BINDURA UNIVERSITY OF SCIENCE EDUCATION

FACULTY OF SCIENCES AND ENGINEERING DEPARTMENT OF DISASTER RISK REDUCTION

Application Of Machine Learning To Map Flood Susceptibility In Chadereka Ward, Muzarabani District, Zimbabwe

BY

IRENE ARGIE TAWONEZVI

B201705B

SUPERVISOR: MR PEDZISAI

A DISSERTATION SUBMITTED TO BINDURA UNIVERSITY OF SCIENCE EDUCATION IN PARTIAL FULFILLMENT OF THE REQUIREMENT OF THE BACHELOR OF SCIENCE HONOURS DEGREE IN DISASTER MANAGEMENT SCIENCES

DECLARATION

I, Irene Argie Tawonezvi (B201705B), hereby affirm that this dissertation is the result of my original research and findings. All information sourced from existing literature has been fully acknowledged, and a comprehensive reference list is included.

Signature Date 30/05/2024

RELEASE FORM

Bindura University of Science Education

NAME OF STUDENT: IRENE ARGIE TAWONEZVI

STUDENT NUMBER : B201705B

TITLE OF PROJECT: APPLICATION OF MACHINE LEARNING IN FLOOD MAPPING SUSCEPTIBILITY IN CHADEREKA WARD, MUZARABANI, ZIMBABWE

DEGREE TITLE : BACHELOR OF SCIENCE HONOURS DEGREE IN DISASTER MANAGEMENT SCIENCES

YEAR GRANTED : 2024

SIGNATURE DATE 28/05/2024

PERMANENT ADDRESS 11686 10TH STREET

GLEN VIEW 2

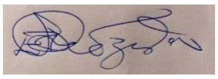
HARARE

APPROVAL

The undersigned approves that the research was supervised and was approved for final submission to the Department of Disaster Risk Reduction

Pedzisai, E.

Supervisor



28/05/2024

Signature (Supervisor) **(Date)**

24/09/24

Prof E. Mavhura (Chairperson, DRR Dpt) **(Date)**

DEDICATION

I dedicate this dissertation to God Almighty, who has been my strength and guide throughout this journey. To my beloved parents Mr. and Mrs. Tawonezvi, whose unwavering support, love, and encouragement have been my constant source of inspiration through the challenges and triumphs of this academic journey.

"For I know the plans I have for you," declares the Lord, "plans to prosper you and not to harm you, plans to give you hope and a future."

Jeremiah 29:11

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Bindura University of Science Education for providing the opportunity and resources to pursue my studies. My sincere thanks go to my supervisor, Mr. E. Pedzisai, for his patience, dedication, and unwavering support. His guidance and introduction to the world of machine learning have been instrumental in shaping this research. I would like to extend my gratitude to Mr. Musodza and Mr. Gomo for their technical assistance throughout this study. Their support and guidance have been instrumental in the completion of this research. I am also grateful to my friends Isheanesu and Ruvimbo for their constant encouragement and collaboration, which made this journey more enjoyable and manageable. Finally, I extend my heartfelt appreciation to my family for their financial and emotional support, which has been a constant source of comfort and strength throughout this process.

ABSTRACT

Floods are a leading disaster that often result in loss of thousands of human lives, affect millions of people, damage property and infrastructure worth several billions of US dollars globally. The increase in the intensity, duration and frequency hence number of flooding events due to a combination of climate change and anthropogenic factors motivated the need to explore accurate flood mapping using a robust model. The aim of the study was to model flood susceptibility using robust machine learning in the flood-prone Chadereka Ward in Muzarabani District, Zimbabwe. Fieldwork was conducted using transects across floodplain for ground truthing. Landsat 8 image was secondary data downloaded from online repository. The study calculated the normalised difference water index (NDWI) and implemented Random Forest algorithm, a decision tree machine learning using Semi-automatic Plugin (SCP) in QGIS to model flood-prone areas validated by field observations. The study used boundary data sets from Diva GIS, digitised physical infrastructure in Google Earth image in order to visualise the spatial distribution of phenomena. The spatial overlay of the physical infrastructure on a Random Forest produced flood extent map defined vulnerability and identified the infrastructure at risk of flooding. The study achieved its first objective by identifying key infrastructures such as the clinic, business centre, Chadereka primary school, and nearby homesteads predominantly located between the Hoya and Nzoumvunda rivers, indicating significant flood risk due to their proximity to these water bodies. For the second objective, the study utilized the Random Forest algorithm to model flood extent in Chadereka Ward, revealing increased flood risk along river areas with an impressive overall model accuracy of 97.75%, thus providing reliable predictions of flood-prone zones. The study addressed the third objective by overlaying the flood susceptibility map with the distribution of physical infrastructure map, revealing that infrastructures along the river banks are particularly vulnerable to flooding, effectively determining flood vulnerability in Chadereka Ward. The study demonstrated that Random Forest machine learning flood susceptibility modelling has great potential for proactive flood risk management in similar flood-prone areas. Therefore, the study recommends the testing of transfer learning techniques to model flood susceptibility in similar areas.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ACRONYMS

ML	Machine Learning
NDWI	Normalized Difference Water Index
RF	Random Forest
SCP	Semi-Automatic Classification Plugin
SVM	Support Vector Machine

CHAPTER 1: INTRODUCTION

1.1 Introduction

Flooding is a significant natural disaster that poses serious risks to communities. The increasing frequency and intensity of floods, exacerbated by climate change and human activities have led to devastating socio-economic consequences. Thus, flood susceptibility mapping is essential for effective disaster risk management particularly in flood prone areas. The application of machine learning techniques offers a promising approach to enhance the accuracy and efficiency of flood susceptibility assessments. This research proposes to investigate the effectiveness of machine learning in mapping flood susceptibility in Chadereka Ward 1, Muzarabani District, Zimbabwe, as a way to minimise the impacts posed by floods.

1.2 Background of the study

Disasters results from complex interactions of hazards and vulnerability conditions. Reducing human exposure and susceptibility to risks can lessen the severity of disaster consequences. The majority of climate change-induced disasters are impacted by shifts in land use, rising population densities, geological variables, and geographical locations.(Quesada-Román et al., 2020). Therefore, this shows that disasters are very common. As such there is need to minimise negative impacts due to disaster occurrence.

Machine learning algorithms have been employed to assess the hazard, exposure, and vulnerability to natural disasters, adopting either descriptive or predictive approaches (Gerasimos et al., 2022). Descriptive machine learning techniques are mainly applied during the response and recovery stages of the disaster management cycle. In contrast, predictive machine learning techniques are used for forecasting natural disasters, enhancing preparedness and mitigation efforts in the disaster management cycle (Yu, et al., 2018). As a result, the use of machine learning algorithms may assist in various stages of the disaster management cycle, including emergency responders, the general public who are directly affected and government authorities, allowing for more efficient evaluation of the impact of disaster events.

1.3 Statement of the problem

Muzarabani is flood prone. The area is occupied by people who have built settlements, schools, clinic, bridges and roads that are affected by floods almost annually. Their crops and livestock are also affected, while their fields are destroyed (Mavhura, 2017). Particularly in Chadereka , there are far reaching challenges of floods due to its location. Chadereka ward, located largely in between the Hoya and Nzoumvunda rivers in low lying Muzarabani area. As such, human and animal population, property, infrastructure, and the environment are often affected by frequent flash floods. Although traditional methods have been applied to mitigate these challenges, they often lack granularity and adaptivity required to handle the increasingly complex nature of flood events. This limitation highlights the potential of machine learning, particularly through the deployment of robust random forest algorithms to evaluate flood susceptibility using flood extent mapping. This innovative approach not only addresses the shortcomings of traditional methods but also strengths community resilience and preparedness against future flood events.

1.4 Justification

The results obtained from this research have benefits to the communities in Chadereka Ward and other areas by providing them with information that may help them to gain knowledge and understanding of the flood disaster dynamics hence empower them to efficiently respond to events. Since the National Policy for Disaster Management in Zimbabwe emphasizes that every citizen should assist wherever possible to avert or limit the effects of disasters (Madamombe, 2004), then this also may provide information to the development planners, and policy makers including the Government of Zimbabwe Civil Protection Department and other relevant stakeholders for example, NGOs to formulate or revise disaster legislations. Policy implementers such as Environmental Management Agency and Muzarabani Rural District Council through this research may also gain knowledge on their land use planning. The results attained from this research may contribute to the body of knowledge as they show how machine learning algorithm, particularly Random Forest can be used in mapping flood extent.in Chadereka Ward, Muzarabani. Therefore, this research sheds light on how robust artificial intelligence algorithms can expedite flood mapping so as to minimise the disastrous impacts of floods.

1.5 Aim of the study

To map flood susceptibility using random forest machine learning as a disaster management effort that seeks to minimise the impacts of floods Chadereka Ward, in Muzarabani, Zimbabwe.

1.5.1 Specific objectives

i. Model physical distribution of physical infrastructure across floodplain
ii. Model disaster susceptibility using flood extent mapping in Chadereka Ward applying robust RF machine learning
iii. Determine the extent of flood vulnerability in Chadereka Ward

1.6 Research questions

i. Which physical infrastructures are vulnerable to flood?
ii. What is the flood extent in Chadereka Ward?

iii. How is flood vulnerable characterised in Chadereka Ward?

1.7 Definition of terms

Flooding occurs when excessive rain or other sources of water accumulate in areas that are not typically submerged, leading to the inundation of land and potentially causing displacement and economic losses. (Kolen and Van Gelder, 2018)
Machine learning is an encompassing term that refers to algorithms that generate intelligent predictions based on a data set. (Nicholas, et al., 2019).

Random forest (RF) is a machine learning ensemble technique that utilizes multiple decision tree classifiers applied to diverse subsets of the dataset, with random feature subsets chosen for node splitting (Sipper and Moore, 2021).

Susceptibility refers to the probability of occurrence of an event within a selected type during a given geographical area. (Wubalem, 2022).

Vulnerability refers to the characteristics and circumstances of a society, system, or asset that render it vulnerable to the negative effects of a hazard. (UNISDR, 2009)

1.8 Organisation of the thesis

The study is organized into five chapters. The first chapter introduced the research study, including the study's background, research aims and objectives, research questions, and essential term definitions, among other elements. The second chapter reviewed related literature, providing a global, regional, and national perspective on flood mapping studies that apply machine learning. The third chapter presented the study area, and concisely described the data, data collection and analysis. It also highlighted the limitations of the study and the validity and reliability of the study. The fourth chapter presents results, interprets them and discusses them in comparison with other related studies. Finally, the fifth chapter summarises the research results, presents a conclusions and proffer recommendations.

1.9 Chapter Summary

This research emphasizes the aim of developing effective flood susceptibility mapping in Chadereka Ward to address the escalating risks of flooding. by leveraging machine learning techniques, the study seeks to analyse the physical distribution of physical infrastructure and the flood extent providing valuable insights that will enhance community resilience and improve local flood management strategies.

CHAPTER 2:LITERATURE REVIEW

2.1 Introduction

This chapter is a review of the literature on utilising machine learning to map flood susceptibility. The literature examined issues at the international, regional and national levels, drawing attention to information gaps that support this study. The chapter starts-off with a theoretical framework as outlined in the following section.

2.2 Theoretical framework

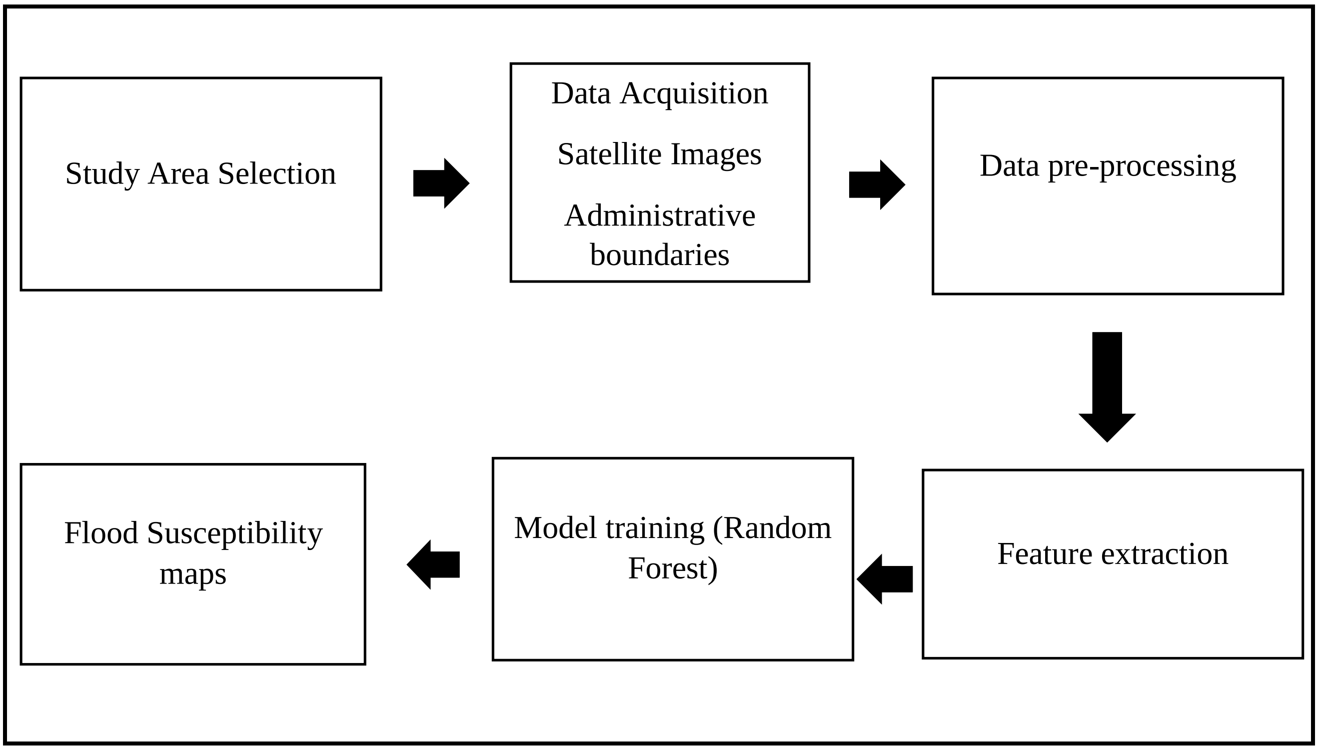


Figure 2.1: Machine Learning theoretical framework (Source: this study)

As illustrated on Figure 2.1, a Machine Learning Framework that comprises of various aspects including data collection, processing data and modelling was relevant to apply in the study. The processes involved in machine learning include data collection, data pre-processing, feature extraction, model training, model reliable performance in real-world applications. In the case of this study, it is used to assess the area of interest in relation to floods. The flood area can be dynamically be assessed by employing machine learning algorithm (Random Forest) that is able to fuse multimodal data that is generated by the analysis of satellite images and GIS based data. the framework is utilized to assess flood susceptibility within Chadereka Ward. The study used a decision-tree based the Random Forest method, which effectively incorporates multimodal data collected from satellite images and GIS-based data, to dynamically analyse flood-prone areas. Thus, offering useful insights for flood management and mitigation activities.

2.3 Global vulnerability to floods

Floods have been recognized as one of the most catastrophic hydro-meteorological hazards globally, causing significant human fatalities and socio-economic damage, drastically impacting livelihoods as well as critical infrastructure worldwide. Conversely, developing countries face heightened vulnerability to flash floods compared to developed nations due to limited financial resources, technological advancements, and infrastructure necessary for anticipating and mitigating flood disasters (Kovacs et al., 2020). Therefore, the development of a robust prediction model for mapping flood susceptibility in vulnerable regions becomes imperative for effective disaster management and risk reduction efforts.

From 2000 to 2019, the Emergency Events Database (EM-DAT), a well-known international platform for documenting such events, tracked more than7,000 disasters around the world. Over the last two decades, there have been an average of 367 disasters per year, compared to 210 events per year from 1980 to 1999, when the total number of disasters was more than 4,000 (CRED, 2019).This high frequency and significant impact highlight the critical need for effective flood risk management and mitigation strategies to protect vulnerable populations and reduce the adverse effects of flooding. Flooding and storms were the most common natural disasters during this time period. The years 2000, 2008, 2011, 2015, and 2017 stand out as having experienced the most significant floods in the world (CRED, 2018). Furthermore, given climate change predictions, changes in land use patterns, and population growth, it is expected that flood occurrence rates and intensity will increase by 2050, potentially resulting in a massive loss of approximately US$ 1 trillion (Alexander et al., 2019). Climate change, along with anthropogenic influences, has an essential role in escalating the severe impacts of flood catastrophes, resulting in massive economic loss, social upheaval, and environmental damage.

The deployment of real-time solutions such as machine learning for mapping flood hazard and the estimation of potential consequences of flood events might be extremely valuable towards confronting emergency response and mitigating the impact of those events (Dottori et al., 2015). Recognizing the need for effective flood management, the European Union enacted European Directive 2007/60/EC on flood risk assessment and management, which became effective on November 26, 2007. Flood mapping was recognized as a critical component of flood risk management in this Directive and EU Member States were required to prepare two types of emergency maps, specifically flood hazard and risk maps, by 2013 and upgrade them every six years thereafter (EXCIMAP, 2007). Disaster and development is a two-sided coin thus both aspects have to be acknowledged so that there more progress.

2.4 Regional vulnerability to floods

In the Southern Africa region, floods pose significant challenges and threats to both rural and urban populations, exacerbating existing vulnerabilities and perpetuating cycles of poverty and food insecurity. This region, like much of Africa, relies largely on floodplains and rivers for its livelihood. A structural approach to flood risk management cannot be justified on both environmental and economic grounds in large floodplains, such as those found in the Zambezi River basin, with low human density (Lumbroso et al., 2008). As a result of unplanned urbanization in Africa, the number of people living in floodplains has increased, as so has the number of fatalities from floods. These include estimating high flood flows in ungauged catchments in Africa's desert regions. Thus, floods in the Southern Africa region have far-reaching impacts, affecting not only individuals and communities but also broader regional socio-economic dynamics and development trajectories.

Floods can also cause community displacement, forcing families to flee their homes and seek safety in improvised shelters or overcrowded evacuation centres, where they face increased health risks and limited access to basic services. The 2018-2019 South-West Indian Ocean Cyclone Season, which included Cyclones Idai and Kenneth, resulted in an unprecedented level of flood damage in Africa, particularly affecting Mozambique, Zimbabwe, and Malawi (Masters, 2019). This catastrophic event led to widespread displacement, with hundreds of thousands of people losing their homes and livelihoods. Implementing machine learning algorithms for real-time flood prediction and early warning systems in the future could substantially improve disaster preparedness and response strategies, thereby protecting vulnerable communities and mitigating the devastating effects of floods.

2.5 Vulnerability to floods in Zimbabwe

In Zimbabwe, flooding is caused by both natural events, such as tropical cyclones, storms, and intense rainfall, and human actions, such as dam failures. These flood events have severe consequences for communities across the country. For example, in 2007, the government reported that about 200 families (approximately 1,200 people) were displaced by floods, and 400 huts were destroyed in Chadereka Ward (DCP, 2007). The impact of floods in Zimbabwe extends beyond immediate displacement and property damage. Floods can lead to the contamination of water sources, resulting in the spread of waterborne diseases such as cholera and typhoid (Noureen et al., 2022) Additionally, the destruction of agricultural land and infrastructure hampers food production and access to markets, exacerbating food insecurity and economic instability (Mavhura, 2019). Floodwaters often damage roads, bridges, and public utilities, disrupting communication and access to essential services (de Bruijn et al., 2019). Moreover, the psychological impact on affected populations, including trauma and stress, can have long-term effects on community resilience and well-being (Mudavanhu, 2014). These recurring flood disasters highlight the urgent need for comprehensive flood management strategies in Zimbabwe, including improved infrastructure, effective early warning systems, and community-based disaster preparedness programs to mitigate the adverse effects of future flood events.

2.5.1 Effects of floods in Mbire and Muzarabani districts

Mbire and Muzarabani floods result from both tropical and convective storms as well as backflows along river systems. Thus, Mbire and Muzarabani experience frequent flooding, particularly during the rainy season, which typically occurs from November to March. The confluence of multiple rivers and the flat topography of the region contribute to the susceptibility to flooding. In 2015, convective storms induced riverine floods and backflows along the tributaries of Hunyani River, which killed livestock, destroyed or damaged roads, crops, dwellings, bridges, schools and clinics in Mbire (Mucherera and Mavhura, 2020). Floods the intensity of flooding in Mbire and Muzarabani vary from minor inundations to catastrophic events. Factors such as rainfall intensity, river discharge, and the condition of water catchment areas influence the severity of flooding. In Mbire, floods result from both tropical and convective storms as well as backflows along river systems. Prioritizing these efforts will enable Muzarabani and Mbire to mitigate the impacts of floods and improve its resilience to future challenges.

2.5.2 Effects of floods in Chadereka ward

During the period from 2001 to 2010, the Chadereka community encountered significant flooding, ranging from moderate to severe. Regarding the depth of floodwater (average height from the base of homes), most residences in Chadereka were inundated with more than one and a half meters of water. (Mavhura et al., 2019). Muzarabani is well-known for floods, although there appears to be a distinct type of flooding occurring between and immediately adjacent to the NzouMvunda and larger Hoya rivers. The Chadereka area, located between the Hoya and Nzoumvunda rivers, is particularly prone to flash flooding. In the flood of December 2007, the NzouMvunda bridge and two Hoya bridges (one upstream and one downstream) were destroyed by the floodwaters. Despite the risk posed by periodic floods, the Chadereka village continues to thrive since their livelihoods are dependent on the agricultural opportunities afforded by flooding.

Given the dual challenge of flood vulnerability and dependence on the area for livelihoods, assisting the Chadereka community requires a holistic approach. This involves implementing flood-resistant infrastructure, establishing early warning systems, promoting sustainable agriculture practices, and providing alternative income-generating opportunities to reduce reliance on flood-dependent activities.

2.6 Factors influencing flood susceptibility

There are several factors that can influence an area's susceptibility to flooding. The shape and height of the land surface influence the flow and accumulation of water. Areas with low elevation, steep slopes, and significant curvature are more prone to floods than areas with high elevation and gentle slopes. Additionally, flooding is more likely in areas with heavy precipitation, high river discharge, high drainage density, high flow accumulation, and a close proximity to the river. For example, the Muzarabani flood plain's small topographical variation leads to bankful discharge, which is further worsened by backflow from the Cahora Bassa Dam (Chingombe et.al., 2015). Moving on to human factors, land use changes can contribute to flooding. Human activities such as agriculture, mining and urbanisation can alter the natural characteristics of the land making it prone to erosion, sedimentation and runoff (Goudie, 2018). Infrastructure development can contribute to flood susceptibility (Jamshed et al., 2020). Road, bridges and other structures can block the natural flow of water, increasing the risk of flooding. These structures can create bottlenecks, causing water to build up and overflow during heavy rains. Overall, physical and human factors influence the risk of floods.

2.7 Methods used in studying floods

2.7.1 Traditional methods to study floods

The transition from traditional methods to machine learning on mapping flood susceptibility mapping is a process of using more advanced and data driven techniques to assess and predict the likelihood and impacts of flooding in different areas. Traditional methods used to study floods have historically relied heavily on fieldwork, observational data and hydrological modelling. These methods have provided valuable insights into flood dynamics but also come with certain limitations (Malakeel et.al, 2021). Fieldwork involves direct observation and data collection in flood prone areas, river channels and floodplains. However, it can be time consuming, costly labour intensive and dangerous especially during extreme flood extends. Additionally, fieldwork may not capture the entire spatial extent of flooding. Moreover, hydrological models simulate the processes of rainfall, runoff, infiltration and streamflow to predict flood behaviour. For example, the hydrological rainfall and runoff models are the common methods for estimating flood prone areas (Lin et al., 2019). However, traditional hydrological models often rely on simplifying assumptions and may struggle to accurately represent complex real-world conditions, such as urbanisation and land use changes.

In recent years, machine learning has emerged as an effective method for flood mapping. Machine learning algorithms are able to analyse large amount of data including satellite imagery and historical flood data to create more accurate and timely flood maps. Machine learning algorithms have significantly improved flood prediction and susceptibility mapping (Bui et al., 2020). Moreover, machine learning based techniques can improve accuracy while shortening computation time and lowering model construction expenses. The transition to modern flood models which include hydraulic modelling, remote sensing, GIS and machine learning integration, represents a promising step forward in flood risk assessment and management.

2.7.2 Machine learning in predicting floods

Disaster and development are interconnected facets that are particularly evident in the context of floods. While floods can devastate communities and hinder development efforts through infrastructure damage, displacement, and economic losses, they also provide an opportunity for development interventions and renewal of ecosystems. Thus, machine learning is suggested as one of the tools that may be used to help reduce the impacts of floods by assessing flood damage and planning response efforts by analysing satellite imagery and sensor data to prioritize resources effectively (Wagenaar, et al.2020). Machine learning can forecast floods in order to minimize their harmful effects, necessitating the development of flood susceptibility maps to identify areas susceptible to flooding. Numerous researches have recently been proposed to develop flood susceptibility maps as a tool for effective flood risk management (Bui et al., 2020). Flood susceptibility indicates an area's inclination to be flooded, as determined by its physical geographical factors. Furthermore, flood susceptibility mapping can be regarded as a quantitative and qualitative assessment of an area with a high likelihood of flooding, while also providing the spatial distribution of the specific natural event (Bui et al., 2020). Flood susceptibility analysis and mapping identify the most vulnerable locations, making it one of the most significant components of early warning systems or approaches for preventing and mitigating future floods (Vojtek and Vojteková, 2019). As a result, flood susceptibility mapping is recognized as an important flood management method.

With the advancement of technology, machine learning has been suggested with the goal of efficiently mapping, monitoring and managing floods. It makes use of data to discover patterns and forecast new outcomes. Machine learning has been extensively utilized in flood mapping and prediction, showcasing its relevance in addressing flood-related challenges. To mimic the complex mathematical expressions of physical processes of floods, during the past two decades, machine learning methods contributed highly in the advancement of prediction systems providing better performance and cost-effective solutions (Amir et al., 2018). Supervised machine learning algorithms, particularly those for classification and regression tasks, have been employed to create flood susceptibility maps. These maps are crucial for identifying areas prone to flooding based on historical data and relevant features, aiding in proactive disaster management and risk mitigation strategies.

Furthermore, researchers have utilized supervised learning algorithms such as random forest (RF), support vector machines (SVM), and neural networks to develop flood susceptibility models based on various environmental and geographical factors. By training these models on labelled data, which includes past flood occurrences and associated environmental features, they can accurately predict areas at high risk of flooding in the future. In other researches, machine learning was successful in predicting flood areas. A study was conducted in Frederiction, Canada in mapping flood susceptibility using Random Forest. Through rigorous analysis and model training, the RF algorithm demonstrated exceptional performance, achieving a high accuracy rate of 97.57%.(Esfandiari et al., 2018). A similar. Another study was conducted in two flood prone districts of Assam in India with aim of mapping flood susceptibility using Random Forest and MLP Classifier to Predict Vulnerable Areas (Chakraborty and Kumar, 2024). The study developed a flood susceptibility map using two machine learning models which are Random Forest Classifier and the MLP Classifier and the findings show that Random forest has potential in large scale flood risk mapping. This predictive capability enables authorities to implement preventive measures, such as land use planning and infrastructure development, to mitigate the impact of floods on vulnerable communities.

2.8 Machine Learning in Disaster Management Cycle

The concept of disasters follows a cycle in which the management comes to play to minimise and reduce disaster risk. In 2017, the United Nations Office for Disaster Risk Reduction

(UNISDR) established a definition of disaster risk that incorporates the Sendai Framework for Disaster Risk Reduction, which operates from 2015 to 2030. Disaster management can be interpreted as a process of preparing and implementing strategies to enable people to effectively respond to and recover from disasters. Therefore, disaster models are loaded with actual exercises, there are more than just theories to increase knowledge.

One of the most popular models widely used by social scientists since the beginning of the 20th Century which brings out the life cycle concept delineates disasters into 4 main stages/phases that include mitigation, preparedness, response, and recovery (Smet et al., 2015). The disaster management cycle and emphases on Mitigation and preparedness that is part of the pre-disaster stage which aims to form capacity building. On the pre-disaster stage, machine learning models analyse historical data to identify risk factors, assess vulnerabilities, and prioritize mitigation efforts. It also assists in creating early warning systems by analysing real-time data (e.g., weather patterns).Response and recovery (rehabilitation/reconstruction) is part of the post-disaster stage. At this stage-driven decision support systems guide responders during emergencies. It also helps assess damage, prioritize recovery efforts, and allocate resources efficiently.

The primary goal of flood risk management is to reduce human and economic losses. Flood risks cannot be completely avoided; thus, they must be controlled. As a result, flood management seeks to alleviate rather than remove flood hazards. Thus, the disaster management cycle acts as a guide to show the stages and how they link together. In this case machine learning is applied to map flood prone areas in Chadereka Ward.

2.9 Gap in literature

In the field of disaster management, the development of susceptibility maps is crucial for enhancing preparedness and mitigating the impact of severe flood events. Many research studies have been undertaken with the goal of evaluating flood hazards and improving the accuracy of flood hazard mapping. While machine learning algorithms have been noted to be robust, however, they have not been widely used especially in complex micro-terrain such as Chadereka Ward to produce reliable and accurate flood extent maps. In this work, a robust machine learning technique, RF was applied using satellite images to map flood extend in Chadereka Ward so as to address this gap in literature.

2.10 Chapter Summary

Floods are regarded one of the greatest hazards to human survival. In this chapter, floods have endangered the future development. The literature review demonstrated how floods have caused harmful effects on individuals, including fatalities, injuries, psychological aftermaths based on historical estimates at global, regional and national level. Thus, machine learning is suggested as tool that can be used to map flood susceptibility due to its abilities.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter presents the research methodology regarding the application of machine learning in flood susceptibility mapping in Chadereka ward, Muzarabani, Zimbabwe. The applicability of experimental research design is discussed in-depth in this chapter. This approach allows to have a deeper understanding of the importance of flood susceptibility mapping in Chadereka ward in Muzarabani District. The primary components of this chapter include study area, research design, data collection instruments, methods and procedure and data analysis are

3.2 Study area

The study was undertaken in Chadereka Ward 1 of Muzarabani District, Mashonaland Central Province, in Zimbabwe's north-east. Chadereka Ward 1 is 60 km north of the Muzarabani Business Centre. (Kasimba, 2017). The majority of the Zambezi basin consists of a plateau that is between 1000 and 1500 meters above sea level. The basin's topography ranges in altitude from sea level at its delta on the Indian Ocean to more than 1500 meters on the plateau, with mountain areas surpassing 2500 meters. (Madamombe, 2004). Based on the agro-ecological boundaries of the country, the study area is located in region IV, which has a dry climate with an average rainfall of 550 mm per year and extremely high temperatures of over 400C during the spring and summer months (Mugandani et al., 2012). The rainy season primarily relies on a low-pressure system known as the ITCZ, which migrates to the southern region of Zambia around November and reaches its climax in January or February (Madamombe, 2004). This peak period corresponds to when rivers are also at their highest levels, resulting in frequent occurrences of flooding during this time of year. (Mudavanhu, 2014). Chadereka ward suffers from disasters triggered by weather related hazards such as floods, and epidemics such as malaria and cholera. Its vulnerability stems from its geophysical, socio-economic and political conditions.

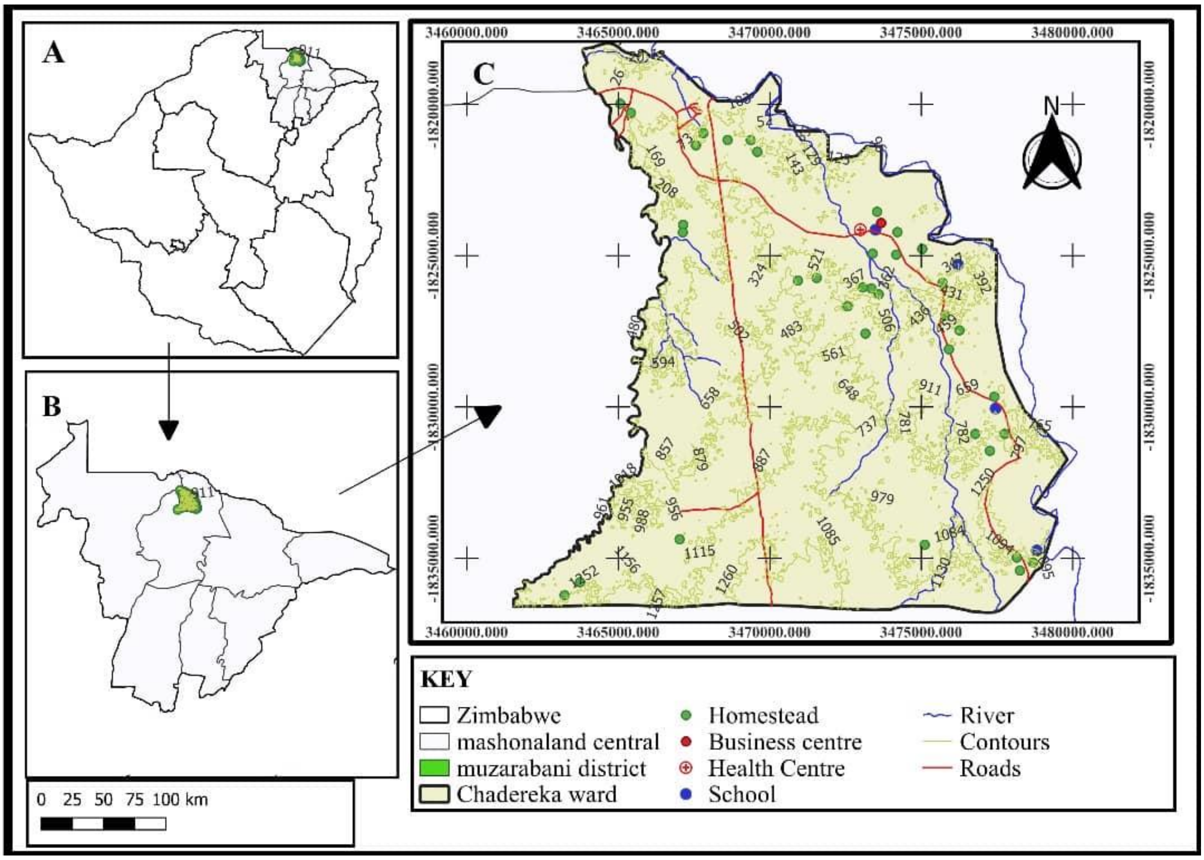


Figure 3.1: Map showing location of Zimbabwe (a), Mashonaland Central Province (b) and Chadereka (c) (source: author).

3.3Research Design

Experimental research design is used to map flood susceptibility by evaluating impacts of floods such as rainfall intensity and land use on the flood extend. Research design discusses to the structure, plan and strategy of investigation comprehended so as to attain answer to the research questions posed in chapter 1.A research design is a systematic framework for answering research questions through the collection and analysis of empirical data (McCombes, 2021). Apart from that, it acts as a blueprint for a study so as it guides data collection and analysis. In this research, random forest was used in the flood mapping susceptibility in Chadereka Ward, Muzarabani District.

3.4 Data used

The author applied a combination of primary and secondary data. in this study. Primary data collection was done during field observations whilst secondary data was acquired from earth explorer.

3.4.1 Primary data

The study used field observation data which is a way of gathering data by observation of vegetation, the infrastructures, human settlements and general terrain of the study area.

3.4.2 Secondary data

Secondary data is vital in quantitative research on flood susceptibility mapping using satellite imagery. It encompasses existing datasets, reports, and scholarly literature that have been collected and analysed by other researchers.

3.5 Data Collection Instruments

The study used more than one data collection instrument to meet the study's objectives and also ensuring data reliability and validity.

3.5.1 Earth Explorer

Earth Explorer is a cloud-based platform which provide access to vast collection of satellite imagery through United States Geological Survey (USGS). Thus, it was used as the principal tool to download the Landsat satellite imagery in the study. Registration and accessing USGS was done through (https://earthexplorer.usgs.gov/). In this research, satellite imagery was obtained through earth explorer, offering valuable insights into the spatial distribution and severity of the flooding.

3.5.2 Google Earth Engine

Google Earth Engine is a valuable tool for flood mapping susceptibility as it provides high resolution satellite imagery that gives a comprehensive view of the landscape including topology, landcover and water bodies through https://earthengine.google.com/. Google Earth Engine (GEE) offers a cloud-based computing platform designed for storing and analysing massive datasets, reaching sizes up to petabytes, to facilitate informed decision-making and comprehensive analysis. (Kumar and Mutanga, 2018). In this case, GEE helps to identify areas with critical infrastructure such as Chadereka clinic, Chadereka primary and secondary schools.

3.5.3 Global Positioning Systems

The GPS Garmin eTrex 10 (Gamin) was used to collect field data on topographical characteristics, land use patterns, infrastructure, and household distribution. Ground truthing was carried out utilizing GPS-guided fieldwork, which involves validating and calibrating remote sensing data through direct observations on the ground.

3.6 Data Collection Methods

3.6.1 Field Observation

Field observations were made to investigate the size and depth of the flood, the area's topography, the type and structure of the built environment, settlement patterns, existing land usage, road conditions, and so on. The transect walks were undertaken during the recognition visit in the first two weeks of February 2024 to supplement the use of Google Earth Engine data and machine learning.

3.7 Data Collection Procedure

3.7.1 Downloading of satellite images

The study employed high-quality, spatially resolved Landsat 8 OLI/TIRS imagery, obtained from the United States Geological Survey's (USGS) free Earth Explorer service (https://earthexplorer.usgs.gov/). First, an account was registered on the platform, creating a username and password and answering security questions to ensure privacy. After logging in, the necessary images were selected and downloaded based on specific study criteria such as date range and cloud cover. The choice of Landsat 8 was driven by its high radiometric resolution, 30-meter spatial resolution, and cost-free accessibility, facilitating detailed and financially feasible environmental and geographical analysis.

3.7.2 Downloading administrative boundaries

The study obtained administrative boundaries from DIVA-GIS. The process involved navigating to the DIVA-GIS website (https://www.diva-gis.org) and accessing the data download section. Specifically, Zimbabwe was selected as the geographic region of interest, and administrative boundaries at the district level were chosen to obtain detailed spatial information relevant to the study. Subsequently, the corresponding shapefile dataset was downloaded, which included administrative boundaries such as province, district and ward boundaries. The accompanying metadata was carefully reviewed to ensure compatibility with the research objectives. Finally, the accuracy and currency of the boundaries were verified before integration into GIS software for further analysis and visualization.

3.8 Sampling

To achieve the intended objectives of this study, Landsat 8 image was required to map flood susceptibility in Chadereka Ward. Purposive non-probability sampling was utilized in this study, which means choosing a sample based on the researcher's assessment and specific goal rather than at random.(Teddlie and Yu, 2007). As a result, the Landsat image from February 7, 2024 was specifically chosen. This was due to the fact that it was rainy season, and the cloud cover was low on this specific day. Furthermore, due to its geographical location and topographical features, Chadereka is more vulnerable to flooding, with even minor flooding events having more severe consequences. Hence, for that reason the ward was purposively selected for the purposes of this study.

3.9 Data Analysis

The Normalized Difference Water Index (NDWI) was used for mapping flood susceptibility in Chadereka Ward.The NDWI was calculated using the formula as shown on Equation (1);

$$NDWI = \frac{(Green - NIR)}{(Green + NIR)} \qquad \text{Equation (1)}$$

Where Green and NIR represent the green and near infrared bands, respectively.

The NDWI is computed through differencing the NIR from the green band divided by the sum of the two bands (Equation 1). The NDWI was proposed by McFeeters in 1996 to detect and hence show water bodies which in disaster management relate to flood inundations

Moreover, Random Forest classification was conducted in QGIS (version 3.36), by obtaining multispectral satellite imagery and gathering training data representing classes of interest, such as flooded and non-flooded areas. Installation of Semi-Automatic Classification Plugin (SCP) within QGIS was done to facilitate machine learning tasks. Furthermore, configuring the Random Forest parameters, including the number of trees and features, so that they suit the classification task as shown in appendix 1. Training the Random Forest classifier using the training data, assigning classes to pixels based on their spectral characteristics. Applying the trained classifier to classify the entire image, producing a classified map. Validating the accuracy of the classification results was done through post-processing accuracy assessment

Furthermore, the spatial overlay of the physical infrastructure on a Random Forest produced flood extent map defined vulnerability and identified the infrastructure at risk of flooding. The process was conducted in QGIS (3.36) by ensuring both layers had the same (Coordinate Reference System) CRS. Thus, enabling targeted risk mitigation strategies and informing future development planning.

3.10 Validity and reliability

The data collected and used in this study was obtained from official sources. Satellite images were downloaded from (USGS) via (www.earthexplorer.USGS). Moreover, shapefiles of wards, rivers and roads was obtained from DIVA GIS website. Google Earth enhances the validity of flood susceptibility mapping by facilitating visual interpretation of high-resolution satellite imagery and enabling ground truthing efforts. QGIS (3.36) software was used on flood susceptibility mapping in Chadereka Ward, as it has robust capabilities for data processing, algorithm implementation, automation, integration with geospatial tools. Ground truthing obtained through field work hence indications of areas flooded from the fields helps to enable validity and reliability of results.

3.11 Limitations of the study

Internet connectivity often affected the downloading of the satellite images. For example, collecting data using computers and mobile phones, the accuracy may be affected by internet connectivity. In this case, accessing websites was difficult due to issues such as latency, pocket loss and connection errors. To reduce these issues, satellite image downloads were scheduled during off peak hours such as mid night where network congestion will be low. Moreover, the size of one image for the study area since Landsat 8 is heavy dataset (over 1 gigabyte per image). Power outrage interrupted and hence affected data collection. For example, during data collection, loss of power led to loss of partially downloaded data. Power outrage was a major limiting factor. In order to overcome this challenge, the author used backup external drives to ensure research continuity. Also, site accessibility of area during rainy season, distance to study area, long protocol to collect data leading to delays, reluctance of community members to help identify flood line. Hence field work had to be delayed but was eventually done when roads were navigable.

3.12 Ethical considerations

In the process of utilising data from USGS and GEE for quantitative research on flood susceptibility mapping , the author created user accounts within these platforms with respect to the ethical principles. Prior to accessing and using the datasets provided by USGS and GEE, the user accounts were created following the platform terms of service and data usage policies. This involved providing accurate information and agreeing to the terms and conditions outlined by the platforms, ensuring compliance with their intellectual property rights and data sharing protocols. By creating user accounts in a transparent and ethical manner, the research-maintained integrity and respect for the rights of data providers, while also facilitating responsible access to valuable geospatial data for scientific analysis. Additionally, during fieldwork, ethical principles guided interactions with local communities, ensuring respectful engagement. These principles included informed consent, respect for autonomy, cultural sensitivity and reciprocity. By upholding ethical principles and standards, researchers can uphold the trust and credibility of their research findings and contribute to positive societal outcomes (Smith, 2019). Therefore, by embracing ethical research practices, it promotes a culture of mutual respect and collaboration between researchers and local communities.

3.13 Chapter Summary

This chapter focuses on the description of the study area, study design, data collection instruments, sampling procedure, and data analysis. Therefore, the following chapter will present, analyse and discuss the results pertaining this study.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents results of application of machine learning in flood susceptibility mapping in Chadereka Ward, Muzarabani, Zimbabwe. These results are analysis centred on three primary objectives namely to examine the spatial distribution of infrastructure within the floodplain, flood extent modelling using the Random Forest (RF) machine learning algorithm to model flood extent and assessment of the vulnerability of Chadereka Ward to flooding disasters.

4.2 Distribution of physical infrastructure in Chadereka ward

The spatial distribution of infrastructure in Chadereka Ward reveals a varied landscape of human settlement and development. The distribution of infrastructure, terrain and drainage in the study area as shown in Figure 4.1

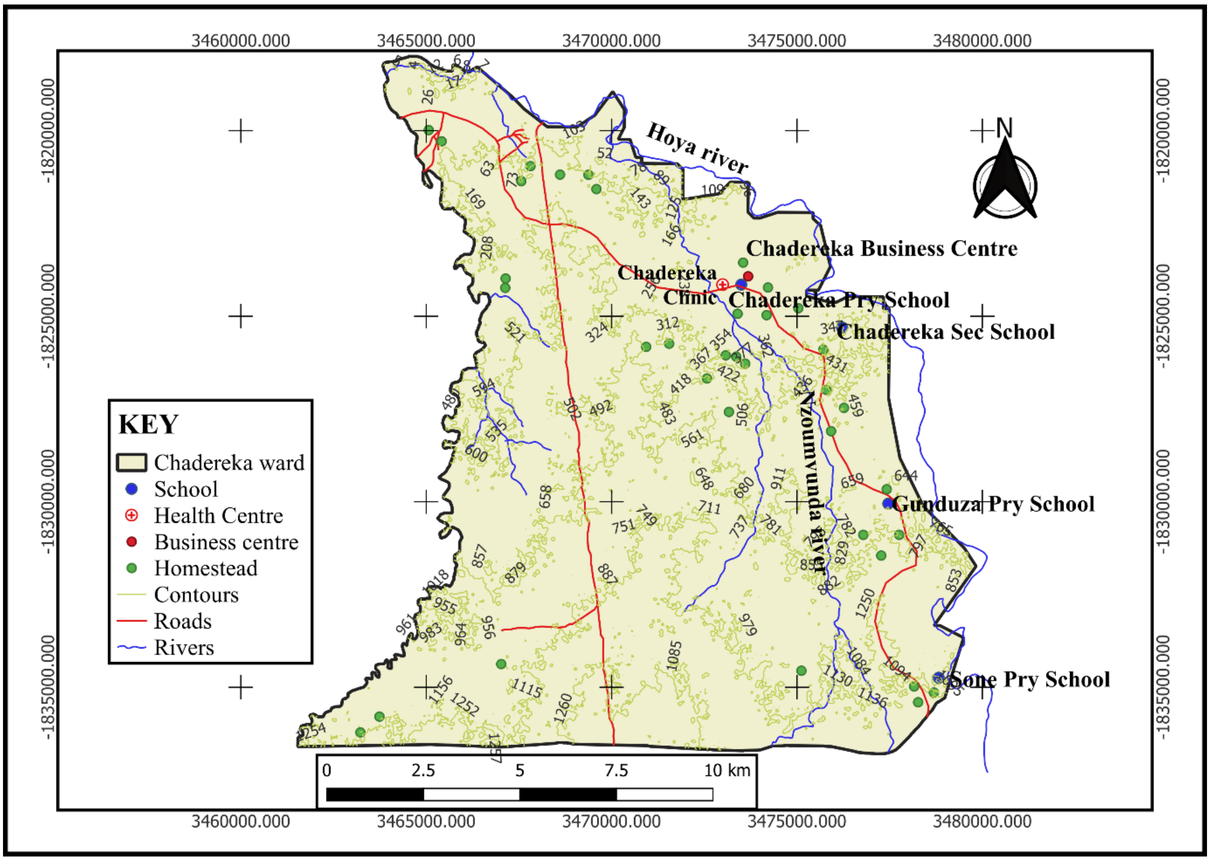


Figure 4.1: Map showing the distribution of infrastructure in Chadereka Ward 1 (source: author)

As illustrated in Figure 4.1, the majority of infrastructure is located near the banks of the Hoya and Nzoumvunda rivers. The proximity of settlements to the riverbanks (Figure 4.1) could be influenced by several factors such as historical settlement patterns, where communities initially developed along water sources for access to water for drinking, irrigation, transportation, and other livelihood activities. However, this proximity also increases the risk of flooding, as settlements situated close to rivers are more susceptible to inundation during periods of high-water levels.

4.3 Charactering flood extent in Chadereka ward

In order to show flood-prone areas, the Normalized Difference Water Index (NDWI) was conducted. The NDWI values ranged from were 0.21159 and -0.324403which were classified into two classes of pixels which are water and non-water. Pixels with NDWI > -0.0564 were classified were occupied with water, hence regarded as flood pixels while those where NDWI ≤ -0.0564 were not flooded ones. This was done using the Raster Calculator in QGIS (3.36). The RF machine learning classification resulted in, a flood extent map showing the areas that are susceptible to flood inundation. A flood extent map derived from a density slicing of the NDWI map of the study is shown on Figure 4.2.

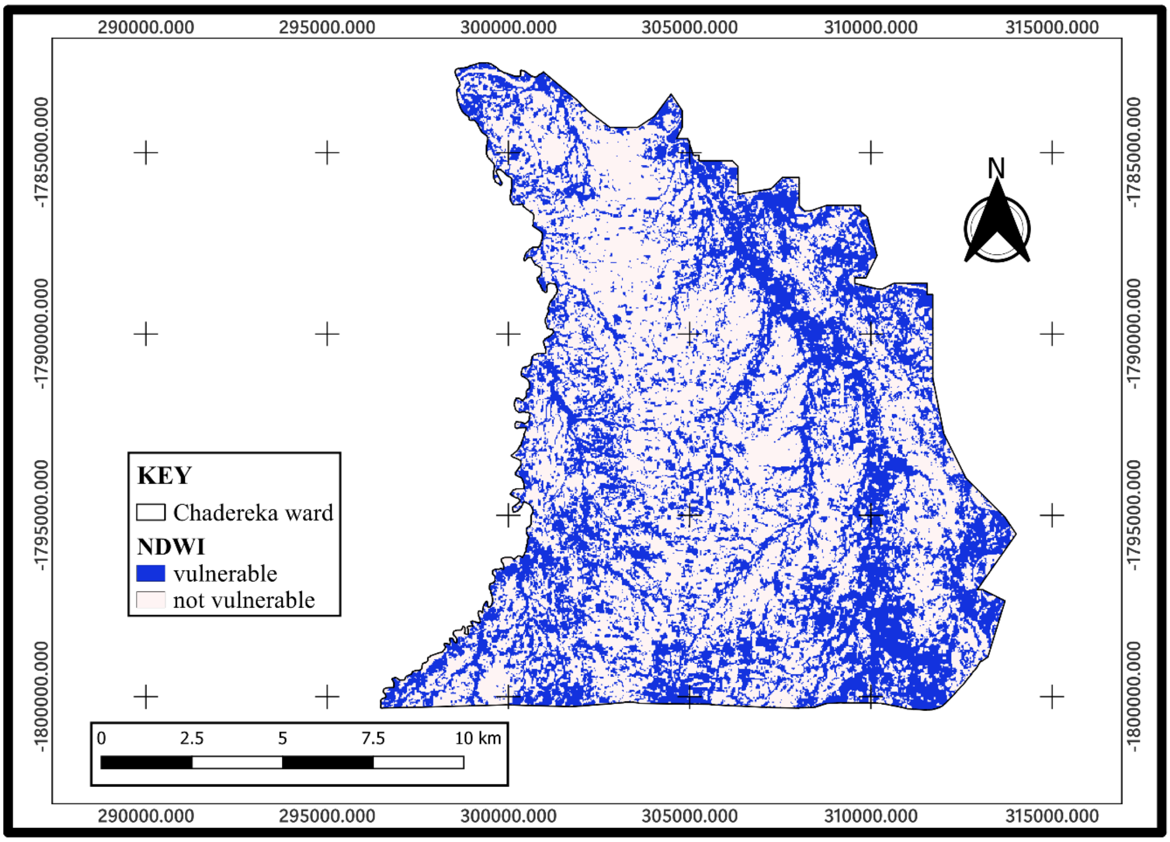


Figure 4.2: The NDWI-derived flood extent map of Chadereka Ward 1 (Source: author)

Based on the NDWI flood extent map analysis (Figure 4.2), the areas classified as not susceptible to flooding appear to be slightly larger than the areas that are susceptible. Although the difference in the proportion of these areas is not substantial, it indicates a slight predominance of non-susceptible regions over susceptible ones.

4.3.1 Classification of Chadereka Ward Land cover using Random Forest Algorithm

In order to establish a map that shows the susceptibility in Chadereka Ward, Landsat 8 OLI/TIRS satellite images were also visualised was utilized and transformed into a false-colour band combination using the 4-5-3 red-green-blue (RGB) This false-colour composition enhances the visualization of features, making water-covered pixels to be more distinguishable.

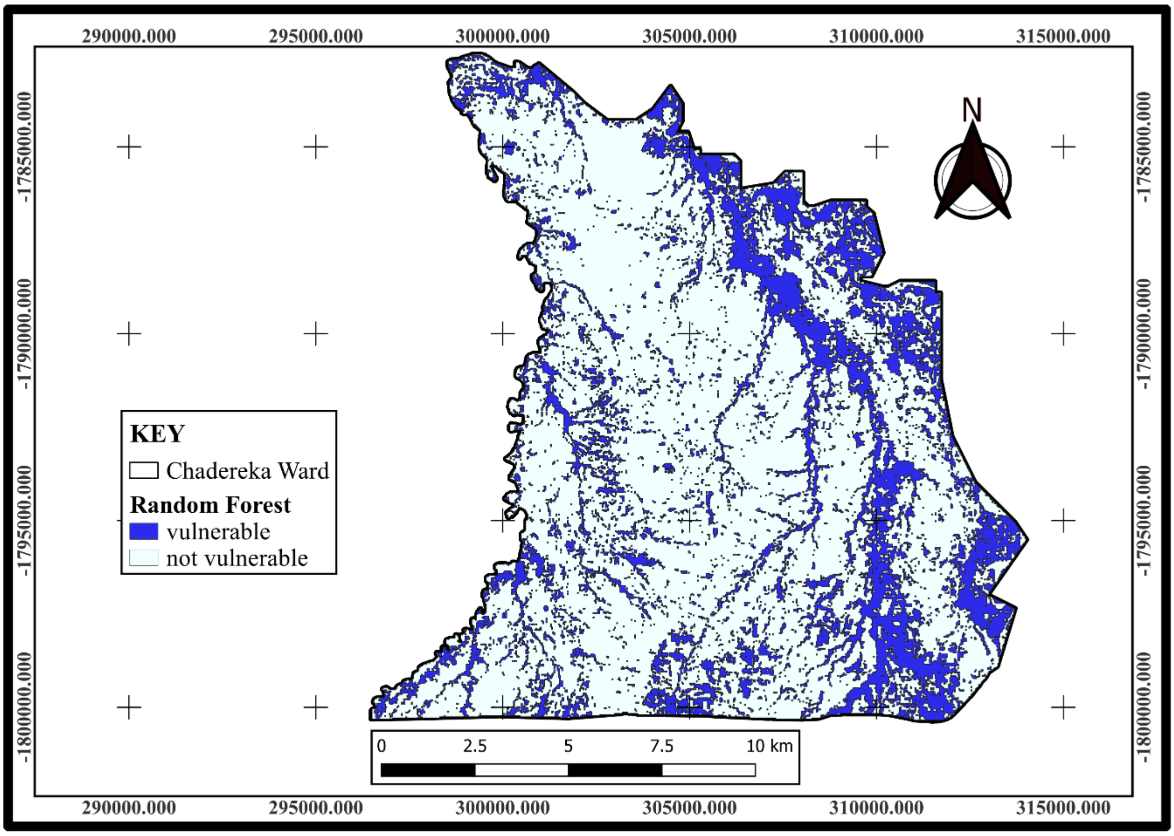


Figure 4.3 Map of flood extent of Chadereka Ward produced using Random Forest algorithm (Source: author)

As shown in Fig 4,3, the classification resulted in the categorization of the study area into two distinct classes: susceptible and not susceptible. This classification enabled the identification and mapping of areas prone to flooding (susceptible) and those less vulnerable to inundation (not susceptible).

4.3.2 Accuracy Assessment Classification

The post-classification area-based error matrix accuracy results for 2024 were used to measure accuracy assessment of the RF model that was used to classify susceptibility in Chadereka ward. Table 4.1 shows the area-based error matrix for Chadereka ward.

Table 4.1 Overall Accuracy

V_Classified	Vulnerable	Not vulnerable	Area
Vulnerable	0.427	0.0225	36000
Not vulnerable	0	0.5506	44100
Total	0.427	0.573	80100
Estimated Area	34200	45900	80100
SE	0.0157	0.0157	
SE area	1256.37	1256.37	
95% CI area	2462.48	2462.48	
PA [%]	100	96.0784	
UA [%]	95	100	

Area unit = Metre^2	SE = Standard error	CI = confidence interval
PA = producer accuracy	AU = user accuracy	

Overall accuracy [%] =97.7528

By selecting the classification to assess and referencing it against the reference raster, the overall accuracy was 97.75%. This high accuracy was attributed to several factors such as the quality of the training data used to build the classification model, the ability of the Random Forest algorithm in handling complex spatial patterns, and the pre-processing of input data to enhance classification performance.

4.4 Flood vulnerability in Chadereka

The flood susceptibility map of Ward 1 Chadereka was developed by the integration of flood-prone factors in QGIS (3.36).

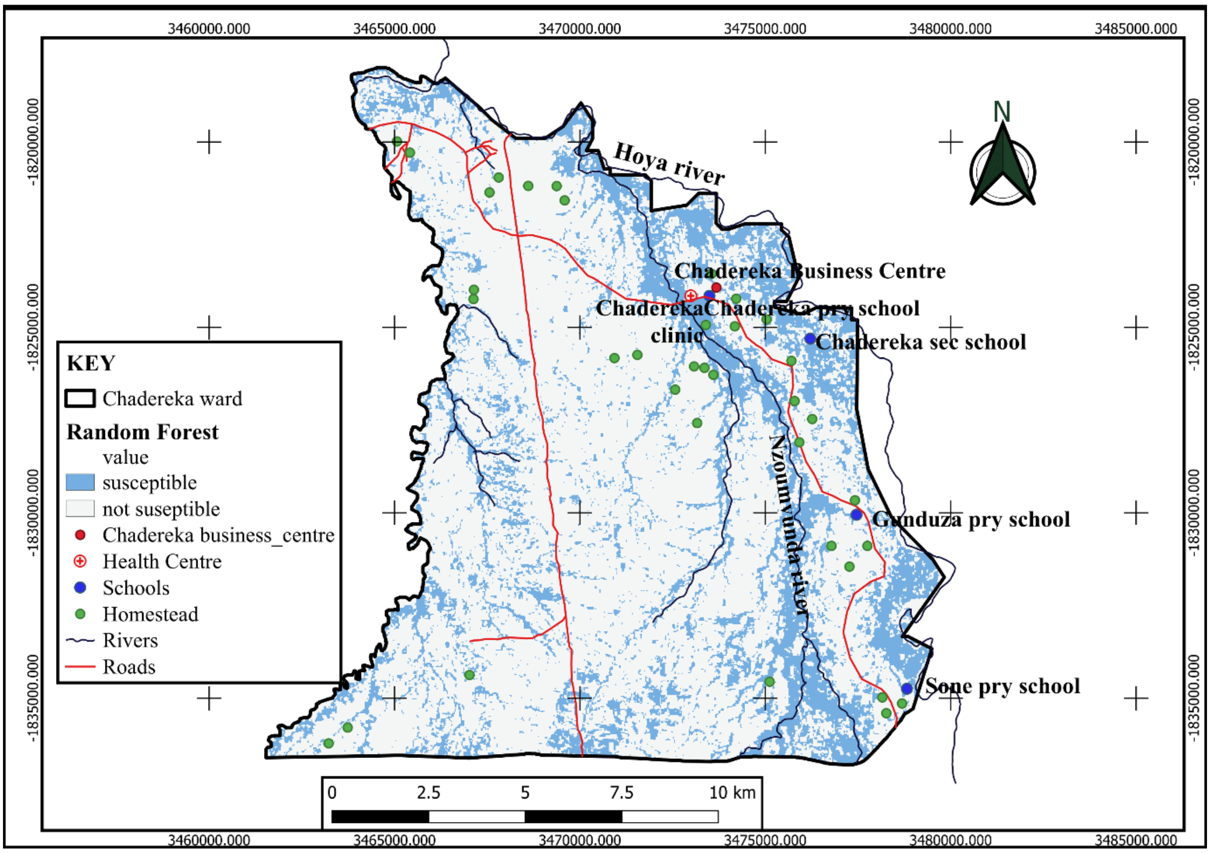


Figure 4.4: Map showing flood susceptibility for Chadereka Ward 1(Source: author)

For determining the flood extent, the study involved overlaying the Random Forest map with the distribution of infrastructure as shown in Figure 4.4. Incorporating these data layers revealed that the majority of the infrastructure lies between the Nzoumvunda and Hoya rivers. This method effectively demonstrated the extent of vulnerability by identifying which infrastructure is at a higher risk of flooding. The Random Forest map's representation, coupled with the physical distribution of infrastructure, provides an analysis of potential flood consequences, allowing for improved disaster management and mitigation planning.

4.5 Discussion

The objectives of the study were effectively addressed through the modelling of physical infrastructure distribution, flood extent using Random Forest machine learning, and the determination of flood vulnerability in Chadereka Ward to reveal the flood-vulnerable areas with a view to pave way for further disaster risk intervention initiatives. This section discusses the findings in relation to the study's objectives.

The study shows that most of the infrastructure is located between the Nzoumvunda and Hoya rivers. The spatial distribution of physical infrastructure provided valuable insights into the built environment of the study area, essential for understanding the potential impact of floods on human settlements and infrastructure. Figure 4.1, shows physical infrastructure highlighting distribution of schools, clinic, shopping centres and homesteads in close proximity to rivers. Similarly, another study conducted by Haque and Rahman (2016) in flood-prone Brahmaputra basin in Bangladesh explored the central role of physical infrastructure in flood susceptibility mapping for riverine communities. It concluded that while natural factors such as rainfall intensity and topography play a significant role in flood susceptibility, the study highlights the equally crucial contribution of physical infrastructure in shaping the vulnerability of riverine communities to inundation events (Haque and Rahman, 2016). Therefore, through spatial analysis of physical infrastructure, it is possible to identify critical infrastructure at risk of inundation during flood events as shown in this study.

Furthermore, the application of Random Forest machine learning algorithms facilitated the modelling of flood extent and vulnerability, revealing that the area is highly susceptible to floods due to its environmental factors. This study conducted Random Forest in SCP and overall accuracy was 97.75% as indicated in Table 4.1. The overall accuracy was the result of leveraging high-quality training data for model development, the Random Forest algorithm's adeptness at addressing complex spatial configurations. Another study was undertaken in Chadereka, with the goal of demonstrating flood extent, although it took a different technique. The study titled "A participatory approach in GIS data collection for flood risk management, Muzarabani district, Zimbabwe" was undertaken in Chadereka ward. The study investigated the spatial extent of the flood zone that occurred during the December 2007 flash flood in the Chadereka area, and the flood extend assessment map showed that the area lies mainly between Hoya and Nzoumvunda (Chingombe et.al., 2015). These findings align with previous research indicating the vulnerability of other regions similar to Chadereka Ward in terms of flooding. In a similar study conducted on flood susceptibility mapping in Frederiction, Canada by Esfandiari et al. (2018) using Random Forest algorithm achieved a high accuracy score (97.57%). The high accuracy was attributed to the selection of key variables, the robustness of the Random Forest algorithm, high-quality data, effective pre-processing, model optimization, and the use of robust validation techniques (Esfandiari et al., 2018). Therefore, these findings confirm the effectiveness of the Random Forest algorithm in mapping flood susceptibility, which provides valuable insights for decision-making processes for policy makers practitioners in flood risk management and mitigation. Although conducted in a different region, this study serves as a reference for the application of the Random Forest algorithm in flood susceptibility mapping studies.

Furthermore, the study finds that majority of the infrastructure lay in between Hoya and Nzoumvunda rivers. The spatial overlay of the physical infrastructure on a Random Forest produced flood extent map defined vulnerability and identified the infrastructure at risk of flooding. For instance, a study conducted in southwestern Iran utilized RF alongside GIS to create detailed flood hazard maps, identifying infrastructure at high risk of flooding (Rahmati et al., 2019). Similarly, research in urban areas highlighted the application of RF in evaluating the spatial distribution of critical infrastructure, such as roads and buildings, to produce accurate flood risk maps (Tehrany et.al., 2014). In Pakistan, RF and remote sensing data were integrated to assess infrastructure risk, demonstrating the method's capability to produce high-resolution flood risk maps (Khan et al., 2014). Furthermore, studies in Egypt's Wadi Qena Basin and in other regions have shown the effectiveness of combining RF with spatial analysis to highlight areas of high flood risk, supporting better disaster management and mitigation strategies (Youssef, et al., 2016). These studies highlight the valuable role of RF in flood risk assessment and infrastructure vulnerability mapping.

4.6 Chapter Summary

This chapter presented results of flood susceptibility mapping using Random forest machine learning algorithm implemented in QGIS software using the SCP tool. The main results highlighted the utility that Random Forest algorithms in delineating flood susceptibility within Chadereka Ward, Muzarabani District, Zimbabwe. Moreover, the physical distribution of the physical infrastructure overlaying with Random forest map showed physical infrastructures that are vulnerable to flooding. A discussion was then proffered to link this research with other similar studies elsewhere.

CHAPTER 5: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1 Introduction

This chapter focuses on major findings, conclusions, and recommendation based on the study's objectives. Blending of significant findings, conclusion, and recommendations is a crucial part of a study since it provides a summary of the study's primary conclusions and aftermaths.

5.2 Summary of key findings

5.2.1 Distribution of physical infrastructure across the floodplain

This study focused on mapping the distribution of physical infrastructure in Chadereka Ward. The major findings of the study demonstrated that the substantial proportion of the settlement are located along the river's banks. Several variables may influence settlement closeness to riverbanks, including historical settlement patterns, in which people built along water sources for access to water for drinking, irrigation and other livelihood activities.

5.2.2 Modelling flood extent in Chadereka using Random Forest

To establish flood susceptibility maps for Chadereka Ward, the NDWI was employed to differentiate between more and less susceptible areas. The NDWI results indicated that regions in close proximity to rivers exhibited higher susceptibility to flooding. Complementing this analysis, a Random Forest algorithm was applied, which corroborated the findings from the NDWI by also identifying these areas as highly susceptible. The model's performance was evaluated through an overall accuracy assessment, yielding a result of 97.75%. This outcome demonstrates the capability of the Random Forest algorithm to effectively handle complex spatial patterns and accurately predict flood-prone areas. Such integrated approaches combining NDWI and advanced machine learning techniques (RF) ensure robust and reliable flood susceptibility mapping.

5.2.3 Determining flood susceptibility in Chadereka Ward

The study used physical infrastructure data and a Random Forest (RF) map to display and analyse the susceptibility of infrastructure to flooding. By overlaying infrastructure such as schools, clinics, and homesteads on the RF map, the study was able to generate detailed vulnerability maps that illustrate the relative risk levels of these assets. This method enabled identification of high-risk locations, allowing for targeted interventions. Furthermore, the susceptibility maps gave critical insights for development planners and policymakers, allowing for the development of more effective flood mitigation techniques and disaster preparedness plans.

5.3 Conclusion

Based on the study findings, it was determined that employing Random Forest for classifying flood susceptibility proved to be highly effective. The Random Forest algorithm demonstrated robust performance in delineating areas prone to flooding, leveraging its ability to handle complex spatial patterns and diverse input variables. Additionally, the study concluded that integrating physical infrastructure data with yielded a comprehensive map illustrating the spatial distribution of infrastructure within the study area. This integrated approach provided insights into the location and density of critical infrastructure elements such as roads, buildings, and utilities

5.4 Recommendations

Basing on the findings of the study, it was therefore, recommended that;

- There is need to put infrastructure resilience measures to reduce the vulnerability of critical infrastructure to flood damage.
- There is need for continuous model refinements and updating the Random Forest model used for flood susceptibility mapping by incorporating new data and refining model parameters.
- Use transfer learning approaches to adapt machine learning models developed in data-rich areas to data-scarce areas in flood-prone areas.

REFERENCES

Alexander, K., Hettiarachchi, S., Ou, Y. and Sharma, A. (2019). Can integrated green spaces and storage facilities absorb the increased risk of flooding due to climate change in developed urban environments? Journal of Hydrology. DOI: 10.1016/j.jhydrol.2019.124201

Amir, M., Pinar, O.,1 and Kwok-wing, C., (2018). Flood prediction using machine learning models: Literature review. Water. https://doi.org/10.3390/w10111536

Bui, D.T., Hoang, N., Martínez Álvarez, F., Ngo, P., Hoa, P., Pham, T. and Costache R. (2020) A novel deep learning neural network approach for predicting flash flood susceptibility: A case study at a high frequency tropical storm area. https://doi.org/10.1016/j.scitotenv.2019.134413

Bui, D.T., Tein Bui, D., Nguyen, Q.P., and Bashir, M. (2020). Flood susceptibility mapping using machine learning algorithms. Science of the Total Environment, 729, 138947. Doi: 10.1016/j.scitotenv.2020.138947

Centre for Research on the Epidemiology of Disasters (CRED). (2018). EM-DAT (The Emergency Events Database). Country profile database—Zimbabwe.

Chakraborty, A., and Kumar, B. (2024). Mapping Flood Susceptibility: Using Random Forest and MLP Classifier to Predict Vulnerable Areas. Medium.

Chingombe, W., Pedzisai, E., Manatsa, D., Mukwada, G., and Taru, P. (2015). A participatory approach in GIS data collection for flood risk management, Muzarabani district, Zimbabwe. *Arabian Journal of Geosciences*, *8*, 1029-1040.

CRED. (2019). Natural disasters in 2019. EMDAT

Darabi, H., Haghighi, A. T., Mohamadi, M. A., Rashidpour, M., Ziegler, A. D., Hekmatzadeh, A. A., and Kløve, B. (2020). Urban flood risk mapping using data-driven geospatial techniques for a flood-prone case area in Iran. *Hydrology Research*, *51*(1), 127-142.

de Bruijn, K. M., Maran, C., Zygnerski, M., Jurado, J., Burzel, A., Jeuken, C., and Obeysekera, J. (2019). Flood resilience of critical infrastructure: Approach and method applied to Fort Lauderdale, Florida. *Water*, *11*(3), 517.

Department of Civil Protection (DCP) (2007). Situation Report on Chadereka Flooding Incident.

Dottori F, Kalas M, Salamon P, Bianchi A, Thielen Del Pozo J, and Feyen L. (2015)A near real-time procedure for flood hazard mapping and risk assessment in Europe. Proceedings of 36th IAHR World Congress; International Association for Hydro-Environment Engineering and Research (IAHR), (Organiser). IAHR; 2015. JRC93988

Esfandiari, A., Rahmati, O., Eslamian, S., and Kazemi, M. (2018). Flood Mapping Using Random Forest and Identifying the Essential Conditioning Factors; A Case Study in Fredericton, New Brunswick, Canada. Journal of Flood Risk Management, 11(3), 879-891. https://doi.org/10.1111/jfr3.12345

EXCIMAP. (2007) Atlas of Flood Maps. Examples from 19 European countries, USA and Japan. European exchange circle on flood mapping (EXCIMAP), European Commission. ec.europa.eu/environment/water/flood risk/flood atlas

Gerasimos, A., Ioannis-Omiros, K., Marios B., Anastasia, M., Ilias G., Anastasios, K., Francesca, L., Roberto, F., Daniele, N., Michele, F., Andreas, S., Stefanos, V. and Ioannis, K. (2022). Flood Hazard and Risk Mapping by Applying an Explainable Machine Learning Framework Using Satellite Imagery and GIS Data. Sustainability.

Goudie, A. S. (2018). *Human impact on the natural environment*. John Wiley and Sons.

Haque, C. E., and Rahman, S. M. (2016). Role of physical infrastructure in flood susceptibility mapping for riverine communities: A case study in the Brahmaputra basin, Bangladesh. *Natural Hazards*, *83*(1), 409-427.

International Federation of Red Cross and Red Crescent Societies (2007) Zimbabwe: Muzarabani floods http://www.ifrc.org/docs/appeals/07/MDRZM002a.pdf

Jamshed, A., Birkmann, J., Feldmeyer, D., and Rana, I. A. (2020). A conceptual framework to understand the dynamics of rural–urban linkages for rural flood vulnerability. *Sustainability*, *12*(7), 2894.

Kasimba, R.osemary. (2017). *Understanding the role of social capital in enhancing community resilience to natural disasters: A case study of Muzarabani District, Zimbabwe* (Doctoral dissertation, PhD thesis, Rhodes University, Grahamstown).

Khan, S. I., Hong, Y., Gourley, J. J., Khattak, M. U., and De Groeve, T. (2014). Multi-sensor imaging and space-ground cross-validation for 2010 flood along Indus River, Pakistan. *Remote Sensing*, *6*(3), 2393-2407.

Kolen, B. and Van Gelder, P. (2018) Risk-Based Decision-Making for Evacuation in Case of Imminent Threat of Flooding, Water.

Kovacs, Y., Doussin, N., Gaussens, M., Pacoud, C. L., and Afd, O. G. (2017). Flood risk and cities in developing countries. *French Development Agency: Paris, France*.

Kumar, L. and Mutanga O. (2018) Google earth engine applications since inception: usage, trends and potential. Remote sensing (Basel).

Lin, L., Di, L., Tang, J., Yu, E., Zang, C., Rahman, M.S., Shrestha, R. and Kang, L. (2019). Improvement and Validation of NASA/MODIS NRT Global Flood Mapping. Remote Sens. 2019, 11(2), 205; https://doi.org/10.3390/rs11020205

Lumbroso, D., Ramsbottom, D., and Spaliveiro, M. (2008). Sustainable flood risk management strategies to reduce rural communities' vulnerability to flooding in Mozambique. Journal of Flood Risk Management. Journal of Flood Risk Management 1(1):34 – 42 DOI:10.1111/j.1753-318X.2008.00005.x

Madamombe, E. K. (2004). Flood management practices- Selected flood prone areas Zambezi Basin. Integrated flood management. Zimbabwe.

Malakeel, G. S., Abdu Rahiman, K. U., and Vishnudas, S. (2021). Flood Risk Assessment Methods—A Review. *Current Trends in Civil Engineering: Select Proceedings of ICRACE 2020*, 197-208.

Masters, J. (2019). Africa's Hurricane Katrina: Tropical Cyclone Idai causes an extreme catastrophe. Weather Underground. Journal of Flood Risk Management 13(L22402)

Mavhura E. (2019) Systems Analysis of Vulnerability to Hydrometeorological Threats: An Exploratory Study of Vulnerability Drivers in Northern Zimbabwe. International Journal of Disaster Risk Science International Journal of Disaster Risk Science 10(10) DOI:10.1007/s13753-019-0217-x

Mavhura E. (2017) Applying a systems-thinking approach to community resilience analysis using rural livelihoods: The case of Muzarabani district, Zimbabwe. International Journal of Disaster Risk Reduction 25, 248-258.

Mavhura, E., Manyena, B., and A. E Collins, (2017) An Approach for Measuring Social Vulnerability in Context: The Case of Flood Hazards in Muzarabani District, Zimbabwe," Geoforum 86:103-117 DOI:10.1016/j.geoforum.2017.09.008

McCombes, S. (2021). What is a research design. Types, guides and examples. Scribbs.

Mucherera, B., and Mavhura, E. (2020) Flood survivors' perspectives on vulnerability reduction to floods in Mbire district, Zimbabwe. Jàmbá Journal of Disaster Risk Studies, 12(1), a663. Available from: https://doi.org/10.4102/jamba.v12i1.663

Mudavanhu C. (2014). The impact of flood disasters on child education in Muzarabani District, Zimbabwe. Jamba: Journal of disaster risk studies. Education, Environmental Science, Sociology Jàmbá: Journal of Disaster Risk Studies 10.4102/JAMBA.V6I1.138

Mugandani, R.; Wuta, M.; Makarau, A. and Chipindu, B. (2012). Re-classification of agro-ecological regions of Zimbabwe in conformity with climate variability and change. African Crop Science Journal.

Nicholas, A., Chan, H. and Baker, B. (2019). Machine learning applications of artificial intelligence to imaging and diagnosis. Biophysical Reviews.

Noureen, A., Aziz, R., Ismail, A., and Trzcinski, A. P. (2022). The impact of climate change on waterborne diseases in Pakistan. *Sustainability and Climate Change*, *15*(2), 138-152.

Quesada-Román, A.; Ballesteros-Cánovas, J.A.; Granados-Bolaños, S.; Birkel, C.;and Stoffel, M. (2020). Dendrogeomorphic reconstruction of floods in a dynamic tropical river.

Sipper, M. and Moore, J. (2021). Conservation machine learning. A case study of random forests. Scientific Reports.

Smet, H. De, Schreurs, B. and Leysen, J. (2015). The Response Phase of the Disaster Journal of Homeland Security and Emergency Management 12(2): DOI:10.1515/jhsem-2015-0005

Smith, J. (2019). Ethical Considerations in Quantitative Research. Journal of Quantitative Studies, 10(2), 45-58.

Teddlie, C., and Yu, F. (2007). Mixed methods sampling: A typology with examples. Journal of mixed methods research, 1(1), 77-100. https://doi.org/10.1177/2345678906292430

Tehrany, M. S., Pradhan, B., and Jebur, M. N. (2014). Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS. Journal of Hydrology, 512, 332-343. https://doi.org/10.1016/j.jhydrol.2014.03.008

UNISDR (2009). Terminology on Disaster Risk Reduction.

United Nations International Strategy for Disaster Reduction (UNISDR) (2015) https:// www.undrr.org/ publication/ sendai- framework- disaster – risk- reduction- 2015 – 2030

Vojtek, M., and Vojteková, J. (2019). Flood susceptibility mapping on a national scale in Slovakia using the analytical hierarchy process. *Water*, *11*(2), 364.

Wagenaar, D., Curran, A., Balbi, M., Bhardwaj, A., Soden, R., Hartato, E. and Lallemant, D. (2020). Invited perspectives: How machine learning will change flood risk and impact assessment. *Natural hazards and earth system sciences*, *20*(4), 1149-1161.

Wubalem, A. (2022). Landslide inventory, susceptibility, hazard and risk mapping. *Landslides*, 164-181.

Youssef, A. M., Pradhan, B., and Sefry, S. A. (2016). Flash flood susceptibility assessment in Jeddah city (Kingdom of Saudi Arabia) using bivariate and multivariate statistical models. *Environmental Earth Sciences*, *75*(1), 12.

Yu, M., Yang, C., and Li, Y. (2018). Big data in natural disaster management: a review. *Geosciences*, *8*(5), 165.

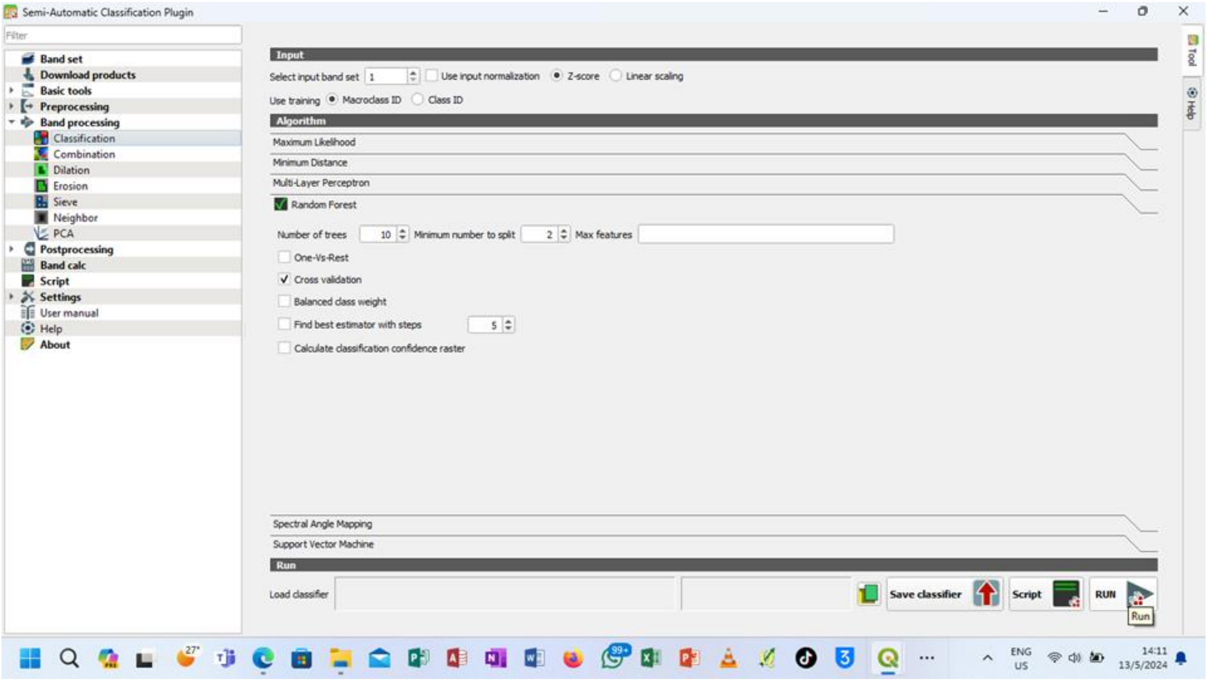
APPENDICES

Appendix 1: Fieldwork Checklist

Date of visitation: 01 to 14 February 2024

Observation	Description	Comments
Settlement pattern	▪ Distribution of houses ▪ Clustering of settlements ▪ Proximity to water bodies	
Infrastructure condition	• Condition of roads • State of bridges • Condition of buildings • Visible floodwater on buildings	
Vegetation and land cover	• Types of vegetation • Extent of forest cover	
Land formation	• Flood plains • Mountainous regions • Plateaus	
Erosion patterns	• Areas of significant erosion • Indicators of water flow patterns • Vulnerable points in the landscape	
River distribution and flow	• Location of rivers and streams • Direction of water flow	

Appendix 2: Random Forest Classification Tab in QGIS



Appendix 3: Overall Classification Accuracy assessment

Raster Value	Reference	Classification	Pixel Sum	Area [metre^2]
1	1	1	38	34200
2	2	1	2	1800
3	2	2	49	44100
ERROR MATRIX [pixel count]				
	> Reference			
V_Classified	1	2	Total	
1	38	2	40	
2	0	49	49	
Total	38	51	89	
AREA BASED ERROR MATRIX				
	> Reference			
V_Classified	1	2	Area	Wi
1	0.427	0.0225	36000	0.4494
2	0	0.5506	44100	0.5506
Total	0.427	0.573	80100	1
Estimated area	34200	45900	80100	
SE	0.0157	0.0157		
SE area	1256.37	1256.37		
95% CI area	2462.48	2462.48		
PA [%]	100	96.0784		
UA [%]	95	100		
Overall accuracy [%] = 97.7528				

Irene Argie Tawonezvi B201705B Project.docx

ORIGINALITY REPORT

5%	4%	6%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.mdpi.com Internet Source	2%
2	www.researchgate.net Internet Source	1%
3	Submitted to Bindura University of Science Education Student Paper	1%
4	W. Chingombe, E. Pedzisai, D. Manatsa, G. Mukwada, P. Taru. "A participatory approach in GIS data collection for flood risk management, Muzarabani district, Zimbabwe", Arabian Journal of Geosciences, 2014 Publication	1%
5	"Climate Crisis, Social Responses and Sustainability", Springer Science and Business Media LLC, 2024 Publication	1%