Integrating Machine Learning and Geospatial Data for GLOF Risk Assessment in Eastern Himalaya

A thesis submitted in partial fulfilment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

by

Anushka Vashistha

(206104003)

Department of Civil Engineering

Indian Institute of Technology, Guwahati

December 2025

Declaration

I hereby certify that the work compiled in this thesis is the outcome of the research work performed by me, as stated elsewhere, under the guidance of Dr. Ajay Dashora and Dr. Afroz Ahmad Shah.

No part of this work has been submitted for the award of any degree, diploma, associate fellowship, fellowship, or its equivalent to any university or institution.

Anushka Vashistha
Registration No.: 206104003
Department of Civil Engineering
Indian Institute of Technology Guwahati

Certificate

It is hereby certified that the research presented in this dissertation, titled "Integrating Machine Learning and Geospatial Data for GLOF Risk Assessment in Eastern Himalaya", by Ms. Anushka Vashistha, a candidate within the Department of Civil Engineering at the Indian Institute of Technology Guwahati, India, has been conducted under my supervision. I further affirm that this work has not been previously submitted elsewhere for any degree.

Date: 22-12-2025

Dr. Ajay Dashora
Supervisor
Assistant Professor
Department of Civil Engineering
Indian Institute of Technology Guwahati
Guwahati-781039, Assam, India

Dr. Afroz Ahmad Shah
External Co-Supervisor
Senior Assistant Professor
Department of Geosciences
Universiti of Brunei Darussalam,
Brunei

Acknowledgement

I stand at the culmination of a remarkable journey–one shaped and enriched by the invaluable contributions and unwavering support of numerous individuals and institutions. As I pause to express my deepest gratitude, I am fully aware that words can scarcely capture the depth of my appreciation.

First and foremost, I wish to express my sincere gratitude to my Dissertation Supervisor, Dr. Ajay Dashora (Indian Institute of Technology Guwahati), and my external Co-Supervisor, Dr. Afroz Ahmad Shah (Universiti Brunei Darussalam, Brunei Darussalam). Your guidance, mentorship, and steadfast belief in the significance of this research have been the cornerstones of my academic journey. Your unwavering commitment to excellence has not only enhanced the quality of this dissertation but has also profoundly enriched my scholarly pursuits. I shall remain forever grateful to you.

Besides my supervisor, I extend my heartfelt appreciation to the members of the Doctoral Committee, Prof. Bimlesh Kumar, Dr. Indu Siva Ranjani G and Dr. Hanumant Singh Shekhawat for their constructive feedback, intellectual insights, and commitment to academic rigour. Your collective wisdom has been instrumental in shaping this dissertation into a meaningful contribution to our field.

IIT Guwahati has been a source of great support for this study, and I am grateful for the resources, research facilities, and vibrant academic environment it provided.

I am truly fortunate to have had the companionship of remarkable peers and colleagues throughout this journey. Dr. Navakanesh Batmanathan, Dr. Nur M. M. Kalimullah, and Dr. Shivam Ojha offered unwavering support and invaluable assistance in strengthening my understanding of computational methods. Our discussions, collaborative efforts, and shared challenges made this endeavor intellectually rewarding and personally fulfilling.

I am sincerely grateful to my senior, Mr. Utla Chandra Sekhar, for his constant encouragement and kindness, and to my lab mates–Mr. P. N. S. Uday Kiran, Mr. Pushpendra Yadav, Mr. Ashutosh Pal, Mr. Umang Mudgal, and Mr. Srimaya Mohapatra–for creating a supportive and pleasant environment in the lab. Their companionship and cooperation have made my time here both productive and enjoyable. Special acknowledgement is due to Mr.

Ayush Thakur, Mr. Antash Kishore Sinha, and Mr. Manish Dalakoti, whose friendship and motivation enriched my academic life with positivity and camaraderie. Their support in navigating work-related challenges and the moments of laughter and relaxation we shared beyond work were invaluable in maintaining balance throughout this journey.

To my parents, Mr. Raju Vashistha and Mrs. Manju Vashistha, brother Tejasva Vashistha, sister Anjaly Vashistha, brother-in-law Maneesh Sharma, and adorable Darshi and Shivaay, all of your unwavering love and support have been my anchor during the turbulent seas of doctoral research. Your sacrifices, encouragement, and belief in my capabilities have sustained me through moments of doubt and exhaustion. I dedicate this accomplishment to you.

This work is also dedicated, with profound respect and affection, to the cherished memory of my late maternal grandfather, Shri Ram Nivas Chaubey, and my maternal grandmother, Smt. Shakuntala Sharma, whose belief, blessings and values continue to guide and inspire me. This accomplishment is as much yours as it is mine, and I remain forever indebted for your love, support, and encouragement.

Lastly, to my friends back home, Madhav Bhatnagar, Madhu Saini, and loved ones who provided moments of respite, laughter, and solace during this rigorous pursuit, I express my gratitude. Your presence in my life has reminded me of the importance of balance, joy, and the pursuit of dreams.

This dissertation stands as a testament to the collective effort of a multitude of individuals and institutions who have illuminated my path. I am deeply humbled and profoundly grateful for their contributions to this work and my personal and academic growth.

Abstract

The Eastern Himalayas, considered one of the most tectonically active and climate-sensitive mountain systems globally, have experienced a rapid increase in the number of glacial lakes over the past thirty years. This trend is driven by accelerated glacier retreat, changing precipitation patterns, and unregulated human activities. Such transformations significantly amplify the risk of Glacial Lake Outburst Floods (GLOFs), which pose severe threats to downstream communities, infrastructure, and ecosystems by releasing large volumes of water and debris. Current risk assessments often depend on a limited set of glaciological and hydrological parameters. They tend to underrepresent the impact of geomorphic, seismic, and multi-hazard factors, and rarely incorporate the quantification of predictive uncertainty. These limitations reduce their effectiveness for decision-making in high-mountain environments that are characterised by scarce data and ongoing dynamism.

This thesis develops a unified, probabilistically informed framework for GLOF hazard and risk assessment in the Eastern Himalayas, integrating geomorphological interpretation, deep learning, multi-criteria decision analysis, and Bayesian inference in a sequential manner. For the selected study area in the Eastern Himalaya, a total of 12,968 gridded open-source images are derived from Google Earth. The work predicts the probability of glacial lake formation (PGLF), targeting the shortcomings of climate-dominated models that neglect geomorphological controls. Erosional features – cirques, valleys, flow channels, retreating glaciers, and neighbouring lakes – were combined with topographic parameters such as elevation, slope, and curvature, derived from high-resolution Google Earth imagery and shuttle radar topography mission (SRTM) 30 m digital elevation model (DEM) data within a 3 × 3 neighbourhood grid. Logistic Regression (LR), Artificial Neural Networks (ANNs), and Bayesian Neural Networks (BNNs) are evaluated, with the BNN achieving the highest accuracy (AUC = 0.878) while quantifying both aleatoric and epistemic uncertainties (10^{-3} to 10^{-4}). Spatial probability maps highlight glacier fore-fields and cirque basins as key formation hotspots, establishing a geomorphology-driven baseline for subsequent hazard assessment.

Building on these predicted glacial lake formation sites, the second component focuses on automating the detection and hazard classification of existing lakes and erosional features – a process traditionally reliant on labour-intensive manual mapping. An EfficientNet-B0 convolutional neural network (CNN) is trained on labelled Google Earth imagery (2015 – 2022) and tested on 12,968 image tiles, achieving intersection-over-union (IoU) scores above

0.88 for all feature classes. The CNN identified 2,647 glacial lakes in the region, which were then classified using a heuristic hazard scoring system into 235 extremely dangerous, 357 very highly dangerous, 175 highly dangerous, 1,269 moderately dangerous, 99 low dangerous, and 512 non-dangerous lakes. It also pinpointed 174 geomorphically suitable locations for future lake development, facilitating proactive monitoring and aiding the shift from potential to actual hazard realisation.

The third component advances from hazard identification to deterministic risk assessment by incorporating a broader range of hazard drivers that have been largely absent from previous models – specifically, seismic and landslide factors. Thirteen parameters, including lake morphometry, proximity to active faults, seismic exposure, landslide occurrence, historical GLOF events, land cover, and anthropogenic exposure, were integrated into a hybrid Multi-Criteria Decision Analysis (MCDA) framework. Parameter weights were derived through a combination of Analytic Hierarchy Process (AHP) and SE, fused using Kullback–Leibler (KL) divergence to balance expert knowledge with data-driven variability. The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and Fuzzy C-Means (FCM) clustering classified 432 lakes as low-risk, 368 as moderate-risk, 227 as high-risk, and 117 as very high-risk, with 95.1% of classifications exceeding 80% confidence.

Recognising that deterministic classifications do not convey uncertainty essential for decision-making, the final component transforms the framework into a probabilistic domain. A BNN was trained on 14 parameters encompassing geomorphology, topography, hazard proximity, climatic extremes, and anthropogenic exposure. Bayesian hyperparameter optimisation achieved 93.4% classification accuracy, with predictive variance below 0.1 for 87.77% of lakes and no case exceeding 0.164. The resulting probabilistic maps revealed that very high-risk lakes are often large, high-elevation proglacial bodies situated near active faults, landslide-prone slopes, and hydropower infrastructure, where seasonal and monsoonal variability amplifies hazard potential.

By progressing from predicting where lakes are likely to form to automated mapping and hazard classification of existing lakes, multi-hazard deterministic risk assessment, and finally uncertainty-aware probabilistic modelling, this thesis delivers the most comprehensive GLOF risk framework to date for the Eastern Himalayas. The results identify spatial hotspots, quantify confidence levels, and provide a transferable methodology for other high-mountain regions

where field-based observation is constrained, thereby supporting targeted monitoring, early warning systems, and climate-resilient infrastructure planning.

Content

List of Figures

List of Tables

Abbreviation

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
AHP	Analytic Hierarchy Process
AWEI	Automated Water Extraction Index
BNN	Bayesian Neural Network
CNN	Convolutional Neural Network
DEM	Digital Elevation Model
FC	Flow Channel
FCM	Fuzzy C-Means
FHI	Flood Hazard Index
GBM	Gradient Boosted Machine
GDN	Graph Deviation Network
GIS	Geographic Information System
GLHE	Glacial Lake at Higher Elevation
GLOF	Glacial Lake Outburst Flood
GPS	Global Positioning System
GPR	Ground Penetrating Radar
HEC-RAS	Hydrologic Engineering Center – River Analysis System
KL Divergence	Kullback–Leibler Divergence
LiDAR	Light Detection and Ranging
LR	Logistic Regression
MC	Monte Carlo
MCDA	Multi-Criteria Decision Analysis
ML	Machine Learning
MNDWI	Modified Normalized Difference Water Index
NB	Naïve Bayes
NDMI	Normalized Difference Moisture Index
NDSI	Normalized Difference Snow Index
NDWI	Normalized Difference Water Index
PCA	Principal Component Analysis
PDGL	Potentially Dangerous Glacial Lake

PGLF	Probability of Glacial Lake Formation
RF	Random Forest
RG	Retreating Glacier
SAR	Synthetic Aperture Radar
SDN	Shallow Decoder Network
SE	Shannon Entropy
SVM	Support Vector Machine

Chapter 1

Introduction

1.1 Background

Glaciers are substantial masses of ice that move under the influence of gravity due to their own weight. They form as a result of snow accumulation, primarily in high-latitude and high-altitude regions (Cuffey et al., 2010). Presently, glaciers occur on every continent except Australia and are distributed across both polar and high-mountain regions. The largest ice masses are located in Antarctica and Greenland, while substantial glacier coverage also exists in the Arctic regions of Canada, Alaska, Iceland, Svalbard, and the Russian High Arctic. Outside the polar zones, major mountain glacier systems persist in the European Alps, Scandinavia, the Caucasus, the Andes, East Africa, New Zealand, and High Mountain Asia, including the Himalaya–Karakoram–Tibetan Plateau region. Together, these glacierized areas cover approximately 14.9 million square kilometres, representing nearly 10% of the Earth's land surface (Flint, 1971; Hamlet et al., 1999; Meier et al., 1996; Rignot et al., 2019; Vaughan et al., 2014). More than 95% of the glaciated area is situated in the polar region (Dyurgerov et al., 2005). Glaciers in non-polar regions support ecosystems involving large populations in many countries. Lakes or glacial water in rivers are the major source of fresh water for fulfilling the needs of the ecosystem, and various activities of agriculture, industry, hydro-power plants, wildlife and domestic works (Barnett et al., 2005; Bolch, 2017; Kaser et al., 2010). For example, during the summer months, precipitation in Central Asia and South Asia is insufficient to meet the water demands for irrigation, domestic use, cattle farming, and other activities. In such circumstances, mountain glaciers serve as a crucial water source for a substantial population and act as a vital resource (Bolch, 2017).

In recent decades, the Earth's temperature has increased at an unprecedented rate, primarily due to the rapid emission of greenhouse gases (carbon emissions) resulting from accelerated industrialisation, deforestation, heightened vehicle pollution, and other anthropogenic activities. As a consequence, glaciers worldwide have been retreating at an accelerated pace since the 20th century. For example, the Siachen Glacier (located in the Eastern Karakoram region, Gangotri, Himachal, Chenab, Parbati, and Baspa basins of the Western Himalaya), the glaciers of the Qinghai-Tibetan Plateau in China, and the northern slope of Qomolangma (Mount Everest) are exhibiting retreat rates ranging from 5 meters per year to 27 meters per

year (Prasad et al., 2009). In the future, glacier retreat is expected to accelerate primarily as a consequence of climate change, including rising temperatures, alterations in precipitation patterns, and increasing anthropogenic activities (Barandun et al., 2022; Chaturvedi et al., 2014; Immerzeel et al., 2012; Lutz et al., 2014; Shea et al., 2015).

Glacier retreat is a fundamental driver in the evolution of high-altitude hydrological systems and geomorphic processes. The progressive down wasting and frontal retreat of glaciers create depressions in the surrounding topography, which gradually evolve into proglacial or supraglacial cavities (Gardelle et al., 2011; Shrestha et al., 2010; Wessels et al., 2002). As meltwater accumulates within these depressions, they develop into glacial lakes, whose size and depth increase in response to sustained glacier retreat and enhanced ablation (Carrivick et al., 2016; Richardson et al., 2000). Over time, the hydrodynamic balance of these lakes is determined by the inflow from glacial melt, precipitation, and ice or snow avalanches, against the storage capacity of the natural dam, typically composed of unconsolidated moraine or ice-cored sediments.

When inflows exceed the hydraulic or structural capacity of the impounding barrier, either by gradual overtopping or sudden failure, the integrity of the lake dam can be compromised. The most common failure mechanisms include piping, slope instability of the dam material, ice-core melting, and shear failure triggered by hydrostatic pressure (Emmer et al., 2016; Westoby et al., 2014). Such failures release large volumes of water and entrained debris in a highly energetic process known as a Glacial Lake Outburst Flood (GLOF). These events can mobilize enormous quantities of sediment and boulders, often producing destructive floods that propagate rapidly downstream, damaging infrastructure, agriculture, and human settlements (Gurung et al., 2017; Sattar et al., 2021).

GLOFs may be triggered by a range of external and internal processes. External triggers include seismic shaking, landslides or rockfalls into the lake, ice and snow avalanches, and extreme meteorological events that accelerate glacier melt (Allen et al., 2016; Richardson et al., 2000). Internal processes such as moraine degradation, permafrost thaw, and freeze–thaw cycling weakens the dam and supply loose sediment to the valley system. Although sediment input can influence lake stability, the presence of easily erodible material within the moraine and along the downstream valley floor largely controls the erosive and transport capacity of the resulting flood. Once a breach develops, this sediment is quickly entrained, which enhances the geomorphic impact and downstream hazard (Haeberli et al., 2017).

In parallel with these physical processes, the exposure of downstream populations to GLOFs has increased in recent decades. Settlements, hydropower facilities, road networks, and tourism infrastructure are expanding into higher-altitude valleys. Many of these developments lie close to potential GLOF flow paths and narrow mountain corridors where flood energy remains high. Recent global assessments estimate that more than nine million people across High Mountain Asia live within zones that could be affected by GLOFs, making the region one of the world's most exposed mountain systems (Taylor et al., 2023). Similar patterns have been documented in Nepal, where approximately 1.5 million people reside within 3 km downstream of moraine-dammed lakes (Chen et al., 2025). This upward expansion of human activity, combined with rapid lake growth, underscores that GLOF risk is shaped not only by changing glacier dynamics but also by the evolving distribution of people, livelihoods, and critical infrastructure in these vulnerable downstream corridors.

Understanding these processes requires situating GLOFs within broader hazard and risk frameworks. Hazard is defined as a process, phenomenon, or human activity that may cause loss of life, injury or other health impacts, property damage, social and economic disruption, or environmental degradation (United Nations Office for Disaster Risk, 2017). Hazards are commonly characterised by attributes such as their location, intensity or magnitude, frequency, and probability (United Nations Office for Disaster Risk, 2017). Risk, by contrast, emerges from the interaction of hazards with the exposure and vulnerability of human or ecological systems. Here, exposure refers to the presence of people, infrastructure, economic activities, or ecosystems in locations that may be adversely affected by a hazard. Vulnerability represents the propensity or predisposition of exposed elements to be harmed, and reflects physical, social, economic, and environmental conditions that influence their sensitivity and capacity to cope. Consistent with the IPCC AR6 framework, risk refers to the potential for adverse consequences and is conceptualised as a function of hazard, exposure, and vulnerability (Intergovernmental Panel on Climate, 2023).

Within this context, GLOFs represent one of the most severe cryosphere hazards in high mountain environments. Their increasing frequency and magnitude–driven by accelerated glacier retreat under contemporary climate change–pose escalating risks to downstream communities, ecosystems, and infrastructure across the Himalaya, Andes, and other glacierized regions worldwide.

These changing exposure patterns have also shifted the nature of impacts associated with GLOFs. Even moderate-sized GLOFs can cause severe damage when they intersect densely settled valleys or major infrastructure such as hydropower plants, bridges, and highways. Events such as the 1985 Dig Tsho flood in Nepal and the 2023 South Lhonak GLOF in Sikkim illustrate how downstream development amplifies the consequences of lake failures (Vuichard et al., 1987; Zhang et al., 2025). As development continues to expand into steep terrain and valley floors, regional exposure is projected to increase, heightening the need for integrated assessments that consider both physical hazard characteristics and the spatial distribution of exposed communities.

Since 1990, there has been a significant global increase in the number of glacial lakes by 53%. Additionally, the areas and volumes (sizes) of these lakes have expanded by 51% and 48%, respectively (Shugar et al., 2020). These scenarios are further increasing the triggering frequency of GLOF events, which are very dangerous for human beings and the environment due to their high speed and potential along the flow path. For instance, based on remote sensing and geomorphological analysis, Cenderelli et al. (2003) find that the GLOF discharges are 7 to 60 times greater than normal floods derived from snowmelt runoff, glacier meltwater, and monsoonal precipitation. This high discharge resulted in specific stream power values ranging from as low as 1900 W/m² in wide, low-gradient valley segments to as high as 51,700 W/m² in narrow, high-gradient valley segments bounded by bedrock. Simultaneously, GLOF significantly modifies the channels and valleys by eroding, transporting, and depositing large quantities of sediment along the flood routes. These transformations channelize the spatial transformation and evolution of glacial lakes at lower elevations.

1.2 Motivation

Accelerating retreat of glaciers under contemporary climate change is contributing to the rapid formation and expansion of glacial lakes, particularly in the Eastern Himalayas (Kaushik et al., 2024; Shugar et al., 2020; Siddique et al., 2023; Wang et al., 2015). Observations indicate that smaller lakes expand at higher rates compared to larger ones, reflecting complex feedback between glacial melt, geomorphology, and hydrology. For example, Bolch et al. (2011) recorded an increase in lakes from 66 to 132 in the Northern Tien Shan region of China between 1970 and 2007. Similarly, Raj et al. (2013) identified 85 new lakes forming in the Sikkim Himalaya, raising the total from 266 to 320 from 2003 to 2010. Kumar et al. (2020) observed that glacial lakes grew from 215 to 255 between 2002 and 2017. Likewise, Ahmed et al.

(2021a) noted an increase from 253 lakes in 1990 to 322 in 2018, with a growth rate of 21.4%. This rising number of glacial lakes presents notable hazards, as demonstrated by the increasing frequency of GLOFs in the Himalayas during recent decades (Nie et al., 2021). Importantly, the Eastern Himalaya shows the highest concentration of such events, with GLOF frequency reported to be nearly three times higher than in other Himalayan sectors (Veh et al., 2020). An example of such destruction is the 1985 Dig Tsho GLOF in the Nepal Himalaya, which resulted in multiple fatalities and extensive damage over an area of more than 60 km downstream. The flood destroyed the Namche hydropower plant, located about 12 km from the lake, significantly impacting the livelihoods and survival of the local population (Vuichard et al., 1987). Another case is South Lhonak Lake in Sikkim, India, known as one of the fastest-growing glacial lakes in the region, which experienced a catastrophic outburst in October 2023. This event led to the destruction of the Teesta III Dam (see Figure 1.1) and caused widespread downstream damage (Sattar et al., 2021; Sharma et al., 2018).

Figure 1.1: Pre and Post GLOF condition of Teesta 3 Dam as indicated by before and after flood image (Source: SANDRP (2023)).

Another recent incident occured in August 2024, when the Sangnga Nehgu Lake in Tawang district, Arunachal Pradesh, experienced a glacial lake outburst flood (GLOF) resulting from

rapid glacier melt, which washed away vital infrastructure such as log bridges and caused damage to traditional Brokpa pathways.

The vulnerability of the Eastern Himalayas to glacial lake outburst floods (GLOFs) is shaped by a combination of climatic, geomorphological, and tectonic factors. Beyond the well-established influence of climate change on glacier retreat, the region's susceptibility is compounded by its geodynamic setting. The ongoing convergence of the Indian and Eurasian plates, with the Indian plate advancing northward at an estimated rate of ~45 mm yr^{-1}, generates one of the most seismically active continental collision zones globally (Bilham et al., 2001; Bilham et al., 2017; Kumar et al., 2019; Liu et al., 2020; Molnar et al., 1975; Afroz Ahmad Shah, 2013). This tectonic interaction drives crustal shortening, uplift, and frequent seismicity, which in turn impose substantial destabilizing forces on glacial systems and their associated proglacial lakes. Earthquakes can directly trigger GLOFs by destabilizing moraine dams, inducing landslides into lakes, or accelerating glacier calving, thereby acting as critical compounding hazards in the region (Avouac, 2003; England et al., 2015).

In addition to seismic activity, other natural processes–including rapid glacier retreat, intense precipitation events, avalanches, and large-scale landslides–are recurrent triggers of GLOFs (Carey, 2005; Jain et al., 2013; Wang et al., 2015; Yongping, 2004). The destructive potential of these events is magnified by the hydrological connectivity between glacial lakes situated along the same valley system. When an upstream lake fails, its floodwaters can propagate downstream, overtopping or breaching successive moraine dams, thereby creating a chain reaction of cascading GLOFs. Such sequential failures have been shown to amplify flood magnitude, extend the hazard footprint over hundreds of kilometres, and increase both geomorphic and socioeconomic impacts downstream (Carrivick et al., 2016; Emmer et al., 2016).

The hazard is further intensified by the rapid proliferation of glacial lakes across the Eastern Himalayas, driven by accelerated glacier mass loss under contemporary warming. In the literature, inventories document both a sharp increase in the number and areal extent of glacial lakes, many of which are situated near densely populated valleys and critical infrastructure (Bajracharya et al., 2009; Bajracharya et al., 2007; Bolch et al., 2011; Gardelle et al., 2011; Song et al., 2016; Wang et al., 2015). This spatial overlap between high-magnitude cryospheric hazards and vulnerable downstream populations highlights the urgent need for comprehensive

monitoring, hazard assessment, and risk evaluation of glacial lakes across the region (Carrivick et al., 2016; Rinzin et al., 2023; Taylor et al., 2023).

However, the Eastern Himalayas pose formidable challenges for field-based monitoring of glacial lakes due to their rugged topography, limited accessibility, and harsh climatic conditions. In this context, remote sensing techniques have emerged as indispensable tools, providing systematic, large-scale, and repeatable observations of glacial features in otherwise inaccessible terrain. The recent integration of machine learning (ML) methods into cryospheric research represents a transformative advancement. ML is a data-driven approach that enables models to learn patterns directly from datasets, unlike conventional statistical methods that rely on predefined relationships. This flexibility allows ML techniques to capture complex, non-linear processes characteristic of glacial and geomorphic systems. ML algorithms have been increasingly applied for automated detection, classification, and temporal tracking of glacial lakes, as well as for predicting lake expansion potential and assessing associated hazards (Qayyum et al., 2020). These models facilitate the efficient processing of large Earth observation datasets, improve the accuracy of lake detection and hazard prediction, and incorporate probabilistic assessments that address uncertainties inherent to GLOF forecasting.

By supporting real-time monitoring, multi-criteria risk assessments, and the design of early warning systems, remote sensing coupled with ML enhances regional disaster preparedness and adaptive capacity. Such integrative, data-driven approaches are essential for safeguarding high-altitude communities and infrastructure that remain increasingly exposed to GLOF hazards under current trajectories of climate change and tectonic activity. The following chapter provides a detailed review of ML-based algorithms applied in GLOF research, with a focus on their methodological strengths, limitations, and emerging applications.

Chapter 2

Literature Review

2.1 Introduction

Glaciers are dynamic bodies of ice that deform and move downslope primarily under the influence of their own mass and gravity (Cuffey et al., 2010). Their movement exerts a powerful erosional force on the underlying bedrock, driven by processes such as abrasion, plucking, and scouring (Boulton, 1979). As glaciers advance, they entrain and transport a wide range of sediments, from fine clay particles to large boulders, thereby reshaping the landscape (Benn et al., 2014). Over time, this erosional activity creates basins and depressions that serve as potential sites for water storage once the glacier retreats (Benn et al., 2014; Sugden et al., 1976). When glaciers extend into lower altitudes or warmer climatic zones, ablation intensifies, and the resulting meltwater accumulates within these depressions, giving rise to glacial lakes (Carrivick et al., 2016).

The spatial distribution of glacial lakes is influenced by both ice dynamics and topography. They typically develop in areas characterized by low surface slope and reduced ice flow velocity, where meltwater can accumulate without being rapidly evacuated (Huybrechts, 2002; Quincey et al., 2007). Such accumulation zones often coincide with topographic depressions formed by erosional processes or by subtle variations in ice thickness caused by differential ice flow speeds and directions (Cook et al., 2012; Huybrechts, 2002). As a result, distinct hydrological and geomorphological conditions govern lake initiation, persistence, and evolution in glaciated environments (Haeberli et al., 2017; Richardson et al., 2000).

In addition to glacial lakes, a variety of other landforms are produced by glacial processes, reflecting both erosional and depositional dynamics. These include moraines, arêtes, horns, foliations, eskers, hanging glaciers, kettles, drumlins, kame terraces, U-shaped valleys, cirques, meltwater channels, and outwash plains. Such features collectively provide critical evidence of past and ongoing glacier–landscape interactions and are central to reconstructing the history of glaciation and associated hazards.

This chapter seeks to synthesize current knowledge on glacial lake formation and associated GLOF hazards, while identifying methodological gaps in the literature. By consolidating state-of-the-art research and analytical approaches, it highlights both the physical

processes leading to glacial lake development and the broader implications of GLOF risk in high mountain regions. The following section discusses the principal factors influencing glacial lake formation and their role in triggering GLOFs.

2.2 Influencing factors for glacial lake formation and GLOF

The formation of glacial lakes is governed by a complex interplay of interdependent factors that can be broadly classified into three categories: (i) above-surface factors, such as atmospheric temperature, rainfall-induced melting, and anthropogenic influences; (ii) surface factors, including geomorphology and topography; and (iii) subsurface factors, such as seismic activity. These categories are not isolated but rather mutually reinforcing, together shaping the initiation, expansion, and persistence of glacial lakes.

2.2.1 Above-Surface Factors

Atmospheric drivers play a critical role in determining the rate of glacier melting and consequent lake development. Rising temperatures and excessive precipitation, influenced by both natural climate variability and anthropogenic activities, are among the most dominant above-surface factors. Temperature increases caused by natural processes such as variations in solar radiation or volcanic activity led to enhanced heat absorption by ice, accelerating ablation and meltwater generation. Long-term studies spanning nearly three decades (1979–2007) indicate that a unit increase in global mean temperature (°K) corresponds to approximately a 95% increase in precipitation intensity, underscoring the strong climatic coupling between temperature and hydrological input to glacier systems (Liu et al., 2009; Liu et al., 2000). Meltwater generated by these processes accumulates in topographic depressions under the influence of gravity, giving rise to glacial lakes.

Rainfall also directly contributes to glacial lake formation in two distinct ways. First, rainfall warmer than the ice surface transfers heat, intensifying surface melt and thereby supplementing meltwater inflows. Second, alterations in precipitation regimes can result in rainfall replacing snowfall at high elevations, a process referred to as rainfall-induced basal melt. This phenomenon not only augments meltwater supply but also reduces snow accumulation, further destabilizing glacier mass balance and promoting lake formation (Liu et al., 2009).

Anthropogenic influences significantly amplify these natural processes. The combustion of fossil fuels, deforestation, and other land-use changes release greenhouse gases, intensifying

the greenhouse effect and driving long-term warming trends. The associated rise in temperatures accelerates glacial melting and enhances precipitation variability, resulting in the expansion and proliferation of glacial lakes. In particular, changes in atmospheric moisture content, driven by human-induced warming, contribute to more frequent and intense precipitation events, further augmenting lake volumes.

The incidence of extreme weather phenomena, notably cloudbursts, has also increased in recent years in response to global warming. Cloudbursts are intense, localized precipitation events typically associated with steep topography, local heating, and rapid uplift of moisture-laden air. The subsequent development of cumulonimbus clouds at high altitudes produces severe thunderstorms and concentrated rainfall. Such events can rapidly increase lake inflows, heightening the risk of overflow or dam instability (Anand et al., 2025; Ganjoo et al., 2025; Nath et al., 2024; S. Singh et al., 2021). Consequently, cloudbursts represent a critical factor linking climate dynamics, regional topography, and glacial lake hazards.

2.2.2 Surface and Subsurface Factors

Surface factors, particularly geomorphology and topography, define the locations where meltwater accumulates. Glacial lakes are often classified into cirque lakes and valley lakes, which are types of glacial erosional lakes. Cirque lakes form within concave, amphitheater-shaped depressions carved by glacial erosion, whereas valley lakes develop along U-shaped valleys created by extensive glacial downcutting. These landforms provide natural depressions that act as reservoirs for accumulating meltwater.

Subsurface processes, particularly seismicity, influence glacial lake systems mainly through their role as external triggers rather than as drivers of lake formation. In steep, tectonically active mountain terrain, earthquakes can initiate landslides, rockfalls, or ice–rock avalanches that enter a lake and generate impact waves capable of overtopping or weakening a moraine dam (Kargel et al., 2016; Keefer, 1984; Roback et al., 2018). These short-term disturbance processes may affect the immediate stability of a lake, but current evidence does not support a significant, long-term reshaping of lake morphology due to seismicity.

In summary, glacial lake formation results from the coupled effects of atmospheric, geomorphological, and tectonic factors, with anthropogenic climate change amplifying their impacts. Above-surface processes such as temperature rise, altered precipitation patterns, and extreme weather events interact with geomorphic depressions and seismic activity to accelerate lake formation and expansion. These interconnected processes, illustrated in Figure 2.1,

highlight the multi-scalar nature of glacial lake dynamics and underscore the necessity of integrated approaches for assessing glacial hazards.

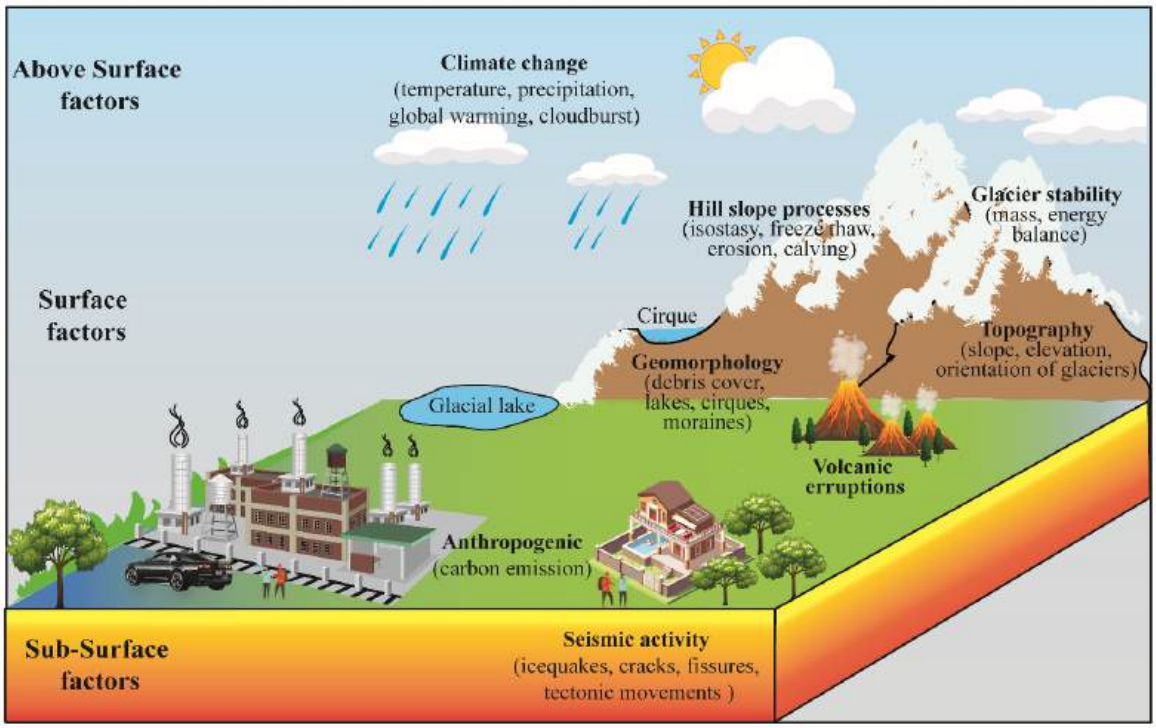


Figure 2.1: Schematic illustration of above-surface, surface, and sub-surface factors influencing glacial lake dynamics and associated hazards.

Figure 2.1 highlights the interconnected nature of glacial lake development, where atmospheric drivers (climate change and precipitation), surface conditions (geomorphology, topography, human activity), and subsurface dynamics (seismicity and tectonics) combine to shape hazard potential. Importantly, it emphasizes that no single factor operates in isolation; instead, the cumulative interaction of climate change, geomorphological setting, and tectonic activity determines the formation, expansion, and eventual outburst risk of glacial lakes.

Glacial lakes are encircled by various geomorphic features like moraines, cirques, debris cover, eroded channels, and floodplains along the flow. Breaching of moraines, which act as natural dams, due to the drainage of meltwater, results in hazardous events like GLOFs. In some regions, geothermal heat from the Earth's interior also melts ice from below, aiding in the formation of subglacial lakes. Thus, increasing global warmth leads to a higher rate of glacier melting, supplying a significant source of water flow to these lakes. Additionally, sub-surface factors like seismic events, causing earthquakes, and impacting glacial lake formation are other minor factors. Earthquakes trigger ice calving or ice avalanches, as well as landslides,

which, in turn, intensify the GLOF potential. The next section discusses the key aspects of the full GLOF hazard–risk continuum in the state-of-the-art literature.

2.3 Research dimensions of GLOF studies

Over the past few decades, research on GLOFs has expanded from basic glacial lake inventories to comprehensive frameworks that address the full hazard–risk continuum. The scientific focus has progressively evolved in four interconnected domains: (i) mapping of GLOF features. After extracting the details of crucial GLOF features, as mentioned in the subsequent section 2.3.1, the study progresses towards (ii) modelling and prediction of glacial lake formation (section 2.3.2), which identifies and monitors existing lakes while forecasting potential new lake sites, (iii) hazard assessment (section 2.3.3), which evaluates the likelihood and potential of a glacial lake outburst based on lake and moraine dam characteristics as well as triggering mechanisms, and (iv) risk assessment (section 2.3.4), which integrates hazard information with downstream exposure to estimate probable impacts on communities, infrastructure, and ecosystems. This framework covers the state-of-the-art GLOF studies, essential for early warning, disaster management and mitigation plans. Together, these research streams form the foundation for proactive monitoring, early warning systems, and mitigation planning, enabling targeted intervention in regions most vulnerable to GLOF impacts. The following sub-sections discuss four domains of glacier research, highlighting various research dimensions and research gaps.

2.3.1 Mapping of GLOF features

Mapping of various features or parameters related to GLOF hazard and risk in the downstream region, such as location of the glacial lake and its feeding glacier, moraine dams, ice cores, mass movements, evolution of the lake, downstream area, population, infrastructure, potential flood propagation parameters, etc, is a critical aspect in GLOF studies. The mapping can be used to create risk evaluations, monitor the spatio-temporal evolution, and guide the creation of mitigation plans for GLOF hazards.

The following three sub-sections present approaches used in mapping the GLOF features associated with GLOF hazard and risk mapping.

2.3.1.1 Remote sensing methods

Glaciers may be mapped up using remote sensing techniques by employing satellite images or aerial images. Based on the dataset retrieval method, remote sensing approaches are broadly

divided into two types: (i) terrestrial remote sensing methods, and (ii) satellite-based remote sensing methods. The terrestrial remote sending method uses Global Positioning System (GPS), Total Station (TS), and Ground Penetrating RADAR (GPR). Whereas, depending on the type of sensor used for mapping, the satellite-based remote sensing method derives information from multiple types of images, namely (i) Optical images (optical bands combination by using sensors like Landsat and Sentinel-2), (ii) Synthetic Aperture Radar (SAR or microwave) images, (iii) Thermal images and (iv) Gravity and satellite altimetry (Bhambri et al., 2011; Karimi et al., 2012; Shukla et al., 2010). In addition, high-resolution maps obtained from airborne LiDAR data are also used. Above, terrestrial-based, aerial-based, and satellite-based techniques are used for reviewing the stability of glaciers. For example, GPS and GPR results are in well agreement with in-situ profiling of glaciers that are used for the monitoring of snow accumulation and snow depth in glaciers, glacier boundary detection, englacial and subglacial debris mapping, ice thickness and volume detection (Ai et al., 2014; Bælum et al., 2011; Ma et al., 2010; Machguth et al., 2006).

Satellite-based methods provide certain benefits over the terrestrial and airborne remote sensing methods, like massive area coverage in a short duration, repeatable assessments, multispectral multi-magnification, better spatial resolution, and safety. Further, optical methods, SAR and LiDAR are used for precisely tracking the transformations in the glacial landscape and lake (Alifu et al., 2020; Bertone et al., 2019; Chelamallu et al., 2014; Feyisa et al., 2014; Fischer et al., 2015; Janke, 2013; Kumar et al., 2012; Panwar et al., 2023; Shen et al., 2010; Zhang et al., 2021). Next, thermal remote sensing has certain advantages over LiDAR and SAR-based approaches, like the detection of internal ice features such as subglacial lakes and ice layers and direct assessment of ice temperature. Further, gravity and satellite altimetry techniques are applied for analysing the glacier mass and volume change (Velicogna et al., 2006).

Indices-based methods are the most primitive remote sensing methods that utilizes the surface reflectance property through spectral band values of the satellite-based image data directly for the development of feature (water, snow, vegetation, and ice) based index. The index method is applied for evaluating the spatiotemporal variation in the surface area of the glacial lake. The value of band ratios was normalized: (i) to avoid the effect of topography (relief), illumination and atmosphere of the pixels, and (ii) to bring index values in a -1 to 1 range. Various methods like band ratios (Wessels et al., 2002) and multi-band index ratio like Normalized Difference Water Index (NDWI) (McFeeters, 1996), Normalized Difference Snow

Index (NDSI) (Hall et al., 1995), modified NDWI (Xu, 2006), Automatic Water Extraction Index (AWEI) (Feyisa et al., 2014), Water Ratio Index (WRI) (Shen et al., 2010), and Land and Water Boundary Enhancement Index (LBWEI) (Khalid et al., 2021) are applied for automatic extraction of waterbody.

Each of these indices represents a step forward in the remote sensing of water bodies, addressing specific limitations of their predecessors and enhancing the capability to monitor and analyze glacial lakes and water dynamics, critical for understanding the impacts of climate change on glacier-fed river systems and the broader environment. The NDWI and the NDSI laid the foundational groundwork by enhancing the detection of water bodies (in comparison to vegetation and soil) and snow cover, respectively. Despite their innovation, the NDWI and NDSI techniques have certain limitations. One major challenge is spectral mixing, where the presence of multiple materials within a single pixel can lead to inaccurate classifications. Additionally, variations in snow grain size can significantly affect the reflectance values captured in satellite imagery, making it challenging to differentiate between snow and various other surface types. These surfaces primarily include densely vegetated areas, bare soil, and urban materials. Thus, snow reflectance varies not only with grain size but also with the snow's age and contamination level, leading to a spectral signature that can sometimes mimic or overlap with those of other surface materials under certain conditions. This spectral similarity can complicate the accurate classification and analysis of snow-covered landscapes, especially in regions where these different surface types are nearby. Another difficulty arises in the detection of water bodies covered by ice, as the spectral characteristics of ice can closely resemble those of snow or clouds, leading to potential misclassification. Furthermore, cloud cover can obscure the view of water bodies, making it difficult to accurately identify and classify them using these indices (Hall et al., 1995; McFeeters, 1996; Zhang et al., 2021; Zhang et al., 2015). Addressing these limitations, the MNDWI emerged, offering improved discrimination of water bodies by adjusting the spectral bands used, thus reducing misclassifications associated with vegetation and built-up areas (Xu, 2006). Building on this progress, the WRI and LBWEI were developed to refine water detection further and enhance the delineation between land and water boundaries, crucial for accurately mapping shoreline changes. The AWEI represents the latest stride in this evolution, incorporating multiple bands to minimize the misclassification of water with shadows and vegetation, showcasing the continuous effort to improve accuracy in water body detection. Despite its advancements, AWEI still struggles to distinguish between turbid water and other surface types like pavements

and agricultural fields flooded for irrigation purposes. Further, LBWEI showed extended accuracy for all types of seasons in comparison to other multi-band water body classification indices (NDWI, MNDWI, NDMI, AWEI, NWI and WRI) that were finding it difficult to eliminate the non-water body pixels like snow, shadow, and dark pixels (Khalid et al., 2021; Sarp et al., 2017).

On the other hand, indices-based methods require the selection of appropriate threshold values to effectively discriminate between water bodies and surrounding landscapes. These threshold values, however, are not universal and can vary significantly depending on the specific geographic location and environmental characteristics of the study area. In general, the value of NDWI varies from 0.1 to 0.5 for the detection of potentially dangerous glacial lakes. For instance, in the Himalayas, researchers set a higher threshold for indices like NDWI and MNDWI to account for the high reflectance of snow and ice, often using values above 0.3 to ensure water bodies are accurately distinguished from the surrounding snow and ice cover (Xu, 2006). In contrast, in the Andes or the Alps, where glacial lakes are often surrounded by dense vegetation, slightly lower thresholds are used, around 0.2 to 0.3, to differentiate water from vegetated areas effectively (Feyisa et al., 2014). Similarly, the NDSI index with the shortwave infra-red band is used in the Himalaya for classifying the debris cover from snow and ice, with a threshold value ranging from 0.17 to 0.20 (Bolch et al., 2011; Kulkarni et al., 2002). For NDSI, a common threshold value is 0.4, as established in studies focusing on snow cover in mountainous regions, including those in North America and Europe (Hall et al., 1995). Moreover, these threshold values are often determined empirically and may be adjusted based on the specific objectives of the research, the temporal aspects of the satellite data (e.g., seasonal variations), and the spectral characteristics of the imagery. Furthermore, the index-based methods show optimal results with satellite imagery that has minimal snow and cloud cover to ensure clear visibility of glacial lakes. Typically, such data is most readily available only during the months of September, October, or November in the Himalayas (Khalid et al., 2021; Nijhawan et al., 2016). Apart from the above limitations, index-based methods are restricted to identifying a particular class or feature at a time.

While remote sensing indices have proven effective for large-scale monitoring, they are limited by spectral ambiguities and threshold dependence. To overcome these challenges, ML approaches have been increasingly adopted, as discussed in the following subsection.

2.3.1.2 Machine learning (ML) based methods

ML-based techniques are an apt and suggestive approach for the regions where the classical approach of classification tends to show limitations. Ensemble learning, ML, and deep learning are the major modern classification techniques. ML techniques like support vector machine (SVM), random forest (RF), naïve bayes (NB), gradient boosted machine (GBM), K-nearest neighbor, and decision tree find a wide applications in image classification (Wu, 2017; Wu et al., 2008; Zhang et al., 2019). The SVM approach shows higher accuracy in water body extraction than the index-based methods like MNDWI, NDWI and AWEI (Sarp et al., 2017). Next, the RF method, which is a pixel-based classification approach, is the most appropriate technique for clearly delineating the object boundaries in comparison to the aforementioned methods (Alifu et al., 2020; Khan et al., 2020). However, the RF method lacks generalization for the classification problem of linear fractionable feature data sets, which can be improved by selecting a large pixel size, leading to high computational cost in the classification of high-resolution images. With large data, deep learning is primarily suitable for object detection, feature similarity-based image segmentation, and image classification by incorporating the repeating patterns and embedded texture information (Lu et al., 2021; Robson et al., 2020; Thati et al., 2022; Xie et al., 2020). Deep learning (DL) is a specialised branch of machine learning that employs neural networks with multiple hidden layers to learn data representations at increasing levels of abstraction. While ML methods rely on input features that are manually selected or engineered, DL models extract hierarchical features automatically from the raw data through stacked non-linear transformations. The fundamental distinction lies in this representation learning capability: ML captures patterns from predefined variables, whereas DL derives both the features and the predictive relationships directly from the data. A convolutional neural network (CNN) is one such deep learning approach applied for the mapping of glacier features. In recent research, a model that combines CNN and object-based image analysis was suggested to automatically map the rock glaciers (Indolia et al., 2018; Robson et al., 2020). The efficiency of CNN for feature extraction depends on the number of hidden layers in the framework. CNN uses a large amount of training data, which leads to over-fitting, and makes it uncertain in detecting the boundaries at some locations (Lu et al., 2021). Recently, an improved method of CNN, named channel attention U-Net, has been used for the cartography of glaciers. Also, relative to basic U-Net and GlacierNet, the findings demonstrate that the channel-attention U-Net distinguishes glaciers with a comparatively high degree of accuracy (Tian et al., 2022; Xie et al., 2020). However, limitations like the presence of debris and seasonal snow cover reduce the accuracy of tracing the clear outline of the glacier in these

methods (Huang et al., 2021; Zhang et al., 2019). Recent research by Lu et al. (2021) finds that the RF-CNN method, which is an ML-based model, is advantageous for the distinct identification of active and inactive debris-covered glaciers. Hence, ML-based techniques can improve the understanding of vital processes of glacier change by supplementing the traditional mapping approach with massive data analysis and interpretations.

Features like glacial lakes, cirque, debris, snow cover, moraines and ice affect the stability of water management resources. Therefore, mapping of the GLOF features is performed in two steps: (i) Feature extraction, and (ii) Classification techniques

A. *Feature extraction*

Color, shape and texture are the basic anatomic characteristics of a feature, and the process of feature extraction complies with these characteristic properties. It formulates the inherent qualities of a feature like entropy, pigment, and pattern in accordance with the spatial distribution of pixel intensities (Mutlag et al., 2020). In general, glacial feature extraction is carried out to extract information about the ice cover, snow, lakes, cirques, and debris. Various tools are developed for extracting glacial features, for example, a texture filter named as Gabor filter is applied over the rock glaciers for the automatic identification of flow patterns. Here, the terrain attributes assist in the mapping of flow patterns with unique surface patterns and little spectral difference from the surrounding areas (Brenning et al., 2012). Further, the mapping of drumlins by object-based technique, which can cluster the pixel values and spatial parameters (shape, texture and surrounding relations) and provides an appropriate substitute for manual digitization (Saha et al., 2011). Similarly, principal component analysis (PCA) fits well for the digitization of the glacier features like crevasses and supra-glacial lakes, providing easy monitoring of the features. Also, PCA-based fusion is better than the instant hue saturation-based fusion technique (Kumar et al., 2012). Further, an automated algorithm, Auto Cirque, developed in Python, has been used for the mapping of cirque bodies. This tool considerably improves the analytical process and reduces subjectivity, providing a generalised framework for cirque delineation (Li et al., 2022). However, its effectiveness remains constrained by DEM resolution, region-specific threshold calibration, and challenges in complex or tectonically modified terrains, limiting its full transferability across mountain regions. Ice, snow, debris, lake, cirque, hills, flow channels, and moraines are some of the features which change with the influence of climate change. Hence, regular monitoring of the spatial and temporal coverage of the features will indeed be beneficial for an early warning.

B. *Classification techniques*

Various approaches of the classification techniques like unsupervised, supervised, and object-based characterization, provide a faster and efficient method for the extraction and identification of glacier surface features. An unsupervised classification technique operates by clustering without using the training datasets. K-means and iterative self-organizing data analysis (ISODATA) are the most common classifiers that employ this clustering approach for the change detection. Next, in supervised learning, each class is given training data (class prototypes) and decision rules are directly established to aid with classification. Further, object-based classification involves a two-stage process: multi-resolution segmentation and segment categorization based on knowledge (linguistic, geographic, spectral and textural data). For example, an unsupervised classification technique based on the interferometric coherence signal is used for analysing the displacement of rock glaciers and distinguishing the moving and non-moving (without discernible displacement) rock glaciers (Bertone et al., 2019). Further, Spectral Unsupervised Feature Extraction (or SpecUFEx) is applied to trace the contribution of englacial and supraglacial lakes to the GLOF, by monitoring the icequakes and seismic noise. In this method, unsupervised learning forms the clusters of cryo-seismic signals with similar traits, demonstrating the state of immediate flood and post-flood melt flow (Sawi et al., 2022). Next, object-based methods are used to determine the change in glacier cover, and their accuracy surpasses the application of sub-pixel-based categorization (shadow problem) and the index-based supervised classification (Nijhawan et al., 2016; Saha et al., 2011). Also, high, mid, and low percolation, exposed ice, and debris-cover zone are some of the essential features of PDGLs that are identified and categorised with relatively high accuracy using GF-DNN and SVM (Panwar et al., 2023). Despite that, precise landform categorization of satellite-based datasets is still a challenge for researchers, owing to the problems raised by satellite datasets, computational errors in image processing, and miscellaneous influences of terrain.

Given the strengths and weaknesses of both classical and ML-based approaches, recent studies increasingly employ hybrid methods that combine multiple techniques for improved accuracy

2.3.1.3 Hybrid methods

The hybrid approach involves the combination of discrete techniques to achieve higher accuracy over the application of the individual method. The classical methods face difficulty in delineating the boundary of features from the surrounding bedrock, shadow and snow cover

(Lu et al., 2021). Multifarious hybrid models are often applied by researchers to extract the features and develop a deep insight into the impacts induced by climate change. For example, a composite model of CNNs and the RF method was applied for demarcating the Himalaya glaciers with debris cover. The accuracy of this approach supersedes the individual application of artificial neural network (ANN), RF and SVM methods (Nijhawan et al., 2018). Further, the producer's accuracy for the maps generated by the composite model frameworks has shown a considerable increment in classification with high-resolution images (Robson et al., 2020). Similarly, Xie et al. (2022) developed the multi-model learning for alpine glacier mapping. This multi-model process images data from two types of CNNs, named GlacierNet, followed by DeepLabV3+, to detect the snow-covered accumulation zone. Moreover, a recent study on the Gyirong Zangbo watershed evaluated both individual ML algorithms and hybrid models–Logistic Regression (LR), SVM, genetic algorithm–SVM (GA-SVM), and certainty factor-GA-SVM (CF-GA-SVM)—to map debris flow susceptibility. Among these, the CF-GA-SVM model demonstrated the highest predictive accuracy (precision = 0.8) and the best model performance (AUC = 0.945), making it the most suitable approach for generating debris flow susceptibility maps (Qiu et al., 2022). Besides this, the hybrid ML method, Particle Swarm Optimization–Extreme Learning Machine–Adaptive Boosting (PSO-ELM_AdaBoost), was compared with conventional ML techniques such as Back-Propagation Neural Network (BPNN), SVM, Extreme Learning Machine (ELM), and PSO-optimized Extreme Learning Machine (PSO-ELM). The PSO-ELM_AdaBoost model demonstrated superior performance in estimating potential debris-flow volumes, achieving a mean absolute percentage error of less than 0.10 (Huang et al., 2020). Overall, hybrid methods of glacier mapping and monitoring can yield very precise and comprehensive briefing about the glacier extent, topography, landforms, and anatomic characteristics of features. By blending diverse forms of datasets and measurement approaches, a problem which was coarser in outputs can be simplified to finer results and will assist the researchers in better understanding of glaciers behavior.

2.3.2 Modelling and predicting Glacial Lake Formation

The prediction of glacial lake formation is a critical first step in understanding and mitigating GLOF hazards. It requires not only the detection and monitoring of existing lakes but also the identification of potential sites where new lakes may develop in deglaciated zones under continued glacier retreat. Advances in remote sensing and ML have greatly enhanced this process, enabling the detection of spatio-temporal trends, quantification of morphometric attributes, and forecasting of lake emergence based on climatic, glaciological, and

geomorphological factors. Remote sensing technologies–particularly satellite and airborne products such as optical imagery, precipitation, and temperature data–have become state-of-the-art tools for glacial studies. In data-rich contexts, field observations and in-situ measurements further support AI-driven semi-automated systems. Such foresight enables early intervention in areas prone to hazardous lake development, thereby strengthening preparedness and mitigation planning.

2.3.2.1 Glacial lake formation

Studies predicting glacial lake formation or lake presence have concentrated on climatic drivers–such as atmospheric warming, changes in precipitation regimes, and glacier mass loss–as primary factors influencing lake genesis and expansion (Veh et al., 2020; Wang et al., 2015). While these drivers govern meltwater production, they do not fully account for the spatial distribution of glacial lakes. Increasing evidence shows that geomorphological and topographic controls play an equally decisive role in determining where lakes form. Erosional and depositional landforms–such as cirques, U-shaped valleys, flow channels, moraines, and glacier snouts–create natural depressions that act as basins for meltwater accumulation. The slope, curvature, and elevation of these basins dictate meltwater routing, lake connectivity, and seasonal hydrological responses. Empirical analysis of over 200 glacial lake sites (Mal et al., 2020) The Eastern Himalaya demonstrates that the coupled dynamics of retreating glaciers, cirques, valleys, and flow channels fundamentally control lake occurrence, growth patterns, and stability. These geomorphic features also influence sediment delivery, which can alter lake turbidity, basin infilling rates, and dam stability–factors that directly affect hazard potential.

Approaches for predicting glacial lake formation can be broadly categorized into geomorphological, statistical, and ML-based methods.

A. *Geomorphological modelling method*

Methods proposed by Furian et al. (2021) reconstruct subglacial bedrock depressions beneath retreating glaciers to locate potential lake sites. While physically grounded, they often apply minimum area thresholds (e.g., > 0.1 km²). Morphological scoring systems (T. Zhang et al., 2022b) use weighted indicators like slope, crevasse density, and glacier proximity but rely heavily on expert judgement, which makes them prone to subjective bias and the omission of key features such as cirques or flow channels.

In addition to morphological scoring frameworks, several studies have applied physically based approaches to predict where new lakes may form beneath retreating glaciers. The most

widely used among these is the GlabTop (Glacier Bed Topography) and GlabTop2 modelling framework, which reconstructs glacier bed geometry by combining glacier outlines, surface slopes, and estimated ice thickness to map potential over-deepenings. These overdeepenings represent sites where future lakes are likely to develop as ice retreats. Linsbauer et al. (2012) applied GlabTop2 across the Himalaya and Karakoram to identify likely future lake basins, highlighting the spatial clustering of overdeepened zones beneath debris-covered glaciers. Drenkhan et al. (2018) extended this approach to the Peruvian Andes, integrating GlabTop outputs with regional glacier-retreat scenarios to map future lake formation hotspots. These studies show that physically based reconstructions of subglacial topography provide critical foresight into lake emergence patterns and complement geomorphology-based and ML-based predictive frameworks.

B. *Statistical models*

LR has been used to relate lake presence to climatic, glaciological, and topographic predictors (Mohanty et al., 2022). However, LR or any linear model often captures a subset of predictors (around 50% in some studies) as statistically significant variables for various factors, namely, multi-collinearity, low predictive power of certain variables, or limited sample variability.

C. *ML-based approaches*

ANNs have demonstrated the ability to capture non-linear dependencies and complex feature interactions, successfully applied in glacier boundary mapping, meltwater modelling, snow cover prediction, and ice thickness estimation (Anilkumar et al., 2023; Baumhoer et al., 2019; Bolibar et al., 2020; Dobreva et al., 2011; Haq et al., 2021; Hou et al., 2014; Kim et al., 2016; Mohajerani, 2019; Zaier et al., 2010). However, their deterministic nature and lack of built-in uncertainty quantification can limit reliability in hazard-sensitive applications. Bayesian Neural Networks (BNNs) address this limitation by integrating probabilistic inference into neural network architectures. They quantify both aleatoric (data-driven) and epistemic (model-driven) uncertainties, capture interdependencies among diverse input variables, and are less prone to overfitting when trained on limited datasets—a common challenge in high-mountain regions. Studies have shown BNNs to be effective in incorporating sparse field measurements, multi-source remote sensing inputs, and derived geomorphic indicators for robust hazard forecasting (Gopalan et al., 2018; Guillet et al., 2023; Hartmann et al., 2021; W. Ren et al., 2018; W. W. Ren et al., 2018; Rounce et al., 2020; Werder et al., 2020).

Given the rapid pace of environmental change, regularly updating glacial lake inventories is crucial for effective predictive modelling. While the integration of high-resolution, multi-sensor remote sensing datasets–such as Google Earth imagery, optical and SAR data, LiDAR, and UAV photogrammetry–with advanced ML frameworks offers significant potential, their application remains limited in many studies, particularly in data-scarce mountain regions. Furthermore, despite the critical role of erosional features in glacial lake evolution, these variables are often underrepresented, further constraining the ability of predictive models to anticipate the formation of hazardous lakes in complex terrains. Addressing these gaps is essential, as robust predictive mapping underpins PDGL classification, hazard assessment, and the development of targeted monitoring strategies.

2.3.3 Hazard assessment of glacial lakes

Hazard assessment of glacial lakes is a critical component of GLOF studies, aimed at identifying lakes with the potential to generate damaging outburst floods. It quantifies the likelihood of failure and the magnitude of potential outburst. Methodologically, GLOF hazard assessment has progressed from simple geomorphic inventories and empirical thresholds, through indicator-based and multi-criteria classification, to more advanced process-chain modelling that integrates triggers, breach mechanics, and flood routing (Allen et al., 2016; Emmer et al., 2016; Huggel, 2004; Huggel et al., 2002; Kougkoulos et al., 2018; Rounce et al., 2017; Worni et al., 2013).

2.3.3.1 Hazard Assessment techniques

A. *Geomorphic screening:*

Early hazard assessments relied on remote sensing to extract morphometric attributes (lake area, volume, freeboard), dam characteristics, and glacier proximity. Empirical rules–such as moraine width-to-height ratios, rapid lake expansion, and steep hanging glaciers nearby–are then used to identify potentially dangerous lakes (Bolch et al., 2008; Huggel et al., 2002; Veh et al., 2020). These approaches remain valuable for rapid regional screening but depend heavily on subjective thresholds and cannot capture complex triggers or internal dam properties.

B. *Multi-Criteria Decision Analysis (MCDA) frameworks*

To manage large lake inventories, researchers have developed indicator-based scoring systems integrated with MCDA. Studies in the Himalaya and Andes combine morphometry, dam composition, dynamic setting (e.g., avalanche/landslide potential), and historical activity into composite hazard indices (Emmer et al., 2016; Kougkoulos et al., 2018). Weighting is

achieved through expert elicitation by the Analytic Hierarchy Process (AHP) method, objective methods by Shannon Entropy (SE), or hybrid schemes. These frameworks produce reproducible hazard rankings, though these are sensitive to indicator selection and weighting assumptions.

C. *Process-chain analysis methods:*

Recognising that many GLOFs are triggered by external forces such as ice/rock avalanches, landslides, or seismic activity, recent studies incorporate process-chain modelling in a series of steps: avalanche dynamics, impact wave generation, dam overtopping and erosion, breach hydrograph, and downstream flood routing (Somos-Valenzuela et al., 2015; Worni et al., 2013). These physics-based models provide scenario-specific hazard estimates, but require high-quality DEMs, dam material data, and computational resources, making them best suited for high-priority lakes identified in earlier screening. Beyond the Himalaya, several studies from the European Alps and the Andes have advanced probabilistic and scenario-based approaches for glacial lake hazard assessment. In the Swiss Alps, Huggel et al. (2004) proposed a semi-quantitative procedure that evaluates both the magnitude of glacial hazards and their probability of occurrence, integrating these elements into hazard maps and risk levels. Nussbaumer et al. (2014) extended this line of research by coupling glacier-retreat modelling, possible future lake formation, and flood simulations to estimate scenario-based GLOF risks in Alpine valleys. In the Peruvian Andes, Somos-Valenzuela et al. (2016) developed a full GLOF process-chain model for Lake Palcacocha, simulating mass movement initiation, lake-impact waves, dam overtopping or breach, and downstream flooding for multiple scenarios. These studies illustrate the global shift toward probabilistic, process-chain, and scenario-based modelling frameworks, providing broader international context for the uncertainty-aware risk evaluation developed in this thesis.

D. *Breach and outburst modelling*

Estimating potential peak discharge and flood volume is central to hazard assessment. Empirical relations link lake volume and dam height to peak discharge (Clague et al., 1973; Huggel, 2004; McKillop et al., 2007), while process-based models (e.g., BREACH, BASEMENT) simulate erosion, widening, and hydrograph evolution (Worni et al., 2013). Empirical tools are efficient for regional use, whereas process models capture erosion dynamics but require detailed input data.

E. *ML methods*

ML methods are being increasingly applied to automate lake outline extraction, detect avalanche source zones, and even rank hazard potential based on multiple predictors. For example, RF and SVM have been used to classify potentially dangerous lakes in the Himalaya (Ahmed et al., 2022a; Ahmed et al., 2021b; Qayyum et al., 2020). Deep learning approaches, including CNNs and U-Net segmentation, show promise for detecting irregular lake geometries and debris-covered glaciers (Xu et al., 2023). However, ML-based hazard models are limited by the scarcity of labelled "failure" vs. "non-failure" datasets and require careful uncertainty treatment.

Despite these advances, many state-of-the-art methods continue to emphasise the geometric relationship between a lake and its neighbouring glacier, often overlooking the influence of broader erosional features that strongly control lake formation and stability. Detecting such complex landforms remains challenging because of their irregular shapes, variable scales, and spectral similarity to surrounding terrain. Recent progress in neural networks—particularly CNNs and segmentation models—has shown improved capabilities in capturing these patterns. For instance, Qayyum et al. (2020) and Xu et al. (2023) successfully apply CNNs to delineate glacial lake shapes from Sentinel (10 m resolution) and PlanetScope (3m resolution) imagery, while Scuderi et al. (2022) demonstrate the detection of irregular cirques using bounding box segmentation. These developments suggest that integrating advanced feature-extraction methods into hazard assessment frameworks could significantly improve the accuracy and comprehensiveness of GLOF hazard evaluation.

2.3.3.2 Hazard classification schemes

A key outcome of hazard assessment is the classification of glacial lakes into hazard categories, which enables prioritisation for monitoring and mitigation. Several approaches exist across the literature. Early frameworks relied on binary classification, where lakes were designated as potentially dangerous glacial lakes (PDGLs) if they exceeded thresholds for area, dam condition, or susceptibility to triggers. (Huggel et al., 2002) and subsequent inventories in the Himalaya and Andes widely applied this approach. To provide greater nuance, researchers later developed ordinal classification systems that rank lakes into low, medium, high, or very high hazard classes, as demonstrated by Allen et al. (2016) for the Himalayan glaciers and Bazai et al. (2021) in the Karakoram glaciers. Building on this, scoring and index-based methods combine multiple indicators into composite hazard indices, such as the 15-parameter system proposed by Emmer et al. (2016) in the Cordillera Blanca and MCDA-based schemes

applied in the Himalaya (Kougkoulos et al., 2018; Rounce et al., 2017). More recently, process-based classification has been employed, where lakes are categorised according to simulated breach scenarios and modelled outcomes of overtopping probability and peak discharge (Somos-Valenzuela et al., 2016; Worni et al., 2013). Together, these approaches demonstrate a methodological evolution from rapid binary screening to physics-based hazard classes, with most studies adopting a tiered approach that uses initial screening to shortlist lakes for more detailed modelling.

Despite these advances, many hazard assessment frameworks continue to focus primarily on measurable morphometric parameters and trigger indicators, while underrepresenting the influence of broader geomorphic and glaciological processes–such as cirque development, valley deepening, moraine dynamics, and glacier retreat–that fundamentally control lake stability. Effective hazard assessment, therefore, requires not only precise quantification of lake characteristics but also explicit integration of these interconnected landscape processes and their temporal evolution.

2.3.4 Risk Assessment

Risk assessment builds upon hazard assessment by integrating information on the physical likelihood and intensity of a potential GLOF with the spatial distribution of downstream exposure. It estimates the probable consequences of an outburst by accounting for the location and vulnerability of populations, infrastructure, and economic assets along potential flood paths. By translating hazard indicators into expected socio-economic and environmental impacts, risk assessment enables the delineation of high-risk zones and supports the prioritisation of mitigation, early warning, and preparedness strategies. While hazard assessment focuses on the likelihood and magnitude of an outburst, risk assessment translates these into consequences by explicitly integrating exposure and vulnerability. The following sub-sections discuss various approaches to risk assessment for GLOF.

2.3.4.1 Glaciological and hydrological-based risk assessment of glacial lakes

Current approaches to glacial lake risk analysis can be broadly classified into Hydrodynamic, statistical, MCDA and ML.

A. *Hydrodynamic and statistical models*

Hydrodynamic models such as Hydrologic Engineering Center – River Analysis System (HEC-RAS) represent a quantitative class of tools that simulate flood wave propagation, downstream inundation extents, and infrastructure impact, using parameters like lake volume,

channel geometry, and flow roughness—most of which can be derived from remote sensing products such as DEMs, optical imagery, and land cover maps. While highly accurate for site-specific analysis, these models require precise DEMs, field calibration, and often struggle in steep or densely vegetated terrain (Hazra et al., 2022; Klimeš et al., 2014). Statistical models, on the other hand, offer semi-quantitative, probabilistic risk estimates based on historical GLOF frequencies, lake area change rates, glacier retreat, and climate variables, but are limited by their dependence on past events and inability to capture novel triggers (He et al., 2022).

B. ***Multi-Criteria Decision Analysis (MCDA)***

MCDA methods offer a scalable and interpretable framework for integrating diverse hazard and exposure parameters—such as lake expansion rate, dam type, proximity to infrastructure, avalanche/snowpack risk, and population density—into a single prioritisation index. Remote sensing plays a key role here, with time-series optical/SAR imagery for lake monitoring, DEMs for topographic and avalanche-prone slope analysis, and nighttime lights for settlement mapping. Within MCDA, subjective techniques like the AHP and Delphi method rely on expert judgment, potentially introducing bias (Agustina et al., 2023; Das et al., 2024; Kumar et al., 2022; Mohanty et al., 2022; Zulkarnain et al., 2021), whereas objective methods such as SE derive weights from data, reducing subjectivity but potentially overlooking nuanced contextual knowledge (Arora et al., 2021; Chen et al., 2023; Zorlu et al., 2023). Hybrid weighting schemes aim to merge these advantages, but many rely on simple linear combinations lacking a rigorous theoretical framework (Kougkoulos et al., 2018; Malczewski et al., 2015). Furthermore, most MCDA-based models do not incorporate uncertainty quantification, limiting their operational reliability for risk-informed decision-making.

C. ***Machine Learning (ML)***

ML methods range from traditional algorithms–such as LR, decision trees, and RFs–to advanced deep learning architectures like CNNs and ANNs (Allen et al., 2019; Jena et al., 2021). These models can integrate large, multi-source datasets and capture complex, non-linear relationships between hazard and exposure variables. ML-based frameworks, which include ML-based methods, have been applied for lake detection, hazard scoring, and susceptibility mapping, sometimes in combination with MCDA to incorporate expert knowledge. For example, (Mondal et al., 2025) used ANNs trained on AHP-derived parameters, improving interpretability. However, the study still lacks explicit uncertainty estimation–an essential component for disaster risk reduction in data-sparse, hazard-prone environments.

Recent advances in probabilistic ML models–including BNNs, Monte Carlo (MC) dropout, and deep ensemble methods–address this gap by jointly modelling predictions and uncertainties, capturing both aleatory and epistemic components (Guillet et al., 2023; Jospin et al., 2022). These approaches generate calibrated confidence intervals on hazard predictions, enabling prioritisation of monitoring and intervention where in-situ validation is infeasible. However, their adoption in GLOF risk assessment is still rare.

Given the strengths and weaknesses of hydrodynamic methods, statistical methods, MCDA methods, and ML methods for the purposes of glacial lake formation and mapping, hazard assessment, and risk assessment, contemporary frameworks increasingly combine them—for example, statistical models for initial high-probability lake screening, hydrodynamic models for downstream impact mapping, and ML models for integrating multi-source hazard and exposure datasets with uncertainty quantification. Such integrated, uncertainty-aware, multi-hazard modelling represents a critical research frontier for advancing GLOF risk assessment.

2.4 **Aims and Objectives of the Study**

GLOFs have emerged as one of the most severe cryospheric hazards in high-mountain environments, particularly in the Eastern Himalaya, where rapid glacier retreat, complex geomorphology, and dynamic climatic conditions converge. The Eastern Himalaya, particularly the state of Arunachal Pradesh and its surrounding areas, remains one of the most understudied yet highly vulnerable regions to GLOFs. Although it reports a GLOF hazard rate nearly three times higher than other Himalayan regions, comprehensive scientific investigations are lacking (Veh et al., 2020). This gap is particularly concerning given that a long-term glacial lake inventory (1970–2013) revealed a dramatic rise in both the number and area of glacial lakes–showing up to a 1,000% increase in area in certain zones—and identified 61 lakes as critical or potentially hazardous (Thomas, 2023). Supporting this, another detailed spatial inventory found that 82.8% of the region's glacial lakes are erosion-type, which are generally more unstable and prone to outbursts (Mal et al., 2020). Furthermore, a sector-wide assessment of the Indian Himalayan Region (IHR) ranked Arunachal Pradesh second only to Jammu and Kashmir (J&K) in the number of very high and high danger lakes, with 388 identified, compared to 556 in J&K and 219 in Sikkim (Mal et al., 2021). The limited and mostly regional nature of existing studies, coupled with significant data scarcity, underscores the urgent need for focused, large-scale research in this region.

In addition to the minimum number of studies, the Himalayan region is characterised by seismic activity, highly undulating terrain, limited accessibility, and scarce data from experts. These factors underscore the urgent requirement for modern, data-efficient methodologies to evaluate GLOF hazards (Afroz A Shah, 2013; Shah et al., 2025; Shah et al., 2024a; Shah et al., 2024b). Such approaches necessitate the integration of earthquake hazards as potential drivers of GLOF events. Though field observations are essential as ground–truth, benchmarking, and validation, they are insufficient to match the requirements of measuring the rate and scale of environmental changes for a large study area. Modern remote sensing technologies, complemented by advanced ML methods, now allow lake prediction, hazard assessment, and risk assessment at local and regional scales and in near-real time, offering a pathway toward more effective disaster risk reduction in vulnerable mountain regions. However, many existing GLOF risk assessment studies in the same regions also (e.g. (Rounce et al., 2016; Rounce et al., 2017; T. Zhang et al., 2022b)) adopt a relatively narrow scope, concentrating primarily on glaciological and hydrological parameters while underrepresenting other critical triggers and amplifiers. For example, T. Zhang et al. (2022b) assessed lake risk using only five parameters and omitted key seismic and geomorphic drivers such as earthquakes, active faults, and land cover (Afroz Ahmad Shah, 2013). Similarly, Rounce et al. (2016), and Rounce et al. (2017) avoided accounting for topographic characteristics (slope, elevation), and tectonic activity. Although Kougkoulos et al. (2018) integrated thirteen parameters, including certain seismic indicators, their framework excluded landslides and proximity to faults–factors known to substantially influence GLOF initiation. As a result, the combined influence of earthquakes, active faults, landslides, and historical GLOF events remains insufficiently explored, limiting the comprehensiveness and predictive strength of current models. The above-mentioned disparities demand the critical need for a GLOF-related comprehensive study in the Eastern Himalayan region, especially for study areas located in Arunachal Pradesh and nearby areas.

In view of the above, this thesis considers rapid changes in glacier dynamics serving as critical indicators of emerging hazards, particularly GLOFs, which pose significant threats to downstream communities. The development, stability, and potential failure of glacial lakes are influenced by multiple interacting and interconnected factors. Among the three categories of factors, (i) climate change drivers such as rising temperatures and precipitation variability, (ii) anthropogenic influences, geomorphological features and topographic characteristics, and (iii) seismic activity (earthquake, tectonic faults, landslides) are prominent and play important roles. The interplay of these multiscale factors determines lake evolution and outburst potential, yet

several remain underrepresented or insufficiently integrated in current hazard and risk assessment studies. Consequently, given the accelerating impacts of climate change – manifested in the expansion of many glacial lakes and the rising likelihood of GLOFs – there is a need for ML-based multiple frameworks that allow prediction of lake formation, hazard assessment, and risk assessment in an innovative, accurate, and scalable manner. Moreover, the developed ML-based frameworks should support robust, scalable, and adaptive early-warning systems in data-scarce high-mountain environments like the Eastern Himalayas. This thesis proposes to develop ML-based GLOF risk assessment frameworks for the Eastern Himalayas. Objectives of the thesis are as follows:

1. Identifying the critical features influencing the glacial lake formation using ML-based mathematical models.
2. To predict the likelihood of glacial lake formation using local neighbourhood (geomorphic and topographic features) influence.
3. To evaluate the hazard level associated with existing glacial lakes under the influence of geomorphology (specifically erosional features) using deep learning and a heuristic framework.
4. To determine risk levels associated with glacial lakes in a tectonically active Eastern Himalayan region through a hybrid weight of influencing factors obtained from the fusion of both data-driven and expert opinion-based approaches.
5. To predict the comprehensive GLOF risk (probabilistic) under the climate change conditions and other variables included in objective 4.

2.5 Organisation of Thesis

This thesis is structured into eight chapters to provide a coherent progression from conceptual foundations to applied analyses and final synthesis. Chapter 1 introduces the research problem, rationale, and scope, situating the study within the broader context of cryospheric hazards and risk management. Chapter 2 presents a critical review of the existing literature, consolidating current knowledge, identifying research gaps, and defining the specific objectives addressed in this work. Chapter 3 outlines the overall methodology, including the theoretical framework, data sources, analytical tools, and computational approaches employed.

Chapter 4 focuses on the identification of key governing features influencing glacial lake formation. In this chapter, future glacial lake development is also predicted using machine learning (ML) techniques, integrating terrain and geomorphic parameters to enhance predictive accuracy. Chapter 5 presents a hazard assessment of existing glacial lakes, employing a hybrid methodological framework that combines ML models with heuristic evaluation to capture both data-driven and expert-informed perspectives. Chapter 6 extends this analysis to deterministic risk assessment through Multi-Criteria Decision Analysis (MCDA), where integrated expert judgment and quantitative data-driven weights are applied to prioritize lakes based on their potential hazard levels.

Building on these findings, Chapter 7 presents a comprehensive probabilistic risk assessment of Glacial Lake Outburst Floods (GLOFs). This chapter incorporates uncertainty quantification in predictive models, thereby providing a robust framework for probabilistic risk estimation. Finally, Chapter 8 synthesizes the overall findings, draws key conclusions, and highlights the broader implications of this research. It also outlines future research directions, emphasizing methodological advancements and applied perspectives for enhancing hazard monitoring, risk assessment, and disaster preparedness in high-altitude regions.

Chapter 3

Methodology and Study Area

3.1 Introduction

The first stage focuses on the identification and visualization of critical features associated with glacial lake formation in the region of interest (ROI). Erosional and topographic features influencing local lake development are systematically selected and analyzed. These identified features serve as predictors or input variables for machine learning (ML) models—specifically Logistic Regression (LR), Artificial Neural Networks (ANN), and Bayesian Neural Networks (BNN)—to estimate the Probability of Glacial Lake Formation (PGLF). Comparative evaluation of model performance enables the selection of the most suitable predictive framework for mapping potential lake formation sites. In parallel, the relative contribution of each input feature is quantified, thereby isolating the most influential drivers of glacial lake formation.

The second stage builds directly on these findings by addressing the hazard potential of existing glacial lakes within the ROI defined in Stage 1. Advanced feature extraction and pattern recognition techniques are employed, with Convolutional Neural Networks (CNNs) applied to remote sensing datasets to capture the spatial coherence between existing lakes and geomorphic features such as erosional landforms. This analysis highlights dynamic processes, including lake expansion, shrinkage, and the emergence of new lakes. The extracted spatial patterns are then incorporated into a heuristic rule-based framework that integrates geomorphic indicators, topographic characteristics, and geometric attributes of lakes to estimate hazard potential. This stage results in the classification of existing glacial lakes into six hazard categories: Non-Dangerous (ND), Low-Dangerous (LD), Moderately-Dangerous (MD), Highly-Dangerous (HD), Very Highly-Dangerous (VHD), and Extremely Dangerous (ED). Simultaneously, potential sites of future lake formation are identified and mapped.

The third stage advances the analysis toward a deterministic and regionally comprehensive risk assessment framework. Here, the scope of the ROI is expanded to incorporate multi-hazard influences, extending beyond local geomorphic factors. This stage accounts for regional drivers such as historical Glacial Lake Outburst Floods (GLOFs), landslide activity, seismicity (faults, landslides, and earthquakes), and anthropogenic exposure. A hybrid weighting scheme–

combining expert-based Analytic Hierarchy Process (AHP) with the objective SE method–is applied to assign relative importance to these variables. The weighted variables are then integrated using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and Fuzzy C-Means (FCM) clustering, resulting in the classification of glacial lakes into four discrete risk categories: Low Risk (LR), Moderate Risk (MR), High Risk (HR), and Very High Risk (VHR). This stage marks a conceptual transition from localized hazard evaluation to regional-scale risk synthesis.

The final stage integrates the findings from all previous stages into a comprehensive, uncertainty-aware probabilistic risk assessment model. Critical geomorphic, topographic, and geometric features identified in Stages 1 and 2, together with anthropogenic and multi-hazard variables derived from Stage 3, are combined with climate-related parameters, including precipitation, temperature, and glacier proximity. These combined predictors account for both geophysical and climatic forcing mechanisms influencing lake stability and outburst potential. At the core of this stage is a BNN model, selected for its superior performance in Stage 1, which is trained using risk class labels generated through the composite MCDA framework developed in Stage 3. The outputs of this stage include both a risk classification of glacial lakes into four levels—Low Risk (LR), Moderate Risk (MR), High Risk (HR), and Very High Risk (VHR)—and an associated uncertainty map. The uncertainty mapping enhances interpretability

and provides critical insights into the robustness and confidence of the classification, thereby strengthening its utility for decision-support and risk-informed policy formulation.

The overall methodology adopted in this thesis is summarized in Figure 3.1, which outlines the four-stage framework for GLOF risk assessment in the Eastern Himalaya using advanced ML-based approaches.

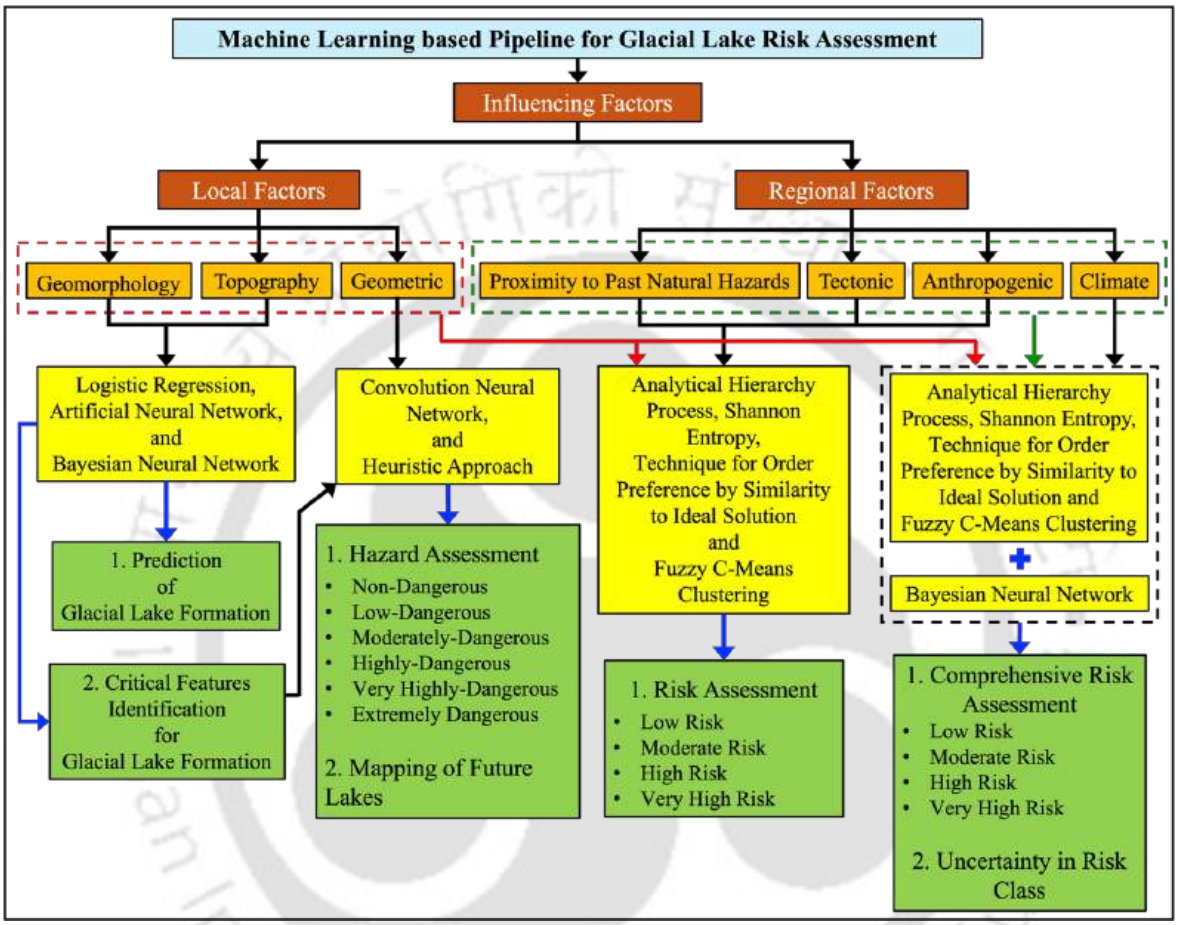


Figure 3.1: ML-based integrated pipeline for Glacial Lake Risk Assessment in the Eastern Himalaya.

The framework incorporates influencing factors categorised into local lake-related factors (geomorphology, topography, geometric attributes) and regional factors (historical hazard proximity, tectonic activity, anthropogenic influences, and climate variables). The pipeline progresses through four major stages for the four objectives of the thesis: (i) prediction of glacial lake formation probability using LR, ANN, and BNN with critical feature identification, (ii) hazard assessment of existing glacial lakes via CNN and heuristic approaches, classifying lakes into hazard levels from ND to ED, (iii) deterministic risk assessment using hybrid MCDA (AHP, SE, TOPSIS, and FCM), producing risk classes from LR to VHR, and (iv) comprehensive probabilistic risk assessment with uncertainty estimation using BNNs, integrating climate and exposure information for robust decision support.

3.2 Study Area

In this thesis, two distinct study areas are considered: a Region of Interest (ROI) representing a local or regional scale, and an Extended-ROI that expands the analysis to incorporate broader, global influences. The work begins with the first two objectives in Stage 1 and Stage 2, focusing on the initial ROI, where the emphasis is on geomorphological and topographical factors within a relatively confined geographic area. For the next two objectives achieved in Stages 3 and 4, the study area broadens to the Extended-ROI, allowing for the integration of additional variables such as geological structures, geomorphic features, tectonic activity, and anthropogenic influences. The ROI used in Stages 1 and 2 is presented in Section 3.2.1, while the Extended-ROI is explained in Section 3.2.2.

3.2.1 Region of Interest (ROI)

The study is performed over the Eastern Himalayan site, having an approximate area of 1,17,796 km^2, spanning across $28°28'N$ to $31°13'N$ latitudes and $92°55'E$ to $98°17'E$ longitudes located in Arunachal Pradesh (India) and some parts of Tibet. The study area is reported to have temperate weather conditions. Also, elevation values of glaciers vary from 4000 to 7600 meters above mean sea level (Bajracharya et al., 2014). The precipitation patterns and glacier dynamics in the study area are primarily influenced by atmospheric moisture flux from the Indian monsoon and, to a lesser extent, the East Asian monsoon, with transitional zones between these systems playing a key role in regulating precipitation variability, and glacier accumulation-melt processes (Yao et al., 2018). In recent decades, glaciers on the site have undergone rapid retreat, increasing the number of glacial lakes (Gardelle et al., 2013; Ojha et al., 2016) formed in the depressions caused by both glacier retreat and erosion. Moreover, the region exhibits diverse landforms that vary with elevation – cirques and retreating glaciers are observed at higher elevations than the other three features (valleys, flow channels, and glacial lakes). The glaciers feed several perennial river streams, which flow through mountainous regions and eventually join the Brahmaputra, supporting agriculture, hydropower generation, and community livelihoods. Figure 3.2 shows the ROI with Google Earth image and elevation profile obtained by using the global Shuttle Radar Topography Mission (SRTM) 30 m (1 Arc Second) digital elevation model (DEM).

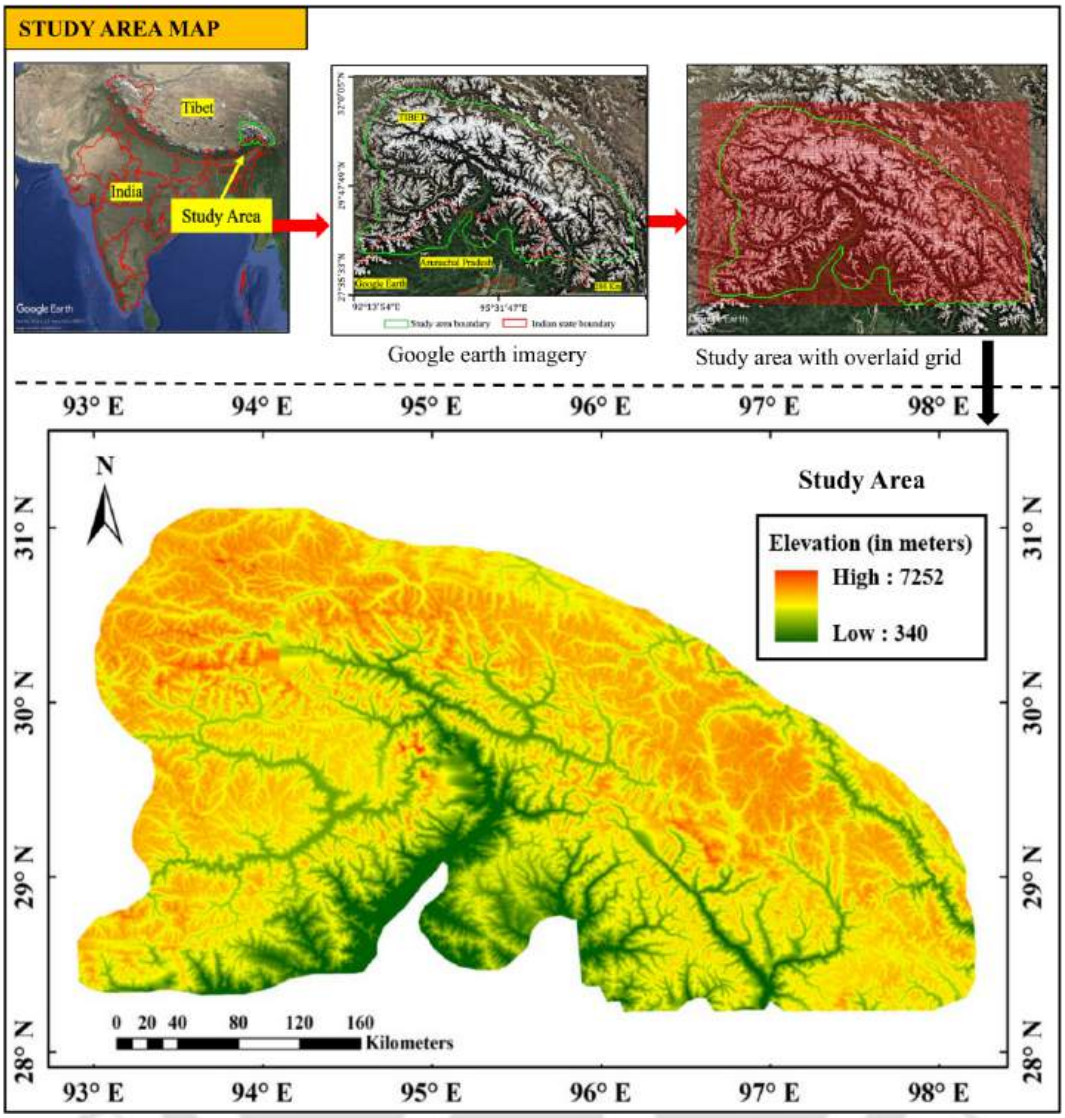


Figure 3.2: Location and topographic overview of the study area in the Eastern Himalaya, covering parts of Arunachal Pradesh (India) and adjacent regions of Tibet. The top panels illustrate the regional context, the specific study boundary, and the grid overlay used for spatial analysis. The bottom panel shows a high-resolution digital elevation model (DEM) of the study area, with elevation ranging from 340 to 7,252 meters. The region exhibits complex mountainous terrain, extensive valley networks, and high-altitude zones where glacier retreat and glacial lake formation processes are active.

The region exhibits complex mountainous terrain, extensive valley networks, and high-altitude zones where glacier retreat and glacial lake formation processes are active. Google Earth images of the ROI are collected from the years 2018 to 2022 by overlaying a grid over the ROI. However, if image scenes captured during the years 2018 to 2022 were obscured by cloud cover or distortion, additional image data from 2015 to 2018 was incorporated to ensure comprehensive coverage. Each image tile within a grid maintained a resolution of 1024 pixels × 768 pixels, corresponding to a size of 3.5 km × 2.63 km (approximately 11m^2 resolution) on ground. This resulted in a dataset comprising 12,924 images for the entire ROI.

3.2.2 Extended-ROI

The extended region of interest (Extended-ROI) is a larger spatial extent that encompasses the ROI and is studied in Chapters 6 and 7 for risk assessment studies. The Extended-ROI lies within the Eastern Himalaya and spans parts of northeastern India, Bhutan, Nepal, northern Myanmar, and the southeastern Tibetan Plateau. This transboundary region is characterized by pronounced elevation gradients and active tectonic systems, including the Main Central Thrust (MCT), Main Boundary Thrust (MBT), and Main Frontal Thrust (MFT), and falls predominantly within Seismic Zone V–India's highest seismic hazard classification. Influenced by a monsoon-dominated climate and subject to accelerated cryospheric changes, the region exhibits rapid glacial retreat and increasing glacial lake activity. As the source area for several major transboundary rivers, it holds critical importance for downstream water security and is highly vulnerable to GLOF-related hazards. The Extended-ROI allowed a comprehensive GLOF risk assessment beyond hazard evaluation to include broader regional influences. Specifically, the extension allows for the integration of key drivers such as tectonic activity, climatic variability, and anthropogenic exposure, which are critical for understanding the spatial heterogeneity of GLOF risk assessment. A detailed discussion about the Extended-ROI is provided in Chapter 6 for reference.

After presenting the methodological overview in this chapter, the thesis advances to address the stage 1 objective in Chapter 4. Subsequently, the objectives of stages 2, 3, and 4 are discussed in detail in Chapter 5, Chapter 6 and Chapter 7, respectively.

Chapter 4

Prediction of Glacial Lake Formation

4.1 Introduction

Glacial erosional features are geomorphic evidence of the glacial processes shaping the landscape and hydrology of a site. A glacial lake forms when flow channels direct glacier meltwater or precipitation into lower-elevation depressions. These depressions may evolve into cirque lakes or valley lakes depending on their geomorphic setting. Consequently, glacial lakes appear in various morphologies and spatial configurations, typically occupying over-deepened basins formed by glacial erosion or subsequent ice retreat. Their irregular or elongated shapes reflect intense glacial sculpting, and their presence highlights regions of significant ice mass loss and topographic modification.

Cirques are steep-walled, amphitheater-like hollows located near the origins of former glaciers. These high-elevation features act as natural accumulation basins for snow and ice, and upon retreat, often give rise to small tarns or nascent glacial lakes. At lower elevations, U-shaped valleys dominate the landscape and represent the primary conduits for meltwater and sediment transport, linking high-altitude glacierized zones with downstream river systems and lowland catchments.

Interpretations drawn from satellite observations also highlight active glacial retreat and related hydrological responses. Retreating glaciers, as evidenced by exposed bedrock and thinning ice tongues, indicate the rapid pace of ice loss driven by regional climate warming and leading to the expansion of proglacial lakes. The flow channels observed in deglaciated and vegetated terrains – ranging from braided streams to incised river networks–demonstrate how meltwater reconfigures the landscape. In several areas, these channels traverse steep gradients and narrow valleys, signaling the potential for high-energy hydrological events such as debris flows, sediment surges, and GLOFs. Sediment transport further modifies lake characteristics, impacting their volume, turbidity, and potential instability. Consequently, the geomorphic configuration of the landscape–including the spatial arrangement and interaction of cirques, valleys, flow channels, and retreating glaciers—regulates water accumulation, drainage pathways, and sediment transport. These processes not only establish the likely locations for the formation of new glacial lakes but also influence the transformation of pre-existing lakes.

Recent efforts to predict glacial lake formation have utilised geomorphological, statistical, or machine learning (ML) approaches, with varying success. Physically based geomorphological methods, such as those employed by Furian et al. (2021)Furian, Loibl14, identify potential lake sites by reconstructing bedrock depressions beneath retreating glaciers. However, these approaches are constrained by size thresholds (area greater than 0.1 km^2) and often exclude smaller, yet hazardous, lakes. Morphological scoring models (e.g., (T. Zhang et al., 2022a))incorporate expert-assigned weights for features such as slope, crevasses, and glacier proximity but are subject to subjective biases and exclude critical geomorphological variables or features such as cirques and flow channels.

ML-based approaches offer improved capacity to capture complex relationships among influencing factors. Mohanty et al. (2022) applied logistic regression (LR) models incorporating climate, topography, and glacier features. However, only 50% of the features used were statistically significant (p-value < 0.05) for lake formation. The insignificance of the remaining features could be attributed to factors such as multicollinearity, weak association with the target variable, or limited data variability. Moreover, LR assumes linear relationships that inadequately capture the inherently non-linear and dynamic processes governing lake formation.

Artificial Neural Networks (ANNs) have demonstrated advantages over linear models by effectively capturing non-linear dependencies and complex interactions in glaciological applications, including glacier boundary detection, meltwater modelling, snow cover forecasting, and ice thickness estimation (Baumhoer et al., 2019; Hou et al., 2014; Kim et al., 2016; Mohajerani, 2019). However, both ANN and traditional statistical regression models (e.g. LR in (Mohanty et al., 2022) study) lack inherent mechanisms to quantify prediction uncertainty or capture correlations among influencing factors, which can lead to biased or unreliable outcomes. In hazard-sensitive scenarios, such as glacial lake formation, this uncertainty may result in either unnecessary resource allocation due to false positives or unanticipated catastrophic events due to false negatives.

Bayesian Neural Networks (BNNs) address these challenges by integrating Bayesian probability principles with the structural advantages of neural networks. BNNs not only address uncertainty inherent in environmental data and predictions but also effectively capture correlations among a broad range of input features in glaciology. Additionally, it uses small amounts of field data combined with Bayesian inference to train a network, which helps prevent

overfitting – a common problem with more 'data-hungry' models (e.g. (Gopalan et al., 2018; Guillet et al., 2023; Jospin et al., 2022)). Therefore, this study presents a novel, data-driven framework for probabilistic prediction of glacial lake formation in the Eastern Himalaya using erosional and topographic features within a BNN modelling approach. To our knowledge, this represents the first application of BNNs for forecasting glacial lake occurrence, explicitly integrating geomorphological controls alongside established topographic variables. Comparative assessments are conducted using LR, ANN, and BNN models, with predictive performance evaluated through receiver operating characteristic (ROC) curves and uncertainty quantification metrics. Our results identify critical geomorphological and topographical features influencing glacial lake formation and demonstrate the superior predictive accuracy and uncertainty management capabilities of the BNN approach. The probabilistic risk maps generated herein provide a valuable tool for hazard mitigation, infrastructure planning, and sustainable water resource management in GLOF-prone regions. On a broader scale, this methodology offers a transferable framework for application in other glaciated mountain systems globally, contributing to improved disaster risk reduction, early warning systems, and climate adaptation strategies. This study is performed over the study area illustrated by Figure 3.2 in the previous chapter.

4.2 Methodology and Materials

The proposed framework adopts a data-driven approach to predict the probability of glacial lake formation (PGLF) by integrating geomorphological and topographic indicators within machine-learning models. High-resolution Google Earth imagery and the 30 m SRTM DEM were used to delineate the features in the study area and extract relevant spatial information. To enable systematic data extraction, the study area was subdivided into a square grid. Google Earth imagery was analysed using a grid-based visualisation format in which each tile (1024 × 768 pixels) represented an area of approximately 3.5 km × 2.63 km. This scale was well suited for identifying macroscale geomorphic features relevant to glacial lake formation—such as cirques, valleys, flow channels, retreating glaciers, and neighbouring lakes—which cannot be interpreted reliably at smaller viewing extents. In contrast, local-scale topographic variables including slope, elevation, profile curvature, and plan curvature, which influence the formation of depressions or cavities and large features, were derived from the SRTM DEM. Together, these datasets provide complementary information: Google Earth imagery captures large-scale geomorphology, while the DEM represents microscale surface characteristics. Glacial lake presence was then mapped to classify each grid cell as either 'lake' or 'no lake'. The flowchart,

representing the complete process of obtaining the likelihood maps of glacial lake formation, is shown in Figure 4.1.

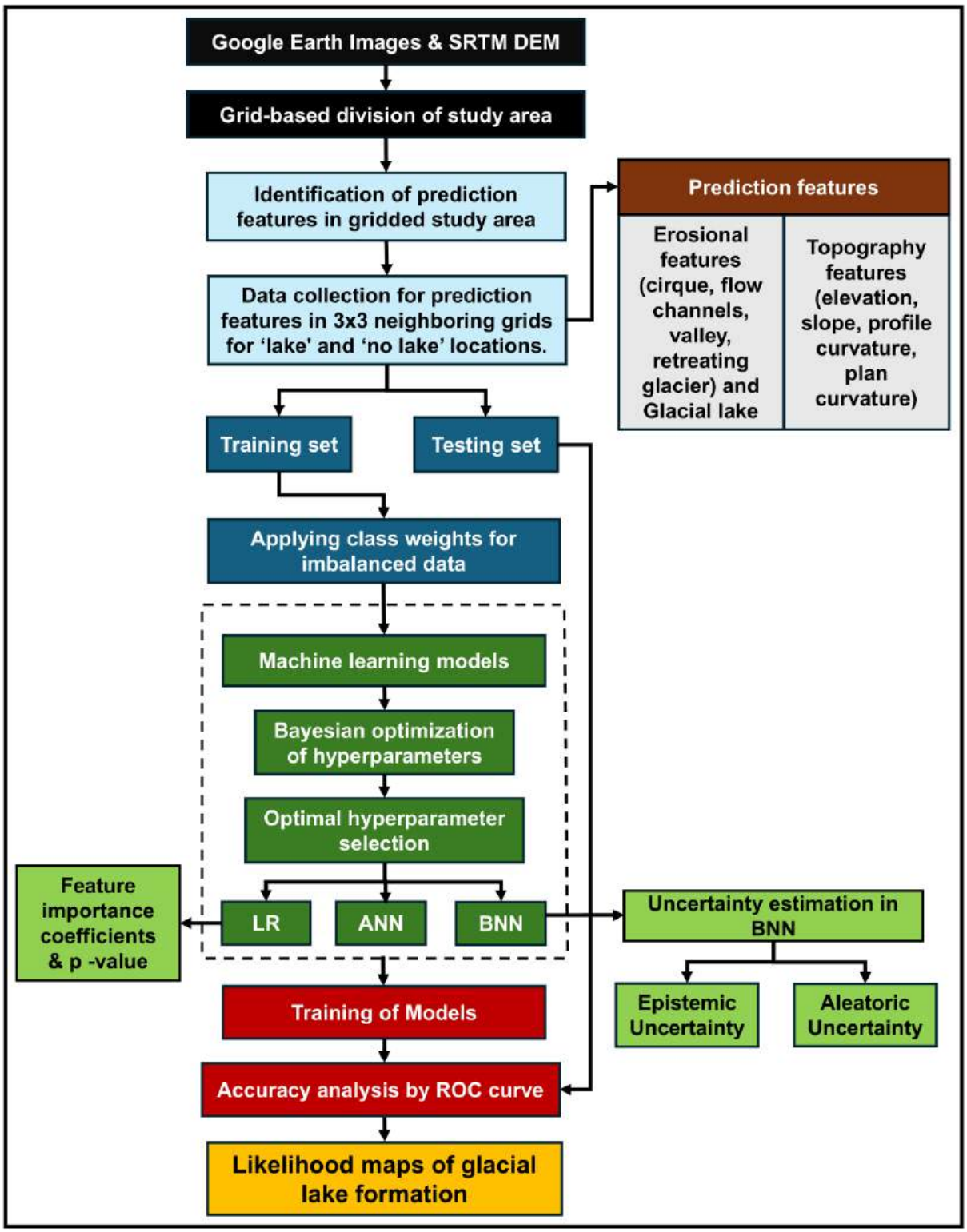


Figure 4.1: Flowchart showing the methodology to model glacial lake formation.

4.2.1 Data Preparation

For the three techniques (LR, ANN, and BNN), input and output data are systematically prepared by measuring the occurrence of erosional features (cirques, valleys, retreating glaciers, flow channels), glacial lakes, and topographical features (elevation, slope, profile curvature, and plan curvature) in each of the 12,924 images. A neighbourhood of 3x3 grids was considered to extract feature values for each grid associated with glacial lake location or non-lake location, forming the primary dataset for model development. Descriptions of erosional prediction features are as follows (Dar et al., 2017; Evans et al., 1974; Flint, 1971; Yongping, 2004):

1. Glacial lakes (GLs) are uniformly textured water bodies of various sizes and forms, ranging in colour from deep blue to green or brown. They display differences in depth, sediment content, and sunlight exposure. This study considers all types of lakes, regardless of their dam conditions, such as bedrock-dammed, ice-dammed, moraine-dammed, and other erosion lakes.
2. For a glacial lake, a neighbouring lake situated at a higher elevation is defined as an upstream lake, whereas a lake at a lower elevation is considered downstream. However, it is less likely that two adjacent lakes share the same elevation. When lakes are at the same elevation, they form due to the partitioning of inflow at that level. Therefore, in this study, lakes at the same elevation are regarded as part of the downstream lake.
3. Cirques (C) are usually identified by a bowl or armchair shape, with a steep headwall, a gently inclined or over-deepened floor, and a convex slope break marking the transition from the cirque to the lower valley.
4. Valleys (V) are broad, elongated U-shaped depressions with gently rising slopes as you move upwards, located among hills and often crossed by meandering or braided flow channels.
5. A retreating glacier (RG) is characterised by rough surfaces, crevasses, striations, and the emergence of rocky terrain and moraines, formed by glacier melting and glacio-fluvial erosion.
6. Flow channels (FC), which serve as pathways for water across glaciers, can have either a smooth or rough texture depending on whether they are filled with water. Additionally, their colour is affected by sediment load, while their flow pattern (such as braided or meandering) is determined by the surrounding topography.

Figure 4.2 and Figure 4.3 show typical examples of feature variables associated with geomorphology.

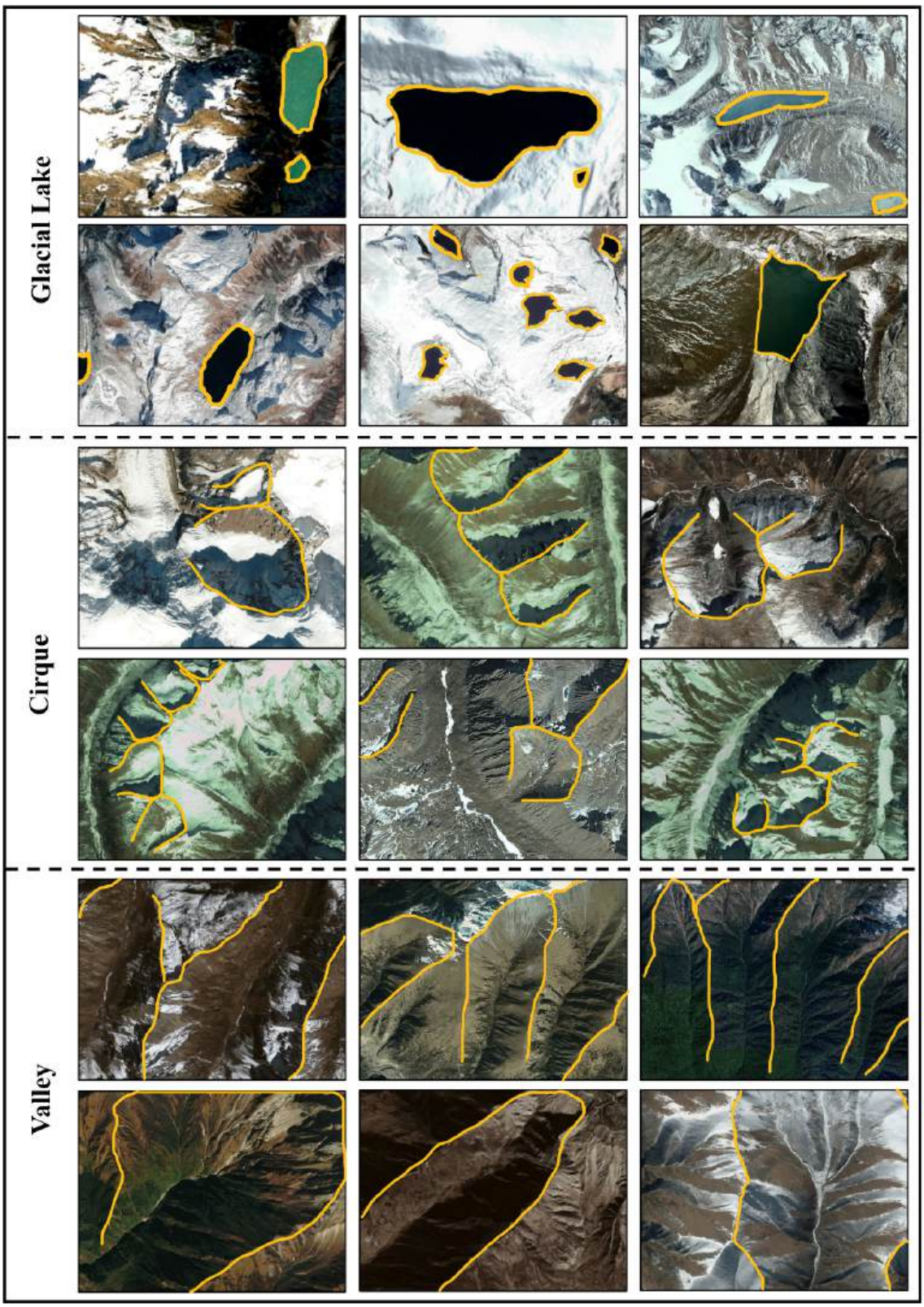


Figure 4.2: Representative examples of mapped glacial erosional features in the Eastern Himalaya based on high-resolution Google satellite imagery. Yellow lines delineate the boundaries of (top) glacial lakes, (middle) cirques, and (bottom) glacial valleys. These

features govern the spatial occurrence, expansion, and hydrological connectivity of glacial lakes, influencing downstream flood hazards and landscape evolution. The systematic mapping of these features forms a critical input for the probabilistic prediction of glacial lake formation..

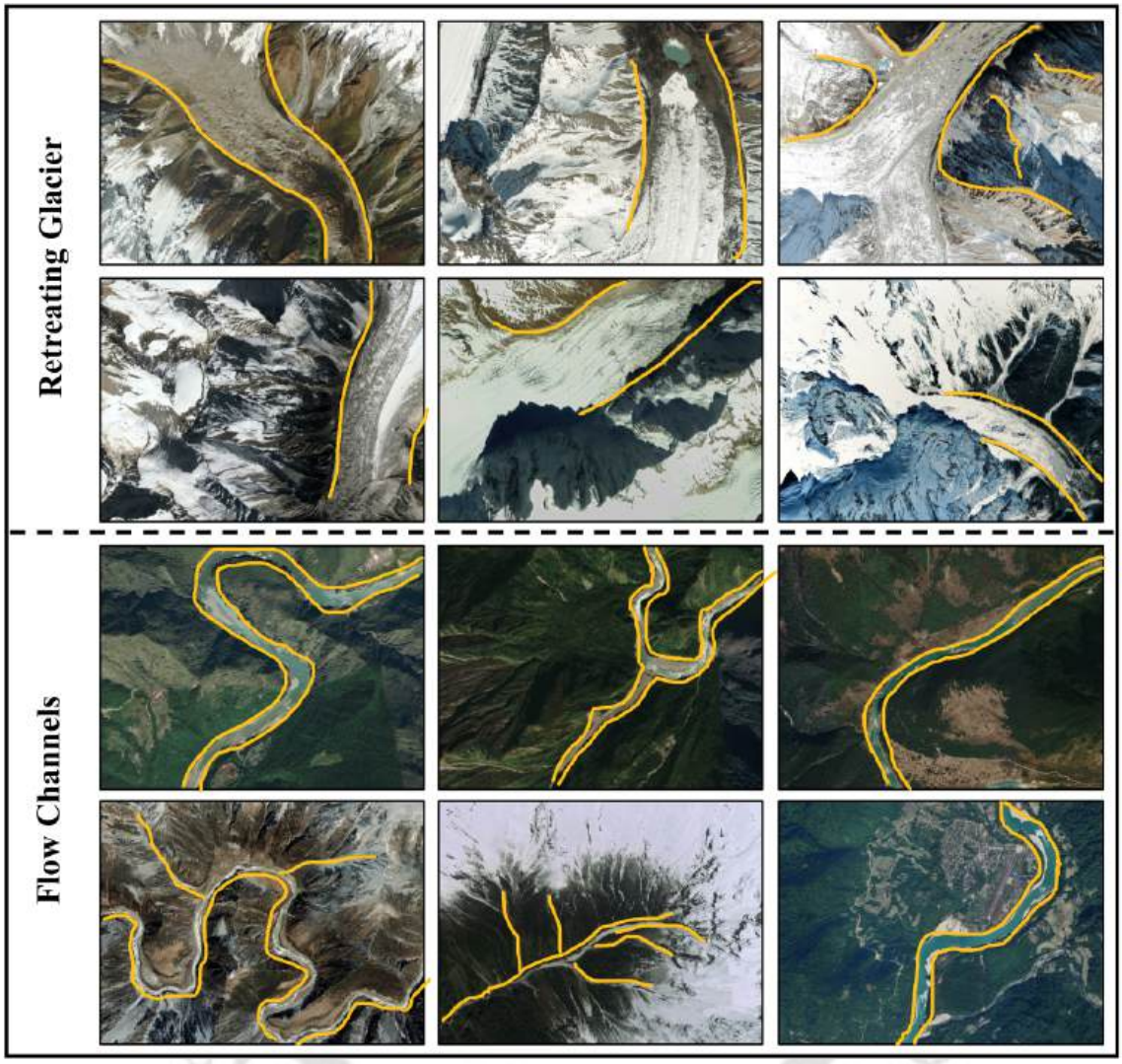
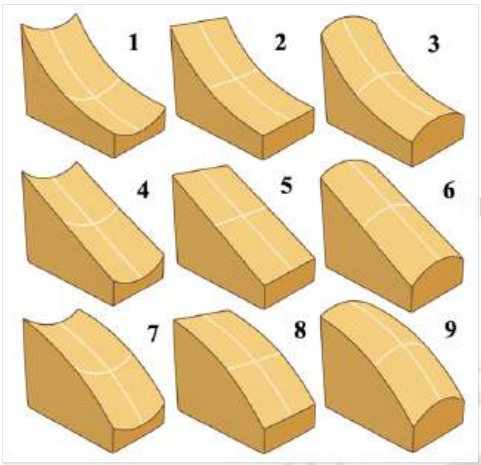


Figure 4.3: Mapped glacial erosional features in the Eastern Himalaya are shown using high-resolution Google satellite images. Visual examples of retreating glaciers (top two rows) and associated flow channels (bottom two rows) in a high mountain region. Yellow lines delineate glacier margins and hydrological flow paths. Retreating glacier images show ice thinning, exposed rock surfaces, and the formation of proglacial lakes, indicative of significant ice loss. Flow channel images illustrate the evolution of drainage networks in response to glacial recession. These visual observations provide contextual support for identifying terrain susceptible to glacial lake formation and hydrological transformation.

In Figure 4.2, yellow lines delineate the boundaries of glacial lakes (upper row), cirques (middle row), and glacial valleys (lower row). Whereas, in Figure 4.3, yellow lines delineate glacier margins and hydrological flow paths. Retreating glacier (RG) images show ice thinning, exposed rock surfaces, and the formation of proglacial lakes, indicative of significant ice loss. Flow channels in images illustrate the evolution of drainage networks in response to glacial

recession. These features govern the spatial occurrence, expansion, and hydrological connectivity of glacial lakes, influencing downstream flood hazards and landscape evolution. Thus, visual observations provide contextual support for identifying terrain susceptible to glacial lake formation and hydrological transformation. Systematic mapping of these geomorphological features forms a critical input for the probabilistic prediction of glacial lake formation.

On the other hand, topography variations influence glacial meltwater dynamics and essentially govern the process of lake formation. Feature variables associated with topography variations, namely slope, elevation, profile curvature and plan curvatures for each grid are estimated using the global SRTM DEM in ArcMap 10.8. Slope influences speed of meltwater flow– steeper slopes cause faster runoff and higher erosion rates, increasing the risk of sediment transport to lower valleys and vice versa. Elevation significantly influences melt rates, as higher elevations retain colder temperatures, slowing the melting of ice. In contrast, lower elevations are warmer, resulting in increased meltwater flow and a higher likelihood of lake formation and downstream impacts. Profile curvature measures the curvature along the direction of maximum slope, which affects how glacial meltwater flows down slopes. Areas with positive profile curvature (upwardly concave, cup-shaped or depression) accelerate flow speed, leading to erosion, while areas with negative profile curvature (upwardly convex, cap-shaped or ridge) decelerate the meltwater movement, encouraging the deposition of sediments. Plan curvature, measuring curvature perpendicular to the slope, affects how meltwater spreads across the terrain. Positive plan curvature (sidewardly convex) areas cause divergent flow, dispersing meltwater and reducing saturation. Conversely, negative plan curvature (sidewardly concave) areas lead to convergent flow, where meltwater collects and increases the potential for ponding or soil saturation (Noël et al., 2014). Notably, if the profile or plan curvature is zero, it represents the linear surface along the associated direction. Accordingly, Figure 4.4 This paper presents the nine different combinations of profile and plan curvature that occur in the hillslope and influence the flow pattern and flow direction.



Hillslope form	Profile Curvature	Plan curvature
1	Concave (+); Accelearated	Concave (-) ; Convergence
2	Concave (+); Accelearated	Linear (0); Uniform
3	Concave (+); Accelearated	Convex (+) ; Divergence
4	Linear (0); Uniform	Concave (-) ; Convergence
5	Linear (0); Uniform)	Linear (0); Uniform
6	Linear (0); Uniform)	Convex (+) ; Divergence
7	Convex (-); Decelerated	Concave (-) ; Convergence
8	Convex (-); Decelerated	Linear (0); Uniform
9	Convex (-); Decelerated	Convex (+) ; Divergence

Figure 4.4: Representation of different combinations of plan curvature and profile curvature of the hillslope.

4.2.1.1 Feature Diagnostics and Multicollinearity Assessment

To ensure uniform scale and stability during model training, all topographic features were standardised to zero mean and unit variance prior to analysis. Simultaneously, erosional features were encoded as binary indicators (1 for presence and 0 for absence). Further, the presence of multicollinearity among predictors was assessed using the variance inflation factor (VIF). Results are summarized in Table 4.1 below.

Table 4.1: Multicollinearity check using VIF value

Features	VIF value
Neighbourhood Lakes	1.5400
Cirque	1.0230
Flow channels	1.2040
Retreating glacier	1.2710
Valley	1.1230
Elevation	0.5130
Slope	0.6460
Profile curvature	2.1580
Plan curvature	2.0720

All features exhibit VIF values below the conservative threshold of 5, confirming the absence of multicollinearity within the dataset. Specifically, neighbourhood lakes (VIF = 1.54), cirques (1.02), flow channels (1.20), retreating glaciers (1.27), and valleys (1.12) show lower VIF values, indicating minimal interdependence while retaining their independent contributions to glacial lake formation modelling. Among the topographic variables, elevation (0.51) and slope (0.65) demonstrate particularly lower VIFs, reflecting their orthogonality and

numerical stability following standardisation. Curvature metrics – profile curvature (VIF = 2.16) and plan curvature (VIF = 2.07) – present higher VIF values, yet their values remain within accepted limits, indicating moderate associations with other features without introducing redundancy or instability into the models. Moreover, the absence of multi-collinearity among features strengthens the interpretability and statistical robustness of the models, ensuring that parameter estimates are stable and free from variance inflation distortions. These results confirm that all selected features contribute as non-redundant, independent, and uncorrelated information to the modelling framework for the prediction of PGLF.

4.2.2 Logistic Regression

LR is a multivariate nonlinear statistical analysis method suitable for binary categorical problems with multiple control features expressing relationships between an output or dependent feature and independent (or explanatory) features (Aloisio et al., 2023; Feby et al., 2020; He et al., 2022; Witteveen et al., 2018; Zhong et al., 2024). In this study, the presence of a lake is an output feature or dependent feature (where 1 indicates the presence of a lake in the grid, 0 represents the absence of a lake in the grid). Further, independent or explanatory features can be either discrete or/continuous values depicting the influencing features of lake occurrence. The LR model uses a logistic function (also known as a sigmoid function), which predicts probability values of the output feature between 0 and 1. The LR model works by fitting a linear equation to the input features and applying the logistic function to the result of the linear equation. The linear equation defines log-odds (Y) as:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_9 x_9 \quad (1)$$

where, $x_1, x_2, \ldots, x_9$ The input features are in binary (1 for the presence of erosional features, else 0) and continuous form (here, it is a continuous value for topography features). In this chapter, nine influencing features or independent variables are used, such as the neighbouring lake, cirque, flow channel, retreating glacier, valley, elevation, slope, profile curvature, and plan curvature.

$\beta_1, \beta_2, \ldots, \beta_9 =$ Coefficients of the corresponding independent variable or influencing features as indicated from $x_1, to\ x_9$,

$\beta_0 =$ Intercept term

$Y =$ Log odds of outcome probability (logit function is the natural logarithm of the odds ratio).

The logistic function is applied to the log-odds (determined by equation 1) to convert them to probabilities $P(Y)$:

$$P(Y) = \frac{1}{1 + e^{-Y}} \tag{2}$$

where, P is the probability of a log-odds of a glacial lake occurrence under the presence of the five erosional features and four topography features.

4.2.3 Artificial Neural Network (ANN)

An ANN is a computational model inspired by the structure and function of the human brain (Aloisio et al., 2023; Kalimullah et al., 2023b). Figure 4.5 shows ANN architecture.

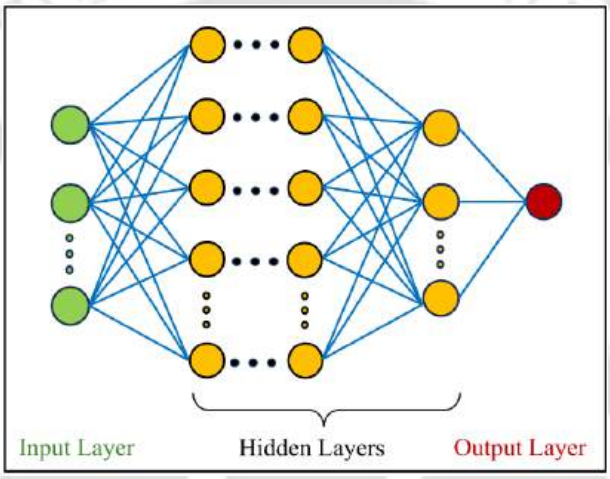


Figure 4.5: ANN architecture.

ANNs aim to determine an underlying function linking inputs and output features by processing a large number of data points through weighted connections and activation functions. The data from the layer of input features is transferred from one layer of interconnected nodes (or neurons) to the next, where each neuron processes the received information and passes it further to the following layer until the final output is generated. To explain an ANN mathematically, consider a dataset $\mathcal{D} = \{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{N}$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ and $\boldsymbol{y}_i \in \{c_1, c_2, \ldots, c_K\}$. Here, c_k represents a one-hot encoded vector, where only the k-th element is one and the others are zero. The number of different classes is denoted by K (two classes: lake and no-lake class), the sample size by N (12924 grids), and the dimension of input features by d (9 influencing features or input variables).

ANNs define model likelihood using a neural network with L hidden layers. For each layer $j \in \{1,2, \ldots, L+1\}$, let $\mathbf{h}^{(j)} \in \mathbb{R}^{d_j}$ be the pre-activated output, expressed as:

$$\mathbf{h}^{(j)} = f_{\boldsymbol{\theta}^{(j)}}\left(\mathbf{h}^{(j-1)}\right) = \boldsymbol{W}^{(j)}\phi\left(\mathbf{h}^{(j-1)}\right) + \boldsymbol{b}^{(j)} \tag{3}$$

where, $\boldsymbol{W}^{(j)} \in \mathbb{R}^{d_j \times d_{j-1}}$ is the weight matrix and $\boldsymbol{b}^{(j)} \in \mathbb{R}^{d_j}$ is the bias vector for the j-th layer, and $\phi(\cdot)$ denotes an element-wise non-linear activation function. By stacking these layers, the output of the ANN is:

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = f_{\boldsymbol{\theta}^{(L+1)}} \circ f_{\boldsymbol{\theta}^{(L)}} \circ \ldots \circ f_{\boldsymbol{\theta}^{(1)}}(\boldsymbol{x}) \tag{4}$$

where, $\boldsymbol{\theta} = \left[\left(\text{vec}\left(\boldsymbol{W}^{(1)}\right), \ldots, \text{vec}\left(\boldsymbol{W}^{(L+1)}\right)\right), \left(\boldsymbol{b}^{(1)}, \ldots, \boldsymbol{b}^{(L+1)}\right)\right]$ represents entire set of network parameters, and vec($\boldsymbol{W}$) denotes vectorization of matrix $\boldsymbol{W}$. Given the final K-dimensional output $f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \left(f_{\boldsymbol{\theta},1}(\boldsymbol{x}), \ldots, f_{\boldsymbol{\theta},K}(\boldsymbol{x})\right)$, the likelihood $p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ is written as:

$$p(\boldsymbol{y} = c_k|\boldsymbol{x}, \boldsymbol{\theta}) = p(\boldsymbol{y} = c_k \mid f_{\boldsymbol{\theta}}(\boldsymbol{x})) = \frac{\exp\left(f_{\boldsymbol{\theta},k}(\boldsymbol{x})\right)}{\sum_{j=1}^{K} \exp\left(f_{\boldsymbol{\theta},j}(\boldsymbol{x})\right)} \tag{5}$$

During training process, the ANN learns to adjust its parameters $\boldsymbol{\theta}$ by minimizing a chosen loss function, which quantifies difference between predicted outputs of the network and measured target outputs. For classification tasks, cross-entropy loss function is preferred, which is expressed as:

$$\mathcal{L}(\boldsymbol{\theta}) = -\sum_{j=1}^{K} c_j \log p\left(\boldsymbol{y} = c_j|\boldsymbol{x}, \boldsymbol{\theta}\right) \tag{6}$$

Training of an ANN involves backpropagation algorithm, which adjusts the network parameters iteratively by propagating error (calculated by loss function) backwards through the network layers. This process enables the network to learn from the error between predicted and labelled outputs to improve its performance over time. The following algorithm expresses functioning of ANN (De et al., 2020).

Algorithm – Training a ANN and probability estimation for a feature class

1: **procedure** Training an ANN for Classification Task

2: Initialize the learnable parameter $\boldsymbol{\theta}$ in the neural network.

3: **while** $\boldsymbol{\theta}$ does not converge **do**

4: # from training dataset $\mathcal{D}$ randomly sample a mini-batch $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{M}$.

5: # Compute cross-entropy loss.

$$\mathcal{L}(\boldsymbol{\theta}) \leftarrow -\frac{1}{M}\sum_{i=1}^{M}\left[\sum_{j=1}^{K} c_j \log p\left(\boldsymbol{y}_i = c_j | \boldsymbol{x}_i, \boldsymbol{\theta}\right)\right].$$

6: # Update $\boldsymbol{\theta}$ with gradient $\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta})$ and learning rate α.

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} - \alpha\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}).$$

7: **end while**

8: **procedure** Probability estimation of a Class

9: # let $\widehat{\boldsymbol{\theta}}$ be the optimised parameter after training the network.

10: # for a new input $\boldsymbol{x}^*$, estimate the probability of each class.

$$p\left(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}\right) = f_{\widehat{\boldsymbol{\theta}}}(\boldsymbol{x}^*).$$

4.2.4 Bayesian Neural Network (BNN)

BNN is a probabilistic approach to neural networks, quantifying uncertainty in predictions and model parameters. Unlike the conventional neural networks (ANN), which provide point estimates for predictions, BNNs provide a distribution of predictions, allowing for more robust decision-making in uncertain environments. Moreover, ANN frequently encounters overfitting during training due to unbounded weight values and fixed fully connected structures (Kalimullah et al., 2023a; Zhang et al., 1998).

To mitigate this issue and enhance generalization capability, the BNN model is developed for predicting the probability of the glacial lake formation ($p(\boldsymbol{y} = c_i)$) using the influencing features of lake occurrence as input features ($\boldsymbol{x}$) to the network. Here, binary classification is performed with classes c_1 and c_2 representing 'lake' and 'no-lake' classes. Similar to ANN, the BNNs consist of interconnected layers, including input, hidden, and output layers. However, in BNNs, each weight and bias parameter in the network is treated as a random variable with a prior distribution.

During training, BNNs learn the posterior distribution, capturing uncertainty in model parameters as well as uncertainty in data. A prior distribution $p(\boldsymbol{\theta})$ is placed on the parameter $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ in the Bayesian framework, which computes posterior distribution as:

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{D})} = \frac{\prod_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathcal{D})}, \tag{7}$$

and for given $p(\boldsymbol{\theta}|\mathcal{D})$ and a new input $\boldsymbol{x}^*$, predictive distribution for a new output $\boldsymbol{y}^*$ is given by:

$$p(\boldsymbol{y}^*|\boldsymbol{x}^*, \boldsymbol{\theta}) = \int_{\boldsymbol{\Theta}} p(\boldsymbol{y}^*|\boldsymbol{x}^*, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}) d\boldsymbol{\theta}. \tag{8}$$

Calculating the posterior $p(\boldsymbol{\theta}|\mathcal{D})$ is often intractable due to need for integrating over the entire parameter space $\boldsymbol{\Theta}$. To address this challenge, studies employ two primary methods for model learning: Markov Chain Monte Carlo (MCMC) and stochastic variational inference. MCMC methods proposed by Welling et al. (2011) and Chen et al. (2014) require significant memory for training deep networks, making them less practical in certain cases. Stochastic variational inference is more attractive as it approximates posterior distribution with a simpler and tractable distribution $q_{\boldsymbol{\psi}}(\boldsymbol{\theta})$, parameterized by $\boldsymbol{\psi}$. It finds the closest distribution to true posterior from a family $Q = \{q_{\boldsymbol{\psi}}(\boldsymbol{\theta}): \boldsymbol{\psi} \in \boldsymbol{\Psi}\}$, making it computationally feasible for large-scale models. This closeness is measured by the Kullback–Leibler (KL) divergence:

$$\mathrm{KL}\big(q_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \| p(\boldsymbol{\theta}|\mathcal{D})\big) = \int_{\boldsymbol{\Theta}} q_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \log \frac{q_{\boldsymbol{\psi}}(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathcal{D})} d\boldsymbol{\theta}. \tag{9}$$

Minimizing the KL divergence equates to minimizing evidence lower bound (ELBO):

$$\mathcal{L}(\boldsymbol{\psi}) = -\mathbb{E}_{q_{\boldsymbol{\psi}}(\boldsymbol{\theta})}[\log p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})] + \mathrm{KL}\Big(q_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \| p(\boldsymbol{\theta})\Big), \tag{10}$$

where, $\mathbb{E}[\cdot]$ is expectation operator.

In practice, it becomes:

$$\mathcal{L}(\boldsymbol{\psi}) \approx -\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} \log p(y_i|\boldsymbol{x_i}, \boldsymbol{\theta_t}) + \mathrm{KL}\Big(q_{\boldsymbol{\psi}}(\boldsymbol{\theta}) \| p(\boldsymbol{\theta})\Big), \tag{11}$$

where, $\boldsymbol{\theta_t} \sim q_{\boldsymbol{\psi}}(\boldsymbol{\theta})$ and T is the number of Monte Carlo samples.

For classification, the network's final layer outputs class (c) logits $z \in \mathbb{R}^k$, followed by a softmax transformation:

$$\mathrm{p}(\mathrm{y} = \mathrm{c} \mid \mathbf{x}, \boldsymbol{\theta}) = \frac{e^{z_c}}{\sum_{j=1}^{k} e^{z_j}}, \qquad c = 1,2, \ldots, k. \tag{12}$$

The softmax output yields normalized class probabilities, representing the model's estimated likelihood for each risk category. These probabilities are used during training to compute the categorical cross-entropy loss, which measures the divergence between the predicted probability distribution and the true class label. During inference, the same softmax outputs provide the predictive class probabilities for new inputs.

After the model hyperparameters optimization, predictions for a new input x^* are obtained by averaging over the T samples from the approximate posterior:

$$\hat{p}(\boldsymbol{y}^* \mid \boldsymbol{x}^*) = \int p(y^* \mid \boldsymbol{x}^*, \boldsymbol{\theta})\, q_{\psi^*}(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \frac{1}{T}\sum_{t=1}^{T} p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \boldsymbol{\theta}_t), \tag{13}$$

where $\boldsymbol{\theta}_t \sim q_{\psi^*}(\boldsymbol{\theta})$ and T is the number of Monte Carlo samples.

The next sub-section discusses methods for quantifying uncertainty in classification tasks using BNNs by the stochastic variational inference introduced by Gal et al. (2016) and further developed by Kendall et al. (2017).

4.2.4.1 Uncertainty Quantification in Lake Prediction using BNN

Variational predictive distribution, which approximates the predictive distribution (posterior probability $p(\boldsymbol{y}^*|\boldsymbol{x}^*, \mathcal{D})$), is defined by Monte Carlo estimator as:

$$\hat{p}_{\psi}(\boldsymbol{y}^* \mid \boldsymbol{x}^*) = \frac{1}{T}\sum_{t=1}^{T} p_{\psi}(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \boldsymbol{\theta}_t), \tag{14}$$

where $\{\boldsymbol{\theta}_t\}_{t=1}^{T}$ are samples drawn from optimized variational distribution $q_{\psi}(\boldsymbol{\theta})$. According to the weak law of large numbers, this estimator converges in probability to $q_{\psi}(\boldsymbol{y}^* \mid \boldsymbol{x}^*)$ as the number of samples T increases. Once the variational parameters $\widehat{\boldsymbol{\psi}}$ are optimized by minimizing the ELBO, the estimator becomes:

$$\hat{p}_{\widehat{\psi}}(\boldsymbol{y}^* \mid \boldsymbol{x}^*) = \frac{1}{T}\sum_{t=1}^{T} p\left(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t\right), \tag{15}$$

Aleatoric and Epistemic Uncertainties: Optimised network parameter $\widehat{\boldsymbol{\theta}}$ is obtained after training a network. Predictive uncertainty $\mathrm{Var}_{p(\boldsymbol{y}^*|\boldsymbol{x}^*,\widehat{\boldsymbol{\theta}})}(\boldsymbol{y}^*)$ can be computed by incorporating negative correlation structures among classes, as proposed by Kwon et al. (2020):

$$\mathrm{Var}_{p(\boldsymbol{y}^*|\boldsymbol{x}^*,\widehat{\boldsymbol{\theta}})}(\boldsymbol{y}^*) \tag{16}$$
$$= \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left[\mathrm{diag}\{p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t)\} - p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t)^{\otimes 2}\right]}_{aleatoric}$$
$$+ \underbrace{\frac{1}{T}\sum_{t=1}^{T}\left[p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t) - \hat{p}_{\widehat{\boldsymbol{\psi}}}(\boldsymbol{y}^* \mid \boldsymbol{x}^*)\right]^{\otimes 2}}_{epistemic}$$

where, $\mathrm{Var}[\cdot]$ is the variance operator, $\mathbf{u}^{\otimes 2} = \mathbf{u}\mathbf{u}^T$ and $\mathrm{diag}(\mathbf{u})$ is the diagonal matrix with elements of $\mathbf{u}$. Predictive uncertainty, $\mathrm{Var}_{p(\boldsymbol{y}^*|\boldsymbol{x}^*,\boldsymbol{\theta})}(\boldsymbol{y}^*)$ is decomposed into aleatoric (data) uncertainty and epistemic (model) uncertainty (as shown in equation 13). Aleatoric uncertainty arises from inherent randomness in the data, while epistemic uncertainty comes from uncertainty in the model parameters (Jospin et al., 2022). Following is the algorithm for calculations of posterior probability and predictive uncertainty.

Algorithm 2 Training BNN and uncertainty quantification with variational inference

procedure Training a BNN

Initialize the learnable parameter $\boldsymbol{\psi}$ in the neural network

while $\boldsymbol{\psi}$ does not converge **do**

from training dataset $\mathcal{D}$ randomly sample a mini-batch $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{M}$.

draw T_{tr} samples $\{\boldsymbol{\theta}_t\}_{t=1}^{T_{tr}}$ from the variational distribution $q_{\boldsymbol{\psi}}(\boldsymbol{\theta})$.

Compute negative ELBO.

$$\mathcal{L}(\boldsymbol{\psi}) \leftarrow \frac{1}{T_{tr}}\sum_{t=1}^{T_{tr}}\left[-\frac{1}{M}\sum_{i=1}^{M}\log p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{\theta}_t) + \mathrm{KL}\left(q_{\boldsymbol{\psi}}(\boldsymbol{\theta}_t)\|p(\boldsymbol{\theta}_t)\right)\right].$$

Update $\boldsymbol{\psi}$ with gradient $\nabla_{\boldsymbol{\psi}}\mathcal{L}(\boldsymbol{\psi})$ and learning rate α.

$$\boldsymbol{\psi} \leftarrow \boldsymbol{\psi} - \alpha\nabla_{\boldsymbol{\psi}}\mathcal{L}(\boldsymbol{\psi}).$$

end while

procedure Uncertainty Quantification

let $\widehat{\boldsymbol{\psi}}$ be the optimised parameter after training network.

11: # draw T samples $\{\widehat{\boldsymbol{\theta}}_t\}_{t=1}^{T}$ from optimised variational distribution $q_{\widehat{\boldsymbol{\psi}}}(\boldsymbol{\theta})$.

12: # for a new input $\boldsymbol{x}^*$, compute the uncertainties

13: # aleatoric uncertainty:

$$\frac{1}{T}\sum_{t=1}^{T}\left[\text{diag}\{p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t)\} - p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t)^{\otimes 2}\right],$$

14: # epistemic uncertainty:

$$\frac{1}{T}\sum_{t=1}^{T}\left[p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t) - \hat{p}_{\widehat{\boldsymbol{\psi}}}(\boldsymbol{y}^* \mid \boldsymbol{x}^*)\right]^{\otimes 2}$$

15: and $p(\boldsymbol{y}^* \mid \boldsymbol{x}^*, \widehat{\boldsymbol{\theta}}_t) = f_{\widehat{\boldsymbol{\theta}}_t}(\boldsymbol{x}^*)$.

4.2.5 Optimization of Hyperparameters

Optimizing hyperparameters is a critical stage in ML algorithms, as it directly influences the model's predictive performance. Configuring the right set of hyperparameters often requires a mix of specialized knowledge, intuition, and trial-and-error experimentation. Bayesian optimization identifies the optimal hyperparameters that minimize the validation error. Let X represent the space of possible hyperparameters and f denote the objective function for minimizing validation error. The optimization process can be expressed as:

$$x^* = \arg\min_{x \in X} f(x), \tag{17}$$

here, x^* represents the set of hyperparameters that produce the minimum value of the objective function, while x refers to any value within the space X.

This process is classified as a global optimization of black-box functions, where the function expressions and their derivatives are unknown (Snoek et al., 2012). It utilises a surrogate probabilistic model based on Bayes' theorem , where the selection of values for subsequent iterations is guided by the outcomes of previous ones (Brochu et al., 2010). Bayesian optimization is especially effective because the surrogate model is computationally cheaper to evaluate than the actual target function. It strategically narrows down the search space by using prior evaluations, avoiding the inefficiencies of grid search or random search, which often waste time exploring poorly performing areas (Wu et al., 2019). This makes Bayesian optimization highly suitable for optimizing expensive or black-box functions, such

as deep learning model tuning or scientific simulations. For a more comprehensive understanding on the Bayes optimization process, a detailed article by Williams et al. (2006) can be referred. There are six stages in Bayesian optimization as follows:

1. Start by evaluating the objective function at a few initial hyperparameter points and construct a surrogate probabilistic model.
2. Use an acquisition function (e.g., Expected Improvement or Upper Confidence Bound) to determine the next hyperparameter set to explore, balancing exploration and exploitation.
3. Evaluate the objective function (e.g., model accuracy or error) for the selected hyperparameters.
4. Update the surrogate model with the new observation, refining its approximation of the objective function.
5. Repeat steps 2–4 for a specified number of iterations or until a stopping criterion is met.
6. Select the hyperparameters with the best-observed performance as the final configuration.

4.3 Simulation

In the ROI, 12,924 gridded images are used. For each of the five erosional features, namely, glacial lake, cirque, valley, flow channel and retreating glacier, images in overlaid grid are visualized from bottom row to top row in a column of the overlaid grid. Once information from all images in a column is collected, the same process is repeated for the next column till all images in the study area are observed.

Next, for a location, a dataset for model training is prepared by observing the occurrence of neighbouring lakes and erosional features within the 3×3 neighbouring grids' images around the central grid. Simultaneously, the dataset for topography features of the central grid is also included for the same location. The central grid image location contains two types of scenarios: either it contains ‘lake’ or it contains ‘no lake’. As a few images contain multiple lakes, it is found that 2,462 image locations contain 2,638 lakes, while the remaining 10,462 images show ‘no lake’ occurrence. For neighbourhood lakes, a lake occurring either upstream or downstream with respect to a central grid plays a unique role in lake formation. Upstream lakes act as a reservoir for the lake in the central grid position and provide water through overflow or drainage for the lake formation, while downstream lakes indicate established

drainage pathways and geomorphic suitability like cirque or valley formations, suggesting favourable conditions for water accumulation in the central grid. Therefore, the input features dataset is prepared by adding information for the presence of neighbourhood lakes, cirque, valley, retreating glaciers and flow channels in the neighbourhood of the central grid.

Input features of topography (slope, elevation, profile curvature and plan curvature) associated with 'lake' and 'no lake' locations are retrieved by using SRTM DEM through ArcGIS. The prepared dataset of the nine features shows information about the surrounding erosional features and topography variations. The dataset shows an imbalance for the two classes, i.e. 2,462 'lake' and 10,462 'no lake'. Out of these 12,924 data locations, 60% data is randomly collected and used to train the three models, and 40% data is used for testing. Training with the imbalanced data may result in a biased model that favours the majority class. However, an ideal model should distinguish between both classes accurately. Therefore, to improve the predictive power of the three models, class weights are considered. Although under-sampling is a valid method for class imbalance, it was not used to avoid discarding valuable spatial patterns present in the majority class (no-lake), which are essential for correctly identifying regions unfavourable for lake formation. Afterwards, the trained models are applied to the remaining dataset for computing the probability of lake formation in the entire study area. The following sub-sections discuss the details of optimisation of model hyperparameters and training of the three models.

4.3.1 Hyperparameter Optimisation

This sub-section discusses the optimisation of hyperparameters using Bayesian optimization, including penalty, solver, regularisation (C), and maximum iterations for LR models, as well as the number of hidden layers, neurons per layer, epochs, learning rate, activation function, and mini-batch size for ANN and BNN models. Table 4.2 shows the option details for LR model hyperparameters, and Table 4.3 presents the varying range of hyperparameters for the ANN and BNN models.

Table 4.2 Description of LR model hyperparameters

Hyperparameters for optimization	Option details
Penalty type	L1 (lasso), L2 (ridge), elasticnet, none
Solver	saga, liblinear, lbfgs, newton-cg
Regularization (C)	$10^{-4} - 10^{4}$
Maximum iterations	$50 - 500$
Class weights	balanced, none

Table 4.3: Range of Hyperparameters for tuning the ANN and BNN models

Hyperparameters for optimization	Range
Number of hidden layer	$1 - 6$
Neurons per hidden layer	$10 - 100$
Epochs	$100 - 500$
Mini batch size	$50 - 150$
Learning rate	$0.0001 - 0.1$
Activation function	Tanh, ReLU, Sigmoid
Class weights	balanced, none

Hyperparameter optimization is conducted by minimizing the cross-entropy loss using *k*-fold cross-validation, where the training data (which corresponds to the complete dataset, excluding the test dataset) is divided into *k* subsets. Each subset is used once as a validation set, while the remaining subsets are used to train the model. This process is repeated until all *k* subsets have been used as the validation set. The average validation error across all *k* folds is calculated and used to identify the optimal hyperparameter combination (Qiu et al., 2022). In this study, *k* is set to 5, determined through an ablation study that balanced computational cost against accuracy for higher *k* values.

The hyperparameter tuning process begins with initial trials using random combinations of hyperparameters within the defined search space. Following the initial trials, each subsequent iteration involves fitting a Gaussian process and utilizing the posterior distribution, combined with an exploration technique, to identify the next set of hyperparameter combinations to evaluate (Snoek et al., 2012; Williams et al., 2006). This process continues until the specified number of trials and iterations is completed. Table 4.4 shows the optimal hyperparameters values for the LR model, while Table 4.5 presents the optimal hyperparameters for the ANN and BNN models, both selected by minimising the negative accuracy (maximising accuracy).

Table 4.4: Optimal hyperparameter values for LR model

Hyperparameters for optimization	Option details
Penalty type	L2 regularization
Solver	Lbfgs
Regularization (C)	0.0001
Maximum iterations	369
Class weights	Balanced

Table 4.5: The optimised values of hyperparameters for the ANN and BNN models

Hyperparameters for optimization	Tuned values for ANN	Tuned values for BNN
Number of hidden layer	3	3
Neurons per hidden layer	99	71
Epochs	173	292
Mini batch size	88	105
Learning rate	0.0021	0.0026
Activation function (Hidden layer)	Sigmoid	Sigmoid
Class weights	Balanced	Balanced

The optimal hyperparameter combinations are then utilized to train the model on the entire training dataset, and the resulting outputs are evaluated on the test set. A detailed discussion of the test results of all the three models (LR, ANN and BNN) are provided in the subsequent section.

4.3.2 Model simulation

4.3.2.1 Logistic Regression (LR) Model for lake prediction:

Using the optimal hyperparameters, the LR model is trained, and the model coefficients (β values) of influencing features or variables (x_i) are determined with their p-values. Table 4.6 mentions model coefficients (β_0 *to* β_9) and p-values show statistical significance.

Table 4.6: Coefficients of the LR model with p-values

Input prediction features (Coefficients)	Feature description	Coefficient estimate	p-value
Intercept (β_0)	not applicable	-2.156	2.55×10^{-15}
Neighbourhood Lakes (β_1)	presence	2.2522	0
Cirque (β_2)	presence	0.40431	1.05×10^{-8}
Flow channels (β_3)	presence	-0.12984	0.01427
Retreating glacier (β_4)	presence	0.16427	0.00143
Valley (β_5)	presence	0.13453	0.02979
Elevation (β_6)	Metre	-0.00022	7.85×10^{-5}
Slope (β_7)	Degree	-0.18355	2.28×10^{-15}
Profile curvature (β_8)	degree per metre	-0.00051	0.01815
Plan curvature (β_9)	degree per metre	0.00042	0.08371

$$Y = -2.156 + 2.2522x_1 + 0.40431x_2 - 0.12984x_3 + 0.16427x_4 + 0.13453x_5 - 0.00022x_6 - 0.18355x_7 - 0.00051x_8 + 0.00042x_9. \quad (18)$$

Probability values of glacial lake formation are calculated by model equation (19) and equation (2). Coefficient values with sign (or beta values) indicate the direction and strengths of the nine predictor features for the lake formation outcome, while p-values assess the statistical significance of the features. A p-value less than 0.05 suggests that a feature significantly contributes to the model. Also, a negative value of the model's intercept (β_0) is −2.156 (with a p-value less than 0.00001), with other predicting coefficients zero, indicates that glacial lake formation is least likely to occur ($P = 0.1$) in the absence of erosional features and appropriate topography.

Among the erosional features contributing positively, neighbourhood lakes have the most substantial positive impact ($\beta_1 = 2.2522$ with p-value= 0). The proximity of neighbourhood lakes can affect hydrological dynamics by contributing to groundwater seepage or surface flow between neighbouring lakes. This clustering of lakes can create a cumulative effect, whereby lakes collectively influence local water levels and stability, supporting lake formation in nearby basins as upstream lakes play a critical role in transferring water to depressions such as cirques or valleys at lower elevations through flow channels or seepage. Also, overflow or drainage from these upstream lakes can lead to the formation of new lakes in downstream regions. Similarly, the occurrence of downstream lakes indicates established flow paths, suggesting the presence of erosional depressions formed during water inflow. These interconnected hydrological processes emphasize the significant role of neighbourhood lakes in facilitating lake formation.

Compared to neighbouring lakes, cirques, retreating glaciers, and valleys have relatively smaller contributions, as reflected by their coefficient values of 0.40431, 0.16427, and 0.13453, respectively. Nevertheless, their positive coefficients indicate that their presence still increases the likelihood of lake formation. Cirques, with the highest coefficient value among the three, provide natural basins that can accumulate meltwater. However, their ability to form lakes depends heavily on water input from external sources, such as upstream lakes. Retreating glaciers contribute meltwater from glaciers, but the rate of production is slower and less concentrated than the substantial reservoir-like releases from upstream lakes.

A small value of the valley coefficient compared to the cirques indicates that the presence of valley landforms has less significance for lake formation. Valleys, shaped as elongated troughs, primarily act as pathways for water flow rather than accumulation, dispersing meltwater and limiting the formation of stagnant water bodies. Moreover, valleys often

undergo significant sediment transport and deposition, which can fill potential basins and prevent lake formation. In contrast, sediment deposition in cirques is less impactful due to their steeper slopes and enclosed nature, preserving the over-deepened basin structure for water retention. A study by Mal et al. (2020) also highlights that the cirque lakes are more dominant in occurrence than the valley lakes or other erosion-related lakes in the Eastern Himalaya.

On the other hand, a negative coefficient value of flow channels ($\beta_3 = -0.12984$) shows that the absence of flow channels will increase the probability of glacial lake formation. This signifies that if flow channels are absent, flow transfers early over a shorter distance to a lake, increasing the likelihood of lake formation. On the contrary, if flow channels are present (they appear more in number and length), it indicates that meltwater will flow for a longer route or pathway, accumulating at downstream regions with substantial delays.

Elevation has a negative coefficient, suggesting that higher elevations decrease the likelihood of lake formation, because most of the lakes occur in the elevation range of 4000 m to 5000 m within the study area. Notably, this elevation is below the equilibrium line of altitude, which ranges from 5500 m to 6000 m in the Eastern Himalaya (Nunchhani et al., 2023). This also indicates that most of our study area with lakes is in the environment of an ablation zone that promotes the likelihood of lake formation. However, the effect of elevation is minimal due to the small coefficient size.

Slope, with a significant negative coefficient ($\beta_7 = -0.18355$), indicates that as the slope increases (steeper slopes), it reduces the likelihood of lake formation as it helps in water drainage rather than retention. Contrarily, flatter areas are more favourable for water accumulation and support lake formation. Profile curvature, with a negative and significant coefficient (p-value = 0.0181), shows that regions with convex surface profiles favour lake formation, as they promote water movement to the depression and retention in one place. Finally, plan curvature (horizontal terrain curvature) has a very small positive coefficient, and it is the only feature which is statistically insignificant (p-value > 0.05), implying that plan curvature has a minor influence on the likelihood of lake formation in this model.

4.3.2.2 Artificial Neural Network (ANN)

ANN architecture was defined with nine features in the input layer, followed by three fully connected hidden layers, each with 99 neurons and sigmoid activation functions to introduce non-linearity. Finally, a fully connected layer contains two neurons corresponding to two classes (lake and non-lake), and the class outputs are converted to class probability values by

a softmax function. The output from the softmax layer is passed to the classification layer, where the model is trained by minimizing the cross-entropy loss as given in the equation (6). Class weights are applied to handle any class imbalance, giving more importance to underrepresented classes in the training process. Training data labels, initially represented by 0 and 1 (for lake and non-lake, respectively) are converted to a one-hot encoded vector. The model is trained using the Adam optimizer for up to 173 epochs with a mini-batch size of 88, and an initial learning rate of 0.0021. Additionally, validation data is provided to monitor the model's performance throughout the training process.

4.3.2.3 Bayesian Neural Network (BNN)

In BNN, weights and biases are varied by a Gaussian distribution. $\mathcal{N}(\mu, \sigma^2)$, where μ represents the mean and σ^2 the variance. During training, the BNN learns the means and the variances of Gaussian distributions, which ultimately determines the distribution of the weights and biases.

In this thesis, scale mixture of two Gaussian distributions is proposed as the prior, where both distributions have a zero mean but different variances. The mixture distribution is given by:

$$p(\boldsymbol{\theta}) = \pi\mathcal{N}(0, \sigma_1^2) + (1 - \pi)\mathcal{N}(0, \sigma_2^2), \tag{19}$$

where, σ_1^2 and σ_2^2 are variances of two mixture components. The first component has a higher variance $(\sigma_1 > \sigma_2)$, leading to a broader prior distribution with a heavier tail. The second component has a small variance $(\sigma_2 \ll 1)$, concentrating many of the prior values around zero (Blundell et al., 2015). The variances are decided by learning during training. The prior distribution follows a Gaussian mixture model with $\mu = 0$, $\sigma_1 = 1$, $\sigma_2 = 0.5$ and mixing proportion $\pi = 0.5$.

The BNN architecture includes a feature input layer with nine units corresponding to the input features. Following this, three Bayesian fully connected layers, each with 71 neurons and Sigmoid activation functions, participated. These layers computed average weights and biases of the weight distribution during training. The Last Bayesian fully connected layer has two neurons corresponding to the two output classes (lake and non-lake), with a softmax function converting the output into class probabilities.

The BNN is constructed, and small values of specific learning rate are applied to ensure the Bayesian parameters remain close to their initial values. An initial learning rate for the priors

is set to 0.25 to update the prior probabilities of the weight distributions, while a learning rate of 0.0026 is used to update the means and variances of layer weights and biases. The training is carried out over 292 epochs with a mini-batch size of 105. The Adam optimizer is employed for updating the network parameters, and both the Evidence Lower Bound (ELBO) loss and validation loss are monitored continuously. After training, the developed BNN model is applied to the entire ROI, providing a spatial probability map for lake formation. Additionally, uncertainty estimates are generated through posterior sampling, aiding in more informed decision-making for predicting lake formation.

4.4 Results and Discussion

4.4.1 Model Performance and Accuracy Assessment

After training the LR, ANN and BNN models using predictor variables, the accuracy analysis of the models is performed by using the ROC curves. Figure 4.6 presents the ROC curves, which present the performance of the LR, ANN and BNN models for classifying 'lake' and 'no lake' instances for the test dataset. The ROC curve represents the trade-off between true positive rate (sensitivity) and false positive rate across varying classification thresholds, with the AUC serving as a metric for overall model discrimination. The LR model achieved an AUC of 0.829, indicating moderate capability in distinguishing 'lake' from 'no lake' locations (Figure 4.6 a). The ANN model improved marginally upon this result, with an AUC of 0.837, reflecting enhanced ability to capture non-linear relationships between predictor variables (Figure 4.6 b). The BNN model exhibited the highest classification accuracy, achieving an AUC of 0.878 (Figure 4.6 c). The ROC curve for the BNN model closely approaches the top-left corner, indicating high sensitivity and a low false-positive rate, both critical attributes for reducing resource expenditure and operational risks during field validation in remote, high-altitude environments. Comparative analysis shows that all models perform substantially above random classification (AUC = 0.5), with the BNN model significantly improving predictive accuracy and reliability over LR and ANN.

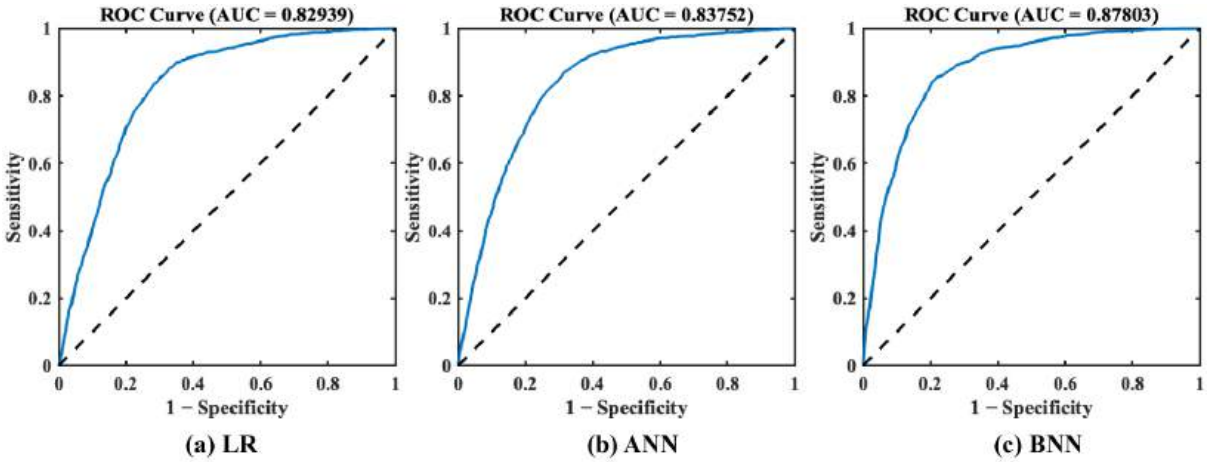


Figure 4.6: Model accuracy based on ROC curve for the test dataset: (a) LR, (b) ANN, and (c) BNN.

4.4.2 Spatial Prediction and Uncertainty Quantification

This chapter uses the spatial distribution of existing glacial lakes as reference labels to train models that predict the likelihood of glacial lake formation in the surrounding terrain. The presence of existing lakes provides empirical examples of where glacial and geomorphic conditions have already favoured lake development. The predictive models developed here therefore do not replicate the current lake inventory; instead, they generate continuous probability surfaces representing the potential for future or incipient lake formation based on terrain characteristics and erosional features.

Continuous probability maps generated using kriging (Vashistha et al., 2023) interpolation illustrate the spatial distribution of glacial lake formation likelihood across the Eastern Himalaya (Figure 4.7, Figure 4.8 and Figure 4.9). These maps classify probabilities into five distinct ranges: very low (0–0.2), low (0.2–0.4), moderate (0.4–0.6), high (0.6–0.8), and very high (0.8–1.0). Regions of very high probability, delineated in red, correspond to areas with geomorphological conditions conducive to lake formation. Contrarily, very low likelihood regions (0 to 0.2) are marked with green colour, indicating the locations where lakes are least likely to form.

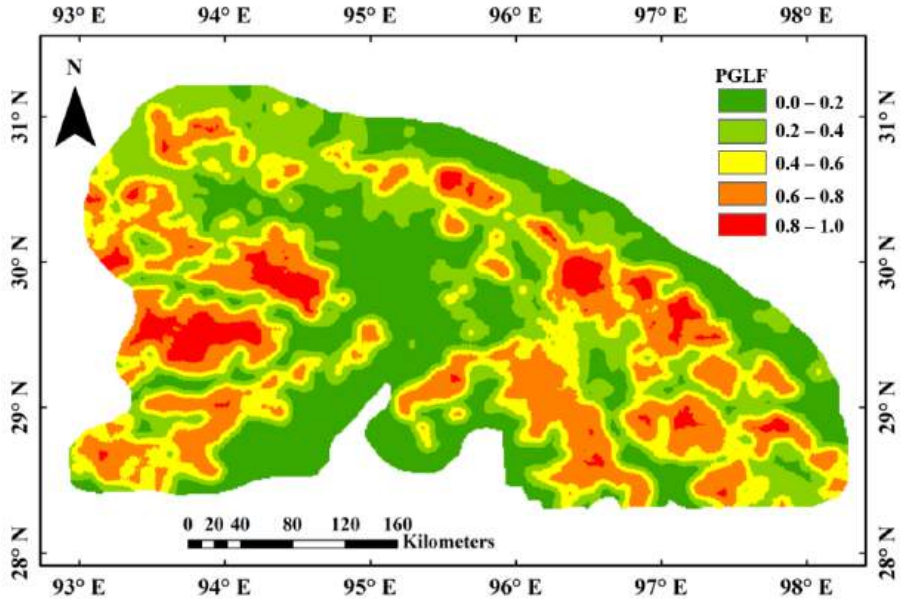


Figure 4.7: LR based probability map of glacial lake formation. The LR-derived map shows colour-coded probability, where red indicates high and dark green low probabilities of lake formation. These maps provide critical information for identifying existing glacial lakes, predicting regions susceptible to future glacial lake formation, and informing hazard mitigation strategies related to GLOFs

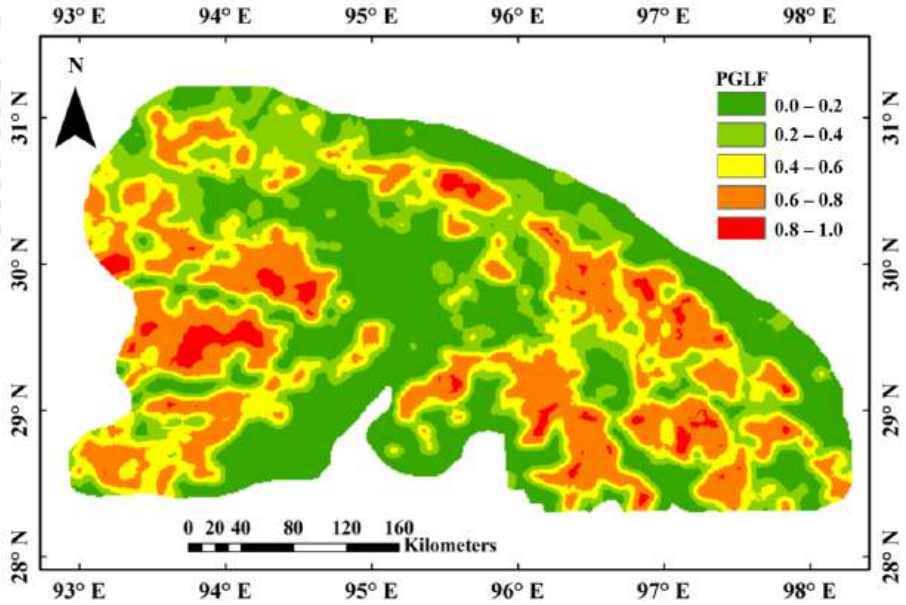


Figure 4.8: ANN model-based probability map of glacial lake formation. It displays colour-coded probabilities, where red indicates high and dark green low chances of lake formation. This map offers essential information for locating existing glacial lakes, forecasting areas vulnerable to future glacial lake development, and guiding hazard mitigation measures related to GLOFs.

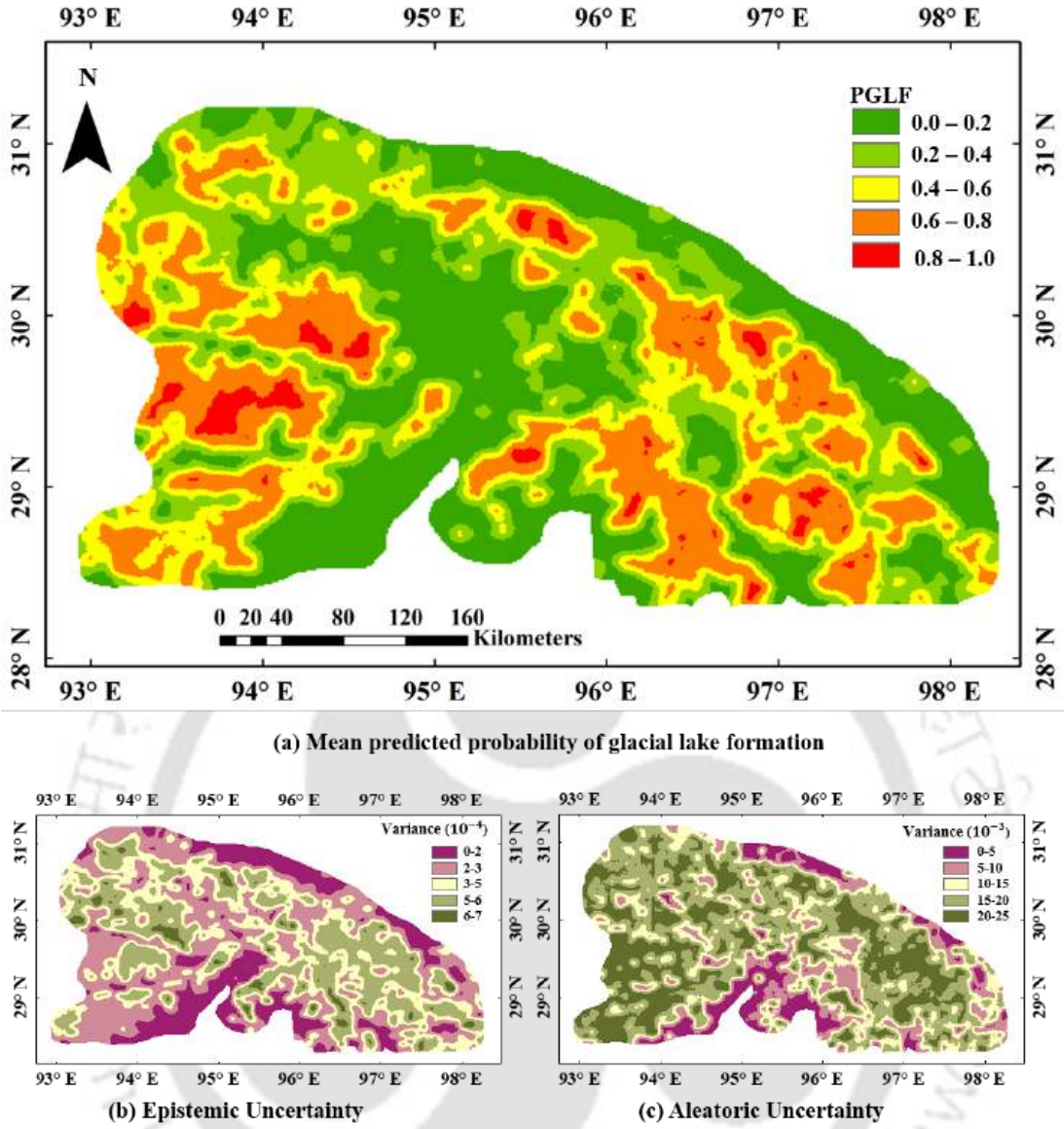


Figure 4.9: BNN model-based PGLF and its associated variance. The top panel shows PGLF values categorised from very low (0.0–0.2, green) to very high (0.8–1.0, red), indicating areas with varying likelihood of glacial lake development. The bottom-left panel depicts variance in PGLF predictions at a scale of 10^{-4}, while the bottom-right panel shows variance at 10^{-3}. Higher variance areas (purple to light brown) indicate greater uncertainty in model predictions and identify zones requiring further investigation.

The LR-based probability map (Figure 4.7) shows a spatially constrained cluster of high-probability locations, whereas the ANN model (Figure 4.8) displays broader, more spatially distributed zones of elevated lake formation likelihood. The BNN model (Figure 4.9a) demonstrates the most extensive and spatially coherent representation of very high-probability regions, reflecting its superior ability to resolve complex spatial patterns associated with glacial lake development.

Figure 4.9 illustrates the spatial distribution and associated uncertainty of the PGLF across the study area. The upper panel (Figure 4.9 a) presents the PGLF values, categorized into five probability classes as mentioned in the previous paragraph. High to very high probabilities (orange to red regions) are predominantly found in central, southeastern, and some southwestern zones, indicating favorable conditions for glacial lake development. These areas are likely influenced by glacial retreat, steep topography, and high meltwater input. In contrast, the northwestern and some central areas show lower PGLF values (green), suggesting a limited likelihood of lake formation due to less favorable geomorphological or climatic conditions. The bottom two maps illustrate the variance associated with the PGLF estimates, which reflects the uncertainty in model predictions. Figure 4.9 b) shows epistemic uncertainty with variance (σ^2)the order of 10^{-4}, while Figure 4.9 (c) represents aleatoric uncertainty with higher magnitude variance on the order of 10^{-3}. The areas with higher variance (light brown to purple shades) are mostly located in the southern and southwestern parts of the study region, suggesting increased uncertainty possibly due to limited input data, higher terrain complexity, or model sensitivity in those areas. Regions with lower variance (green to yellow) correspond to zones with more consistent environmental and topographic characteristics, indicating higher model confidence. The total uncertainty of the prediction is estimated by adding epistemic and aleatoric uncertainty. Estimates of the total uncertainty can be used further for evaluating the upper and lower bounds of the confidence interval. This suggests that the BNN model is the most effective in explaining the complex and non-linear relationship between the erosional features and the lake formation compared to the other two models (LR and ANN).

4.4.3 Probability Distribution and Feature Influence

PGLF distributions for 'no lake' (n = 10,462) and 'existing lake' (n = 2,462) grid locations reveals key patterns. Table 4.7 presents the number of ‘no-lake’ grid locations falling within specific probability ranges of glacial lake formation (PGLF), as predicted by the ML models, thereby representing the spatial likelihood of future lake formation.

Table 4.7: PGLF for ‘no lake grid locations’ (total grids = 10,462)

Model	PGLF-based grid distribution				
	0.0-0.2 (Very Low)	0.2-0.4 (Low)	0.4-0.6 (Moderate)	0.6-0.8 (High)	0.8-1.0 (Very High)
LR	4853	2207	927	1658	817
ANN	5322	1444	1253	1762	678
BNN	4908	2116	1323	1571	544

In 'no lake' grid locations (Table 4.7), all models consistently identify substantial portions of the ROI within the very low (0–0.2) and low (0.2–0.4) probability ranges, with LR predicting 4853 and 2207 locations, ANN identifying 5322 and 1444, and BNN predicting 4908 and 2116, respectively. This indicates that at these locations, the models have predicted a lower chance of likelihood of glacial lake formation, reflecting minimal influence. Notably, 4105 locations in the very low probability range (0.0 to 0.2) are consistently identified by all three models. It is found that neighbouring lakes are not present in all these locations and on steep slopes. In contrast, within moderate probability regions (0.4–0.6), the BNN model predicts 1323 grids, marginally higher than ANN (1253) and LR (927), reflecting increased model sensitivity to transitional conditions where lake formation potential is uncertain. Of particular relevance are the very high probability (0.8–1.0) regions, where 334 locations are consistently identified by all three models within 'no lake' areas, suggesting geomorphological conditions highly favourable for future lake formation despite the current absence of surface water bodies. These locations predominantly feature neighbouring lakes and cirques, with varying occurrences of retreating glaciers (280 locations), valleys (286), and flow channels (196), all coinciding with gentle slopes.

Table 4.8: PGLF for 'existing lake grid locations' (total grids = 2,462)

Model	PGLF based grid distribution				
	0.0-0.2 (Very Low)	0.2-0.4 (Low)	0.4-0.6 (Moderate)	0.6-0.8 (High)	0.8-1.0 (Very High)
LR	117	199	286	1049	812
ANN	126	141	310	1093	792
BNN	78	155	187	1017	1025

Table 4.8 presents the number of existing lake grid locations falling within specific probability ranges of glacial lake formation (PGLF), indicating the model-predicted likelihood of lake presence. In the very low (0 to 0.2) to low (0.2 to 0.4) probability ranges, LR predicts 117 and 199 locations, ANN identifies 126 and 141, and BNN predicts 78 and 155 locations. The predictions with low probability suggest low confidence in lake formation at these grids and indicate that lake-forming characteristics are less pronounced here. Notably, BNN predicts the least number of existing lakes (true positive lakes) locations with a very low probability. Whereas LR and ANN predict a higher number. In the moderate probability range (0.4 to 0.6),

the number of predicted locations increases, with LR identifying 286, ANN predicting 310, and BNN identifying 187, representing transitional regions with moderate favourability for lake formation. In high range (0.6 to 0.8) and very high range (0.8 to 1.0) of probabilities, LR predicts 1049 and 812 locations, ANN predicts 1093 and 792, and BNN identifies 1017 and 1025, suggesting a strong influence from lake-forming features. Further, prediction of a higher number of lake locations (true positive lake locations) by BNN, followed by ANN and LR in a very high probability range, presents the BNN model's superior effectiveness and reliability in understanding the non-linear complex influence of all the predictor variables.

Notably, 492 locations in the very high probability range (0.8 to 1.0) are common across all three models, indicating a strong likelihood of lake formation driven by predictor features, particularly erosional features. As the LR model indicated dominance of erosional features in lake formation, the occurrence of influencing features in these locations is further observed. The neighbouring lake and cirques are the most critical features, which occur in the neighbourhood of all 492 locations. Further, among these 492 locations, 406 feature valleys, 381 contain retreating glaciers, and 304 exhibit flow channels with a gentle slope and varying occurrence of remaining topography features. These findings also highlight the significant role of erosional features, particularly neighbouring lakes, cirques, valleys, retreating glaciers, and flow channels in determining lake formation probabilities.

4.4.4 Discussion

This chapter provides the first quantitative demonstration of the predictive value of erosional and topographic features–cirques, valleys, flow channels, retreating glaciers, and neighbouring lakes–for assessing glacial lake formation likelihood in the study area using an ML-based framework. This builds upon, but significantly extends, earlier efforts that relied on either linear statistical approaches (e.g., (Mohanty et al., 2022) or subjective geomorphological assessments (Furian et al., 2021; T. Zhang et al., 2022a). The limitations of LR models, including assumptions of linear relationships and statistical insignificance of parameters, are evident in the constrained spatial prediction and lower AUC performance observed here. Similarly, while ANNs capture non-linear relationships more effectively, their inability to quantify predictive uncertainty undermines their reliability in operational settings, particularly where hazard management and resource allocation are concerned (Anilkumar et al., 2023; Bolibar et al., 2020). In contrast, BNNs incorporate both epistemic and aleatoric uncertainty into the prediction process (Gopalan et al., 2018; Guillet et al., 2023), offering not only improved classification accuracy but also critical information on model confidence. This

capability is essential for high-altitude hazard assessments, where ground-truth validation is logistically challenging, resource-intensive, and potentially hazardous. Comparison with alternative uncertainty-aware models, including Monte Carlo Dropout and Deep Ensemble Neural Networks, reinforces the advantage of the BNN approach. Although these models improved upon standard ANN performance (AUC of 84.31% and 83.89%, respectively), they remain inferior to BNNs both in predictive accuracy (AUC = 87.8%) and principled uncertainty estimation, which is consistent with previous applications in glaciological hazard forecasting (Rounce et al., 2020; Werder et al., 2020). The findings also provide empirical support for the central role of geomorphological controls in glacial lake development. Specifically, the consistent association of neighbouring lakes and cirques with high lake formation probabilities aligns with established theories of glacial erosion, overdeepening, and meltwater accumulation (Mal et al., 2020; Veh et al., 2020).

4.5 Chapter Conclusion

The rapid growth of glacial lakes in the study site, driven by climate-induced glacier retreat, increases GLOF risk. Genesis, evolution, and alteration of glacial lakes are primarily governed by terrain morphology and erosional features such as cirques, valleys, flow channels, and retreating glaciers. The following are the specific conclusions drawn from this chapter:

(i) Integrating high-resolution geomorphological data with advanced probabilistic modelling offers a scalable, transferable approach to predicting glacial lake formation under data-sparse, high-uncertainty conditions. LR, ANN, and BNN models include the presence of features like cirques, valleys, flow channels, retreating glaciers, nearby lakes, and numerical values of geometric parameters of glacial lakes (elevation, slope, and curvature) to predict the probability of glacial lake formation at a location.

(ii) Comparative analysis of the three predictive models illustrated that BNN provides the most reliable predictions, achieving an AUC of 0.878. This outperforms LR (AUC = 0.829) and ANN (AUC = 0.837), underscoring the advantage of Bayesian frameworks in capturing complex, non-linear processes governing glacial lake formation. Notably, BNN uniquely quantifies both aleatoric (data-driven) and epistemic (model-driven) uncertainties, ranging between 10^{-3} and 10^{-4}, thereby enhancing forecast reliability, a crucial requirement for hazard mitigation planning in the Himalaya.

(iii) LR identifies the relative influence of geomorphic predictors, highlighting neighbouring lakes, cirques, slope, and retreating glaciers as key controls; its limited accuracy restricts its practical application. In contrast, BNN successfully identifies 2,042 true-positive glacial lake locations, particularly within zones where neighbouring lakes, cirques, gentle slopes, and active glacial retreat coincide. Notably, all three models converge to 492 existing glacial lake locations, validating the geomorphic controls influencing the glacial lake formation processes.

(iv) Maps generated from the three ML models provide critical information for identifying existing glacial lakes, predicting regions susceptible to future glacial lake formation, and informing hazard mitigation strategies related to GLOFs. Moreover, these maps offer essential information for locating existing glacial lakes, forecasting areas vulnerable to future glacial lake development, and guiding hazard mitigation measures related to GLOFs. Furthermore, work presented in this chapter advances glacial hazard forecasting by explicitly integrating geomorphology into a transferable, uncertainty-aware modelling framework. The findings not only support improved monitoring of glaciofluvial erosion and lake development but also inform proactive disaster risk reduction strategies in rapidly evolving mountain environments.

(v) Among the three ML-based models, the BNN model provides an essential tool for GLOF risk assessment, early warning system design, and sustainable development planning in the Eastern Himalaya and other glaciated mountain systems globally.

Finally, while this chapter focuses on the likelihood of glacial lake formation, it does not evaluate the hazard levels associated with these lakes. Not all lakes pose equal threat; some may evolve into high-hazard systems, while others remain benign. Next, Chapter 5 extends this work by assessing the hazard potential of existing glacial lakes. This is accomplished through spatial extraction of key geomorphic features and their integration into a heuristic classification framework, thereby transitioning from formation likelihood to hazard categorization within the broader GLOF risk assessment pipeline.

Chapter 5

Hazard Assessment of Glacial Lakes

5.1 Introduction

As evident from Chapter 4, the formation and transformation of glacial lakes are influenced by erosional features, including cirques, retreating glaciers, flow channels, valleys, slopes, and the presence of nearby glacial lakes. Nonetheless, the computational detection of these features presents inherent challenges due to their morphological complexity, spatial variability, and the absence of well-defined geometric boundaries. Unlike clearly segmented land cover classes, erosional features such as cirques and flow channels are characterized by gradual elevation transitions, heterogeneous surface textures, and overlapping terrain signatures. Their forms are often non-uniform and dependent on local topography and glacial dynamics, thereby diminishing the effectiveness of traditional image classification or threshold-based methods. Furthermore, the spectral similarity of erosional features with surrounding terrain in multispectral satellite imagery constrains the discriminative capacity of conventional remote sensing algorithms. To address these limitations, this chapter introduces a hazard assessment methodology utilising deep learning techniques to extract and interpret the dynamic relationships between these complex erosional features and the evolving glacial lakes amidst glacier retreat. Particular attention is given to glacial lakes at higher elevations (GLHE), which provide a steep gradient promoting water flow toward lower-elevation lakes and thereby heightening hazard potential. The proposed approach models the spatial response of glacial lakes to the varying dynamics of key erosional features to assess the hazard potential of existing lakes. Additionally, it extends the spatial coherence of these key features within a logical framework designed to predict locations where glacial lakes are currently absent but are likely to form in the future. The primary objectives of this hazard assessment are as follows:

(i) Extracting glacial lakes and erosional geomorphic features (cirques, valleys, retreating glaciers, and flow channels) by deep learning,

(ii) Integrating combinations of extracted erosional features with glacial lake parameters and implementing the heuristics rules for determining the hazard potential of the glacial lakes, and

(iii) Detecting locations of future glacial lakes (new glacial lakes) using the presence and dynamics of erosional features.

In this chapter, the term ‘existing lake’ is used to represent glacial lakes that are currently present in the ROI. Whereas the term ‘future lake’ is used for the locations where currently no lake is present, yet lakes are likely to emerge in future. The term ‘retreating glacier (RG)’ refers to the geomorphic imprint of a glacier that has already retreated, rather than the active ice body undergoing climatic retreat. RG denotes the newly exposed terrain–such as polished bedrock, proglacial surfaces, and remnant ice-contact features–that records the former extent of the glacier. This surface acts as a key erosional indicator because its morphology and position strongly influence meltwater routing, basin formation, and potential sites of glacial lake development. The work discussed in this chapter has been published in (Vashistha et al., 2025a).

5.2 Methodology

5.2.1 Glacial lake hazard monitoring

A glacial lake is surrounded by any three of the four erosional features (cirque, valley, flow channels, and retreating glaciers), with a GLHE. These erosional features display certain characteristics that can serve as visible clues for their identification on satellite images. Water from melting glacier ice and seasonal precipitation gathers in glacial lakes (located in cirques) at higher elevations, also called cirque lakes. Flow channels (FC), emerging from a cirque lake or GLHE, carry continuous streams of water and sediments that erode and pluck underlying rocks, transforming V-shaped river valleys into broad U-shaped valleys. The water finally reaches the lakes in the valley (valley lakes) and lower elevations. Therefore, lakes in valleys tend to be deeper and larger than cirque lakes because of the greater volume of water resulting from glacio-fluvial erosion and other processes (Liu et al., 2014). As a glacier retreats over time, RG appears as crevasses, striations, or scratch-like features on the exposed ground surface between the glacial lake and the glacier. Thus, RG connects the glacier terminus to an existing lake or a valley. Footprints of ice adjacent to scratch-like marks of RG indicate the existence of glaciated ice in the past on the exposed ground surface. Figure 5.1 illustrates interlinked and interconnected erosional features guiding glacial lake formation.

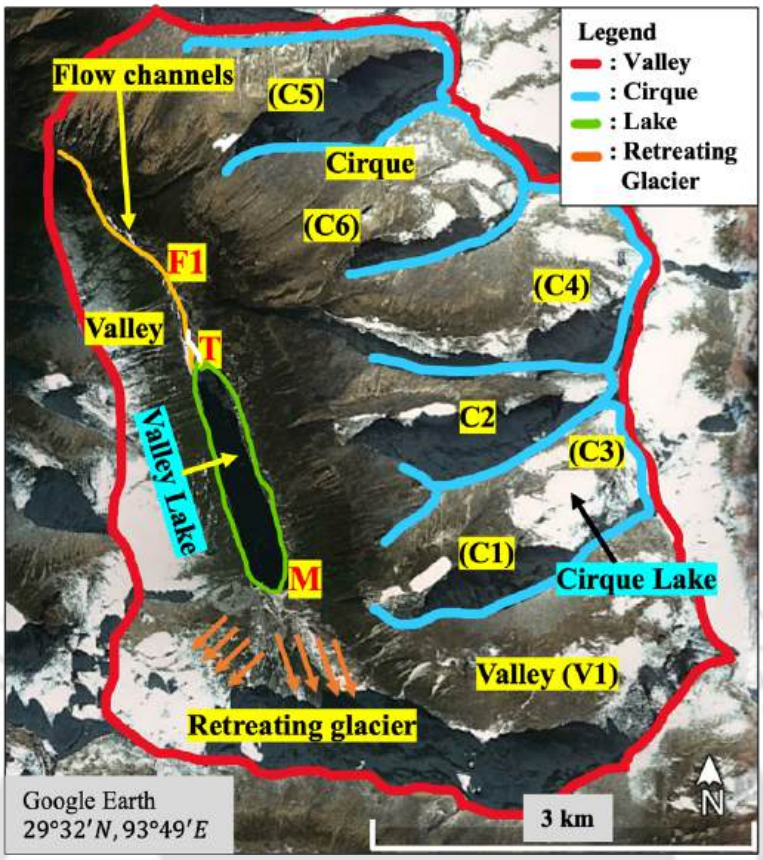


Figure 5.1Erosional features and glacial lake formation are marked on Google Earth imagery of the Bayi district (Nyingchi, Tibet, China) in the Eastern Himalaya.

As shown in Figure 5.1, a wide, large, and elongated valley (V1), which is bounded by the two high hills and marked by a red line, shows evidence of RG movements in the southeast and southwest directions (towards the bottom right corner of the image, indicated by orange arrows). In the valley (V1), meltwater and debris flow enter from point M into the valley lake and exit from point T. Additionally, several cirques (C1, C2, C3, C4, C5, and C6), which appear lateral to the flow in the valley and are surrounded by hills, contribute water through FC to the glacial lake in the valley. Also, an FC (F1) can be observed as an emerging stream from the valley lake (at point T) that carries the lake water towards the downstream region and may form a future lake.

Hazard state of an existing lake is evaluated in a specific order according to the presence of GLHE and RG, the size of the lake, and the presence of the lake itself in the valley. Glacial lakes are first categorized into small, medium, and large sizes based on specified upper and lower threshold values on the lake area. Lakes with an area greater than the upper threshold value are classified as large-sized lakes, those with an area between the lower and upper thresholds are considered medium-sized lakes, and lakes with an area less than the lower

threshold are defined as small-sized lakes. Moreover, the criteria mentioned by Quincey et al. (2007) are adopted. Accordingly, an RG connecting spatially to a glacial lake shows glacier flow movements and thus indicates that the lake is currently active. Conversely, if RG is not present, the existing glacial lake is considered inactive. The following are two fundamental rules defining non-dangerous lakes (or ND lakes), dangerous lakes and locations of future glacial lakes:

(i) 'Inactive lakes', typically where RG is not present, are either stable or shrinking. Therefore, an existing lake, which is inactive and has an area less than a small-sized lake, indicates a non-dangerous lake. Conversely, an active small-sized lake is a dangerous lake. Also, irrespective of being active or inactive, a lake with a minimum area of a medium-sized lake is also a dangerous lake.

(ii) Locations of future glacial lakes are detected by the presence of FC, which are thought to emerge from an existing lake (cirque lake or valley lake), yet not reach other existing lakes at lower elevations. Additionally, the presence of RG or GLHE contributing meltwater, not reaching an existing lake, also indicates the possibility of future lake formation.

It should be noted that we are ignoring contributions by seepage and minor channels below the ground surface. Figure 5.2 demonstrates the methodology of hazard evaluation of existing glacial lakes and the detection of locations of future glacial lakes.

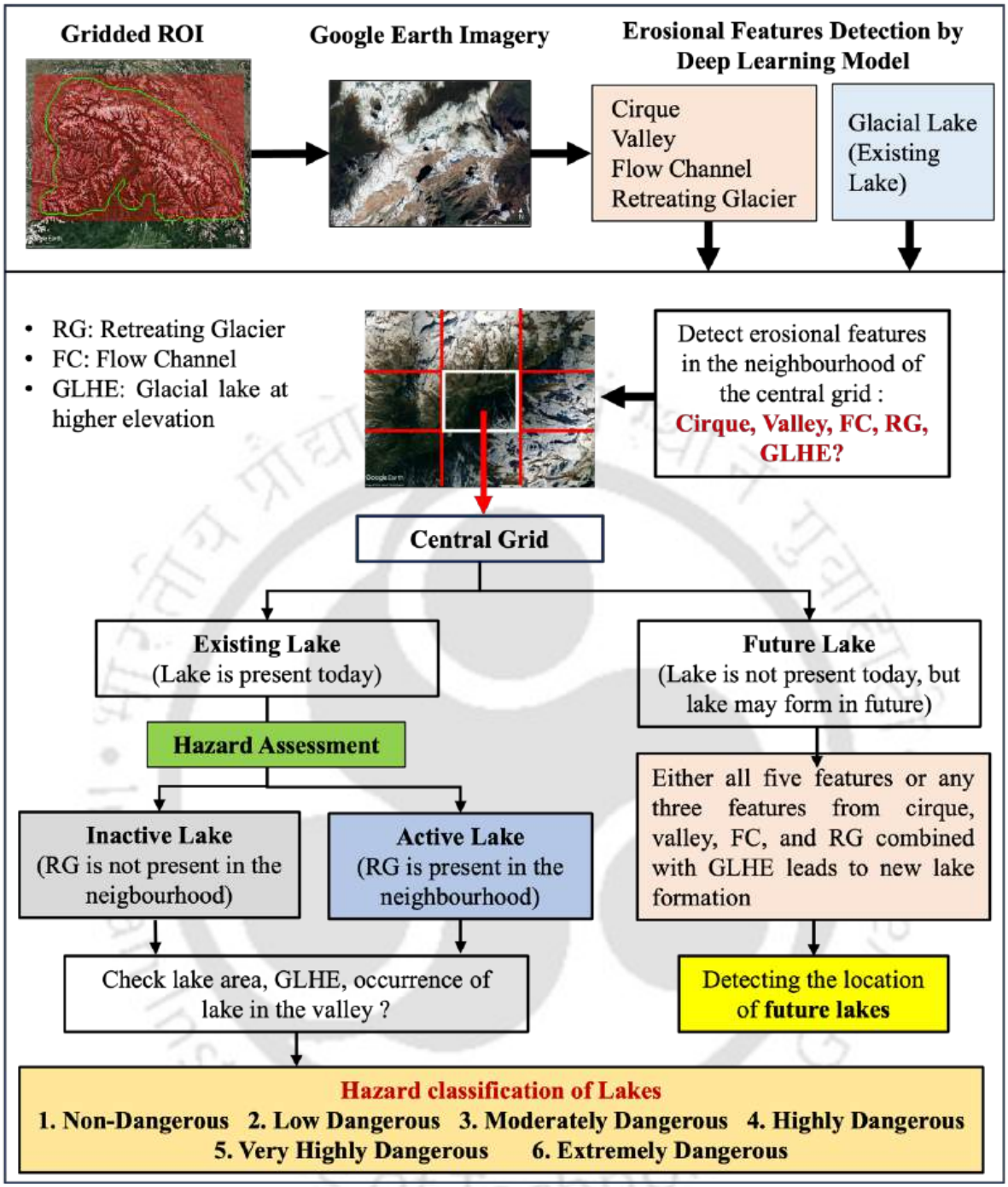


Figure 5.2: Methodology for hazard modelling of glacial lakes.

For a dangerous lake, its hazard potential is increased by the presence of GLHE and RG, as both contribute sediments and water inflow to the lake. Thus, according to the presence of erosional features and GLHE in the neighbourhood, and the size of lakes, dangerous lakes in the ROI are further classified into five classes: i) low dangerous, ii) moderately dangerous, iii) highly dangerous, iv) very highly dangerous, and v) extremely dangerous. Specific rules for classifying a dangerous lake into five classes are as follows:

(i) **Low dangerous (LD):** A small-sized active lake, which is neither present in the valley nor has GLHE in the neighbourhood, is classified as 'low-dangerous'. In addition, if GLHE is not present in the vicinity or neighbourhood of medium-sized and large-sized inactive lakes, these lakes are also classified as 'low dangerous'.

(ii) **Moderately dangerous (MD):** If GLHE is present, inactive medium-sized and large-sized lakes are considered 'moderately dangerous'. Also, medium-sized active lakes, which are not present in the valley and do not have GLHE in the surrounding area, are classified as 'moderately dangerous'. At the same time, a small-sized active lake either present in a valley or having GLHE in the neighbourhood (or both are present) is considered 'moderately dangerous'.

(iii) **Highly dangerous (HD):** A large-sized active lake, not having GLHE in the neighbourhood, or the lake itself is not present in the valley, is classified as 'highly dangerous'. Also, a medium-sized active lake, which is either present in a valley or has GLHE in the neighbourhood, is also categorised as 'highly dangerous'.

(iv) **Very highly dangerous (VHD):** A large-sized active lake with either GLHE in the surrounding area or the lake itself present in a valley is defined as 'very highly dangerous'. In addition, a medium-sized lake present in the valley and having GLHE in the neighbourhood is also classified as 'very highly dangerous'.

(v) **Extremely dangerous (ED):** A large-sized active lake situated in a valley with GLHE in the neighbourhood is classified as 'extremely dangerous'.

Future lakes: With respect to a location, the presence of erosional features at higher elevations and GLHE in the neighbourhood supports the formation of new glacial lakes at that location. The minimal condition for triggering the formation of a new lake is that any three erosional features (among cirque, valley, FC, and RG) in conjunction with GLHE in the neighbourhood are present. Moreover, new lakes start to form at a point of minimum elevation. Figure 5.3 illustrates a flowchart for the hazard assessment of existing lakes and the detection of future lakes.

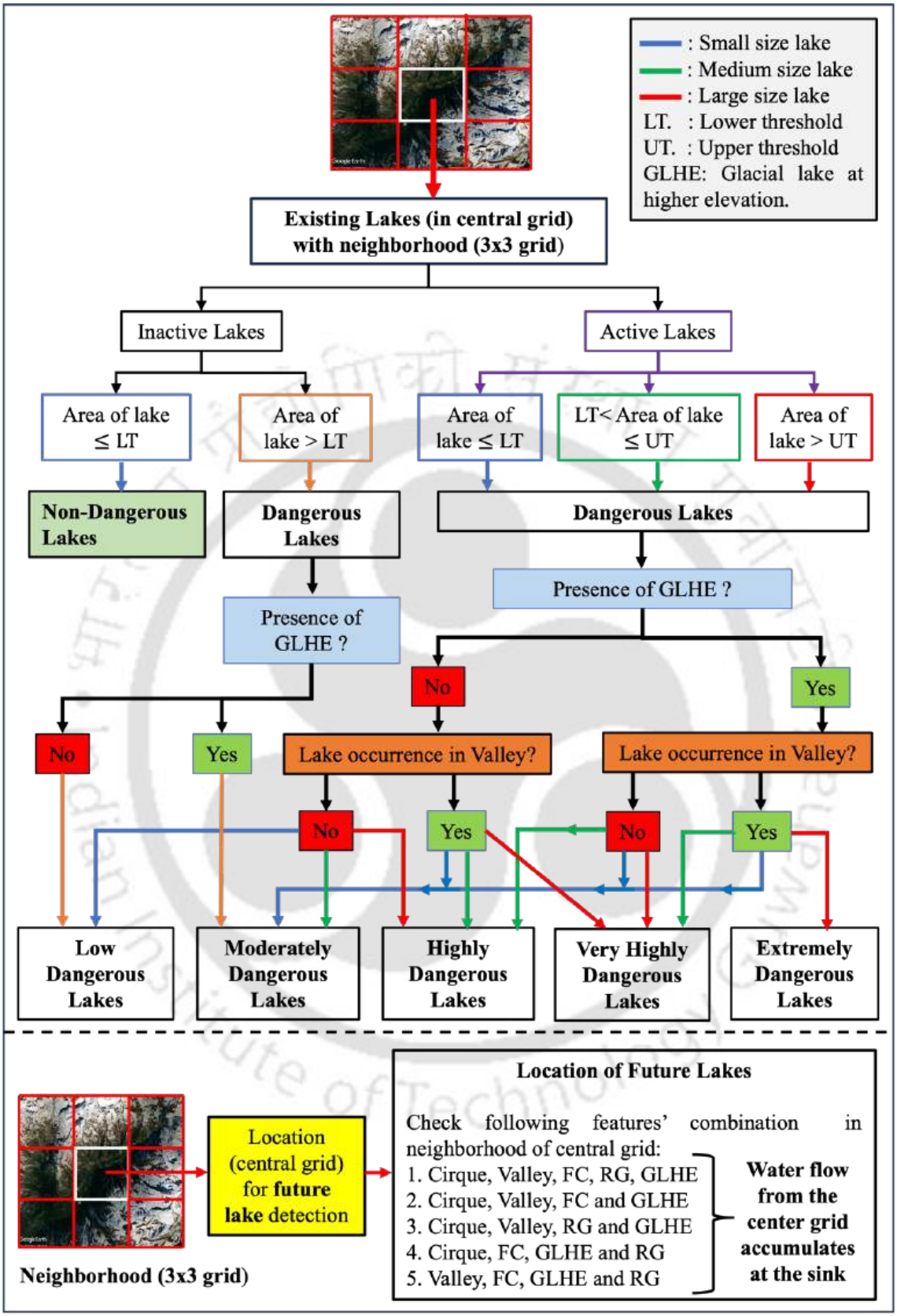


Figure 5.3: Detailed flowchart for the hazard assessment of existing glacial lakes and detection of future lakes.

5.3 Simulation

The methodology is implemented in three steps: (i) extraction of geomorphic erosional features and lakes using CNN, (ii) hazard assessment of the identified lakes using other extracted lakes and erosional features in the neighbourhood, and (iii) predicting locations of future lakes using the extracted lakes and erosional features. Section 5.3.1 discusses the detailed implementation of EfficientNet-B0 CNN for feature extraction. Next, section 5.3.2 discusses the hazard assessment of identified lakes and the detection of future lakes.

5.3.1 Feature extraction

A. *Feature recognition*

The process of feature extraction by the EfficientNet-B0 CNN starts with feature recognition to capture unique textures, patterns, and colour diversity within selected training images. Locations of glacial lakes and erosional features are selected from neighbourhood of 200 locations from the inventory of Mal et al. (2020). To identify, detect, and mark descriptions of erosional features are as follows:

Cirque and Valley: Cirque can be identified as bowl-shaped, steep-walled basins with flat bottoms, typically located at high elevations and occasionally containing a small lake covered with ice. Valleys usually appear as elongated U-shaped depressions with a steep slope that exists between two higher hills. The shape and boundaries of a valley can be marked at the ridge of hills, which may show steep and slanting walls of hills containing FC originating from the ridge portion of the hills. Moreover, a valley is also characterized by either green vegetation, bare brown hillsides, ice-covered slopes, or a combination of these patterns.

Retreating Glacier (RG) and Flow Channel (FC): RG can be identified by striations, which are scratches or grooves on bedrock caused by glacial movement. FC originating from the snout of glaciers due to melting can vary in shape and form, ranging from small streams to large rivers. These channels, while transitioning to dendritic patterns downstream, often exhibit meandering or braided patterns upstream. Depending on the sediment load, water colour and its tonal variations in FC, its colour varies from blue, green, brown or their combinations. Sometimes, FC may emerge from existing lakes and extend downstream.

Glacier Lake: Glacial lakes appear circular, oval, or irregular, with blue or green colours influenced by depth and sediment content.

By recognizing these key characteristics and patterns, erosional features and glacial lakes are marked manually with high accuracy on training images. FC features are marked by water levels on two sides of a channel. Secondary and higher-order FC features are also marked. Further, sediment banks above the water level are not included as part of an FC feature.

B. *Feature extraction by EfficientNet-B0 CNN*

Feature extraction on extracted image tiles is performed using EfficientNet CNN, which is a family of networks, containing 8 models (B0 to B7), and provides a pre-trained CNN. The network also contains deeper networks and complex architectures like ResNet and U-Net. The network uniformly features three scaling dimensions: width, depth, and resolution parameters. The depth represents the total number of layers, the width indicates the number of filters or channels in a layer, and the resolution depicts the quality (number of pixels) or size of input images. The baseline model, B0, contains approximately five million parameters, while the largest model, B7, has 66 million parameters. Also, the network allows for the participation of an image or a set of images with specified probability, i.e. the network will select an image with specified probability. This probability is defined as 'participating probability' for the forthcoming discussion in this chapter.

Among 8 models of CNN (from B0 to B7), the EfficientNet-B0 model is selected for detecting erosional features because it can achieve better results with fewer parameters and shorter training times (Basit et al., 2022; Jain et al., 2024; Qayyum et al., 2020). Moreover, compared to deeper networks like ResNet and more complex architectures such as U-Net, EfficientNet-B0 demonstrates competitive performance on image classification and feature detection tasks with significantly lower computational cost, making it well-suited for large-scale geospatial image analysis where resource efficiency is essential (Rao et al., 2023; Shree et al., 2024; Tan et al., 2019). U-Net is primarily designed for pixel-wise segmentation tasks. However, EfficientNet-B0 performs better for the patch-based detection task, which focuses on identifying the presence or absence of erosional features or lakes in image tiles (Shree et al., 2024). ResNet, although effective, lacks the compound scaling optimization present in EfficientNet, which results in higher parameter redundancy and longer training times for similar performance levels (Rao et al., 2023; Tan et al., 2019). Apart from that, EfficientNet-B0 also achieves faster convergence, reduced overfitting due to its regularized architecture, and robust generalization across diverse input conditions, including both the presence and absence of geomorphic features.

The EfficientNet-B0 model for erosional features and glacial lake detection was trained, validated, and tested on a workstation with an Intel(R) Xenon processor (speed 3.20GHz), 64GB RAM, and an 8GB graphics card. The next sub-sections explain the training, validation, and testing of the EfficientNet-B0 network.

C. *Training and Validating EfficientNet-B0 for Feature Extraction*

To train the EfficientNet-B0 CNN model, binary masks were created by marking each feature on training images manually by precise digitization. A mask (or binary mask) is a binary colour coding of a training image – white representing a feature of interest and black representing a non-feature or background. As mentioned earlier, training images are collected from the neighbourhood of the observed 200 glacial lakes, which were obtained from the glacier lake repository (Mal et al., 2020). Images are selected such that each feature depicts maximum variability for its appearance with respect to colour, texture, pattern and surrounding environment. For the five features (glacial lake, cirque, valley, FC, and RG), 183, 208, 200, 266, and 266 training images and corresponding binary masks were prepared. In addition, an equal number of non-feature images with black masks were also prepared. Model training for the five features is performed using both feature images and non-feature images and corresponding masks (i.e. 336, 416, 400, 532, and 532 number of images and the same numbers of binary masks) for the glacial lake, cirque, valley, FC, and RG, respectively, ensuring an equal representation of feature-present and feature-absent scenarios. This helps the CNN model learn that not all the input images contain features. Moreover, it gives clarity in understanding the environment where the feature is present or absent.

The automated feature extraction process of EfficientNet-B0 includes image pre-processing that adjusts image resolution by resizing each of the original 1024×768 resolution images and the corresponding masks to 224×224-pixel images in each of the three bands to meet the input layer requirement of the CNN model. For generalize feature detection, which may appear in any shape and orientation, a data augmentation process is applied. Accordingly, images and their masks are flipped horizontally and vertically (mirroring about vertical and horizontal edges) with participation probabilities of 0.28 and 0.2, respectively. This introduced variability of the training images in two orthogonally oriented directions, expanding diversity and robustness of the training dataset against a variety of real-world scenarios. Apart from that, implementing random brightness contrast adjustments with a probability of 0.2 enabled the resilience of the EfficientNet-B0 model to variations in lighting and contrast.

Training the EfficientNet-B0 model also includes fine-tuning various hyperparameters, namely learning rate, batch size, and number of epochs, to achieve desired accuracy levels. Various methods like grid search, random search, and Bayesian hyperparameter optimization are commonly used to tune neural network hyperparameters (Kalimullah et al., 2023b). The grid search method and trained the CNN model are employed by varying learning rate (from 0.001 to 0.0001), epochs (from 50 to 200), and batch size (from 12 to 18). The optimal hyperparameter values of 0.001 for the learning rate, 200 for the epochs, and the batch size in the range of 12-14 images were selected, providing minimum training and validation losses. Figure 5.4 plots training loss versus epoch for the trained CNN model.

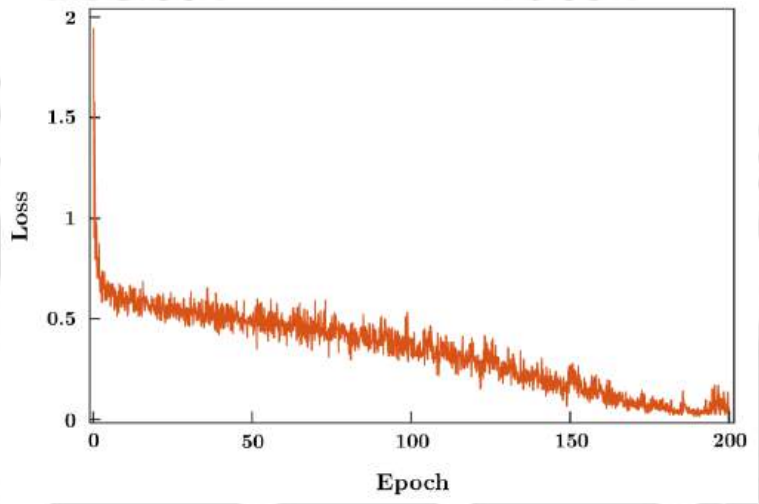


Figure 5.4: Performance summary plot of the optimisation algorithm during the fine-tuning of the EfficientNet-B0 model.

Figure 5.4 illustrates that as epochs tend to 200, training loss converged to a minimum value (< 0.1), showing that the model has appropriately learned the training data. Reducing the learning rate further below 0.001 did not improve the network performance significantly. As a result, the model was trained for 200 epochs with a 0.001 learning rate. To prevent overfitting, early stopping was implemented with an interval of 20 epochs, monitoring the validation loss. However, the validation loss continued to improve throughout training, and early stopping was not ever triggered.

Training performance of the EfficientNet-B0 model for feature detection is evaluated by the IoU (Intersection over Union) score (Tang et al., 2024). The IoU score for a feature is calculated as the ratio of the area of intersection of the feature in the output and input images to the total area occupied by the feature in the two images. Thus, the IoU score is a metric which measures the overall shift of a feature's boundary in the output image with respect to the

original image. Value of the IoU score ranges from 0 to 1, where a value close to 1 represents the efficient prediction (promising performance of the CNN model) while a value close to 0 represents poor performance of the model. For a feature, an average of the IoU score for all input images is calculated to predict the accuracy of the training set and validation set. Values of training loss, validation loss, and IoU scores of training and validation for each feature achieving optimal performance for the EfficientNet-B0 model are mentioned in Table 5.1.

Table 5.1: Hyperparameters optimal values for training the EfficientNet-B0 model

Features	Batch size	Training loss	Validation loss	IoU Score of training set	IoU Score of validation set
Lake	12	0.025	0.032	0.934	0.919
Cirque	12	0.062	0.128	0.843	0.816
Valley	12	0.065	0.195	0.826	0.799
Flow Channel (FC)	14	0.087	0.193	0.896	0.864
Retreating glacier (RG)	14	0.038	0.044	0.902	0.865

The CNN model demonstrates the capability of high accuracy for detecting lakes and RG with IoU scores of 0.934 and 0.902 in the training set, and 0.919 and 0.865 in the validation set, respectively. Higher accuracy of the CNN model for detecting lakes and RG can be attributed to relatively distinct and well-defined characteristics of these features. Lakes generally have clear boundaries and consistent shapes, making them easier for the model to identify. Similarly, RG often exhibits distinct surface patterns and textures that stand out from the surrounding terrain, facilitating accurate detection. On the other hand, the moderate IoU scores, i.e., 0.896 for training and 0.864 for validation, are achieved for FC. This is due to flow variations like contraction and expansion of FC during a flow in a year, causing varying width and less distinct boundaries. Further, lower IoU scores for cirques and valleys are due to the inherent complexity of the two features, due to irregular shapes, varying depths and subtle boundaries that merge into the surrounding terrain, making them difficult for the model to delineate precisely. It should also be noted that the training process of the CNN model for each feature demanded a maximum of 8 hours on the workstation. This indicates the potential utility of the proposed method for practical applications for GLOF feature detection. The next phase deploys the trained CNN model on a large dataset of 12,924 images to identify erosional features and lakes in the ROI.

D. *Feature detection on test images*

With the trained EfficientNet-B0 model, we detected all five features in 12,924 images and analysed accuracy by cross-verification. A confusion matrix is populated with true positive, true negative, false positive and false negative case scenarios for each feature. Subsequently, a series of metrics, namely overall accuracy, precision, recall, and F1-score, are evaluated, demonstrating the quantitative performance of feature detection. Mathematical expressions of the performance indicators are as follows (Goutham et al., 2022; Kaushik et al., 2022):

$$Overall\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{20}$$

$$Precision\ (P_R) = \frac{TP}{TP + FP} \tag{21}$$

$$Recall\ (R_e) = \frac{TP}{TP + FN} \tag{22}$$

$$F1\ score = \frac{2(R_e P_R)}{(R_e + P_R)} \tag{23}$$

Table 5.2 shows specific values of performance metrics for the detection of individual features in test images.

Table 5.2: Performance metric evaluation of 12,924 test images based on the EfficientNet-B0 model

Feature	Overall Accuracy	Precision	Recall	F1-score
Lake	93.26	0.81	0.83	0.82
Cirque	91.31	0.86	0.86	0.86
Valley	86.36	0.77	0.80	0.78
FC	92.67	0.84	0.92	0.88
RG	92.30	0.75	0.84	0.79

Overall accuracy percentage in Table 5.2 presents the ability of the EfficientNet-B0 CNN model to correctly extract features and non-features for a feature class. For instance, the CNN model achieves a high overall accuracy score of 92.67% for FC, indicating the effectiveness of the model for the identification of FC. The precision metric assesses the accuracy of predictions for a specific feature class, while recall signifies the proportion of correctly identified instances within the feature class. It is found that cirques are extracted more precisely with 0.86 precision. On the other hand, FC shows a high recall value of 0.92, suggesting a strong ability of the CNN model to capture the most positive instances of this feature. Additionally, the F1 score is the harmonic mean of precision and recall. If the F1 score is significantly lower (e.g., below 0.5), it indicates a substantial imbalance between precision and recall – either precision or recall is

much lower than the other. F1 scores in the range of 0.78 to 0.86 for all five features indicate a balance between precision and recall and highlight the equilibrium in the model performance. Lakes, cirques, and flow channels achieve F1 scores of 0.82, 0.86, and 0.88, respectively, due to their distinct shapes and clear boundaries, which reflect a good balance between precision and recall. In contrast, retreating glaciers and valleys with their complex dynamic features and less defined edges, as well as their resemblance to bare hillsides, are harder to detect accurately, resulting in lower F1 scores of 0.79 and 0.78, respectively.

Figure 5.5 below depicts an example of feature extraction by the trained CNN (EfficientNet-B0) model for the validation set. The IoU scores of features like cirque, FC, RG, valley and lake are 0.88, 0.94, 0.92, 0.90 and 0.97, respectively, for the case shown by the Figure 5.5.

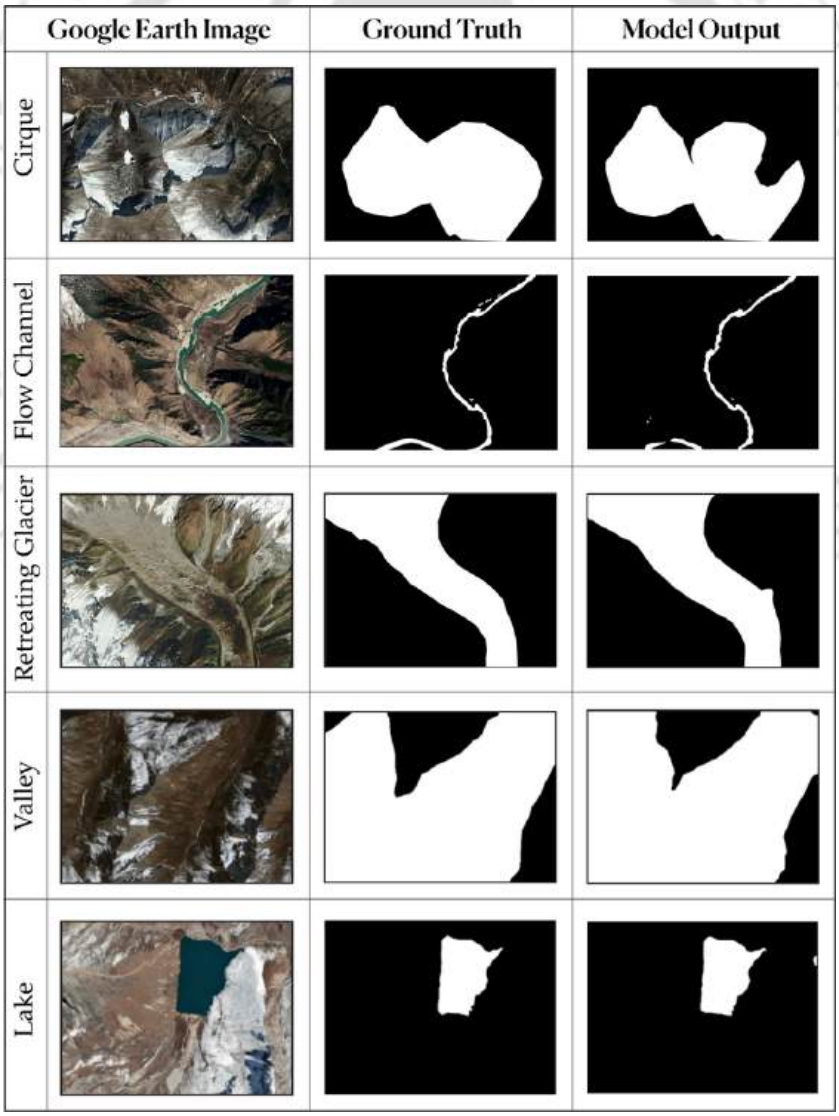


Figure 5.5: Detection of cirque, flow channel (FC), retreating glacier (RG), valley, and glacial lake by trained EfficientNet-B0 model for the validation set.

The next phase of the study analyses the collective presence of the extracted erosional features and the glacial lakes with thresholds to understand how these features, when occurring in specific combinations, indicate hazard classes and potential locations of future glacial lakes.

5.3.2 Hazard assessment and future lakes

This section implements the classification of detected lakes (true positive cases) and identifies the location of future lakes based on neighbourhood features, within a specified grid of image data. In a grid of 3 × 3 images, the neighbourhood is defined as 8 images around the central image. In locations where RG and FC appear in multiple connected images, the neighbourhood is extended to 5 × 5 images for the detection of the two features. A minimum lake area of 0.004 km² is detected, corresponding to at least 341 pixels on an image covering 3.5 km × 2.63 km at a resolution of 1024 × 768 pixels. For hazard classification by lake area, we adopted the criteria of Wang et al. (2015) and Washakh et al. (2019). Accordingly, a lake of area 0.05 km² or less is a small-sized lake and a lake larger than 0.1 km² is a large-sized lake. Further, we defined medium-sized lakes as having an area larger than 0.5 km² and smaller than 0.1 km². Therefore, we adopted 0.05 km² and 0.1 km² values for the lower threshold and the upper threshold of lake areas, respectively. Lakes are further classified into six sub-hazard classes according to heuristic rules discussed in the section 5.2.1.

5.4 Results and Discussion

This section is divided into two sub-sections 5.4.1 and 5.4.2. Section 5.4.1 discusses hazard assessment for glacial lakes. Next, sub-section 5.4.2 discusses future lake predictions in the study area.

5.4.1 Hazard Assessment of Glacial Lakes

The proposed method identified a total of 2647 glacial lakes (each of area > 0.004 km^2) using available Google Earth images (from years 2018-2022). Out of these, 2135 glacial lakes are active lakes, and the remaining 512 lakes are inactive. The identified glacial lakes (true positives scenarios or existing lakes) are further classified based on proposed heuristic criteria, incorporating the erosional features and geometric parameters (area and elevation) of lakes. Figure 5.6 show locations of hazardous lakes with different hazard levels in the ROI.

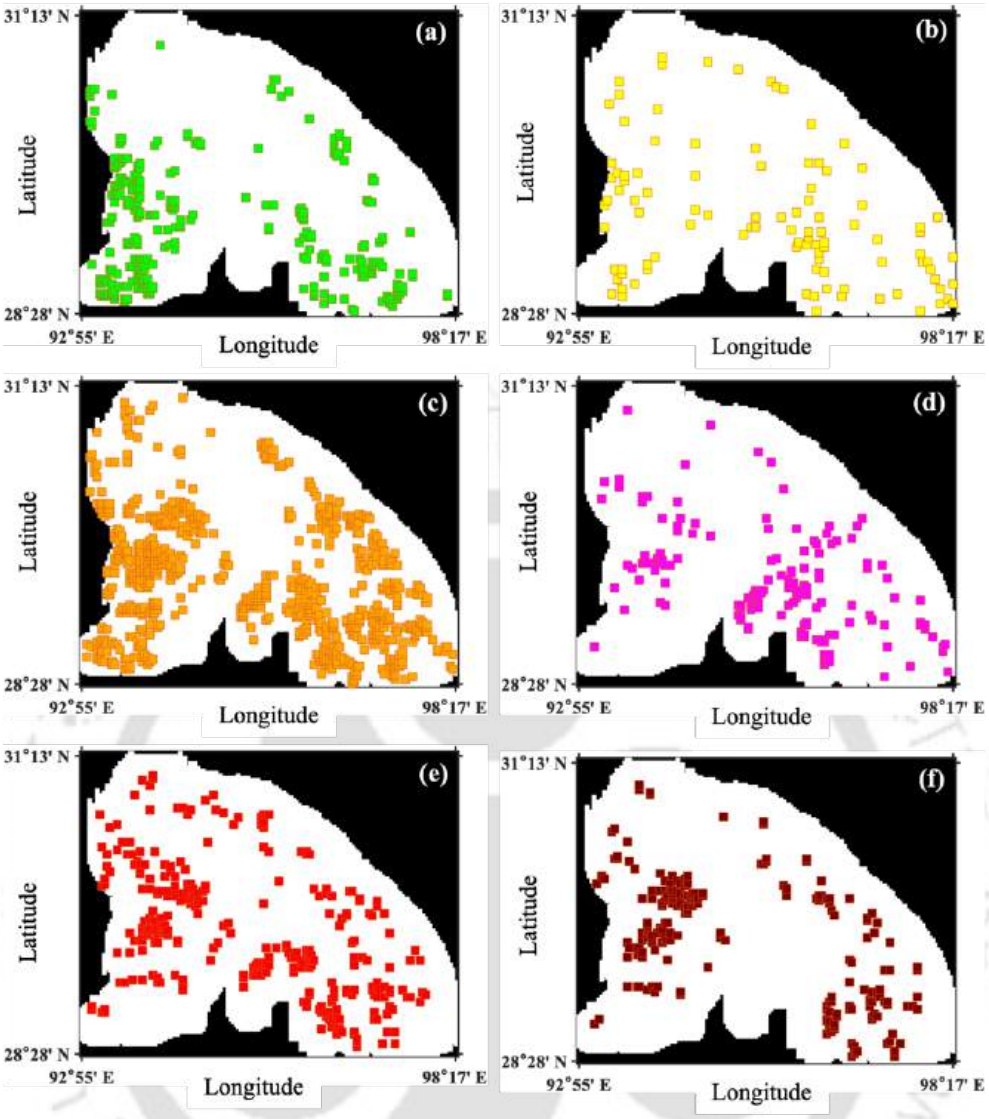


Figure 5.6: Spatial distribution of classified lakes: (a) non-dangerous (ND), (b) low dangerous (LD), (c) moderately dangerous (MD), (d) highly dangerous (HD), (e) very highly dangerous (VHD) and (f) extremely dangerous (ED) lakes.

Table 5.3 shows the distribution of lakes based on area and hazard levels.

Table 5.3: Distribution of lakes based on area and hazard level

Area	ND	LD	MD	HD	VHD	ED
Small-sized	512 (Inactive)	35 (Active)	908 (Active)	0	0	0
Medium-sized	-	39 (Inactive)	188 (Inactive)	147 (Active)	180 (Active)	0
Large-sized	-	25 (Inactive)	173 (18 Active; 155 Inactive)	28 (Active)	177 (Active)	235 (Active)
Total Lakes	512	99	1269	175	357	235

The following inferences are drawn from the results presented in Figure 5.6 and Table 5.3:

(i) Results indicate that 19.34% of the lakes (512 lakes) are classified as non-dangerous (Figure 5.6(a)), consisting entirely of small-sized inactive lakes with a total area of 12.1 km². Being stable or recently formed, these lakes, therefore, require 'very long-term monitoring'.

(ii) Ninety-nine (99 lakes or 3.74%) are categorized as low-dangerous (Figure 5.6(b)), covering a total area of 10 km², and 47.94% (1,269 lakes) are moderately dangerous (Figure 5.6(c)), spanning 74.9 km² area. Low-dangerous lakes uniformly consist of small, medium, and large-sized lakes, necessitating 'long-term monitoring'. Among these, 64 lakes are inactive, with no nearby GLHE, and are located either in valleys or other areas, while the remaining 35 are small, active lakes situated outside valleys without GLHE in the neighbourhood. In contrast, 'moderately dangerous' lakes are predominantly small, active lakes (908 in total) with 774 occurring in both valleys and cirques with nearby GLHE. The remaining 134 small lakes are located in valleys without GLHE. Additionally, this class includes 361 medium-sized and large-sized lakes, with 343 inactive lakes near GLHE, and 18 active lakes situated in locations outside the valley without nearby GLHE. The moderate hazard posed by these lakes necessitates short-term monitoring to mitigate potential risks.

(iii) Results further reveal that 6.61% (175 lakes) are highly dangerous (Figure 5.6(d)), covering an area of 18 km²; 13.48% (357 lakes) are very highly dangerous (Figure 5.6(e)), encompassing 109.1 km²; and 8.87% (235 lakes) are extremely dangerous (Figure 5.6(f)), with a total area of 71.34 km². Highly dangerous lakes are active, with 147 medium-sized lakes – 53 located in valleys without GLHE and 94 associated with GLHE but outside valleys. Additionally, 28 large lakes fall into this hazard class, all situated outside valleys without nearby GLHE. Given their significant hazard potential, highly dangerous lakes should be prioritized above moderately dangerous lakes and require regular monitoring.

(iv) Very highly dangerous lakes feature a nearly balanced mix of medium and large-sized active lakes, with 180 medium-sized lakes located in valleys with nearby GLHE, and 177 large-sized lakes split between 103 in valleys and 74 near GLHE but outside valleys. Further, the extremely dangerous lakes consist entirely of large-sized active lakes located in valleys with adjacent GLHE. Notably, very highly

dangerous and extremely dangerous lakes demand the highest priority for field investigations and necessitate immediate monitoring.

(v) Study by Ahmed et al. (2022a) identified 12 potentially dangerous glacial lakes (PDGL) situated in the valley of the Dibang river basin by using Landsat-8 and Sentinel 2A data (the authors numbered these lakes from 1 to 12). However, with the help of a high-resolution dataset (Google Earth images), our study assesses the hazard levels of these lakes and identifies 10 more lakes existing in the neighbourhood of these 12 PDGL locations. Characteristic details of these lakes are mentioned in Table 5.4 below.

Table 5.4: Hazard Assessment and characteristics of potentially dangerous glacial lakes

Lake	Latitude	Longitude	Area (km^2)	Elevation (m)	Hazard Class (Ahmed et al. (2022a)	Hazard Class (proposed approach)
1	29°23′31"	96°01′59"	0.121	4522	PDGL	VHD
2	29°23′21"	96°01′45"	0.032	4470	-	MD
3	28°57′02"	96°29′33"	0.133	4513	PDGL	VHD
4	28°57′30"	96°29′50"	0.065	4432	-	VHD
5	28°57′07"	96°33′40"	0.140	4296	PDGL	ED
6	28°56′53"	96°33′49"	0.051	4327	-	HD
7	29°22′28"	95°57′10"	0.142	4312	PDGL	ED
8	29°22′45"	95°57′21"	0.032	4387	-	MD
9	29°18′10"	96°09′24"	0.258	4198	PDGL	VHD
10	29°18′13"	96°08′37"	0.183	4393	-	VHD
11	29°18′40"	96°09′32"	0.047	4164	-	MD
12	29°20′06"	95°52′14"	0.215	4027	PDGL	ED
13	29°19′24"	95°51′26"	0.128	4194	-	ED
14	29°21′47"	96°07′23"	0.147	4015	PDGL	VHD
15	29°20′17"	96°05′03"	0.256	4311	PDGL	VHD
16	29°19′32"	96°05′57"	0.032	4336	-	MD
17	29°13′51"	95°58′59"	0.278	4178	PDGL	VHD
18	29°13′44"	95°57′48"	0.144	4313	-	HD
19	29°13′53"	95°59′32"	0.049	4266	-	MD
20	29°21′13"	95°55′42"	0.372	4373	PDGL	ED
21	29°21′10"	95°54′53"	0.616	4359	PDGL	ED
22	29°22′19"	95°52′22"	0.968	4375	PDGL	VHD

Out of these total 22 lakes, 6 lakes are extremely dangerous, as these are large-sized active lakes situated in valleys and connected with GLHE by FC. Further, 9 lakes are classified as very highly dangerous, consisting of medium and large size active lakes with 2 lakes situated in a valley without nearby GLHE, while the other 7 are located outside of valleys and are

connected to GLHE by small streams. Additionally, the study found 2 highly dangerous lakes and 5 moderately dangerous lakes in the close vicinity of PDGL. In the flow path of the PDGL numbers 1, 3, and 9 in Table 5.4, a lake with moderate to very highly dangerous potential is occurring within a maximum distance of 450 m. Also, except for the PDGL numbers 14, 20, 21, and 22, all lakes are connected to a GLHE through FC. In addition, currently, PDGL number 15 has a neighbouring lake (number 16), which is moderately dangerous, but it is not connected with the former one.

Small and medium-sized lakes with an area of less than 0.1 km², occurring in a flow path from an upstream lake or having a lake in its downstream flow path, can pose significant risks to human settlements and localities downstream, particularly when lakes are combined with hazardous processes such as cloudbursts, ice avalanches, and landslides. This was evident in the 2013 Chorabari disaster in Uttarakhand due to heavy rainfall, resulting in over 6,000 fatalities and extensive damage to infrastructure and property (Mehta et al., 2012; Rafiq et al., 2019). We have identified a total of 293 locations of human settlements occurring in the ROI, and 40 lakes are situated near these settlements within a 3 × 3 grid space. Figure 5.7 shows the grids of spatial occurrence of total settlement and neighbouring 40 lakes in the ROI.

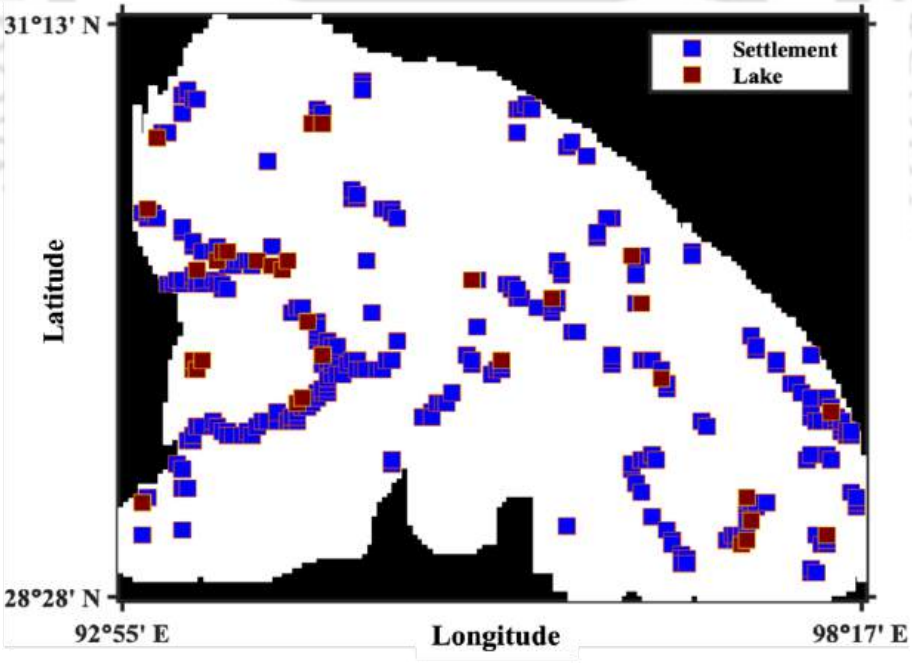


Figure 5.7: Grid based spatial representation of total settlement and 40 lakes near the settlement.

Out of the 40 lakes, 17 are non-dangerous, 4 are low-dangerous, 9 are moderately dangerous, 2 are highly dangerous, 5 are very highly dangerous, and 3 are extremely dangerous

lakes. Lakes falling in the dangerous category (23 out of 40 lakes) require regular monitoring to adopt necessary precautionary measures to safeguard people living downstream.

5.4.2 Future Lakes

Prediction of locations of future lakes in the ROI originated by observing 2638 numbers of glacial lakes detected in 1540 image tiles. The presence of any of the three erosional features with GLHE leads to the formation of a future lake. This rule devises five conditions by combining a minimum of three erosional features and GLHE for the prediction of future lake locations. Among 2638 lakes in 1540 image tiles, 1982 numbers of existing lakes (almost 75.12 %) occupying 1076 image tiles, are covered by these five conditions. Further, we identified 2,738 locations in the test dataset (across 2,738 images) where water reaches and yet the lake has not started to form. Table 5.5 shows the distribution of locations (image tiles) for five governing conditions.

Table 5.5: Grid count of existing and future lakes based on geomorphic features (lake count is shown in brackets for the existing lake grid)

Geomorphic features in neighbourhood tiles	Number of grids (image tiles) of existing lakes	Number of grids (image tiles) of future lakes
Cirque, valley, FC, RG and GLHE	452 (882)	1206
Cirque, FC, RG and GLHE	125 (238)	287
Valley, FC, RG and GLHE	17 (27)	66
Valley, cirque, RG and GLHE	248 (407)	514
Valley, cirque, FC and GLHE	234 (428)	665
Sum of grids or tiles:	1076	2738

Out of 2738 image tiles for future lakes, 1206 image tiles show locations where all the governing features (presence of cirque, valley, FC, RG and GLHE) are present in their neighbourhood. Each of the other four conditions contains less numbers of future lakes - 287 images (for presence of cirque, FC, RG and GLHE), 66 images (for presence of valley, FC, RG and GLHE), 514 images (for presence of valley, cirque, RG and GLHE), and 665 images (for presence of valley, cirque, FC and GLHE) are the remaining possible locations of future lakes.

Analysing flow direction in local neighbourhoods at each of the 2738 image tiles using SRTM DEM in the test dataset reveals that only 174 locations of the lowest elevations are detected, where water will first accumulate and start forming the future lakes. Furthermore,

visualising these locations in Google Earth Pro further revealed moraine formations in these locations, supporting the formation of future lakes. Notably, these 174 locations form a precise subset of the 334 high-probability lake formation locations predicted by all three models (LR, ANN and BNN), as presented in Chapter 3 (see page 33). Figure 5.8 presents the locations and distribution of 174 numbers of predicted future lakes in the ROI.

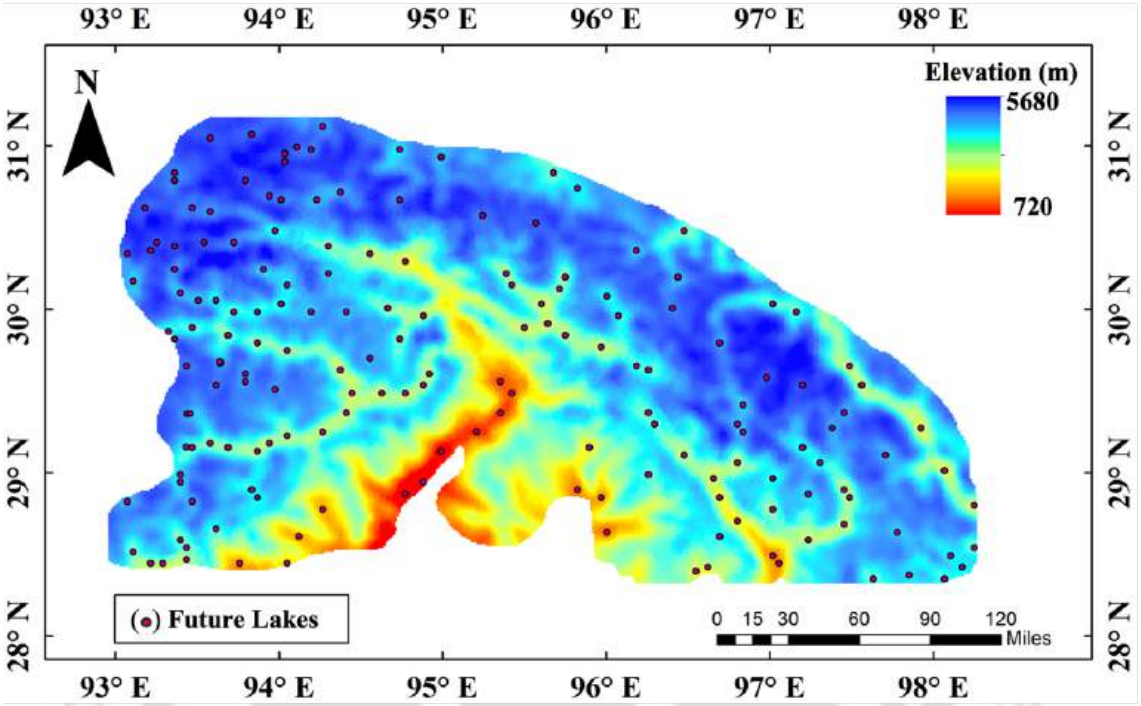


Figure 5.8: Future lakes (174 in numbers) marked on SRTM DEM.

5.5 Chapter Conclusion

This chapter presents a semi-automated framework for glacial lake hazard assessment based on the spatial and morphological characteristics of associated erosional features. The proposed method incorporates the cumulative influence of four key geomorphic features–cirques, valleys, flow channels, and retreating glaciers–that govern glacial lake formation and evolution. These erosional features are extracted from open-source satellite imagery (Google Earth Pro) using the EfficientNet-B0 CNN model, and the resulting outputs are used to assess the hazard level of glacial lakes.

The CNN model is trained by marking erosional features and glacial lakes on image tiles, containing both feature presence and feature absence scenarios. The EfficientNet-B0 provided reliable accuracy with (IoU) scores of 0.97, 0.88, 0.90, 0.96, and 0.92 for lakes, cirques, valleys, FC, and RG, respectively, on the validation set. The variation in accuracy is due to irregular shapes, varying widths, varying depths and subtle boundaries that merge into the surrounding

terrain. The CNN model was then processed on a test dataset of 12,924 image tiles for extracting erosional features and glacial lakes, achieving accuracies of 93.26%, 91.31%, 86.36%, 92.67%, and 92.30% for glacial lakes, cirques, valleys, FC, and RG, respectively. Moreover, the CNN identified 2647 glacial lakes in test images of the study site located in the Eastern Himalayas.

Next, the simulation exercise assesses the hazard levels of 2647 identified glacial lakes (each with an area > 0.004 km^2) by heuristic rules, analysing the presence of erosional features (RG, GLHE, and valley) and geometric parameters of the lake (area and elevation). Out of these lakes, 2135 are active, and the remaining are inactive lakes. Further, the hazard assessment classified 235 lakes as 'extremely dangerous', 357 lakes as 'very highly dangerous', 175 lakes as 'highly dangerous', 1,269 lakes as 'moderately dangerous', 99 lakes as 'low dangerous', and 512 lakes as 'non-dangerous'. Notably, extremely dangerous lakes consist entirely of large-sized active lakes located in valleys with adjacent GLHE. Moreover, 23 dangerous lakes are occurring near the settlements in the ROI. Importantly, lakes classified as extremely dangerous and very highly dangerous require immediate monitoring due to their large size and location in valleys with nearby GLHE, which increases their hazard potential significantly. Similarly, the remaining lakes need to be regularly monitored. Furthermore, extending the derived analogy of the co-occurrence of key erosional features, including valleys, cirques, RG, FC, and GLHE in the surroundings of existing lakes and further refined by considering the local minimum elevation for flow accumulation and moraine formations, 174 new locations are identified where future lakes may occur.

This chapter discussed hazard classification of glacial lakes based on intrinsic geomorphic and geometric factors, yet it avoids external factors—all of which influence the overall risk posed by glacial lakes. Next Chapter 6 includes external hazard drivers such as historical GLOFs, co-occurring landslides, seismic activity, and anthropogenic exposure and expands the assessment framework by incorporating these multi-hazard and risk parameters across an extended ROI. A hybrid MCDA approach is employed to classify lakes into risk categories, thereby transitioning from hazard-focused evaluation to a GLOF risk assessment framework.

Chapter 6

GLOF Risk Assessment

6.1 Introduction

As explained in Chapter 5, the hazard potential of glacial lakes in the ROI is strongly shaped by geomorphic and erosional features such as cirques, valleys, retreating glaciers (RG), and flow channels (FC). These factors play a central role in determining where lakes form and how they evolve. However, these facyores exclude the full range of external processes that may influence lake stability. Dynamic triggers such as landslides, ice–rock avalanches, and seismic shaking can, in principle, modify dam conditions or introduce sudden mass inputs that elevate outburst potential. These triggers are particularly relevant in tectonically active regions of the Eastern Himalaya, where steep topography and high relief make slopes sensitive to disturbance (Afroz A Shah, 2013; Shah et al., 2025; Shah et al., 2024a; Shah et al., 2024b).

Earthquakes are often suspected triggers because ground shaking can destabilise dams or weaken ice–rock interfaces, potentially leading to GLOFs, especially in tectonically active areas like the Himalayas and the Andes (e.g., Veh et al., 2020). However, recent research indicates that earthquakes might be overestimated as triggers. Wood et al. (2024) found that, although many earthquakes affect glacierized regions, only one GLOF has been confidently linked to seismic activity. While this does not prove a direct causal link, a significant earthquake occurring within the shallow crustal depths (0-20 km) near GLOF sites could destabilise lakes and trigger flooding. This potential connection is supported by mapping analyses and the relationships between earthquake faults and GLOF locations (Figure 6.2 and Figure 6.5).

Evidence from past Himalayan earthquakes–including the 1950 Tibet–Assam event (Mw 8.6) and the 2015 Gorkha earthquake (Mw 7.8)–shows that strong shaking can induce widespread landslides and destabilise unconsolidated sediment in high mountain catchments (Bilham et al., 2017; Kargel et al., 2016; Priyanka et al., 2017; Roback et al., 2018). While clear, documented cases of earthquake-triggered GLOFs remain rare in the published literature, seismicity may still act as an indirect modifier of glacial lake stability through its effects on surrounding slopes, moraine material, and sediment flux. For this reason, seismic and tectonic indicators are considered in this study as contextual hazard drivers rather than primary triggers.

Several glacial lakes in the ROI also lie close to major Himalayan structures such as the Main Central Thrust (MCT), Main Boundary Thrust (MBT), and Main Frontal Thrust (MFT), which mark zones of active deformation and heightened seismic hazard (Shah et al., 2025). These regions are characterised by high rates of landslide activity and slope instability, which can amplify local hazard conditions around lakes (Keefer, 1984). Taken together, these factors highlight that GLOF hazard is shaped not only by intrinsic lake characteristics and geomorphic setting but also by the surrounding landscape dynamics, including landslide-prone slopes, tectonic activity, and the spatial distribution of past natural hazard events.

Notably, Chapter 5 only classifies the hazard state of a lake without identifying its potential magnitude and impact. Therefore, a shift from hazard assessment to risk assessment is essential for a comprehensive evaluation of GLOF threats in this seismically and geomorphologically active region, where population clusters and critical infrastructure are increasingly encroaching upon hazard-prone zones. The risk assessment in this chapter accounts for both the potential of GLOF occurrence and the exposure of a glacial lake to vulnerable elements, such as downstream built-up areas and infrastructure.

As seismic factors (proximity to faults, earthquake locations, landslide locations, etc) affect a large area, the study area is expanded. The expanded study area is named ‘extended ROI’. The work discussed in this chapter has been published in (Vashistha et al., 2025b). Next sections describe extended ROI, methodology, results, discussion, and chapter conclusion.

6.2 Extended Region of Interest (Extended ROI)

The Extended-ROI for this study is geographically bounded between 21.63°N to 31.36°N latitude and 85°E to 98.73°E longitude, covering an area of approximately 147.12 million hectares (14,71,200 km^2) in the Eastern Himalaya. This region spans across northeastern India, Bhutan, Nepal, northern Myanmar, and the southeastern Tibetan Plateau, and represents a critical transitional zone between the high-altitude Tibetan Plateau and the tropical lowlands of Southeast Asia. Topographically, the ROI exhibits an elevation gradient ranging from -117m in the southern floodplains to 8781m in the northern mountainous terrain, encompassing Kangchenjunga (8,586 m), the third-highest mountain peak in the world. The region is tectonically active, located at the intersection of the Indian, Eurasian, and Burmese plates, and is traversed by major fault systems such as the Main Central Thrust (MCT), Main Boundary Thrust (MBT), and the Eastern Himalayan Syntaxis. This complex tectonic configuration places the region predominantly in Seismic Zone V, the highest hazard zone classified by the

Bureau of Indian Standards (Nath et al., 2008). Notable historical earthquakes, including the 1897 Assam and 1950 Assam–Tibet events (both > M8.0), underscore the region's seismic vulnerability (Bilham et al., 2001; England et al., 2015; I. Singh et al., 2021). Climatically, the ROI is influenced by a humid, monsoon-dominated system, receiving most of its precipitation between June and September. Compared to the Western Himalaya, the region hosts smaller, fragmented, and debris-covered glaciers, situated on steep slopes. Approximately 35–40% of the area remains snow-covered during summer (Krishna, 2005), supporting numerous glacial and periglacial features. This ROI is undergoing significant and disproportionate warming, with an observed increase in annual mean temperature of approximately 0.01°C per year (Li et al., 2016). This climatic trend is accelerating glacial retreat and contributing to the rapid formation and expansion of glacial lakes. Hydrologically, the Eastern Himalaya serves as the origin of major transboundary rivers, including the Brahmaputra, Teesta, and Manas, which are vital for irrigation, hydropower, and water security across downstream regions in India, Bhutan, Bangladesh, and China. The interplay of climate-induced changes, fragile glaciology, and tectonic instability heightens the region's susceptibility to GLOFs, making it a focal point for both climate resilience planning and hazard risk assessments. Figure 6.1 illustrates the spatial extent of the study area using SRTM 30-meter resolution DEM data.

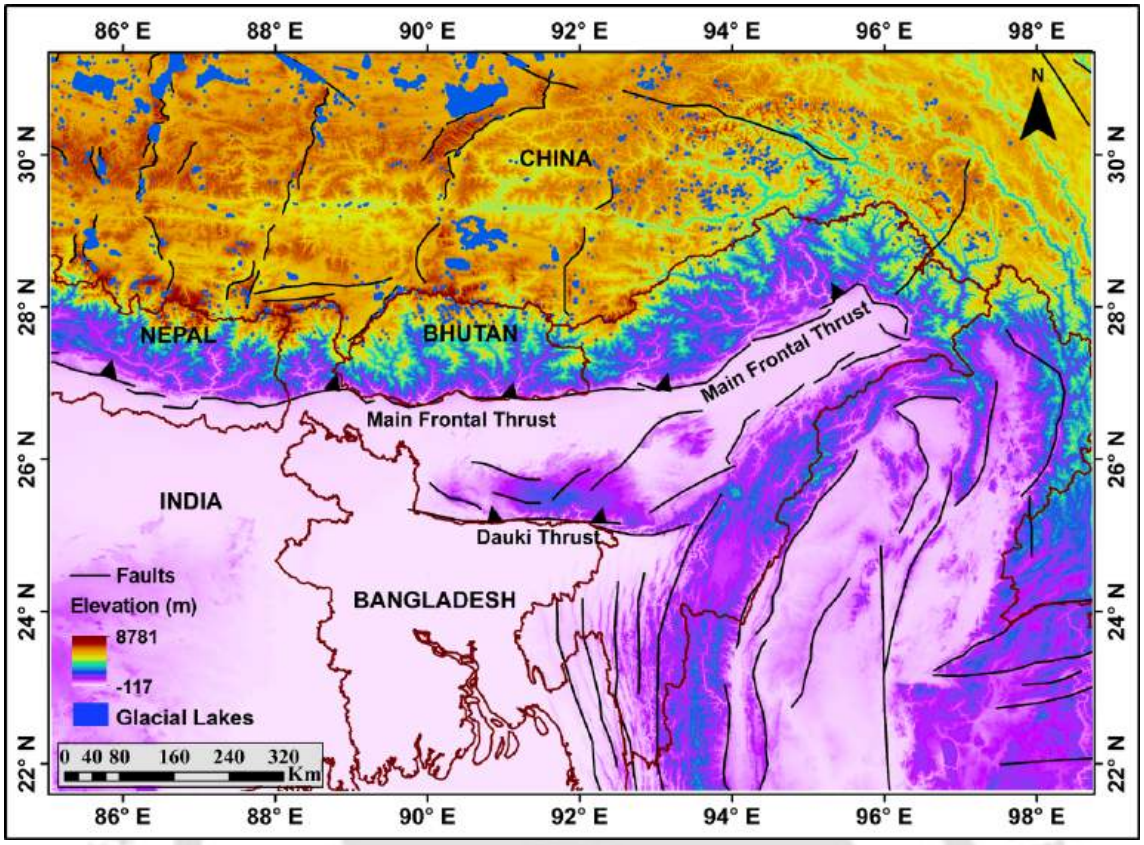


Figure 6.1: Topographic and structural map of the Eastern Himalayan region, including parts of India, Nepal, Bhutan, China, and Bangladesh., derived from SRTM 30-meter resolution DEM data. Elevation ranges from –117 m in the south to 8781 m in the high Himalayan ranges. Major fault structures, including the Main Frontal Thrust (MFT) and Dauki Thrust, are shown as black lines, illustrating the region's tectonic complexity. Glacial lakes, indicated in blue, are predominantly concentrated in high-altitude zones of Nepal, Bhutan, and southeastern Tibet, highlighting areas of potential GLOF risk. The region lies at the convergence of the Indian, Eurasian, and Burmese plates, making it one of the most seismically active zones in the world.

6.3 Methodology and materials

6.3.1 Parameter selection and data generation

Among the geometric parameters of glacial lakes, lake elevation, area of lake, volume of lake, and slope influence GLOF events (Ahmed et al., 2022a). The elevation, area, and volume of a lake directly influence the stability and potential for GLOF events. High-altitude lakes are more hazardous as the potential energy increases with the rise in elevation, which leads to catastrophic outbursts once the breach occurs. Further, larger lakes with significant volumes store more water, which increases the peak discharge during the flooding. Simultaneously, if the area of a lake is larger while the volume is constant, it indicates a shallower depth, suggesting weaker moraine or bedrock constraints and higher susceptibility to wave-induced erosion and thermal exchange. Increased solar radiation absorption in large area lakes

accelerates surface warming, leading to faster glacial melt and potential expansion of the lake. Additionally, warmer water infiltrates moraine dams, weakening their structure and increasing the risk of GLOFs (Harrison et al., 2018). The following are factors associated with seismicity and tectonics, which are observed evidence and included logically to develop the model for the risk assessment:

(i) Proximity to faults: Lakes located near active faults are at higher risk of disruption due to seismic activity. Moreover, fault movement can lead to slope instability, lake expansion, or sudden outbursts, increasing the likelihood of GLOFs. As the influence of fault displacement will be maximum on the closest lake, this thesis considers the influence of the distance between the lakes and the nearest major faults on GLOF risk assessment. Figure 6.2 presents the mapping of faults over the study area. The focal plane solution reveals that thrusting is more prominent on the eastern portion, whereas normal faulting is observed in the central region.

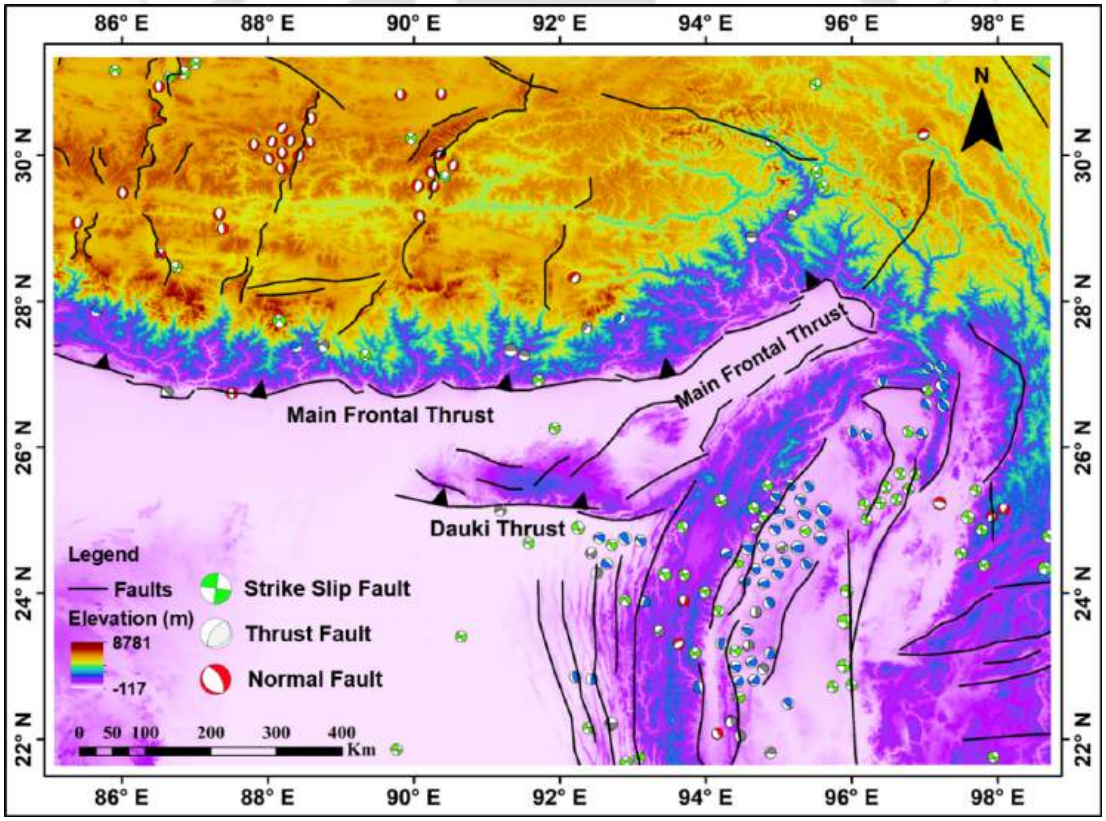


Figure 6.2: The figures illustrate the topography of the Eastern Himalayas, with the distribution of active faults (adapted from Taylor and Yin, 2009) overlaid. Additionally, the earthquake centroid moment tensor (CMT) events are plotted to highlight the varying types of seismic activity in the region, ranging from normal faults to strike-slip faults. These visualizations provide a comprehensive view of the tectonic complexity and the nature of earthquake mechanisms in this geologically dynamic area.

(ii) Elevation difference with landslide location: Elevation difference between a lake and the nearby landslide location also influences the GLOF events. When the elevation difference is small, the landslide debris reaches the lake quickly with minimal energy dissipation, resulting in a sudden and intense impact. This can generate large waves, known as impulse waves, which may overtop natural or artificial lake barriers, increasing the likelihood of a GLOF. In contrast, when the elevation difference is large, debris travels for a longer distance, allowing energy dissipation due to friction, terrain obstacles, and dispersion. As a result, the impact force on the lake is generally reduced, lowering the chances of immediate dam failure. While a greater elevation difference can lead to higher acceleration and velocity of the debris, the energy loss along the slope often mitigates the final impact compared to a shorter, more direct collapse. Therefore, lakes located close to landslide-prone zones with small elevation differences are generally at a higher risk of sudden outburst events. The data for landslide locations is obtained through Kirschbaum et al. (2015) and Kirschbaum et al. (2010). The elevation difference is measured by using the SRTM 30 m DEM. Figure 6.2 presents a spatial distribution of landslides across the study area (extended ROI) and adjoining regions, classified by triggering factors such as earthquakes, cyclones, freeze-thaw processes, rain, floods, human activity, and others.

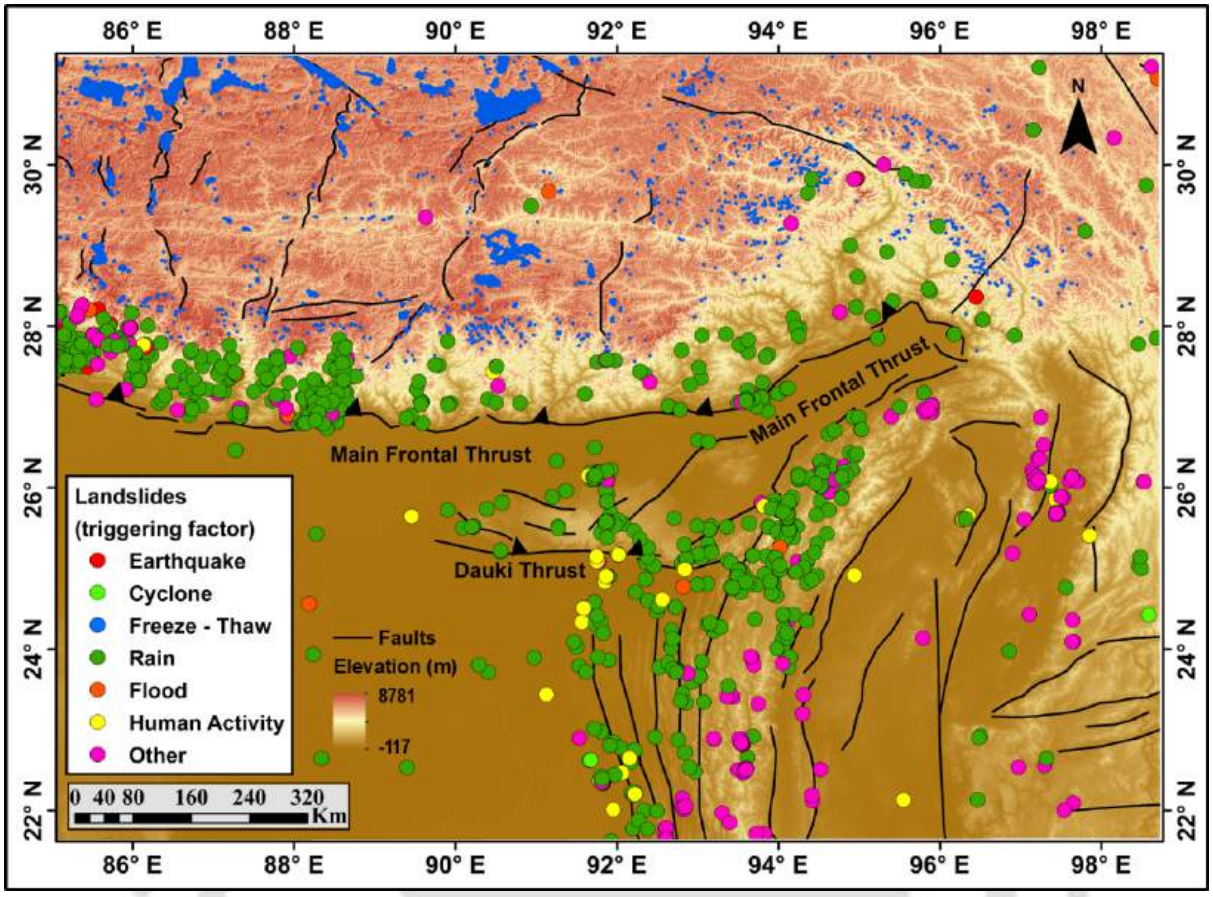


Figure 6.3: Spatial distribution of landslides in the Eastern Himalayas and adjoining regions, classified by triggering factors—earthquake, cyclone, freeze-thaw, rain, flood, human activity, and others—overlaid on a shaded elevation map (−117 m to 8781 m). Major active faults, including the Main Frontal Thrust (MFT) and Dauki Thrust, are marked, showing strong spatial correspondence with landslide occurrences. Rain-induced landslides dominate the region, particularly along the MFT, while earthquake- and human activity-triggered landslides are prominent in seismically and anthropogenically disturbed zones. The figure highlights the interaction between topography, active tectonics, and climatic conditions in governing slope instability across the Himalayan front.

The topography in Figure 6.3 above, shown via a shaded elevation map ranging from –117 m to 8781 m, indicates that most landslides are concentrated along the Himalayan front and interior regions, particularly along the Main Frontal Thrust (MFT) and the Dauki Thrust—two major tectonic structures. Rain-induced landslides (green dots) dominate the dataset, showing high density along the MFT and extending eastwards across Arunachal Pradesh. Earthquake-triggered landslides (red) are less frequent but appear in proximity to known faults, especially near the eastern syntaxis and central MFT. Human activity (yellow) and flood-triggered (orange) landslides are concentrated in the southeastern part, indicating anthropogenic influence and floodplain instability. Cyclone-triggered (green) and freeze-thaw-related (blue) events are sparse, suggesting regionally limited climatic influence. The spatial clustering of

landslides along fault zones underscores the role of active tectonics, particularly the MFT and Dauki Thrust, in preconditioning slopes for failure. The map highlights how both tectonic and climatic drivers interact with topography to control landslide distribution across this seismically active and densely populated mountain front.

(iii) Proximity to landslide locations: The proximity of a lake to landslide-prone areas is a critical parameter for determining its susceptibility to sudden destabilization and potential outburst floods (Kirschbaum et al., 2015; Kirschbaum et al., 2010). Lakes closer to active landslide zones are at a higher risk of being directly impacted by mass movements, which can trigger large impulse waves capable of overtopping or breaching the lake's natural dam. The closer the landslide source, the greater the probability of significant wave generation, leading to increased flood risk (Figure 6.3). In contrast, lakes located farther from landslide-prone regions have a lower likelihood of direct impact, as the debris may dissipate its energy before reaching the water body (Figure 6.3).

(iv) Proximity with earthquake locations: The study site in the Eastern Himalaya encompasses five of the most seismically active regions: the Eastern Syntaxis, the Arakan Yoma Belt, the Shillong Plateau (SP), the Himalayan Frontal Arc (HFA), and South Tibet. Earthquakes in these regions are predominantly of shallow depth, with intermediate magnitudes and spatially variable frequencies. Between 1964 and 2020, the Eastern Syntaxis experienced 247 recorded earthquakes, the Arakan Yoma Belt 727 earthquakes, the Shillong Plateau 83 earthquakes, the HFA 166 earthquakes, and South Tibet 124 earthquakes (Tiwari et al., 2022). The prevalence of shallow, moderate-magnitude earthquakes poses a significant threat to glacial lake stability, as seismic shaking can trigger moraine dam failures or landslides, leading to GLOF events, as observed from the 2015 Gorkha earthquake (Cook et al., 2018). Seismic activity in the vicinity of glacial lakes can induce ground shaking, destabilizing moraine dams and triggering lake outbursts. Strong earthquakes can generate surface ruptures, liquefy sediments, and create cracks in natural dams, increasing seepage and the likelihood of structural failure. Additionally, seismic events can indirectly impact lakes by triggering avalanches or ice calving, leading to sudden water displacement and impulse waves. Therefore, the proximity of lakes to earthquake hotspots is a critical factor influencing their stability and potential for outburst floods. The closer a lake is to a historically active earthquake zone, the higher the probability of seismic-induced failure, making it essential to integrate seismic hazard assessments into GLOF risk evaluations.

(v) Magnitude and depth of nearest earthquakes: The magnitude and depth of an earthquake are crucial in assessing the risk to glacial lakes. Shallow, high-magnitude earthquakes pose a greater threat, as they can trigger moraine failures or destabilize surrounding slopes, significantly increasing GLOF risks. Larger magnitude earthquakes generate stronger ground shaking, potentially destabilizing lake banks and triggering landslides or GLOFs. Shallow earthquakes, occurring at depths less than 70 km, tend to cause more intense surface shaking, increasing the likelihood of landslides or lake ruptures. Deeper earthquakes, though less intense at the surface, can still affect the region's tectonic stability, influencing lake dynamics. For earthquake locations, magnitude, and intensity, updated earthquake data from 1960 to 2024 were sourced from the United States Geological Survey (USGS) database to ensure a comprehensive evaluation. Accordingly, the proximity of glacial lakes to earthquakes is estimated. Figure 6.4 presents the spatial distribution of earthquakes with their magnitude and depth variation over the ROI.

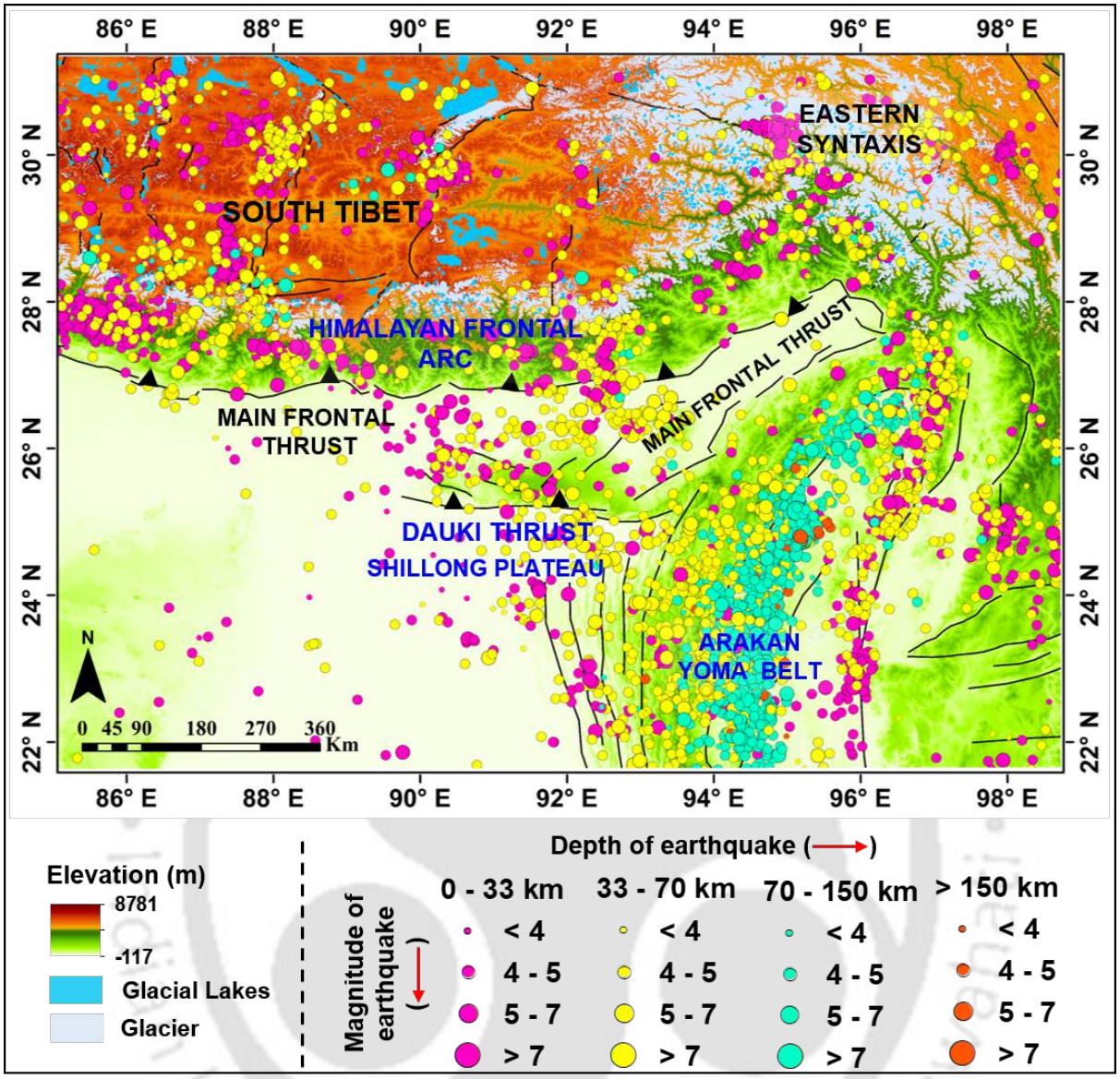


Figure 6.4: Seismotectonic Map of the Eastern Himalaya and Indo-Burmese Arc: Depth-coded and magnitude-scaled earthquake epicentres plotted over shaded relief topography highlight active tectonic boundaries, including the Main Frontal Thrust (MFT) and the Dauki Thrust. The figure shows the distribution of earthquakes, which indicates clustering in the eastern region with some earthquakes >70-100 km, indicating active subduction. Shallow earthquakes (less than 70 km) are concentrated in the Himalayan regions, typical of collisional tectonics. The shallow crustal seismicity initiates ruptures, shaking, and landslides and could be dangerous for GLOFs (this study). Dense seismic clustering in the Indo-Burmese region marks active oblique subduction of the Indian plate beneath the Burma microplate. The variation in earthquake depths and topography illustrates complex plate interactions and crustal deformation across this highly seismically active region.

(vi) Proximity to historic GLOF locations: The nearest distance of lakes from previously occurred GLOF locations is important for assessing the susceptibility of lakes to future outburst events. Lakes in proximity to past GLOF sites may share similar geomorphological, climatic and hydrological conditions, making them more vulnerable to similar triggering mechanisms.

Additionally, past GLOFs changes the terrain, weakening moraine dams and creating pathways for future flood events. Therefore, understanding this distance helps in risk assessment, prioritizing lakes that may be influenced by cascading flood effects or connected hydrological systems. The locations of past GLOF events are obtained by reviewing the Lützow et al. (2023) and T. Zhang et al. (2022b) study, identifying 75 GLOF events within the ROI. Figure 6.5 presents the locations of GLOF with their dam characteristics for reference. Most GLOF-affected lakes are moraine-dam, and their occurrence is closely associated with areas around the main frontal thrust, specifically along the normal faults.

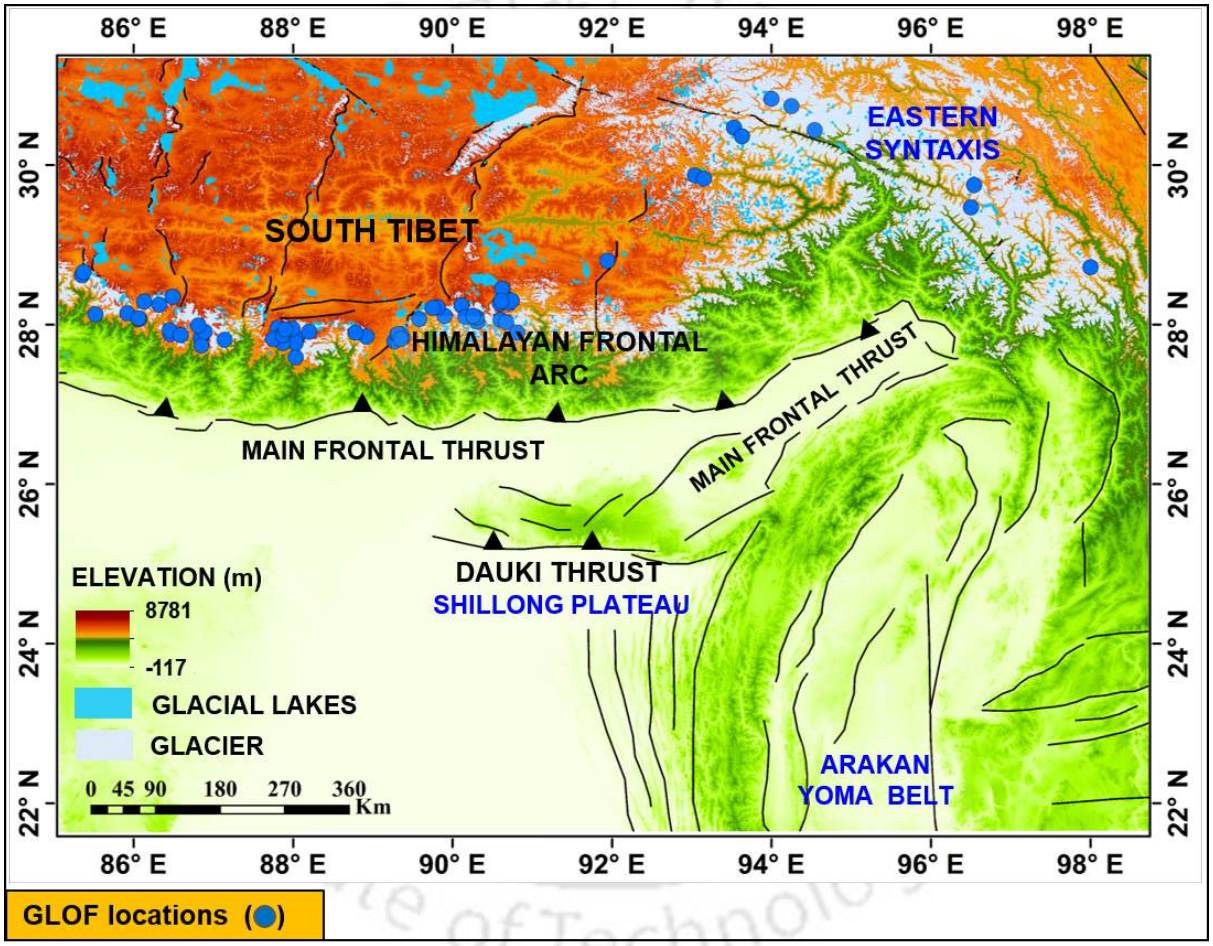


Figure 6.5: Mapping of historically occurred 75 GLOFs in the Extended-ROI. The map also displays lake and glacier extents along with major tectonic structures.

(vii) Slope: The slope around a glacial lake influences water flow, sediment deposition, and the potential for landslides. Further, steeper slopes near glacial lakes can promote rapid downstream movement of floodwaters, increasing their erosive power and destructive impact. For instance, Govindha Raj (2010) found that glacial lakes with downstream slopes greater than 10° are potentially dangerous. Che et al. (2014) suggested that a steep slope (> 8°) between the lake and glacier increases the risk of calving and outburst. Therefore, this study considers

the steepest slope surrounding the lake within a 500 m buffer zone from the outside of the lake. The buffer zone limit is obtained from the Kougkoulos et al. (2018) study. This parameter was estimated from the DEM by using the 'slope analy*sis' and 'buffer' tool* in GIS.

(viii) Proximity to built-up: The distance of lakes from the nearest built-up is crucial in GLOF risk assessment as it determines the potential impact on human settlements, infrastructure, and evacuation time. Lakes closer to urban areas pose a higher risk due to faster flood propagation, greater damage potential, and limited response time. This factor is integrated into hybrid MCDA models to classify lakes into risk categories, aiding in disaster preparedness and urban planning. Shorter distances increase vulnerability, while greater distances allow flood energy dissipation and better emergency response.

(ix) Land cover type: Land cover type significantly affects hydrological dynamics, including surface runoff, infiltration, and water retention. Vegetated areas contribute to slope stability and reduce runoff velocity, thereby limiting erosion and sediment delivery into lakes. In contrast, barren or impervious surfaces tend to accelerate runoff, potentially increasing lake volume and the likelihood of flooding. This thesis uses a classified land cover map (Figure 6.6) of the Eastern Himalaya comprising nine categories: water, built-up areas, rangeland, bare ground, flooded vegetation, trees, crops, ice, and cloud cover. For each lake, the dominant land cover type within a 500-metre buffer is identified using Sentinel-2 land cover data at 10-metre spatial resolution. This analysis was conducted in ArcGIS using the 'Buffer' tool. Figure 6.6 shows the spatial distribution of land cover classes overlaid with major faults across the region.

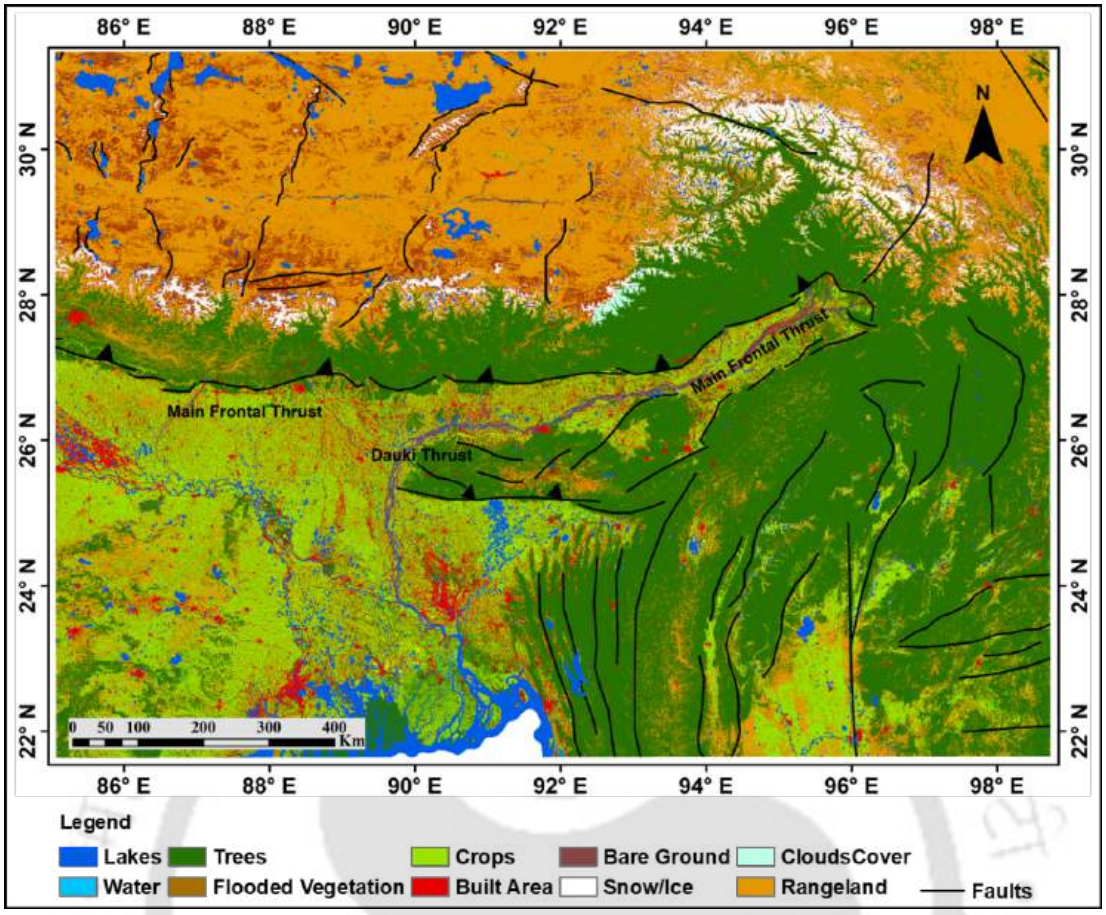


Figure 6.6: Land cover classification across the Eastern Himalaya and adjoining foreland, based on Sentinel-2 imagery at 10-metre spatial resolution. The map shows nine land cover categories, including forest, cropland, built-up area, water bodies, bare ground, ice/snow, and others, overlaid with major tectonic structures such as the Main Frontal Thrust and Dauki Thrust. The distribution of land cover types reflects the influence of elevation, tectonic activity, and human land use, with implications for hydrological processes, ecological stability, and vulnerability to GLOFs.

The landscape is characterised by sharp transitions in vegetation and surface cover across prominent tectonic boundaries, notably the Main Frontal Thrust (MFT) and the Dauki Thrust. Alpine and high-altitude zones north of the MFT are dominated by barren land (orange), exposed rock, and patches of permanent snow cover (white), reflecting glaciated and periglacial environments typical of high Himalayan elevations. Dense forest cover (dark green) blankets the mid-slope regions south of the high mountains, providing stability to hillslopes and playing a significant role in modulating hydrological responses. In contrast, the foothills and foreland basins south of the MFT, including the Brahmaputra and Ganga plains, are dominated by cropland (light green), interspersed with built-up areas (red) and surface water bodies (blue). These lower-elevation regions are more densely populated and are hydrologically connected to upstream glacial lake systems. The presence of impervious surfaces and agricultural land increases surface runoff and reduces infiltration capacity,

potentially exacerbating flood hazards downstream in the event of glacial lake outbursts. The spatial alignment of major faults and thrusts with sharp land cover transitions underscores the role of tectonic uplift and geomorphic partitioning in shaping ecological zones. Notably, the Dauki Thrust, bounding the Shillong Plateau, separates forested uplands from densely cultivated plains, reflecting a strong interaction between tectonic processes, topographic gradients, and land use patterns.

In view of the above evidence, a total of 13 spatial parameters, including lake elevation, area of lake, volume of lake, slope, land cover type, elevation difference between the lakes and the nearest landslide's location, and the nearest distance of lakes from faults, landslides, built-up areas, and GLOF locations, are selected. Next, this chapter proposes a hybrid MCDA framework for the numerical risk assessment of GLOF. Accordingly, the following are challenges and measures adopted for implementing a robust risk assessment:

a) The input parameters are not only high-dimensional and heterogeneous in nature but also differ in scale, units, and physical interpretation—some promote hazard (positive indicators), while others reduce it (negative indicators). To ensure comparability and coherent integration into a single model, all parameters are normalized (both positive and negative indicators), reflecting their contribution to GLOF risk.

b) Multicollinearity among parameters, such as between slope and elevation, can distort weight assignment and affect model robustness. In addition, assigning appropriate parameter weights is also difficult due to the contrast between subjective expert judgment in the literature and objective statistical information. To resolve this, a hybrid weighing approach is employed, combining AHP and SE through KL divergences, thus balancing domain knowledge with data-driven evidence.

c) The Absence of ground-truth risk labels makes supervised learning infeasible. Hence, unsupervised classification is adopted, where risk scores are computed using the TOPSIS. Further, risk scores are categorized into risk levels via FCM clustering. The reliability of this classification is then tested using α-cut sensitivity analysis and a validation test across 30 PDGLs, which are reported in the literature.

6.3.2 Data generation and procedure

Elevation, Area, Volume of Lake: This thesis uses the HydroLAKES v1.0 database to evaluate risk parameters like elevation, area, and volume of the lakes (Messager et al., 2016).

Notably, the study only considers lakes with an area greater than or equal to 0.1 km2 and greater than or equal to 2450m for risk assessment through MCDA, as these lakes are potentially dangerous (Wang et al., 2015; T. Zhang et al., 2022b).

6.3.3 Data processing of qualitative data

The 'land cover type' variable constitutes qualitative data, in contrast to the remaining variables, which are quantitative in nature. To facilitate integration into the analytical framework for GLOF risk assessment, the qualitative land cover data were converted into an ordinal ranking system. This ranking, ranging from 1 to 9, reflects the relative susceptibility of each land cover category to GLOF-related impacts in the vicinity of glacial lakes, with higher values denoting greater risk (i.e. 1 indicates the lowest risk; 9 indicates the highest risk).

Water bodies (assigned a rank of 9) represent the highest risk category, as they facilitate flood propagation by increasing volume and extending the inundation area (Huggel et al., 2002). Built-up areas (rank 8) are particularly exposed due to the vulnerability of infrastructure and human populations, as evidenced by the 2013 Kedarnath disaster (Carrivick et al., 2016). Rangelands (rank 7) contribute to sediment mobilization, intensifying downstream impacts (Sidle et al., 2023). While bare ground (rank 6), lacking vegetation, similarly exacerbates sediment transport (Shugar et al., 2020). Flooded vegetation (rank 5) may temporarily impede flow, but when uprooted, it enhances debris movement (Westoby et al., 2014). Tree cover (rank 4) offers moderate attenuation of flow but can also become a source of debris if destabilized (Vezzola et al., 2016). Croplands (rank 3) tend to mitigate flow speed but primarily incur economic losses without significantly contributing to flood dynamics (Huggel et al., 2002). Ice surfaces (rank 2) pose a limited hazard unless accelerated melting occurs (Carrivick et al., 2016). Finally, cloud cover (rank 1) has no direct influence on GLOF dynamics but impairs remote monitoring and early warning systems.

6.3.4 Risk assessment of glacial lakes

The risk assessment of glacial lakes proceeds through a multi-step analytical framework. First, parameters influencing GLOF susceptibility are identified and classified into either positive or negative indicators, depending on whether they amplify or mitigate risk. Second, the parameters are normalized to ensure comparability, after which multi-collinearity among them is assessed. The relative importance of each parameter is then established through a hybrid weighting approach: subjective expert judgment is operationalized via the AHP, while objective data-driven weighting is derived from theSE method. The two weight sets are

subsequently integrated in a proportional manner to ensure a balanced and comprehensive representation of parameter influence. Finally, glacial lakes are classified into distinct risk categories using a combination of the TOPSIS method and FCM clustering. Sensitivity analysis and validation are undertaken to test the robustness of the model. The overall framework is presented in schematic form in Figure 6.7.

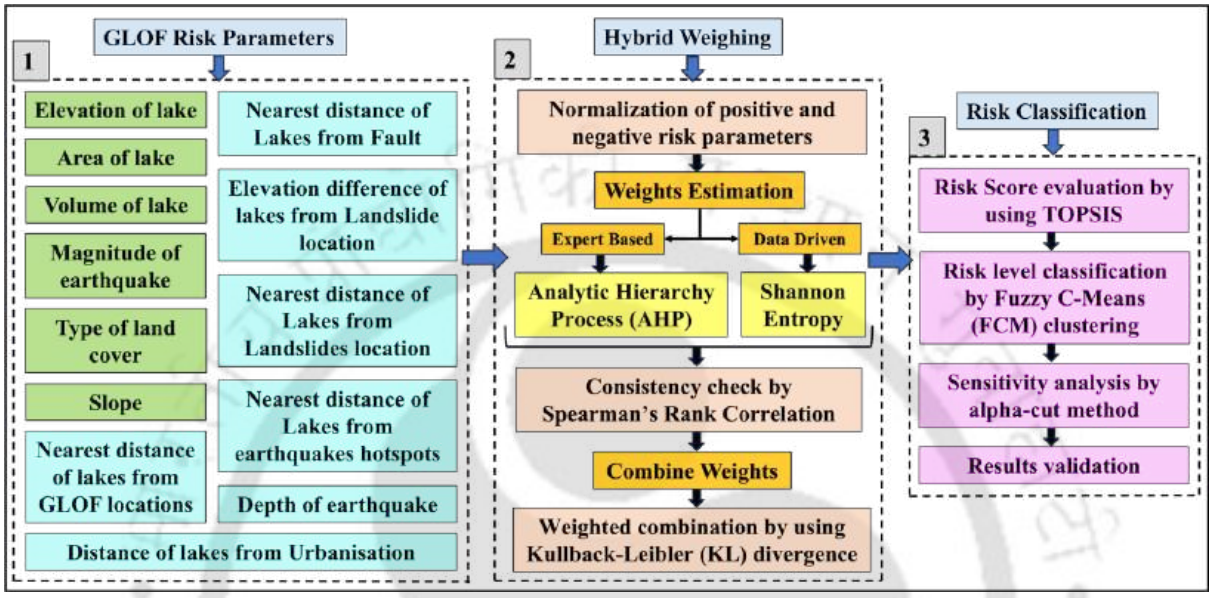


Figure 6.7: Schematic representation of the proposed framework for glacial lake risk assessment. Parameters shown in green denote positive risk indicators, while those in light blue indicate negative risk indicators.

As the 13 parameters are originally expressed in different units and scales (e.g., lake area in km², elevation in meters, and distance from faults in meters), these are rendered dimensionally consistent and directly comparable by normalization. Further, the parameters are categorized into positive (benefit) and negative (cost) indicators, allowing for appropriate normalization and ranking within the risk assessment model. Of the thirteen parameters employed, elevation, lake area, lake volume, slope, earthquake magnitude, and land cover type are treated as positive indicators – meaning that higher values are associated with increased GLOF risk. The land cover type parameter, originally categorical, has been converted into an ordinal scale ranked from one to nine to reflect its relative contribution to hazard risk. This scale ranges from cloud cover (1) to waterbody (9), progressing through ice, crops, trees, flooded vegetation, bare ground, rangeland, and built-up areas, in alignment with their hazard relevance.

On the other hand, the remaining parameters—namely, distance from active faults, historical GLOF events, earthquake epicenters, landslides, and built-up areas, as well as

earthquake focal depth and the elevation differential between landslide source and lake site—are designated as negative indicators. Lower values correspond to increased risk, as closer proximity or shallower depth implies greater susceptibility.

For a dataset comprising *m* lakes and *n* criteria, the normalization of positive indicators is computed using the following formula:

$$r_{ij} = \frac{d_{ij} - \min(d_j)}{\max(d_j) - \min(d_j)} \tag{24}$$

The formula for normalization of negative parameter is as follows:

$$r_{ij} = \frac{\max(d_j) - d_{ij}}{\max(d_j) - \min(d_j)} \tag{25}$$

where, d_{ij} corresponds to the performance value of the alternative i for the risk indicator j. Also, r_{ij} represents the normalized value of the j-th indicator for the i-th evaluation alternative, where $i = 1, 2, 3, \dots, m$ and $j = 1, 2, 3, \dots, n$.

After normalizing the influencing parameters, multi-collinearity was assessed using the variance inflation factor (VIF) to evaluate potential redundancy among the variables. All VIF values were found to be well below 5 (ranging from 0.755 to 3.051), indicating that no significant multicollinearity exists in the model and that each factor contributes unique information to the analysis.

6.3.4.1 Modelling framework

A. *Estimating weights for AHP and Entropy method*

The GLOF risk assessment model, designed to identify potentially hazardous lakes and their associated risk levels, relies on estimating weights for the influencing parameters. This work integrates the AHP, a subjective method based on expert judgment, with the SE method, an objective data-driven approach, to determine parameter weights. Subsequently, MCDA is applied considering multiple hazard factors. The detailed descriptions of these two weighting approaches are discussed as follows:

AHP Method:

The AHP, developed by (Saaty, 2008), is a subjective, structured method based on hierarchical pairwise comparisons. The steps for calculating AHP weights are as follows:

1. Establishing a hierarchical structure: The decision problem is organized into two levels: (i) the Goal Level, representing the main objective – in this case, identifying potentially dangerous lakes and their GLOF risk categories, and (ii) the Criteria Level, comprising thirteen parameters that influence the GLOF risk of glacial lakes. This hierarchy provides a systematic framework for weight estimation based on relative importance.
2. Constructing the pairwise comparison matrix: A pairwise comparison matrix is created to assess the relative importance of each criterion using Saaty (2008) 9-point scale. Each element a_{ij} in the matrix quantifies the importance of criterion i relative to criterion j. The matrix has the general form:

$$A = \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1n}} & \frac{1}{a_{2n}} & \cdots & 1 \end{bmatrix} \quad (26)$$

The values (a_{ij}) follow Saaty (2008)'s scale, where 1, 3, 5, 7 and 9 signifies the equal, moderate, strong, very strong, and extreme importance respectively while 2, 4, 6 and 8 signifies the intermediate importance.

3. Normalization and weight calculation: After constructing the comparison matrix, it is normalized by dividing each element by the column sum. The priority vector (weights) is derived by averaging the values in each row, representing the relative importance of each criterion (risk indicator).
4. Consistency Check: To ensure the reliability of the judgement, the consistency index (CI) and consistency ratio (CR) of the comparison matrix are estimated by using the following equations:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (27)$$

$$CR = \frac{CI}{RI} \quad (28)$$

where λ_{max} is the principal eigen value, n is the total number of criteria, and RI is the random index (determined based on the number of criteria). If CR < 0.1**,** the matrix is considered consistent; otherwise, the comparisons need adjustment.

Shannon Entropy Method:

The SE method is an objective weighting technique used in MCDA to assess the importance of each criterion based on its variability among alternatives. Originating from information theory, entropy quantifies uncertainty or disorder within a dataset. In decision-making, criteria with greater variability contain more information and are assigned higher weights (Arora et al., 2021).

If a criterion shows high uniformity across alternatives, it has a high entropy value, indicating low information content and warranting a lower weight. Conversely, criteria with significant variation have low entropy values, reflecting higher information content and thus higher weights.

The steps for evaluating the SE weights are as follows:

1. Construct the decision matrix: The decision matrix D consists of m alternatives (e.g., lakes) and n criteria (e.g., lake area, volume, slope, etc.):

$$D = \begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix} \quad (29)$$

   Now, compute the normalized values for r_{ij} by using (24 and (25 for positive and negative parameter or indicator, respectively.

2. Compute the entropy for each criterion or risk indicator: The SE **(**E_j**)** for each criterion j is computed as:

$$E_j = -k \sum_{i=1}^{m} r_{ij} \ln r_{ij} \quad (30)$$

   where, $k = \frac{1}{\ln(m)}$ is a scaling factor ensuring $0 \leq E_j \leq 1$.

3. Calculate the entropy weights: The entropy weight w_j for each of the j-th criteria is estimated by using the following equation:

$$w_j = \frac{1 - E_j}{\sum_{k=1}^{n} 1 - E_j} \quad (31)$$

B. *Combined weights estimation*

After obtaining the weights for the AHP and SE methods, Spearman's Rank Correlation is applied to verify the consistency of the results. Spearman's Rank Correlation Coefficient (ρ) measures the strength and direction of the monotonic relationship between two ranked parameters. In this case, it checks the similarity between the rankings of weights from AHP and SE. The formula for Spearman's Rank Correlation Coefficient (ρ) is as follows:

$$\rho = 1 - \frac{6 \sum R_i^2}{n(n^2 - 1)} \tag{32}$$

where, R_i refers to the difference between the rank of weights from AHP andSE and n equals to the total number of criteria or risk indicators. Notably, if $\rho > 0.7$ occurs, the two weight sets are highly consistent and can be combined directly, while if $0.4 < \rho \leq 0.7$ occurs, it indicates the moderate agreement between the two weight sets and is to be added through a weighted combination. However, if $\rho \leq 0.4$ occurs, it signifies the low agreement; further analysis is required before combining the weights.

Further, after checking the consistency of two weight sets, the KL divergence is used for combining the weights with a hybrid weighing approach. It is a statistical measure that quantifies how much one probability distribution differs from another. The combination of weights is performed in such a way that the KL divergence of the two weight sets (W_{AHP} and $W_{Shannon\ entropy}$) from the combined weight ($W_{combined}$) is minimum.

For two discrete probability distributions P (SE weights) and Q (AHP weights), the KL divergence is defined as:

$$D_{KL}(P||Q) = \sum_{j=1}^{n} P_j \ln\left(\frac{P_j}{Q_j}\right) \tag{33}$$

where, P_j refers to theSE weight for the j-th criteria, Q_j refers to the AHP weight for the j-th criteria and n equals to the total number of criteria.

Thus, after minimizing the KL divergence of the two weight sets from the AHP andSE method for a β combination, the combined weight $\left(W_{j(combined)}\right)$ of the j-th criteria is as follows:

$$W_{j(combined)} = \beta\ W_{j(AHP)} + (1 - \beta)\, W_{j(Shannon\ entropy)} \tag{34}$$

6.3.5 Risk classification

TOPSIS is used to compute risk scores by comparing each lake's characteristics against an ideal solution (highest risk) and a negative-ideal solution (lowest risk) (Z. Zhang et al., 2022). It is a ranking technique that selects alternatives by minimizing their distance from the positive ideal solution while maximizing their distance from the negative ideal solution. The process begins by multiplying the normalized values (r_{ij}) by the final hybrid weights ($W_{j(combined)}$).

$$v_{ij} = W_{j(combined)} . r_{ij} \tag{35}$$

Next, compute the distance of i-th alternative from positive ideal solution (S_i^+):

$$S_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^2} \tag{36}$$

where, if j is a positive risk indicator, v_j^+ equals to maximum of v_{ij} else if j is a negative risk indicator, v_j^+ equals to minimum of v_{ij} for $i = 1, \dots, m$ alternatives.

Further, the distance of i-th alternative from the negative ideal solution (S_i^-) is calculated as:

$$S_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2} \tag{37}$$

where, if j is a positive risk indicator, v_j^- equals to minimum of v_{ij} else if j is a negative risk indicator, v_j^- equals to maximum of v_{ij} for $i = 1, \dots, m$ alternatives. Finally, the risk score (C_i) is the relative closeness to the ideal solution such that higher is the C_i value, higher will be the risk. It is evaluated as:

$$C_i = \frac{S_i^-}{S_i^- + S_i^+} \tag{38}$$

These risk scores are further classified into four risk levels such as very high, high, medium, and low risk lakes by using the FCM clustering algorithm. Unlike hard clustering methods, FCM allows soft classification, meaning a lake can belong to multiple categories with varying degrees of membership. Initially, four cluster centers are assigned, corresponding to the risk

categories. Each lake is then assigned a membership value or confidence level (u_{ik}) that represents its likelihood of belonging to a particular category.

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{\|C_i - C_k\|}{\|C_i - C_j\|} \right)^{2/(f-1)}} \quad (39)$$

where, u_{ik} is the membership value of the lake i in cluster k (very high, high, moderate, low risk), C_i is the risk score of the lake i obtained from TOPSIS, C_k is the cluster center for the risk category k, c is the total number of clusters, and f is the fuzziness factor. These membership values are updated iteratively by recalculating cluster centers based on the distribution of risk scores. The cluster centers are recalculated based on membership values as follows:

$$C_k = \frac{\sum_{i=1}^{m} u_{ik}^{f} C_i}{\sum_{i=1}^{m} u_{ik}^{f}} \quad (40)$$

where, C_k is the updated cluster center for the k-th risk category and m equal to the total number of lakes (alternatives). This process continues until the cluster centers stabilize, ensuring accurate classification. Finally, each lake is assigned to the category with the highest membership value.

6.3.6 Sensitivity analysis

The α-cut sensitivity analysis is a technique used in FCM clustering to analyze how changes in the membership threshold (α-cut) affect the classification of alternatives (glacial lakes in this thesis) (Phillips et al., 2011). It helps in determining the stability and robustness of risk classification by controlling the level of uncertainty in membership values. In FCM, each lake has a degree of membership (u_{ik}) in multiple risk categories (Very High, High, Moderate, Low), instead of belonging exclusively to one category. The α-cut threshold defines a minimum required membership value for an alternative to be assigned to a category.

$$\text{Alpha-cut: } u_{ik} \geq \alpha \quad (41)$$

where α is the α-cut threshold. Generally, for high-risk decision making like GLOF risk, a higher α-cut (≥ 0.6) is preferred. Therefore, if u_{ik} is greater than or equal to alpha, the lake is assigned to a risk class k. If a lake has multiple memberships above α, assign it to the highest

membership risk class. This is followed by analysing the number of changes in risk categories as the value of alpha increases. If the lakes frequently change risk classes with small changes in α, it indicates that the classification is unstable and requires adjusting FCM parameters. However, if classification remains stable over a range of α-cut, it signifies that the classification results are robust and reliable.

The proposed method supports a scalable and interpretable GLOF risk assessment, capable of handling diverse lake types, including moraine-dammed, ice-dammed, bedrock-dammed, and composite lakes. The final outputs are mapped in GIS to visualize the spatial distribution of potentially dangerous glacial lakes across the study site in the Eastern Himalaya.

6.4 Results and discussion

Weights of thirteen parameters are determined, and lakes are classified in four risk classes by the proposed method. Results are validated by comparing the assessed risk levels with five GLOF events reported by Rounce et al. (2016) in Nepal. Additionally, risk levels for 25 randomly selected GLOF events from T. Zhang et al. (2022b) are compared to further evaluate the model's performance.

6.4.1.1 Weights distribution of risk parameters

A. ***AHP method***

The AHP was applied to estimate the weights of thirteen parameters influencing GLOF risk assessment. The pairwise comparison matrix (Table 6.1) reflects expert judgments on the relative importance of these criteria.

Table 6.1: Pairwise comparison matrix used in the AHP to estimate weights (W_{AHP}) for various criteria influencing GLOF risk assessment. Each criterion is compared against every other criterion based on expert judgment using Saaty's 1–9 scale (and their reciprocals)

	A	**V**	**H**	**S**	**DF**	**DLs**	**ΔE_{Ls}**	**DE**	**Mag**	**d**	**DG**	**DB**	**LC**	**W_{AHP}**
A	1	1/2	5	5	5	5	5	5	5	5	5	6	6	**0.2004**
V	2	1	7	7	7	7	7	7	7	7	7	7	7	**0.2909**
H	1/5	1/7	1	2	3	3	3	3	3	2	2	2	2	**0.0842**
S	1/5	1/7	1/2	1	3	3	3	3	3	3	3	3	3	**0.0875**
DF	1/5	1/7	1/3	1/3	1	1/2	1	1	1	1	1	2	2	**0.0371**
DLs	1/5	1/7	1/3	1/3	2	1	2	2	2	2	2	2	2	**0.0558**
ΔE_{Ls}	1/4	1/7	1/3	1/3	1	1/2	1	2	2	2	2	2	2	**0.0482**
DE	1/5	1/7	1/3	1/3	1	1/2	1/2	1	1	2	2	2	2	**0.0410**
Mag	1/5	1/7	1/3	1/3	1	1/2	1/2	1	1	2	1	2	2	**0.0382**
D	1/5	1/7	1/2	1/3	1	1/2	1/2	1/2	1/2	1	1	2	2	**0.0328**
DG	1/5	1/7	1/2	1/3	1	1/2	1/2	1/2	1	1	1	2	2	**0.0342**
DB	1/6	1/7	1/2	1/3	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1	2	**0.0258**
LC	1/6	1/7	1/2	1/3	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1	**0.0233**
where, A = Area, V = Volume, H = Elevation, S = Slope, DF = Nearest distance of lakes from Fault, DLs = Nearest distance of Lakes from Landslides, ΔE_{Ls} = Elevation difference of lakes from nearest Landslide, DE = Nearest distance of lakes from Earthquake, Mag = Magnitude of earthquake, d = Depth of earthquake, DG = Nearest distance of lakes from GLOF, DB = Nearest distance of lakes from built-up, LC = Land Cover, and W_{AHP} = Weights for risk criteria by AHP method.														
Consistency Ratio (CR) = 0.0884														

Calculated weights (W_{AHP}) indicate that Volume (0.291) and Area (0.200) are the most significant parameters contributing to the risk, highlighting the critical role of lake size in hazard potential. Other parameters with moderate influence include Slope (0.088), Elevation (0.084), and Distance from Landslides (0.056). Parameters related to earthquake characteristics such as Magnitude (0.038), Depth (0.033), and distance from earthquake (0.041) have lower weights, yet contributing effectively to the overall risk profile. Land Cover and proximity to built-up areas received the lowest weights, indicating comparatively lesser influence in the study area.

The consistency of expert judgments was evaluated using the Consistency Ratio (CR), which yielded a value of 0.0884, below the acceptable threshold of 0.10. This confirms the reliability and coherence of the pairwise comparisons made.

Overall, these results emphasize that the physical attributes of the lakes (volume and area) and their geomorphological context (slope, elevation, and proximity to landslides) dominate

the assessment of GLOF risk, while seismic factors play a less significant role in the current weighting scheme.

B. *Shannon Entropy (SE) method*

The SE method was applied to estimate the weights of thirteen parameters influencing GLOF risk assessment, and the results were cross-compared with expert-derived rankings from AHP. Table 6.2 presents the entropy weights (W_{SE}), standardized ranks (R_{SE}), and corresponding AHP rankings (R_{AHP}).

Table 6.2: SE weights (W_{SE}) of the GLOF risk parameters along with comparative ranks (R_{SE} and R_{AHP}) derived from SE and AHP method. The table highlights the relative influence of each parameter on GLOF risk, with higher weights.

GLOF Risk Parameters	W_{SE}	R_{SE}	R_{AHP}
Area	0.3441	2	2
Volume	0.3984	1	1
Elevation	0.0074	12	4
Slope	0.0815	3	3
Nearest distance of lakes from fault	0.0154	9	9
Nearest distance of lakes from landslides	0.0165	8	5
Elevation difference of lakes from nearest landslide	0.0174	7	6
Nearest distance of lakes from earthquakes	0.0115	10	7
Magnitude of earthquake	0.0337	5	8
Depth of earthquake	0.0057	13	11
Nearest distance of lakes from GLOF	0.0206	6	10
Nearest distance of lakes from built-up	0.0114	11	12
Land Cover	0.0363	4	13

The analysis reveals strong alignment in the top-ranking parameters. Lake volume carries the highest weight (W_{SE}: 0.3984) and is consistently ranked first by both AHP and Entropy methods. Lake area (W_{SE}: 0.3441) and slope (W_{SE}: 0.0815) follow closely, sharing top three ranks across both approaches. This convergence affirms their central role in influencing GLOF risk, particularly in terms of potential outburst magnitude and downstream hazard propagation.

Discrepancies between the two methods (AHP and SE) emerge in parameters with more ambiguous roles. Elevation, for instance, is weighted lower by entropy method (W_{SE}: 0.0074; R_{SE}: 12) but is ranked fourth in AHP, suggesting that expert assessments attribute more significance to terrain positioning than is reflected by its statistical variability. Similarly, land cover receives moderate weight in entropy (W_{SE}: 0.0363; R_{SE}: 4) but is ranked lowest in AHP, possibly reflecting limited perceived relevance to outburst triggering despite its variation across sites.

Seismic parameters – including earthquake magnitude and proximity to epicenters – hold moderate entropy weights but received lower ranks in AHP. This pattern suggests that while seismic factors vary across the region, their perceived contribution to immediate GLOF initiation is considered secondary. Both earthquake depth (W_{SE}: 0.0057;R_{AHP}: 11) and proximity to built-up areas (W_{SE}: 0.0114; R_{AHP}: 12) consistently rank lower, indicating limited influence at the basin level.

The comparative framework underscores the strength of combining quantitative data analysis with expert-informed judgment. The results from both the AHP and SE methods show strong agreement in the prioritization of key parameters such as lake volume, area, and slope, indicating high confidence in their relative importance for GLOF risk. In cases where the two methods differ, the discrepancies reveal the trade-off between statistical significance (captured by entropy) and physical interpretability (informed by expert opinion). This highlights the complementary nature of integrating objective data-driven insights with subjective expert knowledge to achieve a more balanced and comprehensive GLOF risk assessment.

C. *Hybrid weights for risk assessment*

The combined weighting results ($W_{combined}$) and their rankings ($R_{combined}$) provide a synthesized assessment of GLOF risk parameters by integrating both data-driven (SE) and expert-based (AHP) methods. The Table 6.3 presents the final combined weights ($W_{combined}$) and corresponding ranks ($R_{combined}$) of the 13 GLOF risk parameters, obtained using a hybrid weighting approach that integrates both subjective (AHP) and objective (SE) weights through KL divergence minimization.

Table 6.3: Combined weights ($W_{combined}$) and ranks ($R_{combined}$) of GLOF risk parameters

GLOF Risk Parameters	$W_{combined}$	$R_{combined}$
Area	0.2940	2
Volume	0.3610	1
Elevation	0.0342	5
Slope	0.0836	3
Nearest distance of lakes from fault	0.0230	10
Nearest distance of lakes from landslides	0.0302	7
Elevation difference of lakes from nearest landslide	0.0282	8
Nearest distance of lakes from earthquakes	0.0218	11
Magnitude of earthquake	0.0353	4
Depth of earthquake	0.0152	13
Nearest distance of lakes from GLOF	0.0254	9
Nearest distance of lakes from built-up	0.0165	12
Land Cover	0.0317	6

Lake volume (0.3610), area (0.2940), and slope (0.0836) rank as the top three parameters, followed by earthquake magnitude (0.0353) and elevation (0.0342). These five parameters collectively account for 80.81% of the total combined weight, underscoring their critical role in GLOF susceptibility. This suggests that lakes characterized by large volume and area, steep surrounding terrain, high elevations, and exposure to high-magnitude seismic activity are particularly vulnerable to outburst events.

Mid-ranked parameters such as land cover (0.0317) and proximity to landslides (0.0302) play supporting roles, while the remaining factors – including elevation difference from nearby landslides, proximity to faults, previous GLOFs, and earthquake epicenters – have comparatively lower influence (weights between 0.015 and 0.030). Earthquake depth (0.0152) and distance from built-up areas (0.0165) were found to be the least significant. These results suggest that proximity-related parameters may be locally relevant in triggering or amplifying GLOF impacts; their lower spatial variability across the study area reduces their contribution in the overall model.

To reconcile the differing weight estimations from SE and AHP, Spearman's rank correlation coefficient (ρ) was calculated, resulting in a moderate agreement (ρ = 0.4670). This value ($0.4 \leq \rho < 0.7$) indicates partial correlation, insufficient to justify direct averaging. Therefore, a hybrid combination was applied to minimize KL divergence, assigning 34.86% weight to AHP-derived values and 65.14% to entropy-based values (β = 0.3486). This optimized fusion ensures a balanced representation of expert judgment and data variability in the final weighting structure.

$$W_{j(combined)} = 0.3486\ W_{j(AHP)} + 0.6514\ W_{j(Shannon\ entropy)} \quad (42)$$

6.4.2 Risk classification and confidence levels

This thesis applies the TOPSIS method to evaluate the risk score for each of the alternatives or candidate lakes by assessing the corresponding risk parameters. Then, Fuzzy C Means clustering classifies the lakes into the four-risk classes (low, moderate, high and very high risk) with different confidence levels or membership values. Notably, the risk class having the highest confidence is considered the final risk class for a lake. After 49 iterations, there was minimal improvement in the FCM clustering, and the result shows that out of the total 1144 lakes, there are 432, 368, 227, and 117 lakes of low, moderate, high and very high GLOF risk, respectively. The occurrence of very high-risk and high-risk lakes is due to various common factors like high elevation, steep slopes, large area, large volume, closeness to earthquakes and

occurrence of rangeland or bare ground in the vicinity of lakes. On the other hand, moderate risk and low risk lakes are of medium or small size in area, small volume, occurring at lower elevation, and mostly located away from major faults, earthquakes, landslides, GLOF and built-up areas. Figure 6.8 represents the spatial occurrence of these lakes with their different risk levels.

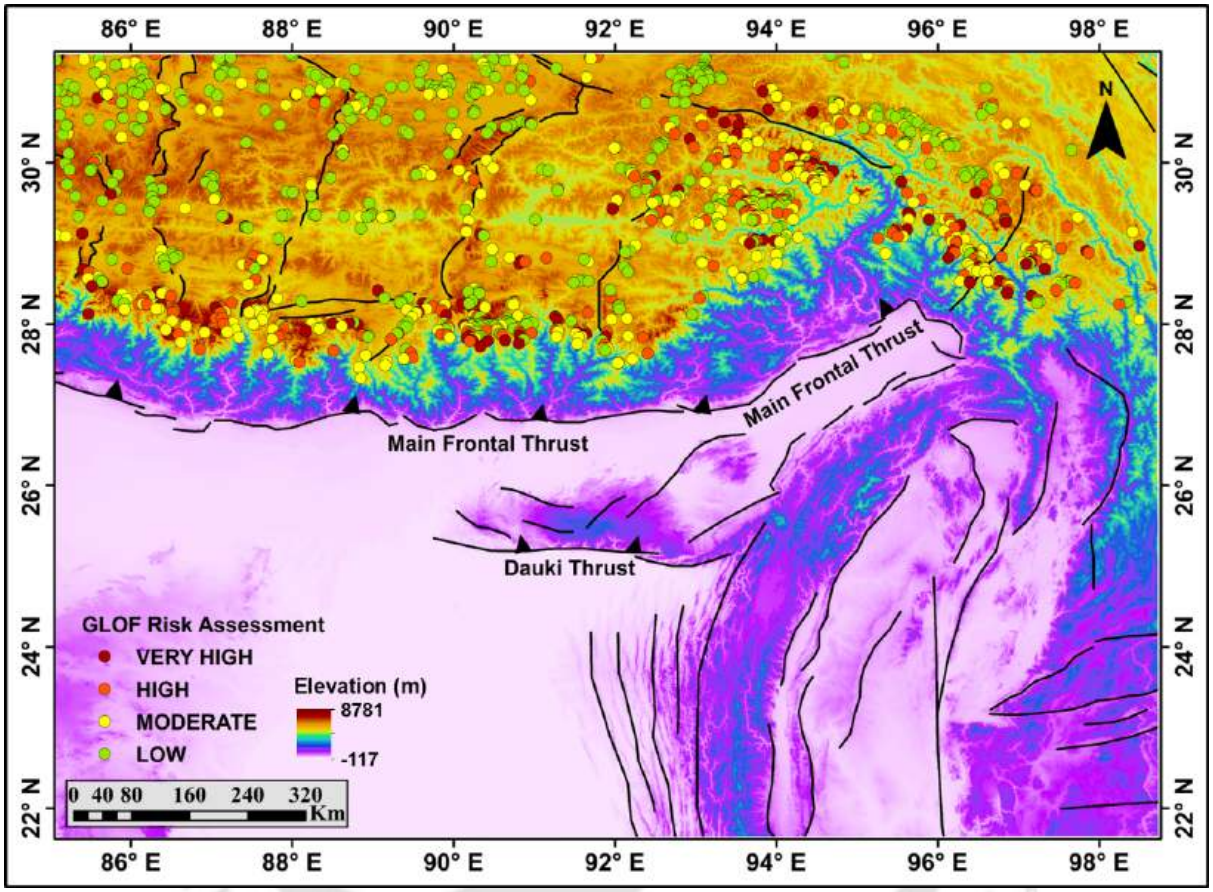


Figure 6.8: GLOF risk assessment map of the Eastern Himalayas and surrounding regions, based on SRTM DEM-derived elevation data. The glacial lakes are categorized into four GLOF risk levels: Very High (dark red), High (orange), Moderate (yellow), and Low (green).

6.4.3 Sensitivity analysis

Following the classification of lakes into pre-defined GLOF risk classes, the thesis assessed the confidence level of these classifications using sensitivity analysis, as presented in Table 6.4. Each risk class includes the number of lakes assigned to it with a classification confidence equal to or exceeding a defined alpha threshold, which in this case ranges from 0.65 to 0.9. An additional "unassigned" category captures lakes with confidence values below the specified alpha, indicating lower certainty in risk classification.

Table 6.4: Sensitivity analysis of risk assessment of lakes for risk confidence ≥α value

≥ α	Low Risk	Moderate Risk	High Risk	Very High Risk	Unassigned
0.65	432	368	227	117	0
0.7	432	367	226	115	4
0.75	431	366	223	115	9
0.8	411	340	223	114	56
0.85	388	309	219	113	115
0.9	333	338	215	110	248

Results in the Table 6.4 show that as the alpha threshold increases, the number of lakes falling into the unassigned category also rises, reflecting a stricter confidence requirement. However, changes in the number of lakes assigned to each risk class remain relatively minor across thresholds. Specifically, 4, 9, 56, and 115 lakes were unassigned at alpha levels of 0.7, 0.75, 0.8, and 0.85, respectively. Despite this increase in unassigned lakes at higher alpha levels, a significant majority – 1088 out of 1144 lakes (95.1%) – were still confidently classified into risk categories at an alpha of 0.8 or higher. This indicates that the risk assessment framework maintains high reliability and stability under varying confidence thresholds.

6.4.4 Validation

Results are compared with those available in the literature to evaluate the robustness of the proposed risk assessment framework. Among the 30 observed lakes, the model classifies 17 as very high risk, 9 as high risk, 2 as moderate risk, and 2 as very low risk. Notably, 27 of these predictions carry a confidence level exceeding 80%, while the remaining 3 are above 70%, indicating strong overall reliability.

Among the 17 lakes in the very high-risk category, most lakes (12) are large (area ≥ 0.1 km²), with the remainder comprising 3 medium-sized (0.05 ≤ area < 0.1 km²) and 2 small lakes (0.01 ≤ area < 0.05 km²). In terms of surrounding land cover, 8 are bordered by bare ground, 7 by rangeland, and the remaining 2 by ice and water, respectively. High-risk lakes similarly exhibit a predominance of large size (7 lakes), with one medium and one small lake. Their immediate surroundings include bare ground (5 lakes), waterbody (2 lakes), rangeland, and ice (1 each).

The two lakes in the moderate-risk class are both small and situated near either bare ground or ice. The two low-risk lakes include one medium and one small lake, both with bare ground in their vicinity. Across the very high and high-risk categories, several common characteristics emerge: steep slopes, higher elevation, and proximity to geological and geomorphological

triggers such as faults, GLOF locations, earthquakes, and landslides. Figure 6.9 illustrates this validation, where part (A) shows the spatial distribution of assessed risk, part (B) depicts the corresponding confidence levels, and part (C) presents the classification outcomes based on the reference studies by T. Zhang et al. (2022b) and Rounce et al. (2016).

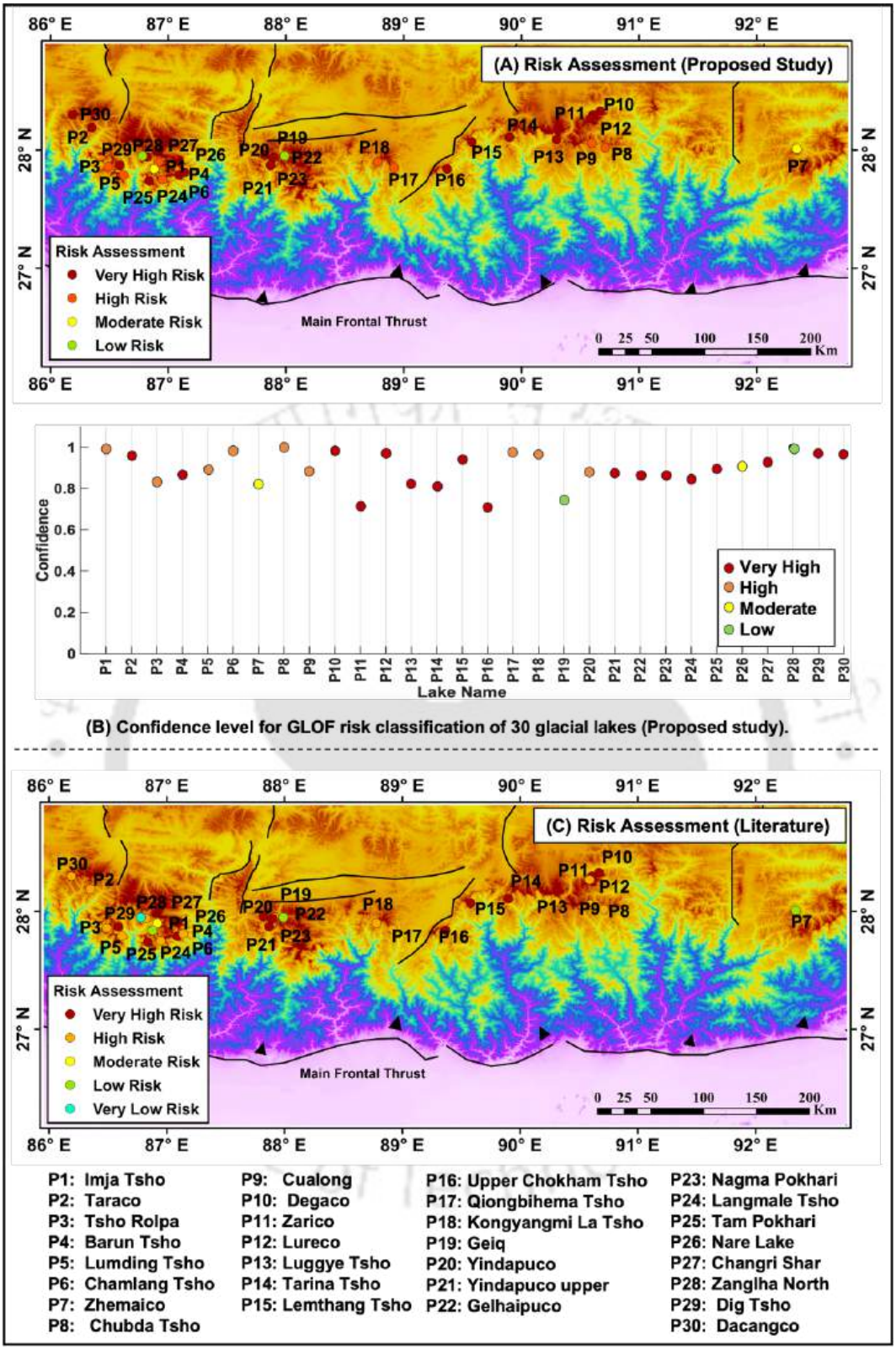


Figure 6.9: Comparison of GLOF risk assessment for 30 glacial lakes in the Central-Eastern Himalayas. (A) Proposed risk classification, (B) Confidence level corresponding to each lake's predicted risk level, and (C) Existing risk assessment from literature (Rounce et al., 2016; T. Zhang et al., 2022b).

The comparative findings align well with those reported by and (Rounce et al., 2016) and (T. Zhang et al., 2022b). The high degree of correspondence between the model predictions and literature-derived classifications supports the reliability and practical applicability of the proposed framework.

Despite the overall alignment, several notable differences emerge. Six lakes – Taraco (P2), Lureco (P12), Luggye Tsho (P13), Langmale Tsho (P15), Changri Sar (P27), and Dacangco (P30) – are identified as very high-risk in this study, whereas T. Zhang et al. (2022b) classified them as high-risk. Similarly, Imja Tsho lake (P1) is categorized here as high-risk, compared to its earlier classification as moderate-risk by Rounce et al. (2016). Zhemaico (P7) and Nare (P26) lakes are now classified as moderate-risk lakes, a shift from their previous classification as low-risk by T. Zhang et al. (2022b). Zanglha North (P28) lake, deemed very low-risk by T. Zhang et al. (2022b), is assessed as low-risk. These variations are due to differences in both type of parameters included and methods adopted.

T. Zhang et al. (2022b) employed a Fuzzy consistent matrix approach, using a limited set of five parameters focused primarily on glacial and geomorphological characteristics, such as glacier slope and moraine geometry. However, the method does not account for critical factors like lake area, volume, land cover, and proximity to GLOFs, earthquakes, landslides, or human settlements. Moreover, it applies to a single weighting approach and lacks integration of both expert judgment and data-driven variability, limiting its adaptability across diverse settings. In contrast, this thesis applies to a hybrid methodology that combines AHP with SE, weighted through KL divergence. This allows for a more flexible and generalized framework that captures both subjective knowledge and objective data patterns. Similarly, Rounce et al. (2016) used a heuristic rule-based model focused on downstream impacts and dam stability, but the analysis was limited to only six lakes and did not incorporate statistical weighting or broader geospatial variables such as fault proximity or elevation.

The higher risk class assigned to lakes in this work are attributed to the inclusion and significant weighting of parameters like lake volume, area, slope, earthquake magnitude, and elevation. Additionally, the proposed model integrates proximity to active geomorphic and seismic hazards, such as faults, GLOFs, landslides, and earthquakes, along with land cover type. With a total of 13 risk parameters when applied as indicators across 1144 lakes, and classification confidence levels exceeding 80% in most cases, this framework provides a more comprehensive and robust basis for GLOF risk assessment. The methodological breadth and

inclusion of a confidence metric support improved reliability and decision-making for hazard management in glacial environments.

6.4.5 Discussion

The results presented here address a critical gap in current GLOF risk assessments in the Eastern Himalayas by incorporating a diverse set of parameters and methodological innovations. As highlighted in the introduction, existing studies often prioritize glaciological and hydrological parameters while underrepresenting the role of tectonic and geomorphological triggers such as active faults, earthquakes, and landslides (Kougkoulos et al., 2018; Rounce et al., 2016; T. Zhang et al., 2022b). This work integrates 13 parameters – including lake volume, area, slope, elevation, seismic attributes, and proximity to hazards – into a unified framework, thereby offering a comprehensive understanding of GLOF susceptibility.

The hybrid weighting methodology proposed in this thesis – combining expert-based AHP and data-driven SE using KL divergence – represents a methodological advancement over conventional MCDA approaches that rely on either subjective or objective techniques in isolation. The moderate Spearman correlation (0.467) between AHP and entropy rankings confirmed the need for such an integrated approach, which allows the model to leverage both expert insight and statistical variation.

The results show a high degree of internal consistency and external reliability. Lake volume (0.3610), area (0.2940), and slope (0.0836) parameters emerged as dominant GLOF predictors, collectively accounting for over 80% of the combined weight. This reaffirms the findings of earlier studies that emphasize the influence of physical lake characteristics, but our model further refines this by demonstrating the additive role of seismic and proximity-based parameters. Through FCM clustering and TOPSIS-based scoring, we classified 1144 lakes into four risk categories, with the majority of high- and very high-risk lakes exhibiting common traits such as large volume, steep slopes, high elevation, and proximity to faults and past GLOF events. The robustness of these classifications is further supported by sensitivity analysis, which shows that 95.1% of lakes retained their classification at an alpha level of 0.8 or higher.

A direct comparison with the literature highlights the novelty and enhanced predictive power of our approach. For example, lakes like Taraco, Lureco, Luggye Tsho, and Langmale Tsho were classified as very high risk by our model. However, these lakes were earlier deemed high-risk by T. Zhang et al. (2022b). These discrepancies stem from our inclusion of seismic

and terrain-related parameters that were omitted in prior studies. Likewise, the classification of Imja Tsho and Zhemaico lakes reflects a more dynamic and realistic risk estimation due to a broader parameter base.

The validation exercise, involving a subset of 30 lakes, underscores the framework's reliability: 27 predictions had confidence levels above 80%, and the remaining three exceeded 70%. This is a substantial improvement over previous heuristic or single-weighting models, which either lacked confidence metrics or were constrained by limited sample sizes and parameter scopes (Rounce et al., 2016; T. Zhang et al., 2022b).

Despite the effectiveness of the proposed method, certain limitations remain. The current framework relies on static geospatial and seismic datasets, which may not fully reflect the dynamic and evolving nature of glacial lake systems. Temporal variability – such as seasonal lake expansion, glacier retreat, or changes in surrounding hazard conditions – is not captured in the present model. Future refinements should benefit from integrating real-time monitoring systems, high-resolution remote sensing, and climate-driven projections of lake evolution. Moreover, the classification component, currently based on FCM clustering, can be further tested using alternative clustering algorithms to assess classification sensitivity and improve robustness. While the combination of TOPSIS and FCM allows flexibility in evaluating composite risk, the model may be expanded by incorporating additional dynamic parameters such as lake growth rates, glacier-lake connectivity, and sediment transport dynamics. These additions would strengthen the adaptability and relevance of the framework in the face of accelerating cryospheric and tectonic changes in the region.

Nonetheless, it is widely recognized that no model can fully encapsulate the complex interactions governing real-world hazards. This limitation applies to the work of this chapter as well. However, the strong validation outcomes and high classification confidence levels demonstrate that the proposed framework yields robust, reliable, and operationally useful results for GLOF risk assessment in the Eastern Himalayas.

6.5 Transferability of the BNN model using transfer learning

The Bayesian Neural Network (BNN) model developed in this study is trained using data from the Eastern Himalaya. The geomorphic and climatic conditions in this region shape the distribution and influence of the predictors. Two methods can be adopted for transferability of the developed knowledge and model: (i) identifying variables influencing the GLOF studies in

another regions and developing a local BNN model, and (ii) adopting the BNN model of this thesis and modifying for another regions. For the latter method, when the model is applied to the Central or Western Himalaya, the underlying prediction task remains the same. The model framework can therefore be reused through transfer learning. In this approach, the trained BNN serves as the base model, and only selected layers or parameters are updated using new region-specific data. This refinement allows the model to adjust to different glacier geometries, debris conditions, and valley forms. Transfer learning reduces the need for training a new model from the beginning and improves robustness. Successful transfer requires representative samples from the target region and recalibration of predictor ranges. With these adjustments, the model can be extended beyond the Eastern Himalaya.

6.6 Conclusion

This chapter presents a comprehensive and scalable framework for GLOF risk assessment in the study area located in the Eastern Himalayas by integrating glaciological, geomorphological, hydrological, and seismic parameters within MCDA approach. Addressing a significant gap in existing literature, the proposed model expands the conventional parameter set by including tectonic and proximity-related variables such as distance from faults, earthquakes, and landslides – factors often overlooked in prior assessments. The novelty lies in the hybrid weighting method, which combines expert-driven AHP and data-driven SE using KL divergence. This integration allows the framework to leverage both subjective expertise and objective data variability, resulting in a more balanced and interpretable assessment.

The key findings demonstrate that lake volume, area, slope, earthquake magnitude, and elevation are the most influential parameters, together contributing over 80% of the total combined weight. These parameters were consistently associated with high- and very high-risk lakes, validating their dominant role in GLOF susceptibility. Using the TOPSIS method for risk scoring and FCM clustering for classification, the study categorised 1144 glacial lakes into four risk levels with a high degree of confidence – 95.1% of lakes remained confidently classified at $\alpha \geq 0.8$. Validation against 30 observed lakes and comparative analysis with previous studies (e.g. (T. Zhang et al., 2022b) and (Rounce et al., 2016)) confirmed the enhanced accuracy and reliability of the proposed framework. The broader parameter base and methodological rigor allowed the reclassification of several lakes to higher risk categories, indicating that earlier assessments may have underestimated potential hazards due to parameter omissions or methodological limitations.

The implications of these findings are significant. By offering a more robust and generalisable framework, the study supports improved early warning systems, policy formulation, and hazard mitigation planning in a region that is sensitive to both cryosphere and tectonics. The spatial GLOF risk map produced can serve as a critical decision-support tool for regional authorities and disaster risk managers.

Despite its strengths, the current framework is inherently deterministic and does not account for model-based uncertainty or dynamic climate interactions – both of which are essential for robust risk computation. In particular, the influence of meteorological conditions such as precipitation and temperature, as well as proximity to the nearest glacier and presence of neighbouring lakes, remains unexplored in the present risk quantification. To address these limitations, the next chapter (Chapter 7) expands the framework by integrating meteorological and glacial proximity variables into a BNN probabilistic model. The BNN is trained using the risk labels generated using the same approach as mentioned in this chapter, enabling probabilistic GLOF risk classification along with quantification of epistemic and aleatoric uncertainties. This uncertainty-aware approach enhances the transparency and reliability of GLOF risk mapping, offering deeper insights for decision-makers operating under high-risk and data-scarce conditions.

Chapter 7

Probabilistic GLOF Risk Assessment

7.1 Introduction

Although Chapter 6 developed a risk assessment framework that explicitly incorporates geometric, geomorphic, and tectonic influences into GLOF hazard modelling. On the one hand, the deterministic MCDA methods (AHP with TOPSYS) provide a structured way to incorporate expert knowledge and multiple risk factors. However, on the other hand, its deterministic classification framework lacks the capacity to quantify predictive uncertainty and remains limited in modelling complex non-linear interactions among parameters. In addition, as the adopted approach also lacks the influence of climate change within the risk assessment framework, the approach is constrained by limited parameter selection, reliance on deterministic threshold-based classification, exclusion of multiple hazard triggers such as seismicity and landslides, and the inability to quantify predictive uncertainty (Vashistha et al., 2025b). More specifically, these models are static in nature, unable to learn from new data or adjust their decision boundaries dynamically. Their reliance on pre-assigned weights and rule-based thresholds introduces subjectivity and rigidity, making them prone to inconsistencies when applied to new geographical regions or evolving climatic conditions. Even hybrid approaches like AHP-based ANNs, though data-driven, typically operate as black-box predictors with fixed architectures trained on point estimates, failing to offer meaningful insights into predictive uncertainty and lacking a predictive formulation to generalize beyond the training dataset (Witteveen et al., 2018)Furthermore, the unification of model performance with uncertainty for assessing risk across all lake types—moraine-dammed, ice-dammed, and bedrock-dammed–while also integrating non-linear interdependencies among diverse contributing factors, is not addressed. These deficiencies pose critical challenges to disaster preparedness, especially in the data-sparse and rapidly evolving environments of the study site situated in the high Himalayas.

This chapter presents an enhanced GLOF risk assessment framework based on a BNN approach. The model is trained using risk labels generated through FCM clustering of TOPSIS scores, which are computed from hybrid weights derived by combining the AHP and SE methods, as detailed in Chapter 6. This integration of expert judgment with objective information theory strengthens the robustness and interpretability of the MCDA process.

The BNN model represents a significant advancement over conventional deterministic approaches by enabling probabilistic risk predictions along with explicit uncertainty quantification – an essential feature for informed hazard mitigation planning. Unlike the static, rule-based models employed in Chapter 6, the BNN provides a flexible and generalizable framework capable of learning complex, nonlinear interactions among geophysical, climatic, and topographic variables. Furthermore, BNNs possess the capacity to adapt as new labelled data become available, enhancing their applicability across diverse and evolving glacial environments. This adaptability ensures that the model does not merely replicate predefined decision rules, but learns underlying patterns from data, thereby improving its predictive performance and enabling early identification of emerging or undocumented glacial lake hazards.

7.2 Methodology

The proposed method consists of a machine-learning-based, data-informed, and uncertainty-aware framework for the automated risk assessment of glacial lakes larger than 0.01 km², suitable across all dam types. The method is implemented for the extended ROI, defined in Chapter 6.

7.2.1 Parameter selection

The GLOF risk assessment is based on six key factors: geomorphology, physical and geometrical characteristics of the lake, topography, spatial proximity to past natural hazards, meteorological conditions, and anthropogenic exposure.

As the geomorphology of the candidate lake is influenced by neighbouring lakes, the geomorphology of the region is incorporated for each candidate lake by including the proximity of the candidate lake and the neighbouring lakes. Further, the factor of physical and geometrical characteristics of the lake is subdivided into two parameters: (i) area of a lake and (ii) proximity of the lake to glaciers. Similarly, the risk associated with topographic variation is influenced by the elevation of a lake and the terrain slope. The following are details of spatial proximity, meteorological conditions, and anthropogenic factors:

Spatial proximity: Spatial proximity to past natural hazards is subdivided into five parameters, namely, (i) proximity to landslide locations, (ii) proximity to earthquake locations, (iii) magnitude of nearest earthquake(s), (iv) depth of the nearest earthquake(s), and (v) proximity to GLOF lakes.

Meteorological conditions: Meteorological conditions are divided into two factors: (i) precipitation seasonality and (ii) temperature seasonality.

Anthropogenic exposure: Anthropogenic exposure is subdivided into two factors: (i) land use land cover (LULC), and (ii) proximity to a hydropower station.

The aforementioned factors lead to the selection of fourteen key parameters for GLOF risk assessment. These parameters are decided based on free accessibility to data, non-recurring influence, and relevance in the literature to update the risk model. Parameters like area of lake, elevation of lake, terrain slope, proximity to landslide locations, magnitude and intensity of earthquake with proximity to earthquake locations, and proximity to historic GLOF locations are adopted from Chapter 6. The remaining parameters derived from the above-mentioned factors are explained as follows:

Proximity of Neighbouring Glacial Lakes: The presence of neighbouring glacial lakes around a candidate lake, either upstream or downstream, is a critical factor in GLOF hazard assessment, as it can create cascading effects on lake failures and substantially amplify flood impacts (Ahmed et al., 2022a; Islam et al., 2022). These events highlight the necessity of considering neighbouring lakes in GLOF risk assessments, as their spatial proximity and hydrological connectivity can significantly affect lake stability and the overall intensity of a GLOF event. Therefore, the distance between a glacial lake and its neighbouring lakes are incorporated. This is based on the rationale that a shorter inter-lake distance indicates a higher potential risk due to the increased likelihood of cascading failures.

Proximity to glacier: Lakes that are located near the glacier terminus are often more dynamic and hazardous due to their direct interaction with active glacial melt, ice calving, and supraglacial water drainage. These pro-glacial lakes tend to expand more rapidly as the glacier retreats, increasing hydrostatic pressure on moraine dams. They are also more vulnerable to sudden inputs such as icefalls or subglacial discharge, which can initiate overtopping waves or destabilize the dam structure. In contrast, lakes farther from the glacier are generally older, potentially more stable, and less influenced by active ice processes. However, they can still present significant hazards depending on factors such as intense precipitation, inflow from neighbouring lakes, or seismic triggers. In this chapter, the distance between each lake and its parent glacier is measured from the outer edge of the lake boundary. Accordingly, glacier extents are delineated by applying the Normalised Difference Snow Index (NDSI) on cloud-free Sentinel-2 imagery, processed within the Google Earth Engine environment. An elevation

threshold of 3500 meters was applied on SRTM DEM (30m) to refine glacier detection and minimize false positives from seasonal snow or non-glaciated high-elevation areas.

Precipitation and temperature seasonality: Precipitation and temperature are key climate change parameters of GLOFs – increased temperature accelerates glacier melt and ice calving, while intense or variable precipitation can rapidly increase lake volume. Both parameters have the potential to increase mass inflow into glacial lakes, thereby elevating the risk of outburst events. This chapter considers two key BioClim variables, particularly BioClim 4 and BioClim 15, for long-term climatic assessment in GLOF-prone regions. BioClim 4 characterizes temperature seasonality as it is calculated as the standard deviation of monthly mean temperatures, and thus it effectively quantifies the degree of temperature fluctuation throughout the year. Higher values of BioClim 4 indicate greater seasonal temperature variability, which can influence glacier stability through repeated freeze–thaw cycles and variable melt rates. BioClim 15, or precipitation seasonality, is defined as the coefficient of variation of monthly precipitation totals—that is, the variability in the total amount of precipitation recorded for each month across a year. This variable captures the extent to which precipitation is unevenly distributed across the year. Higher values suggest more irregular or seasonally concentrated precipitation, which may contribute to the rapid accumulation of lake water, increasing the hydrostatic pressure on moraine dams and the likelihood of a GLOF event. Detailed descriptions of these variables can be found in the study by O'Donnel et al. (2012).

LULC pattern: The type of land cover surrounding a glacial lake plays a key role in influencing the local hydrological behaviour, particularly in terms of how water flows, accelerated and resisted in the terrain. In this thesis, Sentinel-2 LULC data at 10-meter resolution were used to identify the majority of the LULC classes within a 500m buffer distance of the glacial lake boundary. The LULC class may vary as glacier, water, trees, crops, flooded vegetation, rangeland, and built area, providing detailed insight for regional-scale analysis. Further, this qualitative data was converted into a quantitative form by ordering the given classes into an increasing risk order, which varies from 1 to 7, with 1 indicating the lowest risk LULC class and 7 indicating the maximum risk LULC class.

Water bodies (7) pose the highest risk by expanding flood volumes, while built-up areas (6) are highly vulnerable due to infrastructure and population exposure. Rangelands (5) lack dense vegetation, leading to increased soil erosion and sediment transport into lakes, which can weaken natural dams. Flooded vegetation (4) offers temporary hydraulic resistance but can

worsen debris flow if uprooted during a flood. Tree cover (3) reduces the flood velocity through canopy and root resistance, but uprooted trees may become hazardous debris. Croplands (2) generally reduce flow speed but mainly contribute to economic loss rather than structural risks. Ice-covered areas (1) are considered the least hazardous unless subjected to rapid melting, which may raise lake volumes. Figure 7.1 illustrates the LULC class and glacial lakes across the Eastern Himalayas, overlaid with major faults, glaciers, and hydropower stations, highlighting regions where LULC may influence glacial lake stability and associated hazards.

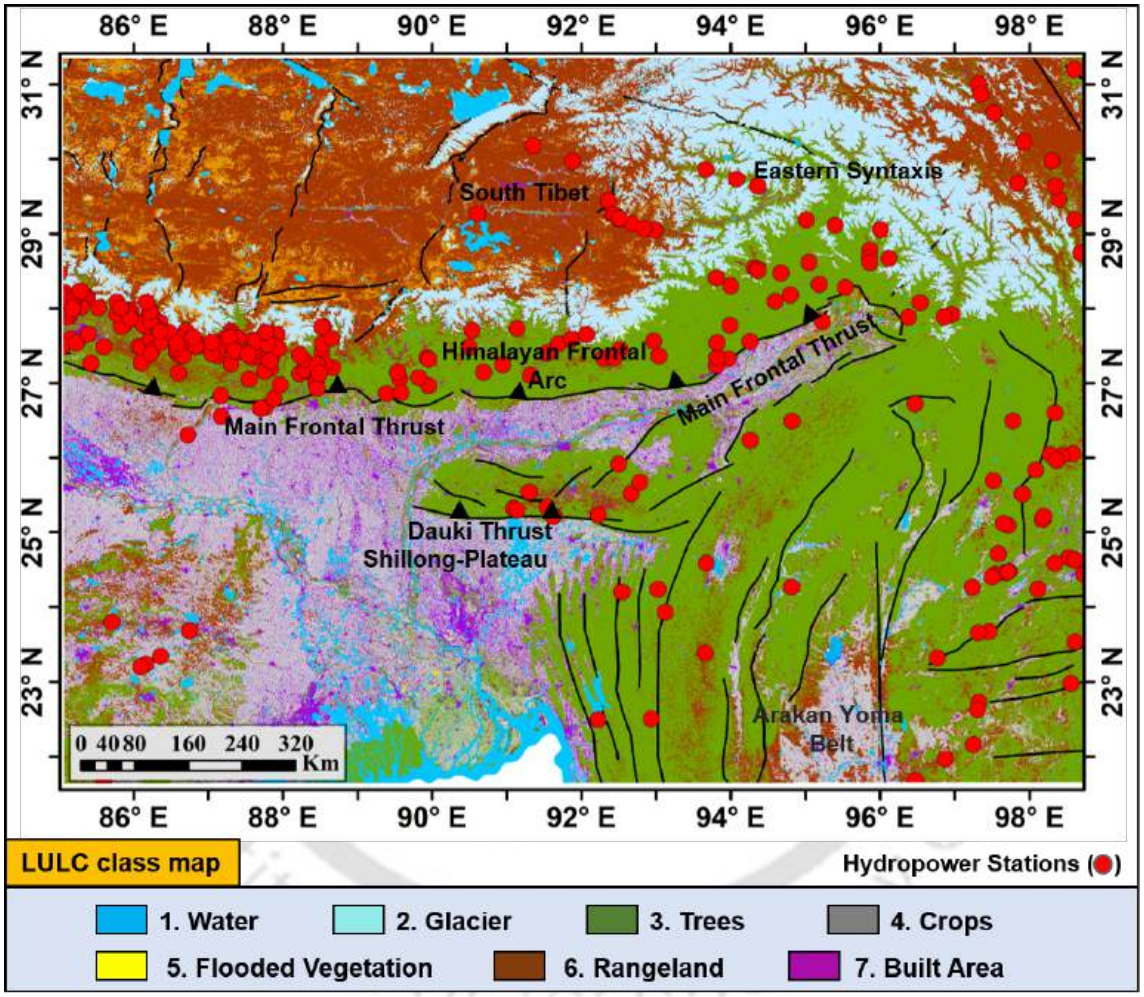


Figure 7.1: LULC map with overlaid hydropower stations (red circle) and major faults.

Proximity to hydropower station: GLOF can cause catastrophic damage to turbines, dams, tunnels, and transmission systems if the flood reaches these installations. A hydropower station located downstream of a potentially unstable glacial lake faces high economic, structural, and human safety risks. Moreover, disruption of the electricity supply can affect large populations and critical services. Therefore, assessing how close a lake is to such infrastructure helps prioritize lakes for monitoring, early warning systems, and mitigation

planning. Figure 7.1 presents a comprehensive view of hydropower stations and waterbodies. The dataset for hydropower stations is obtained from Monitor (2024).

7.2.2 Risk assessment of glacial lakes

A two-stage framework is proposed for the GLOF risk assessment that synergistically integrates MCDA with a probabilistic ML approach. In the first stage, a comprehensive set of fourteen parameters influencing GLOF susceptibility was identified and categorized into positive and negative risk indicators. These indicators were normalized, and their relative importance was quantified using a hybrid weighting scheme combining the AHP and SE methods. The resulting hybrid weights were incorporated into a TOPSIS-FCM clustering framework to generate preliminary risk labels for each glacial lake, which subsequently served as training targets for the supervised model. In the second stage, a BNN was trained to learn the nonlinear and complex relationships between the fourteen input parameters and the multi-class GLOF risk levels (low, moderate, high and very high risk). Bayesian optimization was employed to fine-tune the model's hyperparameters, thereby enhancing predictive performance over the validation set. The final BNN model thus offers a data-driven, uncertainty-aware classification tool for robust risk classification of glacial lakes across the study area in the Eastern Himalayas. A schematic flowchart of the proposed methodology is shown in Figure 7.2.

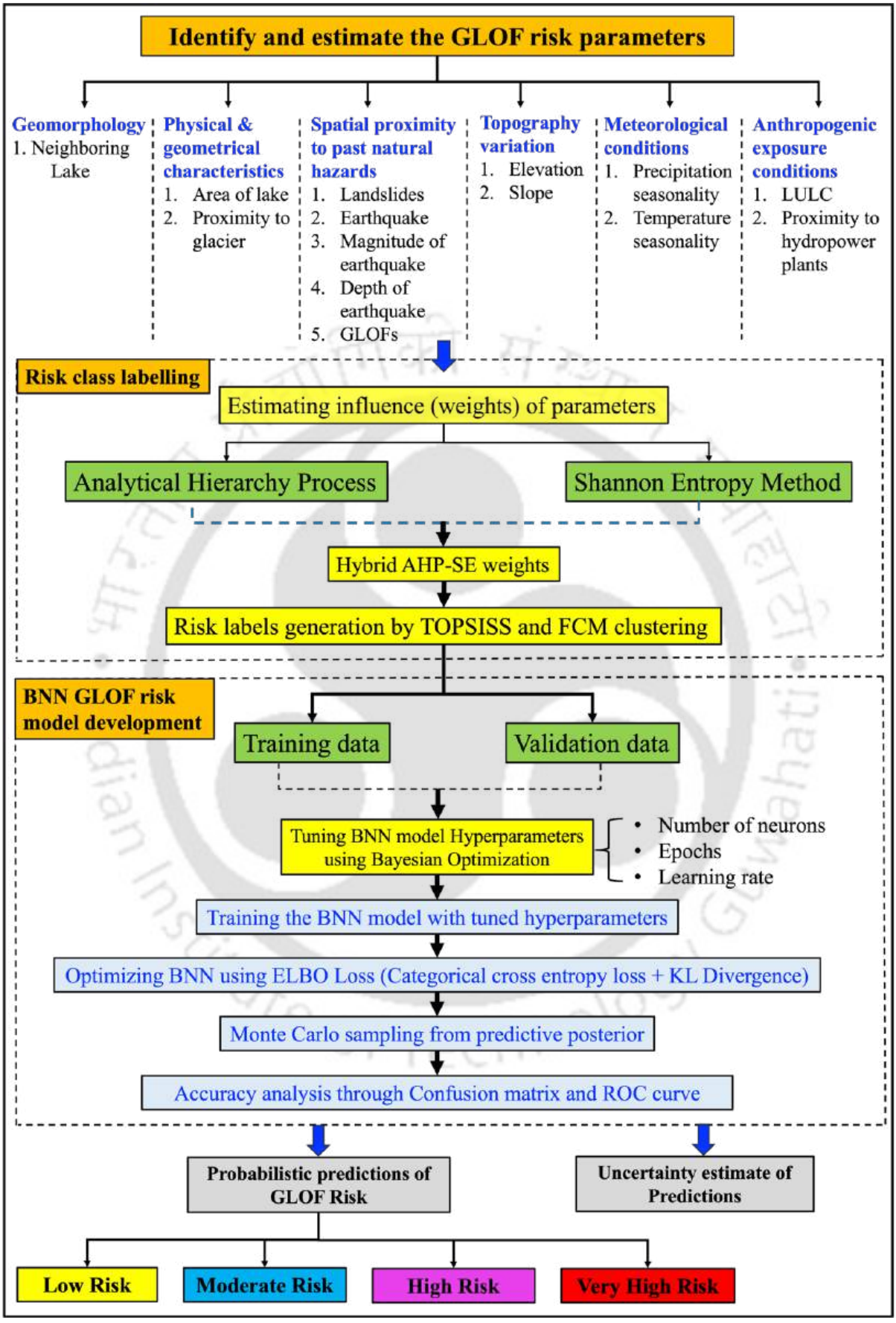


Figure 7.2: Framework for probabilistic GLOF risk predictions using BNN model.

7.3 Simulation

7.3.1 Estimation of hybrid AHP-SE weights

This chapter uses the same approach as mentioned in the section 6.3 of Chapter 6, for generating the ground truth data or label data. The procedure begins by estimating the AHP and SE weights. These weights are then integrated to capture the hybrid influence of parameters, thereby combining expert judgment with data-driven insights. As mentioned in the previous chapter (Chapter 6), this hybrid weighting scheme enhances the model's learning capability by incorporating both subjective and objective information. For training the model, TOPSIS was employed to compute risk scores, which were subsequently classified into discrete risk levels using FCM clustering. The resulting label data were then used to train the BNN model.

7.3.2 BNN for Multi-Class GLOF Risk Classification

A BNN-based risk assessment framework integrates expert-informed and data-driven perspectives through a hybrid weighting scheme, combining AHP and SE. By incorporating Bayesian inference, the model not only captures complex, nonlinear interactions among multiple variables but also provides probabilistic predictions with associated uncertainty estimates. This represents a key advancement over conventional deterministic models, enabling more informed and adaptive decision-making for hazard mitigation in high-risk, data-scarce mountain environments.

7.3.2.1 Probabilistic modelling in BNN

In this chapter, the mechanism of the BNN model is followed from the section 4.2.4 of Chapter 4. Accordingly, the proposed BNN model maps a 14-dimensional input vector $x \in \mathbb{R}^{14}$, comprising key physical characteristics, topography, geomorphological, exposure, and climatic factors, to a probabilistic prediction (output variable y) over the four risk classes. In other words, variable y represents the categorical output of a BNN model (four risk levels – very high, high, moderate, low) with fourteen input parameters. These risk levels are indicated by $y \in \{1,2,3,4\}$ denoting the four risk levels.

For multi-class classification, the network's final layer outputs class (c) logits $z \in \mathbb{R}^4$, followed by a softmax transformation:

$$\mathrm{p}(\mathrm{y}=\mathrm{c}\mid \mathbf{x},\boldsymbol{\theta})=\frac{e^{zc}}{\sum_{j=1}^{4} e^{zj}}\ , \qquad c=1,2,\ldots,4. \tag{43}$$

These probabilities are used during training to compute the categorical cross-entropy loss and, during inference, serve as the basis for both class prediction and uncertainty estimation.

Following optimization of the variational parameters ψ^*, predictions are computed by averaging over T numbers of Monte Carlo samples from the learned posterior distribution $q_{\psi^*}(\boldsymbol{\theta})$, as:

$$\hat{p}(\boldsymbol{y}^* \mid \boldsymbol{x}^*)=\int p(\boldsymbol{y}^* \mid \boldsymbol{x}^*,\boldsymbol{\theta})\, q_{\psi^*}(\boldsymbol{\theta})\boldsymbol{d\theta} \approx \frac{1}{T}\sum_{t=1}^{T} p(\boldsymbol{y}^* \mid \boldsymbol{x}^*,\boldsymbol{\theta}_t), \tag{44}$$

where $\boldsymbol{\theta}_t \sim q_{\psi^*}(\boldsymbol{\theta})$.

The optimum hyperparameters for the model are obtained by using Bayesian optimization. The detailed process of Bayesian optimization is followed from section 4.2.5.

7.3.2.2 Uncertainty analysis

Predictive uncertainty is a combination of aleatoric and epistemic uncertainty. This uncertainty enables risk stratification with confidence intervals in the GLOF, aiding proactive monitoring and response planning. The detailed process of uncertainty analysis is followed from section 4.2.4.1.

7.3.3 Application

After developing the BNN model, this chapter predicts the probabilistic GLOF risk over the 5 most frequently reviewed PDGLs, namely Imja Tsho, Tsho Rolpa, Barun Tsho, Luggye Tsho, and South Lhonak lakes. The selection of these lakes is based on their criticality, which is determined by a combination of factors such as rapid glacial retreat and lake expansion, the presence of unstable moraine dams, historical records of outburst events or strong indications of future risk, and their proximity to downstream populations, infrastructure, and hydropower facilities. These lakes are geographically distributed across Nepal, Bhutan, and India, representing different Himalayan sub-regions, allowing for a broader assessment of the model's applicability.

7.4 Results and Discussion

7.4.1 Estimation of AHP-SE weights

The relative importance of fourteen parameters influencing GLOF susceptibility was determined using a hybrid approach that combines the AHP and SE methods. The AHP weights were derived through 91 pairwise comparisons over eight iterations, resulting in a consistency ratio (CR) of 0.074, which is below the acceptable threshold of 0.1 and confirms the reliability of expert judgments. Simultaneously, the SE method was applied to quantify the variability and information content of each parameter based on the distribution of input data, ensuring an objective estimation of weights without expert bias. Following the individual weight estimation, a consistency check was performed using Spearman's Rank Correlation Coefficient (ρ), which was found to be 0.5209. As the value of ρ falls within the range $0.4 < \rho \leq 0.7, the$ weights from the AHP and SE methods show moderate agreement and cannot be directly averaged. Therefore, a hybrid weighting scheme was adopted using a β-weighted linear combination of SE and AHP weights, where β = 0.2625 was selected to minimize the KL divergence between the combined (AHP-SE) weights and the individual weight distributions. The final AHP-SE weights were obtained by combining 26.258% of the AHP weights and 73.742% of the SE weights, ensuring balanced representation of subjective expert input and objective data variability. Table 7.1 presents the AHP weights, SE weights, and the resulting AHP-SE hybrid weights along with their relative ranking.

Table 7.1: Parameters weights for AHP, SE and AHP-SE, and parameters rank

GLOF risk influencing parameters	AHP weight	SE weight	AHP-SE weight	AHP-SE rank
Neighbouring lake (W1)	0.1064	0.0041	0.0310	11
Area (W2)	0.2136	0.2601	0.2479	1
Proximity to nearest glacier (W3)	0.0498	0.0193	0.0273	12
Elevation (W4)	0.1250	0.2121	0.1892	2
Slope (W5)	0.0247	0.0063	0.0111	14
Proximity to nearest landslides (W6)	0.0659	0.0429	0.0489	7
Proximity to nearest earthquakes (W7)	0.0563	0.0299	0.0368	10
Magnitude of earthquake (W8)	0.0572	0.0877	0.0797	4
Depth of earthquake (W9)	0.0434	0.0149	0.0224	13
Proximity to nearest GLOF (W10)	0.0540	0.0537	0.0538	6
Precipitation seasonality (W11)	0.0524	0.0746	0.0688	5
Temperature seasonality (W12)	0.0491	0.0484	0.0485	8
Land Cover (W13)	0.0666	0.0943	0.0871	3
Proximity to nearest hydropower plant (W14)	0.0352	0.0518	0.0434	9

As presented in Table 7.1, lake area emerged as the most influential parameter in GLOF risk assessment, with a combined AHP–SE weight of 0.2479, followed by elevation (W4 = 0.1892), land cover (W13 = 0.0871), earthquake magnitude (W8 = 0.0797), precipitation seasonality (W11 = 0.0688), and proximity to previous GLOF events (W10 = 0.0538). In contrast, several parameters exhibited relatively lower importance, with weights below 0.05. These include proximity to landslides (W6 = 0.0489), temperature seasonality (W12 = 0.0485), proximity to hydropower plants (W14 = 0.0434), proximity to seismic epicentres (W7 = 0.0368), presence of neighbouring lakes (W1 = 0.0310), proximity to glaciers (W3 = 0.0273), and earthquake depth (W9 = 0.0224). Among all, slope (W5 = 0.0111) received the lowest combined weight, indicating its comparatively limited influence on GLOF susceptibility in the study region. Overall, the results align with the physical understanding of GLOF mechanisms, supporting the effectiveness of the hybrid AHP–SE model in prioritising influential parameters.

After evaluating the hybrid AHP-SE weights, the labelled data of 4 classes (low, moderate, high, and very high risk) is generated for training the BNN model using TOPSIS and FCM clustering. The labelled data was selected after sixteen iterations, following which there was a minimum improvement in FCM outcomes.

7.4.2 BNN model simulation

The BNN was trained on a curated dataset comprising fourteen input parameters (or features) extracted. These features were standardized and mapped to discrete GLOF risk levels – low, moderate, high, and very high, using hard labels derived from initial fuzzy classifications. This dataset was randomly split into training and validation subsets using 80% and 20% ratios to ensure stratified representation of all risk classes. As a result, 915 lakes were used for BNN training and 229 lakes for validation.

In the proposed BNN framework, each weight and bias parameter is modelled as a random variable following a Gaussian distribution, which was defined a *priori* with univariate standard normal distribution $\mathcal{N}(0, 1)$, with 'zero' mean and 'unit' variance. During training, the network learns an approximate posterior distribution for each parameter by optimizing the mean and variance of a variational Gaussian. To ensure optimal predictive performance and robust uncertainty quantification, Bayesian optimization was employed to tune the BNN's hyperparameters. The objective function for the optimization was defined as the average validation loss computed via *k*-fold cross-validation, thereby improving generalization and reducing sensitivity to any single data split. For each sampled hyperparameter configuration,

the BNN was trained and evaluated across $k = 5$ folds, and the mean validation loss was returned to guide the optimization search. A predefined search space was established for tuning of each hyperparameter, including the number of neurons in each hidden layer, learning rate, and training steps (epochs). The optimization process iteratively updated a surrogate probabilistic model to identify promising regions of the hyperparameter space. Table 7.2 presents the option details and the final optimized hyperparameter values used in the BNN architecture.

Table 7.2: Option details and tuned hyperparameters of BNN model

Hyperparameters	Option details	Tuned BNN hyperparameters
Neurons in 1st hidden layer	$64 - 256$	68
Neurons in 2nd hidden layer	$64 - 256$	150
Training steps or Epochs	$5000 - 50000$	20000
Learning rate	$10^{-4} - 10^{-2}$	0.00987

The proposed BNN model, as shown in Figure 7.3, comprises an input layer with 14 parameters, connected to a first hidden layer with 68 neurons, followed by a second hidden layer with 150 neurons. Rectified linear unit (ReLU) activations are applied after each hidden layer to introduce non-linearity. The final output layer maps the 150-dimensional latent representation to 4 output logits, which are passed to a categorical likelihood function for multi-class classification. Training was performed using full-batch optimization (i.e., the entire dataset per gradient step) within the stochastic variational inference function, with the Evidence Lower Bound loss (ELBO) as the optimization objective.

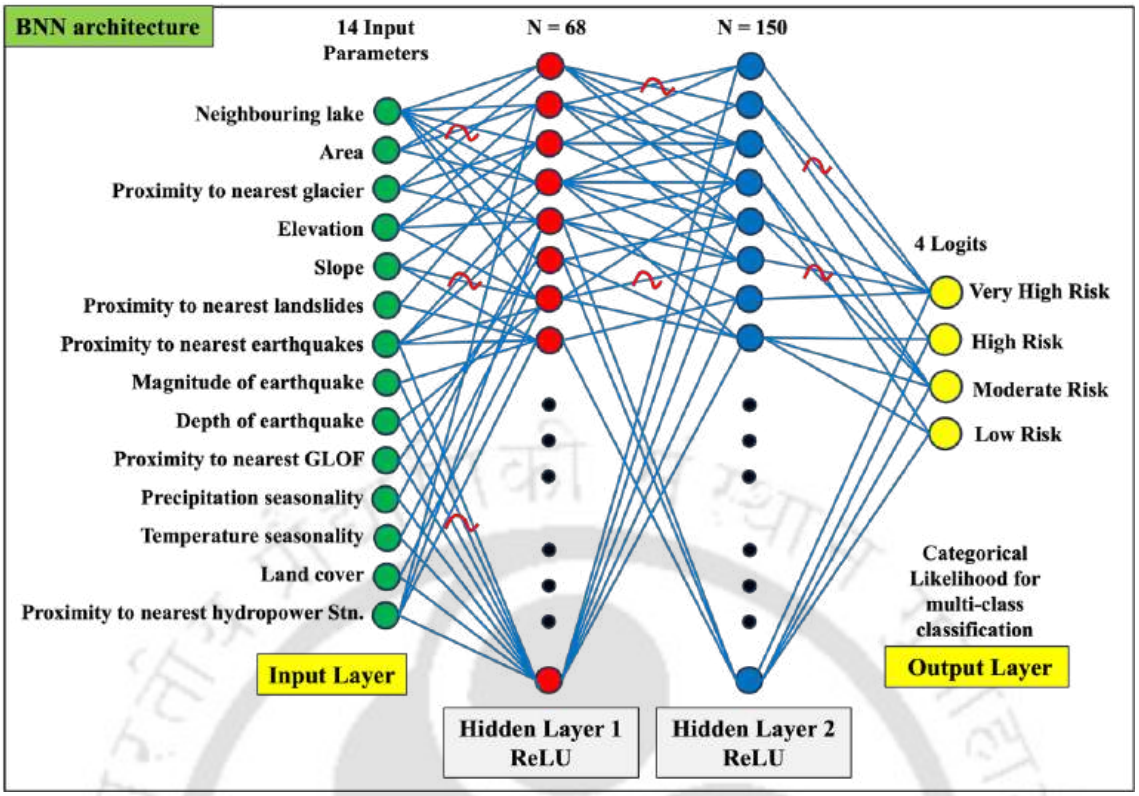


Figure 7.3: Graphical representation of the proposed BNN model (N = 68 number of neurons).

The final model was retrained using these optimal hyperparameters, and the evolution of the ELBO loss over training iterations is shown in Figure 7.4.

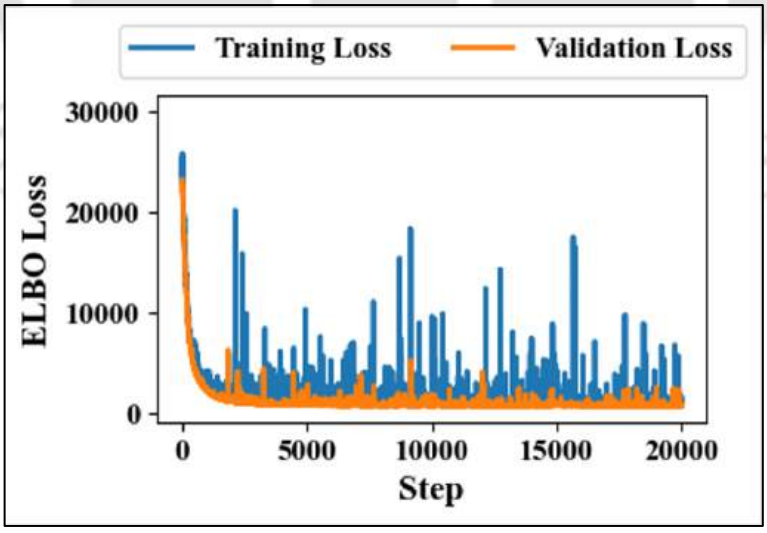


Figure 7.4: Evolution of ELBO loss over training steps for optimal BNN configuration.

With 68 hidden layers, 150 neurons, and a learning rate of 0.00987, the steps are 20,000. Both training and validation losses converge smoothly over 20,000 steps, suggesting stable learning dynamics. As observed in Figure 7.4, both training and validation losses exhibit a

consistent downward trend, with the gap between them remaining narrow, indicating stable convergence and limited overfitting.

7.4.3 Accuracy and uncertainty estimate of GLOF risk predictions

Predictions of the trained BNN were generated using a maximum a posteriori (MAP) estimation approach. Specifically, 100 Monte Carlo forward passes were drawn from the predictive posterior distribution, and the class corresponding to the highest mean predicted logit across these samples was selected as the final prediction. This provides a computationally efficient and practical means of deriving point estimates from the probabilistic model.

Based on these MAP-derived predictions, classification performance was assessed using a confusion matrix, as shown in Figure 7.5(a).

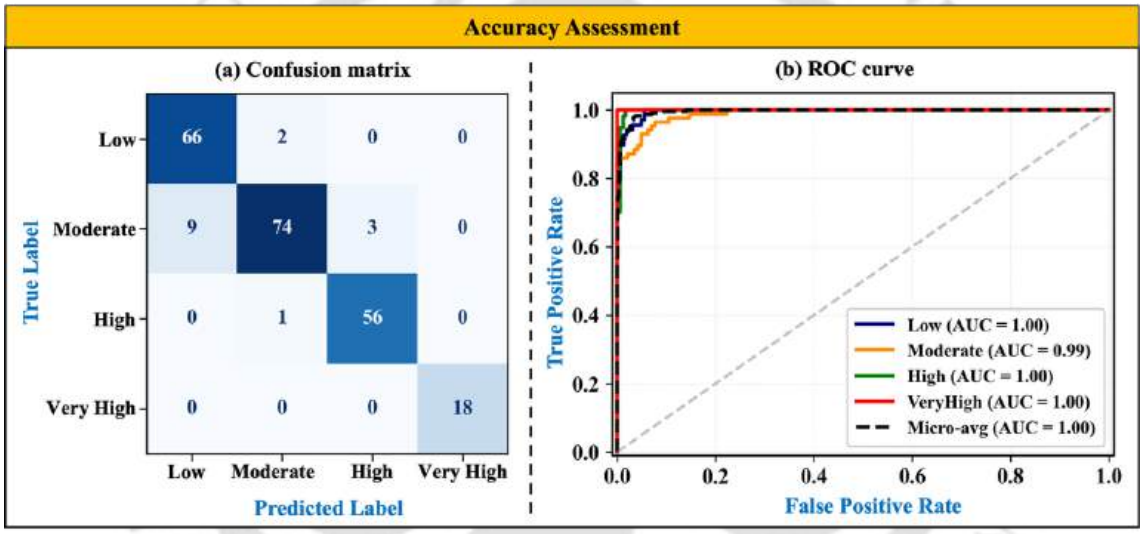


Figure 7.5: Accuracy assessment of the BNN model on the validation set using the confusion matrix (a) and ROC curves with AUC scores (b).

The confusion matrix in Figure 7.5(a) shows strong diagonal dominance, indicating a high level of agreement between the predicted (through BNN) and reference class labels derived from FCM clustering. Minor misclassifications were observed, mostly between adjacent GLOF risk categories such as 'High' and 'Very High', which reflects the gradual transitions in risk severity. Out of the 229 glacial lakes in the validation set (20% of the full dataset of 1144 lakes), the BNN model accurately classified 66 low-risk, 74 moderate-risk, 56 high-risk, and 18 very high-risk lakes, resulting in an overall classification accuracy of 93.4% in the validation set.

To further assess the model's probabilistic performance, receiver operating characteristic (ROC) curves were plotted for each risk class, as shown in Figure 7.5(b). The ROC curve plots

the true positive rate against the false positive rate across varying classification thresholds, providing insight into the model's ability to distinguish between classes based on predicted probability scores.

The model achieved an AUC (Area Under the Curve) of 1.00 for the 'Low', 'High', and 'Very High' risk classes, indicating perfect discrimination between positive and negative samples for these classes. The 'Moderate' class also exhibited excellent performance with an AUC of 0.99, suggesting a near-perfect ranking of probabilities even in the presence of a few misclassifications. The micro-averaged AUC, which aggregates the performance across all classes, was also 1.00, underscoring the model's strong overall probabilistic performance.

Despite minor label-wise misclassifications noted in the confusion matrix, the high AUC values indicate the model's capacity to assign higher probability scores to the correct class consistently. Thus, ROC-AUC presents the ranking performance rather than hard classification accuracy. From Figure 7.5(a) and (b), it is confirmed that the BNN model not only delivers high predictive accuracy but also produces well-calibrated probability distributions, supporting its application in GLOF risk classification and decision-support systems.

Figure 7.6 presents the spatial occurrence of predicted outcomes of the BNN model for the validation set, presenting very high-risk, high-risk, moderate risk and low-risk lakes. It also presents the potentially dangerous glacial lakes such as Imja Tsho, Tsho Rolpa, Barun Tsho, Luggye Tsho and South Lhonak lakes with their predicted risk levels.

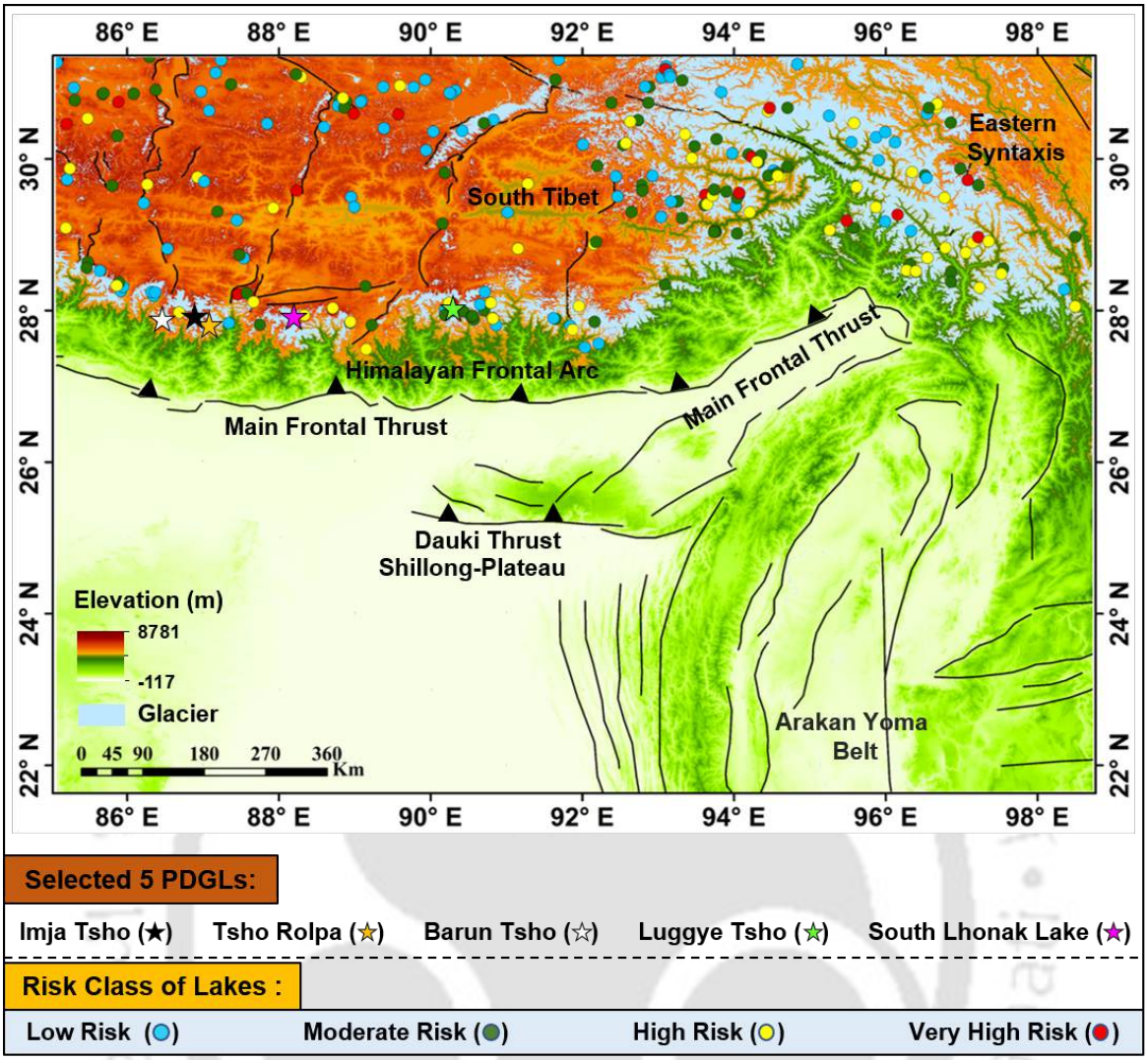


Figure 7.6: Spatial distribution of glacial lakes classified by BNN into four GLOF risk levels across the Eastern Himalaya. Each point represents a lake location, color-coded by predicted risk category: Low (blue), Moderate (green), High (yellow), and Very High (red).

The spatial distribution of 'high' and 'very high' GLOF risk glacial lakes is found to be primarily of a large area, located at higher elevation with majorly rangelands and flooded vegetation in the neighbourhood. Additionally, these lakes are characterized by intense precipitation, relatively higher magnitudes of earthquakes, and regional warming. The glacial lakes farther from GLOFs, earthquakes, landslides with smaller area, mild slope and sparse connectivity to hydropower plants are found to have 'moderate' to 'low' GLOF risk. These findings are consistent with those reported in earlier research. For instance, studies like Sattar et al. (2021) and Kaushik et al. (2024) also documented that glacial lakes such as Khanchung and South Lhonak in Sikkim are particularly susceptible to GLOF events, influenced by factors like glacier retreat, precipitation variability, and seismic activity.

Figure 7.7 shows the spatial distribution of predictive uncertainty for glacial lakes in the validation set, as estimated by the BNN. Each lake is plotted by its geographic coordinates, with colour intensity representing the variance in predicted class probabilities. Most lakes appear in blue shades, indicating low uncertainty and high model confidence, while a few scattered locations in orange to red represent relatively higher uncertainty.

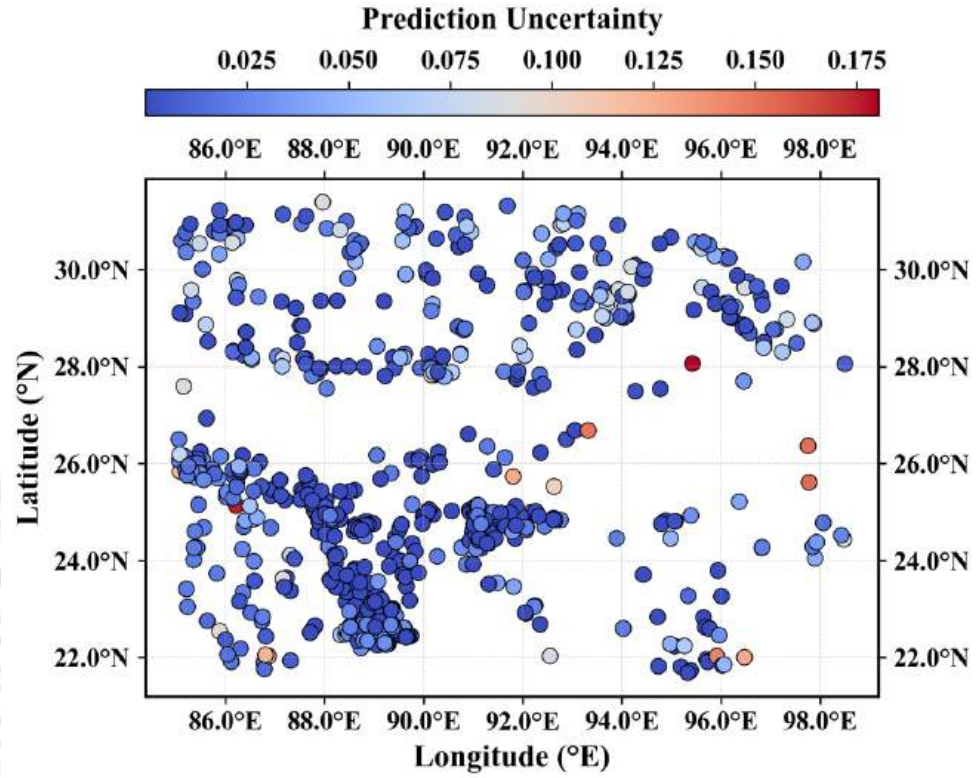


Figure 7.7: Prediction uncertainty of the BNN model for glacial lakes in the validation set across the study area.

Out of the total 229 lakes, 142 (62%) exhibit a variance below 0.05, 59 (25.76%) fall within the range of 0.05 to 0.1, 25 (10.91%) lie between 0.1 and 0.15, and 3 (1.31%) show variance between 0.15 and 0.165. Overall, all predictions lie within a narrow uncertainty range, with the maximum variance not exceeding 0.165, indicating consistent classification performance across the study region. Further examination of the input data reveals that lakes with elevated uncertainty are typically very small in size, located at lower elevations, and situated near past GLOF and landslide sites. Moreover, such lakes may have recently formed or undergone morphological alterations following past failure events, further complicating their identification. The limited representation of these lake types in the training dataset likely contributes to the observed uncertainty. From a practical perspective, these uncertainty values can guide prioritisation for field assessment. Lakes with predictive variance above 0.15 reflect lower model confidence and may warrant ground verification. Lakes in the range of 0.10 to

0.15 can be considered for targeted remote-sensing checks, while lakes below 0.10 reflect higher model confidence and can be treated as lower priority for immediate field inspection. This operational interpretation allows uncertainty metrics to be used alongside predicted likelihoods for informed monitoring and decision-making.

7.4.4 Probabilistic risk predictions of selected PDGLs

The developed BNN model was applied to estimate the probabilistic GLOF risk levels for five glacial lakes identified as highly critical due to their historical evolution, geomorphological instability, and proximity to downstream settlements. These include Imja Tsho, Tsho Rolpa, Barun Tsho, Luggye Tsho, and South Lhonak lakes. The model output provides a probability distribution over four predefined risk classes low, moderate, high, and very high for each lake. To interpret these results, violin plots were generated for each lake (Figure 7.8), where the width of each violin represents the density of the predicted probabilities, and the overlaid boxplot illustrates the interquartile range and median value within each class.

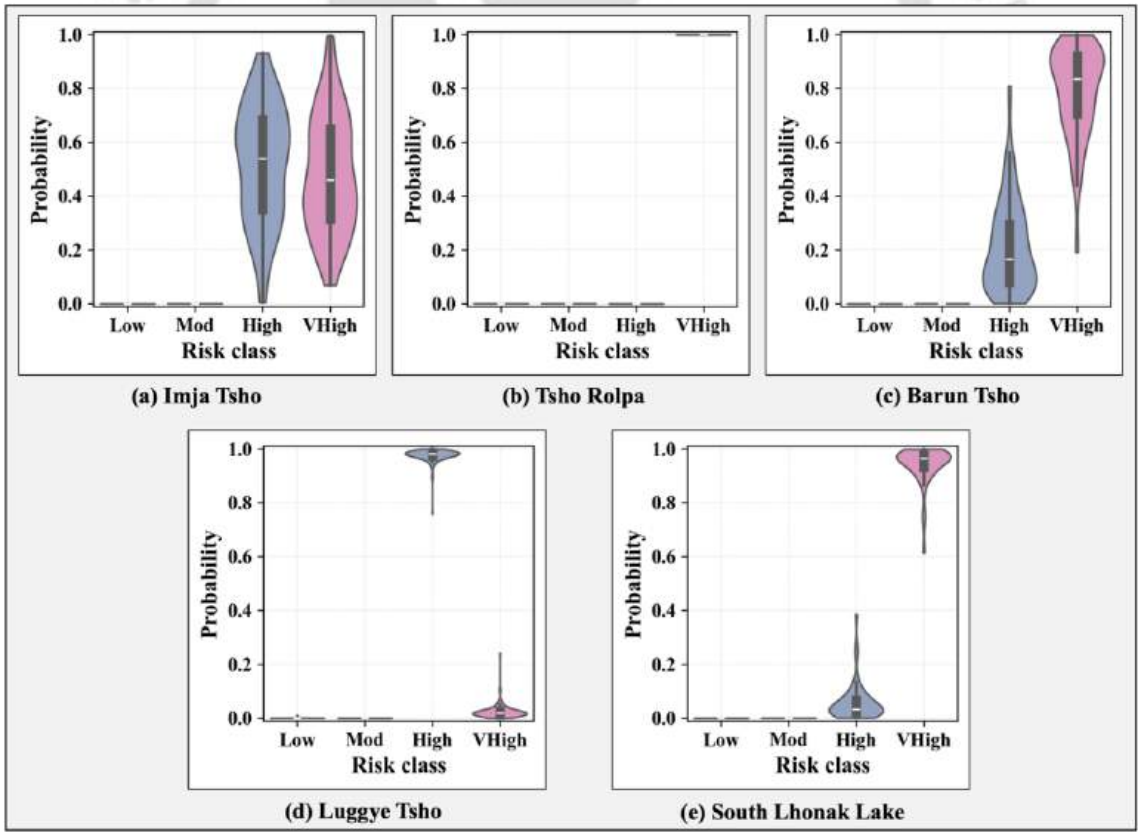


Figure 7.8: Violin plots showing probabilistic GLOF risk classifications for five critical glacial lakes ((a) Imja Tsho, (b) Tsho Rolpa, (c) Barun Tsho, (d) Luggye Tsho, and (e) South Lhonak) using the BNN model.

As shown in Figure 7.8(a), Imja Tsho lake (area = 1.55 km²), located in Nepal, exhibits a nearly bimodal probability distribution between the “High” (0.519) and “Very High” (0.48) risk categories. This split reflects the model’s uncertainty and indicates a transitional stage from ‘high’ to ‘very high risk’. The total prediction uncertainty is medium at 0.0933. The elevated risk is primarily due to its proximity to glaciers and its location in a seismically active zone, necessitating annual monitoring.

The Figure 7.8 (b) presents Tsho Rolpa lake (area = 1.59 km²), also situated in Nepal, as a clearly “Very High” risk lake, with a probability of 0.999. The distribution is sharply concentrated, with negligible support for other risk categories, and an extremely low or negligible uncertainty of 6.922×10^{-12}. Given its location near glaciers, the presence of the Khimti hydropower plant (60 MW) downstream, active earthquakes (M5.4), the Chubung glacial lake in the neighbourhood, and the potential destruction of trekking routes in the event of an outburst, the lake requires quarterly monitoring.

In the Figure 7.8(c), Barun Tsho lake (area = 1.74 km²), located in Nepal, is predominantly classified as “Very High” lake (risk probability = 0.788), with secondary support in the “High” category (0.211). The total prediction uncertainty is 0.0564. This classification is consistent with risk factors such as the presence of the Arun hydropower project (1061 MW) downstream, earthquakes (M5.6), and the threat of cascading outbursts due to the nearby Upper Barun Lake (Seto Pokhari). These factors highlight the need for regular monitoring, preferably quarterly.

Figure 7.8 (d) illustrates Luggye Tsho lake (area = 1.57 km²), located in the Lunana region of Bhutan, as primarily “High” risk with a dominant probability of 0.975. The distribution is tightly clustered with only a minimal tail into the “Very High” class, indicating strong model confidence and a low total uncertainty of 0.0015. The lake's risk is linked to the presence of neighbouring lakes, such as Druk Chung Lake, which could trigger cascading effects and the downstream presence of the PHPA I & II hydropower projects along the Punatsangchhu River. Also, the neighbouring lakes have a history of outburst floods, evidencing the need for annual monitoring.

Finally, Figure 7.8 (e) classifies South Lhonak Lake (area = 1.69 km²), located in Sikkim, India, as a “Very High” risk with a probability of 0.941. The distribution is sharply peaked with a small tail extending into the “High” category (0.058), resulting in a low uncertainty of 0.0099. The lake's very high risk is associated with its setting in the Teesta River basin, a region prone

to intense rainfall, frequent landslides, pronounced effects of climate change, and the occurrence of three hydropower plants downstream, necessitating quarterly monitoring.

In the literature, various studies provide different risk levels to the lakes. For instance, Bajracharya et al. (2009), and Budhathoki et al. (2010) predicted Imja Tsho, Tsho Rolpa and Barun Tsho lakes of Nepal as high-risk lakes. Further, a study by Rounce et al. (2016) predicted Imja Tsho lake of moderate risk, Tsho Rolpa lake with high risk, and Barun Tsho lake with very high risk. Next, studies like Sattar et al. (2021), Sharma et al. (2018), Wangchuk et al. (2024), and Komori et al. (2012), to mention a few examples of evidence that Luggye Tsho and South Lhonak lakes are potentially dangerous and need a risk assessment. Importantly, the state of the lake changes with time, owing to climate change, varying glacier-lake dynamics, topography and downstream exposure. Thus, the predictions of the risk probabilities in this thesis, like Imja Tsho lake and Luggye Tsho lake as high-risk lakes, and Tsho Rolpa, Barun Tsho and South Lhonak lakes as very high-risk lakes, are well agreed with the published literature.

7.4.5 Comparative assessment of the adopted approach

The Bayesian neural network (BNN) framework developed in this study differs from Rounce et al. (2016), Veh et al. (2020), and Ahmed et al. (2022) GLOF assessments in both scope and methodology, while remaining broadly consistent with their main risk patterns.

7.4.5.1 Comparison with Rounce et al. (2016):

Rounce et al. (2016) proposed a fully remote, rule-based hazard and risk assessment for eight proglacial, moraine-dammed lakes in Nepal, focusing on triggers and downstream impacts. Their framework combines (i) dynamic triggers (snow/ice avalanches, rockfalls, and upstream GLOFs), (ii) moraine stability (steep lakefront area and buried ice), and (iii) simple GLOF routing (MC-LCP model) to derive lake-wise *hazard* ratings, and then couples these with qualitative *downstream impact* classes to obtain overall risk (low, moderate, high).

Our approach is conceptually similar in aiming at a transferable, remote assessment, but differs in three main ways:

1. **Broader parameter space and lake types** – The present study includes 14 parameters spanning lake geometry, topography, proximity to landslides, earthquakes and past GLOFs, climatic seasonality, neighbouring lakes, land use and proximity to

hydropower infrastructure, and we apply the model to all dam types (moraine-, ice- and bedrock-dammed) across the Eastern Himalaya. Rounce et al. (2016) focus on eight moraine-dammed lakes and do not explicitly include seismicity, meteorological extremes or anthropogenic exposure in their parameter set.

2. **From heuristic thresholds to hybrid MCDA + ML** – Rounce et al. (2016) rely on deterministic thresholds and expert rules for classifying hazard and risk. In contrast, we derive parameter weights from a hybrid AHP–Shannon Entropy scheme and use TOPSIS + Fuzzy C-Means clustering to generate soft, data-informed risk labels, which are then learned by a BNN. This combination allows non-linear interactions among parameters (e.g. joint effect of earthquake magnitude, lake area and proximity to hydropower) that are difficult to encode in rule-based schemes.
3. **Uncertainty-aware predictions** – Rounce et al. provide single risk labels per lake and do not quantify predictive uncertainty. Our BNN yields full probability distributions over four risk classes (low, moderate, high, very high) and explicitly decomposes total variance into aleatoric and epistemic components, giving both a risk class and a confidence measure (e.g. maximum variance 0.165 across the validation set, overall accuracy 93.4%). This is particularly important for prioritizing follow-up fieldwork and early-warning investments.

For the three lakes common to both studies (Imja Tsho, Tsho Rolpa and Barun Tsho), our results are consistent with the relative risk ordering in Rounce et al. (2016), but with updated, probabilistic classifications. Rounce et al. classify Imja Tsho as moderate risk (with future risk high), Tsho Rolpa as high risk and Lower Barun Tsho as very high risk. Our model classifies Imja as predominantly high risk with a strong tail into very high risk, Tsho Rolpa and Barun Tsho as very high risk with sharply peaked probability distributions. This suggests that the BNN reproduces the established hierarchy of hazardous lakes while reflecting recent changes in lake size and downstream exposure.

7.4.5.2 Comparison with Veh et al. (2020):

Veh et al. (2020) use a Bayesian extreme-value framework to estimate regional-scale GLOF hazard across 5,184 moraine-dammed lakes in the entire Himalaya. This study scales lake area to volume using robust regression, runs billions of dam-break simulations with a physical breach model, and combines these with historic GLOF rates to derive return-period curves for flood volume and peak discharge (e.g. regional 100-year GLOF discharge and volume). Their

key result is that the Eastern Himalayas are a hotspot of GLOF hazard, with 100-year flood discharges ~3× higher than in other Himalayan regions.

Relative to Veh et al. (2020), our study operates at a different, complementary scale:

- Veh et al. quantify **regional hydrologic hazard** (return-period discharge/volume) but do not assign lake-specific risk classes or consider local exposure.
- The present BNN framework provides **lake-scale, categorical risk maps** that integrate both hazard proxies and exposure (LULC, hydropower proximity) across the Eastern Himalaya.

Despite these differences, our spatial patterns agree with Veh et al. (2020): high and very-high risk lakes are concentrated in the Eastern Himalaya (Nepal, Bhutan, Sikkim, Arunachal Pradesh), which is also the region Veh et al. identify as having the highest 100-year GLOF hazard. This cross-scale consistency provides an indirect validation of our risk maps.

7.4.5.3 Comparison with Ahmed et al. (2022b):

Ahmed et al. (2022b) focus on the Dibang River Basin (Eastern Himalaya, India), where they (i) compile an inventory of 403 lakes, (ii) analyze area changes from 1985–2020, and (iii) identify 12 PDGLs and classify them into low, medium and high hazard using a weighted index method. The hazard index is built from six parameters: lake expansion rate, glacier proximity, dam-front slope, potential ice/snow avalanche zone, upstream/downstream lakes and presence of a calving glacier snout; each parameter is assigned discrete weights (0.25, 0.5, 1.0) for low/medium/high hazard levels and summed to obtain cumulative hazard scores. The approach is deterministic and does not explicitly include seismicity, climate variability or socio-economic exposure.

Our framework generalizes and extends this style of weighted-index assessment in several ways:

- **More comprehensive parameterization** – Three of Ahmed et al.'s six parameters (lake expansion, glacier proximity, cascaded lakes) are subsumed within our 14-parameter set, but we additionally incorporate distance to past GLOFs, landslides, earthquake proximity, earthquake magnitude and depth, temperature and precipitation seasonality, land-use/land-cover and distance to hydropower facilities.

- **Objective–subjective hybrid weighting instead of fixed scores** – Ahmed et al. (2022b) allocate fixed weights to parameter classes based on literature and expert judgement with equal-interval classification of cumulative scores. We explicitly combine expert pairwise comparisons (AHP) with data-driven Shannon Entropy weights to minimize divergence between the two, avoiding arbitrary index scaling and improving reproducibility.
- **Probabilistic, multi-class outputs** – Whereas Ahmed et al. (2022b) yield single hazard labels for 12 PDGLs in one basin, we produce probabilistic four-class risk predictions for 229 lakes across the entire Eastern Himalaya, with quantified uncertainty and ROC-based skill diagnostics (AUC ≈ 1 for all classes).

Therefore, the proposed BNN framework can be viewed as a regional, uncertainty-aware evolution of the weighted index methods used by Ahmed et al. (2022b) and similar PDGL studies.

7.5 Chapter conclusions

This chapter presents a unified, uncertainty-aware probabilistic framework that represents a significant methodological advance in GLOF risk assessment. By integrating fourteen critical parameters across geomorphological, topographical, meteorological, and anthropogenic domains, and by combining expert-driven AHP with the objectivity of SE, the chapter establishes a hybrid multi-criteria decision-making model grounded in both domain knowledge and data variability. The subsequent application of the TOPSIS method and FCM clustering produces soft-labeled risk classes, enabling a nuanced classification of glacial lake hazards.

The introduction of a BNN, trained on these risk labels, further strengthens the framework. The model captures nonlinear relationships and generates probabilistic predictions enriched with explicit uncertainty quantification. Results from the assessment of 269 glacial lakes show a high predictive accuracy of 93.4%. A total of 87.77% of lakes exhibit prediction variances below 0.1, indicating stable behaviour of the model. The framework identifies Imja Tsho and Luggye Tsho as high-risk lakes and classifies Tsho Rolpa, Barun Tsho, and South Lhonak as very high-risk lakes. These outcomes are consistent with previous field-based assessments while providing improved statistical confidence.

In addition to its methodological contributions, the framework has practical value for institutions responsible for hazard assessment and planning. Regional governments and State

Disaster Management Authorities can use the predicted likelihood maps to identify lake basins that require periodic inspection. The National Disaster Management Authority (NDMA) can incorporate the model outputs into their monitoring platforms to track changes in potentially hazardous lakes and to update their risk inventories. Hydropower developers can apply these predictions to assess upstream conditions during project planning and to modify safety measures for existing structures where upstream lake development may pose a threat. The uncertainty thresholds provide additional guidance by highlighting lakes that require field verification or targeted remote-sensing checks. These examples show that the framework can be integrated directly into existing monitoring and planning workflows used by government and infrastructure agencies.

Overall, this chapter contributes a scalable and generalisable approach to GLOF risk evaluation. It addresses a critical gap in current hazard modelling practice by linking scientific prediction with operational usability. The framework supports the development of more resilient early-warning systems and adaptive planning strategies to safeguard vulnerable mountain communities in a warming climate.

Chapter 8

Conclusions

Over the past three decades, the Eastern Himalayas have experienced a marked increase in both the number and spatial extent of glacial lakes, driven by a combination of climate change, seismic activity, dynamic hydrology, and evolving geomorphology and topography. These transformations have significantly raised the risk of GLOFs, posing serious threats to downstream communities, infrastructure, and ecologically sensitive mountain environments. In response to the increasing numbers of glacial lakes and continued expansion of the existing glacial lakes across the region, it has become increasingly important to update regional inventories, identify potentially dangerous glacial lakes (PDGLs), and assess associated GLOF hazards and risk levels to safeguard vulnerable high-altitude communities. Yet, most existing studies remain constrained by a narrow set of parameters and limited integration of multi-hazard drivers, while AI- and ML-based approaches, though promising, are still underutilized.

This thesis proposes a comprehensive, data-driven, and probabilistically framework for assessing GLOF hazards and risks in the Eastern Himalayas. The thesis is structured around four key objectives: (i) predicting glacial lake formation using ML, (ii) Hazard assessment of glacial lakes using a heuristic approach through deep learning, (iii) probabilistic and deterministic risk assessment of glacial lakes by integrating MCDA with hybrid weighting techniques and climate variables. These objectives are addressed through a sequence of distinct yet methodologically interconnected studies, culminating in a robust and scalable GLOF risk assessment framework tailored to mountainous environments. Following are the four major contributions of this thesis:

1. Glacial erosional features serve as critical geomorphic indicators of landscape evolution and hydrological dynamics in the Eastern Himalaya. Satellite-based interpretations further reveal active glacial retreat and its associated hydrological responses, emphasizing the region's rapidly transforming cryospheric environment. The spatial configuration and interaction of erosional features directly influence water accumulation, drainage pathways, and sediment transport—all of which govern the likelihood of glacial lake formation. Chapter 4 establishes the geomorphological basis of glacial lake formation, demonstrating that features such as cirques, valleys, flow channel (FC), and retreating glaciers (RG) govern the spatial occurrence of glacial lakes

in the ROI. Through a comparative analysis of LR, ANN, and BNN, the study finds BNN to be the most robust model for predictive mapping, offering superior accuracy (AUC = 0.878) and reliable uncertainty (variance), ranging between 10^{-3} and 10^{-4}. This chapter highlights the value of combining high-resolution erosional and topography features with probabilistic ML to forecast glacial lake formation under data-scarce conditions. The PGLF maps produced by the three ML models offer valuable insights for detecting existing glacial lakes, forecasting areas prone to future lake formation, and supporting hazard mitigation efforts associated with GLOFs in the Eastern Himalaya.

2. The spatial dynamics of glacial lakes in the Eastern Himalayas are strongly influenced by the interconnectedness of key erosional features. In many actively evolving glacial lakes, at least three erosional features, –cirques, retreating glaciers, flow channels, and valleys–are observed in conjunction with nearby high-elevation lakes. This spatial coherence plays a crucial role in shaping GLOF hazard, which depends not only on the intrinsic characteristics of individual lakes but also on their geomorphic context and presence of other glacial lakes at higher elevation (GLHE). However, most state-of-the-art studies have focused on detecting either erosional features or glacial lakes in isolation, without leveraging their spatial interrelationships for hazard assessment. To bridge this gap, Chapter 5 develops a semi-automated hazard assessment framework that incorporates the dynamic relationship between glacial lakes and geomorphic features under ongoing glacier retreat. An EfficientNet-B0 CNN based image classification pipeline is employed to detect critical erosional features, namely cirques, valleys, flow channels, retreating glaciers, and higher-elevation neighbouring lakes, followed by a heuristic rule-based classification of 2,647 glacial lakes using parameters such as area, elevation, and spatial coherence of geomorphic structures. The CNN model achieves a high segmentation accuracy across all features and effectively classifies glacial lakes into ‘extremely dangerous,’ ‘very highly dangerous,’ ‘highly dangerous,’ ‘moderately dangerous,’ ‘low dangerous,’ and ‘non-dangerous’ hazard classes. The framework also uses the proposed geomorphic analogies to identify and predict 174 potential sites, where glacial lakes are likely to form.
3. Dynamic geological processes, seismicity, and cryospheric changes cause sudden releases of water from glacial lakes are typically triggered by external events such as earthquakes, landslides, or rapid glacier melt, often resulting in catastrophic downstream impacts To incorporate both hazard and exposure dimensions, Chapter 6

proposes a hybrid MCDA framework for deterministic GLOF risk assessment across an extended ROI. The framework integrates physical, geomorphic, anthropogenic, historical, and tectonic indicators using a dual-weighting strategy: expert-based AHP and data-driven SE, fused via KL divergence. Risk categorization is performed using TOPSIS and FCM clustering, enabling a nuanced classification of 1,144 glacial lakes. The results, validated against historical GLOF records, identify 432 lakes as low risk, 368 as moderate, 227 as high, and 117 as very high risk. This hybrid methodology significantly enhances previous efforts by addressing key shortcomings such as parameter omission, multicollinearity, and subjectivity in weighting schemes, thereby offering a more robust and interpretable approach to GLOF risk assessment.

4. Despite the proliferation of flood risk assessment models in recent years, a persistent lack of consistency among their outputs has hindered the translation of research into effective policy and on-the-ground action—particularly in Eastern Himalayan region where timely, credible risk information is essential. Chapter 7 presents a unified, probabilistic risk assessment framework that couples the deterministic MCDA model with BNNs. The proposed model captures the nonlinear and uncertain nature of GLOF risk. The probabilistic model accurately classified 66 low-risk, 74 moderate-risk, 56 high-risk, and 18 very high-risk lakes, resulting in an overall classification accuracy of 93.4% and low uncertainty variance in the validation set. The proposed framework can be used as a scalable, generalizable tool for probabilistic risk assessment in glaciated mountain regions globally.

Conclusively, this thesis offers a scalable and uncertainty-aware framework that uses satellite images, geomorphic, seismic, geological, and topographical data for GLOF hazard and risk assessment for a large glaciated terrain. The developed models support scientific decision-making by enabling prioritization of hazardous lakes for monitoring, mitigation, and early warning interventions. The integration of meteorological triggers allows for dynamic, climate-responsive risk evaluation, while predictive uncertainty enhances confidence in remote assessments. The framework also aids in expanding glacial lake inventories by identifying potential future lakes, guiding anticipatory adaptation. The methodologies presented here–combining hybrid weighting, probabilistic deep learning, and fuzzy clustering–are transferable to other high mountain systems such as the Andes, Alps, or Pamir–Tien Shan ranges, and contribute as valuable tools for advancing cryosphere hazard research and climate resilience planning.

8.1 Future Scope:

While this thesis presents a comprehensive framework for predicting, assessing, and probabilistically evaluating glacial lake outburst flood (GLOF) hazards in the Eastern Himalayas, several avenues remain open for further investigation and methodological refinement. A critical extension involves the integration of depositional geomorphological features into hazard assessment. Specifically, detailed analyses of the chronology, geology, and stratigraphy of moraine complexes, sediment transport dynamics, and valley-fill characteristics could provide deeper insights into the long-term stability of glacial lakes and the mechanisms governing potential outburst events.

Future research may also address the simplifying assumption of parameter independence applied during the weighting process. Interactions among correlated variables—such as slope, elevation, and seismic proximity—are complex, and oversimplification may compromise the accuracy of risk assessments. Incorporating multivariate dependency structures or advanced correlation modelling would therefore enhance the reliability of hazard evaluation.

Furthermore, the predictive capacity of probabilistic risk assessment can be expanded through the application of advanced machine learning frameworks. Approaches such as Monte Carlo dropout, deep ensemble neural networks, mixture density networks, and self-organizing maps (SOMs) offer diverse strategies to capture both epistemic and aleatoric uncertainties, thereby improving model robustness and interpretability. Beyond these data-driven methods, physics-informed neural networks (PINNs) present a promising direction for integrating physical constraints into learning architectures. By embedding glaciological and hydrological principles directly into the modelling process, PINNs can ensure greater consistency between computational outputs and real-world dynamics.

List of Publications

Articles from PhD Thesis	
Journal Articles	1. **A. Vashistha**, A. Dashora, A. A. Shah, & A. Pal (2025). Hazard assessment of glacial lakes in Himalaya through erosional features detection using deep learning. ***Natural Hazards, 121, 14847–14870***. 2. **A. Vashistha**, A. A. Shah., N. Batmanathan, & A. Dashora (2025). A data-driven multi-criteria model for GLOF risk in tectonically active Himalayan regions. ***Modeling Earth Systems and Environment, 11(5), 333***. 3. **A. Vashistha**, A. Dashora, & A. A. Shah (2025). Machine Learning-Based Probabilistic Prediction of Glacial Lake Formation Using Erosional and Topographic Features. ***Scientific Reports, 15, 38673.*** 4. **A. Vashistha**, A. Dashora, & A. A. Shah (2025). Assessing Glacial Lake Outburst Flood Risk in the Eastern Himalayas: A Bayesian Neural Network Framework. ***Natural Hazards, 121, 21861-21890.***
Articles beyond PhD Thesis	
Journal Articles	1. A. A. Shah, M. G. Rachman, R. Kumar, **A. Vashistha**, A. Dashora, & M. Mahoor (2024). Pre-disposed tectonic subsidence controls flood hazards and unplanned urbanisation dominates the flood disasters in the Pliocene to Holocene Kashmir basin, NW Himalayas. ***Quaternary Science Advances, 14, 100173***. 2. A. A. Shah, M. G. Rachman, R. Kumar, **A. Vashistha**, & A. Dashora (2024). Satellite Data-based Structural Mapping Reveals Active Panjal Traps Fault (PTF) in Kashmir, NW Himalaya. ***Acta Geologica Sinica - English Edition, 98: 58-61***.
Conference	1. **A. Vashistha**, A. Dashora, & A. A. Shah (2023). Sparse Station Placement for Optimal Reconstruction of Rain Gauge Data and Performance Assessment of Satellite Precipitation Estimates. In Proc. ***World Congress on Disaster Management (WCDM), Dehradun, India.***

References

Agustina, R. D., Putra, R. P., & Susanti, S. (2023). Mapping Greater Bandung flood susceptibility based on multi-criteria decision analysis (MCDA) using AHP method. *Environmental earth sciences*, *82*(15), 370.

Ahmed, R., Ahmad, S. T., Wani, G. F., Mir, R. A., Ahmed, P., & Jain, S. K. (2022a). High resolution inventory and hazard assessment of potentially dangerous glacial Lakes in Upper Jhelum Basin, Kashmir Himalaya, India. *Geocarto International*, *37*(25), 10681-10712.

Ahmed, R., Ahmad, S. T., Wani, G. F., Mir, R. A., Almazroui, M., Bansal, J. K., & Ahmed, P. (2022b). Glacial lake changes and the identification of potentially dangerous glacial lakes (PDGLs) under warming climate in the Dibang River Basin, Eastern Himalaya, India. *Geocarto International*, *37*(27), 17659-17685.

Ahmed, R., Wani, G. F., Ahmad, S. T., Mir, R. A., Almazroui, M., Jain, S. K., & Ahmed, P. (2021a). Spatiotemporal dynamics of glacial lakes (1990–2018) in the Kashmir Himalayas, India using Remote Sensing and GIS. *Discover Water*, *1*, 1-17.

Ahmed, R., Wani, G. F., Ahmad, S. T., Sahana, M., Singh, H., & Ahmed, P. (2021b). A review of glacial lake expansion and associated glacial lake outburst floods in the Himalayan region. *Earth Systems and Environment*, *5*(3), 695-708.

Ai, S., Wang, Z., E, D., Holmen, K., Tan, Z., Zhou, C., & Sun, W. (2014). Topography, ice thickness and ice volume of the glacier Pedersenbreen in Svalbard, using GPR and GPS. *Polar Research*, *33*(1), 18533.

Alifu, H., Vuillaume, J.-F., Johnson, B. A., & Hirabayashi, Y. (2020). Machine-learning classification of debris-covered glaciers using a combination of Sentinel-1/-2 (SAR/optical), Landsat 8 (thermal) and digital elevation data. *Geomorphology*, *369*, 107365.

Allen, S. K., Linsbauer, A., Randhawa, S., Huggel, C., Rana, P., & Kumari, A. (2016). Glacial lake outburst flood risk in Himachal Pradesh, India: an integrative and anticipatory approach considering current and future threats. *Natural Hazards*, *84*, 1741-1763.

Allen, S. K., Zhang, G., Wang, W., Yao, T., & Bolch, T. (2019). Potentially dangerous glacial lakes across the Tibetan Plateau revealed using a large-scale automated assessment approach. *Science Bulletin*, *64*(7), 435-445.

Aloisio, A., Rosso, M. M., De Leo, A. M., Fragiacomo, M., & Basi, M. (2023). Damage classification after the 2009 L'Aquila earthquake using multinomial logistic regression and neural networks. *International Journal of Disaster Risk Reduction*, *96*, 103959.

Anand, V., Gandhi, F., & Songara, J. (2025). Assessing and managing glacial lake outburst flood risks in the Indian Himalayas: a comprehensive review. *Natural Hazards*, *121*(15), 17253-17281.

Anilkumar, R., Bharti, R., Chutia, D., & Aggarwal, S. P. (2023). Modelling point mass balance for the glaciers of the Central European Alps using machine learning techniques. *The Cryosphere*, *17*(7), 2811-2828.

Arora, A., Pandey, M., Siddiqui, M. A., Hong, H., & Mishra, V. N. (2021). Spatial flood susceptibility prediction in Middle Ganga Plain: comparison of frequency ratio and Shannon's entropy models. *Geocarto International*, *36*(18), 2085-2116.

Avouac, J.-P. (2003). Mountain building, erosion, and the seismic cycle in the Nepal Himalaya. *Advances in geophysics*, *46*, 1-80.

Bælum, K., & Benn, D. (2011). Thermal structure and drainage system of a small valley glacier (Tellbreen, Svalbard), investigated by ground penetrating radar. *The Cryosphere*, *5*(1), 139-149.

Bajracharya, S. R., Maharjan, S. B., Shrestha, F., Bajracharya, O. R., & Baidya, S. (2014). Glacier status in Nepal and decadal change from 1980 to 2010 based on Landsat data. *(No Title)*.

Bajracharya, S. R., & Mool, P. (2009). Glaciers, glacial lakes and glacial lake outburst floods in the Mount Everest region, Nepal. *Annals of Glaciology*, *50*(53), 81-86.

Bajracharya, S. R., Mool, P. K., & Shrestha, B. R. (2007). *Impact of climate change on Himalayan glaciers and glacial lakes: case studies on GLOF and associated hazards in Nepal and Bhutan* (Vol. 12). International Centre for Integrated Mountain Development Kathmandu.

Barandun, M., Bravo, C., Grobety, B., Jenk, T., Fang, L., Naegeli, K., Rivera, A., Cisternas, S., Münster, T., & Schwikowski, M. (2022). Anthropogenic influence on surface changes at the Olivares glaciers; Central Chile. *Science of the Total Environment*, *833*, 155068.

Barnett, T. P., Adam, J. C., & Lettenmaier, D. P. (2005). Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, *438*(7066), 303-309.

Basit, A., Bhatti, M. K., Ali, M., Fatima, T., Minchew, B., & Siddique, M. A. (2022). Deep Learning for Monitoring Glacial Lakes Formation using Sentinel 2 Multispectral Data. IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium,

Baumhoer, C. A., Dietz, A. J., Kneisel, C., & Kuenzer, C. (2019). Automated extraction of antarctic glacier and ice shelf fronts from sentinel-1 imagery using deep learning. *Remote Sensing*, *11*(21), 2529.

Bazai, N. A., Cui, P., Carling, P. A., Wang, H., Hassan, J., Liu, D., Zhang, G., & Jin, W. (2021). Increasing glacial lake outburst flood hazard in response to surge glaciers in the Karakoram. *Earth-Science Reviews*, *212*, 103432.

Benn, D., & Evans, D. J. (2014). *Glaciers and glaciation*. Routledge.

Bertone, A., Zucca, F., Marin, C., Notarnicola, C., Cuozzo, G., Krainer, K., Mair, V., Riccardi, P., Callegari, M., & Seppi, R. (2019). An unsupervised method to detect rock glacier activity by using Sentinel-1 SAR interferometric coherence: a regional-scale study in the Eastern European Alps. *Remote Sensing*, *11*(14), 1711.

Bhambri, R., Bolch, T., & Chaujar, R. (2011). Mapping of debris-covered glaciers in the Garhwal Himalayas using ASTER DEMs and thermal data. *International journal of remote sensing*, *32*(23), 8095-8119.

Bilham, R., & England, P. (2001). Plateau 'pop-up'in the great 1897 Assam earthquake. *Nature*, *410*(6830), 806-809.

Bilham, R., Mencin, D., Bendick, R., & Bürgmann, R. (2017). Implications for elastic energy storage in the Himalaya from the Gorkha 2015 earthquake and other incomplete ruptures of the Main Himalayan Thrust. *Quaternary International*, *462*, 3-21.

Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural network. International conference on machine learning,

Bolch, T. (2017). Asian glaciers are a reliable water source. *Nature*, *545*(7653), 161-162.

Bolch, T., Buchroithner, M. F., Peters, J., Baessler, M., & Bajracharya, S. (2008). Identification of glacier motion and potentially dangerous glacial lakes in the Mt. Everest region/Nepal using spaceborne imagery. *Natural Hazards and Earth System Sciences*, *8*(6), 1329-1340.

Bolch, T., Peters, J., Yegorov, A., Pradhan, B., Buchroithner, M., & Blagoveshchensky, V. (2011). Identification of potentially dangerous glacial lakes in the northern Tien Shan. *Natural Hazards*, *59*, 1691-1714.

Bolibar, J., Rabatel, A., Gouttevin, I., & Galiez, C. (2020). A deep learning reconstruction of mass balance series for all glaciers in the French Alps: 1967–2015. *Earth System Science Data*, *12*(3), 1973-1983.

Boulton, G. (1979). Processes of glacier erosion on different substrata. *Journal of Glaciology*, *23*(89), 15-38.

Brenning, A., Long, S., & Fieguth, P. (2012). Detecting rock glacier flow structures using Gabor filters and IKONOS imagery. *Remote Sensing of Environment*, *125*, 227-237.

Brochu, E., Cora, V. M., & De Freitas, N. (2010). A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning. *arXiv preprint arXiv:1012.2599*.

Budhathoki, K. P., Bajracharya, O., & Pokharel, B. (2010). Assessment of Imja Glacier Lake outburst flood (GLOF) risk in Dudh Koshi River Basin using remote sensing techniques. *Journal of Hydrology and Meteorology*, *7*(1), 75-91.

Carey, M. (2005). Living and dying with glaciers: people's historical vulnerability to avalanches and outburst floods in Peru. *Global and Planetary Change*, *47*(2-4), 122-134.

Carrivick, J. L., & Tweed, F. S. (2016). A global assessment of the societal impacts of glacier outburst floods. *Global and Planetary Change*, *144*, 1-16.

Cenderelli, D. A., & Wohl, E. E. (2003). Flow hydraulics and geomorphic effects of glacial-lake outburst floods in the Mount Everest region, Nepal. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group*, *28*(4), 385-407.

Chaturvedi, R. K., Kulkarni, A., Karyakarte, Y., Joshi, J., & Bala, G. (2014). Glacial mass balance changes in the Karakoram and Himalaya based on CMIP5 multi-model climate projections. *Climatic change*, *123*(2), 315-328.

Che, T., Xiao, L., & Liou, Y.-A. (2014). Changes in glaciers and glacial lakes and the identification of dangerous glacial lakes in the Pumqu River Basin, Xizang (Tibet). *Advances in meteorology*, *2014*.

Chelamallu, H. P., Venkataraman, G., Murti, M., Arora, M., & Singh, G. (2014). Comparison of SRM and SNOWMOD models using modis snow cover data for Bhagirathi river basin in the Himalayas. 2014 IEEE Geoscience and Remote Sensing Symposium,

Chen, H., Liang, Q., Zhao, J., & Maharjan, S. B. (2025). Assessing national exposure to and impact of glacial lake outburst floods considering uncertainty under data sparsity. *Hydrol. Earth Syst. Sci.*, *29*(3), 733-752. https://doi.org/10.5194/hess-29-733-2025

Chen, T., Fox, E., & Guestrin, C. (2014). Stochastic gradient hamiltonian monte carlo. International conference on machine learning,

Chen, Y., Wang, D., Zhang, L., Guo, H., Ma, J., & Gao, W. (2023). Flood risk assessment of Wuhan, China, using a multi-criteria analysis model with the improved AHP-Entropy method. *Environmental Science and Pollution Research*, *30*(42), 96001-96018.

Clague, J. J., & Mathews, W. H. (1973). The magnitude of jökulhlaups. *Journal of Glaciology*, *12*(66), 501-504.

Cook, K. L., Andermann, C., Gimbert, F., Adhikari, B. R., & Hovius, N. (2018). Glacial lake outburst floods as drivers of fluvial erosion in the Himalaya. *Science*, *362*(6410), 53-57.

Cook, S. J., & Swift, D. A. (2012). Subglacial basins: Their origin and importance in glacial systems and landscapes. *Earth-Science Reviews*, *115*(4), 332-372.
Cuffey, K. M., & Paterson, W. S. B. (2010). *The physics of glaciers*. Academic Press.
Dar, R. A., Jaan, O., Murtaza, K. O., & Romshoo, S. A. (2017). Glacial-geomorphic study of the Thajwas glacier valley, Kashmir Himalayas, India. *Quaternary International*, *444*, 157-171.
Das, A., Singh, S. K., Kanga, S., Sajan, B., Meraj, G., & Kumar, P. (2024). Glacial Lake Outburst Flood Susceptibility Mapping in Sikkim: A Comparison of AHP and Fuzzy AHP Models. *Climate (2225-1154)*, *12*(11).
De, S., Britton, J., Reynolds, M., Skinner, R., Jansen, K., & Doostan, A. (2020). On transfer learning of neural networks using bi-fidelity data for uncertainty propagation. *International Journal for Uncertainty Quantification*, *10*(6).
Dobreva, I. D., & Klein, A. G. (2011). Fractional snow cover mapping through artificial neural network analysis of MODIS surface reflectance. *Remote Sensing of Environment*, *115*(12), 3355-3366.
Drenkhan, F., Guardamino, L., Huggel, C., & Frey, H. (2018). Current and future glacier and lake assessment in the deglaciating Vilcanota-Urubamba basin, Peruvian Andes. *Global and Planetary Change*, *169*, 105-118.
Dyurgerov, M., & Meier, M. F. (2005). *Glaciers and the changing Earth system: a 2004 snapshot* (Vol. 58). Institute of Arctic and Alpine Research, University of Colorado Boulder.
Emmer, A., Klimeš, J., Mergili, M., Vilímek, V., & Cochachin, A. (2016). 882 lakes of the Cordillera Blanca: An inventory, classification, evolution and assessment of susceptibility to outburst floods. *Catena*, *147*, 269-279.
England, P., & Bilham, R. (2015). The Shillong Plateau and the great 1897 Assam earthquake. *Tectonics*, *34*(9), 1792-1812.
Evans, I. S., & Cox, N. (1974). Geomorphometry and the operational definition of cirques. *Area*, 150-153.
Feby, B., Achu, A., Jimnisha, K., Ayisha, V., & Reghunath, R. (2020). Landslide susceptibility modelling using integrated evidential belief function based logistic regression method: A study from Southern Western Ghats, India. *Remote Sensing Applications: Society and Environment*, *20*, 100411.
Feyisa, G. L., Meilby, H., Fensholt, R., & Proud, S. R. (2014). Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment*, *140*, 23-35.
Fischer, A., Seiser, B., Stocker Waldhuber, M., Mitterer, C., & Abermann, J. (2015). Tracing glacier changes in Austria from the Little Ice Age to the present using a lidar-based high-resolution glacier inventory in Austria. *The Cryosphere*, *9*(2), 753-766.
Flint, R. F. (1971). Glacial and Quaternary geology.
Furian, W., Loibl, D., & Schneider, C. (2021). Future glacial lakes in High Mountain Asia: an inventory and assessment of hazard potential from surrounding slopes. *Journal of Glaciology*, *67*(264), 653-670.
Gal, Y., & Ghahramani, Z. (2016). A theoretically grounded application of dropout in recurrent neural networks. *Advances in neural information processing systems*, *29*.
Ganjoo, R., & Koul, M. (2025). Mountain Geo-Hazards of NW Himalaya. In *The Himalayan Glaciers, Climate Change and Society: A Case Study of the Northwestern Himalayas, India* (pp. 127-146). Springer.
Gardelle, J., Arnaud, Y., & Berthier, E. (2011). Contrasted evolution of glacial lakes along the Hindu Kush Himalaya mountain range between 1990 and 2009. *Global and Planetary Change*, *75*(1-2), 47-55.
Gardelle, J., Berthier, E., Arnaud, Y., & Kääb, A. (2013). Region-wide glacier mass balances over the Pamir-Karakoram-Himalaya during 1999–2011. *The Cryosphere*, *7*(4), 1263-1286.
Gopalan, G., Hrafnkelsson, B., Aðalgeirsdóttir, G., Jarosch, A. H., & Pálsson, F. (2018). A Bayesian hierarchical model for glacial dynamics based on the shallow ice approximation and its evaluation using analytical solutions. *The Cryosphere*, *12*(7), 2229-2248.
Goutham, V., Sameerunnisa, A., Babu, S., & Prakash, T. B. (2022). Brain tumor classification using Efficientnet-B0 model. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE),
Govindha Raj, K. B. (2010). Remote sensing based hazard assessment of glacial lakes: a case study in Zanskar basin, Jammu and Kashmir, India. *Geomatics, Natural Hazards and Risk*, *1*(4), 339-347.
Guillet, G., & Bolch, T. (2023). Bayesian estimation of glacier surface elevation changes from DEMs. *Frontiers in Earth Science*, *11*, 1076732.
Gurung, D. R., Khanal, N. R., Bajracharya, S. R., Tsering, K., Joshi, S., Tshering, P., Chhetri, L. K., Lotay, Y., & Penjor, T. (2017). Lemthang Tsho glacial Lake outburst flood (GLOF) in Bhutan: cause and impact. *Geoenvironmental Disasters*, *4*, 1-13.
Haeberli, W., Schaub, Y., & Huggel, C. (2017). Increasing risks related to landslides from degrading permafrost into new lakes in de-glaciating mountain ranges. *Geomorphology*, *293*, 405-417.

Hall, D. K., Riggs, G. A., & Salomonson, V. V. (1995). Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data. *Remote Sensing of Environment*, *54*(2), 127-140.
Hamlet, A. F., & Lettenmaier, D. P. (1999). Effects of climate change on hydrology and water resources in the Columbia River Basin 1. *JAWRA Journal of the American Water Resources Association*, *35*(6), 1597-1623.
Haq, M. A., Azam, M. F., & Vincent, C. (2021). Efficiency of artificial neural networks for glacier ice-thickness estimation: A case study in western Himalaya, India. *Journal of Glaciology*, *67*(264), 671-684.
Harrison, S., Kargel, J. S., Huggel, C., Reynolds, J., Shugar, D. H., Betts, R. A., Emmer, A., Glasser, N., Haritashya, U. K., & Klimeš, J. (2018). Climate change and the global pattern of moraine-dammed glacial lake outburst floods. *The Cryosphere*, *12*(4), 1195-1209.
Hartmann, A., Davari, A., Seehaus, T., Braun, M., Maier, A., & Christlein, V. (2021). Bayesian U-net for segmenting glaciers in SAR imagery. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS,
Hazra, P., & Krishna, A. P. (2022). Assessment of proglacial lakes in Sikkim Himalaya, India for glacial lake outburst flood (GLOF) risk analysis using HEC-RAS and geospatial techniques. *Journal of the Geological Society of India*, *98*(3), 344-352.
He, X., Wang, Z., Wang, C., Wang, S., Fan, S., & Zeng, Y. (2022). A method for estimating the probability of glacial lake outburst floods based on logistic regression and geodetector: a case study of the Himalayan region. *Earth Science Informatics*, *15*(1), 649-658.
Hou, J., & Huang, C. (2014). Improving mountainous snow cover fraction mapping via artificial neural networks combined with MODIS and ancillary topographic data. *IEEE Transactions on Geoscience and Remote Sensing*, *52*(9), 5601-5611.
Huang, J., Hales, T. C., Huang, R., Ju, N., Li, Q., & Huang, Y. (2020). A hybrid machine-learning model to estimate potential debris-flow volumes. *Geomorphology*, *367*, 107333.
Huang, L., Li, Z., Zhou, J. M., & Zhang, P. (2021). An automatic method for clean glacier and nonseasonal snow area change estimation in High Mountain Asia from 1990 to 2018. *Remote Sensing of Environment*, *258*, 112376.
Huggel, C. (2004). *Assessment of glacial hazards based on remote sensing and GIS modeling* (Vol. 44). University of Zurich.
Huggel, C., Haeberli, W., Kääb, A., Bieri, D., & Richardson, S. (2004). An assessment procedure for glacial hazards in the Swiss Alps. *Canadian Geotechnical Journal*, *41*(6), 1068-1083.
Huggel, C., Kääb, A., Haeberli, W., Teysseire, P., & Paul, F. (2002). Remote sensing based assessment of hazards from glacier lake outbursts: a case study in the Swiss Alps. *Canadian Geotechnical Journal*, *39*(2), 316-330.
Huybrechts, P. (2002). Sea-level changes at the LGM from ice-dynamic reconstructions of the Greenland and Antarctic ice sheets during the glacial cycles. *Quaternary Science Reviews*, *21*(1-3), 203-231.
Immerzeel, W. W., Van Beek, L., Konz, M., Shrestha, A., & Bierkens, M. (2012). Hydrological response to climate change in a glacierized catchment in the Himalayas. *Climatic change*, *110*, 721-736.
Indolia, S., Goswami, A. K., Mishra, S. P., & Asopa, P. (2018). Conceptual understanding of convolutional neural network-a deep learning approach. *Procedia computer science*, *132*, 679-688.
Intergovernmental Panel on Climate, C. (2023). *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. https://doi.org/DOI: 10.1017/9781009325844
Islam, N., & Patel, P. P. (2022). Inventory and GLOF hazard assessment of glacial lakes in the Sikkim Himalayas, India. *Geocarto International*, *37*(13), 3840-3876.
Jain, R., Sharma, R., Tiwari, D., Joshi, K., & Jain, V. (2024). Enhanced Classification of Intel Images Using Refined EfficientNet and MobileNetV2 Frameworks. 2023 4th International Conference on Intelligent Technologies (CONIT),
Jain, S. K., Kumar, V., & Saharia, M. (2013). Analysis of rainfall and temperature trends in northeast India. *International Journal of Climatology*, *33*(4), 968-978.
Janke, J. R. (2013). Using airborne LiDAR and USGS DEM data for assessing rock glaciers and glaciers. *Geomorphology*, *195*, 118-130.
Jena, R., Pradhan, B., Naik, S. P., & Alamri, A. M. (2021). Earthquake risk assessment in NE India using deep learning and geospatial analysis. *Geoscience Frontiers*, *12*(3), 101110.
Jospin, L. V., Laga, H., Boussaid, F., Buntine, W., & Bennamoun, M. (2022). Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, *17*(2), 29-48.
Kalimullah, N. M., Shelke, A., & Habib, A. (2023a). A deep learning approach for anomaly identification in PZT sensors using point contact method. *Smart Materials and Structures*, *32*(9), 095027.

Kalimullah, N. M., Shelke, A., & Habib, A. (2023b). A probabilistic framework for source localization in anisotropic composite using transfer learning based multi-fidelity physics informed neural network (mfPINN). *Mechanical Systems and Signal Processing*, *197*, 110360.

Kargel, J. S., Leonard, G. J., Shugar, D. H., Haritashya, U. K., Bevington, A., Fielding, E. J., Fujita, K., Geertsema, M., Miles, E., & Steiner, J. (2016). Geomorphic and geologic controls of geohazards induced by Nepal's 2015 Gorkha earthquake. *Science*, *351*(6269), aac8353.

Karimi, N., Farokhnia, A., Karimi, L., Eftekhari, M., & Ghalkhani, H. (2012). Combining optical and thermal remote sensing data for mapping debris-covered glaciers (Alamkouh Glaciers, Iran). *Cold Regions Science and Technology*, *71*, 73-83.

Kaser, G., Großhauser, M., & Marzeion, B. (2010). Contribution potential of glaciers to water availability in different climate regimes. *Proceedings of the National Academy of Sciences*, *107*(47), 20223-20227.

Kaushik, S., Rafiq, M., Dharpure, J. K., Howat, I., Moortgat, J., Joshi, P., Singh, T., & Dietz, A. J. (2024). Increasing risk of glacial lake outburst flood in Sikkim, Eastern Himalaya under climate warming. *Remote Sensing Applications: Society and Environment*, *36*, 101286.

Kaushik, S., Singh, T., Joshi, P. K., & Dietz, A. J. (2022). Automated mapping of glacial lakes using multisource remote sensing data and deep convolutional neural network. *International Journal of Applied Earth Observation and Geoinformation*, *115*, 103085.

Keefer, D. K. (1984). Landslides caused by earthquakes. *Geological Society of America Bulletin*, *95*(4), 406-421.

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? *Advances in neural information processing systems*, *30*.

Khalid, H. W., Khalil, R. M. Z., & Qureshi, M. A. (2021). Evaluating spectral indices for water bodies extraction in western Tibetan Plateau. *The Egyptian Journal of Remote Sensing and Space Science*, *24*(3), 619-634.

Khan, A. A., Jamil, A., Hussain, D., Taj, M., Jabeen, G., & Malik, M. K. (2020). Machine-learning algorithms for mapping debris-covered glaciers: the Hunza Basin case study. *IEEE Access*, *8*, 12725-12734.

Kim, J., Yu, J., Wang, L., Liu, H., & Shin, H. (2016). Morphological Characteristics of the Ice Margins of Antarctic Ice Shelf and Outlet Glacier Extracted from ICESat Laser Altimetry Along-Track Profiles. *Terrestrial, Atmospheric & Oceanic Sciences*, *27*(4).

Kirschbaum, D., Stanley, T., & Zhou, Y. (2015). Spatial and temporal analysis of a global landslide catalog. *Geomorphology*, *249*, 4-15.

Kirschbaum, D. B., Adler, R., Hong, Y., Hill, S., & Lerner-Lam, A. (2010). A global landslide catalog for hazard applications: method, results, and limitations. *Natural Hazards*, *52*, 561-575.

Klimeš, J., Benešová, M., Vilímek, V., Bouška, P., & Cochachin Rapre, A. (2014). The reconstruction of a glacial lake outburst flood using HEC-RAS and its significance for future hazard assessments: an example from Lake 513 in the Cordillera Blanca, Peru. *Natural Hazards*, *71*(3), 1617-1638.

Komori, J., Koike, T., Yamanokuchi, T., & Tshering, P. (2012). Glacial lake outburst events in the Bhutan Himalayas. *Global Environmental Research*, *16*(1), 59-70.

Kougkoulos, I., Cook, S. J., Jomelli, V., Clarke, L., Symeonakis, E., Dortch, J. M., Edwards, L. A., & Merad, M. (2018). Use of multi-criteria decision analysis to identify potentially dangerous glacial lakes. *Science of The Total Environment*, *621*, 1453-1466.

Krishna, A. P. (2005). Snow and glacier cover assessment in the high mountains of Sikkim Himalaya. *Hydrological Processes: An International Journal*, *19*(12), 2375-2383.

Kulkarni, A., Srinivasulu, J., & Manjul, S. (2002). Field based spectral reflectance studies to develop NDSI method for snow cover monitoring. *Journal of the Indian Society of Remote Sensing*, *30*, 73-80.

Kumar, P., & Sarkar, P. (2022). A comparison of the AHP and TOPSIS multi-criteria decision-making tools for prioritizing sub-watersheds using morphometric parameters' analysis. *Modeling Earth Systems and Environment*, *8*(3), 3973-3983.

Kumar, R., Bahuguna, I., Ali, S. N., & Singh, R. (2020). Lake inventory and evolution of glacial lakes in the Nubra-Shyok basin of Karakoram Range. *Earth Systems and Environment*, *4*(1), 57-70.

Kumar, S., & Sharma, N. (2019). The seismicity of central and north-east Himalayan region. *Contributions to Geophysics & Geodesy*, *49*(3).

Kumar, V., & Venkataraman, G. (2012). High Resolution TerraSAR-X Image Speckle Suppression and its Fusion with Multispectral IRS LISS–III Data for Himalayan Glacier Feature Extraction. *Journal of the Indian Society of Remote Sensing*, *40*, 325-334.

Kwon, Y., Won, J.-H., Kim, B. J., & Paik, M. C. (2020). Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation. *Computational Statistics & Data Analysis*, *142*, 106816.

Li, H., Jiang, J., Chen, B., Li, Y., Xu, Y., & Shen, W. (2016). Pattern of NDVI-based vegetation greening along an altitudinal gradient in the eastern Himalayas and its response to global warming. *Environmental monitoring and assessment*, *188*, 1-10.

Li, Y., & Zhao, Z. (2022). AutoCirque: An automated method to delineate glacial cirque outlines from digital elevation models. *Geomorphology*, *398*, 108059.
Linsbauer, A., Paul, F., & Haeberli, W. (2012). Modeling glacier thickness distribution and bed topography over entire mountain ranges with GlabTop: Application of a fast and robust approach. *Journal of Geophysical Research: Earth Surface*, *117*(F3).
Liu, J.-J., Cheng, Z.-L., & Su, P.-C. (2014). The relationship between air temperature fluctuation and Glacial Lake Outburst Floods in Tibet, China. *Quaternary International*, *321*, 78-87.
Liu, M., Chen, N., Zhang, Y., & Deng, M. (2020). Glacial lake inventory and lake outburst flood/debris flow hazard assessment after the Gorkha earthquake in the Bhote Koshi Basin. *Water*, *12*(2), 464.
Liu, S. C., Fu, C., Shiu, C. J., Chen, J. P., & Wu, F. (2009). Temperature dependence of global precipitation extremes. *Geophysical Research Letters*, *36*(17).
Liu, X., & Chen, B. (2000). Climatic warming in the Tibetan Plateau during recent decades. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *20*(14), 1729-1742.
Lu, Y., Zhang, Z., Shangguan, D., & Yang, J. (2021). Novel machine learning method integrating ensemble learning and deep learning for mapping debris-covered glaciers. *Remote Sensing*, *13*(13), 2595.
Lutz, A., Immerzeel, W., Shrestha, A. B., & Bierkens, M. (2014). Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nature Climate Change*, *4*(7), 587-592.
Lützow, N., Veh, G., & Korup, O. (2023). A global database of historic glacier lake outburst floods. *Earth System Science Data Discussions*, *2023*, 1-27.
Ma, L., Tian, L., Pu, J., & Wang, P. (2010). Recent area and ice volume change of Kangwure Glacier in the middle of Himalayas. *Chinese Science Bulletin*, *55*, 2088-2096.
Machguth, H., Eisen, O., Paul, F., & Hoelzle, M. (2006). Strong spatial variability of snow accumulation observed with helicopter-borne GPR on two adjacent Alpine glaciers. *Geophysical research letters*, *33*(13).
Mal, S., Allen, S. K., Frey, H., Huggel, C., & Dimri, A. (2021). Sectorwise assessment of glacial lake outburst flood danger in the Indian Himalayan region. *Mountain Research and Development*, *41*(1), R1.
Mal, S., Kumar, A., Bhambri, R., Schickhoff, U., & Singh, R. (2020). Inventory and spatial distribution of glacial lakes in Arunachal Pradesh, Eastern Himalaya, India. *Journal of the Geological Society of India*, *96*, 609-615.
Malczewski, J., Rinner, C., Malczewski, J., & Rinner, C. (2015). Introduction to GIS-mcda. *Multicriteria decision analysis in geographic information science*, 23-54.
McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425-1432.
McKillop, R. J., & Clague, J. J. (2007). A procedure for making objective preliminary assessments of outburst flood hazard from moraine-dammed lakes in southwestern British Columbia. *Natural Hazards*, *41*, 131-157.
Mehta, M., Majeed, Z., Dobhal, D., & Srivastava, P. (2012). Geomorphological evidences of post-LGM glacial advancements in the Himalaya: A study from Chorabari Glacier, Garhwal Himalaya, India. *Journal of Earth System Science*, *121*, 149-163.
Meier, M., & Bahr, D. (1996). Counting Glaciers: Use of Scaling Methods to. *Glaciers, ice sheets and volcanoes: a tribute to Mark F. Meier*, *96*(27), 89.
Mohajerani, Y. (2019). *Understanding Regional Ice Sheet Mass Balance: Remote Sensing, Regional Climate Models, and Deep Learning*. University of California, Irvine.
Mohanty, L. K., & Maiti, S. (2022). Glacial lake formation probability mapping in the Himalayan glacier: A probabilistic approach. *Journal of Earth System Science*, *131*(1), 54.
Molnar, P., & Tapponnier, P. (1975). Cenozoic Tectonics of Asia: Effects of a Continental Collision: Features of recent continental tectonics in Asia can be interpreted as results of the India-Eurasia collision. *Science*, *189*(4201), 419-426.
Mondal, S. K., Narayan, J., Sharma, C., Bharti, R., Dwivedy, S. K., Chowdhury, P. R., & Singh, R. P. (2025). Glacial lakes outburst susceptibility and risk in the Eastern Himalayas using analytical hierarchy process and backpropagation neural network models. *Geomatics, Natural Hazards and Risk*, *16*(1), 2449134.
Monitor, G. E. (2024). Global hydropower tracker. In: Global Energy Monitor. Available at: https://globalenergymonitor. org ….
Mutlag, W. K., Ali, S. K., Aydam, Z. M., & Taher, B. H. (2020). Feature extraction methods: a review. Journal of Physics: Conference Series,
Nath, M., & Choudhury, D. P. (2024). Cloudburst in Sikkim. In: Geological Society of India.
Nath, S. K., Thingbaijam, K. K. S., & Raj, A. (2008). Earthquake hazard in Northeast India—A seismic microzonation approach with typical case studies from Sikkim Himalaya and Guwahati city. *Journal of Earth System Science*, *117*, 809-831.

Nie, Y., Pritchard, H. D., Liu, Q., Hennig, T., Wang, W., Wang, X., Liu, S., Nepal, S., Samyn, D., & Hewitt, K. (2021). Glacial change and hydrological implications in the Himalaya and Karakoram. *Nature reviews earth & environment*, *2*(2), 91-106.

Nijhawan, R., Das, J., & Balasubramanian, R. (2018). A hybrid CNN+ random forest approach to delineate debris covered glaciers using deep features. *Journal of the Indian Society of Remote Sensing*, *46*, 981-989.

Nijhawan, R., Garg, P., & Thakur, P. (2016). A comparison of classification techniques for glacier change detection using multispectral images. *Perspectives in Science*, *8*, 377-380.

Noël, P., Rousseau, A. N., Paniconi, C., & Nadeau, D. F. (2014). Algorithm for delineating and extracting hillslopes and hillslope width functions from gridded elevation data. *Journal of Hydrologic Engineering*, *19*(2), 366-374.

Nunchhani, V., Mekro, V., Bandyopadhyay, A., & Bhadra, A. (2023). Spatio-temporal variability of glacier surface area, ELA, and AAR in the Mago river basin of Arunachal Pradesh. *Journal of Water and Climate Change*, *14*(1), 253-271.

Nussbaumer, S., Schaub, Y., Huggel, C., & Walz, A. (2014). Risk estimation for future glacier lake outburst floods based on local land-use changes. *Natural Hazards and Earth System Sciences*, *14*(6), 1611-1624.

O'Donnel, M. S., & Ignizio, D. A. (2012). *Bioclimatic predictors for supporting ecological applications in the conterminous United States* (2327-638X).

Ojha, S., Fujita, K., Asahi, K., Sakai, A., Lamsal, D., Nuimura, T., & Nagai, H. (2016). Glacier area shrinkage in eastern Nepal Himalaya since 1992 using high-resolution inventories from aerial photographs and ALOS satellite images. *Journal of Glaciology*, *62*(233), 512-524.

Panwar, R., & Singh, G. (2023). Classification of glacier with supervised approaches using PolSAR data. *Environmental Monitoring and Assessment*, *195*(1), 1-17.

Phillips, T., Leyk, S., Rajaram, H., Colgan, W., Abdalati, W., McGrath, D., & Steffen, K. (2011). Modeling moulin distribution on Sermeq Avannarleq glacier using ASTER and WorldView imagery and fuzzy set theory. *Remote Sensing of Environment*, *115*(9), 2292-2301.

Prasad, A. K., Yang, K.-H., El-Askary, H., & Kafatos, M. (2009). Melting of major Glaciers in the western Himalayas: evidence of climatic changes from long term MSU derived tropospheric temperature trend (1979–2008). Annales Geophysicae,

Priyanka, R. S., Jayangondaperumal, R., Pandey, A., Mishra, R. L., Singh, I., Bhushan, R., Srivastava, P., Ramachandran, S., Shah, C., & Kedia, S. (2017). Primary surface rupture of the 1950 Tibet-Assam great earthquake along the eastern Himalayan front, India. *Scientific reports*, *7*(1), 5433.

Qayyum, N., Ghuffar, S., Ahmad, H. M., Yousaf, A., & Shahid, I. (2020). Glacial lakes mapping using multi satellite PlanetScope imagery and deep learning. *ISPRS International Journal of Geo-Information*, *9*(10), 560.

Qiu, C., Su, L., Zou, Q., & Geng, X. (2022). A hybrid machine-learning model to map glacier-related debris flow susceptibility along Gyirong Zangbo watershed under the changing climate. *Science of the Total Environment*, *818*, 151752.

Quincey, D., Richardson, S., Luckman, A., Lucas, R. M., Reynolds, J., Hambrey, M., & Glasser, N. (2007). Early recognition of glacial lake hazards in the Himalaya using remote sensing datasets. *Global and Planetary Change*, *56*(1-2), 137-152.

Rafiq, M., Romshoo, S. A., Mishra, A. K., & Jalal, F. (2019). Modelling Chorabari lake outburst flood, kedarnath, India. *Journal of Mountain Science*, *16*(1), 64-76.

Raj, K. B. G., Remya, S., & Kumar, K. V. (2013). Remote sensing-based hazard assessment of glacial lakes in Sikkim Himalaya. *Current Science*, 359-364.

Rao, M. V. R., Puligundla, S., Ekkala, N. S., & Chebrolu, S. (2023). Image Classification of Ischemic Stroke Blood Clot Origin using Stacked EfficientNet-B0, VGG19 and ResNet-152. 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC),

Ren, W., Yang, T., Shi, P., Xu, C.-y., Zhang, K., Zhou, X., Shao, Q., & Ciais, P. (2018). A probabilistic method for streamflow projection and associated uncertainty analysis in a data sparse alpine region. *Global and Planetary Change*, *165*, 100-113.

Ren, W. W., Yang, T., Huang, C. S., Xu, C. Y., & Shao, Q. X. (2018). Improving monthly streamflow prediction in alpine regions: integrating HBV model with Bayesian neural network. *Stochastic Environmental Research and Risk Assessment*, *32*, 3381-3396.

Richardson, S. D., & Reynolds, J. M. (2000). An overview of glacial hazards in the Himalayas. *Quaternary International*, *65*, 31-47.

Rignot, E., Mouginot, J., Scheuchl, B., Van Den Broeke, M., Van Wessem, M. J., & Morlighem, M. (2019). Four decades of Antarctic Ice Sheet mass balance from 1979–2017. *Proceedings of the National Academy of Sciences*, *116*(4), 1095-1103.

Rinzin, S., Zhang, G., Sattar, A., Wangchuk, S., Allen, S. K., Dunning, S., & Peng, M. (2023). GLOF hazard, exposure, vulnerability, and risk assessment of potentially dangerous glacial lakes in the Bhutan Himalaya. *Journal of Hydrology*, *619*, 129311.

Roback, K., Clark, M. K., West, A. J., Zekkos, D., Li, G., Gallen, S. F., Chamlagain, D., & Godt, J. W. (2018). The size, distribution, and mobility of landslides caused by the 2015 Mw7. 8 Gorkha earthquake, Nepal. *Geomorphology*, *301*, 121-138.

Robson, B. A., Bolch, T., MacDonell, S., Hölbling, D., Rastner, P., & Schaffer, N. (2020). Automated detection of rock glaciers using deep learning and object-based image analysis. *Remote sensing of environment*, *250*, 112033.

Rounce, D. R., Khurana, T., Short, M. B., Hock, R., Shean, D. E., & Brinkerhoff, D. J. (2020). Quantifying parameter uncertainty in a large-scale glacier evolution model using Bayesian inference: application to High Mountain Asia. *Journal of Glaciology*, *66*(256), 175-187.

Rounce, D. R., McKinney, D. C., Lala, J. M., Byers, A. C., & Watson, C. S. (2016). A new remote hazard and risk assessment framework for glacial lakes in the Nepal Himalaya. *Hydrology and Earth System Sciences*, *20*(9), 3455-3475.

Rounce, D. R., Watson, C. S., & McKinney, D. C. (2017). Identification of hazard and risk for glacial lakes in the Nepal Himalaya using satellite imagery from 2000–2015. *Remote Sensing*, *9*(7), 654.

Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, *1*(1), 83-98.

Saha, K., Wells, N. A., & Munro-Stasiuk, M. (2011). An object-oriented approach to automated landform mapping: A case study of drumlins. *Computers & geosciences*, *37*(9), 1324-1336.

SANDRP. (2023, 12-08-2025). Glacial Lake FLood destroys Teesta-3 Dam in Sikkim, brings wide-spread destruction. https://sandrp.in/2023/10/04/glacial-lake-flood-destroys-teesta-3-dam-in-sikkim-brings-wide-spread-destruction/

Sarp, G., & Ozcelik, M. (2017). Water body extraction and change detection using time series: A case study of Lake Burdur, Turkey. *Journal of Taibah University for Science*, *11*(3), 381-391.

Sattar, A., Goswami, A., Kulkarni, A. V., Emmer, A., Haritashya, U. K., Allen, S., Frey, H., & Huggel, C. (2021). Future glacial lake outburst flood (GLOF) hazard of the South Lhonak Lake, Sikkim Himalaya. *Geomorphology*, *388*, 107783.

Sawi, T., Holtzman, B., Walter, F., & Paisley, J. (2022). An Unsupervised Machine-Learning Approach to Understanding Seismicity at an Alpine Glacier. *Journal of Geophysical Research: Earth Surface*, e2022JF006909.

Scuderi, L. A., & Nagle-McNaughton, T. (2022). Automated neural network identification of cirques. *Physical Geography*, *43*(1), 24-51.

Shah, A. A. (2013). Earthquake geology of Kashmir Basin and its implications for future large earthquakes. *International Journal of Earth Sciences*, *102*(7), 1957-1966.

Shah, A. A. (2013). Tectonic geomorphology of the eastern extent of the Kashmir Basin Fault (KBF) zone. 4th International INQUA Meeting on Paleoseismology, Active Tectonics and Archeoseismology (PATA),

Shah, A. A., & Batmanathan, N. M. (2025). Active faulting and lake formation in the Himalaya-Tibet orogenic belt. 18th World Lake Conference: Governance, Resilience and Sustainability of Lakes for a Better Society,

Shah, A. A., Rachman, M. G., Kumar, R., Vashistha, A., Dashora, A., & Mahoor, M. (2024a). Pre-disposed tectonic subsidence controls flood hazards and unplanned urbanisation dominates the flood disasters in the Pliocene to Holocene Kashmir basin, NW Himalayas. *Quaternary Science Advances*, *14*, 100173.

Shah, A. A., Rachman, M. G., & Mahoor, M. (2024b). Transition of Earthquake Hazards into Disasters. In *Natural Hazards and Risk Mitigation: Natural Hazards in Himalaya and Risk Mitigation* (pp. 173-190). Springer.

Sharma, R., Pradhan, P., Sharma, N., & Shrestha, D. (2018). Remote sensing and in situ-based assessment of rapidly growing South Lhonak glacial lake in eastern Himalaya, India. *Natural Hazards*, *93*, 393-409.

Shea, J., Immerzeel, W., Wagnon, P., Vincent, C., & Bajracharya, S. (2015). Modelling glacier change in the Everest region, Nepal Himalaya. *The Cryosphere*, *9*(3), 1105-1128.

Shen, L., & Li, C. (2010). Water body extraction from Landsat ETM+ imagery using adaboost algorithm. 2010 18th International Conference on Geoinformatics,

Shree, A., Sharma, R., Singla, J. G., Dhaka, V. S., Verma, S. S., & Malik, A. (2024). Comparative Evaluation of Deep Learning Models for Image Captioning of Satellite Images. International Conference on Innovations in Computational Intelligence and Computer Vision,

Shrestha, A., Eriksson, M., Mool, P., Ghimire, P., Mishra, B., & Khanal, N. (2010). Glacial lake outburst flood risk assessment of Sun Koshi basin, Nepal. *Geomatics, Natural Hazards and Risk*, *1*(2), 157-169.

Shugar, D. H., Burr, A., Haritashya, U. K., Kargel, J. S., Watson, C. S., Kennedy, M. C., Bevington, A. R., Betts, R. A., Harrison, S., & Strattman, K. (2020). Rapid worldwide growth of glacial lakes since 1990. *Nature Climate Change*, *10*(10), 939-945.

Shukla, A., Arora, M., & Gupta, R. (2010). Synergistic approach for mapping debris-covered glaciers using optical–thermal remote sensing data with inputs from geomorphometric parameters. *Remote Sensing of Environment*, *114*(7), 1378-1387.
Siddique, F., & Rahman, A.-u. (2023). Genesis and spatio-temporal analysis of glacial lakes in the peri-glacial environment of Western Himalayas. *International Journal of Engineering and Geosciences*, *8*(2), 154-164.
Sidle, R. C., Caiserman, A., Jarihani, B., Khojazoda, Z., Kiesel, J., Kulikov, M., & Qadamov, A. (2023). Sediment sources, erosion processes, and interactions with climate dynamics in the Vakhsh River Basin, Tajikistan. *Water*, *16*(1), 122.
Singh, I., Pandey, A., Mishra, R. L., Priyanka, R. S., Brice, A., Jayangondaperumal, R., & Srivastava, V. (2021). Evidence of the 1950 great Assam earthquake surface break along the Mishmi Thrust at Namche Barwa Himalayan Syntaxis. *Geophysical Research Letters*, *48*(11), e2020GL090893.
Singh, S., & Kansal, M. L. (2021). Cloudburst—a major disaster in the Indian Himalayan states. In *Civil engineering for disaster risk reduction* (pp. 115-126). Springer.
Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, *25*.
Somos-Valenzuela, M., McKinney, D., Byers, A., Rounce, D., Portocarrero, C., & Lamsal, D. (2015). Assessing downstream flood impacts due to a potential GLOF from Imja Tsho in Nepal. *Hydrology and Earth System Sciences*, *19*(3), 1401-1412.
Somos-Valenzuela, M. A., Chisolm, R. E., Rivas, D. S., Portocarrero, C., & McKinney, D. C. (2016). Modeling a glacial lake outburst flood process chain: the case of Lake Palcacocha and Huaraz, Peru. *Hydrology and Earth System Sciences*, *20*(6), 2519-2543.
Song, C., Sheng, Y., Ke, L., Nie, Y., & Wang, J. (2016). Glacial lake evolution in the southeastern Tibetan Plateau and the cause of rapid expansion of proglacial lakes linked to glacial-hydrogeomorphic processes. *Journal of Hydrology*, *540*, 504-514.
Sugden, D. E., & John, B. S. (1976). *Glaciers and landscape: a geomorphological approach* (Vol. 365). Edward Arnold London.
Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning,
Tang, Q., Zhang, G., Yao, T., Wieland, M., Liu, L., & Kaushik, S. (2024). Automatic extraction of glacial lakes from Landsat imagery using deep learning across the Third Pole region. *Remote Sensing of Environment*, *315*, 114413.
Taylor, C., Robinson, T. R., Dunning, S., Rachel Carr, J., & Westoby, M. (2023). Glacial lake outburst floods threaten millions globally. *Nature Communications*, *14*(1), 487.
Thati, J., & Ari, S. (2022). A systematic extraction of glacial lakes for satellite imagery using deep learning based technique. *Measurement*, *192*, 110858.
Thomas, R. B. (2023). *4-decadal Glacier Lake Inventory for Arunachal Pradesh, India,*. HydroShare. Retrieved 22/08/2025 from http://www.hydroshare.org/resource/d463068e0286409c8dd5c65b554b6ed5
Tian, S., Dong, Y., Feng, R., Liang, D., & Wang, L. (2022). Mapping mountain glaciers using an improved U-Net model with cSE. *International Journal of Digital Earth*, *15*(1), 463-477.
Tiwari, R. K., & Paudyal, H. (2022). Frequency magnitude distribution and spatial correlation dimension of earthquakes in north-east Himalaya and adjacent regions. *Geologos*, *28*(2), 115-128.
United Nations Office for Disaster Risk, R. (2017). The Sendai Framework Terminology on Disaster Risk Reduction: Disaster risk. https://www.undrr.org/terminology/disaster-risk
Vashistha, A., Dashora, A., & Shah, A. A. (2023). Sparse Station Placement for Optimal Reconstruction of Rain Gauge Data and Performance Assessment of Satellite Precipitation Estimates. World Congress on Disaster Management, Dehradun, India.
Vashistha, A., Dashora, A., Shah, A. A., & Pal, A. (2025a). Hazard assessment of glacial lakes in Himalaya through erosional features detection using deep learning. *Natural Hazards*, 1-24.
Vashistha, A., Shah, A. A., Batmanathan, N., & Dashora, A. (2025b). A data-driven multi-criteria model for GLOF risk in tectonically active Himalayan regions. *Modeling Earth Systems and Environment*, *11*(5), 333.
Vaughan, D. G., Comiso, J., Allison, I., Carrasco, J., Kaser, G., Kwok, R., Mote, P., Murray, T., Paul, F., & Ren, J. (2014). Observations: cryosphere.
Veh, G., Korup, O., & Walz, A. (2020). Hazard from Himalayan glacier lake outburst floods. *Proceedings of the National Academy of Sciences*, *117*(2), 907-912.
Velicogna, I., & Wahr, J. (2006). Measurements of time-variable gravity show mass loss in Antarctica. *Science*, *311*(5768), 1754-1756.

Vezzola, L. C., Diolaiuti, G. A., D'Agata, C., Smiraglia, C., & Pelfini, M. (2016). Assessing glacier features supporting supraglacial trees: A case study of the Miage debris-covered Glacier (Italian Alps). *The Holocene*, *26*(7), 1138-1148.
Vuichard, D., & Zimmermann, M. (1987). The 1985 catastrophic drainage of a moraine-dammed lake, Khumbu Himal, Nepal: cause and consequences. *Mountain Research and Development*, 91-110.
Wang, W., Xiang, Y., Gao, Y., Lu, A., & Yao, T. (2015). Rapid expansion of glacial lakes caused by climate and glacier retreat in the Central Himalayas. *Hydrological Processes*, *29*(6), 859-874.
Wangchuk, T., & Tsubaki, R. (2024). A glacial lake outburst flood risk assessment for the Phochhu river basin, Bhutan. *Natural Hazards and Earth System Sciences*, *24*(7), 2523-2540.
Washakh, R. M. A., Chen, N., Wang, T., Almas, S., Ahmad, S. R., & Rahman, M. (2019). GLOF risk assessment model in the Himalayas: A case study of a hydropower project in the Upper Arun River. *Water*, *11*(9), 1839.
Welling, M., & Teh, Y. W. (2011). Bayesian learning via stochastic gradient Langevin dynamics. Proceedings of the 28th international conference on machine learning (ICML-11),
Werder, M. A., Huss, M., Paul, F., Dehecq, A., & Farinotti, D. (2020). A Bayesian ice thickness estimation model for large-scale applications. *Journal of Glaciology*, *66*(255), 137-152.
Wessels, R. L., Kargel, J. S., & Kieffer, H. H. (2002). ASTER measurement of supraglacial lakes in the Mount Everest region of the Himalaya. *Annals of Glaciology*, *34*, 399-408.
Westoby, M. J., Glasser, N. F., Brasington, J., Hambrey, M. J., Quincey, D. J., & Reynolds, J. M. (2014). Modelling outburst floods from moraine-dammed glacial lakes. *Earth-Science Reviews*, *134*, 137-159.
Williams, C. K., & Rasmussen, C. E. (2006). *Gaussian processes for machine learning* (Vol. 2). MIT press Cambridge, MA.
Witteveen, A., Nane, G. F., Vliegen, I. M., Siesling, S., & IJzerman, M. J. (2018). Comparison of logistic regression and Bayesian networks for risk prediction of breast cancer recurrence. *Medical decision making*, *38*(7), 822-833.
Worni, R., Huggel, C., & Stoffel, M. (2013). Glacial lakes in the Indian Himalayas—From an area-wide glacial lake inventory to on-site and modeling based risk assessment of critical glacial lakes. *Science of The Total Environment*, *468*, S71-S84.
Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26-40.
Wu, Q. (2017). GIS and remote sensing applications in wetland mapping and monitoring.
Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., & Yu, P. S. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*, 1-37.
Xie, Z., Haritashya, U. K., Asari, V. K., Bishop, M. P., Kargel, J. S., & Aspiras, T. H. (2022). GlacierNet2: A hybrid Multi-Model learning architecture for alpine glacier mapping. *International Journal of Applied Earth Observation and Geoinformation*, *112*, 102921.
Xie, Z., Haritashya, U. K., Asari, V. K., Young, B. W., Bishop, M. P., & Kargel, J. S. (2020). GlacierNet: A deep-learning approach for debris-covered glacier mapping. *IEEE Access*, *8*, 83495-83510.
Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025-3033.
Xu, J., Feng, M., Sui, Y., Yan, D., Zhang, K., & Shi, K. (2023). Identifying Alpine Lakes in the Eastern Himalayas Using Deep Learning. *Water*, *15*(2), 229.
Yao, X., Liu, S., Han, L., Sun, M., & Zhao, L. (2018). Definition and classification system of glacial lake for inventory and hazards study. *Journal of Geographical Sciences*, *28*, 193-205.
Yongping, S. (2004). An overview of glaciers, retreating glaciers and their impact in the Tibetan Plateau. *Cold and Arid Regions Environmental and Engineering Research Institute (CAREERI) Chinese AcadY emy of Sciences (CAS). Lanzhou, China*.
Zaier, I., Shu, C., Ouarda, T. B., Seidou, O., & Chebana, F. (2010). Estimation of ice thickness on lakes using artificial neural network ensembles. *Journal of Hydrology*, *383*(3-4), 330-340.
Zhang, B., Liu, G., Zhang, R., Fu, Y., Liu, Q., Cai, J., Wang, X., & Li, Z. (2021). Monitoring dynamic evolution of the glacial lakes by using time series of Sentinel-1A SAR images. *Remote Sensing*, *13*(7), 1313.
Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, *14*(1), 35-62.
Zhang, G., Yao, T., Xie, H., Wang, W., & Yang, W. (2015). An inventory of glacial lakes in the Third Pole region and their changes in response to global warming. *Global and Planetary Change*, *131*, 148-157.
Zhang, J., Jia, L., Menenti, M., & Hu, G. (2019). Glacier facies mapping using a machine-learning algorithm: The Parlung Zangbo Basin case study. *Remote Sensing*, *11*(4), 452.

Zhang, T., Wang, W., & An, B. (2025). A massive lateral moraine collapse triggered the 2023 South Lhonak Lake outburst flood, Sikkim Himalayas. *Landslides*, *22*(2), 299-311. https://doi.org/10.1007/s10346-024-02358-x

Zhang, T., Wang, W., An, B., Gao, T., & Yao, T. (2022a). Ice thickness and morphological analysis reveal the future glacial lake distribution and formation probability in the Tibetan Plateau and its surroundings. *Global and Planetary Change*, *216*, 103923.

Zhang, T., Wang, W., Gao, T., An, B., & Yao, T. (2022b). An integrative method for identifying potentially dangerous glacial lakes in the Himalayas. *Science of the Total Environment*, *806*, 150442.

Zhang, Z., Liu, Y., Li, Y., Wang, X., Li, H., Yang, H., Ding, W., Liao, Y., Tang, N., & He, F. (2022). Lake ecosystem health assessment using a novel hybrid decision-making framework in the Nam Co, Qinghai-Tibet Plateau. *Science of The Total Environment*, *808*, 152087.

Zhong, H., & Liang, D. (2024). A study of road closure due to rainfall and flood zone based on logistic regression. *International Journal of Disaster Risk Reduction*, *102*, 104291.

Zorlu, K., & Dede, V. (2023). Assessment of glacial geoheritage by multi-criteria decision making (MCDM) methods in the Yalnızçam Mountains, Northeastern Türkiye. *International Journal of Geoheritage and Parks*, *11*(1), 100-117.

Zulkarnain, S. H., & Razali, M. N. (2021). The Delphi method to identify attributes for a valuation approach for residential property exposed to flood risk. *Property Management*, *40*(1), 62-82.