Flash flood susceptibility prediction mapping using Remote Sensing and Machine Learning

By

MUHAMMAD TAHA BIN SARDAR

Reg No : 00000361093

Supervisor

Dr. Muhammad Tariq Saeed

(Associate Professor)

Department of Sciences

MS Systems Engineering

School of Interdisciplinary Engineering and Sciences (SINES)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

September 2023

Annex A to NUST Letter No.
0972/102/Exams/Thesis-Cert
dated 23 Dec 16.

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Mr. **M. Taha Bin Sardar** Registration No. **00000361093** of **SINES** has been vetted by undersigned, found complete in all aspects as per NUST Statutes/Regulations, is free of plagiarism, errors, and mistakes and is accepted as partial fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said thesis.

Signature with stamp: (DR. MUHAMMAD TARIQ SAEED) Assistant Professor Research Centre for Modeling & Simulation NUST, Sector H-12 Islamabad

Name of Supervisor:

Date: 20/09/2023

Signature of HoD with stamp: Dr. Mian Ilyas Ahmad HoD Engineering Professor SINES - NUST, Sector H-12 Islamabad

Date: 20/Sep/2023

Countersign by

Signature (Dean/Principal): Dr. Hammad M. Cheema Principal & Dean SINES - NUST, Sector H-12 Islamabad

Date: 21 SEP 2023

"I would like to dedicate this work to my family whose constant support has been a part of my journey. I am grateful, for the mentors who have come into my life in roles. As family members, friends, teachers and even strangers who have left an impact. Their valuable guidance and acts of kindness have significantly shaped the person I'm today. I cannot express gratitude for their influence and the positive effect they have had on my path." ."

Declaration

I, **MUHAMMAD TAHA BIN SARDAR** declare that this thesis titled **"Flash flood susceptibility prediction mapping using Remote Sensing and Machine Learning"** and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work
5. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself

NAME,

Reg No: 00000361093 MSSYSE Fall(2021)

Copyright Notice

Acknowledgments

In the name of Allah, the compassionate and merciful I would like to begin by expressing my gratitude to the key individuals who have played crucial roles in helping me successfully complete my Masters thesis.

First and foremost I would like to extend my appreciation to my Supervisor, Dr. Muhammad Tariq Saeed. His invaluable guidance and support were instrumental, in finalizing my thesis. I am equally thankful to my Co Supervisor, Dr. Zamir Hussain for providing assistance during the phases of data collection and feature selection.

I would also like to convey thanks to the members of my Guidance and Examinations Committee Dr. Ammar Mushtaq and Dr. Mian Ilyas Ahmad. Their expertise and dedication in sharing their time for my pursuits are greatly appreciated.

Furthermore I am deeply grateful for our principal, Dr. Hammad Cheemas encouragement and effective leadership, throughout my journey which significantly contributed to its culmination.

Lastly I want to acknowledge the support and resources provided by my institution that facilitated my research endeavors whenever needed. Moreover I would like to express my appreciation to the administrators and experienced teachers, in our school district. Their enthusiasm and willingness to share their insights have truly made the process of conducting this study enjoyable and enriching. In conclusion I am deeply grateful to my seniors, M. Ibraheem Akram, M. Junaid Saleem and Fawaz Nadeem, for their support and invaluable guidance that enabled me to complete my work within the given timeframe.

Contents

List of Tables

List of Figures

Abstract

Pakistan has witnessed recurring, devastating floods attributed to extreme rainfall, causing loss of life and significant economic consequences. Studies have been conducted with regards to flood prediction mapping in Pakistan using various remote sensing and GIS techniques, the gap which has been identified is that the findings for the previous studies conducted do not include a simulation aspect, only include results obtained using past data. In our study we used a combination of GIS tools, remote sensing and machine learning techniques to generate susceptibility maps for our region of interest. We have focused on the area of Swat District in Khyber Pakhtunkhwa, Pakistan, a flash flood-prone region, as our study area. Datasets were collected through reliable sources such as Climate Research Unit, Open Topography, Global Flood Database, Geological Survey database, Ensuring the quality of the data, preprocessing was applied to cater for outliers, null data and redundant values. The central research question pertains to flood susceptibility prediction within Swat District. The Frequency Ratio method was employed for feature extraction, demonstrating the influence of factors such as slope, flow accumulation, LULC, distance to rivers and precipitation patterns. After analysis, a wide range of factors were examined to understand the vulnerability of the area to sudden floods. This resulted in the development of a set of characteristics that portrays the regions susceptibility to flash floods. Machine learning models, such, as Random Forest (RF) k Nearest Neighbors (KNN) Support Vector Machine (SVM) and XGBoost (XGB) were then applied on these features. The results of those models based on hyperparameter tuning indicate high performance of those models with recall value of 0.90, f1-score of 0.95 and AUC ROC values of 0.99 for RF and XBG and 0.96, for KNN and SVM showcasing their capabilities. In conclusion, the results obtained used

a combination of weighted features, from the Frequency Ratio method and machine learning models to create a map showing the susceptibility of Swat District to floods. The predictions generated by simulating rainfall patterns across our study area we can predict which regions are prone, to flooding and estimate the damage caused by such events. This work will offer valuable guidance and aim to enhance flood risk management strategies, ultimately contributing to the preservation of lives and the reduction of flood-related damages.

Chapter 1

Introduction

One of nature's most destructive disasters, flooding may seriously harm people, agriculture, and infrastructure. As a result of climate change and urbanization, floods are becoming more frequent and intense, necessitating the creation of efficient flood management measures[1]. Flood susceptibility mapping is a key technique in flood management because it enables decision-makers to pinpoint regions that are most susceptible to flooding and implement the necessary precautions to reduce the risks. Remote sensing and machine learning approaches have become effective tools for forecasting flood susceptibility in recent years[2]. This study focuses on a method for mapping flood susceptibility that combines machine learning and remote sensing techniques to identify areas that are most vulnerable to flooding. This study focuses on the application of these methods in a specific geographic region and discuss their potential for improving flood management strategies.

1.1 What are Floods?

Floods are disasters that occur when water exceeds its boundaries and spreads onto usually dry land. They are, among the financially devastating natural calamities worldwide affecting millions of people every year and causing significant economic damage. Floods can be caused by factors such, as rainfall melting snow storm surges, dam or levee failures and tsunamis[3].

Floods can bring about consequences, for the environment, economy and society at large. They possess the ability to result in soil erosion harm crops and disrupt the balance of ecosystems[4]. Floods have the potential to cause harm to homes and infrastructure resulting in both setbacks and the displacement of people. Unfortunately in situations floods can also tragically lead to loss of life[5].

1.2 Types of floods

There are several types of floods, each with unique characteristics and causes.

1. Flash Floods
2. River Floods
3. Coastal
4. Urban
5. Other

Flash Floods are intense floods that happen in areas, with elevation usually due to heavy rain or a sudden discharge of water, from a dam or levee. These flash floods can be highly hazardous since they can happen unexpectedly and the flowing water has the potential to carry away individuals, animals and vehicles[6].

River Floods happen when the water levels, in a river or stream exceed their boundaries because of intense rainfall or melting snow. The speed at which river floods occur can vary depending on the size of the river and the volume of water coursing through it. These floods have the potential to cause harm to buildings, roads and other infrastructure. Can also disrupt transportation and communication networks[3].

Coastal Floods occur due, to storm surges or tsunamis leading to the flooding of regions and causing harm to buildings and residences. These types of floods can be especially perilous because they can happen suddenly and unexpectedly leaving opportunity for individuals to evacuate[7].

Urban Floods are a frequent occurrence in urban environments, including cities and towns, and is frequently caused by elements including heavy rainfall, poorly planned drainage systems, and growing urban expansion. These floods have the ability to seriously damage structures and infrastructure, interfere with transportation and communication systems, and endanger public health by spreading waterborne diseases[8].

Other types of floods, such as glacier lake outburst floods and dam failures. Each type of flood has unique characteristics and causes and requires specific flood management strategies to mitigate their impacts.

1.3 Impact of Floods

Floods rank as one of the most destructive natural calamities, resulting in extensive harm to infrastructure, agricultural yields, and human lives. Their repercussions extend widely, impacting economies, societies, and the natural world. Floods can result in harm to residences and enterprises, interrupt transportation and communication networks, pollute water sources, and trigger electricity failures [9]. Furthermore, floods can exert enduring effects, including heightened soil erosion, harm to ecosystems, and the displacement of individuals[10].

One of the most significant impacts of floods is on infrastructure. Floods have the potential to cause harm by impacting roads, bridges and other essential transportation structures. This can create challenges, for emergency responders when trying to access areas. Additionally floods can cause damage to buildings resulting in setbacks, for both individuals and businesses[11]. In some cases, floods can cause buildings to collapse, leading to loss of life[9].

Floods can cause damage, to crops and livestock leading to food shortages and higher prices. They also have the ability to erode soil and vital nutrients making it difficult for farmers in affected areas to grow crops. Additionally floods can negatively impact irrigation systems resulting in reduced crop yields and financial difficulties, for farmers[10].

Apart, from the impact of floods there are also effects that can cause significant health risks. Floods can contaminate water sources, which in turn increases the chances of spreading diseases. Additionally they can result in the displacement of communities thereby raising the risk of illnesses, in temporary shelters[12].

1.4 Detection and Prevention of Floods

Floods pose a threat, to both communities and ecosystems making it crucial to establish measures for early detection and prevention. Various strategies can be employed to identify and mitigate the risks associated with floods, such as remote sensing[13][14][15], machine learning[16][17][18], floodplain management[19][20][21], and early warning systems [22][23][24].

Remote sensing is a method used to track variations, in water levels acting as a warning system for potential floods. It involves gathering information about the Earths surface from a distance. Remote sensing data can come from sources like satellites, aircraft and ground based sensors. Moreover the application of machine learning algorithms, to this data enables the identification of patterns and trends that can be utilized for predicting and preventing flood occurrences[25].

Floodplain management is another strategy that can be used to prevent floods. This involves identifying areas that are at risk of flooding and implementing measures to reduce the impact of floods. This can include zoning regulations that limit construction in flood-prone areas, building levees or dams to redirect floodwaters, and restoring wetlands to absorb excess water[21].

Early warning systems play a role, in preventing floods. These systems gather information from sources, like sensors, weather forecasts and hydrological models to evaluate the chances and intensity of flooding. When a flood is expected these systems can send out alerts to communities and emergency responders giving them time to evacuate and prepare for the flood[26].

Apart, from the strategies there are additional approaches to identify and mitigate floods[27]. One of these involves employing drones to collect data on areas to flooding. Another method is the development of technologies for flood monitoring and prediction. Additionally raising awareness and providing education about flood risks is crucial. In summary there are techniques, for detecting and preventing floods. By implementing a combination of these strategies, communities and governments can collaborate to minimize the dangers and consequences associated with floods safeguarding people, infrastructure and ecosystems.

In general there exist approaches to identify and mitigate floods. When these tactics are combined communities and governments can collaborate effectively in order to minimize the dangers and consequences of flooding, safeguarding individuals, infrastructure and the environment.

1.5 Floods in Pakistan

Pakistan is, among the nations that face a vulnerability to flooding. This susceptibility is primarily due to its location, climate conditions and infrastructure[28]. In years the country has witnessed catastrophic floods resulting in substantial loss of life displacement of communities and extensive damage, to residential areas, infrastructure and agricultural yields[29].

Floods, in Pakistan are frequently prompted by monsoon rainfall leading to the overflowing of rivers and the submerging of regions. Inadequate drainage systems, deforestation and various other factors can further worsen the impact of these floods. Over the ten years Pakistan has faced devastating floods notably in 2010, 2011 and 2014[30].

The floods, in Pakistan have had an effect on the people and economy of the country. Based on information from the United Nations, the floods that occurred in 2010 impacted around 20 million individuals. Resulted in damages amounting to over 10 billion dollars. Moreover these floods caused losses in crops, which led to shortages of food and an increase in prices. Additionally the floods have had a lasting impact, on the infrastructure and progress of the country as many roads, bridges and other structures were harmed or completely destroyed[29].

1.6 Problem Statement

Pakistan has been facing a series of flood incidents in the years due, to heavy rainfall resulting in severe human casualties and significant financial losses. The purpose of this study is to examine how advanced technologies and predictive methods can be utilized to forecast and minimize the consequences of floods caused by rainfall, in Pakistan. The ultimate objective is to save lives and reduce damages.

1.7 Objectives

1. To develop dataset of flood influencing factors (elevation, slope, slope aspect, distance from river, land use, rainfall) for Swat.

2. To use modelling approaches for simulation of flood flow and determine water accumulation levels.

1.8 Thesis Layout

The remaining thesis is structured as follows. Chapter 2 will cover the extensive research on climate change and flood susceptibility that has been done, and it will connect our issue to the literature already in circulation. Chapter 3 focuses on the methodology approach taken in this study to accomplish our goals. The outcomes and results that emerged from using the suggested methodology are discussed in Chapter 4. Lastly, in chapter 5 we will conclude our proposed work and address potential future approaches.

Chapter 2

Literature Review

Flash floods are natural disasters that present significant dangers, to human lives and infrastructure. It is vital to identify and map areas to flash floods for effective risk assessment, land use planning and the development of strategies and systems, for early warning. In times researchers have utilized a range of modeling techniques and data sources to enhance our understanding and prediction of flash floods. This review aims to provide an overview of the methodologies and approaches employed in modeling flash flood susceptibility and prediction including Early Warning Systems, Floodplain Management, Machine Learning Models, Hybrid Models and Deep Learning.

2.1 Early Warning System

J Cools et al,2018.[31] discusses the creation of an early warning system (EWS) for flash floods. Using the finest information available, including field measurements, simulations, and expert opinions, the EWS was developed and evaluated despite the lack of available data and scientific uncertainties. The distribution of rainfall, prior event inventory, transmission and infiltration losses, and warning thresholds were identified as important characteristics. Nine flash floods were caused by 20 heavy rainfall events over a 30-year span. Notably, infiltration and transmission losses resulted in a 90 percent reduction in rainfall volume during the 2010 flash flood. For a successful EWS, the study emphasises the necessity of institutional competence and strong communication.

Acosta et al,2018.[32] discusses the rise of pluvial flash floods in urban areas and the requirement for efficient Early Warning Systems (EWS). The article evaluates current EWS architectures, specifies critical elements of an EWS, and identifies the primary factors affecting flash flood intensity. Findings show that present implementations miss key details, leading to the suggestion of a fundamental framework for an effective EWS targeted at rain-induced flash floods.

Zang el al,2022.[33] highlights the value of flood early warning systems (FEWS) in minimising flood losses, particularly in metropolitan settings. To account for geographic variations in flood losses, the proposed multi-information FEWS integrates rainfall, inundation, and disaster information. In high-risk residential and commercial regions, the study emphasises the necessity for preventive measures by highlighting disparities in building property types and their associated losses. The research's conclusions offer useful information for global FEWS development and decision-making.

2.2 Floodplain Management

Olsen et al,2006.[34] investigates how federal agencies employ estimates of the frequency of flooding and questions the typical flood risk analysis's reliance on a fixed climate. The promise of hydro-meteorological models is highlighted, along with analyses of other statistical models for estimating flood risk. The study highlights how important it is for floodplain managers to take into account the uncertainties brought on by climate change and unpredictability in flood risk estimation and incorporate them into decision-making and regulation.

Padi et al,2011.[35] focuses on finding sustainable solutions to Africa's rising risk of floods. Using quality-controlled databases, it gives a thorough statistical analysis of flood data, including maximum discharge values and annual maximum flow time series. Through a comprehensive regional analysis, the study generates probabilistic envelope curves, offering useful insights into the statistical properties of floods in Africa. The findings are quite useful and can be applied to help the continent's flood management efforts.

Ndabula et al,2012.[36] monitored and mapped the trends of floodplain encroachment along the River Kaduna in the Nigerian metropolis of Kaduna. The study found significant levels of encroachment close to the Central Business District and industrial sectors, indicating increasing flood risk, using multiple datasets and ArcGIS software. The urban portion of the floodplain of the River Kaduna has been invaded to a degree of 52.83 percent. The report emphasises the requirement for robust institutional frameworks and financial support for floodplain management.

Kiedrzynska et al,2015.[37] investigates the function of eco-hydrology in managing flood risk and river floodplain water quality. In light of climate change and extreme weather events, it emphasises the significance of flood management and prevention. The ability of floodplains to absorb flood and pollution peaks, hence lowering the risk of flooding, is emphasised in the research. It suggests three methods for boosting water storage capacity: ecohydrological biotechnologies, efficient infrastructure utilisation, and sustainable ecohydrological management.

2.3 Machine Learning Models

Janizadeh et al,2019.[38] utilized five machine learning techniques—alternating decision tree (ADT), functional tree (FT), kernel logistic regression (KLR), multilayer perceptron (MLP), and quadratic discriminant analysis (QDA)—this study sought to determine the susceptibility of the Tafresh watershed in Iran to flash floods. Eight flood affecting elements and 320 historical flood episodes from a geospatial database were used. The FT, KLR, MLP, and QDA techniques came in second place to the ADT method in terms of performance compared to the other approaches. All five machine learning models were shown to be suitable for mapping flood susceptibility in various places to lessen the impact of disastrous floods, despite some differences in performance measures.

Bui et al,2019.[39] used feature selection and ensemble approaches to create a novel method for modelling flash flood susceptibility. Traditional approaches were exceeded by the FURIA-GA approach, and the FURIA-GA-Bagging model had the highest sensitivity (96.94) and accuracy (93.37). The model with the greatest AUC (0.9740) was FURIA-GA-AdaBoost. These models provide useful resources for determining a region's sensitivity to flash floods.

Costache et al,2019.[40] examined how well the Analytical Hierarchy Process (AHP), kNN, K-Star (KS), and their ensembles performed in mapping flash flood susceptibility. 70 percent of the previously damaged sites were used to train the models after the study used remote sensing techniques to identify them. The slope angle was found to be the most reliable of the ten flash flood predictors that were taken into consideration. The normalised weights of the predictors were calculated using the AHP model, and the Flash-Flood Potential Index (FFPI) was calculated using the kNN-AHP and KS-AHP ensemble models. Statistical measures were used to assess the models' performance, and the Receiver Operating Characteristics (ROC) Curve and Area Under the Curve (AUC) values served as validation. The kNN–AHP ensemble model demonstrated the best performance overall.

Hosseini et al,2020.[41] The study focused on developing models to map flash flood hazards, in Iran, a country that frequently experiences floods. To improve the performance of flash flood prediction, state of the art ensemble models like the boosted model (GLMBoost) random forest (RF) and Bayesian generalised linear model (BayesGLM) were recommended. Simulated annealing (SA) a processing technique was used to eliminate unnecessary variables. The accuracy of the models ranged from 90 to 92 percent with Kappa values ranging from 79 to 84 percent. Success ratios ranged from 94 to 96 percent threat scores from 80 to 84 percent and Heidke skill scores from 79 to 84 percent. Both models demonstrated performance. Key factors influencing flash flood modeling included proximity to streams, vegetation cover, drainage density, land use patterns and elevation. These findings are crucial, for identifying high risk areas. Can aid watershed managers in preventing and mitigating flood related damages in data regions.

Band et al,2020.[42]The main goal of the study was to assess the susceptibility of the Kalvan watershed in Irans Markazi Province to floods. By analyzing 15 weather and environmental factors researchers tested five machine learning techniques to identify areas to flooding. The results indicated that the randomised trees (ERT) model performed the best with an area, under the curve (AUC) value of 0.82. The AUC values for models were slightly lower ranging from 0.75 to 0.80 for regularised forest (RRF) parallel random forest (PRF) random forest (RF) and boosted regression tree (BRT). According to the ERT model about 28.33 percent (582.56 km2) of the study area was at risk of flash flooding while most areas exhibited very low vulnerability levels. The study concluded that factors such as altitude, slope, rainfall patterns and proximity, to rivers played roles in determining vulnerability levels in this region.

Costache et al,2020.[43] focused on the development of two new modelling techniques, ADT-IOE and ADT-AHP, for mapping flash flood susceptibility in the Suha river watershed, Romania. These strategies were put up against two independent techniques, IOE and AHP, in the study. The models were trained and assessed using ROC Curve, classification accuracy, and Kappa index by examining 111 torrential points, 111 non-torrential points, and 8 flash-flood conditioning factors. The outcomes demonstrated that the ensemble models, ADT-IOE and ADT-AHP, outperformed the other models and displayed strong prediction performance (AUC = 0.972, CLA = 86.37 percent, Kappa = 0.727 and AUC = 0.926, CLA = 87.88 percent, Kappa = 0.758, respectively). As a result, ADT-IOE and ADT-AHP are viewed as promising techniques for modelling flash flood susceptibility.

El-Magd et al,2021.[44] used machine learning methods, particularly XGBoost and KNN. Key determining elements included elevation, slope, separation from streams, and stream density. The models' performance was enhanced by hyper-parameter optimisation, with XGBoost outperforming KNN 80.7 percent in terms of accuracy with a score of 90.2 percent. Decision-makers can use the generated flash flood prediction map to plan and build new construction projects.

2.4 Hybrid Machine Learning Models

Ha el al,2021.[45] created cutting-edge hybrid machine learning methods for mapping and modelling flash flood vulnerability along National Highway 6 in Vietnam's Hoa

Binh region.which consists of DCREPT, AdaBoostM1-REPT, Bagging-REPT and MultiBoostAB-REPT were applied. The best-performing model, DCREPT, found considerable areas with high- and very high-flash flood susceptibility (12,572-17,660 hectares) and attained high prediction accuracy (AUC=0.991). Other mountainous transit routes can use the proposed technology for mapping flash flood predictions.

Elmahdy et al,2020.[46] tested three machine learning models: boosted regression tree (BRT), classification and regression trees (CART), and naive Bayes tree (NBT), with the goal of improving the mapping of flash flood (FF) susceptibility in an arid region. The outcomes demonstrated that BRT worked better than the other models, enabling precise FF susceptibility mapping. The study also developed new metrics to assess FF magnitude in various basins, demonstrating that hilly and smaller basins had the highest likelihood of FF occurrence and amplitude. This method illustrates how machine learning and geohydrological models can be used to enhance FF mapping and quantify its size.

Chen et al,2019.[47] In order to forecast flood susceptibility, this study used machine learning-based ensemble frameworks, notably Bag-REPTree and RS-REPTree, within a geographic information system (GIS). Thirteen influencing elements and 363 flood locations were used to generate a flood spatial database. The Wilcoxon signed-rank test, standard error, confidence interval, and receiver operating characteristic (ROC) curve were all used to gauge how well the models performed. As a result of having the greatest area under the ROC curve (AUC) value of 0.949 for training datasets and 0.907 for validation datasets, the RS-REPTree model beat the other models, the results showed. In terms of performance, the Bag-REPTree and REPTree models came next. These results demonstrate the superiority of the ensemble method for assessing flood susceptibility over individual methods.

Bui el al,2019.[48] proposed and validated a new soft computing approach. In order to forecast flash floods in a region in Northwest Vietnam that experiences frequent tropical typhoons, this study developed and verified the PSO-ELM soft computing approach. With strong prediction performance (kappa statistics = 0.801, RMSE = 0.281, MAE = 0.079, R2 = 0.829, AUC-ROC = 0.954), the PSO-ELM model beat conventional machine learning techniques. The PSO-ELM model is a viable tool for flash flood prediction in such settings.

Liu et al,2021.[49] discuss the difficulties in effectively modelling and forecasting flash floods. Three hybrid models, which combined machine learning techniques were. Evaluated for the Dadu River Basin. The models utilized support vector machines (SVM) classification and regression trees (CART) and convolutional neural networks (CNN) with membership values. Among these models the CNN FMV hybrid model demonstrated the performance in terms of fitting accuracy and prediction capabilities as, per the findings. Moreover all three hybrid models outperformed their respective single machine learning counterparts in predicting flood susceptibilities in the study region. The study identified areas covering 13.21 percent to 22.03 percent of the region with high to flood susceptibilities. These results suggest that the proposed hybrid models, CNN FMV hold potential for future applications.

Ngo el al,2018.[50] develop a novel method for identifying flash flood-prone locations utilizing Sentinel-1 SAR data and a hybrid machine learning approach known as FA-LM-ANN. The study created a GIS database with 12 input variables using the Bac Ha Bao Yen (BHBY) region of Vietnam as a case study. SAR imagery was used to map the flood inundation zones, and the FA-LM-ANN model performed well in forecasting the likelihood of flash floods. The proposed FA-LM-ANN is a useful tool for flash flood prediction because the firefly algorithm (FA) and LM backpropagation performed together very effectively.

2.5 Research Gap

According to the literature review, while research on flash floods has been done using remote sensing tools and techniques, this field is still largely unexplored in developing nations like Pakistan. There are not enough studies specifically using remote sensing and machine learning to produce maps of flash flood susceptibility[51][52][53]. The few studies that have been done in this area mainly focused on river-based analysis and less focus on machine learning without taking into account the entire Swat district[54][55][56].

Therefore, this study aims to fill this research gap by developing a system tailored for the Swat district area that utilizes remote sensing and machine learning techniques to generate a comprehensive susceptibility map for assessing the impact of floods on the region.

Chapter 3

Proposed Methodology

3.1 Study Area

The Swat study area covers approximately 5,337 square kilometers (2,059 square miles) and is situated in the northwestern part of Pakistan from 35.87° N to 34.94° S latitude and 72.165° W to 72.873° E longitude. It encompasses the Swat River and its tributaries which has a drainage area of 3,584 km^2, along with the high altitude, rugged terrain, and a snowy basin and settlements[57]. The area is known for its landscape, which includes mountains, narrow valleys and flat agricultural plains. The elevation ranges from 1,000 meters (3,281 feet), in the plains to over 5,000 meters (16,404 feet) at the highest peaks. The Swat basin experiences seasons throughout the year with winters and pleasant summers. There are two weather patterns that bring rainfall to this region; the monsoon in summer from the south and the Mediterranean system from the west during winter. The melting of glaciers and monsoonal rains significantly affect water flow, in the Swat basin.

The Swat region is prone to frequent and devastating floods, primarily caused by intense rainfall, snow melt, and rapid land use changes[58].The Swat River Basin has been severely affected by floods leading to loss of life, infrastructure damage, disruption of livelihoods and displacement of communities. Over the years it has faced floods, including incidents, in 1973, 1992, 1993 1994, 1995, 1996, 2001, 2005, 2010 and 2016.[59]. The impacts are particularly severe in low-lying areas and densely populated settlements along the riverbanks. The Flood of 2010 was the most catastrophic in the history of Swat (Fig 3.1), with 86 fatalities, killing 9800 livestock, destroyed 4000 houses, washed several bridges, and damaged the Amandara and Munda Headworks[60][61].

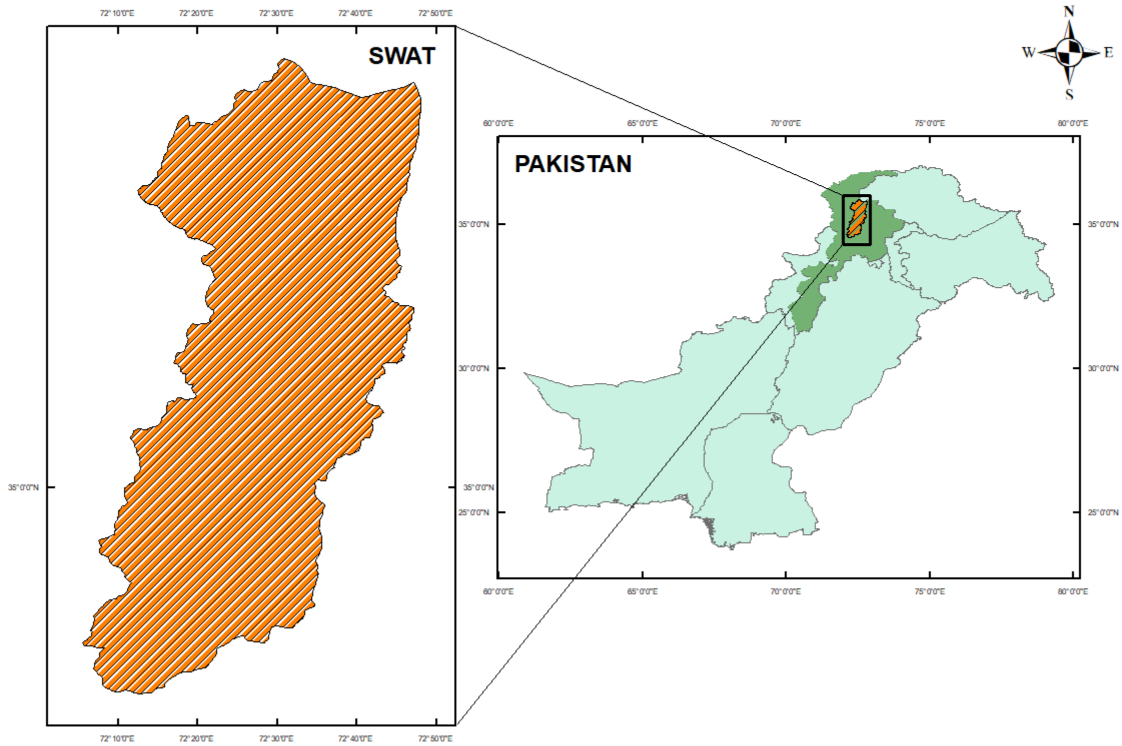


Figure 3.1: Map of Pakistan along with Study Area

3.2 Methodology Workflow

This section discusses the overview of whole methodology. Starting from the data acquisition followed by pre-processing techniques and the subsequent steps that are involved in prediction mapping to the last step which is the generation of flash flood susceptibility map.

The figure below (Fig 3.2) is designed to show the connectivity and link between all the steps to acquire the susceptibility map.

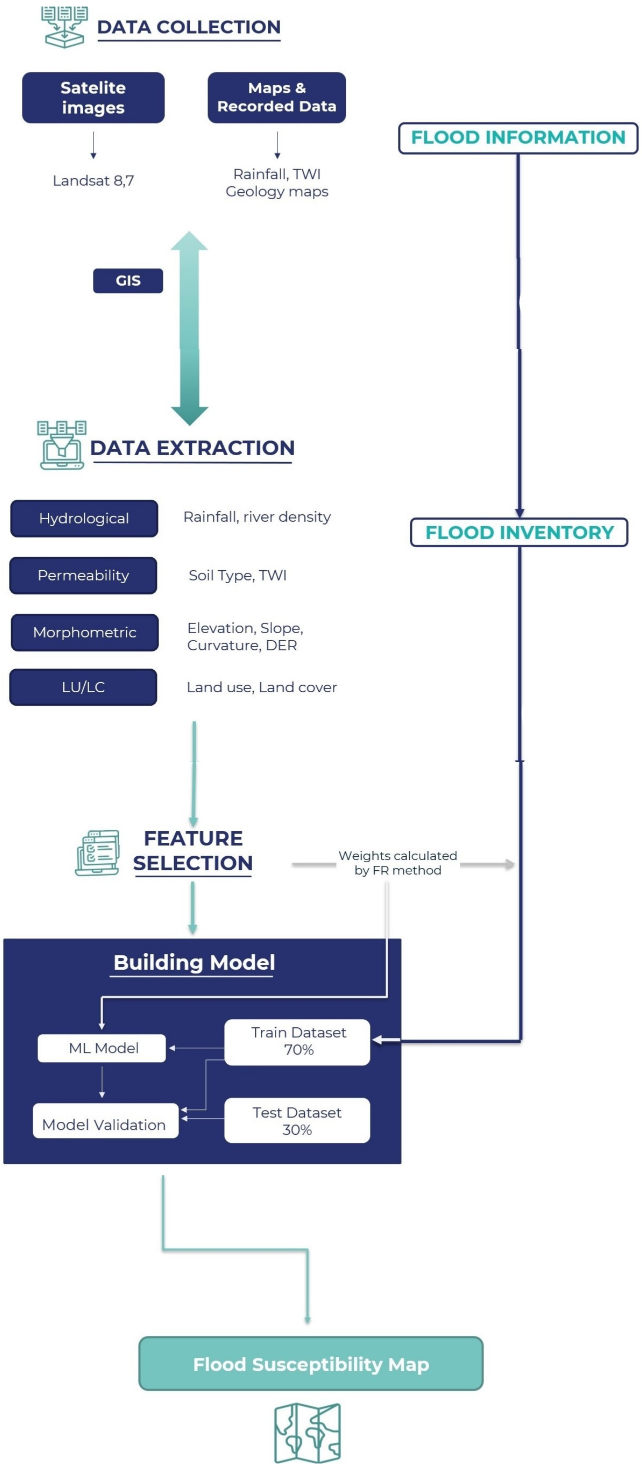


Figure 3.2: Methodology Workflow

3.3 Dataset

3.3.1 The Global Flood Database

The Global Flood Database (GFD) was used to obtain flood data . the area affected by natural catastrophes was pixelated in the data's map representation. The GFD is a thorough and comprehensive database of data about floods from all over the world. The GFD includes a broad range of flood-related factors, such as flash flood occurrences' size, length, frequency, and geographic scope. It contains information from a variety of sources, including regional and national organizations, satellite observations, on-the-ground measurements, and academic works. To ensure a complete and current collection of data on flash floods, the database uses cutting-edge data collection technologies, such as remote sensing, crowd-sourcing, and data mining.

3.3.2 Climate Research Unit

To conduct the study we gathered rainfall data for an area, from the Climate Research Unit (CRU) over a period of 10 years. The data from CRU provides insights into historical precipitation patterns and current climate conditions. Known for their expertise in climate research the CRU maintains a database that includes variables like rainfall, temperature and other meteorological characteristics. Professionals from fields such, as researchers, climatologists, policymakers and those involved in studies related to climate change, water resource management, agriculture and environmental planning frequently rely on CRUs rainfall data.

The rainfall information utilized by CRU is derived from a blend of satellite data, on site observations and simulations from climate models. The collection of rainfall data encompasses both current records. Thanks to its dispersed data users can examine rainfall patterns across scales ranging from the global level down, to regional and local levels.

3.3.3 Exploratory Data Analysis

Exploratory Data Analysis (EDA) refers to the process of examining and summarizing datasets to gain an understanding of the underlying patterns and relationships, within the data. This entails utilizing visual techniques to identify any outliers missing values or anomalies well as exploring the distribution, correlation and trends exhibited by the data.

The flash flood information collected from the Global Flood Database consisted of data encompassing the country over a period of 15 years. This dataset provided an overview of all flash flood incidents that occurred within our designated Area of Interest (AOI) during that timeframe. To efficiently analyze and visualize this flash flood data we utilized a Geographic Information System (GIS) platform, like *QGIS* to combine and filter it.

The technique involved carefully picking and isolating the flash flood occurrences inside the AOI from the wider dataset. To extract pertinent flash flood episodes that occurred inside the defined area, this filtering involved applying particular spatial and temporal characteristics. The flash floods that directly impacted the AOI were the only thing that was of interest when the dataset was refined. the focus was narrowed down to the flash floods that directly impacted the AOI, ensuring the analysis would be specific and relevant to the research objectives.

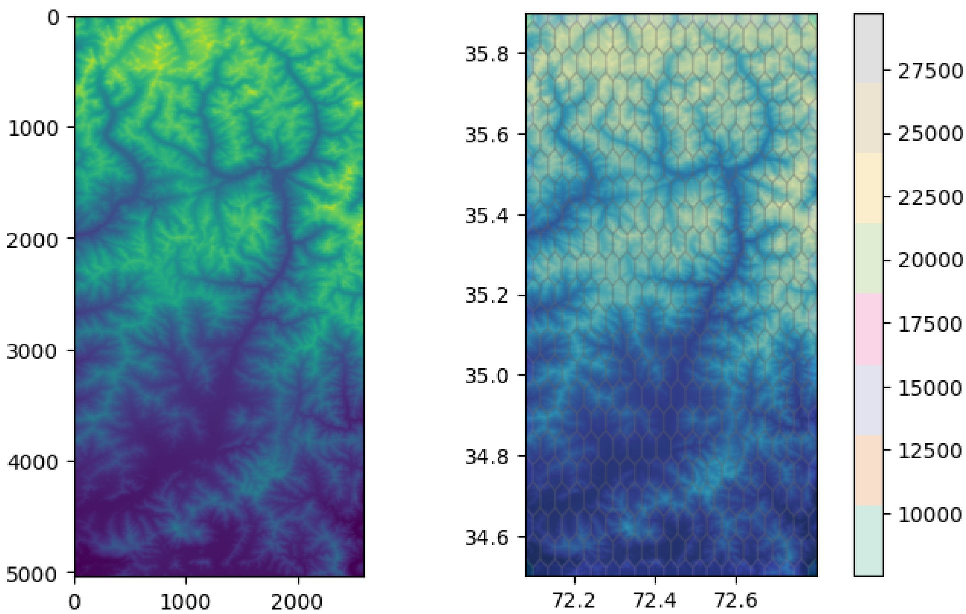
3.4 Data Preprocessing

Pre processing plays a role, in flash flood prediction and susceptibility mapping. It involves transforming and standardizing data to prepare it for analysis. This section explains the processing methods utilized in this study, which encompass data quality assurance, data integration, geographic data manipulation and data collection. These essential steps ensure that the input data is accurate, consistent and suitable for modeling and mapping processes. By addressing data inconsistencies, outliers and spatial discrepancies the pre processing stage establishes a foundation, for flash flood prediction and susceptibility mapping. This significantly enhances the reliability and validity of the studys findings.

3.4.1 Normalization

Normalization in QGIS refers to the process of transforming and rescaling data values within a specified range. Making sure that various layers or attributes are comparable and can be successfully combined or analyzed together is a common preprocessing step in GIS (Geographic Information System) analysis.

When working with datasets which contain raw raster files,they have several scales or measurement units, normalization is very crucial. By bringing all the numbers into a standardized range, usually between 0 and 1 or -1 and 1, normalizing the data makes it simpler to compare and analyze the data. the figure below (Fig 3.3) shows the raw digital elevation model and normalized digital elevation model of the study area.

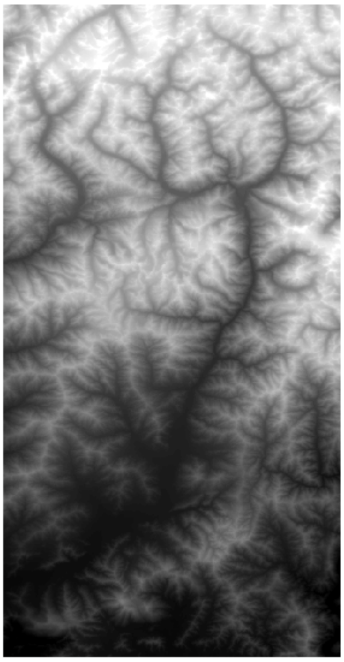


((a)) without normalization ((b)) with normalization

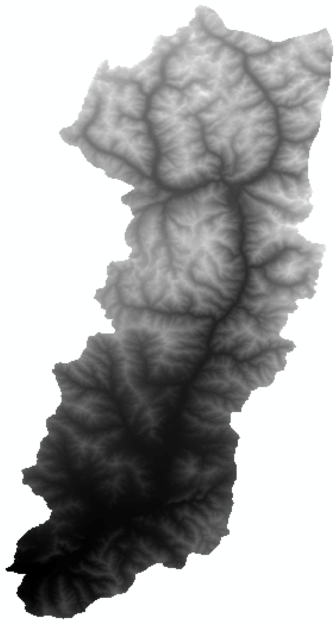
Figure 3.3: Study Area before and after Normalization

3.4.2 Clipping

The raw raster files created earlier do not have defined boundaries. The imagery to be classified must be explicit with no extra region. To achieve this, data clipping is performed in QGIS software. The software runs Geospatial Data Abstraction Library(GDAL) algorithm to perform the task. The raster files are clipped according to the boundaries of respective vector files to generate clipped raster of a specific region. the figure below (Fig 3.4) shows the raw digital elevation model and normalized digital elevation model of the study area.



((a)) normal DEM



((b)) clipped DEM

Figure 3.4: clipping of Study Area

3.4.3 Merging

flash Flood information for certain dates was included in the raw raster files that were downloaded from The Global Flood Database. The raster files were brought together to create a representation of flash flood occurrences, in the Area of Interest (AOI). By merging these files a dataset was obtained that covers the relevant time period. This guarantees that the flood data for both training and testing purposes. Combining the raster files also allowed to generate a representation of flash flood incidents, which will be incredibly helpful for conducting further analysis and modeling in our study, on flash flood prediction and susceptibility mapping. the figure below (Fig 3.5) shows the flash flooded points for the study area.

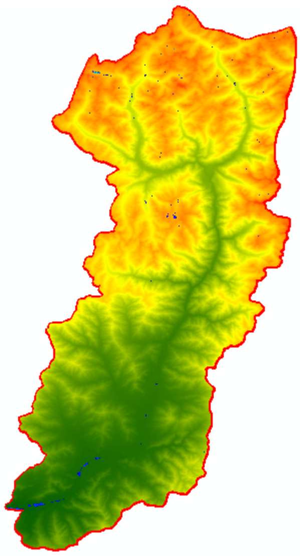


Figure 3.5: flood prone areas of Swat

3.5 Feature Extraction

Feature extraction plays a role, in predicting floods and creating susceptibility maps. It helps identify features and factors that contribute to flood events. To establish the connections between flood occurrences and various terrain variables, the ratio model was employed as a technique for feature extraction, in this section. A statistical technique known as the frequency ratio model calculates the chance of flooding based on the proportion of flooded to unflooded areas with particular topographical characteristics. Important topographical characteristics connected to flood occurrences can be found by analysing these ratios, offering insightful information on the underlying causes of flood risk. The resulting extracted features will serve as input variables for subsequent modeling and mapping processes, aiding in the accurate prediction of areas prone to flooding.

3.5.1 Frequency Ratio (FR) Method

The predictive connection between dependent and independent variables can be measured using the FR model. A fundamental statistical analysis technique called the FR method assesses the impact of each type of conditioning factor on flooding predictions. [62][63][64]. According to Bonham-Carter [65], FR is the possibility of the emergence of a certain event.

This method has been applied in a variety of sectors related to natural disasters, such as the estimation of blast-induced air blast [66] landslide susceptibility mapping [67] and flood susceptibility mapping [68][69]. This methodology has the advantage of being easy to use and producing outcomes that are completely understood. Among numerous bivariate statistical methods, the FR model was chosen for the current analyses' flood susceptibility mapping [70]. This method can be expressed by following Equation

$$FR = \frac{T_o}{I_f} = \frac{(P_f/P_t)}{(A_f/A_t)} \tag{3.1}$$

The frequency ratio (FR) represents the ratio of target occurrence, in each subcategory. The percentage of the category is denoted as $\mathrm{T_o}$ while $\mathrm{I_f}$ represents the percentage of a factor within that category.$\mathrm{P_f}$ and $\mathrm{P_t}$ denote the points within a factor class and total points, respectively. Similarly $\mathrm{A_f}$ and $\mathrm{A_t}$ represent the area, within a factor class and total area, respectively.

The frequency ratio model has proven to be a useful technique for feature extraction in flood prediction and susceptibility mapping. It helps identify key terrain attributes that contribute to flood vulnerability, providing insights into the spatial patterns and factors influencing flood occurrences. The extracted features can then be integrated into predictive models or combined with other spatial datasets to develop comprehensive flood susceptibility maps, aiding in informed decision-making, risk assessment, and the development of effective flood mitigation strategies.

3.6 Machine Learning Models

3.6.1 Random Forrest Classifier

Random forest as the name suggests consists of decision trees that collaborate as a group. The prediction made by the model is determined by the class with the number of votes. Each individual tree, in the forest contributes its class prediction.

The underlying principle behind forest is quite simple yet powerful – it leverages the wisdom of crowds. The reason why random forest performs well is because it combines a number of models (trees) that are generally uncorrelated. Working together these models outperform any model working alone.

The key factor here lies in the correlation between the models. Since they are uncorrelated they can provide forecasts that're more accurate than any single prediction. This concept is similar to how assets with correlations, such, as stocks and bonds combine to form a portfolio whose value exceeds that of its components. Long as the trees don't all lean in the same direction they provide mutual protection, against individual errors resulting in this beautiful outcome. While some trees may be wrong the majority will be right allowing the group of trees to move in the direction. Therefore for a random forest to function effectively two requirements must be met;

- Our features should contain some signal so that models built using those features perform better than guessing
- The predictions made by each tree (and their respective errors) should have correlation, with each other.

3.6.2 K-Nearest Neighbor

K nearest neighbors (KNN) is a machine learning technique that can be used for both regression and classification tasks. Its goal is to predict the class or value for a given test data point by measuring the distances, between that point. All the training data points. It then identifies the K data points to the test data point. The algorithm calculates the probability of the test data belonging to each class among these K training data points. For classification tasks it selects the class with the probability while for regression tasks it computes the value of these K nearest points.

The process of KNN can be summarized in these steps;

- Step-1: Determine the value of K, which represents the number of neighbors to consider.
- Step-2: Calculate the distance, between the test data point and its K neighbors.
- Step-3: Select the K neighbors based on the calculated distances.
- Step-4: Within these K neighbors count how many data points fall into each category (for classification) or calculate their value (for regression).
- Step-5: Assign the data point to the category that has the number of neighbors (for classification) or use the calculated average value (for regression) from, among its K nearest neighbors.
- Step-6: The KNN model is now prepared for use.

3.6.3 Support Vector Machine

Support Vector Machine (SVM) is a powerful machine learning algorithm used for classification and regression tasks. It is frequently utilised in many fields, including the visualisation of flood susceptibility.

SVM is used in flood susceptibility prediction mapping to evaluate and appraise an area's susceptibility to floods. The system takes into account a number of input factors, including slope, land cover types, topographical characteristics, rainfall information, and historical flood records. The SVM model is trained using these characteristics so that it may discover the connections between the input variables and the flood susceptibility levels.

The SVM algorithm finds an ideal hyperplane in a high-dimensional space that maximally divides several classes or categories of data points. In the context of flood susceptibility prediction mapping, the SVM model creates a decision boundary that distinguishes areas with different levels of flood susceptibility. This boundary helps identify regions that are more prone to flooding based on the input features.

3.6.4 Xtreme Gradient Boosting

XGBoost, also known as Extreme Gradient Boosting is an machine learning technique commonly employed for a range of prediction tasks, such, as flood susceptibility mapping. It falls under the category of methods, which involve combining weaker models to generate a robust predictive model.

XGBoost plays a role, in flood susceptibility prediction mapping by examining and modeling the relationship between input features and the likelihood of flooding in a specific area. These input features typically encompass elements, like the landscape, water systems, weather patterns and vegetation cover among others. The system employs a boosting technique that constructs decision trees in succession and combines their predictions to enhance accuracy.

The key benefits of utilizing XGBoost for flood susceptibility prediction mapping are;

- **Accuracy:** XGBoost is renowned for its excellent predicted accuracy because of its capacity to recognise intricate connections and interactions among various input variables.
- **Robustness:** XGBoost is made to properly manage missing data, outliers, and noise. It can handle a variety of data types and manages missing values automatically when training.
- **Feature Importance:** The built-in feature relevance ranking offered by XGBoost aids in identifying the most important elements influencing flood susceptibility. Understanding the underlying mechanisms and allocating resources for

mitigation are made easier with the use of this information.

- **Scalability:** XGBoost is effective and scalable, it can manage huge datasets with a variety of input attributes. It can be parallelized to leverage the power of modern hardware architectures, making it suitable for processing vast amounts of data.

3.7 Flood Risk Map Generation

This methodology includes several crucial elements in the process of creating flood risk maps. First, from the preprocessed DEM data, pertinent terrain attributes are chosen and extracted. A training dataset is then created, consisting of examples of flood occurrences that have been labelled and matched with the associated terrain attributes. Using this dataset, machine learning models are developed and assessed before being used to forecast flood risk throughout the study area. With the help of predictions, continuous flood risk surfaces are created, which can then be divided into discrete risk levels. The generated maps of flood risk offer useful details on areas vulnerable to flooding, assisting in efficient flood management and decision-making processes.

3.8 Performance Evaluation

After applying the machine learning models to the data, different evaluation measures can be used to check the accuracy or correctness of the model. For that various performance metrics are available, but for this problem, since it falls under the category of Supervised Classification so the methods used in this methodology are accuracy, kappa, F1-score and AUC-ROC.

Accuracy is assessed by comparing its predictions to the flood events. This is done by calculating the ratio of identified instances, including both flooded and flooded areas to the total number of instances.

$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{total predictions}} \tag{3.2}$$

Kappa is a performance parameter commonly used in machine learning and classification tasks to assess the agreement between predicted and observed outcomes. It measures the agreement beyond what would be expected by chance alone. Kappa takes into account both the accuracy of the predictions and the possibility of random agreement. it provides a reliable and robust measure of performance, especially in situations where class imbalance or chance agreement may affect the interpretation of other performance metrics.

$$\kappa = \frac{P_{Observed} - P_{bychance}}{1 - P_{bychance}} \tag{3.3}$$

F1-score is a harmonic mean of precision and recall. It provides a single metric that balances the trade-off between precision and recall. A higher F1 score indicates a better balance between accurate positive predictions and the coverage of actual flood occurrences.

$$F1\,score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall} \tag{3.4}$$

The ROC curve are depictions of a models performance showcasing the relationship, between the positive rate (sensitivity) and the false positive rate (1. Specificity) at various classification thresholds. The area, under the ROC curve (AUC) is commonly employed as a metric to gauge the effectiveness of the model, where higher AUC values suggest predictive capabilities.

Chapter 4

Results and Discussion

This section provides a comprehensive discussion of the outcomes derived from the application of the proposed methodology. It includes a detailed study that begins with the process of choosing features and explains the significance of the features chosen in connection to the area of study. The analysis's techniques are also discussed in detail, along with a thorough review of the outcomes they produced.

4.1 Dataset Summary

In this section we will provide an overview of the dataset that formed the basis of our research. We will discuss the origins and categories of the data we used to develop our flood susceptibility prediction map. When conducting analysis it is crucial to have an carefully curated dataset that is accurate and relevant. To ensure the reliability and strength of our modeling and mapping endeavors we diligently. Integrated geographic, hydrological and meteorological statistics, from multiple sources. This synthesis of data was a step, in ensuring the credibility and robustness of our work. We will provide a summary below outlining the sources, types and processing techniques applied during our investigation.

Factor	Dataset source	Spacial Resolution	Type of Data
Elevation	DEM	30x30m	GRID
Slope	DEM	30x30m	GRID
Slope Aspect	DEM	30x30m	GRID
Rainfall	Climatic Research Unit	30x30m	POINT
Curvature	DEM	30x30m	GRID
Flow Direction	DEM	30x30m	GRID
Flow Accumulation	DEM	30x30m	GRID
TWI	DEM	30x30m	GRID
LULC	ESRI	10x10m	GRID
Dist to River	DEM	30x30m	GRID
Dist to Road	DEM	30x30m	GRID
Soil	FAO DSWM	1:50,000	VECTOR
Flood Inventory	Global Flood Database	30x30m	POINT

Table 4.1: Summary Table

4.2 Selection of Features

Finding key factors that have a major impact on flood occurrences is necessary for effectively predicting flood vulnerability. The selected features are anticipated to work together to create an efficient framework for predicting flood occurrences in Pakistan's Swat District of KPK Province.

The features selected cover a broad range of geographic properties, including topographical features, climate variables, and anthropogenic causes. Each of these characteristics significantly affects how susceptible a region is to flooding incidents. The extracted features include:

4.2.1 Elevation

The chosen features encompass a range of characteristics, such, as topography, climate factors and human influences. These attributes significantly impact the vulnerability of a region to flooding incidents. The extracted features include; Elevation plays a role in the occurrence of floods[39]. To create an elevation map for this study we utilized an elevation model (DEM) derived from a Sentinel 1 image with a resolution of 30 meters obtained through OpenTopography. The elevation levels, within the study area vary between 720 and 5841 meters. Have been divided into five classes as shown in Fig 4.1a.

4.2.2 Slope

The incline factor influences both how water flows on the ground and how well the soil absorbs water[71]. We created a map using ArcGIS Pro software using DEM data with a grid cell size of 30x30 meters to show the different slope angles in our study area. The slope angle map ranges from 0° to 90°, as shown in Fig 4.1b.

4.2.3 Aspect, Curvature

In the context of flash flood modeling, we also consider additional topographic factors such as aspect, curvature, and Topographic Wetness Index (TWI)[48] [72] [73]. These factors were created using ArcGIS Pro software. Were derived from the DEM with a grid cell size of 30x30 meters. The aspect map has been divided into ten classes as shown in Fig 4.1c while the curvature map has been grouped into three levels as depicted in Fig 4.2b.

4.2.4 Rainfall

In areas, with mountains extended periods of rainfall have the potential to trigger destructive floods [74]. To address this concern we gathered rainfall information from the Climate Research Unit covering the years 2008 to 2018. Using the Inverse Distance Weighted method we generated a map that visualizes rainfall patterns [75]. The rainfall map is shown below in Fig 4.2a.

4.2.5 Flow Direction

Flow direction is crucial in flood prediction as it determines how water will move across a landscape, guiding evacuation plans, infrastructure protection [76]. the feature was produced in ArcGIS using the DEM with 30x30 m resolution. It is classified into 8 classes. The Flow Direction map is shown in Fig 4.2c.

4.2.6 Flow Accumulation

Flow Accumulation represents the horizontally topographic cleavage due to flow movement [72]. It is essential to consider this aspect when preparing for the modeling of flash flood susceptibility. To determine the length of the network we calculate the length of rivers, within a specific river basin. Flow Accumulation refers to the ratio between the length of drainage, in kilometers and an area of 1 kilometer. The drainage density within the study area is visually represented in Fig 4.2d.

4.2.7 TWI

The Topographic Wetness Index (TWI) is commonly employed to assess the influence of topography, on processes[77]. The TWI is represented in a raster format with a size of 30x30 meters as depicted in Fig 4.2e.

4.2.8 Land Use

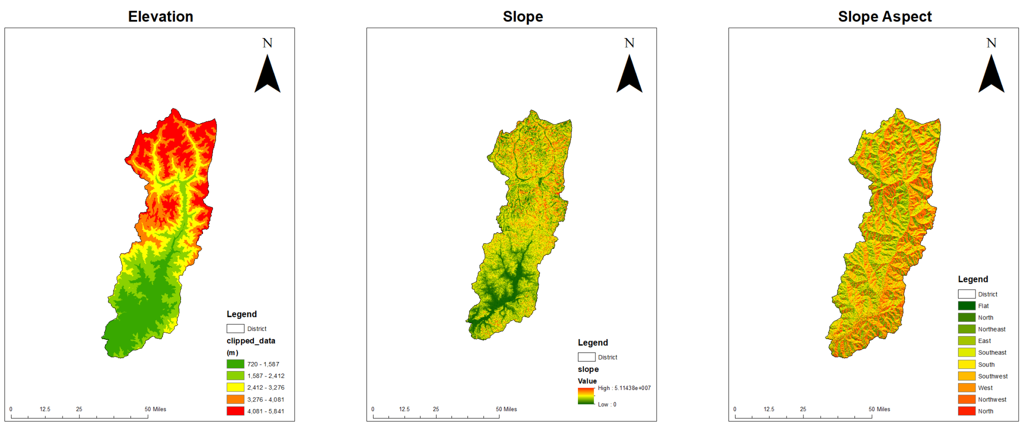
Land-use types influence some hydrological process components such as infiltration, evaporation transpiration, and runoff generation [78]. We collected the land-use category map of the study area ESRI and Impact Observatory institute which was released in 2020. the map was of 10m resolution as shown in Fig 4.2f.

4.2.9 Soil

Soil is another factor which can affect flood sensitive areas. The Soil units based on the rock permeability is also required in flood hazard assessment. The Soil data was obtained from Food and Agriculture Organization Soil map of the World (FAO-DSMW). the figure is shown in Fig 4.2i.

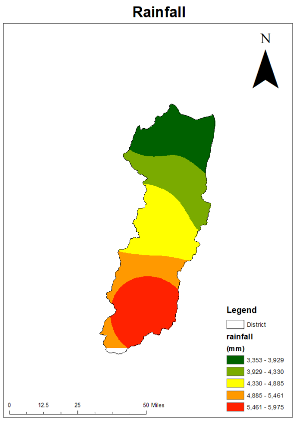
4.2.10 Distance to River and Roads

The Distance to River (Fig 4.2g) and Distance to Road (Fig 4.2h)is also a commonly used factor for identifying the flood susceptibility As river flooding primarily takes place in proximity to streams and rivers, the distance from these water bodies serves as a geomorphological factor of significance when mapping flood susceptibility[79]. Urban roads and the adjacent surfaces reduce the terrain's ability to absorb water, leading to increased runoff, which significantly impacts flood levels. Therefore, the distance from roads is an important factor to consider in flood susceptibility mapping. Additionally, since river flows serve as the primary channels for flood discharge, regions in close proximity to rivers are at a heightened risk of flooding[80]. the maps were generated in ArcGIS using the Euclidean distance tool.

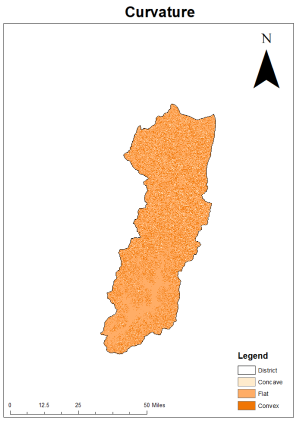


(a) Elevation (b) Slope (c) Slope Aspect

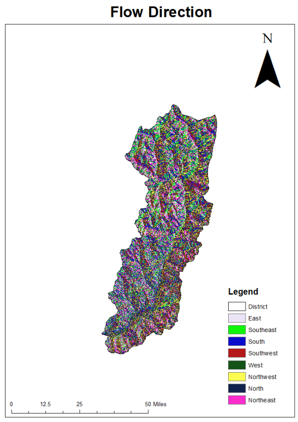
Figure 4.1: Visualization of Key Extracted Features from a Digital Elevation Model (DEM) (a) elevation, (b) slope, (c) slope aspect



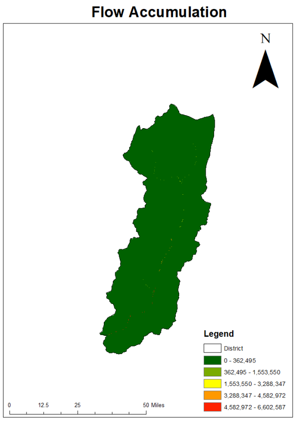
(a) Rainfall



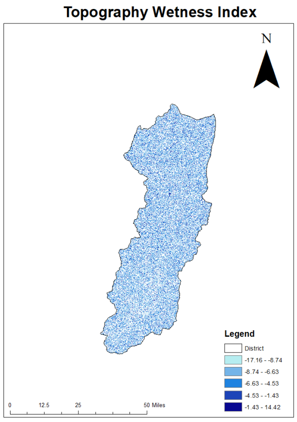
(b) Curvature



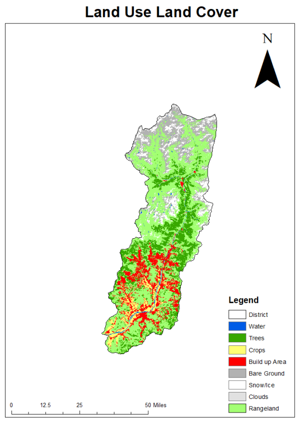
(c) Flow Direction



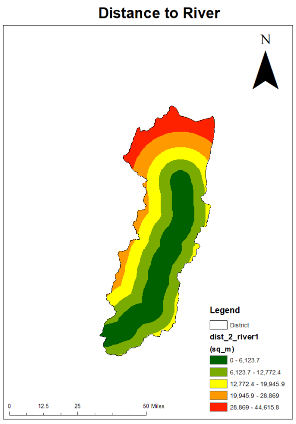
(d) Flow Accumulation



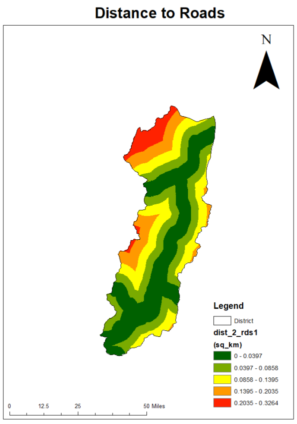
(e) TWI



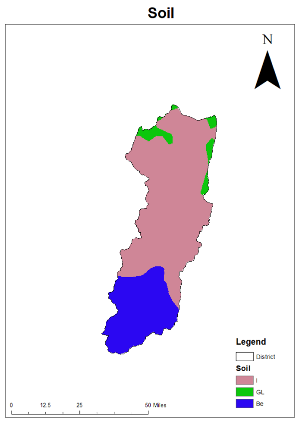
(f) LULC



(g) Distance to river



(h) Distance to road



(i) Soil

Figure 4.2: Visualization of Key Extracted Features from a Digital Elevation Model (DEM). (a) rainfall distribution, (b) curvature and (c) flow direction (d) Flow Accumulation, (e) TWI, (f) LULC, (g) Distance to River, (h) Distance to Road and (i) Soil.

the above features were selected after a comprehensive literature review and their prediction rate was calculated using the Relative Frequency Ratio method. the prediction rate of all the features generated is mentioned in the table below.

4.3 Flood Risk Prediction Model

The technique and results of our Flood Risk Prediction Model, which is essential to improving flood risk assessment and management, are presented in this section. This model was developed through a series of steps beginning with the identification of features and concluding with the implementation of machine learning techniques. Now each phase of our approach providing explanations and discussing the outcomes in the sections will be explored.

4.3.1 Feature Selection Using Frequency Ratio Method

To ensure the precision and effectiveness of our Flood Risk Prediction Model we started by choosing the valuable characteristics. We utilized the Frequency Ratio technique, a established approach, in modeling for this purpose. This technique enables us to evaluate the significance of geographical factors when determining flood risk.

The initial phase of our analysis involved the selection of relevant features using the Frequency Ratio method. This method allowed us to assign weights to each feature, thereby assessing their relative importance in the context of flood prediction. These weighted features served as a foundation for our subsequent modeling endeavors.

Features	**Frequency Ratio**
Slope	6.6
Rainfall	5.75
TWI	2.84
Soil	3.91
LULC	5.41
Flow Dir	1.0
Curvature	2.34
Aspect	5.81
Elevation	2.65
Flow Accumulation	4.83
Dist to River	4.62
Dist to Road	2.53

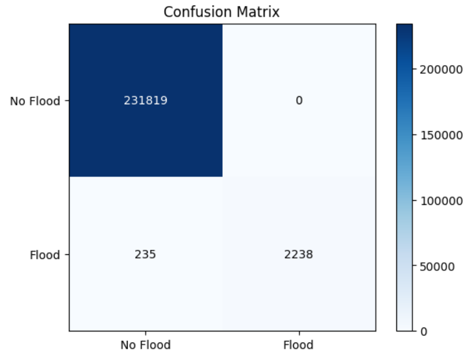
Table 4.2: Frequency Ratio Table

4.3.2 Machine Learning Model Results

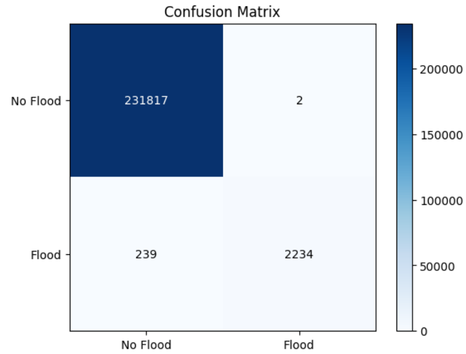
We used machine learning algorithms to forecast flood risk in the Swat region using the weighted features. Precision, recall, F1-score, and AUC-ROC performance measures were carefully assessed to assess how well these models classified locations susceptible to flooding. The results of our Flood Risk Prediction Model are presented in Table 4.3.

Model	Precision	Recall	F1-Score
Random Forest	1.0	0.90	0.95
K Nearest Neighbor	1.0	0.891	0.945
Support Vector Machine	1.0	0.889	0.949
eXtreme Gradient Boosting	1.0	0.91	0.95

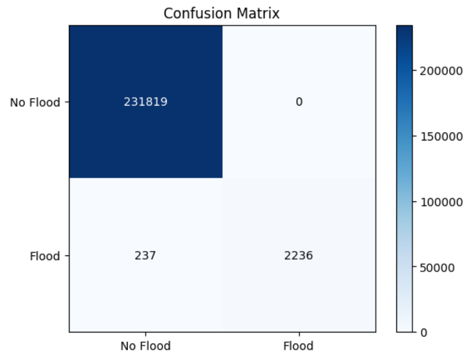
Table 4.3: Model Performance Metrics



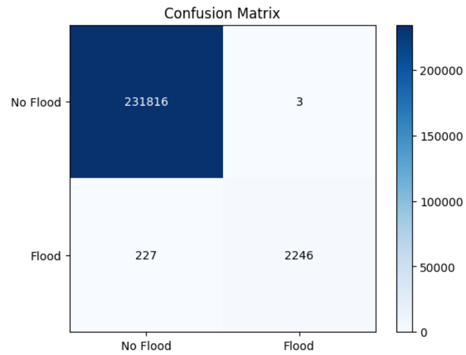
(a) Random Forest



(b) K Nearest Neighbor



(c) Support Vector Machines



(d) eXtreme Gradient Boosting

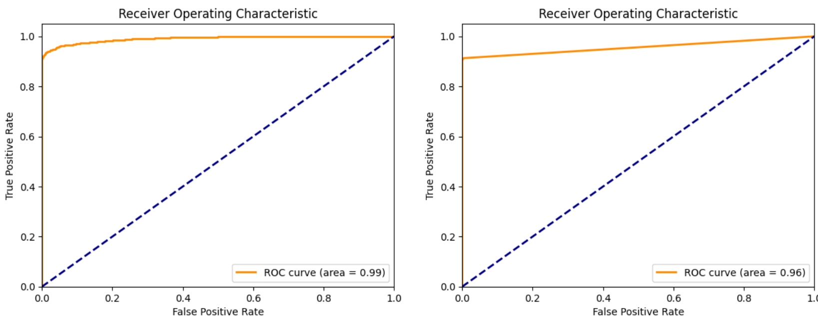
Figure 4.3: Comparison of Confusion Matrix for four machine learning models: Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and XGBoost (XGB).

With an emphasis on both precision (the capacity to accurately predict floods), recall (the capacity to record actual floods), F1-score, and the confusion matrix, to evaluate the model's effectiveness in predicting flood risks., these performance indicators offer insights into the efficiency of our model in identifying locations at danger of flooding.

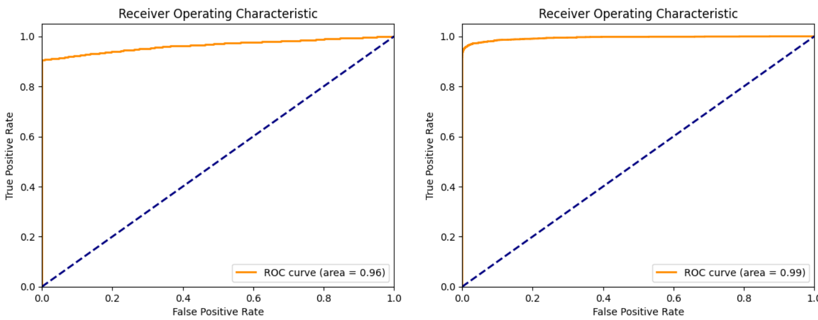
In summary, the Frequency Ratio technique led our feature selection procedure and helped us identify the major geographic features that affect the likelihood of flooding. The performance metrics mentioned above show that these characteristics' integration into our machine learning models produced positive flood risk prediction outcomes.

4.4 AUC-ROC Curve plot

The AUC-ROC scores from our Flood Risk Prediction Model, which was developed utilising a number of machine learning techniques, including Random Forest (RF), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and XGBoost (XGB), are presented in this section. These results show how well the models can distinguish between locations that are at risk of flooding and those that are not.



(a) Random Forest (b) K Nearest Neighbor

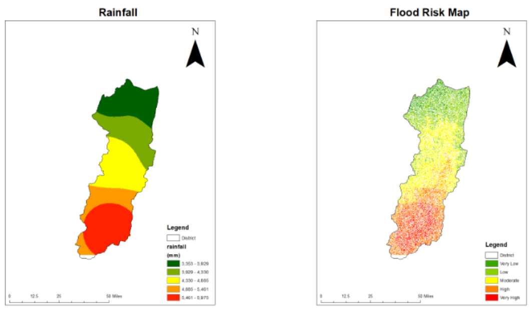


(c) Support Vector Machines (d) eXtreme Gradient Boosting

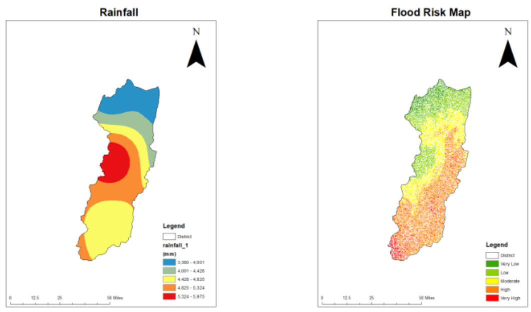
Figure 4.4: Comparison of AUC-ROC curves for four machine learning models: Random Forest (RF), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and XGBoost (XGB).

4.5 Flood Risk Maps

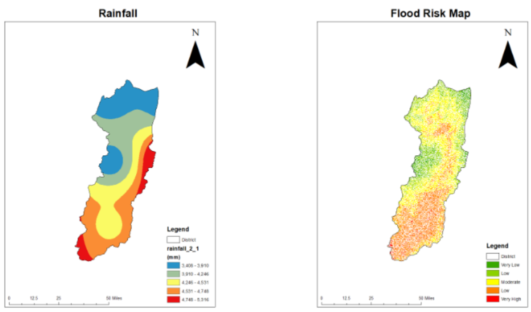
We used the generated feature weights to simulate several flood risk scenarios, drawing on the conclusions drawn from our feature weighting and machine learning investigations. These simulations were run in response to various rainfall scenarios, and the maps of flood risk that were produced are seen below. These maps were made to fit the geography and characteristics of the Swat region in the KPK District.



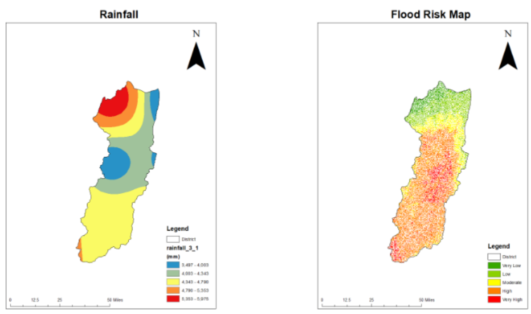
(a) Flood Risk Map 1



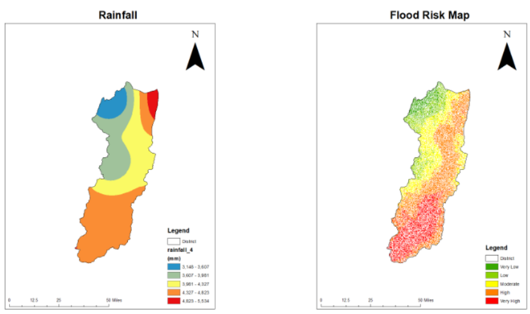
(b) Flood Risk Map 2



(c) Flood Risk Map 3



(d) Flood Risk Map 4



(e) Flood Risk Map 5

Figure 4.5: Five images depicting rainfall and flood risk maps for different districts. Each pair of maps illustrates the correlation between rainfall levels (left) and the corresponding flood risk (right) in Swat district.

The flood risk maps created using feature weighting and machine learning methods have significantly improved our understanding of flood dynamics, in the Swat region. These maps have been scientifically. Provide an empirical foundation for proactive flood risk mitigation plans and emergency response preparations. Additionally they serve as tools, for governments and disaster management organizations enhancing their ability to effectively handle potential floods.

4.6 Results Validation

We conducted a validation process to assess the trustworthiness and accuracy of the results obtained from our methodology. In this section we will compare our findings with those presented in two studies that focus on flood risk assessment and prediction.

Our research findings align closely with those of Meliho et al.2021[81] During their inquiry, into the potential for flooding in a vicinity the study reported an AUC ROC score of 0.95 0.98. The research focused on an area with environmental characteristics demonstrating the models ability to effectively differentiate between locations prone to flooding and those that are not, within the researched domain.

In a separate study conducted by Ha et al.2021[45] In the study they achieved an AUC ROC score of 0.96 which focused on predicting flood risk in a comparable context. This score demonstrates the reliability of their model, in estimating flood risks.

In our investigation we obtained AUC ROC scores for various models; Random Forest scored 0.99 K Nearest Neighbours scored 0.96 and Support Vector Machine also scored 0.96. These results indicate performance. Moreover our XGBoost model achieved an AUC ROC score of 0.99 further reinforcing the accuracy and dependability of our Flood Risk Prediction Model.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

Pakistan is one of the countries that's highly susceptible, to the impacts of climate change. One of the challenges Pakistan faces is flooding, which can be attributed to rainfall, poor management and planning melting glaciers and limited use of technology. Certain regions in Pakistan, such as Swat which is at risk of floods based on the recent 2022 floods. This study focuses on creating flood risk maps for Swat district in the Malakand Division of Khyber Pakhtunkhwa province in Pakistan. By utilizing a Digital Elevation Model (DEM) specific to Swat district relevant characteristics related to flash flooding were extracted. Four different machine learning models including Random Forest (RF) K Nearest Neighbors (KNN) Support Vector Machines (SVM) and eXtreme Gradient Boosting (XGB) were employed with training data to identify these characteristics. The evaluation of performance demonstrates an accuracy rate of 0.99 along with a precision and recall rate of 0.9 each an F1 score of 0.95 and an AUC ROC within the range from 0.956 to 0.988. The outcomes obtained from this study align with research findings[45][81], and highlight areas at risk for flooding such as low lying regions, densely populated urban areas and residential settlements situated close, to rivers. Rainfall is a factor that causes varying amounts of precipitation, in different Swat regions. It has an impact, on all the outcomes thus supporting the conclusion mentioned above.

5.2 Future Work

There are a ways we can expand on this study. Firstly besides using the elevation model and rainfall data we could incorporate detailed data and collect rainfall information, from advanced weather stations. Secondly we could utilize imagery, with resolution to further improve the outcomes.

References

[1] Abu Reza Md Towfiqul Islam, Swapan Talukdar, Susanta Mahato, Sonali Kundu, Kutub Uddin Eibek, Quoc Bao Pham, Alban Kuriqi, and Nguyen Thi Thuy Linh. Flood susceptibility modelling using advanced ensemble machine learning models. *Geoscience Frontiers*, 12(3):101075, 2021.

[2] Seyd Teymoor Seydi, Yousef Kanani-Sadat, Mahdi Hasanlou, Roya Sahraei, Jocelyn Chanussot, and Meisam Amani. Comparison of machine learning algorithms for flood susceptibility mapping. *Remote Sensing*, 15(1):192, 2022.

[3] Bruno Merz, Günter Blöschl, Sergiy Vorogushyn, Francesco Dottori, Jeroen CJH Aerts, Paul Bates, Miriam Bertola, Matthias Kemter, Heidi Kreibich, Upmanu Lall, et al. Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth & Environment*, 2(9):592–609, 2021.

[4] Anil Kumar Acharya and Narayan Kafle. Land degradation issues in nepal and its management through agroforestry. *Journal of Agriculture and Environment*, 10:133–143, 2009.

[5] WJ Wouter Botzen, Olivier Deschenes, and Mark Sanders. The economic impacts of natural disasters: A review of models and empirical studies. *Review of Environmental Economics and Policy*, 2019.

[6] Kevin Sene and Kevin Sene. *Flash floods*. Springer, 2016.

[7] Jonathan D Woodruff, Jennifer L Irish, and Suzana J Camargo. Coastal flooding by tropical cyclones and sea-level rise. *Nature*, 504(7478):44–52, 2013.

[8] TV Ramachandra and Pradeep P Mujumdar. Urban floods: Case study of bangalore. *Disaster Dev*, 3(2):1–98, 2009.

[9] Shannon Doocy, Amy Daniels, Sarah Murray, and Thomas D Kirsch. The human impact of floods: a historical review of events 1980-2009 and systematic literature review. *PLoS currents*, 5, 2013.

[10] Geographic Society National. The many effects of flooding, 2022. URL https://education.nationalgeographic.org/resource/many-effects-flooding/.

[11] Olivia Lai. What are the main causes and effects of floods around the world?, Feb 2023. URL https://earth.org/what-are-the-main-causes-and-effects-of-floods/.

[12] Azar Shokri, Sadaf Sabzevari, and Seyed Ahmad Hashemi. Impacts of flood on health of iranian population: Infectious diseases with an emphasis on parasitic infections. *Parasite epidemiology and control*, 9:e00144, 2020.

[13] Victor Klemas. Remote sensing of floods and flood-prone areas: An overview. *Journal of Coastal Research*, 31(4):1005–1013, 2015.

[14] Li Lin, Liping Di, Eugene Genong Yu, Lingjun Kang, Ranjay Shrestha, Md Shahinoor Rahman, Junmei Tang, Meixia Deng, Ziheng Sun, Chen Zhang, et al. A review of remote sensing in flood assessment. In *2016 Fifth International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, pages 1–4. IEEE, 2016.

[15] Jaroslaw Chormanski, T Okruszko, S Ignar, Okke Batelaan, KT Rebel, and MJ Wassen. Flood mapping with remote sensing and hydrochemistry: A new method to distinguish the origin of flood water during floods. *Ecological Engineering*, 37(9):1334–1349, 2011.

[16] Amir Mosavi, Pinar Ozturk, and Kwok-wing Chau. Flood prediction using machine learning models: Literature review. *Water*, 10(11):1536, 2018.

[17] Kishan Kumar Ganguly, Nadia Nahar, and BM Mainul Hossain. A machine learning-based prediction and analysis of flood affected households: A case study

of floods in bangladesh. *International journal of disaster risk reduction*, 34:283–294, 2019.

[18] Lamovec Peter, Mikoš Matjaž, and Oštir Krištof. Detection of flooded areas using machine learning techniques: Case study of the ljubljana moor floods in 2010. *Disaster Advances*, 6(7):4–11, 2013.

[19] Raymond J Burby. Flood insurance and floodplain management: the us experience. *Global Environmental Change Part B: Environmental Hazards*, 3(3):111–122, 2001.

[20] Interagency Floodplain Management Review Committee (US) and United States. Federal Interagency Floodplain Management Task Force. *Sharing the challenge: Floodplain management into the 21st century: Report of the Interagency Floodplain Management Review Committee to the administration floodplain management task force*. The Committee, 1994.

[21] Bob Freitag, Susan Bolton, Frank Westerlund, and Julie Clark. *Floodplain management: a new approach for a new era*. Island Press, 2012.

[22] Changjun Liu, Liang Guo, Lei Ye, Shunfu Zhang, Yanzeng Zhao, and Tianyu Song. A review of advances in china's flash flood early-warning system. *Natural hazards*, 92:619–634, 2018.

[23] Jan Cools, Demetrio Innocenti, and Sarah O'Brien. Lessons from flood early warning systems. *Environmental science & policy*, 58:117–122, 2016.

[24] Lorenzo Alfieri, Peter Salamon, Florian Pappenberger, Fredrik Wetterhall, and Jutta Thielen. Operational early warning systems for water-related hazards in europe. *Environmental Science & Policy*, 21:35–49, 2012.

[25] Guy J-P Schumann. Remote sensing of floods. In *Oxford Research Encyclopedia of Natural Hazard Science*. 2017.

[26] Valeria V Krzhizhanovskaya, GS Shirshov, Natalia B Melnikova, Robert G Belleman, FI Rusadi, BJ Broekhuijsen, Ben P Gouldby, Julien Lhomme, Bartosz Balis,

Marian Bubak, et al. Flood early warning system: design, implementation and computational modules. *Procedia Computer Science*, 4:106–115, 2011.

[27] Albandari Alsumayt, Nahla El-Haggar, Lobna Amouri, Zeyad M Alfawaer, and Sumayh S Aljameel. Smart flood detection with ai and blockchain integration in saudi arabia using drones. *Sensors*, 23(11):5148, 2023.

[28] Zaira Manzoor, Muhsan Ehsan, Muhammad Bashir Khan, Aqsa Manzoor, Malik Muhammad Akhter, Muhammad Tayyab Sohail, Asrar Hussain, Ahsan Shafi, Tamer Abu-Alam, and Mohamed Abioui. Floods and flood management and its socio-economic impact on pakistan: A review of the empirical literature. *Frontiers in Environmental Science*, 10:2480, 2022.

[29] Relief Islamic. Pakistan monsoon floods 2022 islamic relief pakistan (12 october, 2022) - pakistan, Oct 2022. URL https://reliefweb.int/report/pakistan/pakistan-monsoon-floods-2022-islamic-relief-pakistan-12-october-2022.

[30] Encyclopedia Wikipedia. 2022 pakistan floods, May 2023. URL https://en.wikipedia.org/wiki/2022_Pakistan_floods.

[31] Jan Cools, P Vanderkimpen, G El Afandi, Ahmed Abdelkhalek, S Fockedey, M El Sammany, G Abdallah, M El Bihery, Willy Bauwens, and M Huygens. An early warning system for flash floods in hyper-arid egypt. *Natural Hazards and Earth System Sciences*, 12(2):443–457, 2012.

[32] Melisa Acosta-Coll, Francisco Ballester-Merelo, Marcos Martinez-Peiró, and Emiro De la Hoz-Franco. Real-time early warning system design for pluvial flash floods—a review. *Sensors*, 18(7):2255, 2018.

[33] Yawen Zang, Yu Meng, Xinjian Guan, Hong Lv, and Denghua Yan. Study on urban flood early warning system considering flood loss. *International Journal of Disaster Risk Reduction*, 77:103042, 2022.

[34] J Rolf Olsen. Climate change and floodplain management in the united states. *Climatic Change*, 76(3-4):407–426, 2006.

[35] Philip Tetteh Padi, Giuliano Di Baldassarre, and Attilio Castellarin. Floodplain management in africa: Large scale analysis of flood data. *Physics and Chemistry of the Earth, Parts A/B/C*, 36(7-8):292–298, 2011.

[36] C Ndabula, GG Jidauna, K Oyatayo, PD Averik, and EO Iguisi. Analysis of urban floodplain encroachment: Strategic approach to flood and floodplain management in kaduna metropolis, nigeria. *Journal of Geography and Geology*, 4(1):170, 2012.

[37] Edyta Kiedrzyńska, Marcin Kiedrzyński, and Maciej Zalewski. Sustainable floodplain management for flood prevention and water quality improvement. *Natural Hazards*, 76:955–977, 2015.

[38] Saeid Janizadeh, Mohammadtaghi Avand, Abolfazl Jaafari, Tran Van Phong, Mahmoud Bayat, Ebrahim Ahmadisharaf, Indra Prakash, Binh Thai Pham, and Saro Lee. Prediction success of machine learning methods for flash flood susceptibility mapping in the tafresh watershed, iran. *Sustainability*, 11(19):5426, 2019.

[39] Dieu Tien Bui, Paraskevas Tsangaratos, Phuong-Thao Thi Ngo, Tien Dat Pham, and Binh Thai Pham. Flash flood susceptibility modeling using an optimized fuzzy rule based feature selection technique and tree based ensemble methods. *Science of the total environment*, 668:1038–1054, 2019.

[40] Romulus Costache, Quoc Bao Pham, Ehsan Sharifi, Nguyen Thi Thuy Linh, Sani Isah Abba, Matej Vojtek, Jana Vojteková, Pham Thi Thao Nhi, and Dao Nguyen Khoi. Flash-flood susceptibility assessment using multi-criteria decision making and machine learning supported by remote sensing and gis techniques. *Remote Sensing*, 12(1):106, 2019.

[41] Farzaneh Sajedi Hosseini, Bahram Choubin, Amir Mosavi, Narjes Nabipour, Shahaboddin Shamshirband, Hamid Darabi, and Ali Torabi Haghighi. Flash-flood hazard assessment using ensembles and bayesian-based machine learning models: Application of the simulated annealing feature selection method. *Science of the total environment*, 711:135161, 2020.

[42] Shahab S Band, Saeid Janizadeh, Subodh Chandra Pal, Asish Saha, Rabin Chakrabortty, Assefa M Melesse, and Amirhosein Mosavi. Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. *Remote Sensing*, 12(21):3568, 2020.

[43] Romulus Costache and Dieu Tien Bui. Identification of areas prone to flash-flood phenomena using multiple-criteria decision-making, bivariate statistics, machine learning and their ensembles. *Science of The Total Environment*, 712:136492, 2020.

[44] Sherif Ahmed Abu El-Magd, Biswajeet Pradhan, and Abdullah Alamri. Machine learning algorithm for flash flood prediction mapping in wadi el-laqeita and surroundings, central eastern desert, egypt. *Arabian Journal of Geosciences*, 14:1–14, 2021.

[45] Hang Ha, Chinh Luu, Quynh Duy Bui, Duy-Hoa Pham, Tung Hoang, Viet-Phuong Nguyen, Minh Tuan Vu, and Binh Thai Pham. Flash flood susceptibility prediction mapping for a road network using hybrid machine learning models. *Natural hazards*, 109(1):1247–1270, 2021.

[46] Samy Elmahdy, Tarig Ali, and Mohamed Mohamed. Flash flood susceptibility modeling and magnitude index using machine learning and geohydrological models: A modified hybrid approach. *Remote Sensing*, 12(17):2695, 2020.

[47] Wei Chen, Haoyuan Hong, Shaojun Li, Himan Shahabi, Yi Wang, Xiaojing Wang, and Baharin Bin Ahmad. Flood susceptibility modelling using novel hybrid approach of reduced-error pruning trees with bagging and random subspace ensembles. *Journal of Hydrology*, 575:864–873, 2019.

[48] Dieu Tien Bui, Phuong-Thao Thi Ngo, Tien Dat Pham, Abolfazl Jaafari, Nguyen Quang Minh, Pham Viet Hoa, and Pijush Samui. A novel hybrid approach based on a swarm intelligence optimized extreme learning machine for flash flood susceptibility mapping. *Catena*, 179:184–196, 2019.

[49] Jun Liu, Jiyan Wang, Junnan Xiong, Weiming Cheng, Huaizhang Sun, Zhiwei Yong, and Nan Wang. Hybrid models incorporating bivariate statistics and machine learning methods for flash flood susceptibility assessment based on remote sensing datasets. *Remote Sensing*, 13(23):4945, 2021.

[50] Phuong-Thao Thi Ngo, Nhat-Duc Hoang, Biswajeet Pradhan, Quang Khanh Nguyen, Xuan Truong Tran, Quang Minh Nguyen, Viet Nghia Nguyen, Pijush Samui, and Dieu Tien Bui. A novel hybrid swarm optimized multilayer neural network for spatial prediction of flash floods in tropical areas using sentinel-1 sar imagery and geospatial data. *Sensors*, 18(11):3704, 2018.

[51] Hafiz Suliman Munawar, Ahmad Hammad, Fahim Ullah, and Tauha Hussain Ali. After the flood: A novel application of image processing and machine learning for post-flood disaster management. In *Proceedings of the 2nd International Conference on Sustainable Development in Civil Engineering (ICSDC 2019), Jamshoro, Pakistan*, pages 5–7, 2019.

[52] Ammara Nusrat, Hamza Farooq Gabriel, Sajjad Haider, Shakil Ahmad, Muhammad Shahid, and Saad Ahmed Jamal. Application of machine learning techniques to delineate homogeneous climate zones in river basins of pakistan for hydro-climatic change impact studies. *Applied Sciences*, 10(19):6878, 2020.

[53] Hafiz Suliman Munawar, Fahim Ullah, Siddra Qayyum, Sara Imran Khan, and Mohammad Mojtahedi. Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection. *Sustainability*, 13(14):7547, 2021.

[54] Kashif Ullah and Jiquan Zhang. Gis-based flood hazard mapping using relative frequency ratio method: A case study of panjkora river basin, eastern hindu kush, pakistan. *Plos one*, 15(3):e0229153, 2020.

[55] Talha Ahmed Khan, Zeeshan Shahid, Muhammad Alam, MM Su'ud, and Kushsairy Kadir. Early flood risk assessment using machine learning: A comparative study of svm, q-svm, k-nn and lda. In *2019 13th International Conference on*

Mathematics, Actuarial Science, Computer Science and Statistics (MACS), pages 1–7. IEEE, 2019.

[56] Hassan Waqas, Linlin Lu, Aqil Tariq, Qingting Li, Muhammad Fahad Baqa, Jici Xing, and Asif Sajjad. Flash flood susceptibility assessment and zonation using an integrating analytic hierarchy process and frequency ratio model for the chitral district, khyber pakhtunkhwa, pakistan. *Water*, 13(12):1650, 2021.

[57] Population Census Organization, Statistics Division, Government of Pakistan. *1998 District Census report of Buner*, volume 98. Population Census Organization, Statistics Division, Government of Pakistan, Islamabad, 2000.

[58] Muhammad Irfan Malik and F Ahmad. Flood inundation mapping and risk zoning of the swat river pakistan using hec-ras model. *Lasbela Univ J Sci Technol*, 3:45–52, 2014.

[59] Islam Bahadar, Muhammad Shafique, Tazeem Khan, Iffat Tabassum, and Muhammad Zeeshan Ali. Flood hazard assessment using hydro-dynamic model and gis/rs tools: A case study of babuzai-kabal tehsil swat basin, pakistan. *Journal of Himalayan Earth Sciences*, 48(2):129, 2015.

[60] Federal Flood Commission. Annual Flood Report 2010, 2010. URL https://www.britannica.com/event/Pakistan-Floods-of-2010. Accessed 6 Aug 2018.

[61] Amir Nawaz Khan. Analysis of flood causes and associated socio-economic damages in the hindukush region. *Natural hazards*, 59(3):1239–1260, 2011.

[62] Saro Lee and Biswajeet Pradhan. Landslide hazard mapping at selangor, malaysia using frequency ratio and logistic regression models. *Landslides*, 4(1):33–41, 2007.

[63] Moung-Jin Lee, Jung-eun Kang, and Seongwoo Jeon. Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using gis. In *2012 IEEE international geoscience and remote sensing symposium*, pages 895–898. IEEE, 2012.

[64] Debabrata Sarkar and Prolay Mondal. Flood vulnerability mapping using frequency ratio (fr) model: a case study on kulik river basin, indo-bangladesh barind region. *Applied Water Science*, 10(1):1–13, 2020.

[65] Graeme Bonham-Carter. *Geographic information systems for geoscientists: modelling with GIS*. Number 13. Elsevier, 1994.

[66] Behrooz Keshtegar, Mahdi Hasanipanah, Iman Bakhshayeshi, and Mehdi Esfandi Sarafraz. A novel nonlinear modeling for the prediction of blast-induced airblast using a modified conjugate fr method. *Measurement*, 131:35–41, 2019.

[67] Sujit Mondal and Ramkrishna Maiti. Integrating the analytical hierarchy process (ahp) and the frequency ratio (fr) model in landslide susceptibility mapping of shiv-khola watershed, darjeeling himalaya. *International Journal of Disaster Risk Science*, 4:200–212, 2013.

[68] Omid Rahmati, Hamid Reza Pourghasemi, and Hossein Zeinivand. Flood susceptibility mapping using frequency ratio and weights-of-evidence models in the golastan province, iran. *Geocarto International*, 31(1):42–70, 2016.

[69] Khabat Khosravi, Ebrahim Nohani, Edris Maroufinia, and Hamid Reza Pourghasemi. A gis-based flood susceptibility assessment and its mapping in iran: a comparison between frequency ratio and weights-of-evidence bivariate statistical models with multi-criteria decision-making technique. *Natural hazards*, 83: 947–987, 2016.

[70] Omar F Althuwaynee, Biswajeet Pradhan, Hyuck-Jin Park, and Jung Hyun Lee. A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *Catena*, 114:21–36, 2014.

[71] Wei Chen, Yang Li, Weifeng Xue, Himan Shahabi, Shaojun Li, Haoyuan Hong, Xiaojing Wang, Huiyuan Bian, Shuai Zhang, Biswajeet Pradhan, et al. Modeling flood susceptibility using data-driven approaches of naïve bayes tree, alternating

decision tree, and random forest methods. *Science of The Total Environment*, 701: 134979, 2020.

[72] Bahram Choubin, Ehsan Moradi, Mohammad Golshan, Jan Adamowski, Farzaneh Sajedi-Hosseini, and Amir Mosavi. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. *Science of the Total Environment*, 651:2087–2096, 2019.

[73] Alireza Arabameri, Sunil Saha, Wei Chen, Jagabandhu Roy, Biswajeet Pradhan, and Dieu Tien Bui. Flash flood susceptibility modelling using functional tree and hybrid ensemble techniques. *Journal of Hydrology*, 587:125007, 2020.

[74] Sara Abuzied, May Yuan, Samia Ibrahim, Mona Kaiser, and Tarek Saleem. Geospatial risk assessment of flash floods in nuweiba area, egypt. *Journal of Arid Environments*, 133:54–72, 2016.

[75] Patrick M Bartier and C Peter Keller. Multivariate interpolation to incorporate thematic surface data using inverse distance weighting (idw). *Computers & Geosciences*, 22(7):795–799, 1996.

[76] Sharad Kumar Jain, Pankaj Mani, Sanjay K Jain, Pavithra Prakash, Vijay P Singh, Desiree Tullos, Sanjay Kumar, SP Agarwal, and AP Dimri. A brief review of flood forecasting techniques and their applications. *International Journal of River Basin Management*, 16(3):329–344, 2018.

[77] Seyed Vahid Razavi Termeh, Aiding Kornejady, Hamid Reza Pourghasemi, and Saskia Keesstra. Flood susceptibility mapping using novel ensembles of adaptive neuro fuzzy inference system and metaheuristic algorithms. *Science of the Total Environment*, 615:438–451, 2018.

[78] Khabat Khosravi, Binh Thai Pham, Kamran Chapi, Ataollah Shirzadi, Himan Shahabi, Inge Revhaug, Indra Prakash, and Dieu Tien Bui. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at haraz watershed, northern iran. *Science of the Total Environment*, 627:744–755, 2018.

[79] Dieu Tien Bui, Biswajeet Pradhan, Haleh Nampak, Quang-Thanh Bui, Quynh-An Tran, and Quoc-Phi Nguyen. Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibilitgy modeling in a high-frequency tropical cyclone area using gis. *Journal of Hydrology*, 540:317–330, 2016.

[80] William D Shuster, James Bonta, Hale Thurston, Elizabeth Warnemuende, and DR Smith. Impacts of impervious surface on watershed hydrology: A review. *Urban Water Journal*, 2(4):263–275, 2005.

[81] Modeste Meliho, Abdellatif Khattabi, and Joseph Asinyo. Spatial modeling of flood susceptibility using machine learning algorithms. *Arabian Journal of Geosciences*, 14(21):2243, 2021.