Available online at www.CivileJournal.org

Civil Engineering Journal

(E-ISSN: 2476-3055; ISSN: 2676-6957)

Vol. 10, No. 10, October, 2024

Intelligent Forecasting of Flooding Intensity Using Machine Learning

Abraham Ayuen Ngong Deng [1], Nursetiawan [1, 2], Jazaul Ikhsan [1, 2*], Slamet Riyadi [2, 3], Ahmad Zaki [2, 3]

[1] *Master Program of Civil Engineering, Universitas Muhammadiyah Yogyakarta, 55183 Yogyakarta, Indonesia.*

[2] *Department of Civil Engineering, Universitas Muhammadiyah Yogyakarta, 55183 Yogyakarta, Indonesia.*

[3] *Department of Informatic Engineering, Universitas Muhammadiyah Yogyakarta, 55183 Yogyakarta, Indonesia.*

Received 28 May 2024; Revised 23 September 2024; Accepted 29 September 2024; Published 01 October 2024

Abstract

This innovative study addresses critical flood prediction needs in Bor County, South Sudan, utilizing machine learning to develop an intelligent forecasting model. The research integrates diverse analytical techniques, including land use analysis and rainfall calculations, with a decade of weather data to understand complex hydrological dynamics. This research employs machine learning classifiers such as Support Vector Machines, Decision Trees, and Neural Networks. Findings reveal promising results, with the Linear SVM classifier achieving 87.5% prediction accuracy for raw data and 100% accuracy for high-velocity flooding events. The Naive Bayes classifier matched this performance, while Artificial Neural Networks showed a slight advantage in runoff estimation. The study's novelty lies in its holistic approach, combining machine learning with advanced visualization tools and geographic information systems. This creates a dynamic, real-time forecasting system bridging sophisticated analysis and practical flood management strategies. Focusing on model interpretability and multi-scale forecasting enhances its value to policymakers and disaster management authorities. This research significantly advances the application of AI to flood prediction and disaster management in offering future studies on humanitarian challenges. By enhancing early warning capabilities, this system substantially reduces flood-related losses and transforms disaster preparedness in vulnerable regions worldwide, potentially saving lives and mitigating economic impacts.

Keywords: Support Vector Machines; Flood Intensity; Rainfall Data; Classification Learners; Confusion Matrix.

1. Introduction

The more significant part of Bor County in the Jonglei State of South Sudan is lowland and tropical, with a tiny area of raised ground in the north.

Because of the region's shallow topography, especially in the southern parts, a sustainable urban drainage system must be implemented to lessen the likelihood of flooding, brought about by the lack of staged evaluation and monitoring systems to address catastrophic situations like river overflowing, which highlights this necessity [1]. The county's susceptibility to floods was brought to light on July 12, 2020, when a flash flood destroyed the surrounding countryside. The dyke along the Nile River burst because of a combination of high rainfall and following conditions. Such events highlight the critical importance of developing environmentally conscious urban waterways and water control systems

* Corresponding author: jazaul.ikhsan@umy.ac.id

http://dx.doi.org/10.28991/CEJ-2024-010-10-010

that can provide ecological services to both metropolitan and natural environments while counteracting the effects of gentrification and ensuring sustainable water flow management. Flooding in Bor County is primarily attributed to sub-factorial etiologic factors, encompassing both artificial and natural causes. These include extensive deforestation and prolonged periods of heavy rainfall that can persist for days, weeks, or even months, leading to the overflow of the Nile River. Periodic heavy rains and elevated flows in large bodies of water typically cause seasonal flood events [2]. At the same time, rapid flooding occurs when water cannot penetrate dry soil or hard surfaces quickly enough. Severe flooding has had a devastating effect on South Sudan, especially Bor County, causing terrible fatalities, extensive displacement, and severe harm to livelihoods. The nation lacks adequate early warning systems that can produce precise forecasts and encourage proactive preparation actions despite the lethal and recurring nature of this hazard. This deficiency is particularly acute in isolated and insecure areas like Bor County, where limited data, resources, and technological capacity pose significant challenges. In response to these pressing issues, this research explores the potential of machine learning techniques in developing life-saving flood forecasting capabilities for vulnerable communities. Using Bor County as a pilot case study, the project aims to create artificial intelligence models that can generate accurate flood intensity forecasts by leveraging existing data on rainfall, river levels, terrain, and other relevant factors [3].

By implementing these cutting-edge ML-enabled intelligent forecasting methods, the study seeks to address traditional approaches' limitations and demonstrate AI's potential to save lives in flood-prone areas of South Sudan and other unstable developing countries. The study's findings should have a significant impact not only on Bor County but also on other areas with comparable problems [4]. This study aims to pave the way for the broader application of machine learning in addressing issues related to sustainable development and climate adaptation for humanitarian assistance on a global level by offering crucial warnings, directing disaster response efforts, and establishing evidence of concept [5]. Using cutting-edge computational methods, Intelligent Forecasting of Flooding Intensity Using Machine Learning is a comprehensive strategy for estimating and evaluating flood hazards [6]. Fundamentally, this approach depends on thoroughly gathering and fusing several data sources, such as historical flood logs, meteorological data, topographical data, and real-time satellite imagery. Subsequently, these diverse inputs undergo intricate feature engineering procedures, which extract pertinent attributes and utilize derived variables to capture intricate correlations among various components. A vital element of this strategy is identifying suitable machine-learning models. Algorithms like random forests and support vector machine architectures are assessed to determine which best suits the flood forecasting problem [7]. To guarantee their robustness and dependability, these models undergo extensive training and validation of historical data while utilizing cross-validation techniques. One of this intelligent forecasting system's main strengths is its capacity to process real-time data and update predictions as new information becomes available. This dynamic adaptation is significant given that typical flood patterns may be changing due to climate change. Additionally, recognizing the intrinsic unpredictability in natural systems, adding uncertainty quantification through probabilistic forecasting adds a crucial degree of nuance to the forecasts. This approach is new in several ways, including better accuracy by capturing intricate, non-linear interactions and multi-scale forecasting ranging from short-term local predictions to long-term regional evaluations. These systems provide a more thorough understanding of flood risk variables by combining many data sources, including cutting-edge inputs like social media data and Internet of Things sensor networks. Machine learning-based flood forecasting is advantageous since it can be customized to specific geographic locations, such as Bor County, to respond quickly in flooding situations [8]. Importantly, because it helps stakeholders make informed decisions, efforts to guarantee the interpretability of model predictions are essential. The forecasts' communicative potential is further enhanced by integrating machine learning predictions with advanced visualization tools and geographic information systems, allowing flood risk information to be effectively disseminated to specialists and the public [9].

This manuscript follows a well-structured academic format, effectively guiding readers through its innovative approach to flood prediction. Beginning with a title that succinctly captures the research focus, the paper progresses through a concise abstract, an introduction setting the context, and a literature review establishing the scholarly foundation. The methodology section, likely detailed with subtitles, outlines the comprehensive approach to data collection and machine learning application. This is followed by results and discussion, where findings are presented and interpreted, highlighting the performance of various algorithms in flood prediction. The paper concludes by summarizing key insights and potential future directions, followed by necessary declarations and a comprehensive reference list. This structure ensures a logical flow of information, from the study's conception to its conclusions. It enables readers to grasp the significance of using machine learning for flood forecasting in vulnerable regions like Bor County, South Sudan.

2. Literature Review

Applying sophisticated algorithms and diverse data sources has allowed machine learning to show tremendous promise in flood prediction. A detailed analysis of using artificial neural networks (ANNs), support vector machines (SVMs), decision trees, and ensemble approaches in flood prediction and mitigation is conducted [10]. The research examines machine learning techniques to predict floods, including more complex models like random forests, confusion matrices, deep learning, and more conventional models like linear regression. Multiple studies aim to create real-time flood prediction models based on machine learning and use rainfall, river flow, and soil moisture data to train these models [11]. Some studies look at machine learning models for flood monitoring and prediction, considering variables like temperature, precipitation, and river discharge data [12]. Hybrid models with several machine learning techniques, such as support vector machines (SVM) and k-nearest Neighbors (k-NN), have demonstrated higher accuracy than single models. Because artificial neural networks can record intricate correlations between input variables and predict future flood outputs, they are commonly employed in flood forecasting [13]. Another supervised machine learning technique, support vector machines, has proven highly effective in flood prediction for both linear and non-linear data correlations. Because they can handle complicated datasets and offer insights into feature relevance, ensemble approaches like Random Forests and decision trees are highly regarded. Decision trees or ANNs combined with SVMs are examples of hybrid models that mix various techniques and have demonstrated enhanced prediction performance [14]. The use of machine learning approaches for early warning systems connected to severe weather-induced flooding is also explored in the literature. Mapping and evaluating the risk of flooding are two aspects of this research. Machine learning techniques are used to assess flood susceptibility, accounting for many factors such as socioeconomic status, elevation, and land use [15].

More thorough and accurate evaluations of flood risk are made possible by combining machine learning with geographic data, hydrological models, and socioeconomic considerations, which makes it easier to create mitigation and adaptation strategies that work [16]. Machine learning has drawn interest in flood control due to its capacity to evaluate enormous datasets, produce forecasts, and assist in decision-making. Research highlights the significance of training models using high-quality datasets that comprise data on terrain, rainfall, river flow, land use, and past flood incidents. Aerial photographs and satellite imaging are remote sensing data frequently used to improve flood prediction models' accuracy. Support vector machines, artificial neural networks, decision trees, random forests, deep learning methods, and convolutional and recurrent neural networks are just a few of the machine learning algorithms investigated for flood prediction [17]. The properties of the data and the necessary degree of prediction accuracy are often considered when choosing a model. Using Bayesian techniques to evaluate uncertainty and offer more dependable decision support, researchers increasingly emphasize comprehending and quantifying the uncertainty underlying flood predictions. When geographical information systems (GIS) and machine learning models are integrated, flood-prone areas can be more thoroughly investigated while considering their geographic context. Machine learning-based decision support systems are being created to help emergency responders organize evacuations, allocate resources, and make real-time choices during flood situations. There are still issues, though, such as the requirement for sizable, identified datasets and worries over the model's interpretability and management of uncertainty. Research is still being done to solve these problems and create machine learning-based flood control systems that are more dependable and durable. The topic of machine learning-based flood management decision support systems is constantly changing. Therefore, reading journal articles and conference proceedings is essential to stay current on the newest advancements [18].

The literature review in a study on intelligent forecasting of flooding intensity using machine learning provides a comprehensive overview of existing research in the field [19]. This review examined traditional flood forecasting methods, highlighting their strengths and limitations. Eventually, the review will continue with how different machine-learning techniques have been applied to water and flood prediction and how these approaches have improved upon traditional methodologies [20]. It would examine machine learning algorithms like neural networks, support vector machines, and random forests that have demonstrated potential in flood forecasting, outlining how they are used and successful in various situations. Additionally, the review would address current challenges faced in flood intensity prediction, setting the stage for identifying research gaps and the proposed approach. Gaps in the literature for such a study might encompass several areas where current research falls short. There could be a limited exploration of advanced machine learning techniques tailored explicitly for flood intensity forecasting, presenting an opportunity for innovation. A significant shortcoming is the need for real-time prediction models capable of rapidly processing incoming data and updating forecasts. Another gap could be the need for more integration of multiple data sources, which prevents a holistic approach to flood prediction. The scientific literature also highlights difficulties in managing flood prediction's temporal and spatial components, which are critical to precise forecasts [21]. Moreover, there is a need to do more research on the interpretability of machine learning models in flood forecasting, which is crucial in establishing these systems' credibility and encouraging decision-makers to use them [22]. The proposed method to close these gaps would be a multipronged plan to propel flood intensity predictions forward. The researchers may present a new machine-learning or ensemble method created to capture the intricate dynamics of flood occurrences. This may involve merging data from multiple sources, such as topographic maps, rainfall measurements, soil moisture content, and land use data, into a single model to provide a more comprehensive view of the factors influencing flood severity. One possible tactic to increase the predictability and timeliness of the outcomes would be to develop a real-time prediction system that can update forecasts as new data becomes available. The researchers could suggest ways to improve the predictability of the machine learning model so that end users can better understand and perceive it to close the interpretability gap. Finally, the strategy may use cutting-edge methods to address the temporal and spatial elements of flooding, guaranteeing that the model can reliably forecast flood severity over a range of temporal and spatial scales [23].

3. Research Methodology

This research methodology encompasses several key components, starting with an explanation of the research location. This foundational step provides context and background for the study area, helping readers understand its geographical and environmental characteristics. Next, the methodology examines land use patterns in Bor County, highlighting the interactions between natural landscapes and human activities. These interactions are crucial as they can significantly influence flood risks and hydrological processes.

Following this, the data collection process is detailed, including the strategies and techniques used to gather relevant information. This phase is critical to ensure the study's validity and reliability, as the quality of data directly affects the research outcomes. The methodology then describes the calculation of areal rainfall, a process that determines the average precipitation over a specific area. This calculation is vital for understanding the region's hydrological dynamics and assessing flood risks.

Subsequently, the design rainfall distribution analysis is addressed, focusing on the temporal patterns of rainfall. This analysis is essential for predicting the intensity and duration of precipitation events, which is critical for effective flood forecasting and management. An innovative approach is then introduced, utilizing machine learning to intelligently forecast flood intensity in Bor County. This advanced method leverages computational algorithms to provide more accurate and efficient predictions compared to traditional techniques.

Finally, the methodology concludes with the classification of model testing outcomes to validate the accuracy of predictions and identify potential areas for improvement. This step evaluates the performance and precision of the machine learning models employed in flood forecasting, ensuring their effectiveness in real-world applications. Figure 1 illustrates the methodology framework, providing a visual representation of the processes and steps outlined in this study.

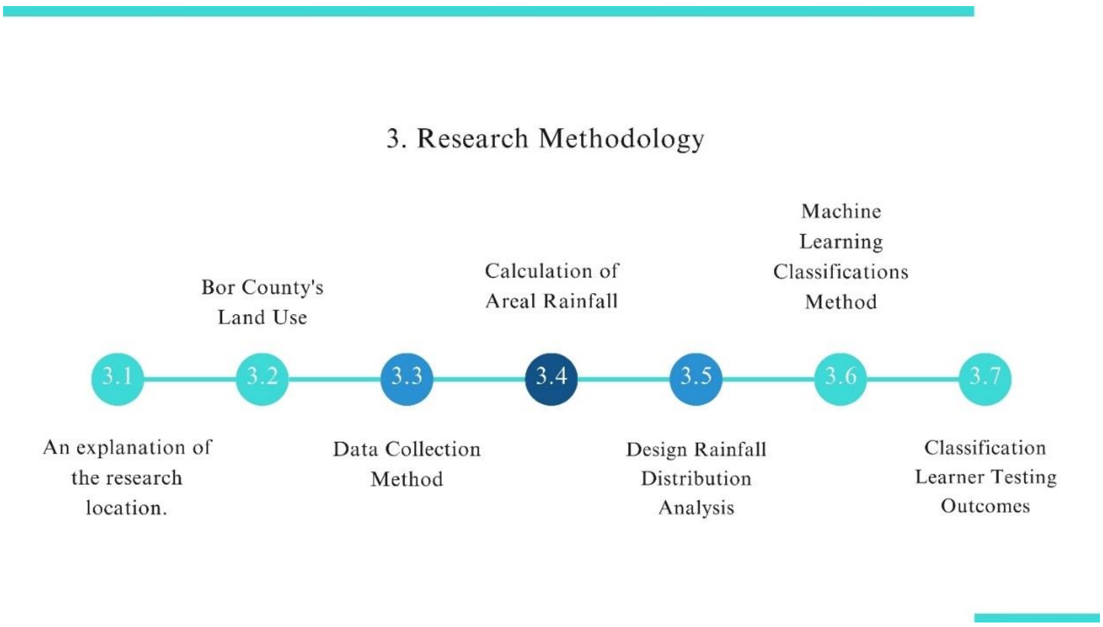


Figure 1. Steps of Research Methodology on Intelligent Forecasting of Flooding Intensity

3.1. An Explanation of the Research Location

The study area is located in Bor County, a town in the northeastern part of South Sudan, within Southern Jonglei State. Situated at an elevation of 434 meters above sea level, Bor's geographic coordinates are 6° 12' 42.45" N and 31° 34' 54.42" E. The total research area spans 22.56 km². Figure 2 illustrates the study area, showing photographs taken in 2013 during a flooding event in Bor, captured by camera operators aboard an aircraft. This map provides a visual representation of Bor's key features, including its natural elements, such as the Nile River, and its infrastructure in Bor Town. The topographical characteristics of the land, including its shape and slopes, are discussed in detail [24].

Bor County's plains are encircled by the Nile River, and the relationship between the land's pitch and its form is also highlighted in Figure 2. Most of the research area consists of lowlands with slightly inclined terrain sloping toward the north. In contrast, the southern part of Bor features shallower geography, where the Pibor River includes a dam known as the Bor County-Pibor Dam. Figure 1 provides an additional overview of the region's geography.

Figure 2. Bor's Areal Perspective

According to the 2008 South Sudanese census, Bor County had a population of 25,216 residents. However, due to conflicts between the government and opposition parties, as well as intraclan warfare, the population has undergone significant changes in recent years. The most recent estimate indicates that Bor County now has a population of 50,316 people. Following intensified fighting between the government and opposition groups on December 15, 2015, more people relocated to Bor South from other parts of the country. Data collected in 2019 shows that the total population in the research area was 50,312, with Mach-deng, a small village, being the most populous, housing 12,344 individuals. The overall population of all villages combined was 3,251. Table 1 provides detailed information about the population and geographic distribution across Bor County.

Table 1. Bor County population and area

Village	Area (squarekm) 1	Residential Population	Population Density
Juet Boma	1.90	3,271	181.7
Mathiang Boma	5.39	8,008	148.8
Machdeng Boma	4.05	4,653	115.5
Mareng Boma	4.41	11,851	491.7
Warkhor Boma	2.97	12,344	310.9
Bor Town	6.21	10,189	317.4
TOTAL	**24.93**	**50,316**	**2,905.1**

Source: Bor County Population from Ministry of Public Works and Housing in Jonglei State

3.2. Bor County's Land Use

The population distribution in Bor County is closely linked to land usage patterns. This study categorizes land use into two phases: the current land use and the projected plan for 2032. As shown in the table, significant changes in land use are anticipated, with 75% of the land transitioning from gardens, rice fields, and vacant land to residential areas, offices, commercial spaces, and educational facilities.

The land use in Bor County follows a "rectilinear" pattern, reflecting the alignment of settlements with the transportation network. The downtown area exhibits moderate intensity and diverse land use due to the presence of vacant land. Figure 3 illustrates the current land usage alongside the areas slated for changes or development under the 2032 land-use plan.

Figure 4 illustrates the infrastructure of South Sudanese cities and emphasizes the importance of land use in Jonglei State. Bor, a city in Jonglei State, serves as the focal point where land use plans were integrated into ArcMap for spatial analysis. Relevant data were gathered to support research efforts in areas such as channel development planning, hydrology, and hydraulics.

The Bor Weather Forecasting, Climatology, and Environmental Conservation Agency (AMCECA, 2021) provided essential data, including topographic maps, channel dimensions and conditions, and average daily precipitation. These data are based on annual rainfall records collected from multiple sites between 2012 and 2021.

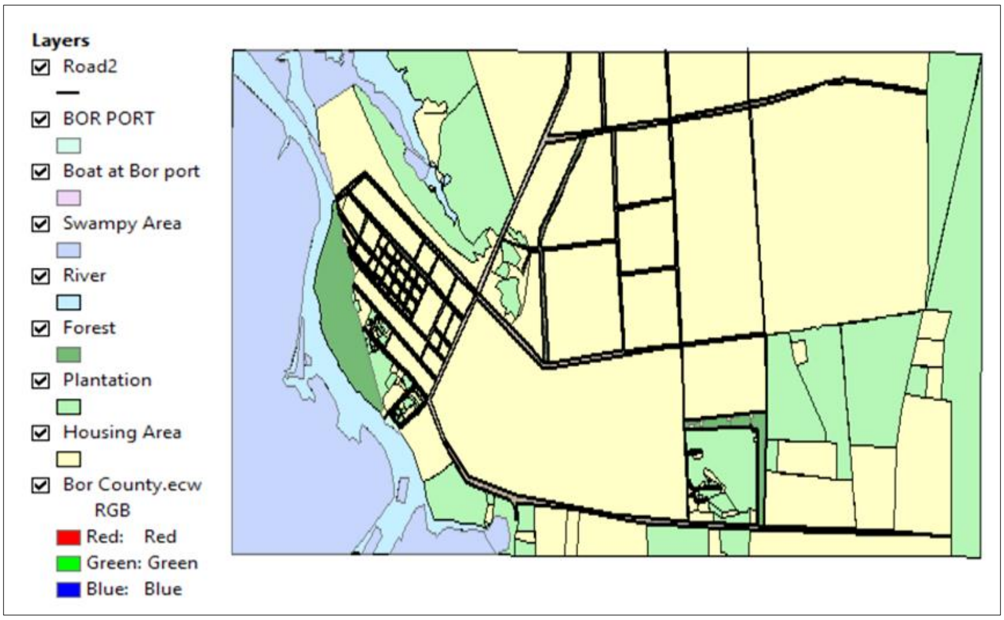


Figure 3. Current Bor Town Map Land Use

Figure 4. Map of South Sudan showing Jonglei State (Bor City) and other states

3.3. Data Collection Method

Two methods were used to collect the data for the study: field surveys, literature reviews, and other secondary data collection methods. The essential secondary and primary statistics are presented below, based on information from pertinent departments and government agencies. The rain list must include data since the Meteorology and Geophysics Center occasionally receives partial rain measurement reports from rain observation stations. The adapted design approximates the gaps in the data to close the gaps in the data. The estimate is predicated on complete rainfall data from nearby and surrounding observation stations. Information on precipitation for the last ten years from 2012- 2021. This study uses a digital terrain model (2012-2021) and topographic maps to examine rainfall data for sustainable urban drainage systems.

3.4. Calculation of Areal Rainfall

Figure 5 illustrates annual rainfall, representing the total precipitation over a given area. Sustainable Urban Drainage Systems (SUDS) are designed to manage surface water runoff in urban areas in an environmentally friendly and sustainable manner. SUDS employ various strategies, including swales, detention basins, permeable pavements, and green roofs, to mitigate surface water runoff. These methods allow water to infiltrate the ground, be absorbed by vegetation, or be temporarily stored, reducing runoff effectively.

Annual rainfall plays a crucial role in planning SUDS, as it determines the volume of surface water runoff that must be managed. Higher rainfall in a given area results in an increased need for runoff control. To address this, SUDS must be designed to handle varying rainfall intensities, from light showers to heavy downpours. This requires careful consideration of factors such as the size and capacity of SUDS features, soil permeability, and the region's expected rainfall patterns.

In summary, annual rainfall is a vital factor in SUDS design, as it dictates the amount of surface water runoff requiring management. By incorporating strategies for infiltration, absorption, and storage, SUDS aim to reduce surface runoff in a sustainable and efficient manner. Properly designed SUDS enhance the resilience and sustainability of urban areas, helping them adapt to changing weather patterns and rainfall intensities.

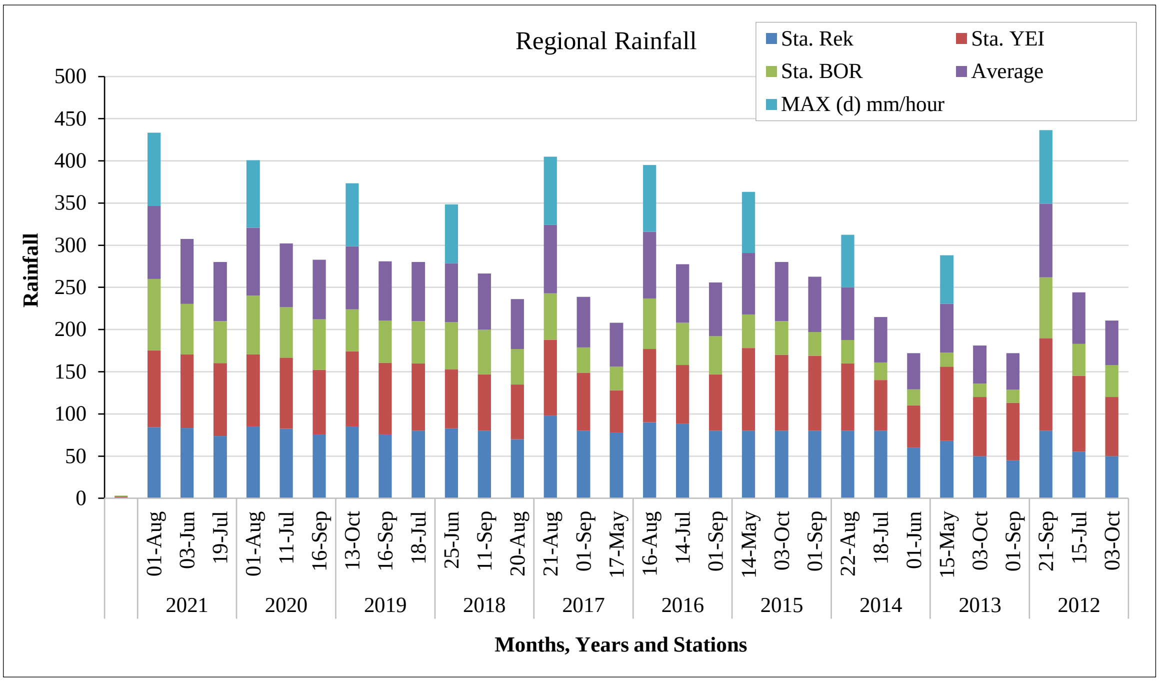


Figure 5. Maximum Regional Rainfall

It tracks the maximum daily rainfall over time at three representative weather stations - Rek, Yei, and Bor. This gives a comprehensive overview of extreme rainfall events across the region. The average and maximum columns condense the data into valuable metrics to see yearly peak rainfall levels. This makes it easy to visualize increasing or decreasing trends. The three-year average columns show if extreme rainfall events are becoming more frequent and intense. For Bor County, this helps assess rising flood risks. The Bor station data is critical for Bor County flood analysis, representing local conditions. Years like 2012 and 2016, when Bor had very high peak rainfall over 110-72 mm, show Bor County's vulnerability. Comparing Bor data to Rek and Yei helps contextualize if rainfall events are localized or widespread events that increase flood risk. The date and total amount of rainfall at each station on the day with the highest rainfall for each year are displayed in the above graph. Additionally, it determines the annual rainfall totals for the entire region. The highest daily rainfall recorded during the year at all three sites is the maximum regional rainfall. For instance, on August 1, 2021, Rek Station reported 87 mm/hour of rainfall.

The averages depict the general rainfall patterns; some years are wetter than others. The maximum regional rainfall column, which helps assess helpful for assessing flood risk, displays the highest annual daily rainfall in 2012; the maximum rainfall was 110 mm/hour. Based on the interpretation of this graph from the three long-term weather stations, the graph summarizes the region's extreme rainfall events over the last ten years. The rainfall statistics in the graph may specifically assist in lowering the risk of floods in the Bor area in the following ways: Food defenses are significant to Bor, which continuously receives less rainfall than Rek and Yei stations, according to the data. It still experiences occasional extremely high rainfall, such as 72 mm per hour in 2012. Overall, having long-term organized data on peak regional rainfall helps Bor County officials assess flood susceptibility in different areas, plan infrastructure like flood drains, issue timely warnings during heavy rain, and be prepared for emergency response. It is a crucial reference to reduce loss of life and property from flood damage. Continuous tracking and analysis of this data should be a priority for Bor County shortly.

3.5. Design Rainfall Distribution Analysis

In the study of hydrology, probability distributions are often employed to evaluate and predict precipitation events. The amount of rain expected to fall at a specific return time is known as the design rain value, and it can be determined using the probability distribution approach. Two joint probability distributions employed in frequency analysis were the Gumbel distribution and the Log Pearson III distribution. The rainfall frequency analysis shows that. Figure 6 can compute the minimal deviation and find the acceptable rainfall distribution. It contains the design rainfall values for return periods of 1, 2, 5, 10, 25, 50, and 100 years. The anticipated time frame within which rainfall of a particular amount is expected to occur or be exceeded is known as the return period.

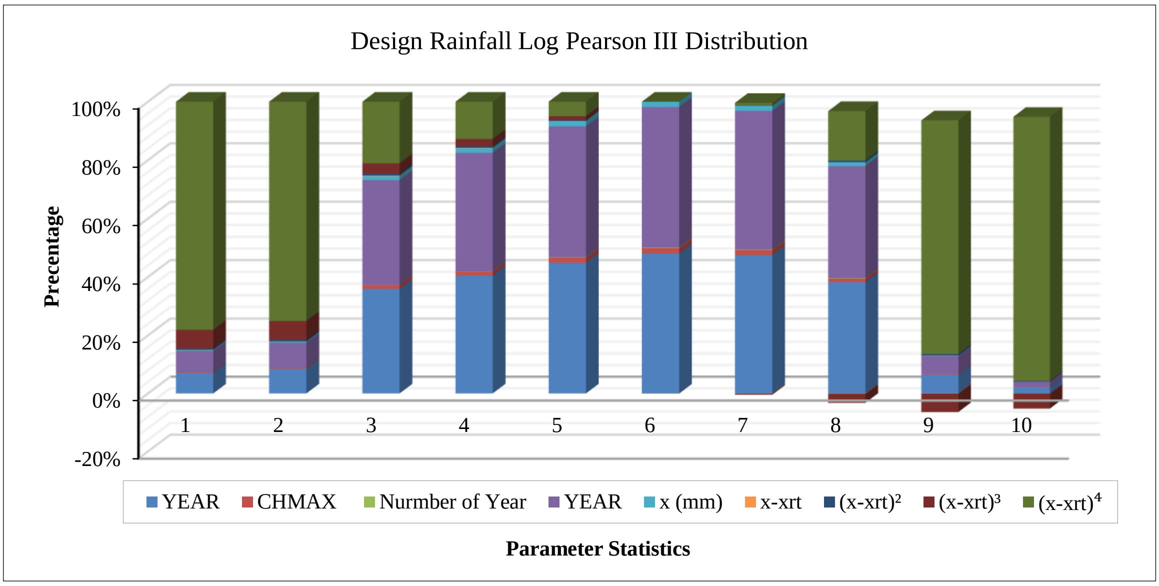


Figure 6. Design Rainfall Log Pearson III Distribution

Figure 6 shows how the rainfall distribution analysis graph can help reduce flooding risk in Bor County: The graph presents a statistical analysis of maximum annual rainfall data (CHMAX) for Bor County, spanning ten years from 2021 to 2012. Tracking the maximal yearly rainfall over this multi-year timeframe provides a representative sample to assess flood risk and design appropriate flood control infrastructure. The analysis quantifies above-average and below-average annual peak rainfall by calculating the deviation of each year's CHMAX from the 10-year average maximum (X-Xrt). Squaring and cubing these deviations progressively amplifies the impact of intense rainfall years in the statistical calculations. This effectively focuses the distribution on years with heavy rainfall that require greater flood management capacity. The larger the positive deviation, the more that year's rainfall intensity must be accommodated by increased drainage, retention basins, levees, and other flood controls. This way, the tailored statistical distribution models the rainfall input for flood risk analysis. It ultimately enables the appropriately sized flood infrastructure design to withstand worst-case rainfall events and reduce flooding hazards in Bor County.

The Log Pearson Type III distribution in Figure 7 is a statistically generated model that effectively forecasts the likelihood of extreme rainfall events resulting in flooding. Extrapolation beyond the measured data points is possible by linearly fitting the Pearson Type III distribution by converting the recorded annual maximum rainfall data to a logarithmic space. Based on the small 10-year rainfall dataset for Bor County, the computed empirical exceedance probabilities explicitly assess the likelihood of previous flood-level precipitation. These empirical probabilities can be smoothed out, and even more violent storms that were not directly witnessed but can occur in the future might be predicted probabilistically by fitting a theoretical distribution. The chi-squared and absolute probability difference metrics confirm a sufficient fit between the theoretical distribution and the observed data. The likelihood of various rainfall amounts being surpassed is modeled by this validated flood frequency curve, providing hydrologic design specifications for infrastructure capacity that can withstand floods within a given return period. Engineers can reduce flooding risks to levels the county can tolerate by designing and implementing flood control measures based on the fitted Log Pearson distribution. These methods include appropriately sizing drainage canals, retention basins, levees, and other systems to handle statistically expected extreme storms. The computed fit of the Log Pearson Type III model offers the statistical quantification of heavy precipitation occurrences required for data-driven flood analysis and infrastructure design aimed at successfully reducing flood risk to Bor County. Fitting a Log Pearson Type III distribution allows for the statistical modeling of extreme rainfall events based on sparsely observed data, which is necessary for the hydrologic design of flood control infrastructure in Bor County to manage runoff better and reduce the risk of inundation, and this used to enable reliable prediction of intense storm magnitudes [25].

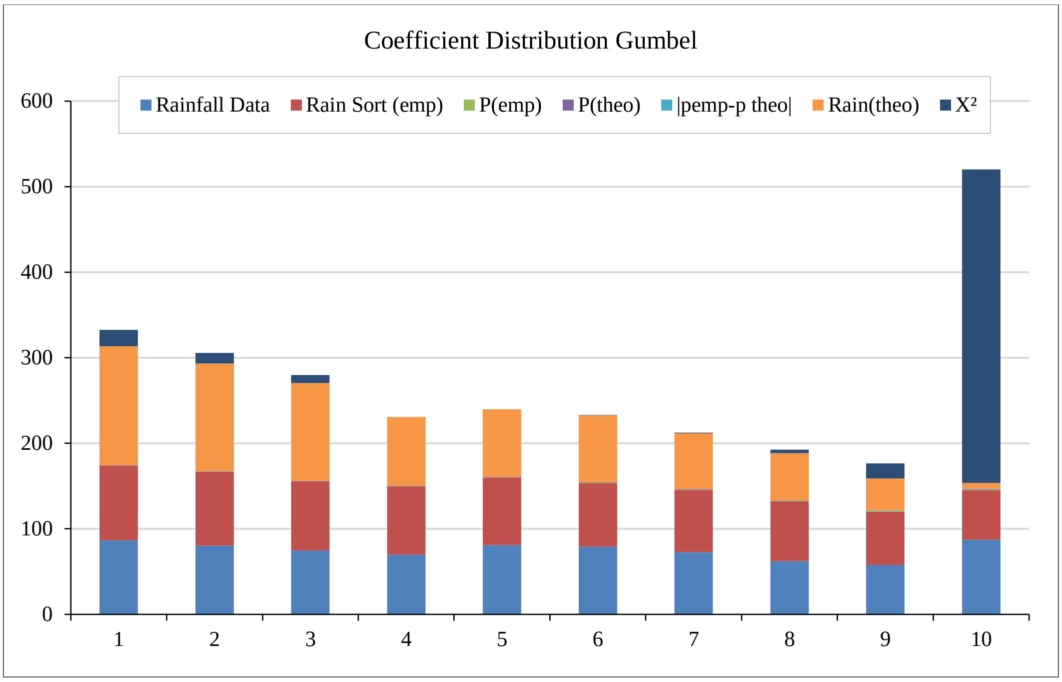


Figure 7. Coefficient Distribution Gumbel Method

3.6. Classification Learner Method

The process of categorizing data collection is called the purpose of classification, which can be applied to ordered or unorganized data. It also projects the classes of data points in which classes can be referred to as objectives, designations, or groups. Classifying data involves determining which kind the incoming data will belong to. Therefore, a classification model makes predictions or inferences about the class or category of the data based on the input data used for training. One helpful tool in MATLAB is the classification learner, which includes the ensemble trees, decision tree, Confusion matrix, k-nearest-neighbor models (KNNs), support vector system (SVM), and some essential classifiers [26]. It also simplifies the completion of supervised learning activities, including feature selection, dynamic data investigation, training model creation, validation scheme development, and outcome evaluation. Several classifiers can be used for supervised machine learning by giving them an established set of data (records) and a recognized output (responses) as class labels. The pathways of a new dataset can be predicted using the exported trained model. The Statistics and Machine Learning Toolbox in MATLAB includes the Classification Learner application. It offers an interactive interface to train and compare different machine-learning classification models. This tool walks users through the various options and procedures involved in the workflow, making the process of selecting and training a categorization model easier [27].

Table 2 shows that Bor, Yei, and Renk provided the information used to classify the data. The amount of rainfall in each sample was calculated to assess the potential for floods. Table 3 indicates that 210 samples were used in this investigation. The training and testing data were divided into two categories. After training on 143 data samples, the classification model was tested using multiple datasets to evaluate its prediction accuracy. Table 1 presents the four categories used to classify the rainfall data: heavy flood, medium flood, mild flood, and no flood. The sample data distribution is shown in Table 4, which is based on the rainfall data class limits. The data distribution, consisting of 143 training samples, was determined using rainfall calculations. Out of the 143 samples, 41 had moderate flooding, 26 had mild flooding, 134 had intense flooding, and 0 had no flooding. The four test samples were categorized as follows: one sample had a heavy flood state, another had a medium flood state, and the remaining two samples had mild and no flood states. A portion of the data used to build the training model is presented in Table 3. Table 2 also shows the input parameters used by the classifiers, which include location, years, months, total rainfall, supply 1, period 2, period 3, maximum precipitation, rainy days, flood percentage, and flooding status. The model's output is the value corresponding to the rainfall data, with the final row representing the amount of rainfall, which applies to each subclass.

Table 2. Training data samples

Location	Year	Months	The sumof the rainfall	Period1	Period2	Period3	Max Rainfall	Rain day	Percent of Flood	Condition of the Flooding
Renk	2012	JAN	0	0	0	0	0	0	0%	no floods
Renk	2012	PEB	0	0	0	0	0	0	0%	no floods
Renk	2012	MAR	14	8	0	4	5	5	1.0%	Low floods
Renk	2012	APR	16	4	6	6	4	4	1.9%	Low floods
Renk	2012	MAY	88	72	33	50	10	10	35.6%	medium Floods
Renk	2012	JUN	173	120	14	50	9	9	78.8%	Heavy floods
Renk	2012	JUL	200	120	83	50	13	13	79%	Heavy floods
Renk	2012	AGS	174	90	35	50	4	4	75%	Heavy floods
Renk	2012	SEP	339	175	151	80	11	11	88%	Heavy floods
Renk	2012	OCT	255	108	41	55	14	14	85%	Heavy floods
Renk	2012	NOP	20	4	9	6	5	5	9.6%	Low floods
Renk	2012	DEC	0	0	0	0	0	0	0%	no floods
Renk	2013	JAN	4	2	2	0	2	2	1.9%	Low floods
Renk	2013	PEB	0	0	0	0	0	0	0%	no floods
Renk	2013	MAR	8	4	2	2	4	4	2.9%	Low floods
Renk	2013	APR	61	15	25	21	10	11	20%	Low floods
Renk	2013	MAY	104	62	40	2	15	11	35%	medium Floods
Renk	2013	JUN	53	14	8	31	20	8	30%	Low floods
Renk	2013	JUL	130	24	45	61	45	6	45%	medium Floods
Renk	2013	AGS	170	125	15	30	50	9	64%	Heavy floods
Renk	2013	SEP	234	116	88	30	68	8	70%	Heavy floods
Renk	2013	OCT	166	75	62	29	40	10	68%	Heavy floods
Renk	2013	NOP	18	6	6	6	6	6	4.8%	Low floods
Renk	2013	DEC	8	5	0	3	4	3	3.8%	Low floods
Renk	2014	JAN	0	0	0	0	0	0	0%	no floods

Table 3. Distribution of the rainfall data sample

Condition of the Floods	Training of Sample Data	Testing Sample Data
Levels of Floods	**Data Training Numbers**	**Data Testing Numbers**
Heavy floods	143	0.9929
Medium flood	41	0.9999
Low floods	26	0.957
No floods	0	1
Training datasamples	210	3.9498

Table 4. Class limit prediction

Floods range Percentage		Predicted class			
		Heavy Floods	**Low Floods**	**Medium Floods**	**No Floods**
Floods> 100	Heavy floods	0	1	1	1
0<Floods<50	Low floods	1	0	1	1
50<Flood<100	Medium Floods	1	1	0	1
0	No Floods	1	1	1	0

The data in Table 5 was used exactly as it was tested, with no changes made. Each value in every column was divided by the highest value in that column to standardize the data further. The accuracy of each classifier in classification was determined using both the normalized and raw testing data samples within the Classification Learner tool in MATLAB, as shown in Table 3. The forecast precision of each classifier, including accuracy and total cost validation, is presented in Table 5. Machine learning analysis indicates that the linear kernel Support Vector Machine (SVM) provided excellent predictive performance regardless of whether the data was normalized, with the prediction accuracy of the training data at 87.5%. In contrast, the prediction accuracy of the normalized data was 68%. On the other hand, the second-best prediction accuracy on the training data was achieved by the coarse tree classifier, with a forecast accuracy of 81.39% for both standardized and unstandardized data. Based on the results of the accuracy tests, the newly constructed linear regression model was used as part of the algorithm to evaluate incoming data and assign it to a class.

Table 5. A comparison of various classifiers

Machine Learning Classifiers for Trained Data		Total Cost Validation
Classifier	**Accuracy (Validation) RawData Percentage**	**Normalizationn Total cost (Validation)**
1. Decision Tree (DT)		
Fine tree	81.11%	68
Medium tree	81.11%	68
Coarse tree	81.39%	67
Linear discriminant	81.39%	67
Quadratic discriminant	81.11%	68
2. Support Vector Machine (SVM)		
Linear SVM	87.5%	68
Quadratic SVM	81.11%	68
Cubic SVM	81.11%	67
Fine Gaussian SVM	63.61%	131
Medium Gaussian SVM	81.39%	67
Coarse Gaussian SVM	78.33%	78
3. Neural Network Classifiers (NNC)		
Neural Network (NN)	79.44%	74
Medium Neural (MN)	80.56%	70
Wide Neural (WN)	79.72%	73
Bilayer Neural (BN)	79.17%	75
Trailered Neural (TN)	80.28 %	71
All Neural Network (ANN)	81.11%	68
4. Ensemble Trees		
Ensemble Boosted Trees	39.72%	217
Ensemble Bagged Trees	81.39%	67
Ensemble SubspaceDiscriminant	81.11%	68
Ensemble Subspace KNN	81.39%	67
Ensemble RUS Boosted Trees	81.11%	68

For several reasons, the prediction may be off because of the inclusion of relevant and insufficiently instructive elements. The classifiers thus need help to generate accurate forecasts because of a discrepancy in the training data; one class has significantly more examples than the others. The machine learning system may learn too much from its initial data and be able to generalize to new data if the method is simple and if there is noise in the training set. The classification method may be overly simplistic and fail to identify the underlying trends in the data, whether the model needs to be more complex or if the amount of training data is inadequate [28].

4. Results and Discussion

4.1. Training Outcomes Based on the Raw Data

The article discusses a study on flood prediction using various machine-learning classifiers. The linear Support Vector Machine (SVM) emerged as the best classifier, achieving a maximum predictability of 87.5%. This high accuracy suggests that the linear SVM is particularly effective at distinguishing between different flood conditions based on the given predictors. Figure 8 presents a data analysis of flood prediction of the raw data analysis, showing different colors representing correct predictions, while short bars indicate incorrect ones. This visual representation allows a quick assessment of the classifier's performance, showing the distribution of correct and incorrect predictions across the dataset. The study utilized eleven predictors to calculate rainfall data, including location, year, total rainfall for three different periods, precipitation days, percentage of flooding, and flooding conditions. With 210 data samples in the preliminary set, this comprehensive approach ensures a robust dataset for training and testing the classifiers. The response classes are categorized into four groups: no floods (0), heavy floods (143), medium floods (41), and low floods (26). This classification system provides a nuanced view of flood severity, allowing for more precise predictions and risk assessments. The paper employed double verification to ensure the reliability of the training procedure and its outcomes. They divided the data into 17 groups, with 16 as training sets and the remaining as a validation group. This cross-validation approach helps to minimize overfitting and provides a more accurate assessment of the model's performance across different subsets of the data. The analysis in MATLAB showcases the model's predicted 2000 observed /second speed with a test time of 44.983 seconds. This information gives insight into the computational efficiency of the model, which is crucial for real-time flood prediction applications. Figure 8 illustrates the classifier's performance on training data using a color-coded confusion matrix. Blue cells represent significant proportions of correct predictions, while orange cells indicate areas where the classifier's performance could be improved. The intensity of the color corresponds to the percentage of accurate predictions, providing a visual guide to the classifier's strengths and weaknesses.

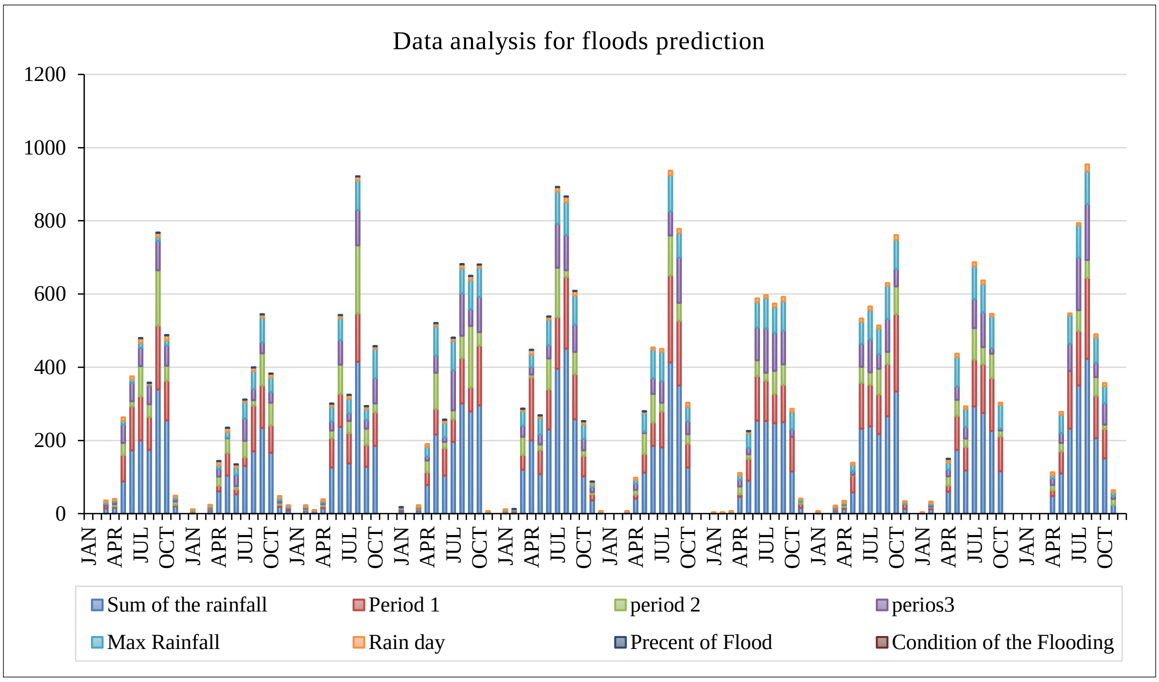


Figure 8. Data Analysis for Flood Prediction

The confusion matrix in Table 3 and Figure 9 offers a detailed breakdown of the classifier's performance. For Class 1 (heavy floods), the linear SVM achieved 34% accuracy, correctly predicting all 143 samples of high-velocity flooding. This exceptional performance for severe flood events is precious for early warning systems and emergency preparedness. However, the classifier showed some limitations in distinguishing between no-flood and low-flood conditions. It achieved 59% accuracy for no-flood samples, correctly identifying 34 out of 39 specimens. Classifying some low-flood samples as no-flood events suggests further improvement in the model's sensitivity to significant flooding events. In this paper, the study demonstrates the potential of machine learning techniques, particularly linear SVM, in flood prediction. The high accuracy achieved for severe flood events is promising for practical flood management and risk assessment applications. The results show more accuracy; however, some areas require improvement, particularly in differentiating only heavy flood circumstances. This study offers significant contributions to the fields of hydrological modeling and preparedness for disasters [29].

Figure 10 Provides a data analysis of flood predictions, detailing four categories: Heavy Floods, Low Floods, Medium Floods, and No Floods. The percentages represent the distribution of actual classes within these categories. The analysis shows that the value 100% for heavy floods indicates that this category occurred with complete certainty in the dataset, while "0.446" for low floods signifies a 44.6% occurrence rate. Medium Floods have a 70% rate, and No Floods have no occurrences. The predicted class percentages indicate the model's output for each category. A prediction rate of "0.243" which is 24.3% for Medium Floods, indicates a low prediction rate, "0.554" for Heavy Floods indicates a 55.4% probability that the model predicted this class, "5.4" which is 5.4% for Low Floods indicates a much higher expected rate than actual, suggesting potential overestimation, and "1" which is 100% for No Floods indicates a perfect prediction. The percentage of actual positives the model accurately detected is called the True Positive Rate, or TPR. A TPR of "1," which is 100%, denotes flawless identification for heavy flooding and no flooding, whereas "0.554," which is 55.4%, and "0.243," which is 24.30%for low and medium flooding, respectively, imply moderate and poor identification, rates. The False Negative Rate (FNR), inversely related to TPR, shows the rate of actual positives incorrectly classified as negatives. A value of "0" for Heavy Floods and No Floods indicates no false negatives, implying perfect detection. However, "0.446" for Low Floods, which is 44.6 %, and "0.757," which is 75.7% for Medium Floods, suggest significant misclassifications, highlighting areas for model improvement. This analysis offers insights into the model's strengths and weaknesses in flood prediction, showcasing areas of high accuracy and potential improvement.

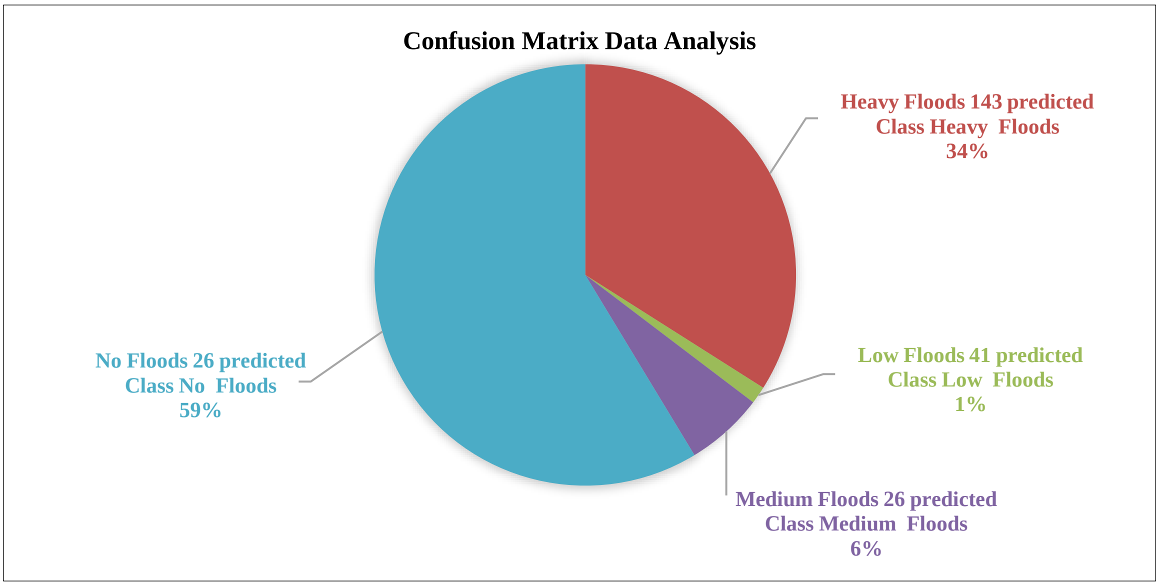
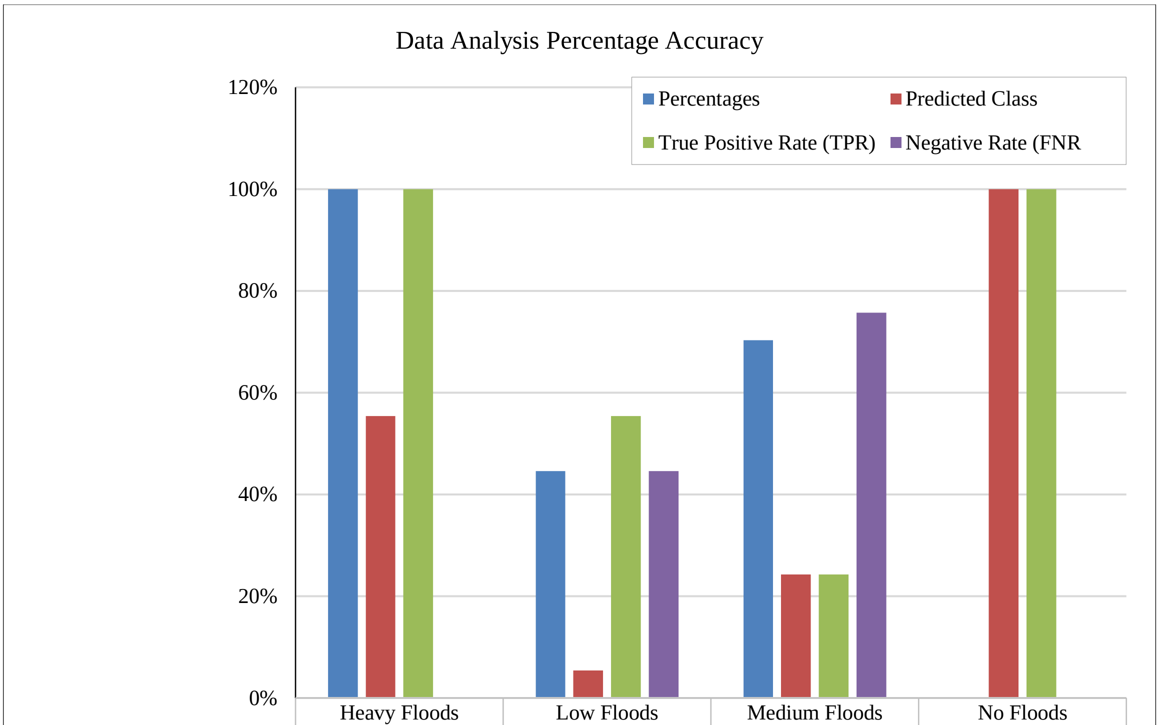


Figure 9. Confusion Matrix Data Analysis



	Heavy Floods	Low Floods	Medium Floods	No Floods
Percentages	100%	44.60%	70.30%	0%
Predicted Class	55.40%	5.40%	24.30%	100%
True Positive Rate (TPR)	100%	55.40%	24.30%	100%
Negative Rate (FNR	0%	44.60%	75.70%	0%

Figure 10. Confusion Matrix Data Analysis

The confusion matrix for the raw data, shown in Figure 11, provides a detailed analysis of both correct and erroneous predictions. This figure highlights the accuracy of the predictions and the estimation of error rates. The analysis revealed that the model correctly predicted the outcome for class 4 (no flooding) ten times. One case showed an error rate of

3.8% for class 2 (medium floods), indicating high accuracy (96.2% to 100%) for the other predictions. Additionally, there were fifteen instances of heavy floods being predicted, with thirteen correctly identified, resulting in a positive predictive value (PPV) of 68.1%. However, two samples had an error rate of 12.4%, being mistakenly predicted as medium floods. Figure 11 further illustrates the model's performance by displaying the distribution of predictions. The colors in Figure 11 range from blue to dark orange, indicating the accuracy of forecasts, with blue representing accurate predictions and dark orange representing errors. For example, the trained linear Support Vector Machine (SVM) model accurately predicted every class 1 (heavy floods) sample with 100% accuracy for high-velocity flooding samples. However, for low flood conditions, there was a misclassification rate for 51 out of 210 samples, although no-flood samples were correctly identified with 88% accuracy. Table 3 summarizes the overall evaluation of the model during the training phase using 210 samples. The trained model's predictions were assessed for accuracy, with an updated test showing that all 11 samples were correctly classified, achieving an accuracy rate of 96.2% for accurate class predictions. The confusion matrix analysis in Table 6 shows the predicted class results for various flooding conditions (heavy, low, medium, no floods) and metrics like PPV and False Discovery Rate (FDR). The data indicate that the model exhibits varying degrees of accuracy across different flood levels, with strengths in predicting no floods and some inaccuracies in distinguishing between medium and low floods.

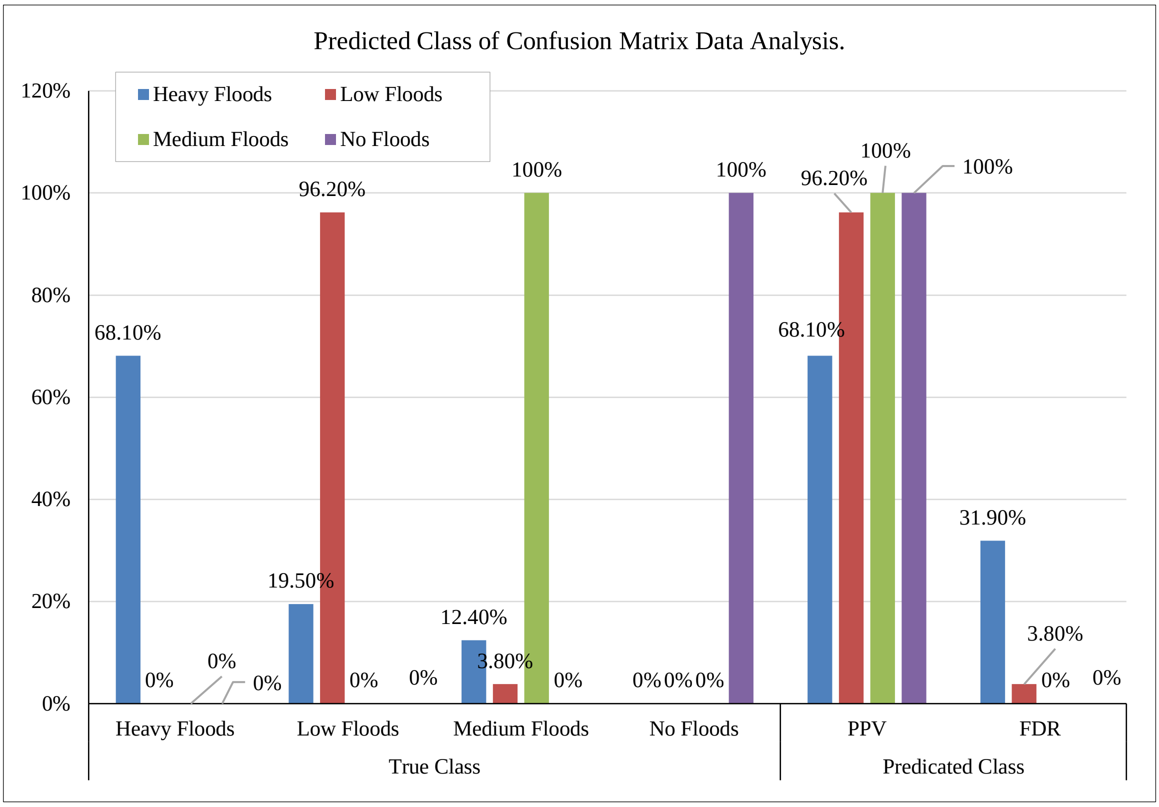


Figure 11. Predicted Class of Confusion Matrix Data Analysis

A ROC (receptor operating characteristic in the bar chart, which illustrates the efficacy of four machine learning models in forecasting flood episodes, is shown in Figure 12. The ROC curve bar chart evaluates the diagnostic ability of binary classifiers, such as these machine learning models, by plotting the actual positive rate (sensitivity) versus the false positive rate at various threshold values. The ability to illustrate the trade-off between actual positive events that are accurately detected and false positive events that are incorrectly identified is essential for assessing the model's effectiveness. The ROC bar chart illustrates the models' ability to distinguish in this context between heavy flood events (classified as class 3) and no flood events (classified as class 1). The shape of the bar chart and the area under the curve (AUC) are two crucial indicators of model success. Higher AUC values indicate more excellent model performance. The value ranges from 0 to 1. An AUC of 1, which means perfect class separation without false positives or negatives, would characterize an ideal model. The four models whose performances are highlighted in Figure 12 are the heavy flood model, low flood model, medium flood model, and no flood model. In that order, the AUC values for these models are 1, 0.8776, 0.8896, 0.7922, and 0.8776. With a perfect AUC of 1, these data suggest that the no-flood model has the highest accuracy, perfectly differentiating between low-flood and no-flood events. With an AUC of 0.8776, the heavy flood model also exhibits high predictive accuracy, trailed by the medium flood model (0.7922) and the Heavy flood model (0.8896). The model performs better when the ROC curve bar chart is closer to the upper left corner, showing higher sensitivity and specificity and fewer false alarms. Each curve's specified operating points show the ideal threshold for each model, striking a balance between specificity and sensitivity.

The critical choice of operational points determines the trade-off between true and false favorable rates, which might vary based on the application and the implications of false alarms versus missed occurrences. For this research, having a higher sensitivity in flood prediction may be more important, as missing a significant flood occurrence can have dire repercussions. Because it offers a threshold-independent metric for comparing various models, ROC bar chart analysis is crucial for assessing machine learning-based flood prediction algorithms. The model's ability to distinguish between multiple classes over all feasible thresholds is summarized by a single number, the AUC. One can choose the best model and operating point by examining the ROC curves based on the requirements and application scenario. In conclusion, the ROC curve and the corresponding AUC values are essential for evaluating how well machine learning models predict floods. They offer a thorough and lucid comprehension of the models' functionality, assisting in choosing and refining models for real-world uses. The conversation emphasizes how crucial it is to have precise flood-predicting systems, which may be attained by carefully examining and interpreting ROC curves.

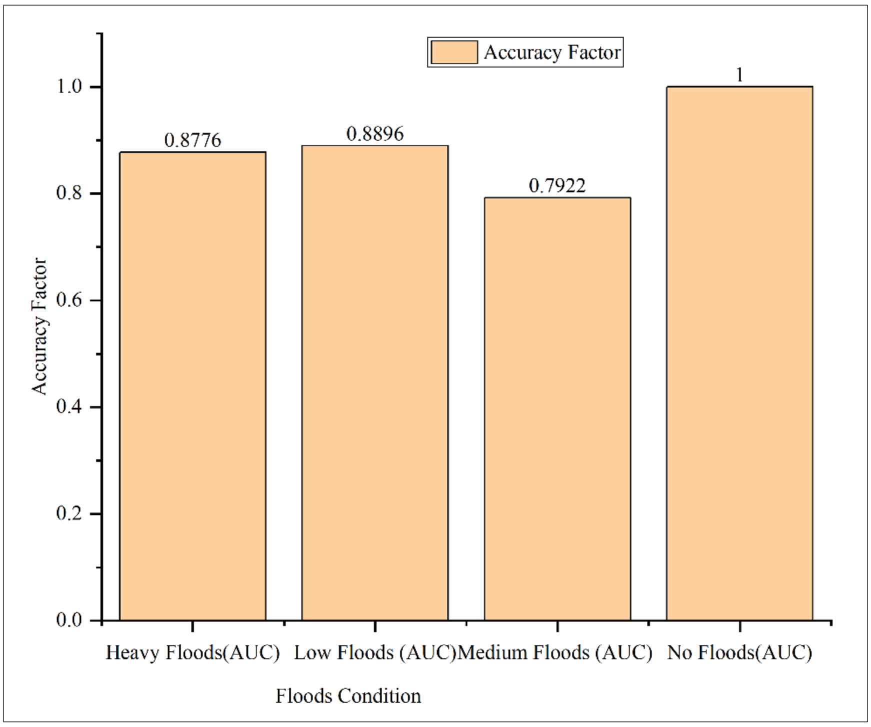


Figure 12. ROC curve comparing class 3 to class 1, the antagonistic class, and the positive class

Table 6 comprehensively compares various classifiers used for the intelligent forecasting of flooding intensity. The accuracy of each classifier and the quantity of data points are used to gauge how well it performs. The assessed classifiers were Decision Trees, Neural Networks, Ensemble approaches, Support Vector Machine (SVM), Logistic Regression, and Naive Bayes. With 81 iterations, Logistic Regression displays a decent accuracy of 77.50%, suggesting a fair trade-off between predictability and simplicity. The Linear SVM classifier slightly outperforms Logistic Regression, achieving an accuracy of 78.61% with fewer iterations (77), suggesting its effectiveness in handling linear decision boundaries. The Naive Bayes classifier demonstrates the highest accuracy among the classifiers, with the two models achieving 87.50% and 87.22% accuracy across 45 and 46 iterations. This high accuracy indicates that the model's strong feature independence assumption suits the dataset.

The SVM classifier performs differently and can handle both linear and non-linear data. The accuracy varies between 77.50% and 80.00% in several iterations, demonstrating its adaptability and resilience to various levels of data complexity. The accuracy of ensemble approaches differs significantly, from a low of 39.72% with iterations of the highest (217) to a high of 79.72%. The wide range suggests that the effectiveness of ensemble techniques heavily depends on the specific models and their combinations. Neural Networks display consistent performance, with accuracies hovering around 78.89% to 79.17% over multiple iterations. This consistency highlights the model's capacity to learn and generalize from complex patterns in the data. Kernel-based methods have comparable performance to Neural Networks, achieving an accuracy of up to 79.44%. They are beneficial in handling non-linear relationships in the data. Decision Trees show promising results, with accuracies consistently around 80.00% to 80.28% over multiple iterations. Their performance indicates they can effectively capture and interpret the data's structure. The Naive Bayes classifier is generally the most accurate, with Decision Trees and SVMs coming in second and third. In intelligent forecasting systems, classifier selection is crucial since it directly affects the prediction quality and, in turn, the flood management decision-making process [30].

Table 6. A comparison of various classifiers

Efficient LogisticRegression	Trained	77.50 %	81
Linear SVM	Trained	78.61 %	77
Naive Bayes	**Trained**	**87.50 %**	**45**
Naive Bayes	Trained	87.22 %	46
SVM	Trained	79.72 %	73
SVM	Trained	80.00 %	72
SVM	Trained	79.72 %	73
SVM	Trained	77.78 %	80
SVM	Trained	79.17 %	75
SVM	Trained	77.50 %	81
Ensemble	Trained	39.72 %	217
Ensemble	Trained	79.72 %	73
Ensemble	Trained	79.44 %	74
Neural Network	Trained	78.89 %	76
Neural Network	Trained	78.89 %	76
Neural Network	Trained	79.17 %	75
Neural Network	Trained	78.61 %	77
Neural Network	Trained	79.17 %	75
Kernel	Trained	79.44 %	74
Kernel	Trained	78.89 %	76
Naive Bayes	Trained	87.50 %	45
Tree	Trained	80.00 %	72
Tree	Trained	80.28 %	71
Tree	Trained	80.28 %	71
Tree	Trained	80.28 %	71

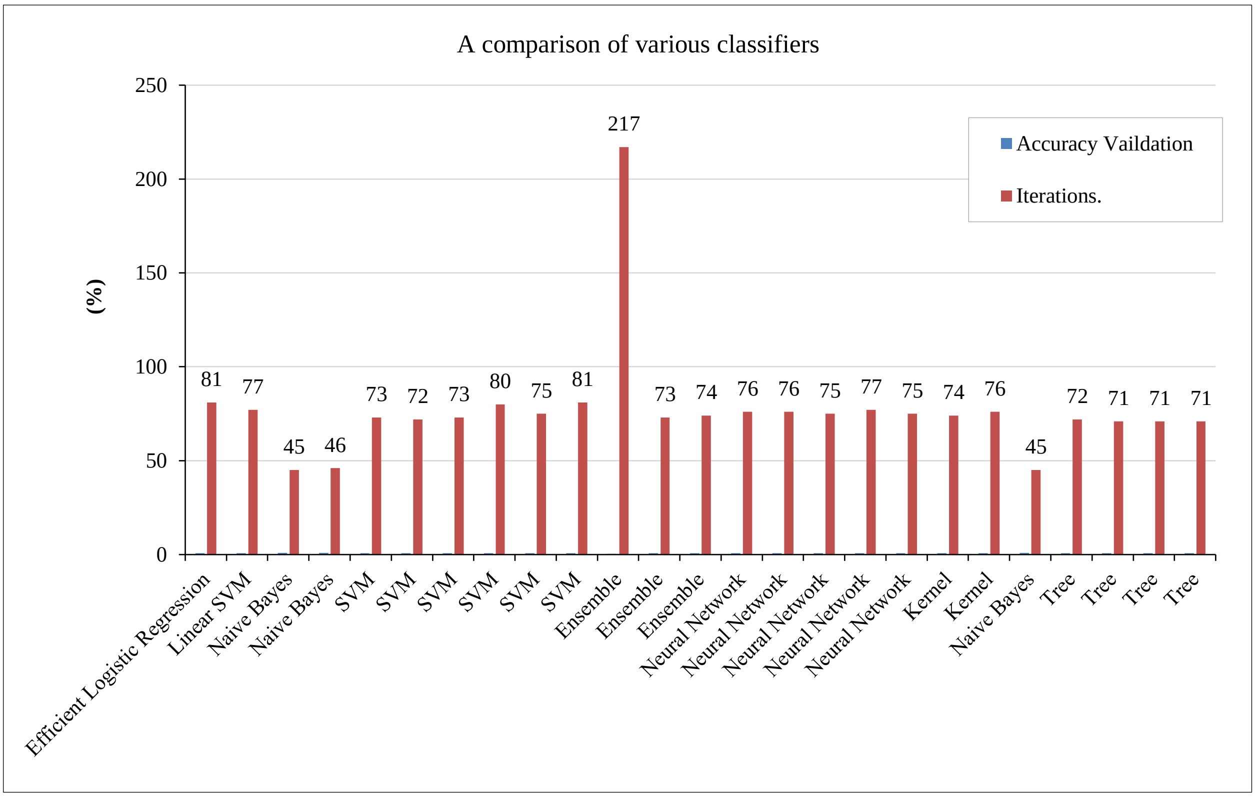
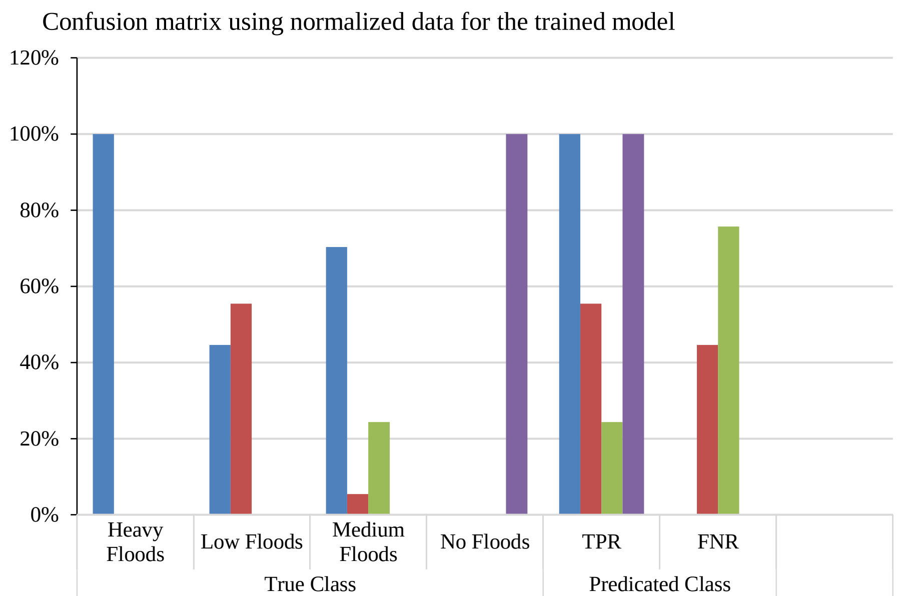


Figure 13. A comparison of various classifiers

4.2. Findings By the Data Normalization Training

Figure 14 illustrates the confusion matrix for the intelligent forecasting of flooding intensity using a trained model on normalized data. The matrix details the prediction accuracy across four classes: heavy floods, low floods, medium floods, and no floods, emphasizing the significance of correct predictions for effective flood management. The model's prediction accuracy for major floods was 100%, accurately identifying every incident with zero errors. This illustrates how well the model can forecast catastrophic flooding, which is essential for early warning systems and disaster relief efforts. The model had an error rate of 44.6% but correctly predicted 55.4% of the low flood cases. This indicates a moderate level of accuracy, suggesting that while the model can identify low flooding conditions, there is room for improvement in distinguishing these events from others. The prediction accuracy for medium floods was 70.3%, showing that the model can reliably identify medium flooding events, though less accurately than heavy floods. This is a significant finding, as medium floods can often be challenging to predict due to their less pronounced characteristics. The model's performance for medium floods could be more satisfactory, with a lower TPR and a higher FNR, indicating that medium flooding events are more challenging to predict accurately. This might be due to the subtler nature of these events, which can be harder to distinguish from low and heavy floods. For no floods, the model demonstrated an accuracy of 100%, successfully identifying all instances without any false positives. This high accuracy is crucial for avoiding unnecessary alarms and ensuring that resources are well-spent on accurate warnings. Overall, the confusion matrix in Figure 14 underscores the importance of data normalization in enhancing the model's prediction accuracy. Using normalized data, the model provided more precise predictions across different flood intensities, highlighting the effectiveness of the linear SVM approach. This accurate forecasting is vital for timely interventions and efficient resource allocation in flood-prone areas [31].



	Heavy Floods	Low Floods	Medium Floods	No Floods	TPR	FNR	
True Class Heavy Floods	100.00%	44.60%	70.30%	0%	100.00%	0.00%	0
True Class Low Floods	0%	55.40%	5.40%	0%	55.40%	44.60%	0
True Class Medium Floods	0%	0%	24%	0%	24%	76%	0
True Class No Floods	0%	0%	0%	100%	100%	0%	0

Figure 14. Confusion matrix using normalized data for the trained model

The comparison of various machine learning classifiers presented in Table 7 and Figure 15 offers valuable insights into the intelligent forecasting of flooding intensity. The development of efficient prediction systems that can lessen the effects of flood disasters depends on the results of this investigation. Table 7 displays various classifiers, such as ensemble methods, Neural Networks, Support Vector Machines (SVM), Tree-based models, Logistic Regression, Naive Bayes, and Kernel-based techniques. Each classifier's performance is evaluated using metrics such as trained data, validation accuracy, and input data requirements. Among the diverse classifiers, Naive Bayes emerges as the standout performer, achieving the highest validation accuracy of 87.50%. This exceptional performance suggests that Naive Bayes is adept at capturing the underlying patterns in flooding intensity data, demonstrating a solid ability to generalize from the training set to new, unseen instances. Other classifiers, including Tree-based models, SVMs, and Neural Networks, exhibit relatively consistent performance with validation accuracies ranging from approximately 77% to 80%. This consistency across different algorithms indicates that the dataset contains predictable patterns that can be captured by multiple approaches, reinforcing the reliability of the forecasting system. The input data column provides insight into each model's efficiency in processing new instances. Lower values in this column suggest that the model requires fewer data points to make accurate predictions. Notably, Naive Bayes, with an input data value of 45, demonstrates high efficiency and superior accuracy, making it an attractive option for real-time forecasting applications. An interesting outlier in the results is one Ensemble method that shows a significantly lower validation accuracy of 39.72% and a much higher input data requirement of 217.

This anomaly could indicate an overfitting issue or a potential error in the model's configuration, as it deviates substantially from the performance of other classifiers. The importance of this comparative analysis for intelligent forecasting of flooding intensity is multifaceted. It aids in model selection by identifying the most suitable algorithm for flood prediction, with Naive Bayes emerging as a vital element. The consistent performance across multiple classifiers suggests that the underlying data contains reliable patterns for flood prediction, enhancing confidence in the forecasting system. Furthermore, the input data metric helps understand each model's computational requirements, which is crucial for real-time forecasting systems where rapid predictions are necessary. The analysis also highlights the potential for balancing accuracy and complexity, as some models like SVMs and Neural Networks offer good accuracy with potentially more complex decision boundaries, which might be beneficial for capturing nuanced patterns in flooding data. The varying performances of different classifiers suggest that an optimized ensemble approach, combining multiple models, could yield even better results by leveraging the strengths of various algorithms. This paper's comparative analysis provides valuable insights for developing a robust and accurate intelligent forecasting system for flooding intensity. It underscores the effectiveness of Naive Bayes for this problem while also demonstrating the overall predictability of flooding patterns using machine learning approaches. This information can guide further refinement and implementation of flood prediction models, ultimately contributing to more effective flood management and mitigation strategies [32].

Table 7. Built model's prediction accuracy

Numbers	Machine Learning Classifiers	Trained Data	Validation	Input Data
1	Tree	Trained	80.00 %	72
2.1	Tree	Trained	80.28 %	71
2.2	Tree	Trained	80.28 %	71
2.3	Tree	Trained	80.28 %	71
2.4	Efficient LogisticRegression	Trained	77.50 %	81
2.5	Efficient Linear SVM	Trained	78.61 %	77
2.6	**Naive Bayes**	**Trained**	**87.50 %**	**45**
2.7	Naive Bayes	Trained	87.22 %	46
2.8	SVM	Trained	79.72 %	73
2.9	SVM	Trained	80.00 %	72
2.10	SVM	Trained	79.72 %	73
2.11	SVM	Trained	77.78 %	80
2.12	SVM	Trained	79.17 %	75
2.13	SVM	Trained	77.50 %	81
2.14	Ensemble	Trained	39.72 %	217
2.15	Ensemble	Trained	79.72 %	73
2.16	Ensemble	Trained	79.44 %	74
2.17	Neural Network	Trained	78.89 %	76
2.18	Neural Network	Trained	78.89 %	76
2.19	Neural Network	Trained	79.17 %	75
2.20	Neural Network	Trained	78.61 %	77
2.21	Neural Network	Trained	79.17 %	75
2.22	Kernel	Trained	79.44 %	74
2.23	Kernel	Trained	78.89 %	76
3	Naive Bayes	Trained	87.50 %	45

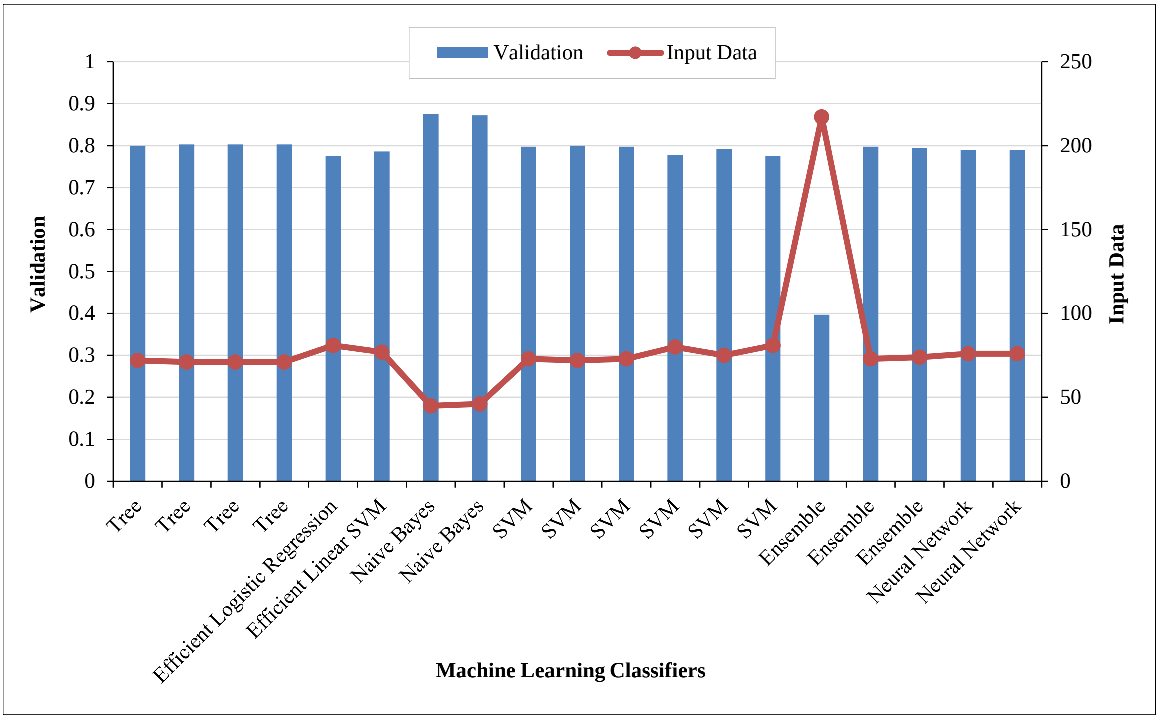


Figure 15. Built model's prediction accuracy

4.3. Comparison Between the Previous Studies and the Present Study

A comparison of the present study and previous studies on Intelligent Forecasting of Flooding Intensity Using Machine Learning for the Support Vector Machine (SVM) and Artificial Neural Network (ANN) models for runoff estimation is shown in Table 8. Because these machine learning algorithms can estimate discharge rates directly related to flood potential, they are essential for intelligent flooding intensity forecasts. Various statistical indices utilized to evaluate the models' performance during the training and testing phases are shown in Figure 16. These metrics include the Normalized Root Mean Square Error (NRMSE), coefficient of determination (R2), Root Mean Square Error (RMSE), Nash-Sutcliffe Efficiency (NSE), and percentage bias (PBIAS). The ANN model demonstrates excellent performance in both the training and testing phases. It shows lower RMSE values (9.56 and 7.01 Cfs for training and testing, respectively) than the SVM model. The NSE values (0.95 for training and 0.97 for testing) are very close to 1, indicating a high level of agreement between predictions and actual data. The PBIAS values (1.34% for training and 0.15% for testing) show a slight underestimation of peak discharge but are still within acceptable ranges.

The SVM model also performs well but shows slightly higher RMSE values (13.15 and 10.41 m3/s for training and testing, respectively) than the ANN model. Its NSE values (0.92 for training and 0.94 for testing) are good but slightly lower than the ANN model. The PBIAS values (-1.49% for training and -0.39% for testing) indicate a tendency to underestimate discharge, particularly during extreme events. Both models demonstrate high R^2 values, indicating strong correlations between predicted and observed data. However, the ANN model also slightly outperforms the SVM model [33]. The capacity of these models to precisely estimate discharge rates, which is essential for flood prediction and control, makes them important for intelligent forecasting of flooding intensity.

The ANN model seems especially good at capturing non-linear interactions between different hydrological parameters impacting discharge because of its multi-layered structure and activation functions. This makes it especially valuable for flood forecasting in complex watershed systems. While both models perform well under normal conditions, they tend to underestimate discharge during extreme flooding events. This limitation highlights the need to refine these models further to improve their accuracy during high-intensity flood scenarios. In comparison to this present and previous studies on intelligent forecasting of flooding intensity using machine learning, the ANN and SVM models demonstrate significant potential for intelligent forecasting of flooding intensity, especially in areas with limited data availability. However, the ANN model performs slightly superior overall, making it an up-and-coming tool for flood prediction and water resource management.

Table 8. Comparison between ANN and SVM models for runoff estimation [33]

Statistical Index	ANN Model		SVM Model	
	Training	Testing	Training	Testing
RMSE (Cfs)	9.56	7.01	13.15	10.41
NSE	0.95	0.97	0.92	0.94
PBIAS (%)	1.34	0.15	−1.49	−0.39
R2	0.96	0.97	0.92	0.94
NRMSE	1.68	0.91	0.036	0.0328

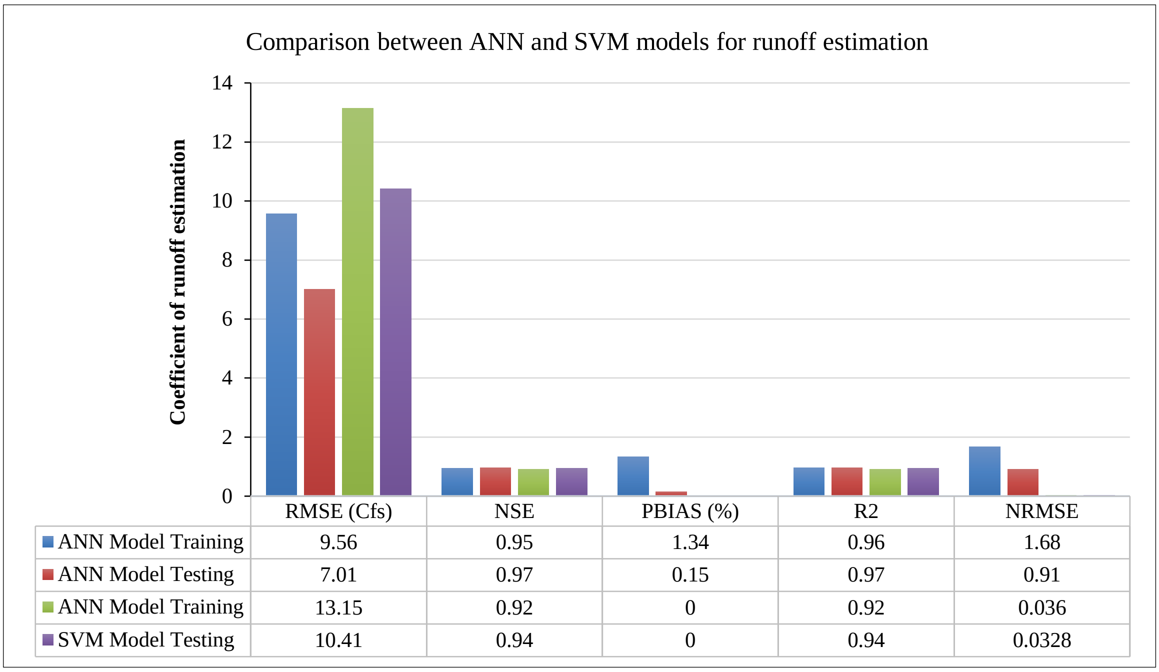


Figure 16. Comparison between ANN and SVM models for runoff estimation

5. Conclusions

This study presents significant advancements in intelligent flood intensity forecasting using machine learning techniques, focusing on Bor County, South Sudan. The study tackles a pressing requirement for enhanced forecasting and administration in an area highly susceptible to floods. The thorough methodology of the analysis incorporates a variety of data sources, such as current satellite imagery, historical flood records, meteorological data, and topographical information. This comprehensive approach enhances prediction accuracy and allows a more profound comprehension of flood risk elements. The research investigates various approaches to show how modern machine learning methods, such as random forests, support vector machines (SVM), and artificial neural networks (ANN), can capture intricate, non-linear correlations between flood-related variables. A significant innovation is creating a dynamic forecasting system that can process data in real time and update predictions when new information becomes available. This feature is precious in changing climate patterns and evolving flood risks. The suggested methodology tackles immediate local forecasts and extended regional evaluations, offering a comprehensive instrument for managing flood risks across many temporal and spatial dimensions. The study highlights how crucial model interpretability is, especially when deciding on flood control. This focus enhances the credibility and practical applicability of the system. The research also proposes integrating machine learning predictions with advanced visualization tools and geographic information systems, improving the communication of flood risk information to experts and the public. The findings indicate that the Linear Support Vector Machine (SVM) classifier performed exceptionally well, achieving 87.5% prediction accuracy on raw data. This result suggests this approach's potential for accurate flood forecasting in Bor County. The linear SVM effectively predicted heavy flood events, achieving 100% accuracy for high-velocity flooding samples. Emergency preparedness and early warning systems depend on this competence. Surprisingly, out of all the tested models, the Naive Bayes classifier had the best overall accuracy (87.50%). The discovery above implies that the Naive Bayes technique is especially appropriate for flood forecasting assignments because of its capacity to manage the uncertain character of flood occurrences efficiently. The research also explored the impact of data normalization on model performance. Normalized data improved prediction accuracy across different flood intensities, highlighting the importance of proper data reprocessing in enhancing model effectiveness.

In comparing ANN and SVM models for runoff estimation, the ANN model slightly outperformed the SVM model, demonstrating lower Root Mean Square Error (RMSE) values and higher Nash-Sutcliffe Efficiency (NSE) scores. This superiority suggests the ANN model's enhanced ability to capture non-linear interactions between hydrological parameters, making it particularly valuable for complex watershed systems. However, the study identified a standard limitation across models: the tendency to underestimate discharge during extreme flooding events. This finding underscores the need for further research to improve model accuracy in high-intensity flood scenarios. The novelty and improvement of this research lie in several areas. First, it uses cutting-edge machine-learning techniques to address a pressing environmental issue in a developing nation where conventional flood-predicting tools can have limitations. The study offers more dynamic and thorough risk assessment capabilities than static flood prediction models because it strongly emphasizes real-time adaptation and multi-scale forecasting. Integrating machine learning with visualization tools and geographic information systems presents a novel approach to enhancing flood risk information's communication and practical application. The research's focus on model interpretability also addresses a common challenge in machine learning applications, particularly in critical decision-making contexts like flood management. The comprehensive comparison of various models and exploration of data normalization effects provide valuable guidance for future research and practical applications in flood management and disaster preparedness. Enhancing flood prediction accuracy and supporting more effective flood mitigation tactics, the research presents a complete framework by integrating varied data sources, utilizing modern machine learning algorithms, and emphasizing interpretability and real-time adaptation. To sum up, this study makes a substantial contribution to the field of intelligent flood forecasting by showcasing the ability of machine learning approaches to anticipate the occurrence and intensity of floods precisely. The study shows the present strengths of these models and highlights areas that need to be improved in the future, especially in managing extreme flood occurrences. This will pave the way for future developments in this crucial environmental modeling and disaster mitigation field.

6. Declarations

6.1. Author Contributions

Conceptualization, A.A.N.D. and J.I.; methodology, A.A.N.D.; software, A.A.N.D.; validation, S.R., A.Z. and N.; formal analysis, A.Z.; investigation, S.R.; resources, J.I.; data curation, A.Z. and S.R.; writing—original draft preparation, A.A.N.D.; writing—review and editing, A.A.N.D. and J.I.; visualization, A.Z. and S.R.; supervision, J.I. and N.; project administration, A.A.N.D.; funding acquisition, J.I. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This research paper is supported financially by the Universitas Muhammadiyah Yogyakarta for publication of this Article.

6.4. Acknowledgements

The author thanks the University Muhammadiyah Yogyakarta and the lecturers of the Scientific Research Study Program for their support of this study.

6.5. Conflicts of Interest

The authors declare no conflict of interest.

7. References

[1] Mayen, J. V., Wood, E., & Frazier, T. (2022). Practical flood risk reduction strategies in South Sudan. Journal of Emergency Management, 20(8), 123–136. doi:10.5055/jem.0669.

[2] Food Security Cluster. (2024). South Sudan Flood Preparedness and Response Plan-June to December 2024. Food Security Cluster, Rome. Available online: https://fscluster.org/sites/default/files/South%20Sudan%20Floods%20Preparedness%20and%20Response%2022%20June%202024.pdf (accessed on September 2024).

[3] Blöchl, G., Ardoin-Bardin, S., Bonell, M., Dorninger, M., Goodrich, D., Gutknecht, D., Matamoros, D., Merz, B., Shand, P., & Szolgay, J. (2007). At what scales do climate variability and land cover change impact on flooding and low flows? Hydrological Processes, 21(9), 1241–1247. doi:10.1002/hyp.6669.

[4] Yaseen, Z. M., Sulaiman, S. O., Deo, R. C., & Chau, K. W. (2019). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. Journal of Hydrology, 569, 387–408. doi:10.1016/j.jhydrol.2018.11.069.

[5] Smith, A., Boyd, D., & Veale, B. (1999). Flood management on the Grand River, Ontario, Canada: a watershed conservation perspective. Environments, 27(1), 23.

[6] Abou Rjeily, Y., Abbas, O., Sadek, M., Shahrour, I., & Hage Chehade, F. (2017). Flood forecasting within urban drainage systems using NARX neural network. Water Science and Technology, 76(9), 2401–2412. doi:10.2166/wst.2017.409.

[7] Tang, Y., Sun, Y., Han, Z., Soomro, S. e. hyde., Wu, Q., Tan, B., & Hu, C. (2023). Flood Forecasting Based on Machine Learning Pattern Recognition and Dynamic Migration of Parameters. Journal of Hydrology: Regional Studies, 47, 101406. doi:10.1016/j.ejrh.2023.101406.

[8] Nevo, S., Morin, E., Gerzi Rosenthal, A., Metzger, A., Barshai, C., Weitzner, D., Voloshin, D., Kratzert, F., Elidan, G., Dror, G., Begelman, G., Nearing, G., Shalev, G., Noga, H., Shavitt, I., Yuklea, L., Royz, M., Giladi, N., Peled Levi, N., … Matias, Y. (2022). Flood forecasting with machine learning models in an operational framework. Hydrology and Earth System Sciences, 26(15), 4013–4032. doi:10.5194/hess-26-4013-2022.

[9] Chang, F., Hsu, K., & Chang, L. (2019). Flood Forecasting Using Machine Learning Methods. In Flood Forecasting Using Machine Learning Methods. doi:10.3390/books978-3-03897-549-6.

[10] Kunverji, K., Shah, K., & Shah, N. (2021). A Flood Prediction System Developed Using Various Machine Learning Algorithms. SSRN Electronic Journal. doi:10.2139/ssrn.3866524.

[11] Faruq, A., Hussein, S. F. M., Marto, A., & Abdullah, S. S. (2022). Flood River Water Level Forecasting using Ensemble Machine Learning for Early Warning Systems. IOP Conference Series: Earth and Environmental Science, 1091(1). doi:10.1088/1755-1315/1091/1/012041.

[12] Yang, S. N., & Chang, L. C. (2020). Regional inundation forecasting using machine learning techniques with the internet of things. Water (Switzerland), 12(6). doi:10.3390/W12061578.

[13] Nakhaei, M., Nakhaei, P., Gheibi, M., Chahkandi, B., Wacławek, S., Behzadian, K., Chen, A. S., & Campos, L. C. (2023). Enhancing community resilience in arid regions: A smart framework for flash flood risk assessment. Ecological Indicators, 153. doi:10.1016/j.ecolind.2023.110457.

[14] Shehzadi, M., Ali, R. H., Abideen, Z. ul, Ijaz, A. Z., & Khan, T. A. (2023). Enhancing Flood Resilience: Streamflow Forecasting and Inundation Modeling in Pakistan †. Engineering Proceedings, 56(1), 1–8. doi:10.3390/ASEC2023-16612.

[15] Mardian, J., Champagne, C., Bonsal, B., & Berg, A. (2023). A Machine Learning Framework for Predicting and Understanding the Canadian Drought Monitor. Water Resources Research, 59(8), 1–23. doi:10.1029/2022WR033847.

[16] Filipova, V., Hammond, A., Leedal, D., & Lamb, R. (2022). Prediction of flood quantiles at ungauged catchments for the contiguous USA using Artificial Neural Networks. Hydrology Research, 53(1), 1–17. doi:10.2166/NH.2021.082.

[17] Zhang, Y., Pan, D., Van Griensven, J., Yang, S. X., & Gharabaghi, B. (2023). Intelligent flood forecasting and warning: a survey. Intelligence and Robotics, 3(2), 190–212. doi:10.20517/ir.2023.12.

[18] Zhang, B., Ouyang, C., Cui, P., Xu, Q., Wang, D., Zhang, F., Li, Z., Fan, L., Lovati, M., Liu, Y., & Zhang, Q. (2024). Deep learning for cross-region streamflow and flood forecasting at a global scale. Innovation, 5(3). doi:10.1016/j.xinn.2024.100617.

[19] Kumar, V., Azamathulla, H. M., Sharma, K. V., Mehta, D. J., & Maharaj, K. T. (2023). The State of the Art in Deep Learning Applications, Challenges, and Future Prospects: A Comprehensive Review of Flood Forecasting and Management. Sustainability (Switzerland), 15(13). doi:10.3390/su151310543.

[20] Yang, Y., & Chui, T. F. M. (2021). Modeling and interpreting hydrological responses of sustainable urban drainage systems with explainable machine learning methods. Hydrology and Earth System Sciences, 25(11), 5839–5858. doi:10.5194/hess-25-5839-2021.

[21] Xu, H., Ragno, E., Tan, J., Antonini, A., Bricker, J. D., Jonkman, S. N., Liu, Q., & Wang, J. (2023). Perspectives on Compound Flooding in Chinese Estuary Regions. International Journal of Disaster Risk Science, 14(2), 269–279. doi:10.1007/s13753-023-00482-1.

[22] Hu, R., Fang, F., Pain, C. C., & Navon, I. M. (2019). Rapid spatio-temporal flood prediction and uncertainty quantification using a deep learning method. Journal of Hydrology, 575, 911–920. doi:10.1016/j.jhydrol.2019.05.087.

[23] Moon, H., Yoon, S., & Moon, Y. (2023). Urban flood forecasting using a hybrid modeling approach based on a deep learning technique. Journal of Hydroinformatics, 25(2), 593–610. doi:10.2166/hydro.2023.203.

[24] Deng, A. A. N., Nursetiawan, N., & Ikhsan, J. (2024). Evaluating Flood Hazard Mitigation through Sustainable Urban Drainage Systems in Bor, Jonglei State, South Sudan. Journal of Civil and Hydraulic Engineering, 2(1), 31–50. doi:10.56578/jche020103.

[25] Orton, P. M., Conticello, F. R., Cioffi, F., Hall, T. M., Georgas, N., Lall, U., Blumberg, A. F., & MacManus, K. (2020). Flood hazard assessment from storm tides, rain and sea level rise for a tidal river estuary. Natural Hazards, 102(2), 729–757. doi:10.1007/s11069-018-3251-x.

[26] Puspasari, R. L., Yoon, D., Kim, H., & Kim, K. W. (2023). Machine Learning for Flood Prediction in Indonesia: Providing Online Access for Disaster Management Control. Economic and Environmental Geology, 56(1), 65–73. doi:10.9719/eeg.2023.56.1.65.

[27] Maspo, N. A., Bin Harun, A. N., Goto, M., Cheros, F., Haron, N. A., & Mohd Nawi, M. N. (2020). Evaluation of Machine Learning approach in flood prediction scenarios and its input parameters: A systematic review. IOP Conference Series: Earth and Environmental Science, 479(1). doi:10.1088/1755-1315/479/1/012038.

[28] Nyoagbe, M., Ayer, J., & Yevugah, L. L. (2023). Flood Prediction using Machine Learning and GIS Flood Prediction Using Machine Learning and GIS as an Early Warning System (12205) Michael Nyoagbe, John Ayer, Lily Lisa Yevugah and Yaw Mensah Asare (Ghana). FIG Working Week 2023, Protecting Our World. 28 May-1 June, 2023, Orlando, United States.

[29] Yang, Y., Yin, J., Zhang, W., Zhang, Y., Lu, Y., Liu, Y., Xiao, A., Wang, Y., & Song, W. (2021). Modeling of a compound flood induced by the levee breach at Qianbujing Creek, Shanghai, during Typhoon Fitow. Natural Hazards and Earth System Sciences, 21(11), 3563–3572. doi:10.5194/nhess-21-3563-2021.

[30] Razali, N., Ismail, S., & Mustapha, A. (2020). Machine learning approach for flood risks prediction. IAES International Journal of Artificial Intelligence, 9(1), 73–80. doi:10.11591/ijai.v9.i1.pp73-80.

[31] Sekulić, P., Regina, P., Spadafina, L., Dentamaro, G., Porcelli, A., Bove, C., Kovačević, S., & Kalezić, M. (2020). Real-time flood prediction using recurrent neural networks and random forest. 24th IMEKO TC4 International Symposium and 22nd International Workshop on ADC and DAC Modelling and Testing, 14-16 September, 2020, Palermo, Italy.

[32] Ogbuene, E. B., Eze, C. A., Aloh, O. G., Oroke, A. M., Udegbunam, D. O., Ogbuka, J. C., Achoru, F. E., Ozorme, V. A., Anwara, O., Chukwunonyelum, I., Nebo, A. N., & Okolo, O. J. (2024). Application of Machine Learning for Flood Prediction and Evaluation in Southern Nigeria. Atmospheric and Climate Sciences, 14(03), 299–316. doi:10.4236/acs.2024.143019.

[33] Asadollahi, A., Magar, B. A., Poudel, B., Sohrabifar, A., & Kalra, A. (2024). Application of Machine Learning Models for Improving Discharge Prediction in Ungauged Watershed: A Case Study in East DuPage, Illinois. Geographies, 4(2), 363–377. doi:10.3390/geographies4020021.