




Article

Integrating Machine Learning and Geospatial Data for Mapping Socioeconomic Vulnerability to Urban Natural Hazard

Esaie Dufitimana ^{1,2,*} , Paterne Gahungu ², Ernest Uwayezu ³ , Emmy Mugisha ⁴ and Jean Pierre Bizimana ³ 

¹ School of Information Communication Technology, College of Science and Technology, University of Rwanda, Kigali P.O. Box 3900, Rwanda

² African Institute for Mathematical Sciences (AIMS) Research and Innovation Centre, Kigali P.O. Box 6428, Rwanda

³ Department of Spatial Planning, School of Architecture and Built Environment, College of Science and Technology, University of Rwanda, Kigali P.O. Box 3900, Rwanda

⁴ African Center of Excellence in Internet of Things (ACEIoT), College of Science and Technology, University of Rwanda, Kigali P.O. Box 3900, Rwanda

* Correspondence: edufitimana@aimsric.org

Abstract: Rapid urbanization and climate change are increasing the risks associated with natural hazards, especially in cities where socio-economic disparities are significant. Current hazard risk assessment frameworks fail to consider socio-economic factors, which limits their ability to effectively address vulnerabilities at the community level. This study introduces a machine learning framework designed to assess flood susceptibility and socio-economic vulnerability, particularly in urban areas with limited data. Using Kigali, Rwanda, as a case study, we quantified socio-economic vulnerability through a composite index that includes indicators of sensitivity and adaptive capacity. We utilized a variety of data sources, such as demographic, environmental, and remotely sensing datasets, applying machine learning algorithms like Multilayer Perceptron (MLP), Random Forest, Support Vector Machine (SVM), and XGBoost. Among these, MLP achieved the best predictive performance, with an AUC score of 0.902 and an F1-score of 0.86. The findings indicate spatial differences in socio-economic vulnerability, with central and southern Kigali showing greater vulnerability due to a mix of socio-economic challenges and high flood risk. The vulnerability maps created were validated against historical flood records, socio-economic research, and expert insights, confirming their accuracy and relevance for urban risk assessment. Additionally, we tested the framework's scalability and adaptability in Kampala, Uganda, and Dar es Salaam, Tanzania, showing that making context-specific adjustments to the model improves its transferability. This study offers a solid, data-driven approach for combining assessments of flood susceptibility and socio-economic vulnerability, filling important gaps in urban resilience planning. The results support the advancement of risk-informed decision-making, especially in areas with limited access to detailed socio-economic information.

Keywords: urban flooding; socio-economic vulnerability; flood susceptibility; machine learning



Academic Editor: Wolfgang Kainz

Received: 30 December 2024

Revised: 25 March 2025

Accepted: 4 April 2025

Published: 8 April 2025

Citation: Dufitimana, E.; Gahungu, P.; Uwayezu, E.; Mugisha, E.; Bizimana, J.P. Integrating Machine Learning and Geospatial Data for Mapping Socioeconomic Vulnerability to Urban Natural Hazard. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 161. <https://doi.org/10.3390/ijgi14040161>

Copyright: © 2025 by the authors.

Published by MDPI on behalf of the

International Society for

Photogrammetry and Remote Sensing.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the Creative Commons

Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Urbanization and climate change are significantly impacting human health, socio-economic stability, and sustainability. Urbanization, the process resulting from the increase in urban residents and expansion of the built-up areas, is frequently driven by economic opportunities and improved living standards, including access to employment, education, and healthcare [1]. However, many urban residents, particularly in the Global South, face

critical challenges, such as inadequate housing, contaminated water, insufficient sanitation, and poor waste management [2]. The rapid urbanization has outpaced the capacity of urban planning processes to effectively address these challenges, exacerbating risks faced by urban residents [3]. As natural hazards such as earthquakes, floods, and wildfires become more frequent due to climate change, the risk of these events escalating into disasters grows significantly, particularly for vulnerable communities [4]. A disaster occurs when a hazard interacts with exposed populations or systems, leading to widespread disruption that exceeds the ability of those affected to respond and recover using their own resources, necessitating outside assistance [5]. In this framework, risk arises from the relationship between hazard, exposure, and vulnerability, with exposure defining the presence of people and assets in areas at risk during specific times.

In this study, the primary focus is on social and economic vulnerability, as these dimensions are critical for understanding how different individuals or groups experience and cope with natural hazards. Social vulnerability is influenced by demographic characteristics and resource access, while economic vulnerability involves financial losses and recovery resources [6]. These dimensions help to understand the complex interactions between hazards and vulnerability, for informing strategies to reduce risk and enhance resilience in urban environments [4]. As cities expand, increasing population densities, unregulated urban development, and inadequate infrastructure place more people in harm's way, especially in informal settlements and marginalized communities [7–9]. Urban areas are particularly susceptible to hazards due to their geographic positioning and the high concentration of human activity, which can strain existing resources and emergency response systems [10]. The consequences of these hazards can lead to widespread loss of life, displacement, economic instability, and disruption of essential services, including healthcare, transportation, and utilities. Exposure to hazards refers to the extent to which urban residents, assets, and infrastructure are subjected to hazardous events [11].

Urban socio-economically deprived individuals and communities usually reside in areas prone to natural hazards, which significantly heightens their exposure to various risks [9]. This exposure is especially hazardous because these populations possess limited capacity for mitigation or adaptation, rendering them highly vulnerable to natural hazards [12]. The frequent occurrence of natural hazards poses a severe threat to public health, particularly in Global South areas with high socio-economic sensitivity and poor adaptive capacity [13]. Understanding socio-economic vulnerability associated with the climate change related hazards is crucial for developing effective strategies aimed at preventing risks and mitigating damages caused by natural hazards toward enhanced socio-economic conditions and public health outcomes [9,14]. However, this understanding is missing due to the high reliance on extensive data that is unavailable in many cities of the Global South. Consequently, the assessment of socio-economic vulnerability to natural hazards is frequently ignored despite its importance to identify how natural hazards affect the urban population [9,15].

Countries in the Global South lack the required data for assessing socio-economic vulnerability to natural hazards, which complicates their ability to implement effective disaster risk reduction strategies [16]. One of the primary issues is the lack of comprehensive and reliable data, which is critical for understanding the socio-economic conditions of vulnerable populations [17]. Many areas do not have systematic data collection mechanisms, leading to gaps in information regarding population demographics, income levels, health outcomes, and access to basic services [16]. This absence of data hinders the development of accurate vulnerability assessments and makes it difficult for policymakers to identify the most at-risk communities and tailor interventions accordingly [18]. Moreover, the quality of available data is questionable. In many cases, existing data are outdated or incomplete,

failing to capture the changing socio-economic conditions. For instance, socio-economic data of informal settlement dwellers are mostly not included in national statistics, leading to an underestimation of the risks faced by their residents [18]. In this regard, reliance on national-level data obscures local vulnerabilities, as small spatial scale disparities in socio-economic conditions and exposure to hazards are not adequately represented at large spatial scales. Additionally, socio-political factors further complicate the data collection efforts. In many Global South countries, political instability, corruption, and inadequate governance can impede effective data gathering and sharing [19]. Furthermore, the data challenge results from a lack of political will to prioritize socio-economic vulnerability in risk assessment and management or a lack of budget for data collection initiatives [19]. As natural hazards become more prevalent in urban areas, reliable data is needed to support risk assessments and adaptation strategies.

Recent advancements in technology offer potential solutions to data challenges. Remote sensing and machine learning techniques have shown a promising capacity to obtain and process large-scale datasets for hazard management. Several studies have applied these technologies to address various urban challenges such as flooding, landslides, and gullies development in urban areas. For instance, Refs. [9,20,21] have used remote sensing-based satellite imagery in combination with machine learning for measuring flood susceptibility. Refs. [22–25] applied machine learning to evaluate multi-hazards including flooding and landslides. Despite these advancements, most of the studies utilizing remote sensing and machine learning have primarily focused on hazard assessment rather than integrating socio-economic vulnerability into their risk frameworks. While few studies have attempted this integration [26–29], they developed approaches and frameworks tailored to specific input data sets and localized areas, which limits their broader applicability in both data-rich and data-scarce regions. This underscores the need for new approaches that can bridge the gap between hazard management and socio-economic vulnerability assessment.

Therefore, this study proposed a flexible and scalable framework for mapping socio-economic vulnerability to natural hazards in urban areas, specifically designed to support more effective and equitable public health interventions in data-scarce urban environments. The significance of this research lies in its potential to address gaps in understanding how socio-economic factors influence vulnerability to natural hazards, particularly in rapidly urbanizing regions of the Global South. By focusing on the City of Kigali, the proposed framework was applied to map flood susceptibility and socio-economic vulnerability to flooding. The obtained maps were locally validated through a combination of historical flood data comparison and comparative analysis with existing socio-economic studies. Validation also included adherence to established methodologies and qualitative assessments using local knowledge. In order to evaluate the scalability of the proposed framework and enable comparative analysis across several urban contexts, it was applied to the cities of Dar es Salaam, and Kampala, in Tanzania and Uganda, respectively. The results derived from this study contribute to a deeper understanding of socio-economic vulnerabilities in urban areas prone to natural hazards, ultimately informing policy decisions and resource allocation for targeted interventions for enhanced resilience and improved public health outcomes.

The rest of this paper is structured as follows: materials and methods for this study are presented in the next part, which presents a summary of the proposed framework and how it was utilized to map Kigali's socioeconomic susceptibility to floods. This is followed by a section on the scalability and transferability of the proposed framework, which highlights its relevance in different urban settings. The results and discussion section present key findings and their implications, and limitations of the study. Finally, a concluding section summarizes the results and emphasizes the importance of the study and

integration of socio-economic vulnerability assessments into urban planning and disaster risk management strategies.

2. Materials and Methods

2.1. Description of the Proposed Framework

The proposed framework is a result of a review of recent research on hazard risk modeling and mapping and socio-economic vulnerability assessment [9,15,21,25,28–33]. As shown in Figure 1, the framework is composed of three primary components. The first component, which is shown in blue on the left side of Figure 1, consists of estimating hazard susceptibility by combining machine learning models with data from remote sensing. This component enables users to model susceptibility for one or multiple hazards by leveraging data that is readily available for the area of interest. The flexibility inherent in this component allows practitioners to select the most suitable machine learning algorithms depending on the specific characteristics of geographical area, the hazards being modeled, and available data. Various machine learning methods, including Random Forest (RF), Gradient-Boosted Decision Trees (XGBoost), Support Vector Machines (SVM), and Artificial Neural Networks (ANN), have demonstrated good performance in hazard susceptibility assessments in various studies [20,21,34–36].

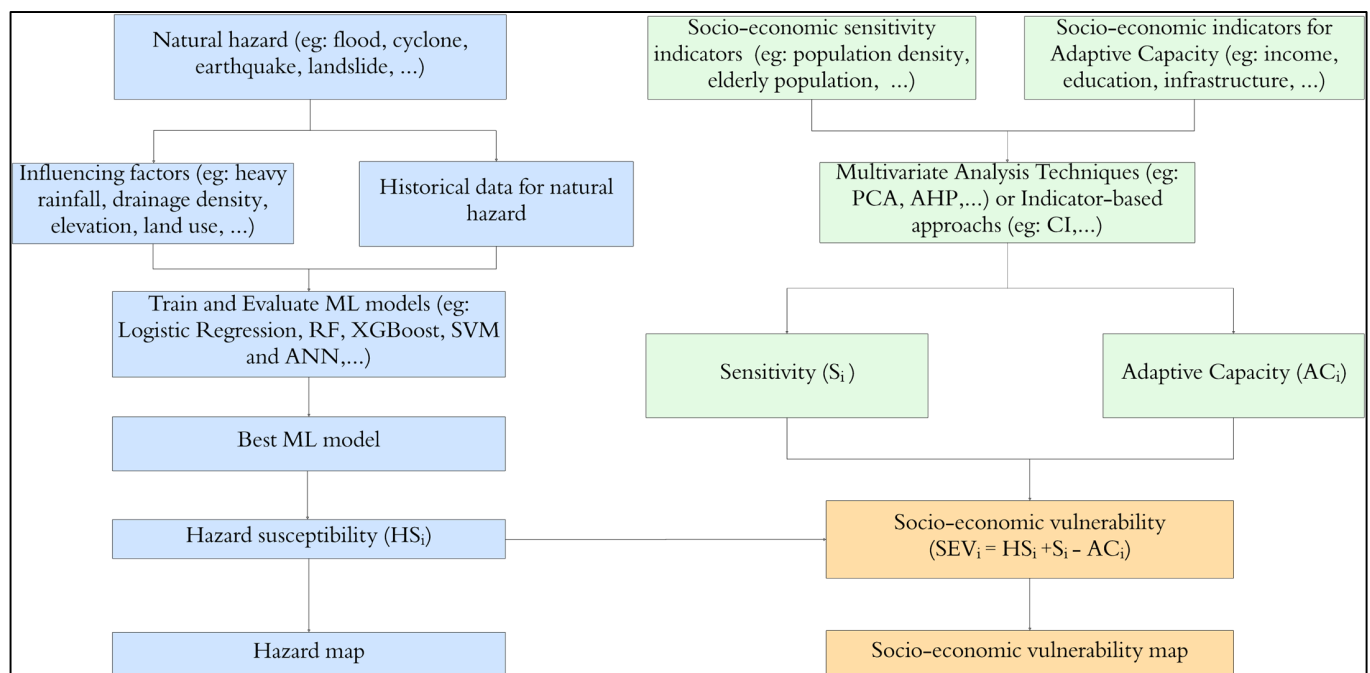


Figure 1. Proposed framework for mapping urban socio-economic vulnerability to natural hazards.

The second component, presented in green on the Figure 1's right side, involves analyzing socio-economic data to evaluate socio-economic vulnerability. This component emphasizes the use of diverse socio-economic indicators and multivariate analysis techniques, such as regression models, Principal Component Analysis (PCA), Analytic Hierarchy Process (AHP), and Composite Indicator (CI) approaches [14,28,32]. These methodologies enable to quantify sensitivity and adaptive capacity effectively. The final component, represented at the bottom in orange, creates a socio-economic vulnerability index by combining hazard susceptibility with quantified sensitivity and adaptive capacity. This combination is crucial for assessing and mapping socio-economic vulnerability associated with natural hazards, as demonstrated by various studies [9,26,37,38]. Therefore,

the proposed framework would support the development of effective hazard management strategies and inform policy decisions aimed at enhancing community resilience.

The flexibility and scalability of the proposed framework stem from its ability to integrate multiple data sources, making it applicable to both data-rich and data-scarce regions. Unlike previous studies that develop frameworks tailored to specific datasets and localized contexts, the proposed framework leverages machine learning models that can generalize across different geographic areas by adapting to varying data availability. For example, while some frameworks rely exclusively on high-resolution satellite imagery or extensive ground-truth data, our framework can also incorporate open-access datasets, such as Sentinel-2 imagery and crowd-sourced geospatial data, ensuring broader applicability. This adaptability enhances its scalability, allowing it to be implemented for different hazards in diverse urban environments with varying data infrastructures.

2.2. Application of the Proposed Framework to Mapping Socio-Economic Vulnerability to Flooding in the City of Kigali

The case study of the City of Kigali, Rwanda, where flooding is a frequent natural hazard, was used to evaluate the proposed framework. The following sections present the case study area, a description of the data used and their sources, historical flooding data, data about factors influencing floods, and socio-economic data. These are followed by the estimation of flood susceptibility by training and testing various machine learning models. The best-performing model was used to obtain the flood susceptibility index. Additionally, the study applied Indicator-based approaches to compute sensitivity and adaptive capacity using socio-economic data. Details on each step are presented in the following sections.

2.2.1. Description of City of Kigali

The City of Kigali is the largest city of Rwanda, and serves as the capital and focal point of economic activities. In terms of administration, the City of Kigali is comprised of three districts (Figure 2), which are further subdivided into 35 sectors, which account for 161 cells and 1176 villages. The village represents the lowest and smallest administrative unit, which is referred to neighborhood in this study [39]. Located near the geographic center of the country, Kigali is a city that is exhibiting rapid urban growth and economic transformation. Spanning over 730 square kilometers, it is home to more than 1.7 million people and is pivotal to Rwanda's socio-economic landscape [39]. The city exhibits a diverse array of land-use types, including commercial, residential, industrial, agricultural, and public facilities, alongside wetlands and water bodies [40]. Over the past two decades, Kigali has undergone significant urban expansion and development. The city has seen a rapid increase in built-up areas whereby the urban landscape is a mosaic of modern high-rise buildings, residential neighborhoods, commercial zones, and informal settlements [41,42]. This is particularly true in the urban central core, where continuous development is observed through the presence of modern buildings and upgraded road networks. This rapid urbanization has outpaced the development of adequate infrastructure [43]. Consequently, majority of informal settlements are frequently found in the most vulnerable areas, which are highlighted by Kigali's urban fabric and include steep hillsides and flood-prone valleys [40]. These areas lack proper sanitation, drainage, and other basic services, exacerbating the vulnerability of their residents to natural hazards [44]. Despite significant economic progress, substantial socio-economic inequalities persist, with large segments of the population living in poverty [41,45]. Many residents, particularly those living in informal settlements, do not have access to essential services including sanitation facilities, clean water, and healthcare [44,46]. These socio-economic disparities mean that the poorest and most vulnerable populations are disproportionately affected by natural

hazards. They often reside in the most at-risk areas and have the least capacity to recover from adverse events, creating a cycle of vulnerability and poor health outcomes.

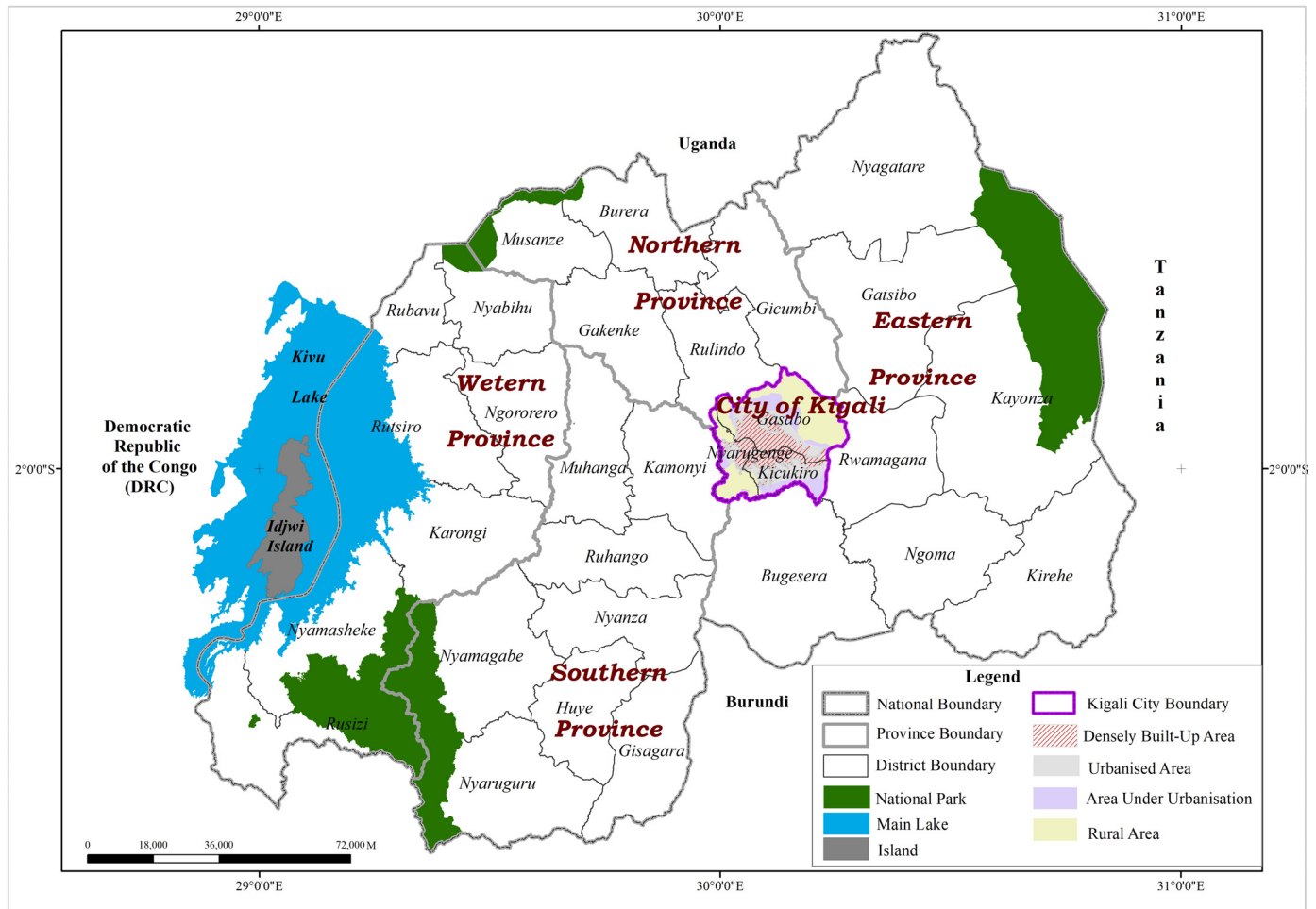


Figure 2. Location map of the City of Kigali in Rwanda. Data source: National Land Authority, National Institute of Statistics of Rwanda, and Ministry of Infrastructure.

Kigali's geography is characterized by its hills and valleys, with elevations ranging from approximately 1300 m to over 1600 m above sea level, which influences the city's drainage patterns [40]. The steep slopes accelerate water flows, resulting in a higher volume of water that accumulates in valleys, which, when combined with seasonal severe rainfalls, frequently leads to flooding [47]. Flooding in Kigali, like in other areas, is a complex natural hazard characterized by the overflow of water and excessive pluvial water beyond its normal limits, resulting from heavy and prolonged rainfall [48]. Flooding is particularly noticeable in urban areas due to rapid urban growth, which has reduced vegetation cover and increased impervious surfaces and runoff, and hence exacerbating the risk of flooding in places with lower elevation. Furthermore, the combination of increased surface runoff due to urbanization and inadequate drainage systems hinders groundwater recharge and leads to the accumulation of excess water during rain events [49].

The consequences of flooding in Kigali, like in other urban areas, are severe and multifaceted. Immediate impacts include loss of life and destruction of buildings, utilities, roads, bridges, and other infrastructure. Additionally, flooding can compromise water supply systems, increasing exposure to contaminated water and facilitating the spread of infectious diseases such as dengue, malaria, measles, meningitis, and typhoid [50].

Furthermore, flooding impacts individuals with chronic health conditions by damaging critical infrastructure and creating barriers to accessing essential health services [51,52].

2.2.2. Overview of Data

The study used a data-driven approach following the proposed framework to map socio-economic vulnerability to flooding for public health interventions in Kigali. In the City of Kigali, historical flood data has not been recorded as geospatial data, making it challenging to extract geospatial flood information directly from reports. To overcome this, data on previous floods was extracted using Synthetic Aperture Radar (SAR) imagery from Sentinel-1 using the same methods as studies for flood modeling [53,54]. Based on flood incidents reported by the Rwandan Ministry in Charge of Emergency Management, two SAR images were selected: one acquired prior to a period of heavy rainfall on 22 December 2019, and the other taken following a flood event on 25 December 2019. Image ratioing and Otsu's thresholding methods [55] were applied to detected and delineate floodwater. Image ratioing enhances flood mapping by highlighting water presence through spectral band division, while Otsu's thresholding automatically segments flooded areas by determining an optimal threshold that separates water from non-water regions [55]. Therefore, 456 flood points and 484 non-flood points were randomly generated (A figure illustrating their distribution is presented in Figure A1 in Appendix A). These were divided into 80% training and 20% testing datasets. Non-flooded points were assigned a value of 0, whereas the target class value points were given a value of 1. In addition, nine flood-influencing factors were identified after a review of various studies on flood susceptibility assessment and considering the geographic appearances of the study area. Table 1 describes the factors that influence floods and the associated data sources used in this study.

Table 1. Flood influencing factors.

Flood-Influencing Factor	Description	Data Source
Elevation	Lower elevation areas are more prone to water accumulation, which increases the likelihood of flooding, while higher elevations typically experience less flooding as water drains downhill [56].	Extracted from DEM (10 m resolution) obtained from the National Land Authority (NLA) of Rwanda.
Slope	Moderate slopes may lead to water accumulation, increasing flood risk, while steep slopes promote rapid runoff, potentially resulting in flash floods [56].	Extracted from DEM (10 m resolution) obtained from the National Land Authority (NLA) of Rwanda.
Aspect	Different aspects can influence vegetation growth and soil moisture levels, impacting flood dynamics; for example, south-facing slopes may dry out faster than north-facing ones [35,57–59].	Extracted from DEM (10 m resolution) obtained from the National Land Authority (NLA) of Rwanda.
Land cover	Land cover influences the flow and accumulation of water. For instance, vegetation is important in reducing water runoff and enhancing soil infiltration, which helps mitigate flooding [60]. In contrast, impervious surfaces and barren or open land exacerbate flooding by accelerating water runoff and decreasing water infiltration [61].	Data were obtained from land cover map of the City of Kigali
Normalized Difference Vegetation Index (NDVI)	High NDVI values indicate dense vegetation that can absorb and slow water movement and mitigate flooding effects; low NDVI values suggest sparse vegetation cover correlating with higher flood susceptibility [62].	Extracted from Sentinel-2 satellite image.

Table 1. Cont.

Flood-Influencing Factor	Description	Data Source
Normalized Difference Built-up Index (NDBI)	High NDBI values indicate extensive urban development with impermeable surfaces that exacerbate flooding by increasing surface runoff during heavy rains [63].	Extracted from Sentinel-2 satellite images.
Cumulative Rainfall	Excessive cumulative rainfall can overwhelm drainage systems, particularly in areas with low drainage density or poor soil permeability, leading to increased flooding risks [64].	Computed from Climate Hazards Group Infrared Precipitation with Station (CHIRPS) data.
Drainage Density	Low drainage density can hinder effective water channeling during floods, increasing the likelihood of flooding in those areas [65].	Computed from drainage networks data obtained from the City of Kigali.
Distance from drainage	Areas that are close to drainage systems, including rivers and streams, are more prone to experience flooding in the event that the drainage system is overloaded with water [62].	Computed based on drainage network data obtained from the City of Kigali. We considered a distance of 10 m from each river and stream based on Law n°48/2018 of 13 August 2018 on the environment in Rwanda [66].

The min-max approach was used to normalize all factors to the range [0, 1] after they were transformed to raster with a spatial resolution of 10 m (Equation (1)).

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

where X is the original value, X_{norm} is the normalized value, X_{min} is the dataset's minimal value, and X_{max} is its highest value. This provides data ranging from 0 as low value and 1 as maximum value for all factors except for land cover, whereby low values represent class water class followed by forest, green spaces, and agricultural land toward built-up and bare land classes. Normalization is used in data analysis as a critical preprocessing step because the study used various datasets containing variables that were measured on different scales. Therefore, by applying normalization, all variables in datasets were transformed to a common scale to ensure that no single variable disproportionately influences the results due to its scale [67]. In addition, a correlation analysis was used to detect multicollinearity among these factors (Figure A2 in the Appendix B illustrates the correlation matrix for all factors). The goal of this correlation analysis was to identify and remove highly correlated factors that could adversely affect model performance and interpretation, leading to unreliable predictions [68]. Multicollinearity was quantified by correlation coefficient, which provides insight into the relationships between input factors. A correlation coefficient greater than 0.7 typically indicates a strong correlation, suggesting potential multicollinearity issues. Thus, any factor exceeding a correlation coefficient of 0.7 was not included. Figure 3 presents factors that were included in susceptibility modeling. The results on the correlation analysis are presented in Appendix B.

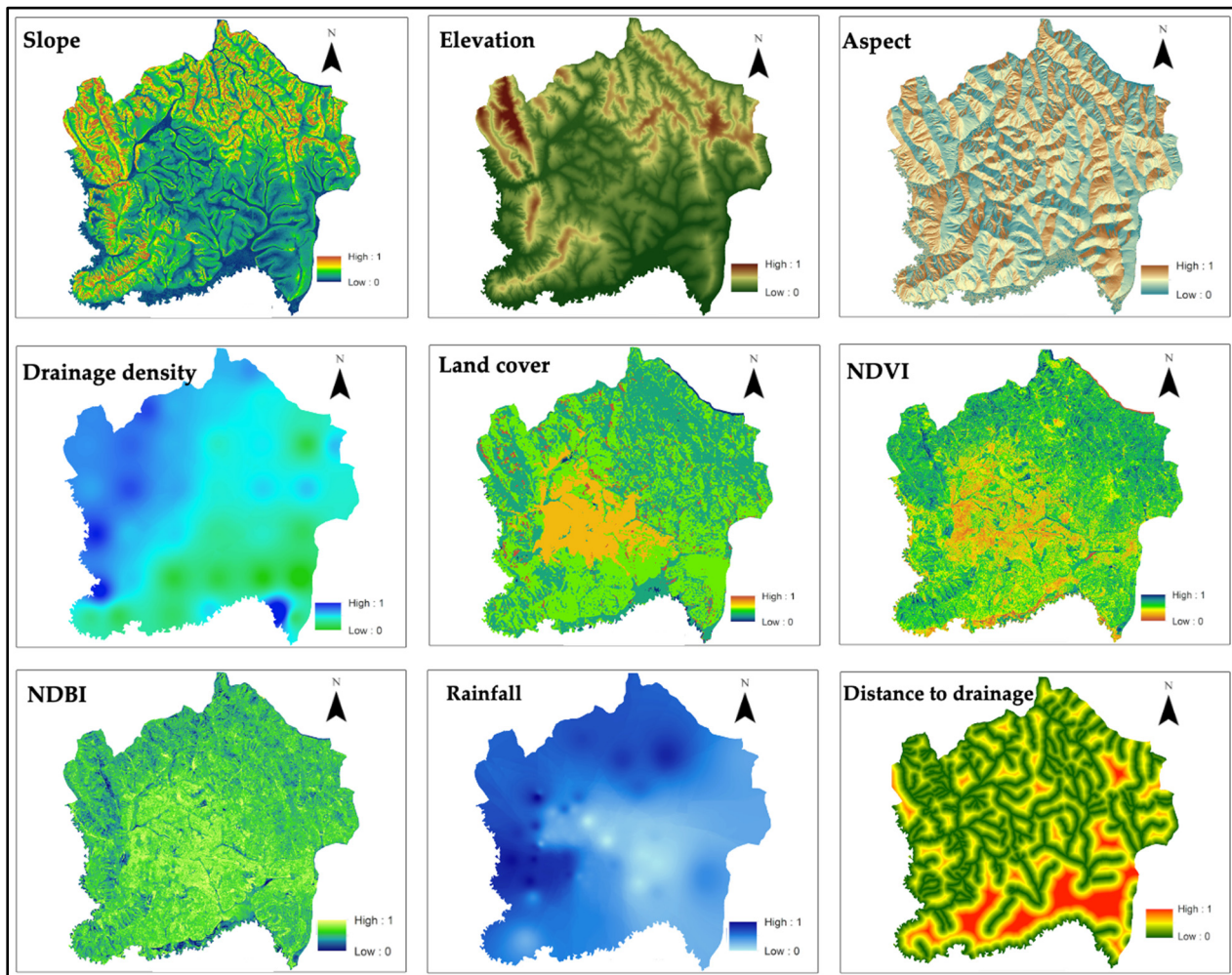


Figure 3. Visualization of flood influencing factors.

2.2.3. Flood Susceptibility Estimation with Machine Learning Models

Machine Learning models like Naïve Bayes (NB), K-Nearest Neighbors (KNN), Logistic Regression (LR), RF, SVM and XGBoost and ANN, have been widely used for analyzing and assessing natural hazards risks [22–25,29,69]. All models present advantages as well as disadvantages, and no single model is known to be the best generalized model for the assessment and analysis of hazard risks. Four machine learning models—RF, SVM, XGBoost, and Multilayer Perceptron (MLP)—were employed in this study. The choice of these models was guided by existing literature on modeling hazard risks, and which highlighted their ability to combine the raster images of environmental factors influencing flood occurrence [21]. RF and SVM were preferred for their robustness in handling high-dimensional data, making them suitable for complex assessments [22–25,29,69]. XGBoost offers superior performance in complex datasets but requires fine tuning, which involves adjusting model parameters to achieve optimal performance while avoiding overfitting [22,24]. Fine tuning involves iterative adjustments to model parameters to ensure that the model performs well on both training and unseen data [23]. The MLP learns complex patterns but demands significant computational resources and data. Additionally, these models were selected based on the datasets that were locally available for the study.

Accuracy, Precision, Recall, F1-Score, and the area under the receiver operating characteristic curve (AUC) were used to assess the performance of models. These metrics are widely used in machine learning to assess the performance of models, especially in classification tasks [21,22,29,34,68]. Each model was optimized using k-fold cross-validation and

hyperparameter tuning. Consequently, flood susceptibility indices were obtained using the model that performed the best based on the evaluation metrics on the testing dataset. The obtained indices were used to create a map of flood susceptibility index, whereby for comprehensive visualization, indices obtained were grouped in intervals for ease based on the natural breaks classifying method. In addition, flood susceptibility was aggregated at neighborhood-level. This aggregation facilitates understanding of susceptibility at small administrative level, which help the framework to provide more actionable information required for local authorities and stakeholders to formulate targeted interventions [70].

2.2.4. Mapping Socio-Economic Vulnerability to Flood

Mapping of socio-economic vulnerability to natural hazards consisted of the selection of socio-economic factors, data collection, data analysis/construction socio-economic vulnerability index, and socio-economic vulnerability mapping. Socio-economic indicators employed in this study were selected after a review of literature and available data sets. However, due to a lack of detailed spatial data and the difficulties in obtaining it, several important socio-economic variables, like household income, sources of livelihood, education, and employment, were not included in this study. The existing statistical data, especially at more localized levels, were not available, making it hard to perform a comprehensive analysis of community vulnerabilities. Consequently, the study had to depend on available proxy variables, such as population density, access to primary healthcare, and road networks, which are frequently used as indirect measures of vulnerability in similar studies [15,26,28,33,71]. Although these proxies are not as precise as direct socio-economic indicators, they still capture broader socio-economic trends like population concentration and the availability of essential services that affect community resilience. For instance, population density, which is linked to exposure and sensitivity to hazards, along with access to healthcare and road networks, which influence a community's ability to adapt, were included in the vulnerability matrix. While these proxies are not perfect replacements for more detailed socio-economic data, they were essential given the data limitations and provide a meaningful assessment of vulnerability within the study's context. Table 2 outlines these socio-economic factors included in the study.

Table 2. Socio-economic indicators.

Categories	Socio-Economic Factors/Indicators	Effect on Vulnerability	Data Source
Exposure sensitivity	Population density	Higher population density often leads to increased exposure to hazards such as flooding [6]. In densely populated regions, the concentration of individuals exacerbates the effects of these hazards, as more people are simultaneously affected by limited resources and emergency services during disasters [70].	Obtained from Worldpop a database for global population and their characteristics at high resolution.
	Population below 5 years	Young children are not physically able to resist during the flood event since their bodies adapt less efficiently than adults, increasing their risk during flood event [38,71].	Obtained from Worldpop.
	Population above 65 years	Older people are particularly sensitive to natural hazards people are not physically able to resist during the flood event and are likely suffering from pre-existing health conditions that can be exacerbated by environmental factors, making them a high-risk group during disasters [71,72].	Obtained from Worldpop.

Table 2. Cont.

Categories	Socio-Economic Factors/Indicators	Effect on Vulnerability	Data Source
	Road network	The road network is crucial for understanding human and socio-economic interactions, particularly in accessing essential services [73]. Access to road networks facilitates quicker responses during emergencies and enhances the overall adaptive capacity of communities [26].	Extracted from OpenStreetMap (OSM), a global open-source database where volunteers map geographic elements [73].
Adaptive capacity	Access to primary healthcare facilities,	Access to healthcare facilities enables quicker medical responses during disasters. When facilities are within reach, individuals can receive timely treatment for injuries or health issues that arise during emergencies [74–76]. Primary healthcare facilities serve as the initial point of entry for individuals seeking healthcare services.	Computed from the spatial distribution of primary healthcare facilities available from the Ministry of Health of Rwanda and downloaded from the national spatial data geoportal.
	Points of interest (POIs)	Socio-economic related POIs, including economic and social activities, were used to describe the availability of socio-economic activities across the city [77]. In total, 804 POIs were extracted and grouped into eight categories, namely hospitality services, education, amenities, shopping centers, financial services, culture and recreation, auto services, and health.	POIs were obtained from OSM.

Following the selection of indicators and the collection of data for each indicator, the values were transformed to raster with 10 m resolution, and normalized using the Min-Max method (see Equation (1)), which results in values between 0 and 1, with 0 denoting the lowest value and high denoting the highest value for each indicator. This allowed for the generation of comparable datasets. Figure 4 illustrates these normalized values and their geographic implications, providing a clear depiction of how various socio-economic factors contribute to overall vulnerability levels.

To construct a socio-economic vulnerability index, the study initially tried PCA, which is known as the best for avoiding expert bias while speeding up the process of the assessment [26]. To ensure that the available datasets were suitable for PCA, the study conducted the Kaiser-Meyer-Olkin (KMO) test, which is used to assess the sampling adequacy for factor analysis, ensuring that the data is suitable for such analysis [78]. The results indicated that the datasets were not suitable for PCA whereby KMO was below 0.8, a threshold for which the PCA is considered reliable [78]. Therefore, the study employed an indicator-based approach through Composite Index to compute the socio-economic vulnerability index as an alternative. The overall socio-economic vulnerability (SEV_i) was calculated by combining flood susceptibility (FS_i), sensitivity (S_i), and adaptive capacity (AC_i) into a unified index, following Equation (2) adopted from [28]:

$$SEV_i = FS_i + S_i - AC_i \quad (2)$$

Whereby sensitivity S_i for each area i was calculated as the sum of the proportions of Pop_d , $Pop_{<5}$, and $Pop_{>65}$ represent the normalized proportions of the population density, population under 5 years and over 65 years in each area, respectively, given by Equation (3) adopted from [14]:

$$S_i = Pop_d + Pop_{<5} + Pop_{>65} \quad (3)$$

Adaptive capacity AC_i was determined by access to key infrastructure and services that support community resilience to natural hazards. Therefore, the adaptive capacity index was calculated by following Equation (4):

$$AC_i = PHF_i + POI_i + RN_i \quad (4)$$

where PHF_i , POI_i , and RN_i are the normalized values representing access to healthcare, the density of POIs, and road network infrastructure in area i . Obtained SEV_i was normalized using Equation (1). Since flood susceptibility was aggregated at neighborhood level, socio-economic vulnerability was also aggregated at the same scale to keep spatial consistency, allow understanding of spatial distribution of vulnerability at small administrative unity for facilitating communication of information.

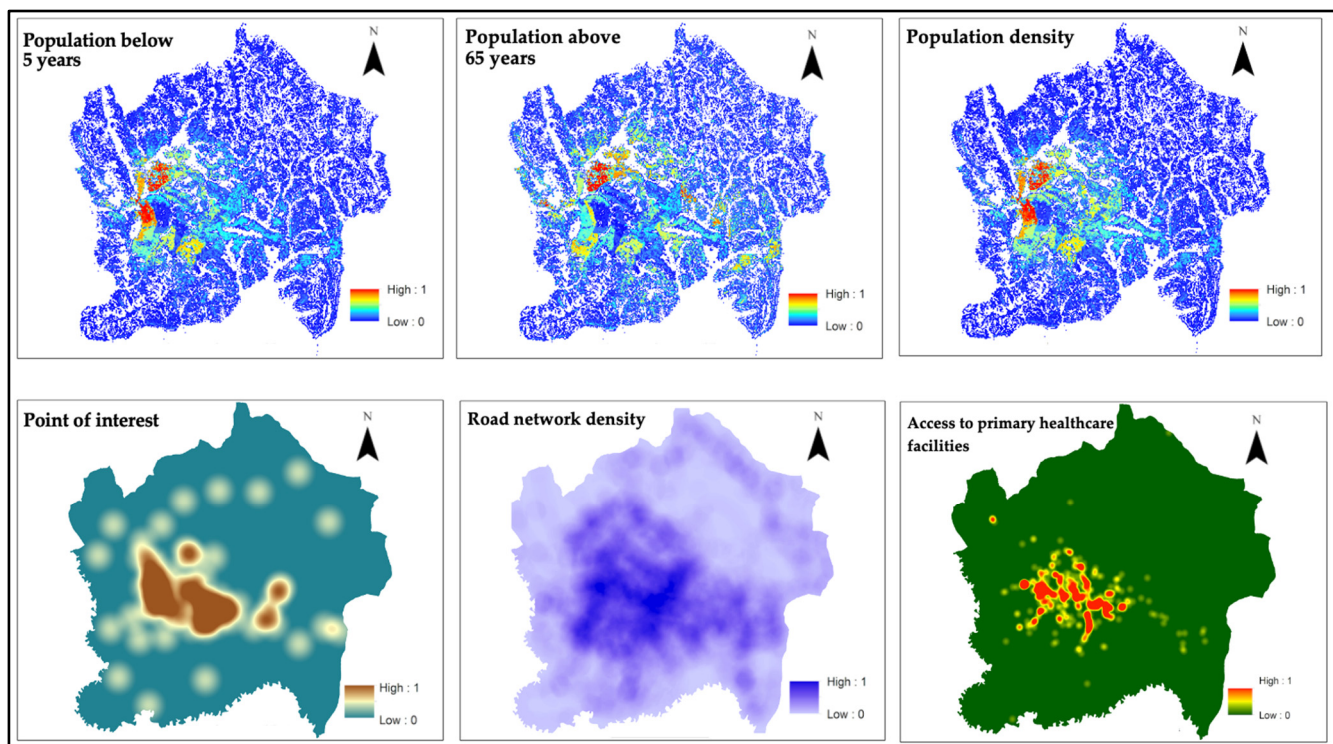


Figure 4. Socio-economic factors employed in the study.

2.2.5. Validation of Flood Susceptibility and Socio-Economic Vulnerability Maps

The flood susceptibility map was validated against historical flood data by assessing the correspondence between identified susceptible areas and actual past flood events, as illustrated in Figure A3 of Appendix C. Additionally, the socio-economic vulnerability map was validated through comparative analysis with existing studies that have mapped socio-economic inequalities and poverty within the same study area. This validation process was further strengthened by referencing methodologies from other successful flood susceptibility modeling and socio-economic vulnerability assessments, ensuring adherence to established standards. Furthermore, qualitative validation was achieved through visual inspections and local knowledge, which helped confirm that the outputs were consistent with the area's social, economic, geographic, and environmental conditions.

2.3. Scalability and Transferability of the Framework

Scalability testing in this study consisted of assessing the ability of framework to handle an increasing amount of data or its capacity to be enlarged to accommodate more

models. In contrast, transferability was used to assess efficiency of the models applied in different contexts and settings, beyond its original design. To evaluate whether the proposed framework is transferable, the study focused on historical flooding data from Kampala and Dar es Salaam, which were extracted from Sentinel-1 imagery, specifically targeting flood events that occurred between May 2019 and September 2020 in Kampala and October 2020 in Dar es Salaam (<https://floodlist.com/africa>, accessed on July 2024). Following the same image ratioing and Otsu's thresholding approach as applied to the City of Kigali, 469 and 384 flood location points were extracted and used for Dar es Salaam and Kampala, respectively. To facilitate this analysis, Digital DEMs from the Shuttle Radar Topography Mission (SRTM) were utilized to derive essential topographic features, including slope, elevation, aspect, and drainage density. Furthermore, cumulative rainfall data were sourced from CHIRP, while land cover information was obtained from ESRI. The NDVI and NDBI were calculated using Sentinel-2 images. The scalability and transferability were limited to testing machine learning models and tuning and validation to ensure accurate predictions.

Initially, the MLP model trained on data from Kigali was applied to predict flooding in both Kampala and Dar es Salaam. This step aimed to evaluate how the model can adapt to different geographical contexts. Following this initial application, the model trained on Kigali was fine-tuned by using subsets of data specific to Kampala and Dar es Salaam. This iterative process was aimed at learning how model capability improves while being exposed to local conditions through data variation for flood susceptibility mapping. Furthermore, the scalability of the proposed framework was evaluated by training MLP, SVM, RF, and XGBoost models in both cities, systematically splitting the available data into training (80%) and testing (20%) sets. This methodological approach allows for a comprehensive analysis of how effectively the framework can be adapted and applied across different urban contexts, even when the transferability is not well-suitable.

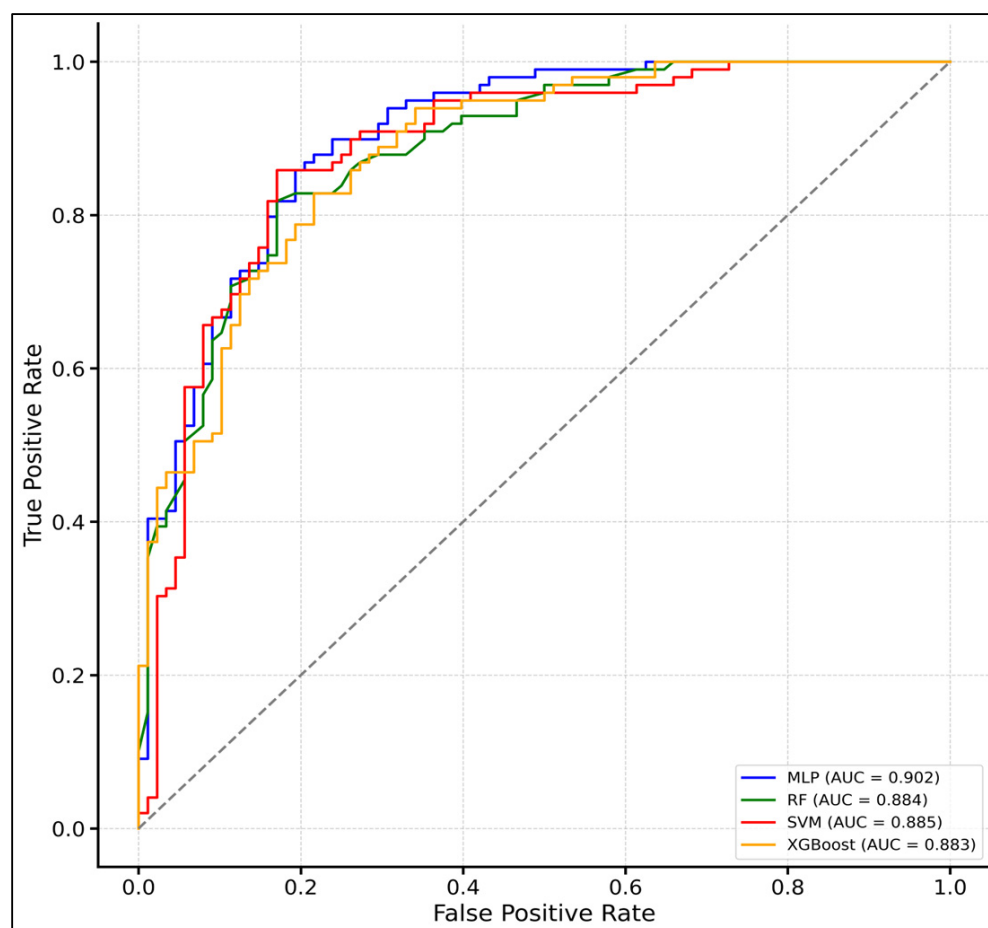
3. Results and Discussion

3.1. Flood Susceptibility Map

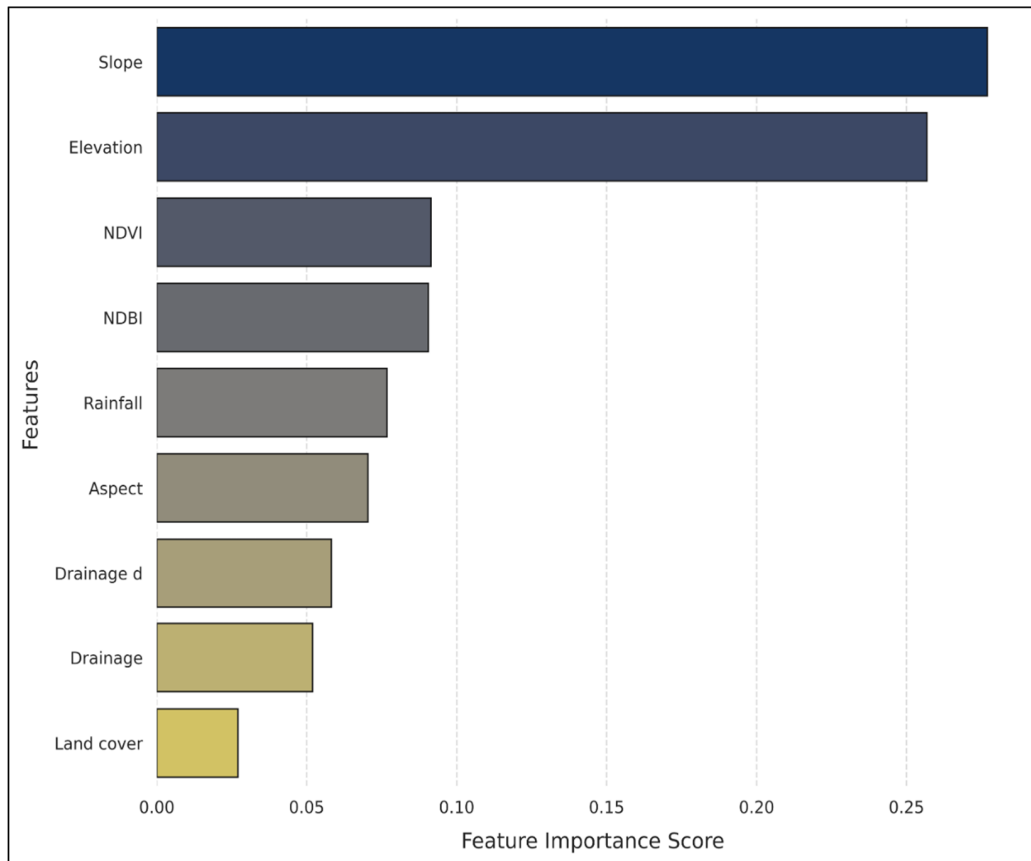
The results presented in Table 3 indicate the performance of the model based on AUC, Accuracy, Precision, Recall, and F1-Score metrics. Figure 5 presents variation of AUC on test data for all models. The results show that the MLP model exhibits the best performance with an AUC of 0.902, indicating it has the highest ability to distinguish between positive and negative cases. This performance is complemented by its Accuracy of 0.85, meaning it accurately predicts 85% of cases. MLP also excels with a Precision score of 0.83. MLP also presents a high Recall of 0.90, indicating they are equally effective at capturing true positive cases. Finally, the MLP again outperforms others with an F1 Score of 0.86, reflecting its overall effectiveness in balancing precision and recall. Following closely, the SVM model demonstrates nearly equivalent performance to MLP, though it has slightly higher prediction errors. The RF model performs slightly behind SVM in terms of AUC but still shows excellent classification ability, as witnessed by its metrics. Finally, the XGBoost model is marginally less effective than the other models, though its performance is still commendable. While all models show strong performance with minimal differences, making them all viable candidates for classification tasks, MLP stood out slightly in all metrics. Thus, it was selected as the best model and was applied to the entire study area to compute flood susceptibility.

Table 3. Performance of Models Based on AUC, Accuracy, Precision, Recall, and F1-Score.

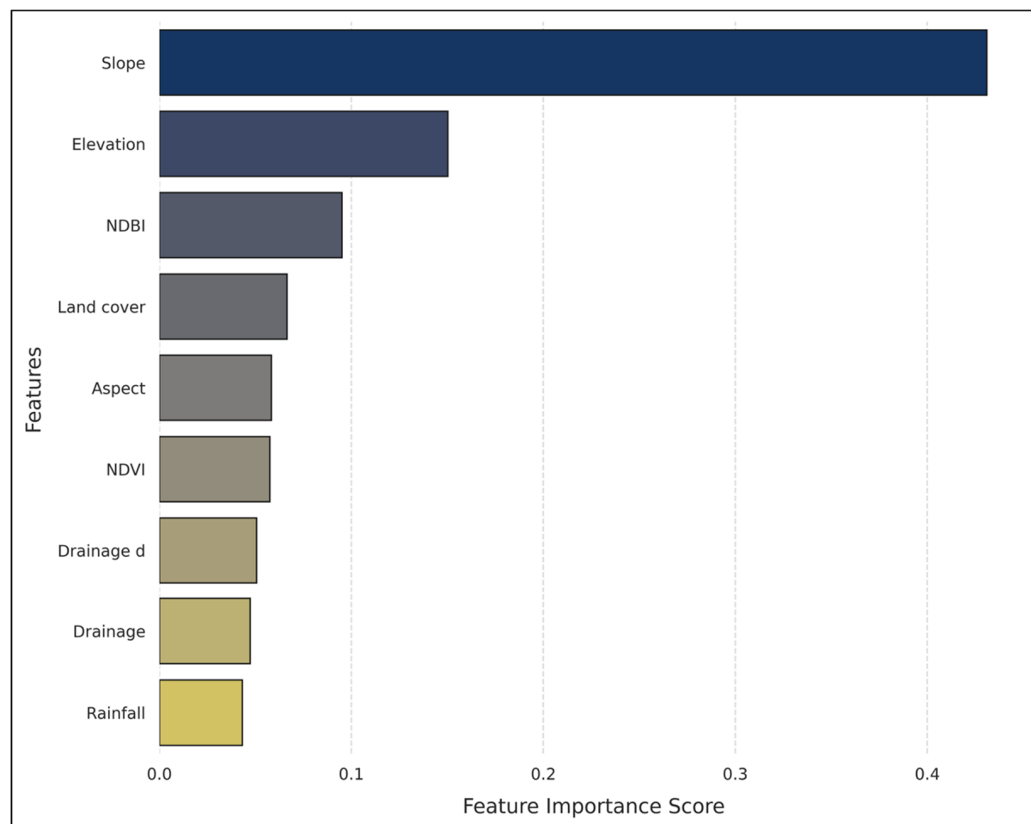
Model	AUC	Accuracy	Precision	Recall	F1-Score
MLP	0.902	0.85	0.83	0.90	0.86
SVM	0.885	0.82	0.79	0.90	0.84
RF	0.884	0.80	0.78	0.87	0.82
XGBoost	0.883	0.80	0.77	0.88	0.82

**Figure 5.** The receiver operating characteristic curves (AUC) on the testing dataset for the models.

While MLP and SVM demonstrate high performance, they are less interpretable and require additional methods to understand the factors that significantly contribute to their predictive capabilities. MLPs, like other artificial neural networks, are regarded as black box models due to their complex architectures, which offer minimal inherent interpretability [79,80]. Similarly, the non-linear kernel SVM model used in this study also lacks transparency and interpretability [80]. This makes their adoption to the domain application for decision makers very hard. In contrast, RF and XGBoost emerge as more interpretable models. Both models have a straightforward analysis of feature importance, allowing users to easily identify which features most significantly influence predictions. Figure 6a,b illustrate the feature importance for RF and XGBoost, respectively. The results shown in these figures indicate that slope and elevation play crucial roles in predicting flood susceptibility across the City of Kigali.



(a)



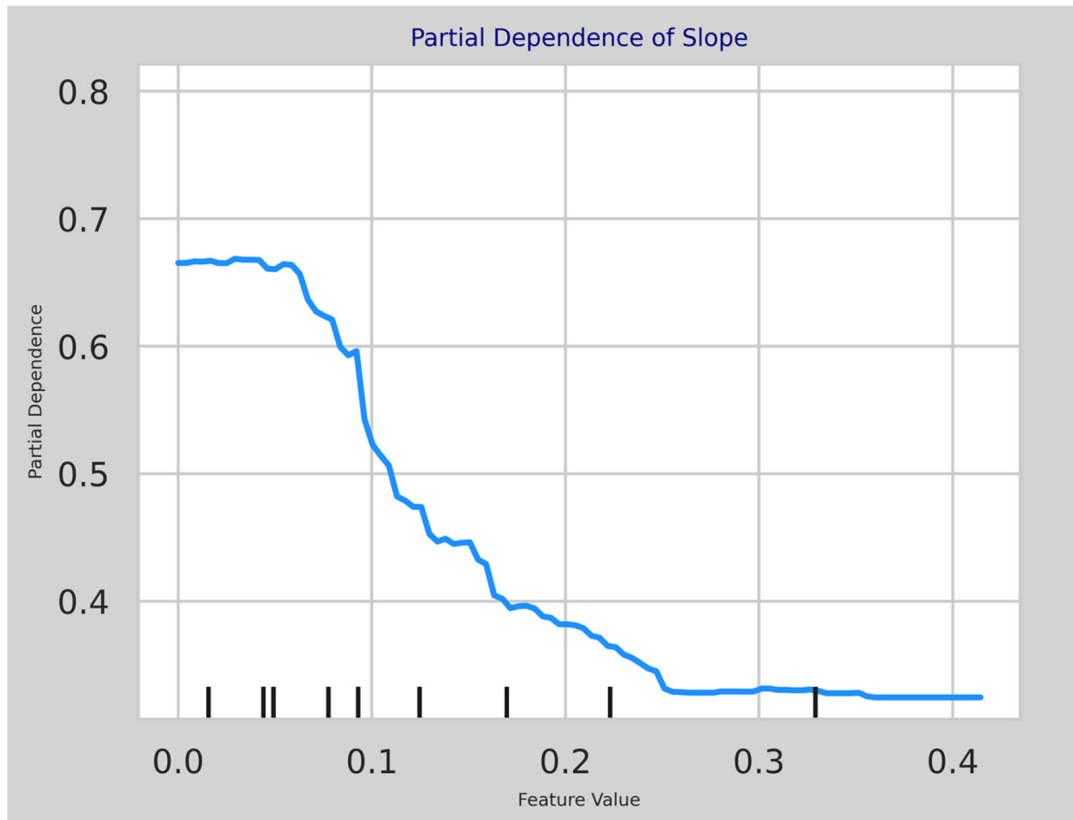
(b)

Figure 6. Features/factors importance score for (a) RF model, and (b) XGBoost model.

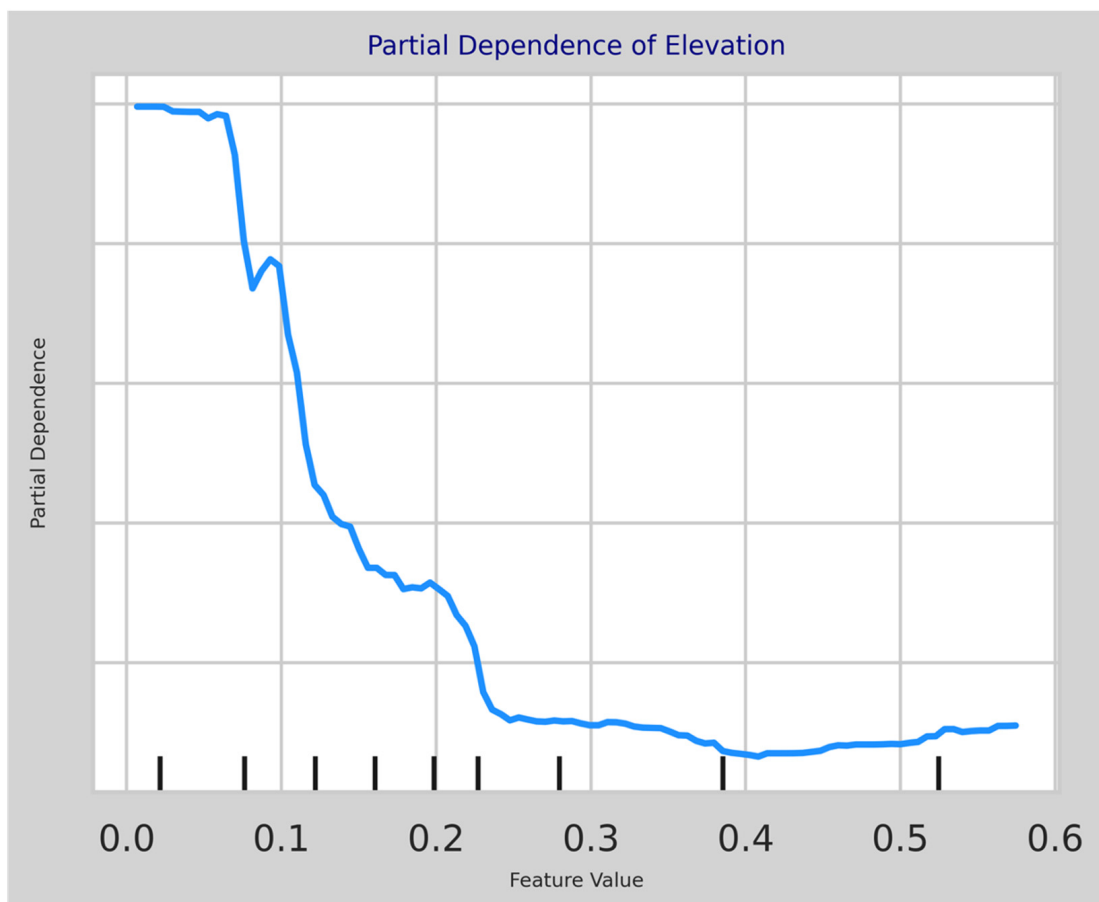
Both RF and XGBoost use different methods to evaluate feature importance. In RF, feature importance is calculated using the Gini impurity or mean decrease in accuracy, while XGBoost calculates it based on metrics such as gain (which measures the contribution to the model's prediction accuracy), coverage, and frequency. These methods allow for a clear ranking of feature contributions. As for the observation in Figure 6, the lower importance of rainfall in the XGBoost model may be due to how XGBoost handles complex, nonlinear relationships between features and the target variable, while Random Forest might capture such relationships better. Regarding the difference in elevation importance between RF and XGBoost, RF may consider the global relationship of elevation with the target variable, whereas XGBoost, being a gradient boosting model, might focus more on local, higher-order interactions that make elevation appear less significant. We have addressed these differences more critically in the discussion.

For Figure 7, the partial dependence plot represents the marginal effect of each feature on the predicted outcome, computed by averaging model predictions over a range of feature values while holding other variables constant. The results presented in the Figure 7a illustrate that for the slope, the predicted probability of the flooding is high for lower slope values but decreases sharply as the slope exceeds approximately 0.1, eventually stabilizing at higher values. While for Elevation, the Figure 7b indicates a strong negative relationship with the predicted outcome, where the probability is high at very low elevations but decreases substantially as elevation increases, remaining constant at higher elevation levels.

The result presented in Figure 8a,b present flood susceptibility map generated using MLP model, and its respective aggregation at the neighborhood level. These results highlight part of the city with darker brown shades, which represent higher susceptibility to flooding, while lighter shades indicate lower susceptibility. The central and southern parts of the region show the highest susceptibility, which makes them more prone to flooding. On the contrary, the northern and northeastern parts show lower flood susceptibility. This result illustrates that the MLP model was able to identify areas with high flood susceptibility, which closely aligns with historically flooded locations. This model's strong predictive capabilities are rooted in its ability to map complex non-linear relationships between environmental factors such as topography, land use, and hydrological conditions, as highlighted by [22,24,25]. By integrating diverse data inputs such as topography, land use, and hydrological conditions, the model generates a comprehensive susceptibility map, which not only identifies high-risk zones but also provides critical insights for developing further actions targeted to protect the public against health risks associated with flood exposure. This ability of the model to predict flood susceptibility would allow decision-makers to enhance disaster preparedness, mitigate health risks, and implement community-level responses. Thus, the result emphasizes the need for targeted public health actions, as floods can significantly impact both physical infrastructure and public health by increasing the risk of waterborne diseases, injuries, and disruptions to healthcare access, as shown by malaria [51,81]. The resulting map would serve as a valuable tool for planners and public health officials, helping them prioritize flood prevention measures such as enhanced drainage systems, flood barriers, and land use regulations in high-risk areas while also guiding emergency preparedness and healthcare resource allocation.



(a)



(b)

Figure 7. Partial dependence plots for (a) slope and (b) elevation (RF model).

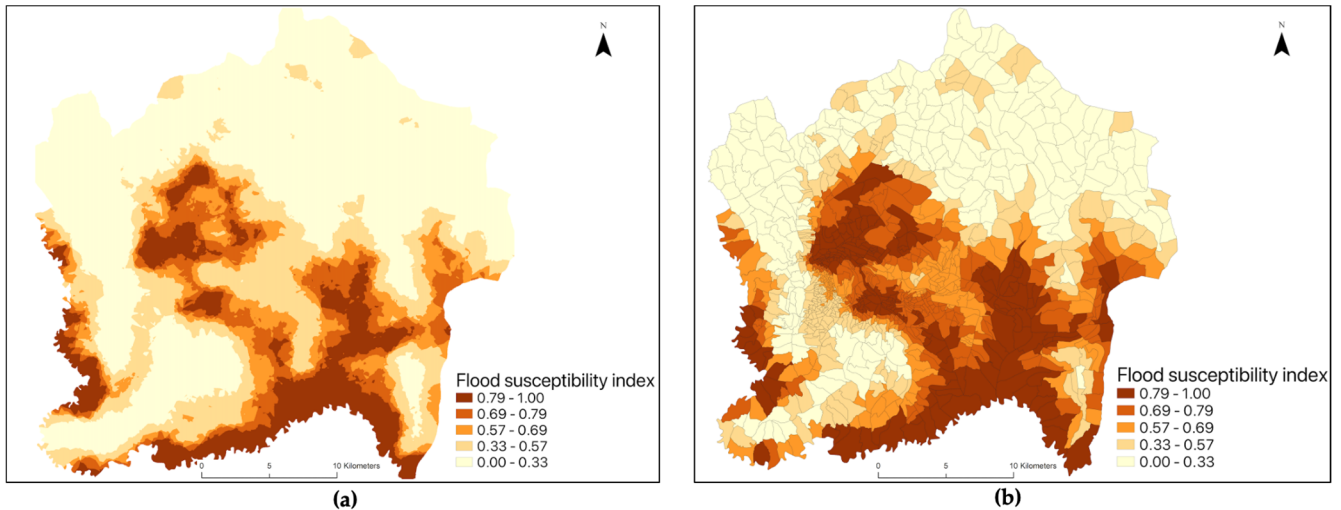


Figure 8. (a) Flood susceptibility map generated using MLP model. (b) Flood susceptibility aggregated at the neighborhood level.

3.2. Socio-Economic Vulnerability Map

The result presented in Figure 9a,b present socio-economic vulnerability to flooding across the City of Kigali. They reveal how flood susceptibility, a prevalent natural hazard in Kigali due to its hilly terrain and frequent heavy rainfall, disproportionately affects socio-economically vulnerable populations. The darker shaded areas on the map represent higher socio-economic vulnerability resulting from high flood susceptibility and sensitivity and with relatively low adaptive capacity. These areas are characterized by informal settlements, lower income levels, inadequate infrastructure, and limited access to essential services such as access to healthcare facilities, as highlighted in [44,46,82]. In contrast, the central parts of Kigali, which exhibit lower vulnerability, benefit from improved urban planning, robust infrastructure, and more resilient housing, as shown by [41,45,47].

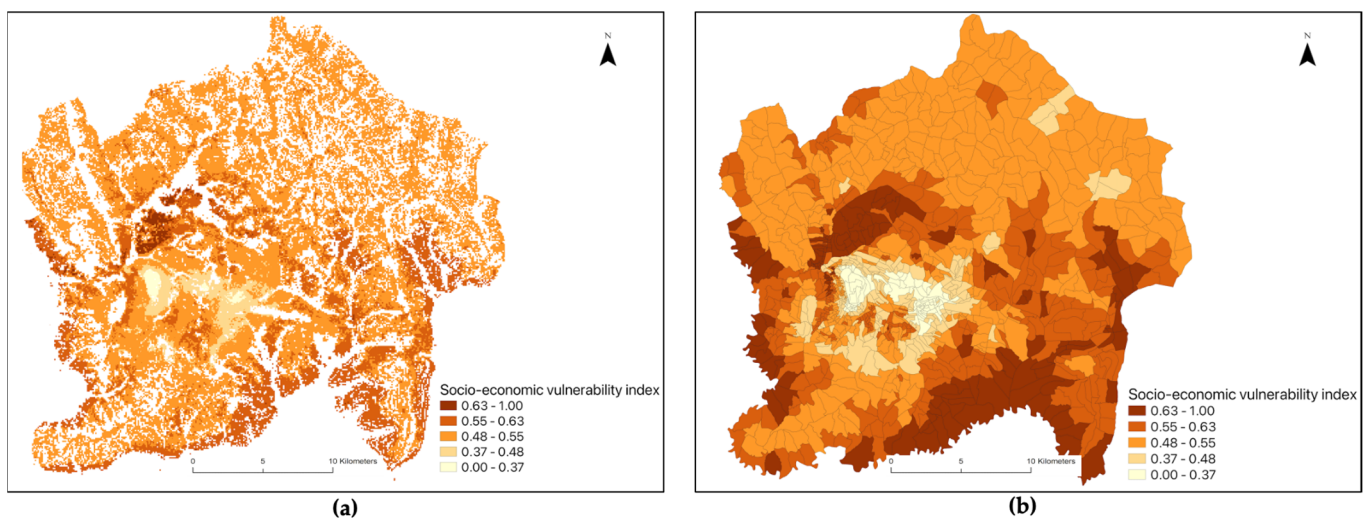


Figure 9. (a) Socio-economic vulnerability map generated using the Composite Index approach. (b) Socio-economic vulnerability aggregated at the neighborhood level.

The spatial variation of socio-economic vulnerability across the city indicates that the presence a higher concentration of the population, especially young and elderly residents, who are concentrated in the most vulnerable areas, such as next to wetlands or in informal settlements with poor drainage systems, are at higher risk of negative health outcomes

due to inadequate drainage and flood management. These locations amplify the risks associated with heavy rainfall and flooding, potentially leading to catastrophic outcomes like property loss, displacement, and increased exposure to health hazards. This exposes them to risks, such as exposure to outbreaks of waterborne diseases such as cholera, typhoid, and dysentery [9,26]. Floodwaters in these areas often contaminate drinking water supplies and overwhelm fragile sanitation systems, creating conditions favorable for disease transmission [49]. Additionally, floodings lead to stagnant water, increasing the risk of mosquito-borne diseases like malaria [50].

3.3. Scalability and Transferability

First, the study applied models trained on Kigali data to make predictions in Kampala and Dar es Salaam; the AUC and MAE obtained are presented in Table 4. In Kampala, the XGBoost model achieved the highest AUC of 0.519, whereas in Dar es Salaam, the MLP model had the highest AUC at 0.402. This indicates that the models trained on the Kigali dataset struggled to generalize across different urban settings, highlighting the need for continuous refinement of machine learning algorithms by incorporating local data and expert knowledge into model training processes. As urban environments evolve and data becomes more accessible, leveraging this information can lead to improved predictive accuracy for flood events, ultimately contributing to better urban planning and flood disaster management strategies.

Table 4. AUC and MAE values for Kampala and Dar es Salam.

City	Model	AUC	MAE
Kampala	MLP	0.475	0.511
	RF	0.473	0.530
	SVM	0.455	0.547
	XGBoost	0.519	0.484
Dar es Salaam	MLP	0.402	0.523
	RF	0.403	0.590
	SVM	0.447	0.535
	XGBoost	0.387	0.605

Then we fine-tuned network weights of MLP, which performed well on Kigali using small sub-sets of data from target cities (using 10%, 20%, 30%, 40%, and 50% of all available data for each city), performances reduced slightly to AUC of 0.491 for Kampala but improved slightly to AUC of 0.590 for Dar es Salaam. These performance changes correspond to 50% subset of data used in fine-tuning for each target city. This lower performance of the employed models highlights potential limitations in their transferability, which can be due to the fact that cities are different in morphology and do not exhibit similar geographic/topographic characteristics. However, environmental conditions and urban dynamics differ significantly between cities [17,82]. For instance, all the cities used in this study to test the proposed framework exhibit different topographic/geographic patterns, socio-economic conditions, and climate. Thus, differences in the spatial distribution of flood hazard, infrastructure, or population density could affect model performance. Thus, the success of transferring models across cities may be constrained by variations in data availability, quality, or level of details.

However, despite the difficulties of transferability of the machine learning model trained on Kigali to Kampala and Dar es Salaam, the framework itself demonstrated

scalability whereby when we trained the machine learning models using data from each respective city, the results on the test set revealed good performance of the models as shown in Figure 10a,b. Additionally, the framework is designed to incorporate various data sources and machine learning models, enabling it to be applied in diverse urban environments. Its architecture allows for the integration of city-specific data, which shows its adaptability to different urban contexts.

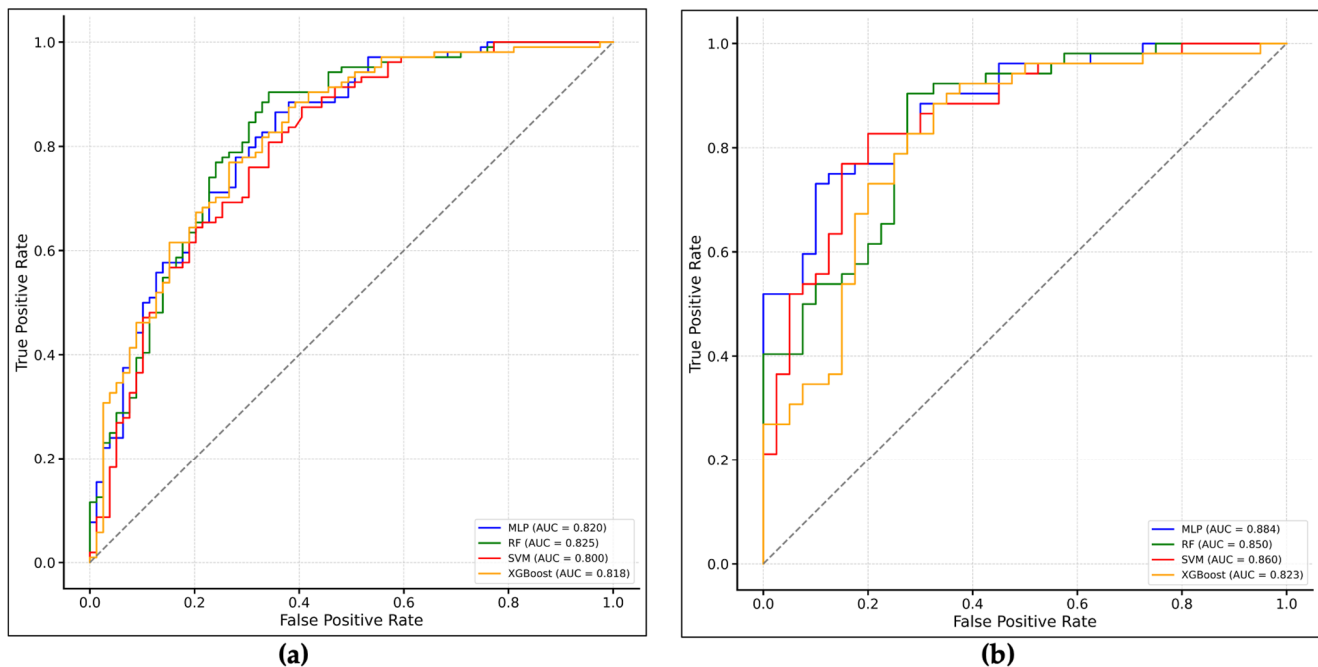


Figure 10. The receiver operating characteristic curve (AUC) curves on the testing dataset for the models (a) Kampala and (b) Dar es Salaam.

3.4. Limitations of the Study

The study assessed and mapped social-economic vulnerability related to flood hazards using an assessment framework that leverages machine learning and indicator-based approaches. One major limitation of this study is the availability and detail of socioeconomic data, which greatly affected the choice of indicators used to evaluate vulnerability (as illustrated in Table 2). This limitation suggests that while our model offers a valuable approximation of vulnerability, it may not fully reflect the complex dynamics that shape how various communities experience and respond to flood hazards. The lack of such data can result in an oversimplification of vulnerability patterns, potentially leading to an underestimation or overestimation of the actual flood-related socioeconomic risks in certain regions. Another crucial factor to consider is the model's transferability and applicability across different geographical contexts. Our framework's flexibility allows it to be tailored to areas with varying levels of data availability. However, caution is necessary when using the model in regions where socioeconomic and environmental conditions differ significantly from those in the study area. Variations in urban infrastructure, governance structures, disaster preparedness measures, and social inequalities can all impact the model's effectiveness. Therefore, any application of this framework in a new context should be accompanied by a thorough evaluation of local data sources and the relevance of the chosen indicators to ensure meaningful and context-specific vulnerability assessments. Although the above limitation exists in the current work, the proposed framework is designed to be flexible to allow application to areas with limited data. However, the availability of more data would

provide more valuable outputs for supporting decision-making for planners and managers to positively deal with hazardous urban environments.

4. Conclusions

This study introduces a scalable and transferable framework designed to map socio-economic vulnerability to urban natural hazards in areas with limited data. The framework combines socio-economic factors with environmental data utilizing machine learning and indicator-based approaches for susceptibility and vulnerability assessment. The integration of machine learning was particularly effective in identifying non-linear patterns of vulnerability, which are often difficult to capture using traditional methods. Moreover, the framework's reliance on freely available data allows for wide applicability, especially in regions with limited access to high-quality socio-economic data. By applying the proposed framework to the City of Kigali, the findings reveal that the central, southern, and western regions of Kigali are particularly vulnerable to flooding. Notably, while the central area faces a high risk of flooding, it shows lower socioeconomic vulnerability due to better economic conditions that facilitate adaptation. In contrast, regions with poorer socioeconomic conditions are more severely impacted by flooding hazards, emphasizing the link between environmental risk and social inequality. Although the model developed in Kigali did not initially perform well when applied to Kampala and Dar es Salaam, fine-tuning it with localized data enhanced its effectiveness. This highlights the necessity of making context-specific adjustments when transferring models to different urban environments. Additionally, our results indicate that the proposed framework can be adapted to evaluate socioeconomic vulnerability to various hazards in urban settings.

Author Contributions: Conceptualization: Esaie Dufitimana, Paterne Gahungu, Ernest Uwayezu, Emmy Mugisha and Jean Pierre Bizimana; Methodology: Esaie Dufitimana; data curation, Esaie Dufitimana; Analysis, Esaie Dufitimana; Writing—original draft preparation: Esaie Dufitimana, Paterne Gahungu, Ernest Uwayezu, Emmy Mugisha and Jean Pierre Bizimana; Writing, Review and editing: Esaie Dufitimana, Paterne Gahungu, Ernest Uwayezu, Emmy Mugisha and Jean Pierre Bizimana; Supervision: Paterne Gahungu, Ernest Uwayezu, Emmy Mugisha and Jean Pierre Bizimana. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge the funding from National Institute of Health (NIH) (Grant #s U2RTW012122 and UE5 HL172181) provided through Research training in Data Science for Health in Rwanda, collaborative projects between the Regional Centre of Excellence in Biomedical Engineering and E-Health (CEBE), at the University of Rwanda, African Institute for Mathematical Sciences (AIMS) and Washington University in Saint Louis. The statements made, including study design, data acquisition & analysis, and decision to publish are solely the responsibility of the authors. The APC was waived by the journal.

Data Availability Statement: The DEM is available on request from the National Land Authority of Rwanda, land cover is available on request at the City of Kigali, Sentinel-1 and 2 images are available from the European Space Agency (ESA) at <https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data> (accessed on 19 June 2024), Rainfall data are available from CHIRPS at <https://www.chc.ucsb.edu/data/chirps> (accessed on 22 June 2024), drainage networks data available on request at City of Kigali, Population data available from Worldpop at <https://www.worldpop.org/datacatalog/> (accessed on 7 June 2024), road network and POIs available from OSM at <https://www.openstreetmap.org> (accessed on 5 May 2024), primary healthcare facilities are available from the Ministry of Health of Rwanda at <https://geodata.rw/portal/home/> (accessed on 19 January 2024).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

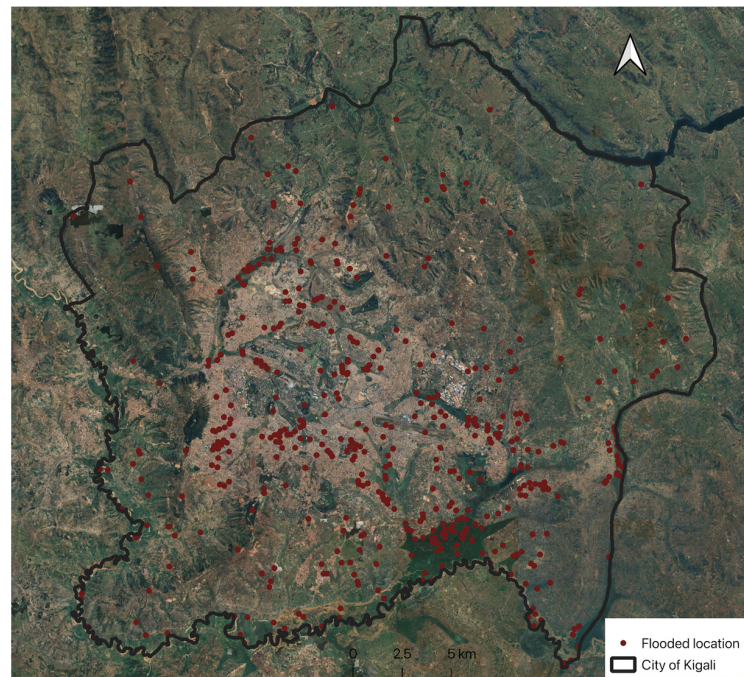


Figure A1. Spatial distribution of 456 flood events recorded between 22 December 2019, and 25 December 2019, across the City of Kigali.

Appendix B

Figure A2 illustrates the correlation matrix among environmental factors. The strength and direction of the correlations are represented by the color gradient, with red indicating strong positive correlations, blue for strong negative correlations, and lighter shades for weaker correlations.

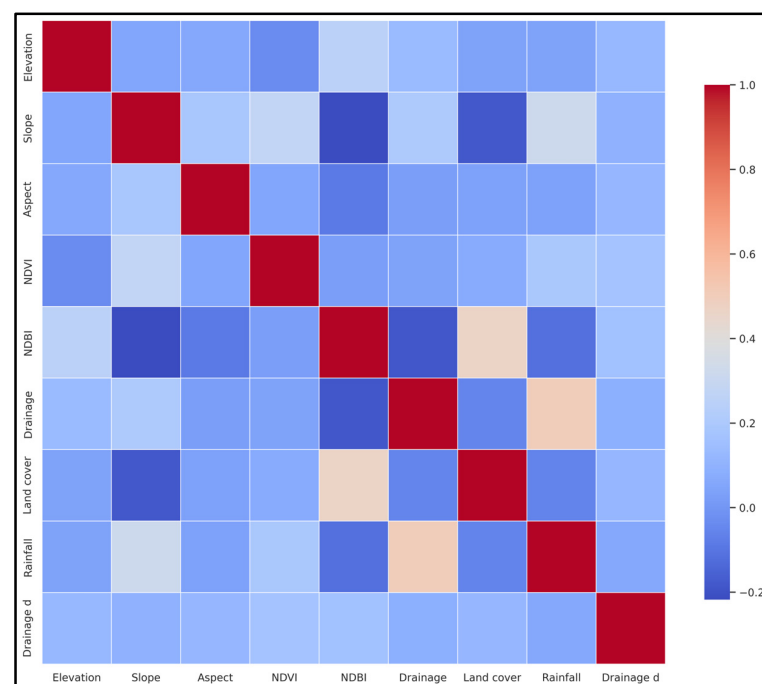


Figure A2. Heatmap depicting the correlation matrix among environmental factors. (Note: Drainage d is distance to drainage).

Appendix C

Figure A3 presents the validation of flood susceptibility map generated in this study against historical flood data.

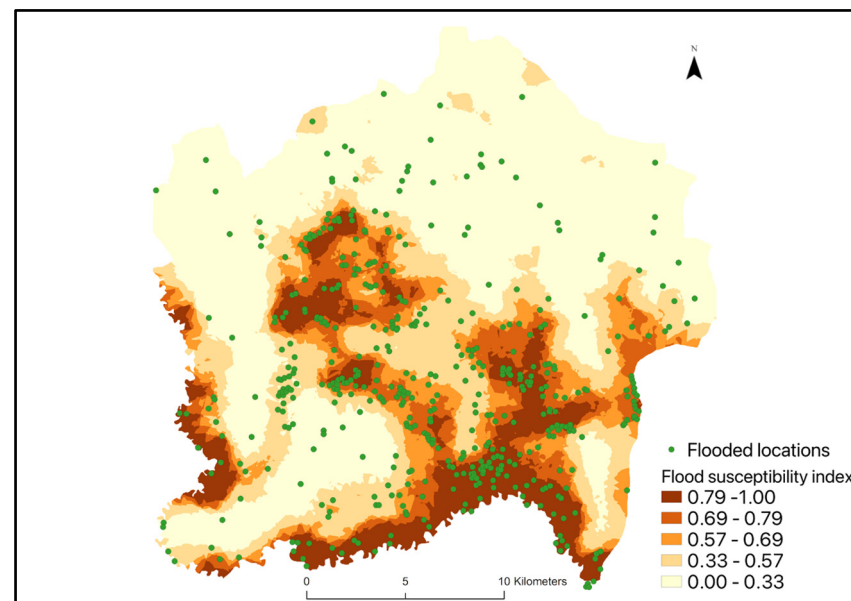


Figure A3. Spatial distribution of flood events overlaid on flood susceptibility map.

References

1. United Nations, World Social Report 2020: Inequality in a Rapidly Changing World. 2020. Available online: <http://www.un.org/development/desa/dspd/wp-content/uploads/sites/22/2020/02/World-Social-Report2020-FullReport.pdf> (accessed on 12 June 2023).
2. UN-Habitat. Urbanization and Development: Emerging Futures. Nairobi. 2016. Available online: <https://unhabitat.org/sites/default/files/download-manager-files/WCR-2016-WEB.pdf> (accessed on 22 May 2020).
3. Mahabir, R.; Crooks, A.; Croitoru, A.; Agouris, P. The study of slums as social and physical constructs: Challenges and emerging research opportunities. *Reg. Stud. Reg. Sci.* **2016**, *3*, 399–419. [CrossRef]
4. Alves, B.; Angnuureng, D.B.; Morand, P.; Almar, R. A review on coastal erosion and flooding risks and best management practices in West Africa: What has been done and should be done. *J. Coast. Conserv.* **2020**, *24*, 1–22. [CrossRef]
5. Wisner, B.; Blaikie, P.; Cannon, T.; Davis, I. *At Risk*; Routledge: London, UK, 2014. [CrossRef]
6. Cutter, S.L. Vulnerability to environmental hazards. *Prog. Hum. Geogr.* **1996**, *20*, 529–539. [CrossRef]
7. Prana, A.M.; Dionisio, R.; Curl, A.; Hart, D.; Gomez, C.; Apriyanto, H.; Prasetya, H. Informal adaptation to flooding in North Jakarta, Indonesia. *Prog. Plann.* **2024**, *186*, 100851. [CrossRef]
8. Prana, A.M.; Curl, A.; Dionisio, M.R.; Gomez, C.; Hart, D.; Apriyanto, H.; Prasetya, H. Urban planning approaches to support community-based flood adaptation in North Jakarta Kampung. *Disaster Prev. Manag. Int. J.* **2024**, *33*, 383–405. [CrossRef]
9. Deroliya, P.; Ghosh, M.; Mohanty, M.P.; Ghosh, S.; Rao, K.H.V.D.; Karmakar, S. A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions. *Sci. Total Environ.* **2022**, *851*, 158002. [CrossRef]
10. Adger, W.N. Vulnerability. *Glob. Environ. Change* **2006**, *16*, 268–281. [CrossRef]
11. Berkes, F. Understanding uncertainty and reducing vulnerability: Lessons from resilience thinking. *Nat. Hazards* **2007**, *41*, 283–295. [CrossRef]
12. Hagenlocher, M.; Schneiderbauer, S.; Sebesvari, Z.; Bertram, M.; Renner, K.; Renaud, F.; Wiley, H.; Zebisch, M. Climate Risk Assessment for Ecosystem-Based Adaptation A Guidebook for Planners and Practitioners. Bonn, 2018. Available online: www.giz.de (accessed on 16 February 2025).
13. United Nations, Revision of World Urbanization Prospects. United Nations Department of Economic and Social Affairs. 2018. Available online: <https://population.un.org/wup/assets/WUP2018-Report.pdf> (accessed on 4 March 2023).
14. Aroca-Jiménez, E.; Bodoque, J.M.; García, J.A. How to construct and validate an Integrated Socio-Economic Vulnerability Index: Implementation at regional scale in urban areas prone to flash flooding. *Sci. Total Environ.* **2020**, *746*, 140905. [CrossRef]
15. Biswas, S.; Nautiyal, S. A review of socio-economic vulnerability: The emergence of its theoretical concepts, models and methodologies. *Nat. Hazards Res.* **2023**, *3*, 563–571. [CrossRef]

16. McCallum, I.; Kyba, C.C.M.; Bayas, J.C.L.; Moltchanova, E.; Cooper, M.; Cuaresma, J.C.; Pachauri, S.; See, L.; Danylo, O.; Moorthy, I.; et al. Estimating global economic well-being with unlit settlements. *Nat. Commun.* **2022**, *13*, 2459. [CrossRef] [PubMed]
17. Yeh, C.; Perez, A.; Driscoll, A.; Azzari, G.; Tang, Z.; Lobell, D.; Ermon, S.; Burke, M. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nat. Commun.* **2020**, *11*, 2583. [CrossRef]
18. Kuffer, M.; Thomson, D.R.; Boo, G.; Mahabir, R.; Grippa, T.; Vanhuyse, S.; Engstrom, R.; Ndugwa, R.; Makau, J.; Darin, E.; et al. The role of earth observation in an integrated deprived area mapping 'system' for low-to-middle income countries. *Remote Sens.* **2020**, *12*, 982. [CrossRef]
19. Skinner, C. Issues and Challenges in Census Taking. *Annu. Rev. Stat. Appl.* **2018**, *5*, 49–63. [CrossRef]
20. Kazemi, M.; Mohammadi, F.; Nafooti, M.H.; Behvar, K.; Kariminejad, N. Flood susceptibility mapping using machine learning and remote sensing data in the Southern Karun Basin, Iran. *Appl. Geomat.* **2024**, *16*, 731–750. [CrossRef]
21. Seleem, O.; Ayzel, G.; de Souza, A.C.T.; Bronstert, A.; Heistermann, M. Towards urban flood susceptibility mapping using data-driven models in Berlin, Germany. *Geomat. Nat. Hazards Risk* **2022**, *13*, 1640–1662. [CrossRef]
22. Javidan, N.; Kavian, A.; Pourghasemi, H.R.; Conoscenti, C.; Jafarian, Z.; Rodrigo-Comino, J. Evaluation of multi-hazard map produced using MaxEnt machine learning technique. *Sci. Rep.* **2021**, *11*, 6496. [CrossRef]
23. Pourghasemi, H.R.; Kariminejad, N.; Amiri, M.; Edalat, M.; Zarafshar, M.; Blaschke, T.; Cerda, A. Assessing and mapping multi-hazard risk susceptibility using a machine learning technique. *Sci. Rep.* **2020**, *10*, 3203. [CrossRef]
24. Sakti, A.D.; Deliar, A.; Hafidzah, D.R.; Chintia, A.V.; Anggraini, T.S.; Ihsan, K.T.N.; Virtriana, R.; Suwardhi, D.; Harto, A.B.; Nurmaulia, S.L.; et al. Machine learning based urban sprawl assessment using integrated multi-hazard and environmental-economic impact. *Sci. Rep.* **2024**, *14*, 13385. [CrossRef]
25. Yousefi, S.; Pourghasemi, H.R.; Emami, S.N.; Pouyan, S.; Eskandari, S.; Tiefenbacher, J.P. A machine learning framework for multi-hazards modeling and mapping in a mountainous area. *Sci. Rep.* **2020**, *10*, 12144. [CrossRef]
26. Ajtai, I.; Ștefănie, H.; Maloș, C.; Botezan, C.; Radovici, A.; Bizău-Cârstea, M.; Baci, C. Mapping social vulnerability to floods. A comprehensive framework using a vulnerability index approach and PCA analysis. *Ecol. Indic.* **2023**, *154*, 110838. [CrossRef]
27. Alabbad, Y.; Demir, I. Comprehensive flood vulnerability analysis in urban communities: Iowa case study. *Int. J. Disaster Risk Reduct.* **2022**, *74*, 102955. [CrossRef]
28. Sun, Y.; Li, Y.; Ma, R.; Gao, C.; Wu, Y. Mapping urban socio-economic vulnerability related to heat risk: A grid-based assessment framework by combing the geospatial big data. *Urban Clim.* **2022**, *43*, 101169. [CrossRef]
29. Zhang, T.; Wang, D.; Lu, Y. Machine learning-enabled regional multi-hazards risk assessment considering social vulnerability. *Sci. Rep.* **2023**, *13*, 13405. [CrossRef]
30. Brower, A.E.; Ramesh, B.; Islam, K.A.; Mortveit, H.S.; Hoops, S.; Vullikanti, A.; Marathe, M.V.; Zaitchik, B.; Gohlke, J.M.; Swarup, S. Augmenting the Social Vulnerability Index using an agent-based simulation of Hurricane Harvey. *Comput. Environ. Urban Syst.* **2023**, *105*, 102020. [CrossRef]
31. Davino, C.; Gherghi, M.; Sorana, S.; Vistocco, D. Measuring Social Vulnerability in an Urban Space Through Multivariate Methods and Models. *Soc. Indic. Res.* **2021**, *157*, 1179–1201. [CrossRef]
32. Hadipour, V.; Vafaie, F.; Kerle, N. An indicator-based approach to assess social vulnerability of coastal areas to sea-level rise and flooding: A case study of Bandar Abbas city, Iran. *Ocean Coast. Manag.* **2020**, *188*, 105077. [CrossRef]
33. Streifeneder, V.; Kienberger, S.; Reichel, S.; Hölbling, D. Socio-Economic Vulnerability Assessment for Supporting a Sustainable Pandemic Management in Austria. *Sustainability* **2023**, *16*, 78. [CrossRef]
34. Zhu, K.; Wang, Z.; Lai, C.; Li, S.; Zeng, Z.; Chen, X. Evaluating Factors Affecting Flood Susceptibility in the Yangtze River Delta Using Machine Learning Methods. *Int. J. Disaster Risk Sci.* **2024**, *15*, 738–753. [CrossRef]
35. Al-Kindi, K.M.; Alabri, Z. Investigating the Role of the Key Conditioning Factors in Flood Susceptibility Mapping Through Machine Learning Approaches. *Earth Syst. Environ.* **2024**, *8*, 63–81. [CrossRef]
36. Khosravi, K.; Pham, B.T.; Chapi, K.; Shirzadi, A.; Shahabi, H.; Revhaug, I.; Prakash, I.; Bui, D.T. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at Haraz watershed, northern Iran. *Sci. Total Environ.* **2018**, *627*, 744–755. [CrossRef] [PubMed]
37. Terna, I.P. Vulnerability: Types, Causes, and Coping Mechanisms. *Int. J. Sci. Manag. Stud. (IJSMS)* **2021**, *4*, 187–194. [CrossRef]
38. Chakraborty, L.; Rus, H.; Henstra, D.; Thistlethwaite, J.; Scott, D. A place-based socioeconomic status index: Measuring social vulnerability to flood hazards in the context of environmental justice. *Int. J. Disaster Risk Reduct.* **2020**, *43*, 101394. [CrossRef]
39. NISR. *Fifth Rwanda Population and Housing Census, 2022*; National Institute of Statistics of Rwanda, Ministry of Finance and Economic Planning; Ministry of Health; The DHS Program, ICF International: Kigali, Rwanda, 2022. Available online: https://www.statistics.gov.rw/publication/main_indicators_2022 (accessed on 2 December 2024).
40. City of Kigali. *Zoning Regulations: Kigali Master Plan 2050*; City of Kigali: Kigali City, Rwanda, 2019.
41. Baffoe, G.; Malonza, J.; Manirakiza, V.; Mugabe, L. Understanding the concept of neighbourhood in Kigali City, Rwanda. *Sustainability* **2020**, *12*, 1555. [CrossRef]

42. Hafner, S.; Georganos, S.; Mugiraneza, T.; Ban, Y. Mapping Urban Population Growth from Sentinel-2 MSI and Census Data Using Deep Learning: A Case Study in Kigali, Rwanda. Available online: <http://arxiv.org/abs/2303.08511> (accessed on 2 August 2023).
43. Nikuze, A.; Sliuzas, R.; Flacke, J. Towards Equitable Urban Residential Resettlement in Kigali, Rwanda. In *GIS in Sustainable Urban Planning and Management*; CRC Press: Boca Raton, FL, USA, 2018; pp. 325–344. [CrossRef]
44. Uwizeye, D.; Irambeshya, A.; Wiehler, S.; Niragire, F. Poverty profile and efforts to access basic household needs in an emerging city: A mixed-method study in Kigali's informal urban settlements, Rwanda. *Cities Health* **2022**, *6*, 98–112. [CrossRef]
45. Dufitimana, E.; Gahungu, P.; Uwayezu, E.; Mugisha, E.; Poorthuis, A.; Bizimana, J.P. Measuring urban socio-economic disparities in the global south from space using convolutional neural network: The case of the City of Kigali, Rwanda. *Geojournal* **2024**, *89*, 107. [CrossRef]
46. Nduwayezu, G.; Ingabire, E.; Bizimana, J.P. Measuring disparities in access to district and referral hospitals in the city of Kigali, Rwanda. *Rwanda J. Eng. Sci. Technol. Environ.* **2023**, *5*, 2617–2321. [CrossRef]
47. Manirakiza, V.; Mugabe, L.; Nsabimana, A.; Nzayirambaho, M. City Profile: Kigali, Rwanda. *Environ. Urban. ASIA* **2019**, *10*, 290–307. [CrossRef]
48. Naeem, A.; Zaheer, Z.; Tabassum, S.; Nazir, A.; Naeem, F. Diseases caused by floods with a spotlight on the present situation of unprecedented floods in Pakistan: A short communication. *Ann. Med. Surg.* **2023**, *85*, 3209–3212. [CrossRef]
49. Haque, A.N. Climate risk responses and the urban poor in the global South: The case of Dhaka's flood risk in the low-income settlements. *Int. J. Disaster Risk Reduct.* **2021**, *64*, 102534. [CrossRef]
50. Liu, Q.; Yuan, J.; Yan, W.; Liang, W.; Liu, M.; Liu, J. Association of natural flood disasters with infectious diseases in 168 countries and territories from 1990 to 2019: A worldwide observational study. *Glob. Transit.* **2023**, *5*, 149–159. [CrossRef]
51. Paterson, D.L.; Wright, H.; Harris, P.N.A. Health Risks of Flood Disasters. *Clin. Infect. Dis.* **2018**, *67*, 1450–1454. [CrossRef] [PubMed]
52. Nagendra, H.; Bai, X.; Brondizio, E.S.; Lwasa, S. The urban south and the predicament of global sustainability. *Nat. Sustain.* **2018**, *1*, 341–349. [CrossRef]
53. Singh, G.; Kishan; Rawat, S. Mapping flooded areas utilizing Google Earth Engine and open SAR data: A comprehensive approach for disaster response. *Discov. Geosci.* **2024**, *2*, 1–12. [CrossRef]
54. Rahman, M.R.; Thakur, P.K. Detecting, mapping and analysing of flood water propagation using synthetic aperture radar (SAR) satellite data and GIS: A case study from the Kendrapara District of Orissa State of India. *Egypt. J. Remote Sens. Space Sci.* **2018**, *21*, S37–S41. [CrossRef]
55. Dhanabalan, S.P.; Rahaman, S.A.; Jegankumar, R. Flood monitoring using Sentinel-1 SAR data: A case study based on an event of 2018 and 2019 Southern part of Kerala. In Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLIV-M-3-2021 ASPRS 2021 Annual Conference, Virtual, 29 March–2 April 2021. [CrossRef]
56. Kalisch, H.; Lagona, F.; Roeber, V. Sudden wave flooding on steep rock shores: A clear but hidden danger. *Nat. Hazards* **2024**, *120*, 3105–3125. [CrossRef]
57. Dodangeh, E.; Panahi, M.; Rezaie, F.; Lee, S.; Bui, D.T.; Lee, C.W.; Pradhan, B. Novel hybrid intelligence models for flood-susceptibility prediction: Meta optimization of the GMDH and SVR models with the genetic algorithm and harmony search. *J. Hydrol.* **2020**, *590*, 125423. [CrossRef]
58. Liu, J.; Wang, J.; Xiong, J.; Cheng, W.; Sun, H.; Yong, Z.; Wang, N. Hybrid Models Incorporating Bivariate Statistics and Machine Learning Methods for Flash Flood Susceptibility Assessment Based on Remote Sensing Datasets. *Remote Sens.* **2021**, *13*, 4945. [CrossRef]
59. Siahkamari, S.; Haghizadeh, A.; Zeinivand, H.; Tahmasebipour, N.; Rahmati, O. Spatial prediction of flood-susceptible areas using frequency ratio and maximum entropy models. *Geocarto Int.* **2018**, *33*, 927–941. [CrossRef]
60. Lee, S.; Rezaie, F. Data used for GIS-based Flood Susceptibility Mapping. *GEO Data* **2022**, *4*, 1–15. [CrossRef]
61. Darabi, H.; Choubin, B.; Rahmati, O.; Haghighi, A.T.; Pradhan, B.; Kløve, B. Urban flood risk mapping using the GARP and QUEST models: A comparative study of machine learning techniques. *J. Hydrol.* **2019**, *569*, 142–154. [CrossRef]
62. Yariyan, P.; Avand, M.; Abbaspour, R.A.; Haghighi, A.T.; Costache, R.; Ghorbanzadeh, O.; Janizadeh, S.; Blaschke, T. Flood susceptibility mapping using an improved analytic network process with statistical models. *Geomat. Nat. Hazards Risk* **2020**, *11*, 2282–2314. [CrossRef]
63. Li, Y.; Osei, F.B.; Hu, T.; Stein, A. Urban flood susceptibility mapping based on social media data in Chengdu city, China. *Sustain. Cities Soc.* **2023**, *88*, 104307. [CrossRef]
64. Pham, Q.B.; Pal, S.C.; Saha, A.; Chowdhuri, I.; Albanai, J.A.; Janizadeh, S.; Ahmadi, K.; Khedher, K.M.; Anh, D.T.; Duan, W. Current and future projections of flood risk dynamics under seasonal precipitation regimes in the Hyrcanian Forest region. *Geocarto Int.* **2022**, *37*, 9047–9070. [CrossRef]
65. Breinl, K.; Lun, D.; Müller-Thomy, H.; Blöschl, G. Understanding the relationship between rainfall and flood probabilities through combined intensity-duration-frequency analysis. *J. Hydrol.* **2021**, *602*, 126759. [CrossRef]
66. Government of Rwanda. Law n°48/2018 of 13/08/2018 on Environment; Government of Rwanda. 2018. Available online: https://rema.gov.rw/fileadmin/templates/Documents/Law_on_environment.pdf (accessed on 2 December 2024).

67. Izonin, I.; Tkachenko, R.; Shakhovska, N.; Ilchyshyn, B.; Singh, K.K. A Two-Step Data Normalization Approach for Improving Classification Accuracy in the Medical Diagnosis Domain. *Mathematics* **2022**, *10*, 1942. [CrossRef]
68. Chen, Y.; Zhang, X.; Yang, K.; Zeng, S.; Hong, A. Modeling rules of regional flash flood susceptibility prediction using different machine learning models. *Front. Earth Sci.* **2023**, *11*, 1117004. [CrossRef]
69. El-Magd, S.A.; Soliman, G.; Morsy, M.; Kharbush, S. Environmental hazard assessment and monitoring for air pollution using machine learning and remote sensing. *Int. J. Environ. Sci. Technol.* **2023**, *20*, 6103–6116. [CrossRef]
70. Henry, D.; Gorman-Smith, D.; Schoeny, M.; Tolan, P. 'Neighborhood Matters': Assessment of Neighborhood Social Processes. *Am. J. Community Psychol.* **2014**, *54*, 187–204. [CrossRef]
71. Warembourg, C.; Nieuwenhuijsen, M.; Ballester, F.; De Castro, M.; Chatzi, L.; Esplugues, A.; Heude, B.; Maitre, L.; McEachan, R.; Robinson, O.; et al. Urban environment during early-life and blood pressure in young children. *Environ. Int.* **2021**, *146*, 106174. [CrossRef]
72. Jiang, T.-B.; Deng, Z.-W.; Zhi, Y.-P.; Cheng, H.; Gao, Q. The Effect of Urbanization on Population Health: Evidence From China. *Front. Public Health* **2021**, *9*, 706982. [CrossRef]
73. Yap, W.; Biljecki, F. A Global Feature-Rich Network Dataset of Cities and Dashboard for Comprehensive Urban Analyses. *Sci. Data* **2023**, *10*, 667. [CrossRef] [PubMed]
74. Ndayishimiye, P.; Uwase, R.; Kubwimana, I.; Niyonzima JD, L.C.; Dzekem Dine, R.; Nyandwi, J.B.; Ntokamunda Kadima, J. Availability, accessibility, and quality of adolescent Sexual and Reproductive Health (SRH) services in urban health facilities of Rwanda: A survey among social and healthcare providers. *BMC Health Serv. Res.* **2020**, *20*, 697. [CrossRef] [PubMed]
75. Jimoh, M.; Bikam, P.; Chikoore, H. The Influence of Socioeconomic Factors on Households' Vulnerability to Climate Change in Semiarid Towns of Mopani, South Africa. *Climate* **2021**, *9*, 13. [CrossRef]
76. Galderisi, A.; Limongi, G. A Comprehensive Assessment of Exposure and Vulnerabilities in Multi-Hazard Urban Environments: A Key Tool for Risk-Informed Planning Strategies. *Sustainability* **2021**, *13*, 9055. [CrossRef]
77. Ganter, M.; Toetzke, M.; Feuerriegel, S. Mining Points-of-Interest Data to Predict Urban Inequality: Evidence from Germany and France. 2022. Available online: www.pricehubble.com (accessed on 24 January 2024).
78. Shrestha, N. Factor Analysis as a Tool for Survey Analysis. *Am. J. Appl. Math. Stat.* **2021**, *9*, 4–11. [CrossRef]
79. Gulum, M.A.; Trombley, C.M.; Kantardzic, M. A Review of Explainable Deep Learning Cancer Detection Models in Medical Imaging. *Appl. Sci.* **2021**, *11*, 4573. [CrossRef]
80. Hall, O.; Ohlsson, M.; Rögnvaldsson, T. A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. *Patterns* **2022**, *3*, 100600. [CrossRef]
81. Fazeli, D.; Zeynab, S.; Khatami, S.M.; Ranjbar, E. The Associations Between Urban Form and Major Non-communicable Diseases: A Systematic Review. *J. Urban Health* **2022**, *99*, 941–958. [CrossRef]
82. Persello, C.; Kuffer, M. Towards Uncovering Socio-Economic Inequalities Using VHR Satellite Images and Deep Learning. In *IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway Township, NJ, USA, 2020; pp. 3747–3750. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.