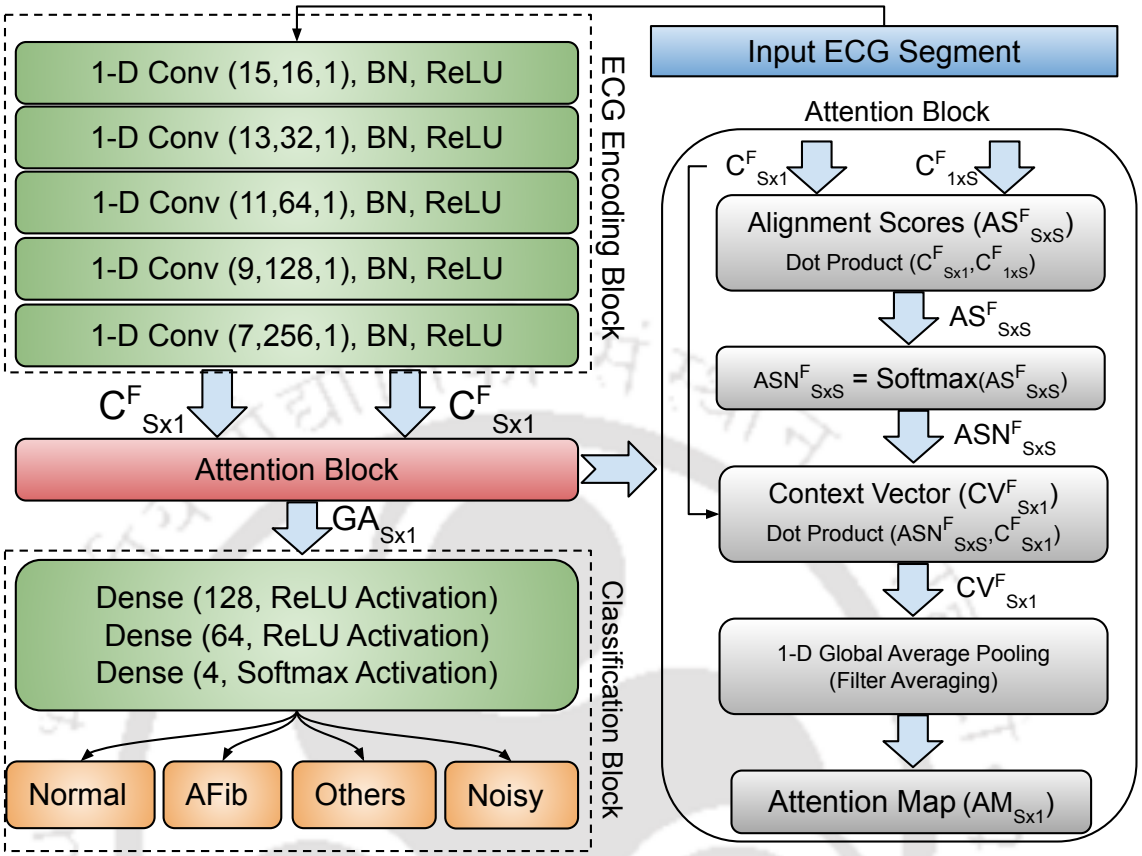


Figure 6.4: Attentive Convolution Neural Network

2. Then the relative importance between signal encoding is calculated by passing $AS^F_{S\times S}$ through softmax function to generate the weights, representing probabilities using Equation 6.2.

$$ASN^F_{S\times S} = \frac{\exp\left(AS^F_{S\times S}\right)}{\sum_{s=1}^{S}\exp\left(AS^F_{S\times s}\right)} \tag{6.2}$$

3. The probabilities scaled with encodings generate a context vector using Equation 6.3.

$$CV^F_{S\times 1} = ASN^F_{S\times S}\cdot C^F_{S\times 1} \tag{6.3}$$

4. The context vector across all filters (F) are averaged using the 1-D global average pooling layer to generate the Attention Map ($AM_{S\times 1}$) using Equation 6.4.

$$AM_{S\times 1} = \sum_{f=1}^{F} CV^f_{S\times 1} \tag{6.4}$$

The attention map outputs are provided to the Classification block consisting of two fully connected layers with ReLU activation followed by one fully connected layer with softmax activation function for classification. The attention map highlights relatively important signal encodings responsible for generating the output prediction class for the input ECG.

6.2.5 Transformer Neural Network

The last investigated model is transformer neural network (TNN) that captures long and short-term signal dependencies to identify important signal timestamps with high training efficiency [143]. TNN architecture depends on the parallel self-attention mechanism that learns global dependencies between input and output. The multi-head self-attention blocks introduce parallelism in the model. TNN consists of an encoder block and decoder block, we employed only the encoder. The architecture of TNN is illustrated in Figure 6.5.

The input ECG is provided to the encoding block, and the output of the last convolution filter followed by BN and ReLU activation is added with positional encoding (PE) output and provided to the transformer block. The PE block also uses the convolution filter output and establishes the relationship between the timestamp and its relative position in the encoded output. PE is computed using Equation 6.5, where, pos is position, i is encoded dimension, and d_{model} is the number of convolution filters in last layer.

$$\begin{aligned} PE_{(pos,2i)} &= \sin\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right) \\ PE_{(pos,2i+1)} &= \cos\left(\frac{pos}{1000^{\frac{2i}{d_{model}}}}\right) \end{aligned} \tag{6.5}$$

The output is provided to the multi-head attention (MHA) module as Query (Q), Key (K), and Value (V). MHA allows the model to jointly attend the information from different representation subspaces at different positions, whereas the single attention inhibits this behavior. Rather than only computing the attention once, the multi-head mechanism runs through the scaled dot-product attention multiple times in parallel. MHA uses Q, K, and V to compute independent attention outputs that are concatenated to generate context vector (CV) using Equation 6.6. Here, Q, K, V use

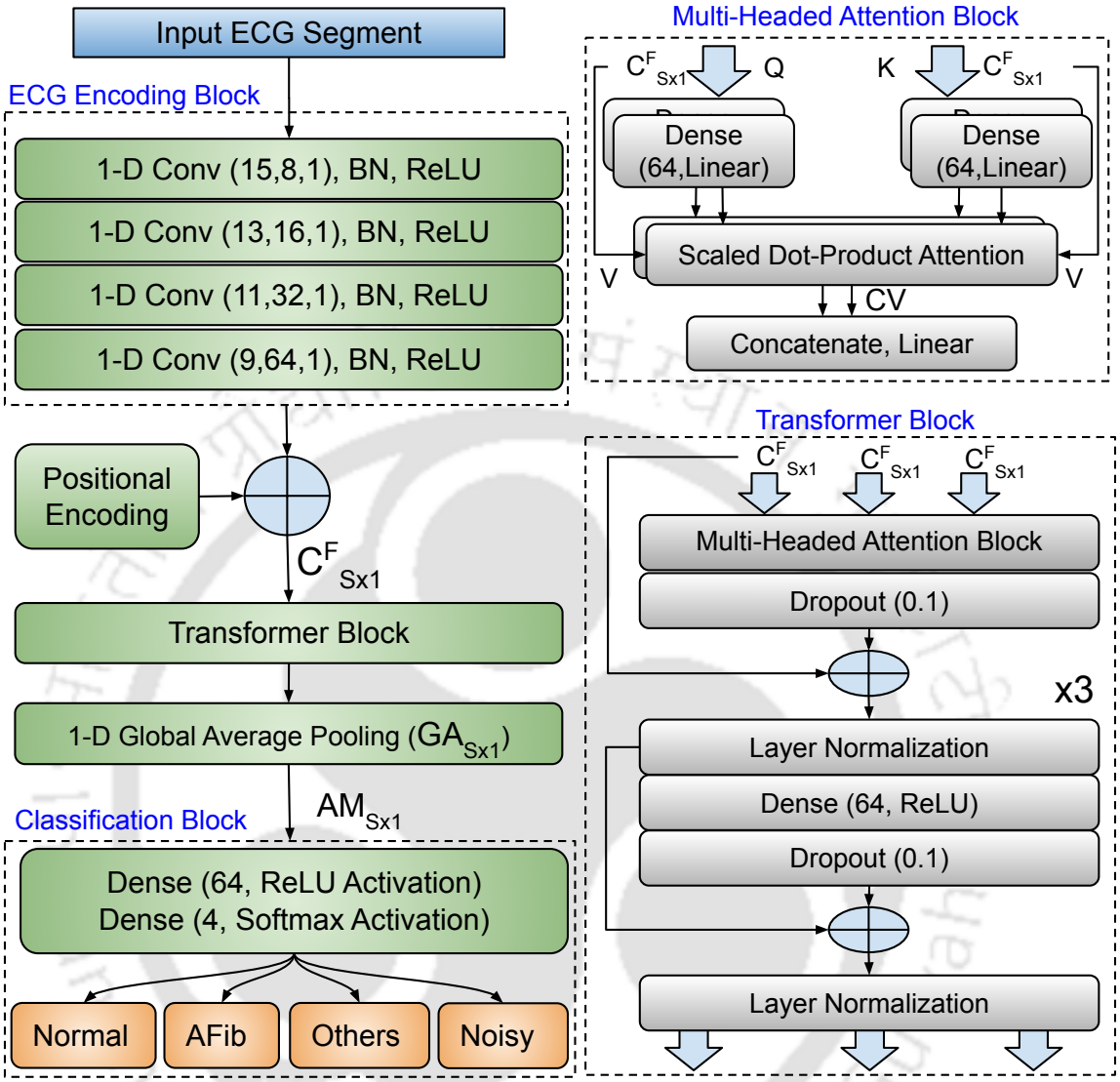


Figure 6.5: Transformer Neural Network

$C^F_{S\times 1}$.

$$CV = softmax\left(\frac{QK^T}{\sqrt{d_{model}}}\right) \times V \tag{6.6}$$

The output of the MHA module is provided to the dropout layer, and the output of the dropout layer is added with MHA output through an additive residual connection and passed through a normalization layer. The normalization layer is similar to batch normalization, but instead of considering the whole minibatch of data for calculating the normalization statistics, all hidden units of the same layer are considered, overcoming the drawback of estimating the statistics for the summed input to any neuron over a minibatch of the training samples. The output is provided to

a feedforward layer (Dense, ReLU, Dropout). This mechanism is repeated several times, and the final output is provided to a 1-D GAP layer that performs filter averaging to generate an attention map. A classification block follows the attention map to classify the ECG segment.

6.2.6 Interpretation

The trained models, ACNN and TNN, are used during inference to predict the ECG segments, and the output of the 1-D GAP layer ($AM_{S\times1}$) is extracted to describe the contribution of essential timestamps responsible for the prediction. The output of the 1-D GAP layer is similar to the input ECG segment due to the convolution filter stride of unity. The self-attention map is plotted against the original signal to visualize the important regions for the corresponding ECG segment.

6.3 Experiments and Results

The system configuration that we used during experimentation had 32 GB of RAM, an Intel Xeon Processor and a NVIDIA TITAN Xp GPU. All the codes were implemented using Keras [329] library in python. Since, the test set is not publicly available, the models are verified on publicly available validation set. The training dataset is split into two parts: training set (80%) and validation set (20%). Validation set is used to tune the hyperparameters of the model such as convolution filter size, number of filters, pooling stride and dropout probability. All the convolution kernels are initialized by Xavier initialization [330] that initializes the weights through a Gaussian distribution to maintain activation and gradient variances. The model performance is measured using Accuracy (Acc), Sensitivity (Se), Specificity (Sp), and Precision (Pr) as described in Equation 4.3.2.

6.3.1 Quantitative Performance Evaluation

The model architectures with internal details are described in their respective sections. The MHA parameters in TNN were tuned to achieve high F1 scores by varying number of heads and layers with and without PE as described in Table 6.2 for three second ECG segment. It is evident that PE is essential, as merely increasing the number of layers without PE deteriorated model performance. The best F1 score was 0.8157 without PE with single layer, whereas, with PE, best F1 score was 0.8722 with three layers and two heads. The other parameters of TNN are as follows: channels in

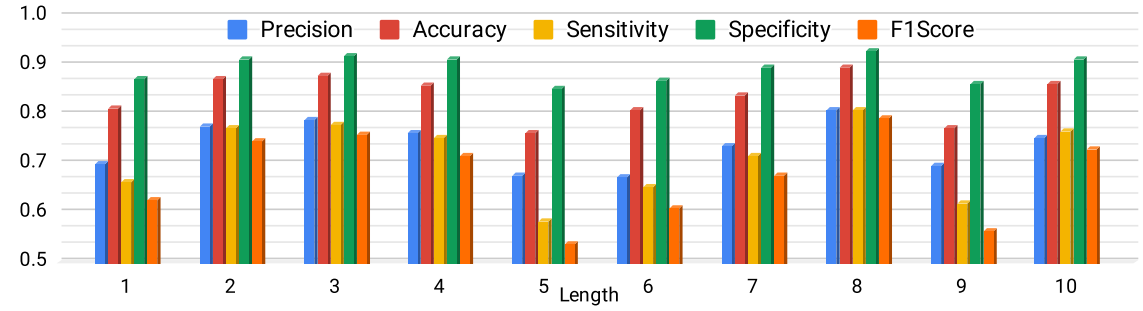
convolutional layers ($d_{model} = 64$), hidden layers in FFNN ($d_{ff} = 64$) and dimension of key/query/value depend on the convolution filter output.

Table 6.2: TNN parameter selection for three second ECG signal.

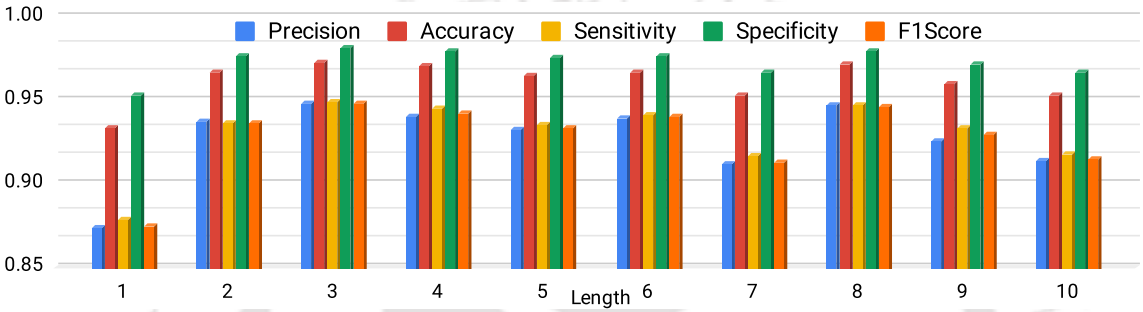
Layers	Heads	F1 without PE	F1 with PE
1	**2**	**0.8157**	0.7335
1	3	0.6746	0.7447
1	4	0.6500	0.6917
2	2	0.7576	0.7428
2	3	0.7272	0.7793
2	4	0.6911	0.8047
3	**2**	0.5500	**0.8722**
3	3	0.5500	0.5500
3	4	0.5500	0.5500

The performance of all the models using different segment length (ranging from 1 to 10 seconds duration) ECG signals is provided in Figure 6.6 using Precision (Pr), Accuracy (Acc), Sensitivity (Se), Specificity (Sp), and F1 Score (F1). All the models were trained for 100 epochs during the training phase and were tested on the validation set. It can be observed that ResNet achieved the highest performance among all the models for all length ECG signals followed by TNN, BCNN, and ACNN as can be seen in Figures 6.6b, 6.6d, 6.6a, and 6.6c, respectively. The optimum length of ECG segment that resulted in best model performance was found between 1 to 3 second length ECG. The results show a clear improvement in performance when the signal length was increased from 1 to 3 seconds and the performance deteriorated for larger length ECG signals. In longer segment lengths, the number of instances were relatively inadequate thereby obstructing the model to familiarize itself with the training set and generalizing on the validation set. The lower performance in 1 second ECG was because the segment size does not contain enough information to allow the model to distinguish between the four classes and due to the fact that AF detection is comparatively easier for relatively larger length ECG signal clinically. Table 6.3 describes the overall F1 score of all the models for varying length ECG signals.

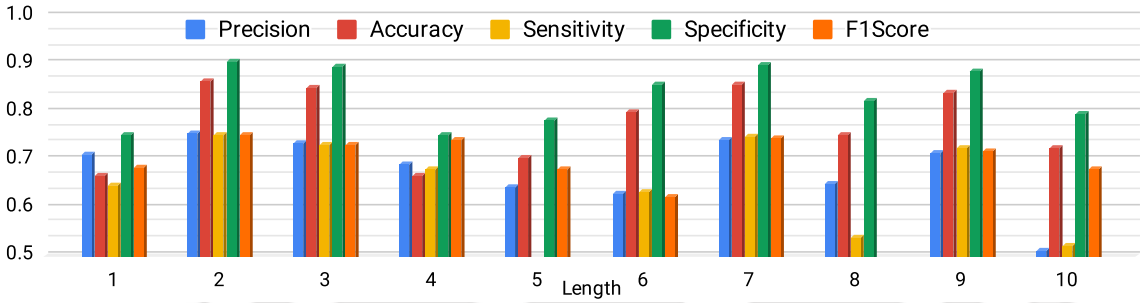
The confusion matrix for all models for 2 and 3 second segment are provided in Figure 6.7 and 6.8. The figures depict less false positives and false negatives as compared to true positives and true negatives. However, for a large population this might not be suitable because a higher percentage of normal subjects might have to go through unnecessary treatment procedures and diseased subjects might be left untreated. The noisy and normal rhythms were having more correct classifications as



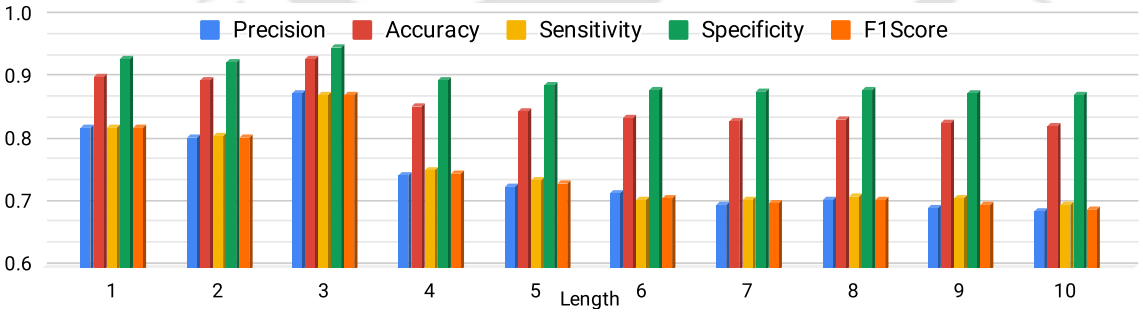
(a) Baseline Convolution Neural Network



(b) Residual Neural Network



(c) Attentive Convolution Neural Network



(d) Transformer Neural Network

Figure 6.6: Overall Model performance for all length signals.

compared to AF and other class rhythms. Most of the misclassification happens in other rhythms as the morphologies vary from one cardiac disease to another which hinders the model to tune its weights and reduces its performance. This problem can be solved if the classification is performed in fine grained ECG classification rather

Table 6.3: F1 score of all the models for varying length signals.

Length	BCNN	ResNet	ACNN	TNN
1	0.63	0.87	0.68	0.82
2	0.75	0.94	0.75	0.81
3	0.76	0.95	0.73	0.87
4	0.71	0.94	0.74	0.75
5	0.54	0.93	0.68	0.73
6	0.61	0.94	0.62	0.71
7	0.68	0.91	0.74	0.70
8	0.79	0.95	0.48	0.70
9	0.56	0.93	0.72	0.70
10	0.73	0.91	0.68	0.69

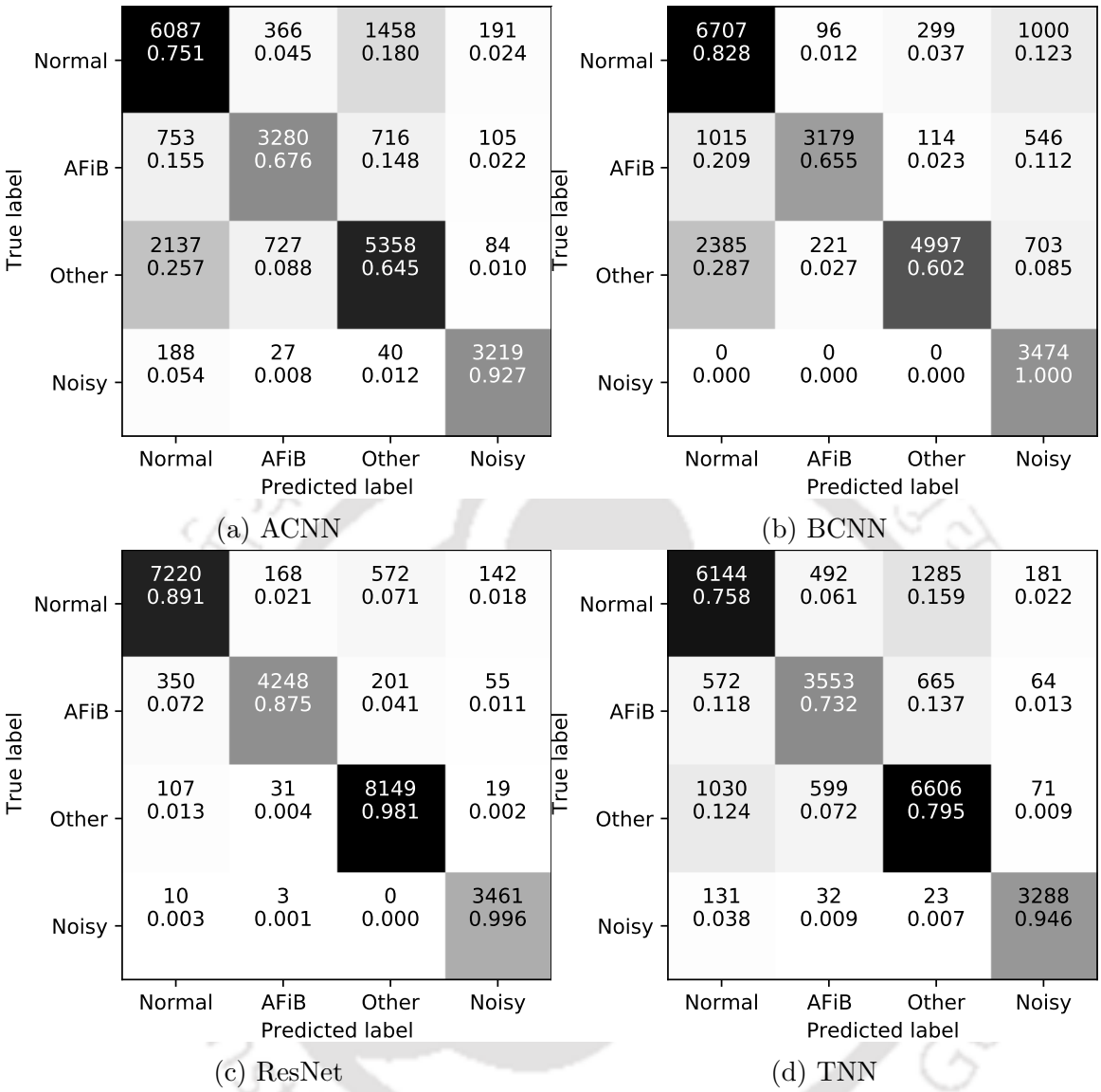
than coarse grained ECG classification. A similar trend could be observed for the 3 second length ECG signal.

6.3.2 Comparison with Existing Methods

The comparison of the ACNN, BCNN, ResNet and TNN models with the state-of-the-art models is provided in Table 6.4. It is worth mentioning that the developed models were tested on validation set as the test set was not publicly available. Therefore, for fair comparison between the models, the F1 scores described in Table 6.4 represent the scores achieved by the state-of-the-art models on the validation set. It can be observed that ResNet performed better than other models. Although attention based models achieved lower performance, the explanations provided by the models make real-world deployment easier.

6.3.3 Qualitative Prediction Interpretation

The model interpretations generated by ACNN and TNN models representing clinically relevant signal timestamps for 2 and 3 second ECG segment are provided in this section. The interpretations of correctly classified two second ECG signals are illustrated in Figure 6.9. The darker dots represent relatively higher clinical importance of the signal timestamps. Figure 6.9a and 6.9b represent correctly classified AFib signal as predicted by ACNN and TNN models. The model focuses on QR segment with a slight focus on f-waves, TP-segment throughout the signal. In Figure 6.9c and 6.9d correctly classified Normal signals are illustrated as predicted by ACNN and TNN models. The ACNN model do not focus on f-waves or fibrillation frequencies



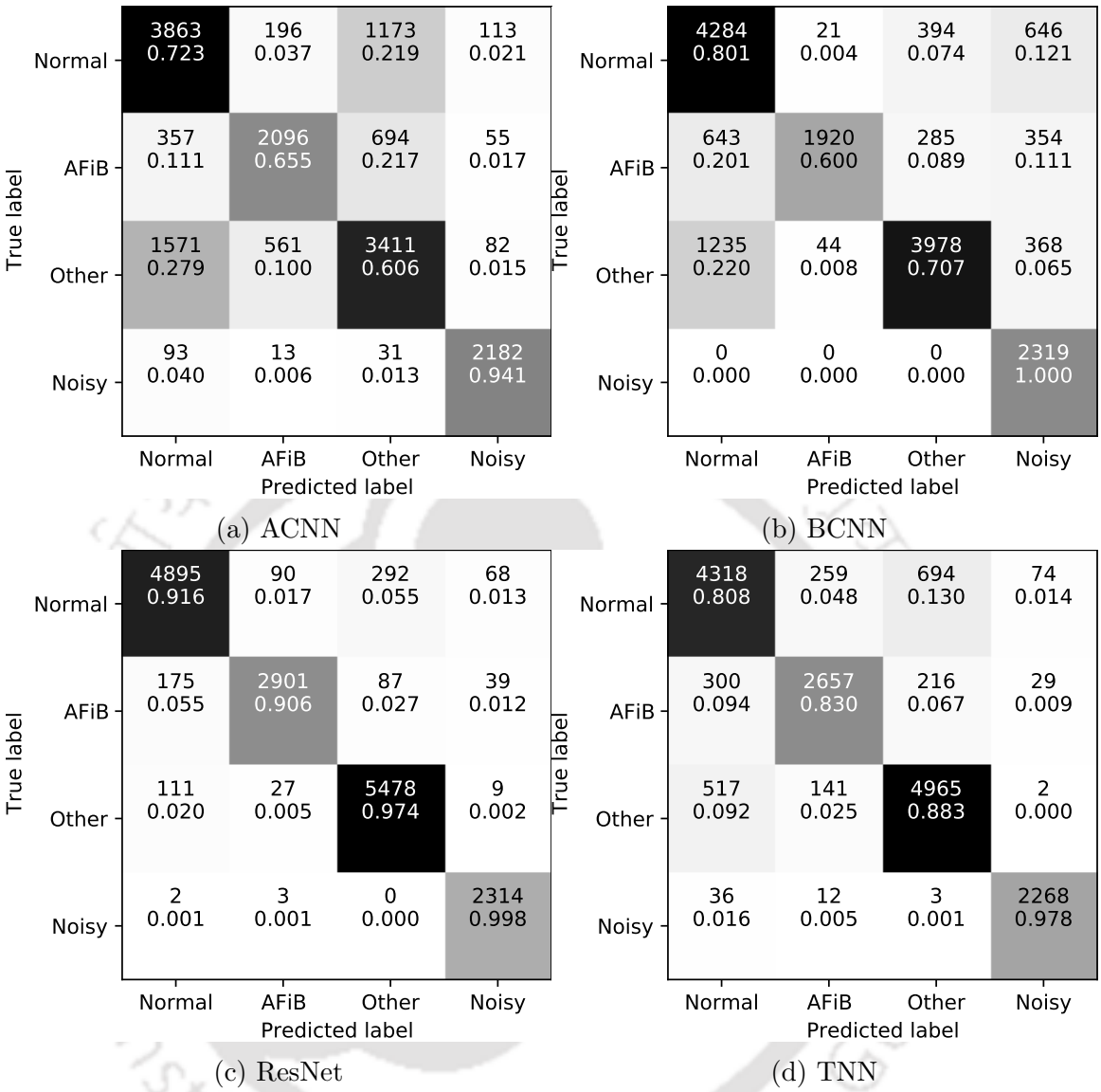
(a) ACNN (b) BCNN

(c) ResNet (d) TNN

Figure 6.7: Confusion Matrix of all models for 2 second ECG signal.

but rather focus on P-waves with slight focus on QRS complexes as seen in Figure 6.9c. The TNN model highly focuses on the presence of P-waves and ST-segment as seen in Figure 6.9d.

In Figure 6.10, miss classified two second ECG signals are represented. Figure 6.10a and 6.10b represented AF segments predicted as NSR by ACNN and TNN models. The models somewhat focus on f-waves but also focus on QRS complex and maybe that is the reason model confuses between AF and NSR and predicts the segment as NSR. The TNN puts more focus on f-waves and less on QRS complex and only QRS complex of one beat is highlighted as illustrated in Figure 6.10b, making TNN a better choice. Figure 6.10c and 6.10d represent the Normal segments predicted



(a) ACNN (b) BCNN

(c) ResNet (d) TNN

Figure 6.8: Confusion Matrix of all models for 3 second ECG signal.

as AF by ACNN and TNN models, respectively. The models clearly highlight the P-waves but do not highlight the QRS complex.

In Figure 6.11, correctly classified three second ECG signals are illustrated. Figure 6.11a and 6.11b represent correctly classified AF rhythms as predicted by ACNN and TNN models, respectively. The models focus on f-waves without focusing on QRS complexes while classifying the segments as AF. The focus on three second segments is better as compared to two second segments. The TNN is better as it completely focuses on fibrillation frequencies and not on QRS complexes. Figure 6.11c and 6.11d represent Normal rhythms that are correctly classified by ACNN and TNN models. Although ACNN correctly classifies, it focuses on f-waves with a strong focus on T-

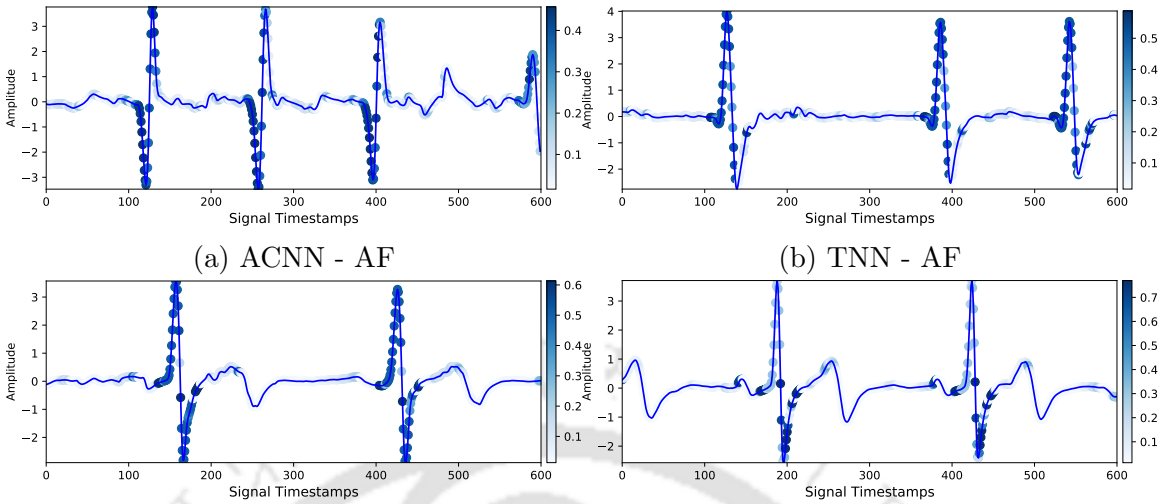
Table 6.4: Comparison with existing techniques.

Paper	Method	F1
Rubin et al. [331]	Low-quality ECG determined by signal quality index is classified as noisy Good quality ECG is transformed to the spectrogram, classified using CNN	0.82
Fan et al. [332]	U-Net for locating R peaks, calculating RR intervals 34-layer ResNet for capturing morphological features Classifying both features using 2-layer FCN	0.85
Zhao et al. [333]	CNN to classify the 2D spectro-temporal ECG obtained from time-varying coefficients of Fourier series using Kalman filter	0.80
Tran et al. [322]	Multiplicative fusion net that combines ResNet-LSTM-FCN using ECG with FCN Network that uses morphological and RR interval features	0.80
Cao et al. [334]	Augmentaed data through duplication, concatenation, resampling 2-layer LSTM trained with augmented dataset.	0.82
Nguyen et al. [335]	Classified features extracted from ECG using a CNN with Stacked support vector machine.	0.83
Andreotti et al. [336]	Time, frequency, and HRV features, clustering of beats on Poincare plots. Ensemble of 50 trees and MLP (2-layer, 10 hidden neurons) output averaged Compared with a ResNet model	0.79 0.83
Behar et al. [319]	Features: time domain, coefficient of sample entropy, Morphological Trained support vector machine in a one-vs.-rest approach	0.80
Billeci et al. [313]	RR interval features reduced through linear discriminant analysis and classified using Least Squares SVM	0.90
Bin et al. [337]	AF features: RR intervals. Shannon Entropy, Kolmogorov-Smirnov test value Morphology Features, RR Interval Features, Similarity Index Between Beats Beat similarity calculated through correlation of QRS wave and R amplitude Classified using decision tree ensemble (adaBoost. M2)	0.86
Christov et al. [314]	HRV, morphology analysis, detection of atrial activity A Linear discriminant classifier	0.79
Ghiasi et al. [338]	Features: correlation coefficient, fractal dimension, variance of R peaks, RR-intervals classified using a decision tree	0.80
Our Approach	**Baseline Convolution Neural Network**	0.76
Our Approach	**Residual Neural Network**	0.95
Our Approach	**Attentive Convolution Neural Network (with interpretation)**	0.73
Our Approach	**Transformer Neural Network (with interpretation)**	0.87

wave. TNN focus on P-wave and not much on TP segment, making TNN a better choice.

In Figure 6.12, miss classified three second ECG rhythms are depicted. Figure 6.12a and 6.12b represent AF rhythms predicted as NSR by TNN and ACNN models. The models highlight f-waves without a clear demarcation of P-wave. TNN model precisely highlights the f-waves as compared to ACNN model. Figure 6.12c and 6.12d represent NSR rhythms predicted as AF by TNN and ACNN models. The models are unable to focus on P-waves but rather focus on f-waves. So the models predicted AF by focusing on AF related clinical characteristics. These missclassified examples provide an insight on where the ACNN and TNN models are making mistakes, knowing these caveats may aid the user to make changes and reduce false alarms and increase the chances model deployment in real world scenarios.

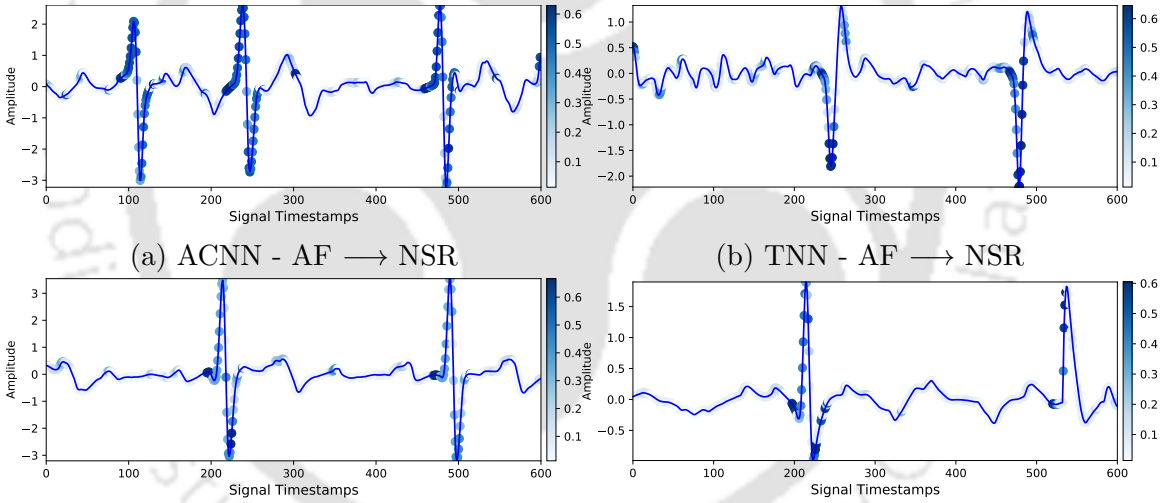
Overall, the three second ECG signal provides better interpretation for both



(a) ACNN - AF (b) TNN - AF

(c) ACNN - Normal (d) TNN - Normal

Figure 6.9: Correctly Classified two second ECG signal.

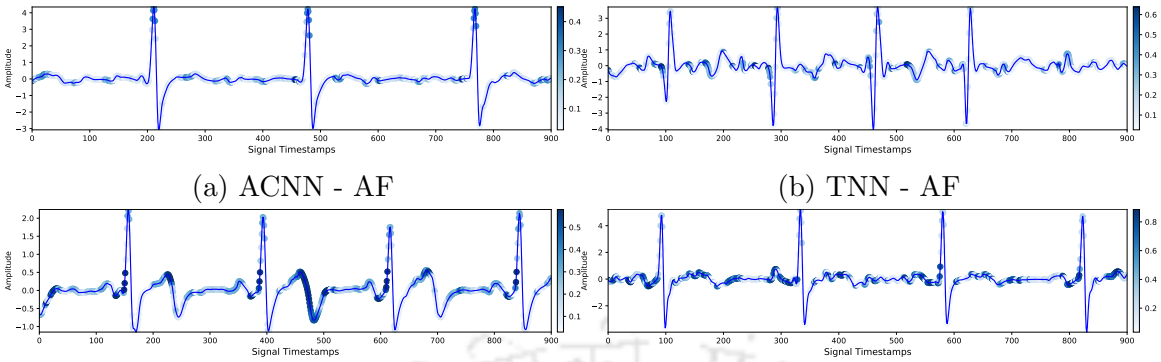


(a) ACNN - AF $\longrightarrow$ NSR (b) TNN - AF $\longrightarrow$ NSR

(c) ACNN - NSR $\longrightarrow$ AF (d) TNN - NSR $\longrightarrow$ AF

Figure 6.10: Miss Classified two second ECG signal.

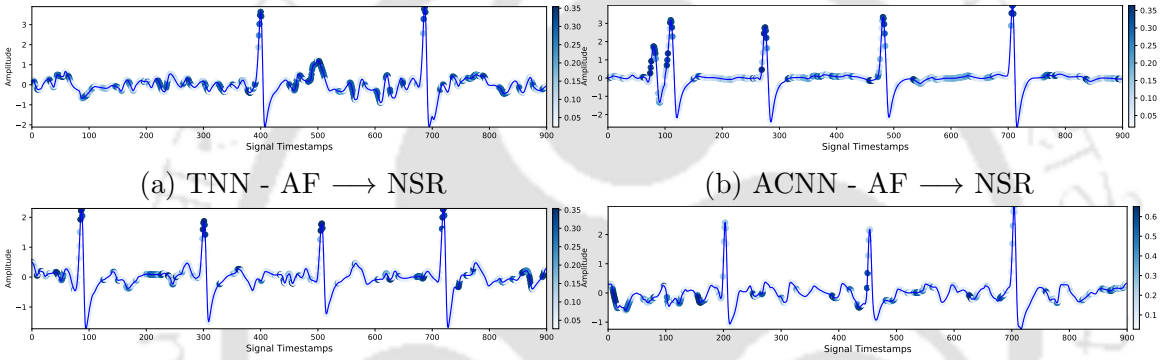
rhythms as the model highlights more relevant timestamps for both correct and miss classified samples. The TNN and ACNN emphasize on F-waves, the clinical features used to recognize AF. The TNN model relatively emphasises more on relevant timestamps as compared to ACNN model, making it a better choice for interpreting model diagnosis. For miss-classified samples, model focused on T-wave or detected noise causing irregularity between T and P waves creating non-uniform RR intervals. Clinically, the normal class rhythms is characterised through P-wave, QRS complex, T-wave morphology and the AF is characterized by the absence of P-wave.



(a) ACNN - AF (b) TNN - AF

(c) ACNN - Normal (d) TNN - Normal

Figure 6.11: Correctly Classified three second ECG signal.



(a) TNN - AF $\longrightarrow$ NSR (b) ACNN - AF $\longrightarrow$ NSR

(c) TNN - NSR $\longrightarrow$ AF (d) ACNN - NSR $\longrightarrow$ AF

Figure 6.12: Miss Classified three second ECG signal.

6.4 Summary

In this chapter, an attention based convolution neural network and transformer neural network are developed for automatic classification and interpretation of atrial fibrillation from ECG signals. For comparison, a baseline convolution neural network and residual neural network are also developed. The residual neural network achieved the highest F1 Score followed by the transformer neural network, baseline convolution neural network, and attentive convolution neural network on PhysioNet Computing in Cardiology Challenge 2017 database. TNN with three layers, two heads in MHA module, and positional encoding achieved an F1 score of 0.87 on CinC dataset. Positional encoding was necessary when TNN depth is increased. From the experimentation, it is evident that ECG segment length also plays a crucial role in model performance. The optimum length was found between one to three seconds. Both ACNN and TNN highlight clinically meaningful timestamps responsible for prediction, but TNN is better suited for such tasks as demonstrated by the experiments.

The ACNN and TNN were also capable of detecting R-peaks in ECG rhythms. The cardiac abnormalities on single-channel ECG often persist in lower resolution but provide better resolution on the multichannel ECG. In addition, the cardiac abnormality interpreted by one cardiologist may differ when interpreted by another cardiologist, leading to multiple cardiac abnormalities for a single recording. Therefore, the next chapter aims at classifying several cardiac abnormalities from multichannel ECG.

7

Feature-Fused Multichannel Multilabel ECG Classification with Prediction Interpretation using Channel Specific Dynamic CNN

The standard 12-lead ECG is a standard method that is representative of the heart's electrical activity. However, the limited accessibility of 12-lead devices in suburban areas motivates the use of reduced lead mobile or handheld devices that have easy accessibility for recording ECG. However, there is a trade-off between the resolution and device availability. Here, resolution refers to the amount of information captured by ECG from different angles. The reduced lead devices capture limited useful information as they sometimes miss relevant information about heart as compared to standard 12-lead ECG, but the device availability reduces and vice versa [12]. This chapter explores the feasibility of predicting multiple cardiac abnormalities from multichannel ECG (MECG) signals that are captured through different lead combinations. The challenge of multilabel MECG classification is relatively new with respect to the thesis as till now single channel ECG classification has been discussed. Therefore, the task of multilabel MECG classification is performed in different stages as described in Figure 7.1. In the first stage, single label 12-lead ECG classification is performed through investigation of convolution, recurrent, and attention based models. In the second stage, heartbeat and demographic features are incorporated into a parallel static CNN to predict multiple cardiac abnormalities for an MECG segment. The model achieves good performance but lacks the aspect of interpretability. In the third stage, a channel specific dynamically built CNN is employed with demographic features for multilabel MECG classification. In addition, the channel specific dynamically built CNN introduces an interpretability mechanism

that highlights the important leads and relevant signal timestamps responsible for corresponding prediction, thereby enhancing trust in the model.

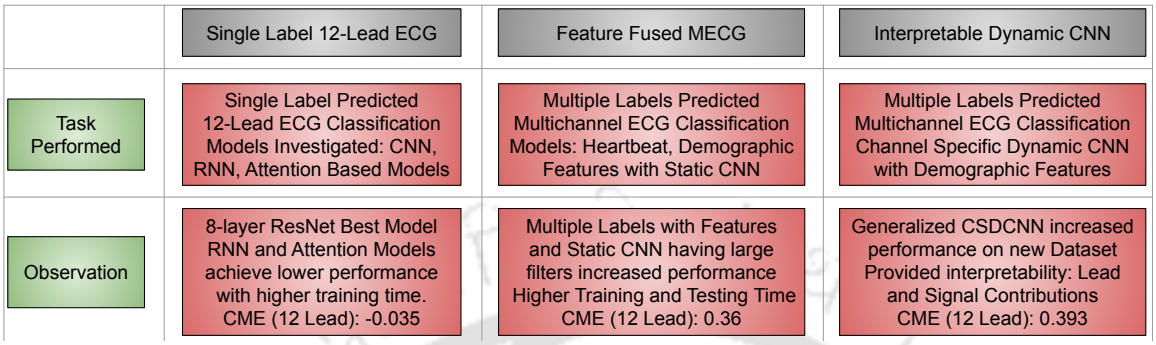


Figure 7.1: Overall Stage Wise Workflow.

The rest of the chapter is organized as follows. Section 7.1 describes the literature review of MECG classification. Section 7.2 describes the experimental setup which includes the problem formulation, dataset description, and evaluation metrics for evaluating the proposed approaches. Section 7.3 discusses the CNN, RNN, and Attention based models investigated for single label 12-lead ECG classification. Section 7.4 describes the usefulness of heartbeat and demographic features fused with MECG and extends the problem to multiple labels using a Static CNN architecture. Section 7.5 describes an interpretability mechanism for multilabel MECG Classification using a channel specific dynamically built CNN and discusses the performance quantitatively and qualitatively. Section 7.6 summarizes the chapter.

7.1 Literature Review

Several studies have been published that aim at solving the problem of 12-lead ECG classification. Traditional machine learning (ML) models such as K-Nearest Neighbour [339] and Support Vector Machines [340] combined with hand-crafted features have achieved significant performance for 12-lead ECG classification. The recent insurgence of Deep learning models outperformed conventional ML models for 12-lead ECG classification [341]. A combination of convolutional layers with Long Short term memory (LSTM) have been proposed to classify ECG signals [324, 342]. Strodthoff et al. [147] found cardiologists' rules for Myocardial Infarction (MI) Detection using an ensemble of Convolution Neural Network (CNN) on Physikalisch-Technische Bundesanstalt (PTB) [343] dataset and also investigated channel selection with its clinical implications. Yao et al. [148] used an attention-based time-incremental CNN

to detect eight types of arrhythmia and sinus rhythm on China Physiological Signal Challenge [344] dataset. The spatial information was fused using CNN and temporal information was combined using LSTM, and an attention module was employed to locate the informative signal segment, thereby providing interpretability. Several studies have applied deep domain adaptation approach to reduce the discrepancy between the training and test distributions caused due to different acquisition methods [345–347]. Ammour et al. [345] used fully connected hidden layers to mitigate disparities among ECG from different datasets. Jin et al. [346] used multi-layer multi-kernel maximum mean discrepancy strategy and Chen et al. [347] used a cluster-aligning loss and a cluster-separating loss.

Recently, the performance of 12-lead ECG classification models has been greatly benefited because of the newly surface online challenge datasets such as the PhysioNet Computing in Cardiology Challenge (CinC) 2020 and 2021 [8, 34, 46]. The challenges allowed investigation of different lead configurations from MECG for detecting a variety of cardiac abnormalities [8,34,46]. The new datasets present a unique opportunity to tackle the problem of multi-class cardiac abnormality detection from large ECG datasets. Table 7.1 provides the methodology and Challenge Metric Scores, a modified accuracy metric, achieved during CinC 2020 [8] using 12-lead ECG signals.

The works demonstrate the application of DLM, including CNN and RNN, for arrhythmia detection using ECG signals [264]. Augmentation techniques such as vertical flip are unsuitable due to left and right axis deviation rhythms, which cause label error during flipping [349, 350]. Researchers have used the same convolution kernels for different ECG leads, and it does utilize the benefit of similarity present in leads. However, since every lead captures a different orientation of the heart's electrical activity, the same filter might not be suitable for different lead signals. In [350], authors demonstrated the effectiveness of demographic and heartbeat features in improving the model performance.

The studies performed in the past have achieved significant performance for MECG classification. However, these methods do not consider the distribution discrepancy within the training set as data has not been collected from multiple sources with different characteristics. The studies assume that test samples are available during the training phase and do not focus on generalization on the unseen test dataset. Moreover, the studies are limited to a few cardiac pathologies. An extensive study for a large group of cardiac pathologies is still lacking. Analysis on reduced lead ECG classification has not been performed. The models lack practical deployment due to lack of interpretability, or the ability of the model to explain its prediction.

Table 7.1: Methods and Challenge Metric Score achieved during CinC 2020 [8].

Paper	Method	CMS
Jia et al. [348]	Combination of Squeeze and Excitation (SE) network and ResNet, which selectively enhance useful features.	0.653
Nonaka et al. [349]	Data augmentation such as flipping, padding and zooming. Efficient Net model for classification.	0.581
Chen et al. [350]	Data enhancement using signal shifting, noise addition. Multi-scale shared convolution kernel with sizes of 5 and 3 followed by SE-Net. Age and gender were also exploited.	0.640
Feng et al. [351]	Encoder based on ResNet; a multi-instance classification module and a decoder to classify through sigmoid.	0.486
Zhao et al. [352]	Embedded SE layer into ResNet with large kernel size.	0.672
Bortolan et al. [353]	Morphological, time-frequency (TF) features extracted from ECG. Rules from features made using physician knowledge.	0.325
Bortolan et al. [353]	CNN trained using scaleogram (TF representation) images generated through continuous wavelet transformed ECG	0.426
Singstad et al. [354]	CNN using Raw ECG, age, gender with a rule-based model using heart rate variability (HRV) features from R-peaks	0.377
Natarajan et al. [149]	Important HRV, Morphological features selected via random forest which are combined with age, gender, and features learned from ECG using Transformer Neural Network.	0.533
Hasani et al. [355]	ECG augmentation through random padding, cropping, addition of low-frequency artifacts, and lead dropout. Combination of 2 parallel CNN followed by LSTM.	0.437

7.2 Experimental Setup

The developed models were trained locally on a workstation with an Intel Xeon processor with 32 GB RAM and an NVIDIA GeForce Titan Xp graphics card with 12 GB GDDR6 VRAM. To avoid loading whole data in memory at once, the training procedure is optimized by employing the generator method available in keras library to perform batchwise training. During testing, the prediction of all segments of the record is averaged and that class of diagnosis is chosen that is predicted most number of times. An early stopping criteria is employed to avoid model overfitting.

7.2.1 Problem formulation

MECG classification could be generalized as a multiclass time-series classification problem. The goal is to develop a highly complex nonlinear decision function that extracts useful information and classify the segment into one or more cardiac abnormali-

ties. The model takes input signal $X = [x_1^1, \ldots, x_k^l]$, and output label $Y = [y_1, \ldots, y_c]$, where k is the number of signal timestamps, l is the number of leads, $c = 30$ for scored cardiac abnormalities, $y_i = 1$ if cardiac pathology is present, and $y_i = 0$ if cardiac pathology is absent. The objective is to minimize the binary cross-entropy loss for each MECG segment between the reference labels and predicted output as provided in Equation 7.1. The loss represents the negative average log of corrected predicted probabilities, where $\sigma(x_i)$ is the probability of the sample x_i belonging to the respective cardiac pathology.

$$l = -\frac{1}{C}\sum_{i=1}^{C} y_i \cdot \log \sigma(x_i) + (1 - y_i) \cdot \log(1 - \sigma(x_i)) \tag{7.1}$$

7.2.2 Dataset Description

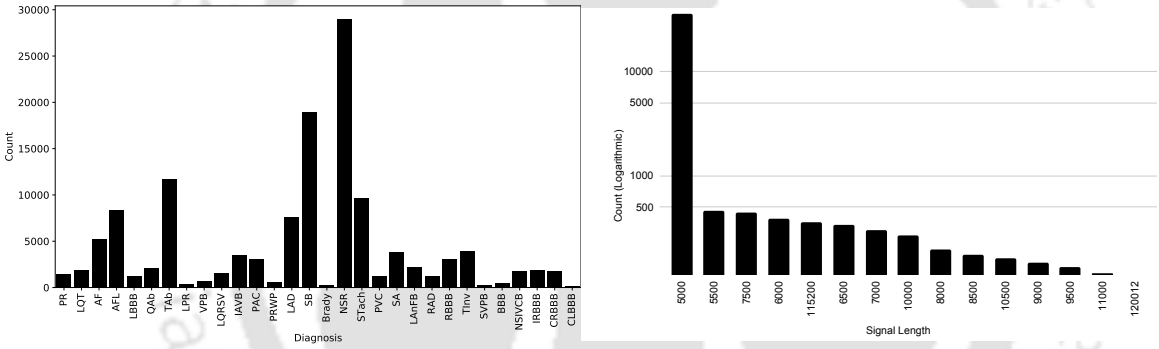
A total of nine databases are used with 1,31,155 12-lead ECG recordings acquired from PhysioNet Computing in Cardiology Challenge 2020 and 2021 [8, 46]. The databases include: China Physiological Signal Challenge in 2018 (CPSC 2018) [344], St Petersburg INCART 12-lead Arrhythmia Database [356], Physikalisch-Technische Bundesanstalt (PTB) [343] and PTB-XL [357] database, Chapman University, Shaoxing People's Hospital (Chapman-Shaoxing) [358] and Ningbo First Hospital (Ningbo) database [359], University of Michigan (UMich) database, Georgia database which represents a unique demographic of the Southeastern United States and an undisclosed American database that is geographically distinct from the Georgia database [8, 46]. The detailed dataset description is provided in Table 7.2.

Table 7.2: Detailed Dataset Description.

	Number of ECG Recordings				**Frequency**	**Duration**
Dataset	**Train**	**Val**	**Test**	**Total**	**(Hertz)**	**(seconds)**
CPSC [344]	10,330	1,463	1,463	13,256	500	6 to 144
INCART [356]	74	0	0	74	257	30 minutes
PTB [343, 357]	22,353	0	0	22,353	500/1,000	10 to 120
Chapman [358]	10,247	0	0	10,247	500	10
Ningbo [359]	34,905	0	0	34,905	500	10
Georgia [8, 46]	10,344	5,167	5,167	20,678	500	5 to 10
UMich [8, 46]	19,642	0	0	19,642	250/500	10
Unknown [8, 46]	0	0	10,000	10000	–	–
Total	107,895	6,630	16,630	**131,155**	–	–

Exploratory Data Analysis: The dataset consists of patient records consisting of 12-lead ECG, Analog to Digital conversion (ADC) gain, the baseline for each lead,

age, gender, patient history, symptoms, medical prescription, and diagnosis or cardiac rhythm information (disease labels). Since patient history, symptoms, and medical prescription are not available for majority of the records, this information is not used. The diagnosis distribution of SNOMED scored classes is described in Figure 7.2a. The labels in the dataset belong to one of the 30 scored classes. The abbreviations of the diagnosis in Figure 7.2a are described in Table 7.3. The challenge specified three identically scored diagnosis pairs: SVPB and PAC, PVC and VPB, CRBBB and RBBB. These diagnosis were not merged and the models were developed considering thirty output labels. Since homogenization of ECG signal length is important before feeding it to a neural network, the signal length that is most common in the dataset is selected. The signal length distribution is described in Figure 7.2b and 5000 sample length is most common. Therefore, 5000 sample length is selected for modelling.



(a) Diagnosis Distribution (b) Signal Length Distribution

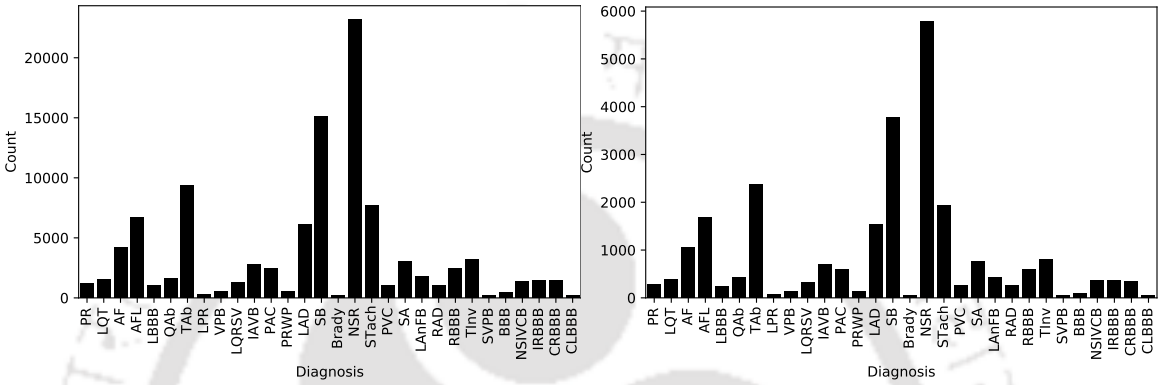
Figure 7.2: Distribution of Diagnosis and Signal Length in the dataset.

Table 7.3: Diagnosis with Abbreviations (Abbrev).

Diagnosis	Abbrev	Diagnosis	Abbrev	Diagnosis	Abbrev
Premature atrial contraction	PAC	Pacing rhythm	PR	Low QRS voltages	LQRSV
Ventricular premature beats	VPB	Poor R wave Progression	PRWP	Atrial fibrillation	AF
Left bundle branch block	LBBB	1st degree AV block	IAVB	Sinus rhythm	NSR
Complete left bundle branch block	CLBBB	Prolonged PR interval	LPR	Bradycardia	Brady
Complete right bundle branch block	CRBBB	Prolonged QT interval	LQT	Atrial flutter	AFL
Premature ventricular contractions	PVC	Q wave abnormal	QAb	T wave abnormal	TAb
Incomplete right bundle branch block	IRBBB	Right axis deviation	RAD	T wave inversion	TInv
Supraventricular premature beats	SVPB	Right bundle branch block	RBBB	Sinus arrhythmia	SA
Left anterior fascicular block	LAnFB	Bundle branch block	BBB	Left axis deviation	LAD
Nonspecific intraventricular conduction	NSIVCB	Sinus bradycardia	SB	Sinus tachycardia	STach

The leads in the reduced set are (I, II, III, aVL, aVR, and aVF) in 6 lead, (I, II, III, V2) in 4 lead, (I, II, V2) in 3 lead, and (I, II) in 2 lead. A multi-label stratified five-fold cross-validation was applied on the combined dataset so that the training and the validation sets in each fold have similar class distribution. The distribution

is similar across all folds as illustrated in Figure 7.3. The dataset consists of 111 abnormalities. Out of these, 30 scored abnormalities were used which reduced around 4000 recordings. Finally, 88253 records were available during training. Out of these, 70602 recordings were used for training, and 17651 recordings were used for validation in all folds.



(a) Training Data Distribution (b) Validation Data Distribution

Figure 7.3: Pathology distribution in training and validation set while splitting.

7.2.3 Evaluation Metrics

The model performance is evaluated using modified accuracy metric or challenge metric (CME) as recommended in [8,46]. Since, few misdiagnoses are more harmful than others, CME awards partial credit to misdiagnoses that result in similar outcomes as correct diagnosis as judged by cardiologists. Let $C = \{c_i\}_{i=1}^{m}$ be a collection of m distinct diagnoses for a database of n recordings. A multi-class confusion matrix $A = [a_{ij}]$, where a_{ij} is the normalized number of recordings in a database that were classified as belonging to class c_i but actually belong to class c_j (c_i and c_j may be same or different). Since recordings can have multiple labels and classifier can predict multiple labels, the challenge organizers normalized the contribution of each recording to the scoring metric by dividing by the number of classes with a positive label and/or classifier output. Specifically, for each recording $k = (1, \ldots, n)$, let x_k be the set of positive labels and y_k be the set of positive classifier outputs for recording k. Multi-class confusion matrix $A = [a_{ij}]$ is defined as

$$a_{ij} = \sum_{k=1}^{n} a_{ijk}, \text{ where, } a_{ijk} = \begin{cases} \frac{1}{|x_k \cup y_k|}, & \text{if } c_i \in x_k \text{ and } c_j \in y_k \\ 0, & \text{otherwise} \end{cases} \tag{7.2}$$

The quantity $|x_k \cup y_k|$ is the number of distinct classes with a positive label and/or classifier output for recording k. The classifiers receive higher score from recordings with multiple labels than from those with a single label, but more predicted positive labels may reduce the score. Next, a reward matrix is defined $W = [w_{ij}]$, where w_{ij} is the reward for a positive classifier output for class c_i with a positive label c_j (where c_i and c_j may be the same class or different classes). The weight matrix provided by challenge organizers [8]. The following class pairs: PAC and SVPB; PVC and VPB; CRBBB and RBBB are considered similar, so a predicted output in one of these classes is considered to be a positive label or classifier output for all of them.

The weight matrix provided by challenge organizers is described in Figure 7.4. The following class pairs: PAC and SVPB; PVC and VPB; CRBBB and RBBB are considered similar, so a predicted output in one of these classes is considered to be a positive label or classifier output for all of them. The highest values of the reward matrix are along its diagonal, associating full credit with correct classifier outputs, partial credit with incorrect classifier outputs, and no credit for labels and classifier outputs that are not captured in the weight matrix.

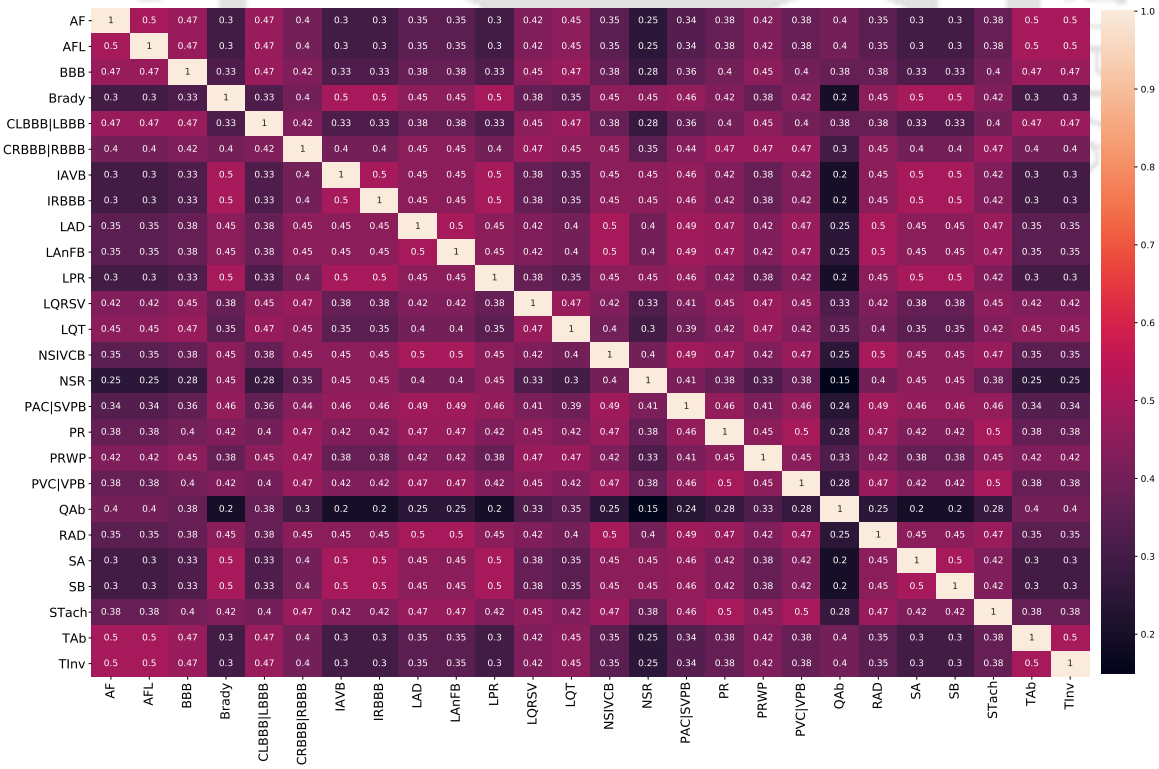


Figure 7.4: Illustration of the Reward Matrix for the scored diagnoses with rows and columns labeled by diagnoses abbreviations.

Finally, the modified accuracy metric or CME is defined in Equation 7.3, where

s_{inactive} is the score for the inactive classifier and s_{true} is the score for ground-truth classifier. A classifier that returns only positive outputs will typically receive a negative score, i.e., a lower score than a classifier that returns only negative outputs, which reflects the harm of false alarms.

$$s_{normalized} = \frac{s_{unnormalized} - s_{inactive}}{s_{true} - s_{inactive}}, \text{ where, } s_{unnormalized} = \sum_{i=1}^{m}\sum_{j\neq 1}^{m} w_{ij}a_{ij} \tag{7.3}$$

In addition the model evaluation is performed using the evaluation metrics: Sensitivity (Se), Specificity (Sp), Accuracy (ACC), and F-measure (FME) described in Section 4.3.2. Area Under the Receiver Operating Characteristics Curve (AUC) is also employed which represents false alarm rate versus the hit rate or false positive rate (FPR) versus the true positive rate (TPR) for $threshold \in (0, 1)$. AUC can be calculated as (Se+Sp)/2. AUP represents precision vs recall curves for different thresholds.

7.3 Model Investigation for Single Label Twelve Lead ECG Classification

In the first stage, single label 12-lead ECG classification is performed through investigation of five categories of models that include convolution neural networks (CNN), recurrent neural networks (RNN), combination of CNN and RNN, RNN with attention based models [142], and lastly the combination of CNN, RNN, and attention based models. Figure 7.5 illustrates the schematic of investigated models. The details of each category of models is provided in subsequent sections. In this stage, only the first label from each diagnosis is considered and similarly a single class is predicted for each record. Hyperparameters such as number of convolution and recurrent layers, size of kernels in CNNs and cells in RNNs were adjusted using the validation data.

7.3.1 Convolution Based ResNet

The first category employs the convolution based Residual neural network (ResNet) [277]. The architecture of the proposed 8-layer ResNet is illustrated in Figure 7.6. The model consists of a variety of layers including one dimensional convolution layer (1-D Conv), batch normalization (BN), dropout, max pooling, and fully connected layer or dense layer. Table 7.4 describe the details of the parameters used in the

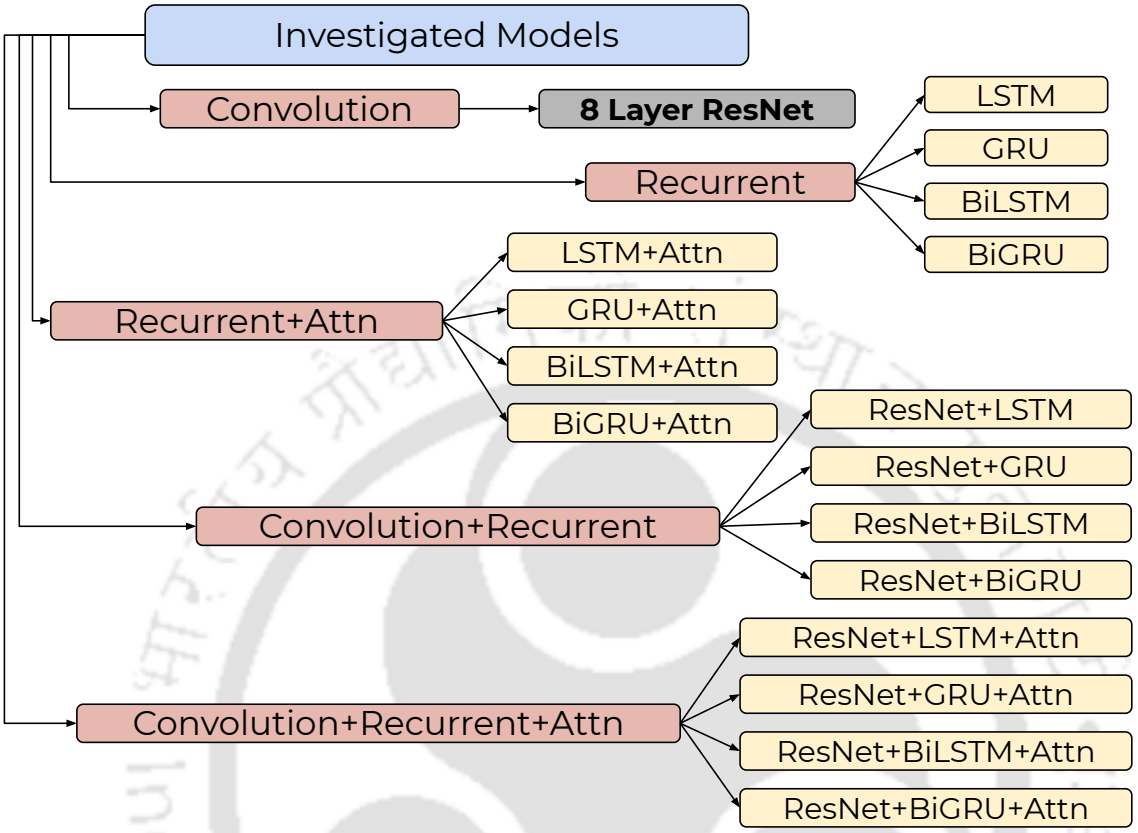


Figure 7.5: A schematic of different investigated models.

8-layer ResNet model. The convolution filters present in the CNN makes the need of peak detection, feature extraction, ranking, and selection insignificant as the model recognises patterns of each rhythm during training as uses the learned representation during inference. Regularization of the model is attained using batch normalization and dropout to avoiding overfitting on the training data. Batch normalization [274] normalizes the inputs of preceding layer. Dropout [276] drops random convolution filters by temporarily removing their contribution during training with dropout probability p and vice versa during the testing phase. The 1-D Conv accounts for temporal relations present in samples of ECG signal. The skip connections skip over the non-contributing layers, allowing the gradient to backpropagate to the initial layers, mitigating the problem of vanishing and exploding gradient. This allows the model to train faster with better accuracy. The non-linearity is introduced using Rectified Linear Unit (ReLU) activation in the model. The dense layer with 27 neurons is employed to provide more abstraction to the model representation. The softmax activation function assigns a probability to the final model prediction. In the network

described in Figure 7.6, 1-D CBR describes a combination of 1-D convolution, BN, and ReLU. BRD describes the combination of BN, ReLU, and dropout. The convolution filter weights were initialised using Xavier initialization [330]. 64 filters were used with the size of 16×1 and a stride of 1. Stride of Max Pooling is kept at 2 and dropout probability of 0.5.

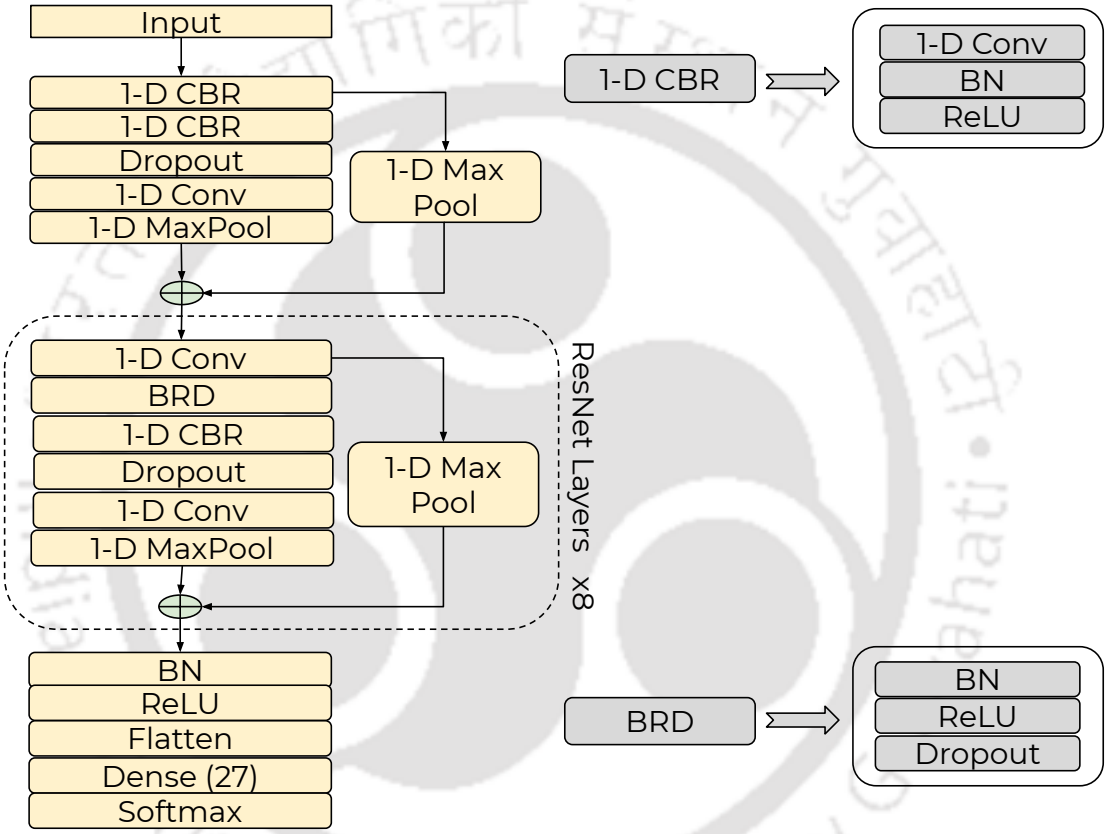


Figure 7.6: Proposed architecture of the 8-layer ResNet.

Table 7.4: Details of the ResNet Model Parameters.

Parameter	**Value**
Convolution Weights	Xavier initialization [330]
Convolution Filter Size	16×1
Convolution Filter Stride	1
Number of Filters	64
MaxPool Stride	2
Dropout Probability	0.5

7.3.2 Recurrent Neural Network Models

The Recurrent Neural Network Models consists a chain of repeating modules of neural network that learns about the 12-lead ECG signal and stores information in its repeating cells. Similar to the CNN, they also ingest a constant sized input. Inherently, RNNs work over sequences that makes it suitable for our application. However, RNNs struggle to remember information for longer period of time which led us to use long short term memory (LSTM), special class of RNN, designed to remember information for long periods of time without any struggle [360]. As arrhythmias occur arbitrarily in long term ECG, making long term information retention important, leading to the use of LSTM. Gated recurrent unit (GRU), special class of RNN, is also used that consists of update, reset, and current memory gates and does not maintain any internal cell state and passes the information to next GRU [361]. Bidirectional LSTM and GRU were also employed as they focus on past and future samples during predictions. The architecture of the Recurrent Neural Network Models is illustrated in Figure 7.7. Since, four models are used, the dotted line represents the other RNN model configurations used for classification.

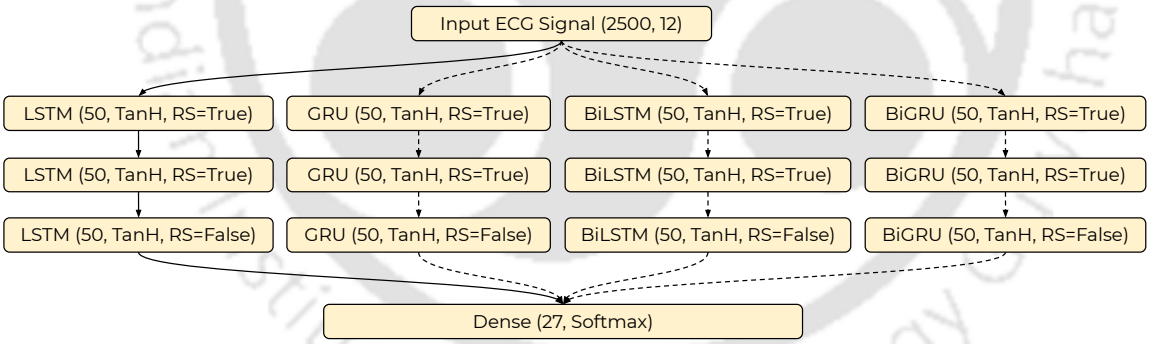


Figure 7.7: Proposed architecture of the RNN.

7.3.3 RNN with Attention Mechanism

The aspect of interpretability is also explored to explain the predictions of our models using attention mechanism along with the recurrent based networks [142]. It is better than conventional RNN based models because in attention mechanism the entire input is passed along the network without creating any bottleneck situation, alleviating the learning problem of neural network. This allows recurrent networks to focus at specific morphologies present in ECG at different timestamps. Attention layer softly searches for parts of the signal that are relevant to predicting a specific class of arrhythmia.

The architecture of the RNN Models with Attention Mechanism module is illustrated in Figure 7.8. The dotted line represents the different RNN model configurations with Attention Mechanism module used for classification.

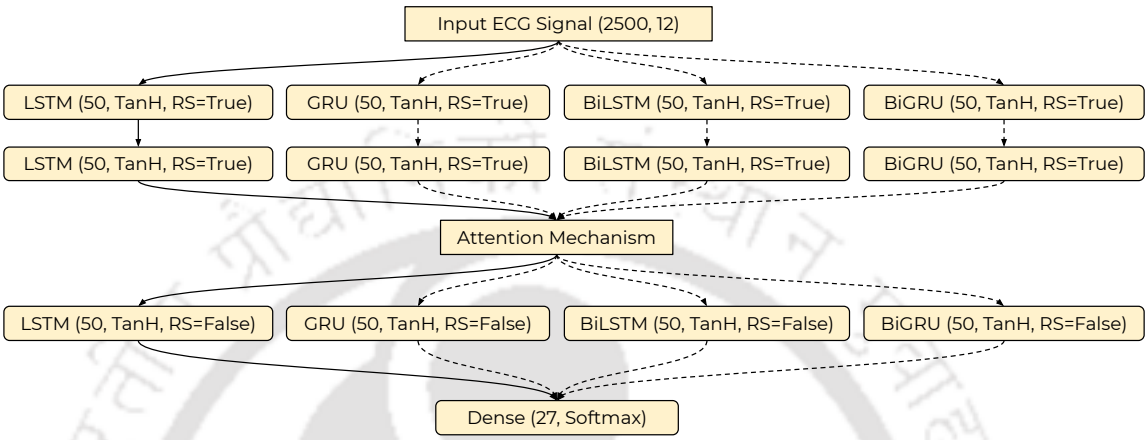


Figure 7.8: Proposed architecture of the RNN and Attention.

7.3.4 ResNet with RNN

In the fourth category of models, the combination of ResNet model with RNN models is investigated. The architecture of eight layer ResNet with different configurations of RNN based models is illustrated in Figure 7.9.

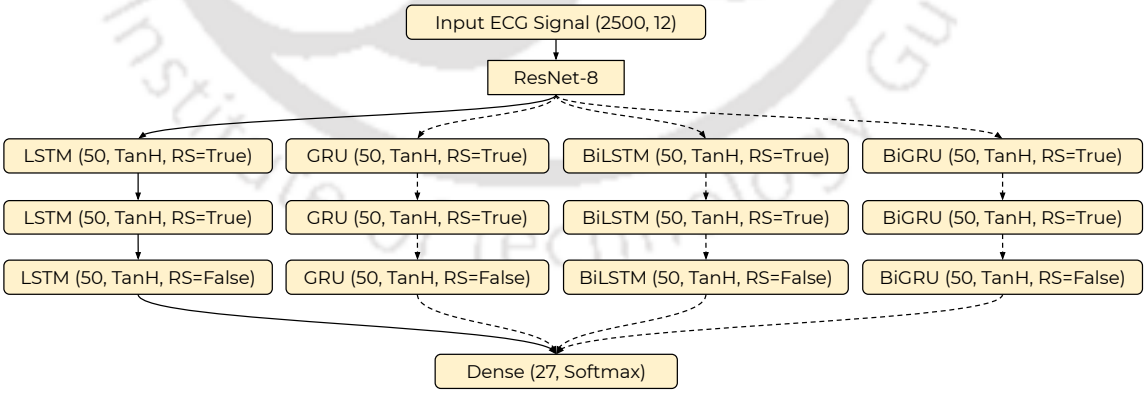


Figure 7.9: Proposed architecture of the 8-layer ResNet and RNN.

7.3.5 ResNet with RNN and Attention Mechanism

The final category of models includes a combination of ResNet followed by RNN variants followed by an attention decoder to provide the models with more flexibility

to learn the data representation. The architecture of the combination of 8-layer ResNet with RNN models and Attention mechanism module is illustrated in Figure 7.10.

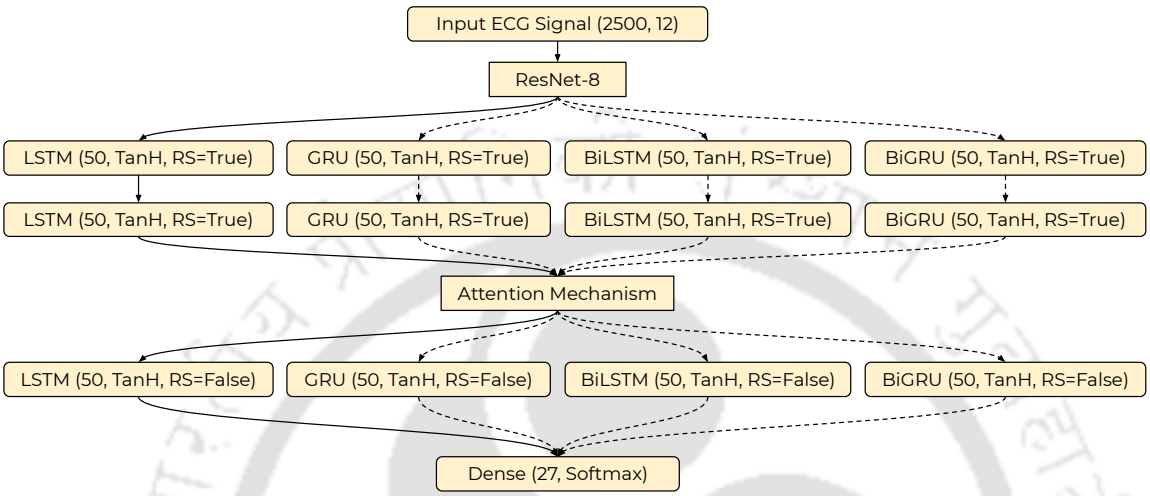


Figure 7.10: Proposed architecture of the 8-layer ResNet with different RNN configurations and attention mechanism module.

7.3.6 Observations

Convolution layers in ResNet, LSTM, GRU layers, and cells in each layer were finalised empirically. The challenge score [362] described in Section 7.2.3 is calculated for all the above mentioned models after training for all training data is described in Figure 7.11. ResNet with LSTM and GRU with attention achieved scores of -0.128 and -0.418. Figure 7.12 describes the training time taken per epoch by each model. The 8-layer ResNet model outperformed other model configurations in terms of challenge score and execution time. The attention based models did not provide better results even after having a larger context vector. Training time per epoch of ResNet was around 110 sec. as compared to 9070 sec. by attention based models because the attention decoder module was not optimized for GPU. Therefore, the convolution based model is selected for further evaluation.

Figure 7.13 describes the results over stratified 10-fold cross validation (CV) on the training data by 8-layer ResNet model. Table 7.5 describes the final results for different evaluation metrics on validation and test data. The main reason behind low performance of models might be the selection of single label for each record. ResNet produced a score of 0.21 ± 0.04 during stratified 10 fold CV on training data, 0.305 on validation data, -0.035 on the full test set.

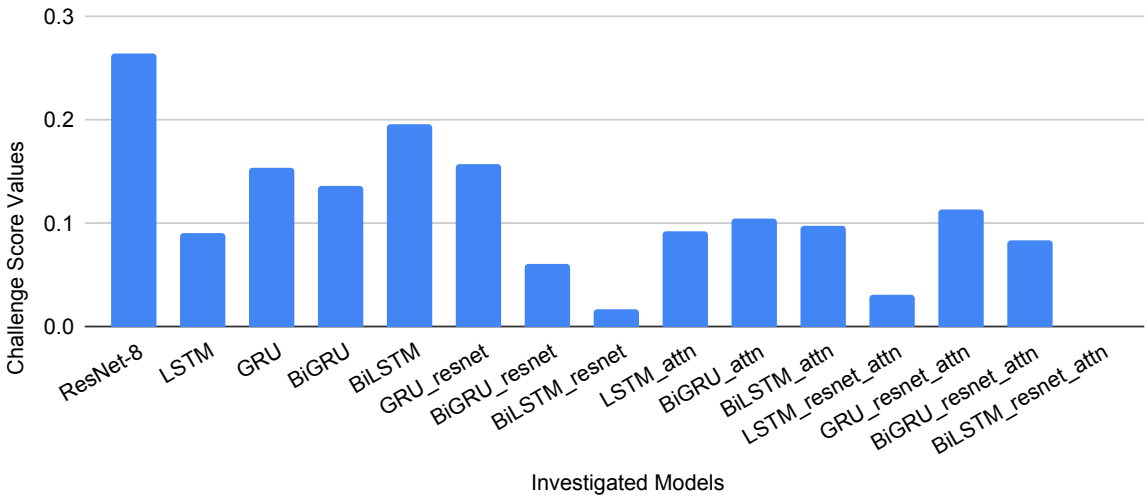


Figure 7.11: Challenge Scores on train data for all models.

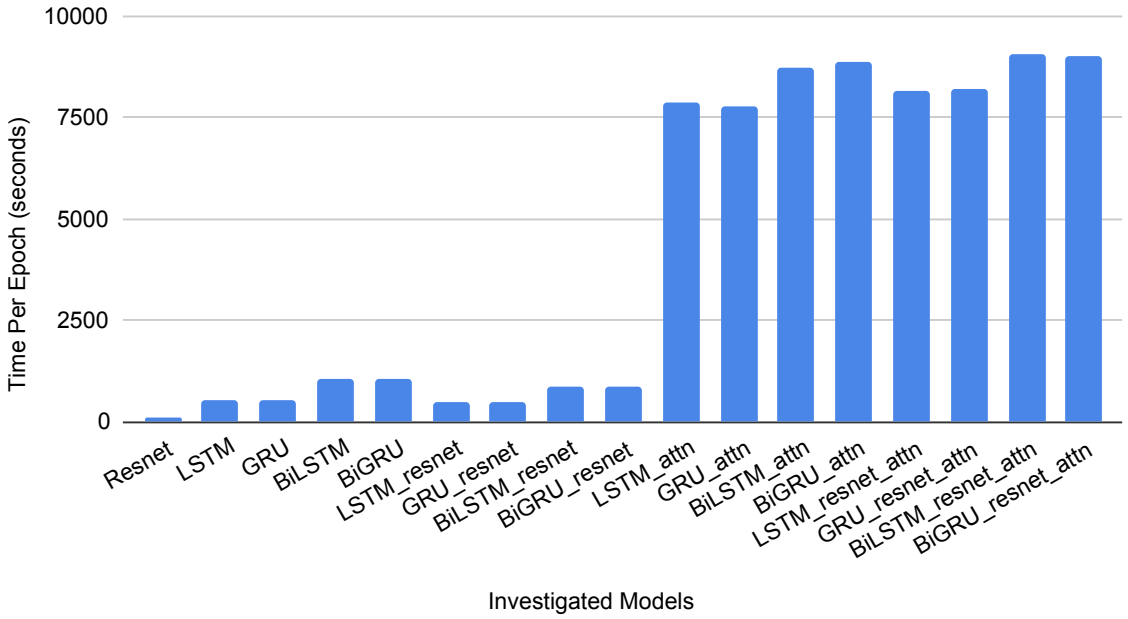


Figure 7.12: Time to train per epoch for all the models.

Table 7.5: Performance for the 8 layer ResNet model on validation and test data.

Metrics	Valid.	CPSC	GE12EC	Undisclosed	Full Test
AUROC	0.825	0.919	0.812	0.694	**0.742**
AUPRC	0.326	0.711	0.318	0.239	**0.229**
Accuracy	0.331	0.527	0.277	0.08	**0.181**
F-measure	0.286	0.228	0.259	0.118	**0.182**
Challenge Score	0.305	0.648	0.25	-0.287	**-0.035**

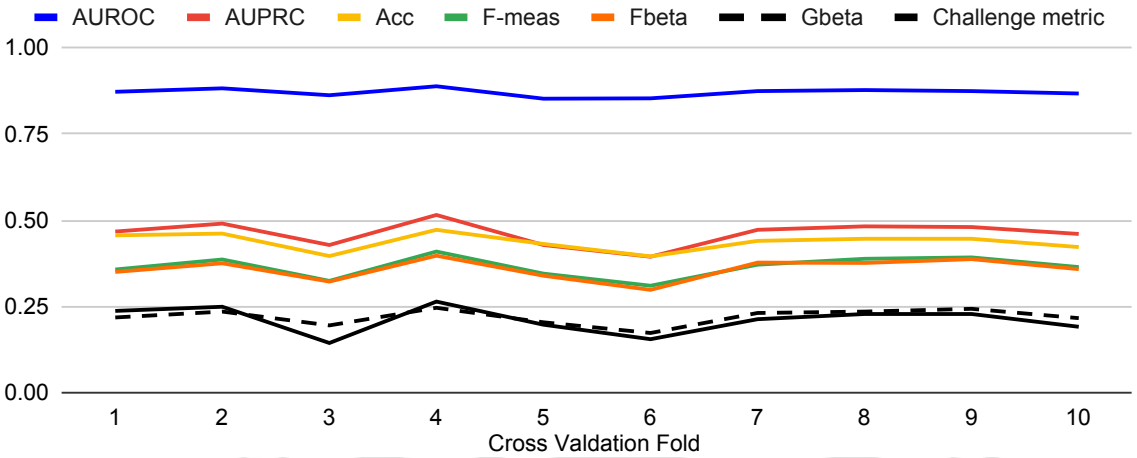


Figure 7.13: 10-fold cross validation results for best performing ResNet model.

It can be observed that the model achieved a negative score on undisclosed test set and overall test set, it happened because of two interrelated reasons: (i) Model always predicts a cardiac abnormality, as softmax activation function is applied in the last layer which always predicts a scored label (30 classes) even for a segment which has an unscored cardiac abnormality (111 total classes); (ii) Design of the challenge metric that assigns a negative score for the models that returns only positive outputs, i.e., a lower score than a classifier that returns only negative outputs, which reflects the harm of false alarms. Since the test data is hidden and not available online, the final validation and test data results for other model configurations are not available. Moreover, as the models achieved significantly less performance than the convolution based ResNet, the convolution neural network is further investigated to achieve better performance.

The shortcomings of the first stage are: (i) single label is considered for an MECG segment; (ii) the models are not generalized as they achieved low performance for test datasets; (iii) Even though attention mechanism module is applied the performance is very less therefore the model lacks interpretability; (iv) investigation on reduced lead ECG is not performed. These shortcomings are improved in the next stage.

7.4 Heartbeat and Demographic Feature Fused Multilabel MECG Classification with Static CNN

In the second stage, the convolution based models are preferred and heartbeat, demographic features are incorporated to predict multiple cardiac abnormalities for an

MECG segment using a static Parallel Convolution Neural Network with Global Average Pooling (PCNN-GAP) network. PCNN-GAP network is fused with age, gender, and patient specific features extracted from Lead II. Several preprocessing steps are employed, including noise removal, resampling, and segmentation before classifying the MECG segment. Lastly, the sigmoid activation function classifies the rhythm into one or more cardiac abnormalities. Multiple channel configurations are explored. The workflow followed in this stage encompasses a preprocessing and classification stage as described in Figure 7.14. The lead information, age, gender, and diagnosis from the patient record are used in this stage. The preprocessing stage prepares the dataset which is classified by the proposed PCNN-GAP classifier.

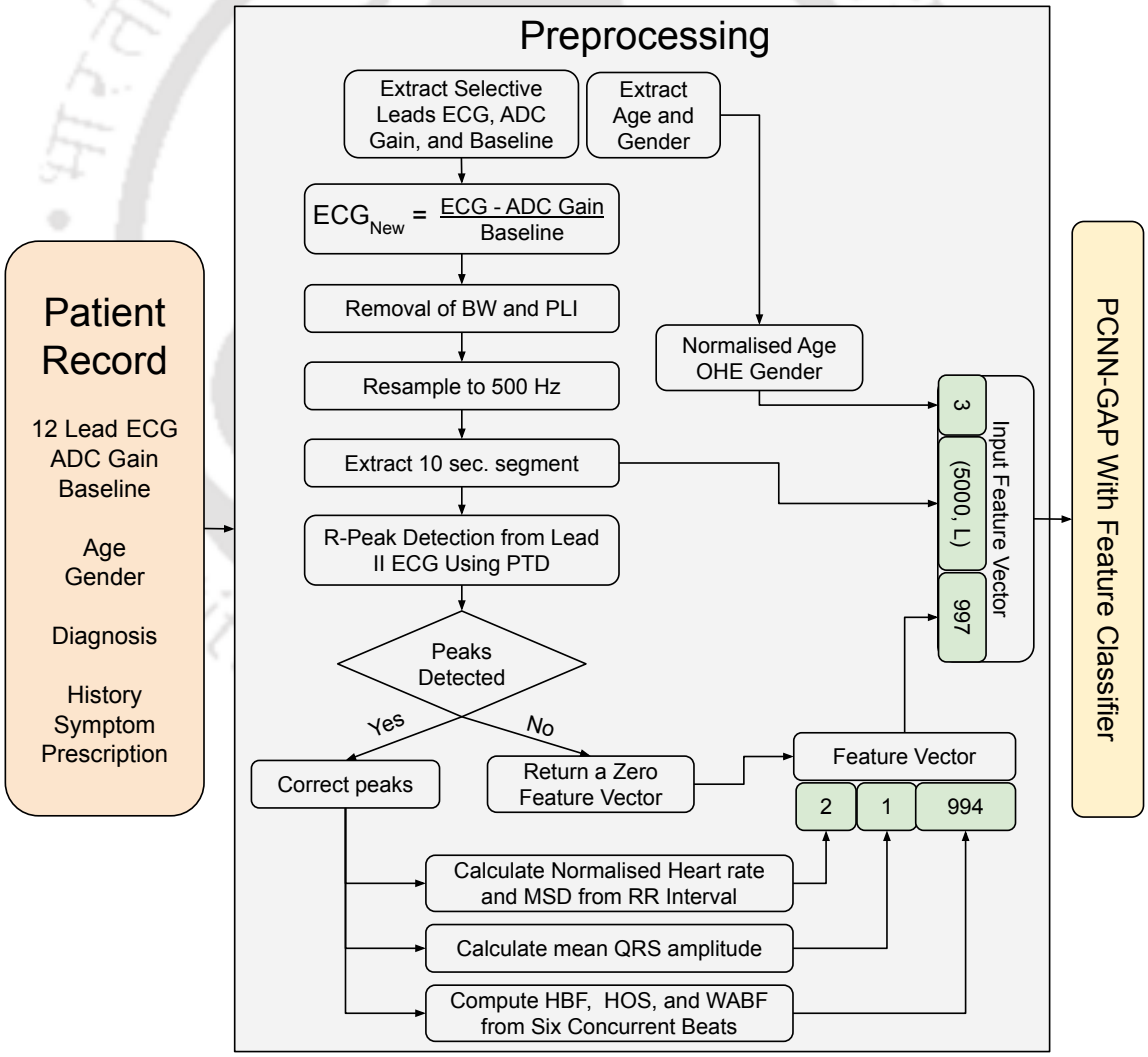


Figure 7.14: Proposed workflow consisting of patient record description, preprocessing, and classification stages.

7.4.1 Preprocessing

The MECG is normalized using the respective ADC gains and baselines. The resultant signal (ECG_{New}) is decluttered from low frequency noise ($\leq 0.6\ Hz$) such as baseline wander (BW) and high frequency noise ($50\ Hz$) such as powerline interference (PLI). The data is procured from different sources with different sampling frequencies, resampling to $500\ Hz$ is performed as 4/6 sources are sampled at $500\ Hz$. Then, segments of 10-second are extracted from resampled signals as it is the most common duration among all the records. The longer length segments are divided into 10 second segments are short segments are padded with zeros. Then heartbeat features and demographic features are extracted that includes age and gender encoding, followed by classification.

Heartbeat Features: The Lead II signal is selected from segmented MECG for R-peak detection using the Pan Tompkins Detection (PTD) algorithm [85]. Lead II provides better depiction of characteristic waveforms present in ECG rhythm and is the common lead in all sets [363]. If PTD algorithm fails to extract R-peaks from the clean signal, then a 997 dimensional zero feature vector is returned. If the peaks are detected using PTD, they are updated with the local maxima present in the close vicinity. The correct R-peaks are used to extract the following features: (i) Average Heart rate; (ii) Mean Square Difference between the RR intervals (MSDRR); (iii) Mean QRS Amplitude; and (iv) Beat Level Features (BLF). The average heart rate (HR_{avg}) and MSDRR are calculated using Equation 7.4. The RR intervals (RRI) are calculated using the difference between adjacent R-peaks. Mean QRS amplitude is the mean of R-Peak amplitude.

$$\begin{aligned} HR &= \frac{60 \times Fs}{R-Peak_{Current} - R-Peak_{Previous}} \\ HR_{avg} &= \frac{\sum(HR)}{100 \times Len(HR)} \\ MSDRR &= \frac{Mean[Diff[RRI]^2]}{100} \end{aligned} \tag{7.4}$$

The BLF are calculated for up to six concurrent beats enclosed inside a window of 130 timestamps left to the R-peak and 230 timestamps right to the R-peak so that at least a heartbeat is enclosed in the window. The BLF include: (i) Hermite Polynomial Coefficients (HPC) [250]; (ii) Statistical Features [96, 250]; and (iii) Wave Amplitude-Based Features (WABF) [96]. The HPC exploits similarity between the beats by representing them with Hermite basis function coefficients. Assume, $x(t)$ denotes the beat, the Hermite series expansion yields Equation 7.5. Here, c_n is the

expansion coefficient, σ is width parameter, $H_0(x) = 1$ and $H_1(x) = 2x$. The higher the Hermite polynomial order, the higher is its frequency of changes within the time domain, and the better is its capability to reconstruct quick changes of the ECG paradigm.

$$\begin{aligned} x(t) &= \sum_{n=0}^{N-1} c_n \phi_n(t, \sigma) \\ \phi_n(t, \sigma) &= \frac{1}{\sqrt{\sigma 2^n n! \sqrt{\pi}}} e^{\frac{-t^2}{2\sigma^2}} H_n\left(\frac{t}{\sigma}\right) \\ H_n(x) &= 2xH_{n-1}(x) - 2(n-1)H_{n-2}(x) \end{aligned} \tag{7.5}$$

For computing the statistical features, the beat is divided into five intervals, and Kurtosis and Skewness value over each interval is computed. For computing WABF, euclidean distance is calculated between R-peak and the characteristic waves such as P, Q, S, T-wave. The BLF produce a 994 dimensional feature vector. If the number of beats are less and the features are less than 994, then zero padding is performed. The obtained feature vector is combined with HR_{avg}, MSDRR, and mean QRS amplitude to make a 997 dimensional feature vector.

Demographic Features: The age and gender values are extracted for each patient. The age is normalised between 0 and 1. The negative and empty age fields are replaced with zero. One hot encoding (OHE) is performed for the gender attribute, where the male and female are encoded as (1,0) and (0,1), respectively. The gender with Nan value is encoded as (0,0). The OHE age and normalised gender attributes account for three features.

The final feature vector consisting of cleaned 10 second MECG of 5000 timestamps with L leads, and a 1000 dimensional feature vector are provided to the proposed PCNN-GAP with feature classifier. L depends on number of leads available in the dataset. For 2-lead ECG, $L = 2$.

7.4.2 PCNN-GAP Classifier

The proposed architecture of Parallel Convolution Neural Network - Global Average Pooling (PCNN-GAP) incorporated with age, gender, and heartbeat level features is illustrated in Figure 7.15. The idea behind the application of parallel convolution layers is that MECG consists of local and global patterns. The global patterns are extracted using the large kernels embedded in the left branch and the local patterns are extracted using the small kernels embedded in the right branch of PCNN-GAP

network. The number of filters, stride, and activation function of each convolution layer are mentioned in Figure 7.15. For instance, the first convolution layer in the left branch encompasses a 1-D convolution with 48 filters of size 19 and stride 4 followed by a Batch Normalization (BN) [274] layer and Rectified Linear Unit (ReLU) activation. Four more layers are added in a cascaded fashion in both branches with more number of filters of reduced size followed by 1-D GAP layer [225].

The GAP layer calculates the spatial average of filters, making it robust to spatial translations of MECG. GAP layer has the following advantages over the combination of flatten and fully connected (FC) layer:

1. Calculates the spatial average of filters, making the GAP layer robust to spatial translations present in ECG.
2. Less prone to overfitting with no dependency on external regularization.
3. Compresses spatial information of feature maps into channel description vectors.
4. No trainable parameters leads to significantly faster training time and reduction in model size, making the final model suitable for low resource devices.

The reduced parameters lead to significantly faster training and reduced model size, making it suitable for low resource devices. The GAP layer reduces the last layer dimensions from $(53, 112)$ to (112). The output of GAP layer is concatenated with other lead encodings along with the demographic features which are provided to the fully connected layer with 100 neurons having linear activation function. Finally, these encodings are classified using fully connected layer consisting 30 neurons with sigmoid activation function. Sigmoid layer generates scores ($s \in [0, 1]$) for 'scored' cardiac pathologies. The predicted scores are thresholded and labels are assigned only if the respective threshold is surpassed.

7.4.3 Observations

The last layer sigmoid predicts probabilities $p \in [0, 1]$ rather than predicting classes directly, thereby providing the flexibility to calibrate the threshold for class prediction. A default threshold of 0.5, leads to a negative outcome of 0 if $p \in [0.0, 0.49]$ and a positive outcome of 1 if $p \in [0.5, 1.0]$. The tuned threshold helps balance the amount of False Positive (FP) and False Negative (FN) during ECG classification. FP and FN represent disease prediction when there was no disease and no disease prediction when there was a disease, respectively. Experiments were performed to finalise sigmoid thresholding to be used for testing as illustrated in Figure 7.16. Thresholding the

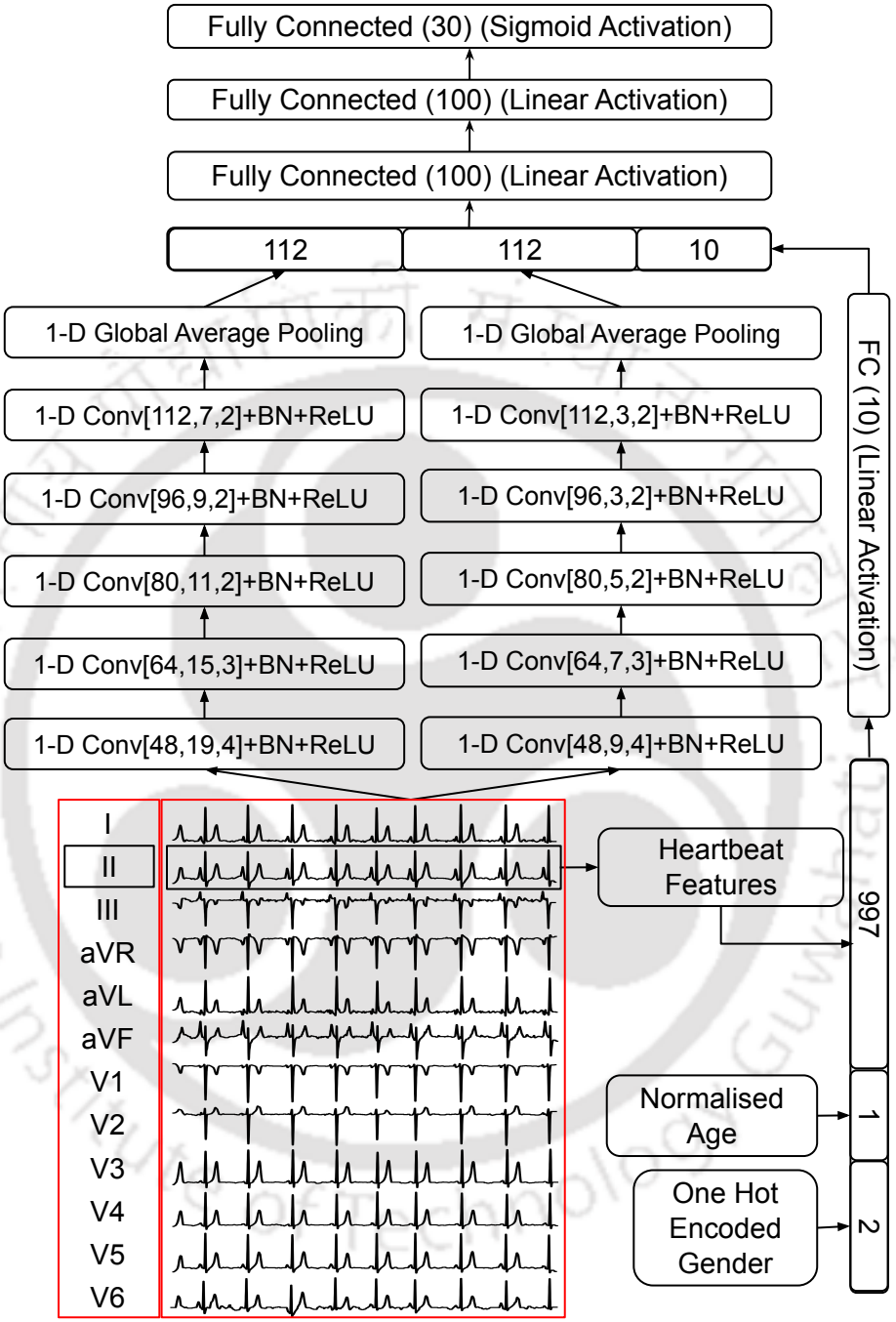


Figure 7.15: Architecture of Feature Fused PCNN-GAP Model.

scores at 0.1 produces optimum results for all evaluation metrics and deteriorates at other intervals. Therefore, thresholding at 0.1 is preferred for further model testing.

During training, early stopping criteria was employed and validation AUC was monitored for 5 epochs to avoid overfitting. Challenge scores for the 5-fold Cross Validation (CV) on the public training set, hidden validation, and test set for the

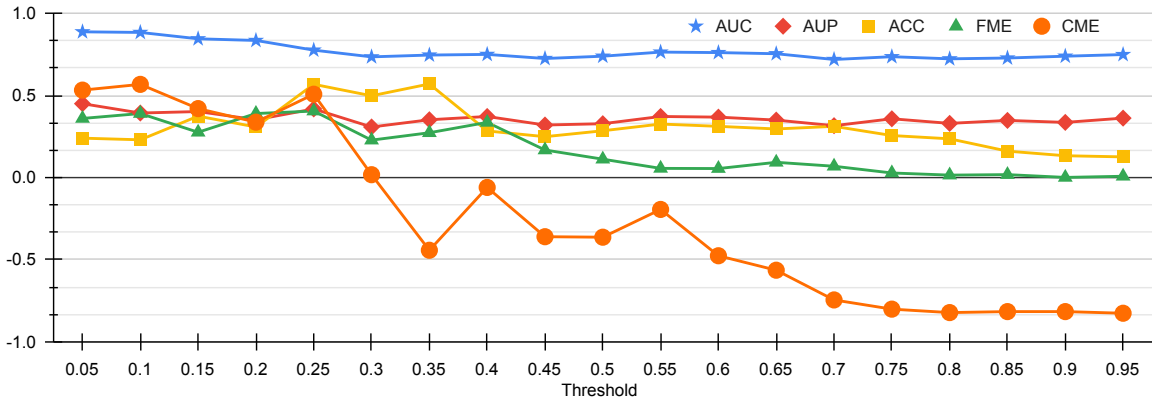


Figure 7.16: Performance deterioration with increasing threshold of sigmoid.

investigated models: PCNN-GAP; PCNN-GAP-Big; and PCNN-GAP-Feature are described in Table 7.6. While performing CV, MECG of the same patient was not included in both the training and validation set. The PCNN-GAP method consists of parallel CNN and GAP layer, PCNN-GAP-Big consists of two additional convolution layers in each branch, PCNN-GAP-Feature incorporates demographic and heartbeat features in PCNN-GAP model. The methods improved over 8-layer ResNet model that performed multi-class classification and predicted only single-label for MECG. In addition multiple lead configurations were also performed to find a reduced lead combination that could be used in sub-urban areas where standard 12-lead ECG is unavailable. Introducing multiple labels improved the score to 0.5. Fusing demographic and heartbeat features with MECG proved to be beneficial for classification for increased number of lead sets. Adding features improved the performance for more number of leads and reduced the performance for reduced lead sets. The model with less number of trainable parameters performed better for reduced lead sets and models with more number of trainable parameters performed better for standard 12-lead ECG. The training time of feature fused models was around 3000 seconds and 200 seconds for non-feature based models. Therefore, the heartbeat features are not used and only the demographic features with the CNN architecture is investigated to further improve the performance.

Table 7.7 describes the detailed performance evaluation of PCNN-GAP-Feature model for different test sets. The datasets in Table 7.7 include China Physiological Signal Challenge Test Database (CPSC2) [344], Georgia database (GE12EC), University of Michigan (UMich) and an Undisclosed American database (UDC) [8,46]. It can be observed that the model performed well for the CPSC2 and GE12EC test sets as these datasets had a similar distribution as the training datasets. But the model

Table 7.6: Comparison of PCNN-GAP-Feature with other investigated approaches.

Models	**PCNN-GAP**		**PCNN-GAP-Big**		**PCNN-GAP-Feature**		
Leads	**Training**	**Valid.**	**Training**	**Valid.**	**Training**	**Valid.**	**Test**
12	0.53±0.06	0.495	0.54±0.07	0.466	0.50±0.03	0.478	0.36
6	0.18±0.07	0.516	0.19±0.05	0.426	0.19±0.04	0.467	0.41
4	0.55±0.04	0.496	0.56±0.02	0.5	0.56±0.04	0.517	0.41
3	0.57±0.02	0.493	0.53±0.10	0.491	0.53±0.04	0.532	0.45
2	0.56±0.04	0.499	0.58±0.04	0.514	0.59±0.04	0.498	0.49

performance deteriorated when the Undisclosed and UMich datasets were used this means that the model is not generalized and needs improvement.

Table 7.7: Performance Evaluation of PCNN-GAP-Feature model for test sets.

Data	**CPSC2**					**G12EC**				
Lead	**AUC**	**AUP**	**ACC**	**FME**	**CME**	**AUC**	**AUP**	**ACC**	**FME**	**CME**
12	0.941	0.773	0.27	0.172	0.543	0.835	0.378	0.14	0.349	0.458
6	0.937	0.753	0.037	0.158	0.355	0.871	0.387	0.051	0.331	0.492
4	0.93	0.732	0.171	0.157	0.51	0.871	0.406	0.122	0.372	0.512
3	0.94	0.755	0.234	0.169	0.572	0.881	0.411	0.177	0.373	0.52
2	0.933	0.751	0.09	0.167	0.424	0.88	0.413	0.162	0.368	0.509
Data	**Undisclosed**					**UMich**				
Lead	**AUC**	**AUP**	**ACC**	**FME**	**CME**	**AUC**	**AUP**	**ACC**	**FME**	**CME**
12	0.907	0.534	0.261	0.335	0.256	0.858	0.383	0.163	0.358	0.376
6	0.896	0.549	0.204	0.349	0.415	0.872	0.405	0.054	0.334	0.391
4	0.902	0.548	0.19	0.345	0.291	0.88	0.426	0.145	0.379	0.429
3	0.91	0.547	0.231	0.344	0.32	0.883	0.429	0.217	0.382	0.474
2	0.908	0.576	0.33	0.369	0.465	0.877	0.426	0.221	0.379	0.506

The shortcomings of the model are as follows: (i) PCNN-GAP-Feature model was preferred for all lead configurations making the approach sub optimal as making the model with varying number of parameters might reduce the training and inference times. (ii) Not generalized as performance reduced on the datasets with different data distributions. (iii) The model still lacks interpretability and therefore fails to highlight the relevant signal timestamps responsible for the prediction.

7.5 Demographic Feature Fused MECG Classification and Interpretation using Dynamic CNN

In this stage, the observations from Section 7.3 and 7.4 are combined and a Channel Specific Dynamically built Convolution Neural Network (CSD-CNN) is developed

that detects cardiac abnormalities from MECG signals. Demographic features such as age and gender are incorporated into CSD-CNN to improve model performance. The dynamic construction eliminates manual effort and continuous modification. It retains less trainable parameters for reduced lead ECG, making the model indisposed to overfitting and memory-efficient. However, CSD-CNN's size and training time increases depending on the lead configuration. CSD-CNN also introduces an interpretability mechanism highlighting the crucial leads and relevant signal timestamps responsible for the prediction. The workflow followed encompasses a preprocessing stage, cardiac abnormality prediction stage, and interpretation stage as described in Figure 7.17. The preprocessing step extracts the selective lead ECG signals, age, and gender attributes which are provided to CSD-CNN that classifies the ECG segment into one or more cardiac abnormalities. The model is further exploited to extract lead and signal contributions towards the prediction.

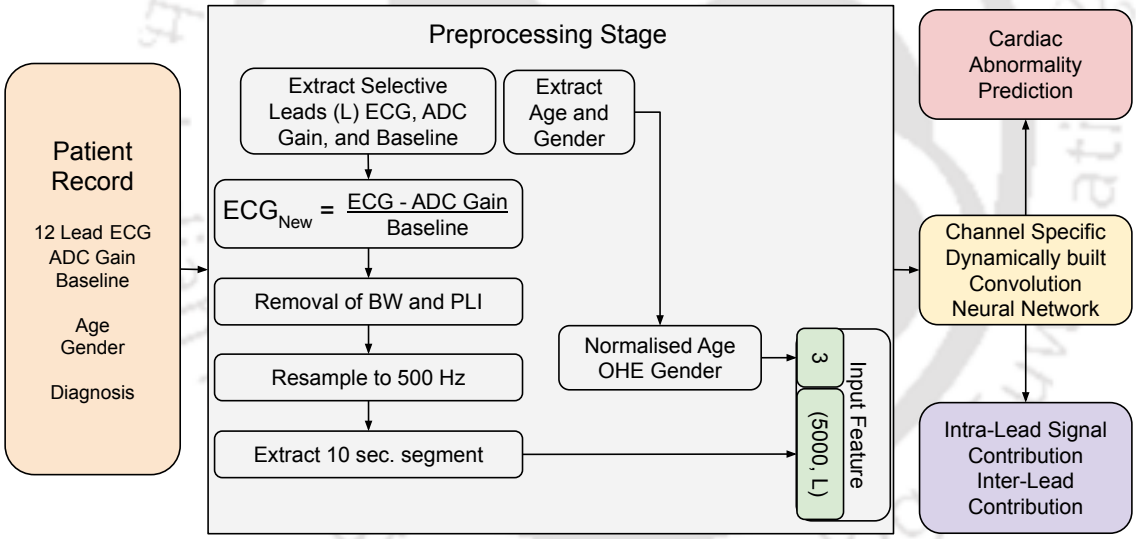


Figure 7.17: Illustration of the workflow.

Preprocessing Stage: The preprocessing applied is similar as described in Section 7.4.1 without the extraction of heartbeat features as they increased the training time without any significant performance gain. The label corresponding to each MECG segment is represented as a One-Hot Encoded (OHE) vector given as $y = [y_1, y_2, \ldots, y_C]$, where $C = 30$, total number of scored classes, and $y_i = 1$ if class i is positive and 0 otherwise. Demographic Features such as age and gender are extracted for each patient wherein age is normalized such that $age \in (0, 1)$ and gender is OHE where the male and female are encoded as (1,0) and (0,1), respectively. The negative and empty age fields are replaced with zero. The gender with Nan value is

encoded as (0,0). The normalized age and OHE gender attributes account for three additional features provided to the classifier. The final feature vector consists of a clean 10-second MECG with L leads and a 3 dimensional feature vector. The feature vector is provided to the CSD-CNN model for classification and interpretation.

7.5.1 Channel Specific Dynamically built CNN

The Channel Specific Dynamically built Convolution Neural Network (CSD-CNN) is inspired by [355], where the authors utilize the same convolution filters for each lead. In contrast, different filters for each lead are employed. The term "dynamic" refers to adding new branches with CNN block depending on the input lead configuration. Multiple parallel CNN blocks followed by the Global Average Pooling (GAP) layer are employed to produce final feature vectors. The architecture of CSD-CNN incorporated with demographic features is illustrated in Figure 7.18. The details of the CNN block are provided in Figure 7.19a. The CNN block uses an additional layer that adds input signal and penultimate layer convolution filter output. The addition layer allowed better loss convergence and improved validation AUC compared to concatenate, merge, and multiplication layers. Therefore, addition layer is chosen as compared to other layers. The intuition of dynamic CNN branches is as follows.

1. Parallel branches are built according to the leads present in the dataset, thereby eliminating manual intervention for different lead configurations.
2. Dynamic memory allocation for fewer leads makes the model computationally cheap in terms of time, memory, and computation power.
3. Model uses independent convolution filters to learn distinct features of each lead rather than using the same filter for all leads as each lead records the electrical activity from different angles and information learned by CNN of one lead might not be relevant for another lead.
4. Independent analysis of individual lead contribution and respective signal timestamps towards the candidate cardiac abnormality alleviates the black-box nature of deep learning models. The interpretations aid in locating regions causing the abnormality in ECG, thereby improving model reliability.

For making the model interpretable, the challenge was to generate the convolution filter outputs similar to the input signal. Therefore, the stride of convolution filter was kept to unity and varying size convolution filters are employed to account for global and local patterns in the ECG signal.

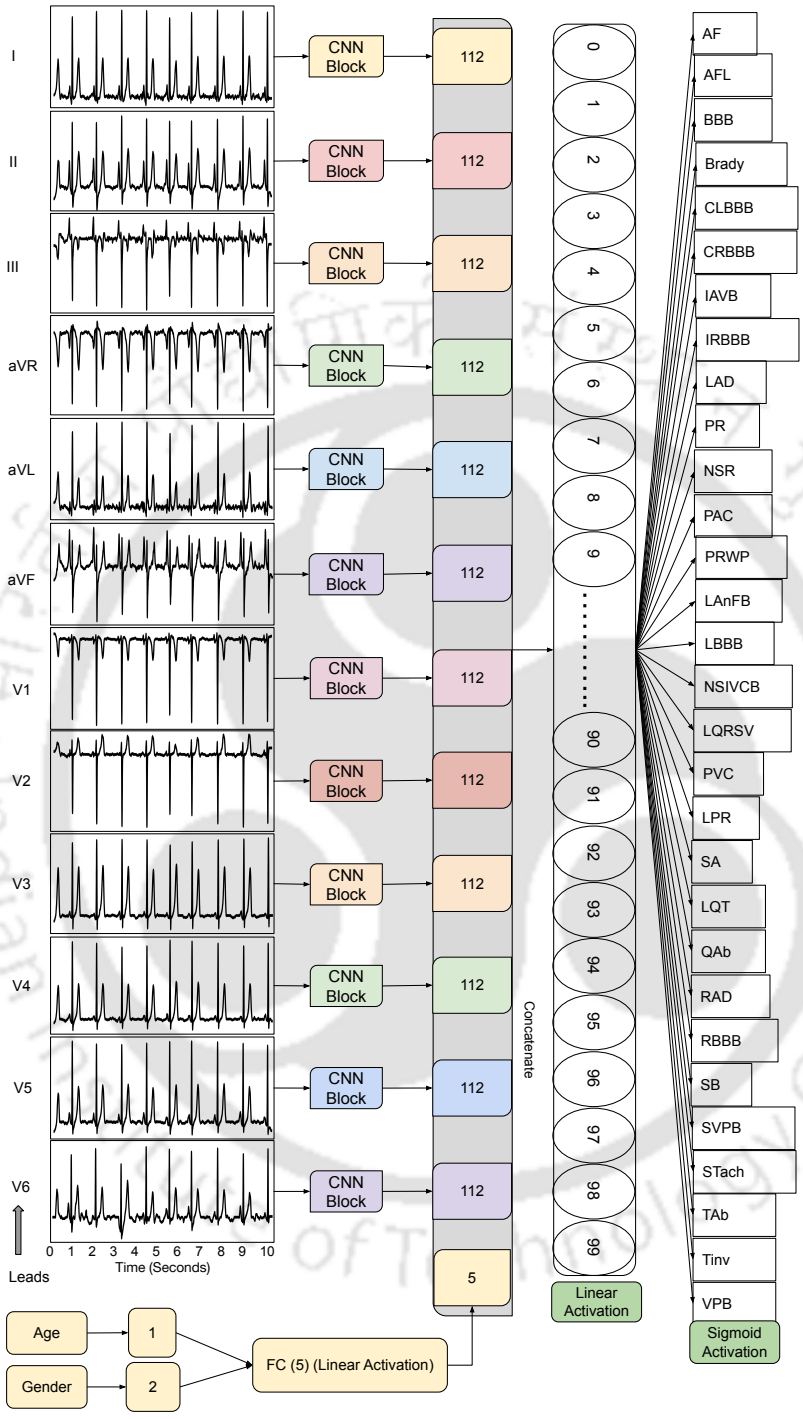
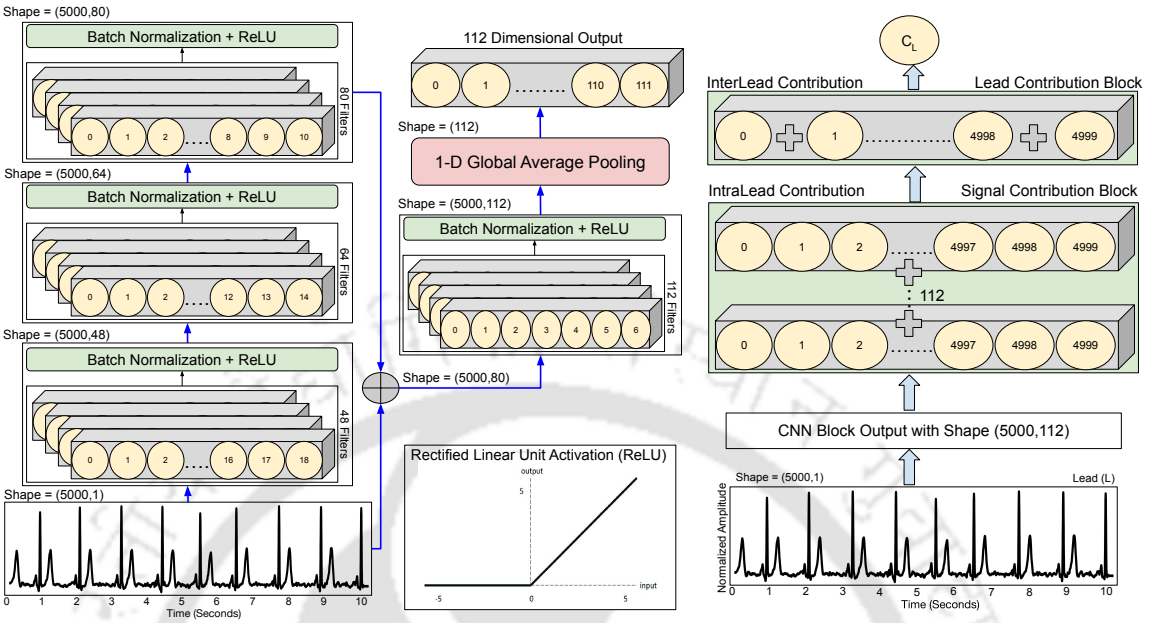


Figure 7.18: Illustration of Model Architecture for Standard 12 lead ECG. Models for reduced lead ECG could also be generated using similar architecture.

CSD-CNN Interpretation: The model generates visual interpretations during inference supporting the candidate cardiac abnormality prediction. The model uses input ECG for each lead, and the output of the penultimate CNN block of shape



(a) Expanded view of CNN Block.

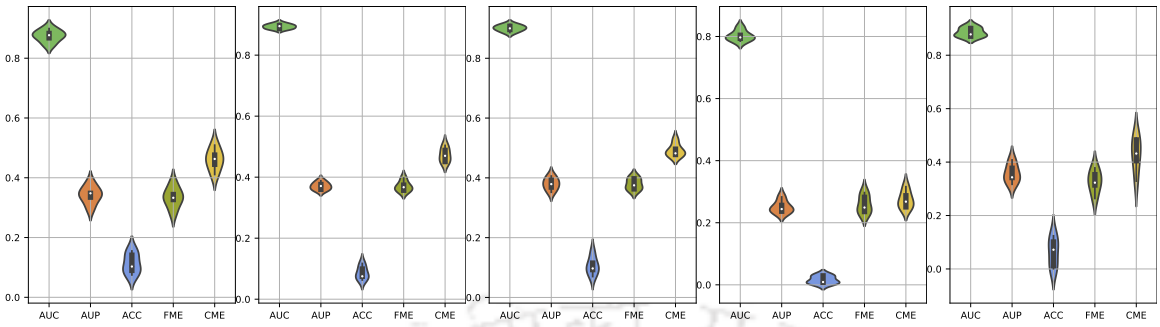
(b) Signal/Lead Contribution.

Figure 7.19: Illustration of Model Architecture and its Interpretation to generate Signal and Lead Contribution.

$(5000, 112)$ is used for interpretation, where 112 is the number of convolution filters. The CNN block summation along the channels generates $(5000, 1)$ dimensional output resembling the individual lead ECG shape depicting signal timestamp contribution towards the cardiac abnormality. Finally, the output summation along the filters generates a scalar value that determines the corresponding lead contribution towards the cardiac abnormality. The process of generating signal and lead contribution towards the cardiac abnormality prediction is described in Figure 7.19b. Theoretically, the highlighted timestamps and leads should correlate with the signs recognized by cardiologists as indicative of the respective candidate cardiac abnormality.

7.5.2 Quantitative Evaluation of CSD-CNN Model

The quantitative evaluation of the CSD-CNN model is provided in this section. The five-fold cross-validation results on training data for all lead configurations are provided in Figure 7.20. The model achieved high AUC but lower ACC for all leads. The AUP and FME were almost equivalent for all leads. All the metrics were consistent for 3, 4, and 6 lead models, but the variance was high for 2 and 12 lead models. The low variance in metrics made the models generalized and reliable for real-world use. The 6-lead model achieved lower performance than other lead models.



(a) 2 Lead (b) 3 Lead (c) 4 Lead (d) 6 Lead (e) 12 Lead

Figure 7.20: Violin Plots for evaluation metrics during five-fold cross-validation on training data for all lead configurations.

The challenge metric scores for all lead configurations for five-fold cross-validation on training data are provided in Table 7.8. While performing cross-validation, the stratified holdout MECG of the same patient was not included in both the training and validation set. The scores in bold highlight the best-performing model that is further used for validation and testing on unseen data. For 2, 3, 4, 6, 12-lead, the best performing model was developed in second, fourth, second, second, and fourth fold during cross-validation. Table 7.8 also describes the average run time per epoch (on our system) with trainable parameters in the respective CSD-CNN model for each lead. The time per epoch and the model parameters increased as CNN branches were dynamically added due to the increasing number of leads in the training dataset. The trainable parameters increased from 0.3 Million to 2.1 Million to account for additional information present in the standard 12-lead ECG configuration. The model adapts to available leads in the dataset and dynamically modifies the CSD-CNN architecture.

Table 7.8: Five fold cross validation results over challenge metric with execution time per epoch and number of trainable model parameters for all lead configurations.

Fold/Leads	**2**	**3**	**4**	**6**	**12**
1	0.476	0.47	0.476	0.248	0.483
2	**0.509**	0.451	**0.527**	**0.316**	0.405
3	0.445	0.488	0.482	0.251	0.329
4	0.463	**0.504**	0.478	0.287	**0.49**
5	0.411	0.45	0.496	0.268	0.432
Average	0.46 ± 0.02	0.47 ± 0.02	0.49 ± 0.02	0.27 ± 0.02	0.42 ± 0.06
RunTime(s)	1598	2154	2827	4161	7952
Parameters	359,938	538,082	716,226	1,072,514	2,141,378

The detailed performance of the best performing models for all lead configurations on different test sets for all evaluation metrics are provided in Table 7.9. The datasets in Table 7.9 include China Physiological Signal Challenge Test Database (CPSC2) [344], Georgia database (GE12EC), University of Michigan (UMich) and an Undisclosed American database (UDC) [8, 46]. Since the models are generalized, they performed well on the Georgia database, University of Michigan database, and the Undisclosed American database. The models achieved less performance for the CPSC2 dataset even though the models were trained on the China Physiological Signal Challenge Test Database. Since CPSC2 dataset was part of original CPSC dataset and early stopping criteria was used, the models did not overfit and therefore the performance deteriorated on CPSC2 dataset.

Table 7.9: Performance of best performing models on unseen test set.

Data	**CPSC2**					**G12EC**				
Lead	**AUC**	**AUP**	**ACC**	**FME**	**CME**	**AUC**	**AUP**	**ACC**	**FME**	**CME**
12	0.867	0.604	0.017	0.127	0.312	0.821	0.287	0.004	0.257	0.434
6	0.858	0.551	0.011	0.112	0.275	0.814	0.249	0.018	0.247	0.406
4	0.868	0.59	0.024	0.122	0.317	0.824	0.282	0.01	0.255	0.428
3	0.859	0.577	0.013	0.123	0.294	0.827	0.277	0.007	0.253	0.422
2	0.848	0.547	0.023	0.118	0.314	0.82	0.258	0.017	0.259	0.421
Data	**UDC**					**UMich**				
Lead	**AUC**	**AUP**	**ACC**	**FME**	**CME**	**AUC**	**AUP**	**ACC**	**FME**	**CME**
12	0.859	0.431	0.055	0.249	0.385	0.827	0.317	0.009	0.275	0.39
6	0.838	0.407	0.048	0.232	0.307	0.808	0.274	0.024	0.262	0.406
4	0.854	0.434	0.067	0.246	0.347	0.822	0.311	0.012	0.277	0.414
3	0.855	0.445	0.051	0.253	0.393	0.821	0.309	0.007	0.271	0.391
2	0.833	0.389	0.042	0.246	0.341	0.811	0.289	0.018	0.276	0.414

Table 7.10 describes the comparison between static CNN with and without features and the CSD-CNN model. The performance is calculated based on challenge metric scores for five-fold cross-validation on training, validation, and unseen test data. A challenge score of 0.393, 0.377, 0.397, 0.394, and 0.394 is achieved for 12, 6, 4, 3, and 2-lead ECG on the unseen hidden test set of CinC Challenge 2021 database [8,46]. Although static CNN models achieved good performance on training and validation data, performance on test data deteriorated which means that the models are not generalized. The CSD-CNN model achieved consistent performance for validation and test data for all lead configurations making the model more generalized and reliable for real-world use. The PCNN-Feature model achieved better scores for reduced lead configuration and lower scores for more leads. It may be due to static model architecture that provides enough parameters for fewer leads, but

the same number of parameters may be insufficient for increased leads. Therefore, the CSD-CNN architecture was made dynamic to account for information available in the standard lead configuration. CSD-CNN does not have a scarcity of trainable parameters and could, in theory, extract more information from the available data.

Table 7.10: Comparison between previously developed models and CSD-CNN model on the basis of challenge metric scores for five-fold cross validation on training data, validation data, and unseen test data.

	PCNN			**PCNN-Feature**			**Proposed CSD-CNN**		
Leads	**Training**	**Valid**	**Test**	**Training**	**Valid**	**Test**	**Training**	**Valid**	**Test**
12	0.53±0.06	0.495	–	0.50±0.03	0.478	0.36	0.43±0.07	0.414	0.393
6	0.18±0.07	0.516	–	0.19±0.04	0.467	0.41	0.27±0.03	0.385	0.377
4	0.55±0.04	0.496	–	0.56±0.04	0.517	0.41	0.49±0.02	0.412	0.397
3	0.57±0.02	0.493	–	0.53±0.04	0.532	0.45	0.47±0.02	0.4	0.394
2	0.56±0.04	0.499	–	0.59±0.04	0.498	0.49	0.46±0.04	0.399	0.394

The detailed challenge metric scores for the proposed CSD-CNN model and Feature-based PCNN for unseen test datasets for all lead models are provided in Table 7.11. Since the data distribution of CPSC2 and G12EC resembles training data, the models achieved lower scores as early stopping was applied during training, making the models generalized to prevent overfitting. The CSD-CNN models are more generalized than PCNN-Feature model, they achieved good performance for UDC and UMich datasets as their distribution is different from training datasets.

Table 7.11: Detailed challenge metric scores on test dataset for all leads.

	PCNN-Feature				**Proposed CSD-CNN**			
Leads	**CPSC2**	**G12EC**	**UDC**	**UMich**	**CPSC2**	**G12EC**	**UDC**	**UMich**
12	0.54	0.46	0.26	0.38	0.312	0.434	0.385	0.39
6	0.36	0.49	0.42	0.39	0.275	0.406	0.307	0.406
4	0.51	0.51	0.29	0.43	0.317	0.428	0.347	0.414
3	0.57	0.52	0.32	0.47	0.294	0.422	0.393	0.391
2	0.42	0.51	0.47	0.51	0.314	0.421	0.341	0.414

7.5.3 Qualitative Evaluation of CSD-CNN Model

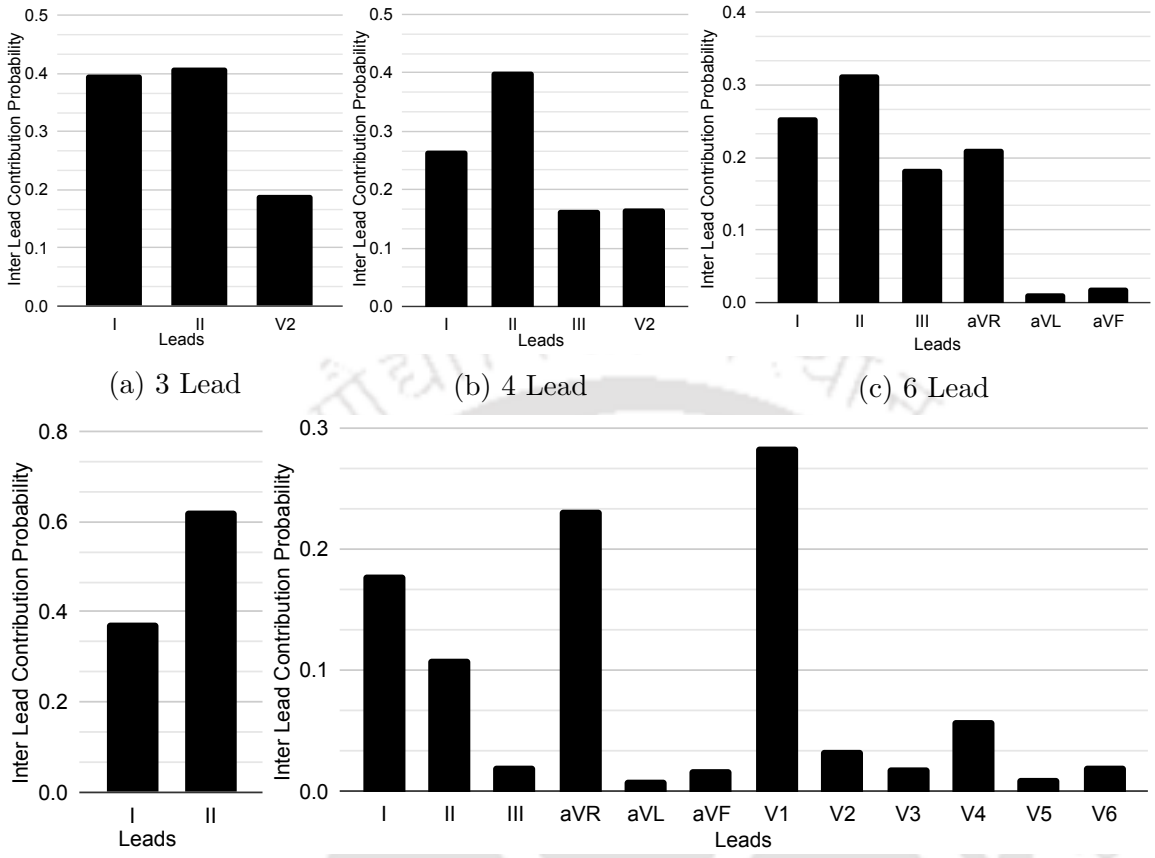
An efficient model with high performance on training data might fail on unseen test data acquired from a different source. Therefore performance alone is not sufficient to deploy a model into the real world. Here, interpretability comes into the picture that provides the reasoning about the model predictions. In the healthcare domain, reasoning plays an essential role as wrong decisions might be hazardous for the patient.

A generalized model should predict cardiac abnormalities based on fundamental ECG characteristics that reflect specific pathology rather than systematic bias present in the dataset. From this perspective, the intra-lead and inter-lead contributions are visualized through the convolution filters of the proposed CSD-CNN model.

Figure 7.21 describes the individual lead contribution during the final prediction over the whole dataset for correct classifications of various cardiac abnormalities for all lead configurations during inference. For two leads, lead II is mainly responsible, as depicted in Figure 7.21d. For three and four leads, lead I and II are primarily responsible, as illustrated in Figure 7.21a and 7.21b. For six leads, lead I, II, and aVR are primarily responsible, as illustrated in Figure 7.21c. For standard 12 leads, lead I, II, aVR, and V1 are mainly responsible, as depicted in Figure 7.21e. It can be inferred that the CSD-CNN model assigns a higher weight to lead II for almost all lead configurations. The physiological reason might be the orientation of the heart and lead II, which captures the most beneficial electrical information generated from heart cells.

However, the lead contributions had few exceptions when a close inspection was performed, as illustrated in Figure 7.22. For four lead configurations, V2 also contributed significantly for pathologies like BBB, LBBB, and LAD as depicted in Figure 7.22a, 7.22b, and 7.22d. Lead III contributed significantly to LAnFB pathology as depicted in Figure 7.22e. For six lead configurations, lead III had a significant contribution in NSR and LAD pathologies as depicted in Figure 7.22c and 7.22f. It can be inferred that lead importance is disease dependent.

The contribution of signal timestamps in each lead for correctly classified AF pathology during inference is illustrated in Figure 7.23. A candidate cardiac abnormality such as AF is selected to describe the model interpretations. The same can be extended to other cardiac abnormalities. The most contributing lead II in the two lead model highlights more relevant signal timestamps than lead I. Although the summation of CNN filter outputs focuses on the regions around R-peaks, the fibrillation frequencies or missing P-wave are also densely highlighted, which correlates with the medical definition of AF [364]. According to medical practitioners, the missing P-wave and F-wave or fibrillation frequencies provide enough information to classify the corresponding ECG segment as AF. For the three lead model, lead II highlights relevant timestamps with high significance compared to other leads. The filters focus from Q-onset to T-onset with moderately high focus F-waves. For the four lead model, lead I, II, and III highlight timestamps with high attribution scores following a similar pattern observed in the two lead model. For the six lead model, lead I,



(a) 3 Lead (b) 4 Lead (c) 6 Lead

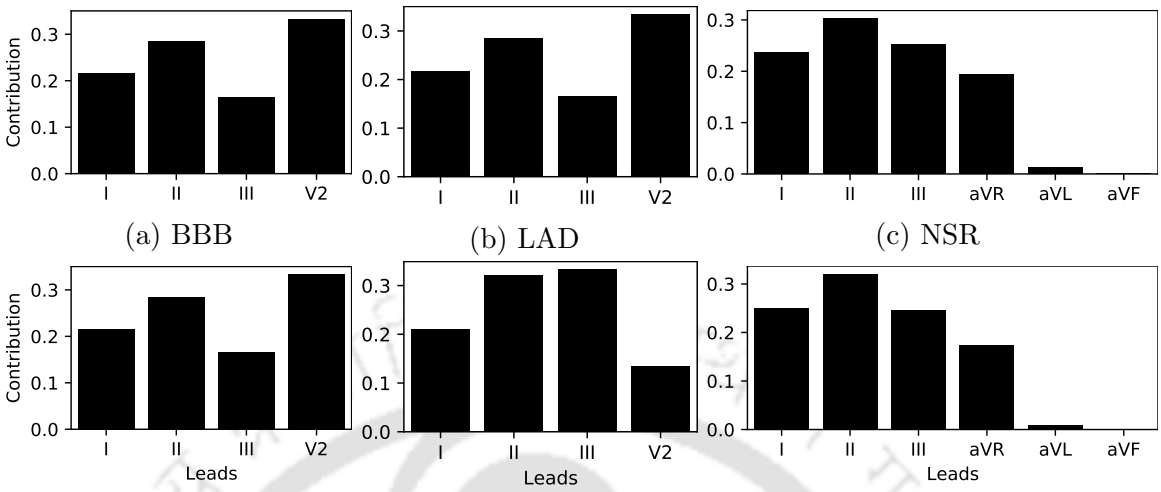
(d) 2 Lead (e) 12 Lead

Figure 7.21: Individual lead contribution over the full dataset for correct classifications for all lead configurations during inference.

II, and aVR have higher attribution scores than lead III, aVL, and aVF. For the standard 12-lead model, lead I, II, aVR, and V1 highlight relevant timestamps with significantly high attribution scores compared to other leads. The extracted attribution scores from convolution filter outputs are inherent properties of the underlying models. For simplicity, the focus lies on a single CSD-CNN model rather than a wide variety of models. Through visual inspection, the most typical attribution pattern for a candidate cardiac pathology, AF, is demonstrated in detail for all lead models, and the same can be extracted for other pathologies.

7.6 Summary

This chapter proposes a novel approach for detecting cardiac abnormalities from MECG signals. The work has been performed in three stages. Firstly, five categories of models are investigated for single-label 12-lead ECG Classification. The ResNet



(a) BBB (b) LAD (c) NSR

(d) LBBB (e) LAnFB (f) LAD

Figure 7.22: Exceptions in individual lead contribution over the full dataset for few pathologies for four and six lead configurations.

achieved better performance and significantly less training and inference time than recurrent and attention-based models. However, the model achieved a negative challenge score for test data due to the last layer softmax. Secondly, Heartbeat and Demographic Feature Fused MECG Classification is performed using Static CNN, wherein the softmax activation was replaced with sigmoid to predict multiple cardiac abnormalities for a segment. Thresholding the sigmoid at 0.1 produced the best results for all evaluation metrics and deteriorated at other thresholds. Incorporating heartbeat features increased the training time by orders of magnitude without providing a significant performance gain. Lastly, Demographic Feature Fused MECG Classification and Interpretation using a CSD-CNN model. CSD-CNN eliminated manual intervention and used less trainable parameters, making the model memory efficient and less prone to overfitting and generalized than existing models as testing is performed on multiple datasets. The model provides information about the most contributing ECG leads and the respective signal contribution towards the pathologies. The highlighted regions of interest often correlate with the cardiologists' interpretation of the abnormalities. The intra-lead and inter-lead modules contribute individually to improve the model performance by focusing on relevant information within and across the channels. The results demonstrate the high prospects of algorithmic ECG analysis for future clinical applications considering both quantitative performance and the possibility of assessing decision criteria on a per-example basis, which enhances the comprehensibility of the approach.

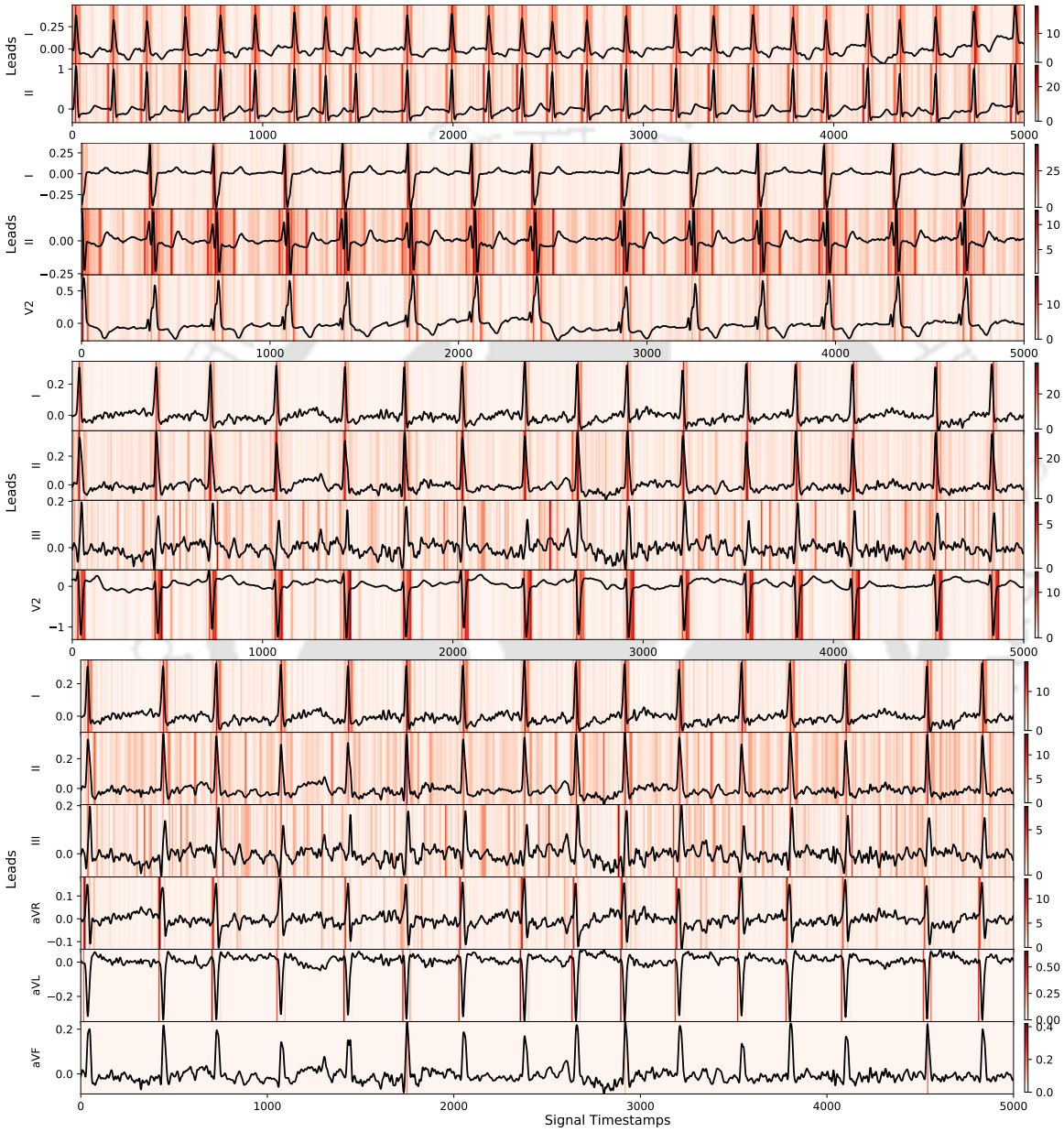


Figure 7.23: Signal contribution in each lead for multichannel ECG for AF pathology for different lead configurations. (Top to Bottom: Two, Three, Four, and Six lead). The attribution scores for each signal timestamp are represented with the amount of red colour assigned to the timestamp.

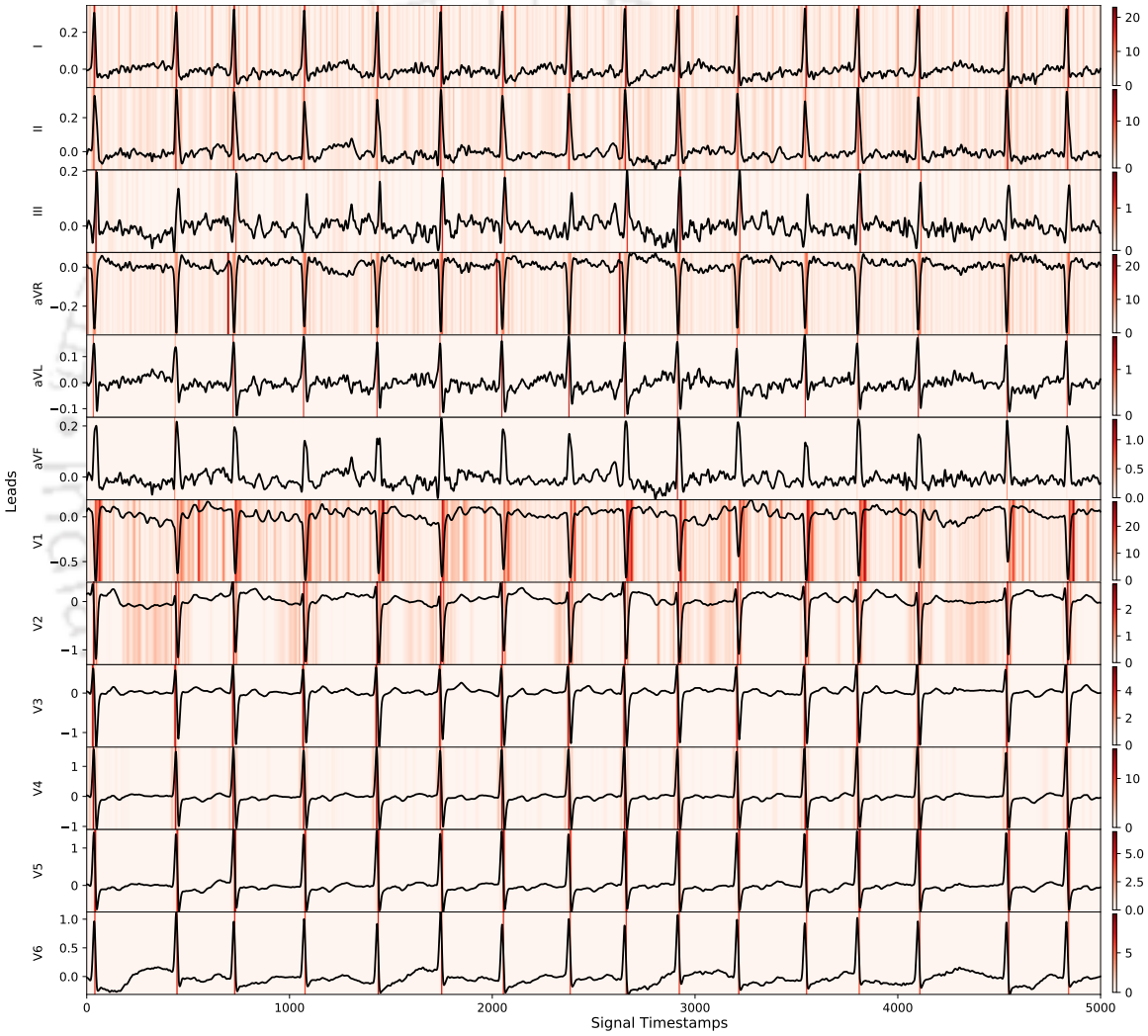


Figure 7.23: (Continued..) Signal contribution in 12-lead ECG for AF pathology.

8

Conclusions and Future Work

In this thesis, the cardiac abnormality detection from ECG signals is performed using deep learning models with an interpretability perspective. The contributions made in this thesis aim to develop a classification system that provides explanations of its predictions. The classification approaches investigated in this thesis can be categorised into: single lead heartbeat classification, single lead rhythm classification, and multichannel ECG classification. Several interpretability based techniques have been investigated for the developed classifier. This chapter presents a summary of the contributions made within the scope of the thesis and then concludes with the future research and directions for the extension of this work.

8.1 Summary of the Thesis

In Chapter 2, background is presented that provides an overview of the cardiovascular system which includes electrophysiology behind heartbeat genesis, history of clinical ECG, its acquisition. Then individual components of an ideal ECG classification framework are discussed, including noise removal, heartbeat segmentation, heartbeat synthesis, cardiac abnormality classification, and the interpretability landscape. In interpretability landscape, definition of interpretability is discussed, the difference between interpretability and explainability, the criteria for developing an interpretable model, taxonomy of interpretability methods, and the perspective of interpretability in cardiology are discussed.

In Chapter 3, ECG preprocessing is described where single lead ECG signal is decluttered from Low-Frequency BW using VMD and compared with signal decomposition techniques such as EMD, EEMD, CEEMDAN, and filtering techniques such as

mean and mean median filtering. The higher the complexity of the baseline, the more execution time the algorithm took for signal decomposition. CEEMDAN execution time exceeds by a huge margin as compared to other approaches. The comparison depicted that VMD estimates better baseline as compared to other techniques in terms of PC, PRMSD, and MAE. However, the time required to decompose the signal is relatively higher than the filtering techniques. The decluttered signal is used to extract heartbeats (R-Peaks) using fractal based mathematical morphological operators such as erosion and dilation, that are implemented using dynamic programming with memoization. The method achieved an overall sensitivity of 95.82%, predictivity of 95.15%, and DER of 10.16% when tested against all the records available in the MIT-BIH dataset. Lastly, a deep convolution conditional generative adversarial network is employed to synthesise different classes of heartbeats, including N, SVEB, and VEB as recommended by AAMI. The generated heartbeats are verified quantitatively using five evaluation metrics and qualitatively by plotting against original beats. Generated heartbeats resemble real beats as they encompass essential characteristics and follow the intricate structure of beats.

In Chapter 4, a penalty induced prototype based explainable residual neural network is proposed for heartbeat classification and prediction explanation. PIPxResNet extracts prototypes from heartbeats using pretrained residual neural network and minimizes resource consumption by eliminating GPU requirement. The method classifies heartbeats into N, SVEB, and VEB and also provides explanations in the form of visual prototypes that support model diagnosis. Quantitatively, PIPxResNet achieved state-of-the-art performance without compromising individual class performance when tested on four publicly available standard datasets. Qualitatively, the model simulates a clinical decision support system for general physicians by providing explanation through extracted prototypes, addressing the black-box nature of deep neural networks. The lower dimension data achieved better performance with significant reduction in training time.

In Chapter 5, three gradient backpropagation based posthoc interpretability techniques are investigated for explaining VT-VF predictions using ResNet from ECG signals. The models were tested on five publicly available datasets. The effect of layers of the ResNet model, augmentation technique, and segment length were analysed and the optimum performance was obtained for single layer ResNet for one second segment using ADASYN augmentation technique. The interpretability techniques include Guided Backpropagation, Gradient Class Activation Map, and Guided Gradient Class Activation Map that highlight the signal timestamps responsible for VT-VF de-

tection. Grad CAM highlighted relevant signal timestamps whereas GBP and Guided Grad CAM mostly worked as peak detectors which aid in heartbeat detection. The validity of these techniques is verified using sanity checks, namely, weight randomization and training data permutation. All techniques passed the sanity checks but Grad CAM saliency maps deteriorated the most, making it a better choice for signal interpretation. Single layer ResNet with ADASYN augmentation and Grad CAM highlighted most relevant signal timestamps responsible for prediction.

In Chapter 6, AF is detected from single lead ECG signals using baseline CNN, ResNet, attentive CNN, and transformer neural network. ResNet achieved the highest F1 Score followed by the transformer neural network, baseline CNN, and attentive CNN on PhysioNet Computing in Cardiology Challenge 2017 database. The length of ECG segment between one to four seconds was found optimum in terms of both quantitative and qualitative analysis. The attentive CNN and transformer models provided interpretability by highlighting the essential characteristic waves responsible for and against the correct and miss classified ECG rhythms. Moreover, the transformer model is better suited than attentive CNN as it highlighted more relevant characteristic waves as demonstrated in the experiments.

In Chapter 7, a novel interpretable methodology is proposed to detect multiple cardiac abnormalities from MECG signals in three stages. In the first stage, five categories of models are investigated for single label 12-lead ECG Classification. The models include CNN, RNN, combination of CNN and RNN, RNN with interpretability based attention mechanism, and combination of CNN, RNN, and Attention Mechanism. It was observed that a ResNet achieved better performance with less training and inference time as compared to recurrent and attention-based models. However, the model consistently predicted a single cardiac abnormality for the records that wither had multiple diagnosis or had other abnormalities than the training labels, thereby achieving a negative score for test data. In the second stage, heartbeat and demographic features are fused with MECG and Multilabel Classification is performed using a Static CNN with softmax replaced with sigmoid activation function in the last layer. The effect of thresholding was observed and a threshold of 0.1 produced best results for all evaluation metrics. It was also observed that large convolution filters achieved higher performance as compared to small kernel sizes. Incorporating heartbeat features increased the training time by orders of magnitude without providing a significant performance gain. In the third stage, only demographic features were incorporated with MECG and a CSD-CNN model was used for multilabel classification. The dynamic nature of CSD-CNN eliminated manual intervention with less

trainable model parameters, less prone to overfitting, and more generalized than previous models. CSD-CNN introduces interpretability by providing information about the most contributing ECG leads and the respective signal contribution towards the pathologies. The intra-lead and inter-lead modules contribute individually to improve the model performance by focusing on relevant information within and across the channels.

8.2 Future Research Directions

The work performed in this thesis opens up several directions for future research.

The fractal-based beat detection approach proposed in Chapter 3 could be tested on additional datasets to test the robustness of the algorithm. The hard thresholding in the post-processing stage could be modified to adaptive thresholding for reduced false negatives or missed beats. A sample-by-sample approach could be investigated to improve the detection accuracy rather than windowing the signal in the post-processing stage.

The DCCGAN proposed in Chapter 3 for heartbeat synthesis could be extended to generate single-lead or multi-lead ECG rhythms to mitigate the data imbalance while performing rhythm classification. The approach could also be extended for other physiological signals such as electroencephalogram, photoplethysmogram, and phonocardiogram signals.

PIPxResNet proposed in Chapter 4 for single-lead heartbeat classification could be extended to include counter-examples to explain the wrong predictions along with the supporting prototype-based explanations for correct predictions. The explanations could also be provided in some human-understandable clinical rules, such as the amplitude and width of ECG characteristic waves. PIPxResNet could also be extended to other classification problems where predictions can be explained through prototypes.

The single-channel rhythm classification performed in Chapter 5 and 6 through convolution neural networks and interpretation by visualizing penultimate layer convolution filters could be extended such that saliency maps of all convolution layers of the network are utilized to generate better interpretations, to increase the confidence of the user in the model predictions and explanations.

Recently, PhysioNet challenges 2020 and 2021 have open-sourced 12-lead ECG databases, enabling the development of generalized classifiers to detect a large group of cardiac abnormalities. It is evident from Chapter 7 that the proposed and state-

of-the-art models have not yet achieved significant results, and there is a scope of improvement. The parallel CNN branches in the proposed CSD-CNN model independently learn the information from each lead ECG. The branches are not influenced by other lead ECGs, deteriorating the model's performance. This could be mitigated by using a guided signal from one CNN branch to another, which could make the model more effective for classification and generate better interpretations for explaining the predicted diagnosis.

The CSD-CNN model in Chapter 7 predicts several cardiac abnormalities for a MECG segment by using a fixed threshold in the last layer sigmoid activation function. A constant threshold might not suit all cardiac abnormalities and could be varied with respect to different cardiac abnormalities to achieve better performance for MECG classification.

Publications

Refereed Conference Publications

1. **Deepankar Nankani** and Rashmi Dutta Baruah. "An End-to-End framework for automatic detection of Atrial Fibrillation using Deep Residual Learning." In IEEE Region 10 Conference (TENCON), pp. 690-695. IEEE, 2019 [263].

2. **Deepankar Nankani** and Rashmi Dutta Baruah. "Effective Removal of Baseline Wander from ECG Signals: A Comparative Study." In International Conference on Machine Learning, Image Processing, Network Security and Data Sciences (MIND), pp. 310-324. Springer, Singapore, 2020 [228].

3. **Deepankar Nankani** and Rashmi Dutta Baruah. "Investigating Deep Convolution Conditional GANs for Electrocardiogram Generation." In International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2020 [91].

4. **Deepankar Nankani**, Pallabi Saikia, and Rashmi Dutta Baruah, "Automatic Concurrent Arrhythmia Classification using Residual Neural Networks", In Computing in Cardiology (CinC), September 13-16, 2020, Rimini, Italy [264].

5. **Deepankar Nankani** and Rashmi Dutta Baruah, "Incorporating Demographic and Heartbeat Features with Multichannel ECG for Cardiac Abnormality Detection using Parallel CNN and GAP Network", In Computing in Cardiology (CinC), September 13-16, 2021, Brno, Czech Republic [365].

6. **Deepankar Nankani** and Rashmi Dutta Baruah, "Improved Diagnostic Performance of Arrhythmia Classification Using Conditional GAN Augmented Heartbeats", In Springer Book entitled "Generative Adversarial Learning: Architectures and Applications", 2021 **(In Press)**.

7. **Deepankar Nankani**, Bhagath Parabattina, Rashmi Dutta Baruah, and Pradip K. Das, "R-Peak Detection from ECG Signals Using Fractal Based Mathematical Morphological Operators", In IEEE Region 10 Conference (TENCON) 2021 **(In Press)**.

8. **Deepankar Nankani** and Rashmi Dutta Baruah, "Ventricular Arrhythmia Classification and Interpretation Using Residual Neural Network with Guided Backpropagation", In IEEE Region 10 Conference (TENCON) 2021 **(In Press)**.

9. **Deepankar Nankani** and Rashmi Dutta Baruah, "Classification and Prediction Explanation of Atrial Fibrillation using Transformer Neural Network", In 2022 International Joint Conference on Neural Networks (IJCNN).

Refereed Journal Publications

1. **Deepankar Nankani** and Rashmi Dutta Baruah, "PIPxResNet: Penalty Induced Prototype-Based eXplainable Residual Neural Network for Heartbeat Classification", In Neural Computing and Applications, Springer. **(Under Review)**

2. **Deepankar Nankani** and Rashmi Dutta Baruah, "Posthoc Interpretability Techniques for Explaining Ventricular Arrhythmia Prediction using Residual Neural Network", In Neural Computing and Applications, Springer. **(Under Review)**

3. **Deepankar Nankani** and Rashmi Dutta Baruah, "Feature Fused Multichannel ECG Classification using Channel Specific Dynamic CNN for Detecting and Interpreting Cardiac Abnormalities", In Special Issue: Classification of Multilead ECGs, Physiological Measurement. **(Under Review)**

Other Publications

1. Pallabi Saikia, **Deepankar Nankani**, and Rashmi Dutta Baruah, "Seismic Signal Interpretation for Reservoir Facies Classification", International Conference on Pattern Recognition and Machine Intelligence (PReMI), December 17-20, 2019, Tezpur, Assam, Lecture Notes in Computer Science, vol 11942. Springer, Cham, pp. 409-417. [366]

2. Pallabi Saikia, **Deepankar Nankani**, and Rashmi Dutta Baruah, "Reservoir Facies Classification using Convolutional Neural Networks", in proceedings of IEEE TENCON Geoscience and Remote Sensing Conference (TENGARSS) October 17-2019, Kochi, Kerala, pp. 34-38 [367].

3. **Deepankar Nankani**, Shivam Maurya, and Rashmi Dutta Baruah, "Model Compression Techniques for Atrial Fibrillation Detection on Mobile Devices", International Conference on Pattern Recognition and Machine Intelligence (PReMI), 2021 **(In Press)**.

Vitae

Deepankar Nankani joined the Dual Degree (M.Tech. + Ph.D.) program at the Department of Computer Science and Engineering of Indian Institute of Technology Guwahati, India in July 2015. Prior to joining PhD, he did his Bachelors of Technology in Electronic and Communication Engineering from Jaipur Engineering College and Research Centre, Rajasthan Technical University, Jaipur, Rajasthan, India. He has keen interests in pursuing the field of machine learning, deep learning, biomedical signal processing and designing interpretable deep learning models for the healthcare domain. He enjoys traveling to the places where nature has bestowed its bounty.

Contact Information

Email	:	d.nankani@iitg.ac.in, dnankani8@gmail.com
Address	:	47, Basant Bahar Colony Gopalpua Mod, Tonk Road, Jaipur Rajasthan - 302018 INDIA

Bibliography

[1] G. A. Roth, G. A. Mensah, C. O. Johnson, G. Addolorato, E. Ammirati, L. M. Baddour, N. C. Barengo, A. Z. Beaton, E. J. Benjamin, C. P. Benziger *et al.*, "Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study," *Journal of the American College of Cardiology*, vol. 76, no. 25, pp. 2982–3021, 2020.

[2] "Ecg conduction system," https://brilliantnurse.com/nclex-cardiac-electrical-conduction-system/.

[3] M. AlGhatrif and J. Lindsay, "A brief review: history to understand fundamentals of electrocardiography," *Journal of community hospital internal medicine perspectives*, vol. 2, no. 1, p. 14383, 2012.

[4] Cables and S. LLC, "12-lead ecg placement guide with illustrations," https://www.cablesandsensors.com/pages/12-lead-ecg-placement-guide-with-illustrations.

[5] U. R. Acharya, H. Fujita, S. L. Oh, U. Raghavendra, J. H. Tan, M. Adam, A. Gertych, and Y. Hagiwara, "Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network," *Future Generation Computer Systems*, vol. 79, pp. 952–959, 2018.

[6] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, "Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[7] R. Mark, "Aami-recommended practice: Testing and reporting performance results of ventricular arrhythmia detection algorithms," *Association for the Advancement of Medical Instrumentation, Arrhythmia Monitoring Subcommittee, AAMI ECAR, 1987*, 1987.

[8] E. A. Perez Alday, A. Gu, A. Shah, C. Robichaux, A.-K. I. Wong, C. Liu, F. Liu, B. A. Rad, A. Elola, S. Seyedi, Q. Li, A. Sharma, G. D. Clifford, and M. A. Reyna, "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiological Measurement*, vol. 41, 2020.

[9] S. S. Virani, A. Alonso, H. J. Aparicio, E. J. Benjamin, M. S. Bittencourt, C. W. Callaway, A. P. Carson, A. M. Chamberlain, S. Cheng, F. N. Delling, M. S. V. Elkind, K. R. Evenson, J. F. Ferguson, D. K. Gupta, S. S. Khan, B. M. Kissela, K. L. Knutson, C. D. Lee, T. T. Lewis, J. Liu, M. S. Loop, P. L. Lutsey, J. Ma, J. Mackey, S. S. Martin, D. B. Matchar, M. E. Mussolino, S. D. Navaneethan, A. M. Perak, G. A. Roth, Z. Samad, G. M. Satou, E. B. Schroeder, S. H. Shah, C. M. Shay, A. Stokes, L. B. VanWagner, N.-Y. Wang, C. W. Tsao, and American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee, "Heart Disease and Stroke Statistics – 2021 Update: a Report from the American Heart Association," *Circulation*, vol. 143, no. 8, pp. e254–e743, 2021.

[10] M. D. Huffman, D. Prabhakaran, C. Osmond, C. H. Fall, N. Tandon, R. Lakshmy, S. Ramji, A. Khalil, T. Gera, P. Prabhakaran, S. D. Biswas, K. S. Reddy, S. K. Bhargava, and H. S. Sachdev, "Incidence of cardiovascular risk factors in an indian urban cohort: Results from the new delhi birth cohort," *Journal of the American College of Cardiology*, vol. 57, no. 17, pp. 1765–1774, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0735109711005717

[11] P. Kligfield, "The Centennial of the Einthoven Electrocardiogram," *Journal of Electrocardiology*, vol. 35, no. 4, pp. 123–129, 2002.

[12] B. J. Drew, M. M. Pelter, D. E. Brodnick, A. V. Yadav, D. Dempel, and M. G. Adams, "Comparison of a new reduced lead set ecg with the standard ecg for diagnosing cardiac arrhythmias and myocardial ischemia," *Journal of electrocardiology*, vol. 35, no. 4, pp. 13–21, 2002.

[13] B.-U. Kohler, C. Hennig, and R. Orglmeister, "The principles of software qrs detection," *IEEE Engineering in Medicine and Biology Magazine*, vol. 21, no. 1, pp. 42–57, 2002.

[14] F. K. Došilović, M. Brčić, and N. Hlupić, "Explainable artificial intelligence: A survey," *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp. 0210–0215, 2018.

[15] A. Abdul, J. Vermeulen, D. Wang, B. Y. Lim, and M. Kankanhalli, "Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda," *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–18, 2018.

[16] M. A. Escalona-Morán, M. C. Soriano, I. Fischer, and C. R. Mirasso, "Electrocardiogram classification using reservoir computing with logistic regression,"

IEEE Journal of Biomedical and health Informatics, vol. 19, no. 3, pp. 892–898, 2014.

[17] T. Ince, S. Kiranyaz, and M. Gabbouj, "A generic and robust system for automated patient-specific classification of ecg signals," *IEEE Transactions on Biomedical Engineering*, vol. 56, no. 5, pp. 1415–1426, 2009.

[18] A. Bibal and B. Frénay, "Interpretability of machine learning models and representations: an introduction." *ESANN*, 2016.

[19] M. A. Ahmad, C. Eckert, and A. Teredesai, "Interpretable machine learning in healthcare," *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*, pp. 559–560, 2018.

[20] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM computing surveys (CSUR)*, vol. 51, no. 5, pp. 1–42, 2018.

[21] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial intelligence*, vol. 267, pp. 1–38, 2019.

[22] A. A. Freitas, "Comprehensible classification models: a position paper," *ACM SIGKDD explorations newsletter*, vol. 15, no. 1, pp. 1–10, 2014.

[23] Z. Lipton, "The doctor just won't accept that! interpretable ml symposium," *31st conference on neural information processing systems (NIPS 2017)*, 2017.

[24] T. Miller, P. Howe, and L. Sonenberg, "Explainable ai: Beware of inmates running the asylum or: How i learnt to stop worrying and love the social and behavioural sciences," *arXiv preprint arXiv:1712.00547*, 2017.

[25] G. Ras, M. Van Gerven, and H. Pim, "Explanation methods in deep learning: Users, values, concerns and challenges," *Explainable and Interpretable Models in Computer Vision and Machine Learning*, pp. 19–36, 2018.

[26] D. Wang, Q. Yang, A. Abdul, and B. Y. Lim, "Designing theory-driven user-centric explainable ai," *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–15, 2019.

[27] B. Y. Lim, Q. Yang, A. M. Abdul, and D. Wang, "Why these explanations? selecting intelligibility types for explanation goals." *IUI Workshops*, 2019.

[28] M. Ribera and A. Lapedriza, "Can we do better explanations? a proposal of user-centered explainable ai." *IUI Workshops*, vol. 2327, p. 38, 2019.

[29] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," *arXiv preprint arXiv:1702.08608*, 2017.

[30] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable ai systems," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 11, no. 3-4, pp. 1–45, 2021.

[31] P. Angelov and E. Soares, "Towards explainable deep neural networks (xdnn)," *Neural Networks*, vol. 130, pp. 185 – 194, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608020302513

[32] R. Mark, P. Schluter, G. Moody, P. Devlin, and D. Chernoff, "An annotated ecg database for evaluating arrhythmia detectors," *IEEE Transactions on Biomedical Engineering*, pp. 600–600, 1982.

[33] R. Mark, G. Moody, and S. Greenwald, "Mit-bih supraventricular arrhythmia database," 1990.

[34] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[35] Z. Cai, C. Liu, H. Gao, X. Wang, L. Zhao, Q. Shen, E. Ng, and J. Li, "An open-access long-term wearable ecg database for premature ventricular contractions and supraventricular premature beat detection," *Journal of Medical Imaging and Health Informatics*, vol. 10, no. 11, pp. 2663–2667, 2020.

[36] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

[37] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 618–626, 2017.

[38] G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[39] S. D. Greenwald, "The development and analysis of a ventricular fibrillation detector," Ph.D. dissertation, Massachusetts Institute of Technology, 1986.

[40] F. Nolle, F. Badura, J. Catlett, R. Bowser, and M. Sketch, "Crei-gard, a new concept in computerized arrhythmia monitoring systems," *Computers in Cardiology*, vol. 13, pp. 515–518, 1986.

[41] J. Zheng, G. Fu, K. Anderson, H. Chu, and C. Rakovski, "A 12-lead ecg database to identify origins of idiopathic ventricular arrhythmia containing 334 patients," *Scientific Data*, vol. 7, no. 1, pp. 1–10, 2020.

[42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[43] H. M. Nguyen, E. W. Cooper, and K. Kamei, "Borderline over-sampling for imbalanced data classification," *International Journal of Knowledge Engineering and Soft Data Paradigms*, vol. 3, no. 1, pp. 4–21, 2011.

[44] H. He, Y. Bai, E. A. Garcia, and S. Li, "Adasyn: Adaptive synthetic sampling approach for imbalanced learning," *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322–1328, 2008.

[45] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," *arXiv preprint arXiv:1810.03292*, 2018.

[46] M. A. Reyna, N. Sadr, E. A. Perez Alday, A. Gu, A. Shah, C. Robichaux, B. A. Rad, A. Elola, S. Seyedi, S. Ansari, Q. Li, A. Sharma, and G. D. Clifford, "Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021," *Computing in Cardiology*, vol. 48, pp. 1–4, 2021.

[47] J. E. Hall, "Cardiac muscle; the heart as a pump and function of the heart valves," *Guyton and Hall Textbook of Medical Physiology*, pp. 101–113, 2011.

[48] L. Schamroth, *An introduction to electrocardiography*. Blackwell Scientific; Chicago, Ill.: Distributors, USA, Year Book Medical . . . , 1990.

[49] F. Morris, W. J. Brady, and A. J. Camm, *ABC of clinical electrocardiography*. John Wiley & Sons, 2009, vol. 93.

[50] W. J. Tompkins, "Biomedical digital signal processing," *Editorial Prentice Hall*, 1993.

[51] J. E. Hall, *Guyton and hall physiology review e-book*. Elsevier Health Sciences, 2015.

[52] A. L. Goldberger, Z. D. Goldberger, and A. Shvilkin, *Clinical electrocardiography: a simplified approach e-book*. Elsevier Health Sciences, 2017.

[53] M. S. Spach and J. M. Kootsey, "The nature of electrical propagation in cardiac muscle," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 244, no. 1, pp. H3–H22, 1983.

[54] G. V. Chow, J. E. Marine, and J. L. Fleg, "Epidemiology of arrhythmias and conduction disorders in older adults," *Clinics in geriatric medicine*, vol. 28, no. 4, pp. 539–553, 2012.

[55] R. McFEE and G. M. Baule, "Research in electrocardiography and magnetocardiography," *Proceedings of the IEEE*, vol. 60, no. 3, pp. 290–321, 1972.

[56] D. Kilpatrick and P. Johnston, "Origin of the electrocardiogram," *IEEE Engineering in Medicine and Biology Magazine*, vol. 13, no. 4, pp. 479–486, 1994.

[57] C. Matteucci, "Sur un phenomene physiologique produit par les muscles en contraction," *Ann Chim Phys.*, vol. 6, pp. 339–341, 1842.

[58] S. S. Barold, "Willem einthoven and the birth of clinical electrocardiography a hundred years ago," *Cardiac electrophysiology review*, vol. 7, no. 1, pp. 99–104, 2003.

[59] W. B. Fye, "A history of the origin, evolution, and impact of electrocardiography," *The American journal of cardiology*, vol. 73, no. 13, pp. 937–949, 1994.

[60] J. Burnett, "The origins of the electrocardiograph as a clinical instrument," *Medical History*, vol. 29, no. S5, pp. 53–76, 1985.

[61] W. Einthoven, "Ueber die form des menschlichen electrocardiogramms," *Pflügers Archiv European Journal of Physiology*, vol. 60, no. 3, pp. 101–123, 1895.

[62] ——, "New methods for clinical investigation]," *Ned T Geneesk*, vol. 29, no. II, pp. 263–286, 1893.

[63] ——, *Un nouveau galvanometre*. Soc. Holl. des Sciences, 1901.

[64] G. E. Burch and N. P. DePasquale, *A history of electrocardiography*. Norman Publishing, 1990, no. 1.

[65] T. Lewis, "Report cxix. auricular fibrillation: a common clinical condition," *British medical journal*, vol. 2, no. 2552, p. 1528, 1909.

[66] G. Burch, "History of precordial leads in electrocardiography." *European journal of cardiology*, vol. 8, no. 2, pp. 207–236, 1978.

[67] E. Goldberger, "A simple, indifferent, electrocardiographic electrode of zero potential and a technique of obtaining augmented, unipolar, extremity leads," *American Heart Journal*, vol. 23, no. 4, pp. 483–492, 1942.

[68] F. N. Wilson, C. E. KOSSMANN, G. E. BURCH, E. GOLDBERGER, A. GRAYBIEL, H. H. HECHT, F. D. JOHNSTON, E. LEPESCHKIN, and G. B. MYERS, "Recommendations for standardization of electrocardiographic and vectorcardiographic leads," *Circulation*, vol. 10, no. 4, pp. 564–573, 1954.

[69] M. Thaler, *The only EKG book you'll ever need*. Lippincott Williams & Wilkins, 2017.

[70] Ausmed, "12 lead ecg placement," https://www.ausmed.com/cpd/articles/ecg-lead-placement.

[71] Y. Zigel, A. Cohen, and A. Katz, "The weighted diagnostic distortion (wdd) measure for ecg signal compression," *IEEE transactions on biomedical engineering*, vol. 47, no. 11, pp. 1422–1430, 2000.

[72] L. Sun, Y. Lu, K. Yang, and S. Li, "Ecg analysis using multiple instance learning for myocardial infarction detection," *IEEE transactions on biomedical engineering*, vol. 59, no. 12, pp. 3348–3356, 2012.

[73] G. D. Clifford, F. Azuaje, P. McSharry *et al.*, *Advanced methods and tools for ECG data analysis*. Artech house Boston, 2006.

[74] G. Han, B. Lin, and Z. Xu, "Electrocardiogram signal denoising based on empirical mode decomposition technique: an overview," *Journal of Instrumentation*, vol. 12, no. 03, p. P03010, 2017.

[75] S. Chatterjee, R. S. Thakur, R. N. Yadav, L. Gupta, and D. K. Raghuvanshi, "Review of noise removal techniques in ecg signals," *IET Signal Processing*, vol. 14, no. 9, pp. 569–590, 2020.

[76] J. A. Van Alste and T. Schilder, "Removal of baseline wander and powerline interference from the ecg by an efficient fir filter with a reduced number of taps," *IEEE transactions on biomedical engineering*, no. 12, pp. 1052–1060, 1985.

[77] C. Levkov, G. Mihov, R. Ivanov, I. Daskalov, I. Christov, and I. Dotsinsky, "Removal of power-line interference from the ecg: a review of the subtraction procedure," *BioMedical Engineering OnLine*, vol. 4, no. 1, pp. 1–18, 2005.

[78] G. D. Clifford, F. Azuaje, and P. Mcsharry, "Ecg statistics, noise, artifacts, and missing data," *Advanced methods and tools for ECG data analysis*, vol. 6, no. 1, p. 18, 2006.

[79] L. Frølich and I. Dowding, "Removal of muscular artifacts in eeg signals: a comparison of linear decomposition methods," *Brain informatics*, vol. 5, no. 1, pp. 13–22, 2018.

[80] A. Velayudhan and S. Peter, "Noise analysis and different denoising techniques of ecg signal-a survey," *IOSR journal of electronics and communication engineering*, vol. 1, no. 1, pp. 40–44, 2016.

[81] S. Agrawal and A. Gupta, "Fractal and emd based removal of baseline wander and powerline interference from ecg signals," *Computers in biology and medicine*, vol. 43, no. 11, pp. 1889–1899, 2013.

[82] A. Fasano and V. Villani, "Ecg baseline wander removal and impact on beat morphology: A comparative analysis," *Computing in Cardiology Conference (CinC), 2013*, pp. 1167–1170, 2013.

[83] R. M. Rangayyan and N. P. Reddy, "Biomedical signal analysis: a case-study approach," *Annals of Biomedical Engineering*, vol. 30, no. 7, pp. 983–983, 2002.

[84] P. S. Hamilton and W. J. Tompkins, "Quantitative investigation of qrs detection rules using the mit/bih arrhythmia database," *IEEE transactions on biomedical engineering*, no. 12, pp. 1157–1165, 1986.

[85] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 3, pp. 230–236, 1985.

[86] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[87] T. Golany and K. Radinsky, "Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 557–564, 2019.

[88] T. Golany, K. Radinsky, and D. Freedman, "Simgans: Simulator-based generative adversarial networks for ecg synthesis to improve deep ecg classification," *International Conference on Machine Learning*, pp. 3597–3606, 2020.

[89] F. Zhuang, X. Cheng, P. Luo, S. J. Pan, and Q. He, "Supervised representation learning: Transfer learning with deep autoencoders," *Proc. of the 24th International Joint Conference on Artificial Intelligence*, pp. 4119–4125, 2015.

[90] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, pp. 2672–2680, 2014.

[91] D. Nankani and R. D. Baruah, "Investigating deep convolution conditional gans for electrocardiogram generation," *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2020.

[92] E. J. d. S. Luz, W. R. Schwartz, G. Cámara-Chávez, and D. Menotti, "Ecg-based heartbeat classification for arrhythmia detection: A survey," *Computer methods and programs in biomedicine*, vol. 127, pp. 144–164, 2016.

[93] M. Moavenian and H. Khorrami, "A qualitative comparison of artificial neural networks and support vector machines in ecg arrhythmias classification," *Expert Systems with Applications*, vol. 37, no. 4, pp. 3088–3093, 2010.

[94] Q. Li, C. Rajagopalan, and G. D. Clifford, "Ventricular fibrillation and tachycardia classification using a machine learning approach," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 6, pp. 1607–1613, 2014.

[95] J. Lee, B. A. Reyes, D. D. McManus, O. Maitas, and K. H. Chon, "Atrial fibrillation detection using an iphone 4s," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 1, pp. 203–206, 2012.

[96] V. Mondéjar-Guerra, J. Novo, J. Rouco, M. G. Penedo, and M. Ortega, "Heartbeat classification fusing temporal and morphological information of ecgs via ensemble of classifiers," *Biomedical Signal Processing and Control*, vol. 47, pp. 41–48, 2019.

[97] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, p. 65, 2019.

[98] U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam, "Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network," *Information sciences*, vol. 405, pp. 81–90, 2017.

[99] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan, and C. K. Chua, "Automated detection of coronary artery disease using different durations of ecg segments with convolutional neural network," *Knowledge-Based Systems*, vol. 132, pp. 62–71, 2017.

[100] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.

[101] F. Nakamura and M. Nakai, "Prediction models—why are they used or not used?—," *Circulation Journal*, pp. CJ–17, 2017.

[102] F. Cabitza, R. Rasoini, and G. F. Gensini, "Unintended consequences of machine learning in medicine," *Jama*, vol. 318, no. 6, pp. 517–518, 2017.

[103] P. Voigt and A. Von dem Bussche, "The eu general data protection regulation (gdpr)," *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 2017.

[104] B. Kim, R. Khanna, and O. O. Koyejo, "Examples are not enough, learn to criticize! criticism for interpretability," *Advances in neural information processing systems*, vol. 29, 2016.

[105] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Definitions, methods, and applications in interpretable machine learning," *Proceedings of the National Academy of Sciences*, vol. 116, no. 44, pp. 22 071–22 080, 2019.

[106] K. Simonyan, A. Vedaldi, and A. Zisserman, "Deep inside convolutional networks: Visualising image classification models and saliency maps," *arXiv preprint arXiv:1312.6034*, 2013.

[107] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," *International Conference on Machine Learning*, pp. 3319–3328, 2017.

[108] G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, pp. 1–15, 2018.

[109] Z. Wang, W. Yan, and T. Oates, "Time series classification from scratch with deep neural networks: A strong baseline," *2017 International joint conference on neural networks (IJCNN)*, pp. 1578–1585, 2017.

[110] S. A. Siddiqui, D. Mercier, M. Munir, A. Dengel, and S. Ahmed, "Tsviz: Demystification of deep learning models for time-series analysis," *IEEE Access*, vol. 7, pp. 67 027–67 040, 2019.

[111] T. Teijeiro, C. A. García, D. Castro, and P. Félix, "Abductive reasoning as a basis to reproduce expert criteria in ecg atrial fibrillation identification," *Physiological measurement*, vol. 39, no. 8, p. 084006, 2018.

[112] Z. C. Lipton, "The mythos of model interpretability," *Queue*, vol. 16, no. 3, pp. 31–57, 2018.

[113] A. Bibal and B. Frénay, "Interpretability of machine learning models and representations: an introduction," *24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 77–82, 2016, 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2016 ; Conference date: 27-04-2016 Through 29-05-2016.

[114] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (xai)," *IEEE access*, vol. 6, pp. 52 138–52 160, 2018.

[115] D. Doran, S. Schulz, and T. R. Besold, "What does explainable ai really mean? a new conceptualization of perspectives," *arXiv preprint arXiv:1710.00794*, 2017.

[116] A. Karim, A. Mishra, M. Newton, and A. Sattar, "Machine learning interpretability: A science rather than a tool," *arXiv preprint arXiv:1807.06722*, 2018.

[117] F. Poursabzi-Sangdeh, D. G. Goldstein, J. M. Hofman, J. W. Wortman Vaughan, and H. Wallach, "Manipulating and measuring model interpretability," *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–52, 2021.

[118] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An overview of interpretability of machine learning," *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pp. 80–89, 2018.

[119] R. R. Hoffman and G. Klein, "Explaining explanation, part 1: theoretical foundations," *IEEE Intelligent Systems*, vol. 32, no. 3, pp. 68–73, 2017.

[120] D. S. Weld and G. Bansal, "The challenge of crafting intelligible intelligence," *Communications of the ACM*, vol. 62, no. 6, pp. 70–79, 2019.

[121] B. Herman, "The promise and peril of human evaluation for model interpretability," *arXiv preprint arXiv:1711.07414*, p. 8, 2017.

[122] E. Ventocilla, T. Helldin, M. Riveiro, J. Bae, V. Boeva, G. Falkman, and N. Lavesson, "Towards a taxonomy for interpretable and interactive machine learning," *2nd Workshop on Explainable AI (XAI-18), 27th International Joint Conferences on Artificial Intelligence (IJCAI) July 13-19, 2018, Stockholm, Sweden*, pp. 151–157, 2018.

[123] M. T. Ribeiro, S. Singh, and C. Guestrin, "" why should i trust you?" explaining the predictions of any classifier," *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 1135–1144, 2016.

[124] A. Vellido, "Societal issues concerning the application of artificial intelligence in medicine," *Kidney Diseases*, vol. 5, no. 1, pp. 11–17, 2019.

[125] C. Molnar, *Interpretable Machine Learning*. Online, 2019.

[126] J. Krause, A. Perer, and E. Bertini, "Using visual analytics to interpret predictive machine learning models," *ArXiv*, vol. abs/1606.05685, 2016.

[127] J. Zerilli, A. Knott, J. Maclaurin, and C. Gavaghan, "Transparency in algorithmic and human decision-making: is there a double standard?" *Philosophy & Technology*, vol. 32, no. 4, pp. 661–683, 2019.

[128] D. Martens, J. Vanthienen, W. Verbeke, and B. Baesens, "Performance of classification models from a user perspective," *Decision Support Systems*, vol. 51, no. 4, pp. 782–793, 2011.

[129] G. Ras, M. van Gerven, and P. Haselager, "Explanation methods in deep learning: Users, values, concerns and challenges," *Explainable and interpretable models in computer vision and machine learning*, pp. 19–36, 2018.

[130] J. Zhu, A. Liapis, S. Risi, R. Bidarra, and G. M. Youngblood, "Explainable ai for designers: A human-centered perspective on mixed-initiative co-creation," *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, pp. 1–8, 2018.

[131] W. Pieters, "Explanation and trust: what to tell the user in security and ai?" *Ethics and information technology*, vol. 13, no. 1, pp. 53–64, 2011.

[132] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[133] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Proceedings of the 31st international conference on neural information processing systems*, pp. 4768–4777, 2017.

[134] P. P. Angelov and X. Gu, *Empirical approach to machine learning*. Springer, 2019.

[135] P. Hall and N. Gill, *An introduction to machine learning interpretability*. OŔeilly Media, Incorporated, 2019.

[136] G. Klein, "Explaining explanation, part 3: The causal landscape," *IEEE Intelligent Systems*, vol. 33, no. 2, pp. 83–88, 2018.

[137] R. Hoffman, T. Miller, S. T. Mueller, G. Klein, and W. J. Clancey, "Explaining explanation, part 4: a deep dive on deep nets," *IEEE Intelligent Systems*, vol. 33, no. 3, pp. 87–95, 2018.

[138] O. Biran and C. Cotton, "Explanation and justification in machine learning: A survey," *IJCAI-17 workshop on explainable AI (XAI)*, vol. 8, no. 1, pp. 8–13, 2017.

[139] P. Ferrando Hernández, "Lighting the black box: explaining individual predictions of machine learning algorithms," Master's thesis, Universitat Politècnica de Catalunya, 2018.

[140] M. T. Ribeiro, S. Singh, and C. Guestrin, "Model-agnostic interpretability of machine learning," *arXiv preprint arXiv:1606.05386*, 2016.

[141] ——, "Nothing else matters: model-agnostic explanations by identifying prediction invariance," *arXiv preprint arXiv:1611.05817*, 2016.

[142] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[143] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[144] R. Duan, X. He, and Z. Ouyang, "Madnn: A multi-scale attention deep neural network for arrythmia classification," *Challenge*, vol. 9, p. 10, 2020.

[145] S. Mousavi, F. Afghah, and U. R. Acharya, "Han-ecg: An interpretable atrial fibrillation detection model using hierarchical attention networks," *Computers in Biology and Medicine*, vol. 127, p. 104057, 2020.

[146] G. Yan, S. Liang, Y. Zhang, and F. Liu, "Fusing transformer model with temporal features for ecg heartbeat classification," *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 898–905, 2019.

[147] N. Strodthoff and C. Strodthoff, "Detecting and interpreting myocardial infarction using fully convolutional neural networks," *Physiological measurement*, vol. 40, no. 1, p. 015001, 2019.

[148] Q. Yao, R. Wang, X. Fan, J. Liu, and Y. Li, "Multi-class arrhythmia detection from 12-lead varied-length ecg using attention-based time-incremental convolutional neural network," *Information Fusion*, vol. 53, pp. 174–182, 2020.

[149] A. Natarajan, Y. Chang, S. Mariani, A. Rahman, G. Boverman, S. Vij, and J. Rubin, "A wide and deep transformer neural network for 12-lead ecg classification," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[150] W. Samek, A. Binder, G. Montavon, S. Lapuschkin, and K.-R. Müller, "Evaluating the visualization of what a deep neural network has learned," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 11, pp. 2660–2673, 2016.

[151] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek, "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation," *PloS one*, vol. 10, no. 7, p. e0130140, 2015.

[152] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 3145–3153, 2017.

[153] A. L. Beam and I. S. Kohane, "Big data and machine learning in health care," *Jama*, vol. 319, no. 13, pp. 1317–1318, 2018.

[154] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health information science and systems*, vol. 2, no. 1, pp. 1–10, 2014.

[155] J. L. Guidi, K. Clark, M. T. Upton, H. Faust, C. A. Umscheid, M. B. Lane-Fall, M. E. Mikkelsen, W. D. Schweickert, C. A. Vanzandbergen, J. Betesh *et al.*, "Clinician perception of the effectiveness of an automated early warning and response system for sepsis in an academic medical center," *Annals of the American Thoracic Society*, vol. 12, no. 10, pp. 1514–1519, 2015.

[156] G. Stiglic, P. Kocbek, N. Fijacko, M. Zitnik, K. Verbert, and L. Cilar, "Interpretability of machine learning-based prediction models in healthcare," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 10, no. 5, p. e1379, 2020.

[157] G. Quer, R. Arnaout, M. Henne, and R. Arnaout, "Machine learning and the future of cardiovascular care: Jacc state-of-the-art review," *Journal of the American College of Cardiology*, vol. 77, no. 3, pp. 300–313, 2021.

[158] C. Krittanawong, K. W. Johnson, R. S. Rosenson, Z. Wang, M. Aydar, U. Baber, J. K. Min, W. W. Tang, J. L. Halperin, and S. M. Narayan, "Deep learning for cardiovascular medicine: a practical primer," *European heart journal*, vol. 40, no. 25, pp. 2058–2073, 2019.

[159] P. A. Iannattone, X. Zhao, J. VanHouten, A. Garg, and T. Huynh, "Artificial intelligence for diagnosis of acute coronary syndromes: a meta-analysis of machine learning approaches," *Canadian Journal of Cardiology*, vol. 36, no. 4, pp. 577–583, 2020.

[160] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of clinical epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996.

[161] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.

[162] J. Petch, S. Di, and W. Nelson, "Opening the black box: the promise and limitations of explainable machine learning in cardiology," *Canadian Journal of Cardiology*, 2021.

[163] G. Shmueli, "To explain or to predict?" *Statistical science*, vol. 25, no. 3, pp. 289–310, 2010.

[164] R. R. Hoffman, S. T. Mueller, and G. Klein, "Explaining explanation, part 2: Empirical foundations," *IEEE Intelligent Systems*, vol. 32, no. 4, pp. 78–86, 2017.

[165] C. A. Umscheid, J. Betesh, C. VanZandbergen, A. Hanish, G. Tait, M. E. Mikkelsen, B. French, and B. D. Fuchs, "Development, implementation, and impact of an automated early warning and response system for sepsis," *Journal of hospital medicine*, vol. 10, no. 1, pp. 26–31, 2015.

[166] G. J. Katuwal and R. Chen, "Machine learning model interpretability for precision medicine," *arXiv preprint arXiv:1610.09045*, 2016.

[167] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong, "Interpretable machine learning: Fundamental principles and 10 grand challenges," *arXiv preprint arXiv:2103.11251*, 2021.

[168] O. Lahav, N. Mastronarde, and M. van der Schaar, "What is interpretable? using machine learning to design interpretable decision-support systems," *arXiv preprint arXiv:1811.10799*, 2018.

[169] S. Tonekaboni, S. Joshi, M. D. McCradden, and A. Goldenberg, "What clinicians want: contextualizing explainable machine learning for clinical end use," *Machine learning for healthcare conference*, pp. 359–380, 2019.

[170] J. Krause, A. Perer, and K. Ng, "Interacting with predictions: Visual inspection of black-box machine learning models," *Proceedings of the 2016 CHI conference on human factors in computing systems*, pp. 5686–5697, 2016.

[171] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature biomedical engineering*, vol. 2, no. 10, pp. 749–760, 2018.

[172] S.-C. Pei and C.-C. Tseng, "Elimination of ac interference in electrocardiogram using iir notch filter with transient suppression," *IEEE transactions on biomedical engineering*, vol. 42, no. 11, pp. 1128–1132, 1995.

[173] G. S. Furno and W. J. Tompkins, "A learning filter for removing noise interference," *IEEE Transactions on biomedical Engineering*, no. 4, pp. 234–235, 1983.

[174] M. Yelderman, B. Widrow, J. M. Cioffi, E. Hesler, and J. A. Leddy, "Ecg enhancement by adaptive cancellation of electrosurgical interference," *IEEE Transactions on Biomedical Engineering*, no. 7, pp. 392–398, 1983.

[175] V. Barbero Romero, "Ecg baseline wander removal and noise suppression analysis in an embedded platform," *Ene*, vol. 16, p. 38, 2010.

[176] C. Papaloukas, D. Fotiadis, A. Liavas, A. Likas, and L. Michalis, "A knowledge-based technique for automated detection of ischaemic episodes in long duration electrocardiograms," *Medical and Biological Engineering and Computing*, vol. 39, no. 1, pp. 105–112, 2001.

[177] J. M. Leski and N. Henzel, "Ecg baseline wander and powerline interference reduction using nonlinear filter bank," *Signal processing*, vol. 85, no. 4, pp. 781–793, 2005.

[178] V. Chouhan and S. S. Mehta, "Total removal of baseline drift from ecg signal," *Computing: Theory and Applications, 2007. ICCTA'07. International Conference on*, pp. 512–515, 2007.

[179] P. J. Huber, "John w. tukey's contributions to robust statistics," *Annals of statistics*, pp. 1640–1648, 2002.

[180] N. V. Thakor and Y.-S. Zhu, "Applications of adaptive filtering to ecg analysis: noise cancellation and arrhythmia detection," *IEEE transactions on biomedical engineering*, vol. 38, no. 8, pp. 785–794, 1991.

[181] K. Park, K. Lee, and H. Yoon, "Application of a wavelet adaptive filter to minimise distortion of the st-segment," *Medical and Biological Engineering and Computing*, vol. 36, no. 5, pp. 581–586, 1998.

[182] E. Ercelebi, "Electrocardiogram signals de-noising using lifting-based discrete wavelet transform," *Computers in Biology and Medicine*, vol. 34, no. 6, pp. 479–493, 2004.

[183] T. He, G. Clifford, and L. Tarassenko, "Application of independent component analysis in removing artefacts from the electrocardiogram," *Neural Computing & Applications*, vol. 15, no. 2, pp. 105–116, 2006.

[184] S. Iravanian and L. Tung, "A novel algorithm for cardiac biosignal filtering based on filtered residue method," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 11, pp. 1310–1317, 2002.

[185] A. K. Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from electrocardiographic signals using independent components analysis," *Neurocomputing*, vol. 22, no. 1-3, pp. 173–186, 1998.

[186] M. Blanco-Velasco, B. Weng, and K. E. Barner, "Ecg signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in biology and medicine*, vol. 38, no. 1, pp. 1–13, 2008.

[187] T. Ji, Z. Lu, Q. Wu, and Z. Ji, "Baseline normalisation of ecg signals using empirical mode decomposition and mathematical morphology," *Electronics letters*, vol. 44, no. 2, p. 1, 2008.

[188] Z. Zhao and J. Liu, "Baseline wander removal of ecg signals using empirical mode decomposition and adaptive filter," *Bioinformatics and Biomedical Engineering (iCBBE), 2010 4th International Conference on*, pp. 1–3, 2010.

[189] M. A. Kabir and C. Shahnaz, "Denoising of ecg signals based on noise reduction algorithms in emd and wavelet domains," *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 481–489, 2012.

[190] K.-M. Chang, "Arrhythmia ecg noise reduction by ensemble empirical mode decomposition," *Sensors*, vol. 10, no. 6, pp. 6063–6080, 2010.

[191] Y. Xu, M. Luo, T. Li, and G. Song, "Ecg signal de-noising and baseline wander correction based on ceemdan and wavelet threshold," *Sensors*, vol. 17, no. 12, p. 2754, 2017.

[192] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*, pp. 903–995, 1998.

[193] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE signal processing letters*, vol. 11, no. 2, pp. 112–114, 2004.

[194] Z. Wu and N. E. Huang, "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 1–41, 2009.

[195] M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, "A complete ensemble empirical mode decomposition with adaptive noise," *Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on*, pp. 4144–4147, 2011.

[196] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE transactions on signal processing*, vol. 62, no. 3, pp. 531–544, 2014.

[197] X. Nie and R. Unbehauen, "Edge preserving filtering by combining nonlinear mean and median filters," *IEEE Transactions on signal processing*, vol. 39, no. 11, pp. 2552–2554, 1991.

[198] Y. Xin, Y. Chen, and W. T. Hao, "Ecg baseline wander correction based on mean-median filter and empirical mode decomposition," *Bio-medical materials and engineering*, vol. 24, no. 1, pp. 365–371, 2014.

[199] C. Li, C. Zheng, and C. Tai, "Detection of ecg characteristic points using wavelet transforms," *IEEE Transactions on biomedical Engineering*, vol. 42, no. 1, pp. 21–28, 1995.

[200] X. Xu and Y. Liu, "Ecg qrs complex detection using slope vector waveform (svw) algorithm," *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2, pp. 3597–3600, 2004.

[201] K. V. Suarez, J. C. Silva, Y. Berthoumieu, P. Gomis, and M. Najim, "Ecg beat detection using a geometrical matching approach," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 4, pp. 641–650, 2007.

[202] Y.-C. Yeh and W.-J. Wang, "Qrs complexes detection for ecg signal: The difference operation method," *Computer methods and programs in biomedicine*, vol. 91, no. 3, pp. 245–254, 2008.

[203] S. Pal and M. Mitra, "Empirical mode decomposition based ecg enhancement and qrs detection," *Computers in biology and medicine*, vol. 42, no. 1, pp. 83–92, 2012.

[204] J. Kim and H. Shin, "Simple and robust realtime qrs detection algorithm based on spatiotemporal characteristic of the qrs complex," *PloS one*, vol. 11, no. 3, p. e0150144, 2016.

[205] R. He, K. Wang, Q. Li, Y. Yuan, N. Zhao, Y. Liu, and H. Zhang, "A novel method for the detection of r-peaks in ecg based on k-nearest neighbors and particle swarm optimization," *EURASIP Journal on Advances in Signal Processing*, vol. 2017, no. 1, p. 82, 2017.

[206] J. Malik, E. Z. Soliman, and H.-T. Wu, "An adaptive qrs detection algorithm for ultra-long-term ecg recordings," *Journal of Electrocardiology*, 2020.

[207] A. Chen, Y. Zhang, M. Zhang, W. Liu, S. Chang, H. Wang, J. He, and Q. Huang, "A real time qrs detection algorithm based on et and pd controlled threshold strategy," *Sensors*, vol. 20, no. 14, p. 4003, 2020.

[208] B. B. Mandelbrot, *The fractal geometry of nature*. WH freeman New York, 1983, vol. 173.

[209] P. Maragos, "Fractal signal analysis using mathematical morphology," *Advances in electronics and electron physics*, vol. 88, pp. 199–246, 1994.

[210] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, "A dynamical model for generating synthetic electrocardiogram signals," *IEEE transactions on biomedical engineering*, vol. 50, no. 3, pp. 289–294, 2003.

[211] G. D. Clifford, S. Nemati, and R. Sameni, "An artificial vector model for generating abnormal electrocardiographic rhythms," *Physiological measurement*, vol. 31, no. 5, p. 595, 2010.

[212] E. Brophy, Z. Wang, and T. E. Ward, "Quick and easy time series generation with established image-based gans," *arXiv preprint arXiv:1902.05624*, 2019.

[213] S. Haradal, H. Hayashi, and S. Uchida, "Biosignal data augmentation based on generative adversarial networks," *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 368–371, 2018.

[214] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," *International conference on machine learning*, pp. 214–223, 2017.

[215] F. Zhu, F. Ye, Y. Fu, Q. Liu, and B. Shen, "Electrocardiogram generation with a bidirectional lstm-cnn generative adversarial network," *Scientific reports*, vol. 9, no. 1, p. 6734, 2019.

[216] B. Zhou, S. Liu, B. Hooi, X. Cheng, and J. Ye, "Beatgan: anomalous rhythm detection using adversarially generated time series," *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pp. 4433–4439, 2019.

[217] A. M. Delaney, E. Brophy, and T. E. Ward, "Synthesis of realistic ecg using generative adversarial networks," *arXiv preprint arXiv:1909.09150*, 2019.

[218] D. Hazra and Y.-C. Byun, "Synsiggan: Generative adversarial networks for synthetic biomedical signal generation," *Biology*, vol. 9, no. 12, p. 441, 2020.

[219] P. Wang, B. Hou, S. Shao, and R. Yan, "Ecg arrhythmias detection using auxiliary classifier generative adversarial network and residual network," *IEEE Access*, vol. 7, pp. 100 910–100 922, 2019.

[220] F. Ye, F. Zhu, Y. Fu, and B. Shen, "Ecg generation with sequence generative adversarial nets optimized by policy gradient," *IEEE Access*, vol. 7, pp. 159 369–159 378, 2019.

[221] N. Wulan, W. Wang, P. Sun, K. Wang, Y. Xia, and H. Zhang, "Generating electrocardiogram signals by deep learning," *Neurocomputing*, vol. 404, pp. 122–136, 2020.

[222] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[223] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[224] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2794–2802, 2017.

[225] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[226] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[227] Z.-D. Zhao and Y.-Q. Chen, "A new method for removal of baseline wander and power line interference in ecg signals," *Machine Learning and Cybernetics, 2006 International Conference on*, pp. 4342–4347, 2006.

[228] D. Nankani and R. D. Baruah, "Effective removal of baseline wander from ecg signals: A comparative study," *International Conference on Machine Learning, Image Processing, Network Security and Data Sciences*, pp. 310–324, 2020.

[229] P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.

[230] H. Wang, H. Shi, K. Lin, C. Qin, L. Zhao, Y. Huang, and C. Liu, "A high-precision arrhythmia classification method based on dual fully connected neural network," *Biomedical Signal Processing and Control*, vol. 58, p. 101874, 2020.

[231] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," *Advances in neural information processing systems*, pp. 6626–6637, 2017.

[232] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A kernel method for the two-sample-problem," *Advances in neural information processing systems*, pp. 513–520, 2007.

[233] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE transactions on acoustics, speech, and signal processing*, vol. 26, no. 1, pp. 43–49, 1978.

[234] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intelligent Data Analysis*, vol. 11, no. 5, pp. 561–580, 2007.

[235] P.-F. Marteau, "Time warp edit distance with stiffness adjustment for time series matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 306–318, 2008.

[236] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.

[237] G. De Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 1, pp. 241–247, 2011.

[238] S. Sahoo, B. Kanungo, S. Behera, and S. Sabut, "Multiresolution wavelet transform based feature extraction and ecg classification to detect cardiac abnormalities," *Measurement*, vol. 108, pp. 55–66, 2017.

[239] Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification," *Computers in biology and medicine*, vol. 96, pp. 189–202, 2018.

[240] S. Banerjee and M. Mitra, "Ecg beat classification based on discrete wavelet transformation and nearest neighbour classifier," *Journal of medical engineering & technology*, vol. 37, no. 4, pp. 264–272, 2013.

[241] L.-Y. Shyu, Y.-H. Wu, and W. Hu, "Using wavelet transform and fuzzy neural network for vpc detection from the holter ecg," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 7, pp. 1269–1273, 2004.

[242] A. Al-Fahoum and I. Howitt, "Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias," *Medical & biological engineering & computing*, vol. 37, no. 5, pp. 566–573, 1999.

[243] R. J. Martis, U. R. Acharya, and L. C. Min, "Ecg beat classification using pca, lda, ica and discrete wavelet transform," *Biomedical Signal Processing and Control*, vol. 8, no. 5, pp. 437–448, 2013.

[244] L. El Bouny, M. Khalil, and A. Adib, "Ecg heartbeat classification based on multi-scale wavelet convolutional neural networks," *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3212–3216, 2020.

[245] R. J. Martis, U. R. Acharya, C. M. Lim, and J. S. Suri, "Characterization of ecg beats from cardiac arrhythmia using discrete cosine transform in pca framework," *Knowledge-Based Systems*, vol. 45, pp. 76–82, 2013.

[246] D. Benitez, P. Gaydecki, A. Zaidi, and A. Fitzpatrick, "The use of the hilbert transform in ecg signal analysis," *Computers in biology and medicine*, vol. 31, no. 5, pp. 399–406, 2001.

[247] N. Zeng, Z. Wang, and H. Zhang, "Inferring nonlinear lateral flow immunoassay state-space models via an unscented kalman filter," *Science China Information Sciences*, vol. 59, no. 11, p. 112204, 2016.

[248] M. K. Das and S. Ari, "Patient-specific ecg beat classification technique," *Healthcare technology letters*, vol. 1, no. 3, pp. 98–103, 2014.

[249] T. Mar, S. Zaunseder, J. P. Martínez, M. Llamedo, and R. Poll, "Optimization of ecg classification by means of feature selection," *IEEE transactions on Biomedical Engineering*, vol. 58, no. 8, pp. 2168–2177, 2011.

[250] M. Lagerholm, C. Peterson, G. Braccini, L. Edenbrandt, and L. Sornmo, "Clustering ecg complexes using hermite functions and self-organizing maps," *IEEE Transactions on Biomedical Engineering*, vol. 47, no. 7, pp. 838–848, 2000.

[251] M. K. Das and S. Ari, "Ecg beats classification using mixture of features," *International scholarly research notices*, vol. 2014, 2014.

[252] G. Garcia, G. Moreira, D. Menotti, and E. Luz, "Inter-patient ecg heartbeat classification with temporal vcg optimized by pso," *Scientific Reports*, vol. 7, no. 1, pp. 1–11, 2017.

[253] F. Melgani and Y. Bazi, "Classification of electrocardiogram signals with support vector machines and particle swarm optimization," *IEEE transactions on information technology in biomedicine*, vol. 12, no. 5, pp. 667–677, 2008.

[254] H. Li, D. Yuan, X. Ma, D. Cui, and L. Cao, "Genetic algorithm for the optimization of features and neural networks in ecg signals classification," *Scientific reports*, vol. 7, p. 41011, 2017.

[255] N. Acır, "Classification of ecg beats by using a fast least square support vector machines with a dynamic programming feature selection algorithm," *Neural computing & applications*, vol. 14, no. 4, pp. 299–309, 2005.

[256] M. Javadi, S. A. A. A. Arani, A. Sajedin, and R. Ebrahimpour, "Classification of ecg arrhythmia by a modular neural network based on mixture of experts and negatively correlated learning," *Biomedical Signal Processing and Control*, vol. 8, no. 3, pp. 289–296, 2013.

[257] Y. H. Hu, S. Palreddy, and W. J. Tompkins, "A patient-adaptable ecg beat classifier using a mixture of experts approach," *IEEE transactions on biomedical engineering*, vol. 44, no. 9, pp. 891–900, 1997.

[258] Y.-C. Yeh, C. W. Chiou, and H.-J. Lin, "Analyzing ecg for cardiac arrhythmia using cluster analysis," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1000–1010, 2012.

[259] B. Zhu, Y. Ding, and K. Hao, "A novel automatic detection system for ecg arrhythmias using maximum margin clustering with immune evolutionary algorithm," *Computational and mathematical methods in medicine*, vol. 2013, 2013.

[260] H. He and Y. Tan, "Automatic pattern recognition of ecg signals using entropy-based adaptive dimensionality reduction and clustering," *Applied Soft Computing*, vol. 55, pp. 238–252, 2017.

[261] S. L. Oh, E. Y. Ng, R. San Tan, and U. R. Acharya, "Automated diagnosis of arrhythmia using combination of cnn and lstm techniques with variable length heart beats," *Computers in biology and medicine*, vol. 102, pp. 278–287, 2018.

[262] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2015.

[263] D. Nankani and R. D. Baruah, "An end-to-end framework for automatic detection of atrial fibrillation using deep residual learning," *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*, pp. 690–695, 2019.

[264] D. Nankani, P. Saikia, and R. D. Baruah, "Automatic concurrent arrhythmia classification using deep residual neural networks," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[265] J. Li, Y. Si, T. Xu, and S. Jiang, "Deep convolutional neural network based ecg classification system using information fusion and one-hot encoding techniques," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[266] T. Ölmez and Z. Dokur, "Application of inp neural network to ecg beatclassification," *Neural Computing & Applications*, vol. 11, no. 3, pp. 144–155, 2003.

[267] N. Zeng, Z. Wang, H. Zhang, W. Liu, and F. E. Alsaadi, "Deep belief networks for quantitative analysis of a gold immunochromatographic strip," *Cognitive Computation*, vol. 8, no. 4, pp. 684–692, 2016.

[268] P. Melin, J. Amezcua, F. Valdez, and O. Castillo, "A new neural network model based on the lvq algorithm for multi-class classification of arrhythmias," *Information sciences*, vol. 279, pp. 483–497, 2014.

[269] M. Korürek and B. Doğan, "Ecg beat classification using particle swarm optimization and radial basis function neural network," *Expert systems with Applications*, vol. 37, no. 12, pp. 7563–7569, 2010.

[270] Y. Ming, P. Xu, H. Qu, and L. Ren, "Interpretable and steerable sequence learning via prototypes," *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 903–913, 2019.

[271] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[272] N. Kalchbrenner, E. Grefenstette, and P. Blunsom, "A convolutional neural network for modelling sentences," *arXiv preprint arXiv:1404.2188*, 2014.

[273] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[274] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[275] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," *International conference on artificial neural networks*, pp. 92–101, 2010.

[276] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[277] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[278] S. Meek and F. Morris, "Introduction. ii—basic terminology," *BMJ*, vol. 324, no. 7335, pp. 470–473, 2002.

[279] P. Pławiak and U. R. Acharya, "Novel deep genetic ensemble of classifiers for arrhythmia detection using ecg signals," *Neural Computing and Applications*, vol. 32, no. 15, pp. 11 137–11 161, 2020.

[280] P. Sharma, S. K. Dinkar, and D. Gupta, "A novel hybrid deep learning method with cuckoo search algorithm for classification of arrhythmia disease using ecg signals," *Neural Computing and Applications*, pp. 1–21, 2021.

[281] M. Alfaras, M. C. Soriano, and S. Ortín, "A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection," *Frontiers in Physics*, vol. 7, p. 103, 2019.

[282] A. Vora, A. Naik, Y. Lokhandwala, A. Chopra, J. Varma, G. Wander, A. Jaswal, V. Srikanthan, B. Singh, D. Kahali, A. Gupta, R. Mantri, A. Mishra, U. Pandurangi, D. Ghosh, J. S. Makkar, S. Sahu, and R. Radhakrishnan, "Profiling cardiac arrhythmia and heart failure patients in india: The pan-arrhythmia and heart failure observational study," *Indian Heart Journal*, vol. 69, no. 2, pp. 226–239, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0019483216301626

[283] M. A. Othman, N. M. Safri, I. A. Ghani, F. K. C. Harun, and I. Ariffin, "A new semantic mining approach for detecting ventricular tachycardia and ventricular fibrillation," *Biomedical Signal Processing and Control*, vol. 8, no. 2, pp. 222–227, 2013.

[284] G. T. Taye, E. B. Shim, H.-J. Hwang, and K. M. Lim, "Machine learning approach to predict ventricular fibrillation based on qrs complex shape," *Frontiers in physiology*, vol. 10, p. 1193, 2019.

[285] C. Figuera, U. Irusta, E. Morgado, E. Aramendi, U. Ayala, L. Wik, J. Kramer-Johansen, T. Eftestøl, and F. Alonso-Atienza, "Machine learning techniques for the detection of shockable rhythms in automated external defibrillators," *PloS one*, vol. 11, no. 7, 2016.

[286] A. Picon, U. Irusta, A. Álvarez-Gila, E. Aramendi, F. Alonso-Atienza, C. Figuera, U. Ayala, E. Garrote, L. Wik *et al.*, "Mixed convolutional and long short-term memory network for the detection of lethal ventricular arrhythmia," *PloS one*, vol. 14, no. 5, 2019.

[287] M. Mohanty, P. Biswal, and S. Sabut, "Ventricular tachycardia and fibrillation detection using dwt and decision tree classifier," *Journal of Mechanics in Medicine and Biology*, p. 1950008, 2018.

[288] S. Mandala, T. Cai Di, and M. S. Sunar, "Ecg-based prediction algorithm for imminent malignant ventricular arrhythmias using decision tree," *Plos one*, vol. 15, no. 5, p. e0231635, 2020.

[289] W. W. Heng, E. S. L. Ming, A. N. B. Jamaluddin, F. K. C. Harun, N. A. Abdul-Kadir, and C. F. Yeong, "Prediction algorithm of malignant ventricular arrhythmia validated across multiple online public databases," *2019 Computing in Cardiology (CinC)*, pp. 1–4, 2019.

[290] G. S. d. Santos, "Towards short-term forecasting of ventricular tachyarrhythmias," Ph.D. dissertation, Massachusetts Institute of Technology, 2006.

[291] E. Prabhakararao and M. S. Manikandan, "Efficient and robust ventricular tachycardia and fibrillation detection method for wearable cardiac health monitoring devices," *Healthcare technology letters*, vol. 3, no. 3, pp. 239–246, 2016.

[292] M. T. Nguyen, B. Van Nguyen, and K. Kim, "Deep feature learning for sudden cardiac arrest detection in automated external defibrillators," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[293] Y. Xu, D. Wang, W. Zhang, P. Ping, and L. Feng, "Detection of ventricular tachycardia and fibrillation using adaptive variational mode decomposition and boosted-cart classifier," *Biomedical Signal Processing and Control*, vol. 39, pp. 219–229, 2018.

[294] R. K. Tripathy, A. Z. Mendez, S. de la O, M. R. Arrieta Paternina, J. G. Arrieta, G. R. Naik *et al.*, "Detection of life threatening ventricular arrhythmia using digital taylor fourier transform," *Frontiers in physiology*, vol. 9, p. 722, 2018.

[295] F. Alonso-Atienza, E. Morgado, L. Fernandez-Martinez, A. García-Alberola, and J. L. Rojo-Alvarez, "Detection of life-threatening arrhythmias using feature selection and support vector machines," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 832–840, 2014.

[296] H. Fujita and D. Cimr, "Decision support system for arrhythmia prediction using convolutional neural network structure without preprocessing," *Applied Intelligence*, vol. 49, no. 9, pp. 3383–3391, 2019.

[297] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, and R. San Tan, "Deep convolutional neural network for the automated diagnosis of

congestive heart failure using ecg signals," *Applied Intelligence*, vol. 49, no. 1, pp. 16–27, 2019.

[298] A. Amann, R. Tratnig, and K. Unterkofler, "Reliability of old and new ventricular fibrillation detection algorithms for automated external defibrillators," *Biomedical engineering online*, vol. 4, no. 1, p. 60, 2005.

[299] F. Alonso-Atienza, J. L. Rojo-Álvarez, A. Rosado-Muñoz, J. J. Vinagre, A. García-Alberola, and G. Camps-Valls, "Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection," *Expert Systems with Applications*, vol. 39, no. 2, pp. 1956–1967, 2012.

[300] A. Amann, R. Tratnig, and K. Unterkofler, "Detecting ventricular fibrillation by time-delay methods," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 1, pp. 174–177, 2006.

[301] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks (2013)," *arXiv preprint arXiv:1311.2901*, 2013.

[302] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2921–2929, 2016.

[303] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[304] R. Tripathy, L. Sharma, and S. Dandapat, "Detection of shockable ventricular arrhythmia using variational mode decomposition," *Journal of medical systems*, vol. 40, no. 4, p. 79, 2016.

[305] F. H. Samie and J. Jalife, "Mechanisms underlying ventricular tachycardia and its transition to ventricular fibrillation in the structurally normal heart," *Cardiovascular research*, vol. 50, no. 2, pp. 242–250, 2001.

[306] M. Zoni-Berisso, F. Lercari, T. Carazza, and S. Domenicucci, "Epidemiology of atrial fibrillation: European perspective," *Clinical epidemiology*, vol. 6, p. 213, 2014.

[307] Y. Hagiwara, H. Fujita, S. L. Oh, J. H. Tan, R. San Tan, E. J. Ciaccio, and U. R. Acharya, "Computer-aided diagnosis of atrial fibrillation based on ecg signals: A review," *Information Sciences*, vol. 467, pp. 99–114, 2018.

[308] G. Y. H. Lip, L. Fauchier, S. B. Freedman, I. Van Gelder, A. Natale, C. Gianni, S. Nattel, T. Potpara, M. Rienstra, H.-F. Tse, and D. A. Lane, "Atrial fibrillation," *Nature Reviews Disease Primers*, vol. 2, 2016.

[309] D. T. Linker, "Accurate, automated detection of atrial fibrillation in ambulatory recordings," *Cardiovascular engineering and technology*, vol. 7, no. 2, pp. 182–189, 2016.

[310] F. Censi, I. Corazza, E. Reggiani, G. Calcagnini, E. Mattei, M. Triventi, and G. Boriani, "P-wave variability and atrial fibrillation," *Scientific reports*, vol. 6, p. 26799, 2016.

[311] S. Dash, K. Chon, S. Lu, and E. Raeder, "Automatic real time detection of atrial fibrillation," *Annals of biomedical engineering*, vol. 37, no. 9, pp. 1701–1709, 2009.

[312] N. Larburu, T. Lopetegi, and I. Romero, "Comparative study of algorithms for atrial fibrillation detection," *2011 Computing in Cardiology*, pp. 265–268, 2011.

[313] L. Billeci, F. Chiarugi, M. Costi, D. Lombardi, and M. Varanini, "Detection of af and other rhythms using rr variability and ecg spectral measures," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[314] I. Christov, V. Krasteva, I. Simova, T. Neycheva, and R. Schmid, "Multi-parametric analysis for atrial fibrillation classification in ecg," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[315] M. Da Silva-Filarder and F. Marzbanrad, "Combining template-based and feature-based classification to detect atrial fibrillation from a short single lead ecg recording," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[316] S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A. D. Choudhury, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, and S. Khandelwal, "Identifying normal, af and other abnormal ecg rhythms using a cascaded binary classifier," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[317] M. García, J. Ródenas, R. Alcaraz, and J. J. Rieta, "Atrial fibrillation screening through combined timing features of short single-lead electrocardiograms," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[318] A. Petrėnas, L. Sörnmo, A. Lukoševičius, and V. Marozas, "Detection of occult paroxysmal atrial fibrillation," *Medical & biological engineering & computing*, vol. 53, no. 4, pp. 287–297, 2015.

[319] J. A. Behar, A. A. Rosenberg, Y. Yaniv, and J. Oster, "Rhythm and quality classification from short ecgs recorded using a mobile device," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[320] M. Bettoni and M. Zimmermann, "Autonomic tone variations before the onset of paroxysmal atrial fibrillation," *Circulation*, vol. 105, no. 23, pp. 2753–2759, 2002.

[321] B. Chandra, C. S. Sastry, S. Jana, and S. Patidar, "Atrial fibrillation detection using convolutional neural networks," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[322] L. Tran, Y. Li, L. Nocera, C. Shahabi, and L. Xiong, "Multifusionnet: atrial fibrillation detection with deep neural networks," *AMIA Summits on Translational Science Proceedings*, vol. 2020, p. 654, 2020.

[323] S. Hong, M. Wu, Y. Zhou, Q. Wang, J. Shang, H. Li, and J. Xie, "Encase: An ensemble classifier for ecg classification using expert features and deep neural networks," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[324] M. Zihlmann, D. Perekrestenko, and M. Tschannen, "Convolutional recurrent neural networks for electrocardiogram classification," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[325] F. Murat, F. Sadak, O. Yildirim, M. Talo, E. Murat, M. Karabatak, Y. Demir, R.-S. Tan, and U. R. Acharya, "Review of deep learning-based atrial fibrillation detection studies," *International journal of environmental research and public health*, vol. 18, no. 21, p. 11302, 2021.

[326] R. Rouhi, M. Clausel, J. Oster, and F. Lauer, "An interpretable hand-crafted feature-based model for atrial fibrillation detection," *Frontiers in Physiology*, vol. 12, p. 581, 2021.

[327] S. Vijayarangan, B. Murugesan, R. Vignesh, S. Preejith, J. Joseph, and M. Sivaprakasam, "Interpreting deep neural networks for single-lead ecg arrhythmia classification," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 300–303, 2020.

[328] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412–1421, Sep. 2015. [Online]. Available: https://aclanthology.org/D15-1166

[329] F. Chollet *et al.*, "Keras: Deep learning library for theano and tensorflow," *GitHub*, 2015. [Online]. Available: https://github.com/fchollet/keras

[330] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256, 2010.

[331] J. Rubin, S. Parvaneh, A. Rahman, B. Conroy, and S. Babaeizadeh, "Densely connected convolutional networks for detection of atrial fibrillation from short single-lead ecg recordings," *Journal of electrocardiology*, vol. 51, no. 6, pp. S18–S21, 2018.

[332] X. Fan, Z. Hu, R. Wang, L. Yin, Y. Li, and Y. Cai, "A novel hybrid network of fusing rhythmic and morphological features for atrial fibrillation detection on mobile ecg signals," *Neural Computing and Applications*, vol. 32, no. 12, pp. 8101–8113, 2020.

[333] Z. Zhao, S. Särkkä, and A. B. Rad, "Kalman-based spectro-temporal ecg analysis using deep convolutional networks for atrial fibrillation detection," *Journal of Signal Processing Systems*, vol. 92, no. 7, pp. 621–636, 2020.

[334] P. Cao, X. Li, K. Mao, F. Lu, G. Ning, L. Fang, and Q. Pan, "A novel data augmentation method to enhance deep neural networks for detection of atrial fibrillation," *Biomedical Signal Processing and Control*, vol. 56, p. 101675, 2020.

[335] Q. H. Nguyen, B. P. Nguyen, T. B. Nguyen, T. T. Do, J. F. Mbinta, and C. R. Simpson, "Stacking segment-based cnn with svm for recognition of atrial fibrillation from single-lead ecg recordings," *Biomedical Signal Processing and Control*, vol. 68, p. 102672, 2021.

[336] F. Andreotti, O. Carr, M. A. Pimentel, A. Mahdi, and M. De Vos, "Comparing feature-based classifiers and convolutional neural networks to detect arrhythmia from short segments of ecg," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[337] G. Bin, M. Shao, G. Bin, J. Huang, D. Zheng, and S. Wu, "Detection of atrial fibrillation using decision tree ensemble," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[338] S. Ghiasi, M. Abdollahpur, N. Madani, K. Kiani, and A. Ghaffari, "Atrial fibrillation detection using feature based algorithm and deep convolutional neural network," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[339] S. Raj and K. C. Ray, "Sparse representation of ecg signals for automated recognition of cardiac arrhythmias," *Expert systems with applications*, vol. 105, pp. 49–64, 2018.

[340] W. Yang, Y. Si, D. Wang, and B. Guo, "Automatic recognition of arrhythmia based on principal component analysis network and linear support vector machine," *Computers in biology and medicine*, vol. 101, pp. 22–32, 2018.

[341] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Computer Methods and Programs in Biomedicine*, vol. 161, pp. 1–13, 2018.

[342] P. Warrick and M. N. Homsi, "Cardiac arrhythmia detection from ecg combining convolutional and long short-term memory networks," *2017 Computing in Cardiology (CinC)*, pp. 1–4, 2017.

[343] R. Bousseljot, D. Kreiseler, and A. Schnabel, "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet," *Biomedizinische Technik*, vol. 40, no. S1, pp. 317–318, 1995.

[344] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He, J. Li, and E. N. Y. Kwee, "An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection," *Journal of Medical Imaging and Health Informatics*, vol. 8, no. 7, pp. 1368—1373, 2018.

[345] N. Ammour, "Atrial fibrillation detection with a domain adaptation neural network approach," *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 738–743, 2018.

[346] Y. Jin, C. Qin, J. Liu, K. Lin, H. Shi, Y. Huang, and C. Liu, "A novel domain adaptive residual network for automatic atrial fibrillation detection," *Knowledge-Based Systems*, vol. 203, p. 106122, 2020.

[347] M. Chen, G. Wang, Z. Ding, J. Li, and H. Yang, "Unsupervised domain adaptation for ecg arrhythmia classification," *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 304–307, 2020.

[348] W. Jia, X. Xu, X. Xu, Y. Sun, and X. Liu, "Automatic detection and classification of 12-lead ecgs using a deep neural network," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[349] N. Nonaka and J. Seita, "Electrocardiogram classification by modified efficientnet with data augmentation," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[350] J. Chen, T. Chen, B. Xiao, X. Bi, Y. Wang, H. Duan, W. Li, J. Zhang, and X. Ma, "Se-ecgnet: Multi-scale se-net for multi-lead ecg data," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[351] Y. Feng and E. Vigmond, "Deep multi-label multi-instance classification on 12-lead ecg," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[352] Z. Zhao, H. Fang, S. D. Relton, R. Yan, Y. Liu, Z. Li, J. Qin, and D. C. Wong, "Adaptive lead weighted resnet trained with different duration signals for classifying 12-lead ecgs," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[353] G. Bortolan, I. Christov, and I. Simova, "Rule-based method and deep learning networks for automatic classification of ecg," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[354] B.-J. Singstad and C. Tronstad, "Convolutional neural network and rule-based algorithms for classifying 12-lead ecgs," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[355] H. Hasani, A. Bitarafan, and M. S. Baghshah, "Classification of 12-lead ecg signals with adversarial multi-source domain generalization," *2020 Computing in Cardiology*, pp. 1–4, 2020.

[356] V. Tihonenko, A. Khaustov, S. Ivanov, A. Rivin, and E. Yakushenko, "St Petersburg INCART 12-lead Arrhythmia Database," *PhysioBank, PhysioToolkit, and PhysioNet*, 2008, doi: 10.13026/C2V88N.

[357] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, "PTB-XL, a Large Publicly Available Electrocardiography Dataset," *Scientific Data*, vol. 7, no. 1, pp. 1–15, 2020.

[358] J. Zheng, J. Zhang, S. Danioko, H. Yao, H. Guo, and C. Rakovski, "A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients," *Scientific Data*, vol. 7, no. 48, pp. 1–8, 2020.

[359] J. Zheng, H. Cui, D. Struppa, J. Zhang, S. M. Yacoub, H. El-Askary, A. Chang, L. Ehwerhemuepha, I. Abudayyeh, A. Barrett, G. Fu, H. Yao, D. Li, H. Guo, and C. Rakovski, "Optimal Multi-Stage Arrhythmia Classification Approach," *Scientific Data*, vol. 10, no. 2898, pp. 1–17, 2020.

[360] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[361] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[362] E. A. Perez Alday, A. Gu, A. Shah, C. Robichaux, A.-K. I. Wong, C. Liu, F. Liu, B. A. Rad, A. Elola, S. Seyedi, Q. Li, A. Sharma, G. D. Clifford, and M. A. Reyna, "Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020," *Physiological Measurement*, vol. In Press, 2020.

[363] S. Meek and F. Morris, "Abc of clinical electrocardiography: Introduction. i—leads, rate, rhythm, and cardiac axis," *BMJ: British Medical Journal*, vol. 324, no. 7334, p. 415, 2002.

[364] L. Staerk, J. A. Sherer, D. Ko, E. J. Benjamin, and R. H. Helm, "Atrial fibrillation: epidemiology, pathophysiology, and clinical outcomes," *Circulation research*, vol. 120, no. 9, pp. 1501–1517, 2017.

[365] D. Nankani and R. D. Baruah, "Incorporating demographic and heartbeat features with multichannel ecg for cardiac abnormality detection using parallel cnn and gap network," *2021 Computing in Cardiology*, pp. 1–4, 2021.

[366] P. Saikia, D. Nankani, and R. D. Baruah, "Seismic signal interpretation for reservoir facies classification," *International Conference on Pattern Recognition and Machine Intelligence*, pp. 409–417, 2019.

[367] ——, "Reservoir facies classification using convolutional neural networks," *2019 IEEE Recent Advances in Geoscience and Remote Sensing: Technologies, Standards and Applications (TENGARSS)*, pp. 34–38, 2019.