Grenze journal name, Volume. 1, Issue.1, May 2015

Flood Prediction Using Machine Learning

Sandhya L Assistant Professor[1], Swarna Lohit[2], Sanjay S[3] and Manoj M[4]
Presidency University, Bangalore, India
Email: saisandhyalax@gmail.com, Email: swarnalohit06@gmail.com
Email: {sanjaysan2324, manojmanu05.d }@gmail.com

***Abstract*—Floods are among the most destructive natural disasters, which are highly complex to model. The complex nature of rainfall, influenced by various atmospheric, oceanic, and geographical factors, makes it a challenging phenomenon to forecast. This research employs data preprocessing techniques, outlier analysis, correlation analysis, feature selection, and several machine learning algorithms. This research on the advancement of flood prediction models contributed to risk reduction, minimization of the loss of human life, and reduction of the property damage associated with floods, during the past two decades, machine learning (ML) methods contributed highly in the advancement of prediction systems providing better performance and cost-effective solutions. This research focuses on leveraging historical meteorological data to find trends using machine learning to estimate rainfall. In this paper, the literature where ML models were benchmarked through a qualitative analysis of robustness, accuracy, effectiveness, and speed are particularly investigated to provide an extensive overview on the various ML algorithms used in the field. This paper aims to reduce the extreme risks of the natural disaster and also contributes to policy suggestions by providing a prediction for floods using different machine learning models. We will use k-nearest neighbors (KNNs), support vector machines (SVMs), random forests (RFs), and decision trees (DTs) to build our ML models. And to resolve the issue of oversampling and low accuracy, a stacking classifier will be used. For comparison among these models, we will use accuracy, f1-scores, recall, and precision. The results indicate that stacked models are best for predicting floods due to real-time rainfall in that area.**

***Index Terms*—Binary Logistic Regression, Support Vector Classifier (SVC), K-Nearest Neighbor (KNN), Decision Tree Classifier (DTC), Flood Prediction, Rainfall.**

I. INTRODUCTION

Natural calamities like hurricanes, earthquakes, floods, wildfires, and tsunamis are caused by the forces of nature and can happen suddenly. Environmental variables, such as climate change, deforestation, and urbanisation, frequently feed these occurrences and increase their frequency and intensity. Natural catastrophes can have severe effects, leading to extensive destruction and fatalities. One of the most critical weather factors that affects many parts of our everyday lives is rainfall [1,2]. However, the unpredictable nature of rainfall patterns can give rise to extreme weather events, such as prolonged droughts or devastating floods, which can have far reaching consequences for ecosystems, agriculture, and human populations [3]. Various kinds of rainfall exist, and unique mechanisms and climatic factors distinguish each. A few typical types of precipitation are mentioned in Figure 1. According to Indian Meteorological Department (IMD) published its "Climate of India" report, precipitation in India increased to 1298.53 mm in 2022 from 1213.82 mm in 2021. Among extreme weather events, floods, heavy rains, and landslides, causing deaths of 759 people [4]. In this regard, the field of weather forecasting has witnessed significant advancements with the integration of data analysis and machine learning techniques. Machine learning, a powerful computational approach, harnesses the potential of vast datasets to uncover intricate patterns, correlations, and trends among various meteorological variables. By leveraging this knowledge, machine learning algorithms can make accurate predictions, aiding in better understanding and anticipation of rainfall patterns [5].

Several well-established rainfall forecasting models are currently employed worldwide. These models include the Weather Research and Forecasting (WRF) model, which combines advanced atmospheric physics with numerical simulations to generate high-resolution weather forecasts. The General Forecasting Model focuses on providing short-term weather predictions, while Seasonal Climate Forecasting aims to anticipate rainfall patterns over longer periods. The Global Data Forecasting Model integrates a wide range of meteorological data from across the globe to produce com prehensive weather forecasts. Although these models offer valuable insights, their computational requirements can be substantial, making them resource-intensive to run and maintain [5].

The field of artificial intelligence concentrates on creating machines that can process data, learn from it, and make judgements. The use of machine learning is an appealing approach for flood forecasting because it holds the promise of revealing intricate, complicated correlations within huge datasets. Its capacity to incorporate data 2 of 40 from numerous sources, including satellite images, river gauge data, and climate models, offers chances to improve floods' precision, predictability, and lead time [6].

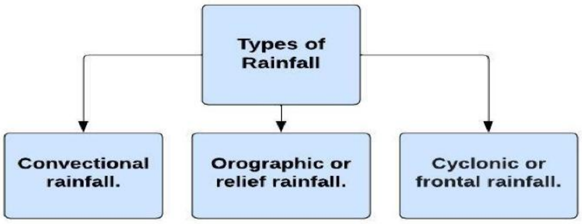


Figure 1. Various Types of rainfall according to literature.

Based on a critical study of literature, it has been observed that work on rainfall prediction is at infancy. Kerala is a region which is highly affected by rainfall and lots of lives are lost yearly because of floods. This study performs a comparative analysis of different ML methods for rainfall prediction. Finally, this paper offers and set apart from the others that compare flood prediction models:

1. Highlight the serious and long-lasting effects that floods have on the socioeconomic system, agriculture, and human life while acknowledging the growing challenge in accurately estimating rainfall because of climate changes, non-linear qualities, and variable attributes.
2. Identify undetectable trends in historical meteorological data to identify machine learning approaches as valuable tool for precisely estimating rainfall with quantitative results to prove the usefulness of the machine learning models.
3. Observation that the Support Vector Classifier and Binary Logistic Regression algorithms achieved high accuracy, 96% and 92%, respectively.

The remaining sections of this work are organised as follows: Section 2 presents a detailed literature review regarding the materials and methods, and partly examines the theoretical background on flood early prediction utilising deep learning and various machine learning techniques. Section 3 discusses the proposed methodology. Sections 4–6 present the results and discussion for comparing various machine learning and deep learning techniques for flood earlier forecasting based on a dataset. Section 7 provides a conclusion and future recommendations.

II. Literature Review

We structured this section as follows. First, the past studies are highlighted. A discussion on the justification of this research follows this.

II. (A) *Related Work*

Kerala State has an average annual precipitation of about 3000 mm. The rainfall in the State is controlled by the South-west and North-east monsoons. About 90% of the rainfall occurs during six monsoon months [7]. Kerala, a state in Southern India, experienced severe rainfall, landslides, and floods between June and August

2018. These were the worst floods in the state since 1924 and the third worst in India since 1900. A total of 504 people died and 23 million people were directly affected [8]. Kerala experienced an abnormally high rainfall from 1 June 2018 to 19 August 2018. This resulted in severe flooding in 13 out of 14 districts in the State. As per IMD data, Kerala received 2346.6 mm of rainfall from 1 June 2018 to 19 August 2018 in contrast to an expected 1649.5 mm of rainfall. This rainfall was about 42% above the normal. Further, the rainfall over Kerala during June, July and 1st to 19th of August was 15%, 18% and 164% respectively, above normal. Month-wise rainfall for the period, as reported by IMD, are given in Table-1.

Period	Normal Rainfall	Actual Rainfall	Departure from normal
	(mm)	(mm)	(%)
June, 2018	649.8	749.6	15
July, 2018	726.1	857.4	18
1-19, August, 2018	287.6	758.6	164
Total	**1649.5**	**2346.6**	**42**

Table-1: Month wise actual rainfall, normal rainfall and percentage departure from normal

The water levels in several reservoirs were almost near their Full Reservoir Level (FRL) due to continuous rainfall from 1st of June. Another severe spell of rainfall started from the 14th of August and continued till the 19th of August, resulting in disastrous flooding in 13 out of 14 districts. The water level records at CWC G&D sites for some of the rivers in Kerala are given at Annex-I. As per the rainfall records of IMD, it has been found that the rainfall depths recorded during the 15-17, August 2018 were comparable to the severe storm that occurred in the year 1924 [7]. Kerala has experienced disasters in the past, resulting in loss of human lives and livestock along with damage to infrastructure including public and private properties. Some major disasters experienced by the State are as follows:

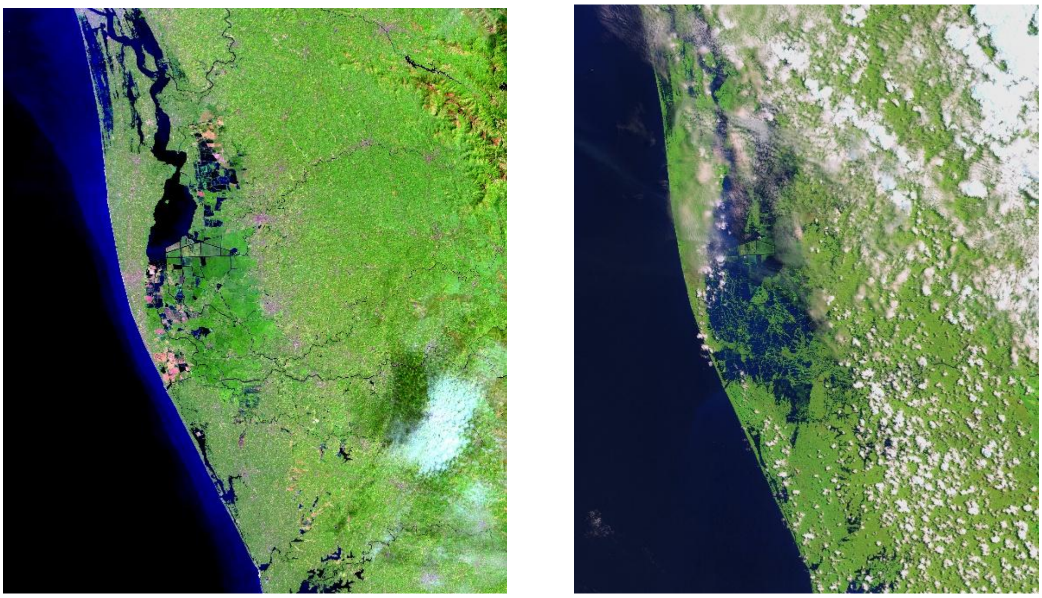
1. **Great flood of 99, (1924):** The great flood of '99 occurred as the result of the flooding of the Periyar River in Kerala in July, 1924. Kerala saw unprecedented rainfall in this incident of ‘flood of 99’ with nearly 3,368 mm of rain was recorded that month. It was 64 per cent higher than the normal rainfall and is the highest recorded rainfall till date. Around 1000 people died in the great flood of ’99 **(Wikimapia, 2013)**

2. **Kerala Floods, 2018:** Kerala experienced the worst floods in its history between 1 June and 19 August, 2018, since the Great flood of ’99. The state that year received 42 % of excess rainfall compared to average rainfall. The state government reported that 1,259 villages out of a total of 1,664 villages in 14 districts were affected. The central government declared the floods as “calamity of a severe nature”. 35 out of 54 dams in the state were opened for the first time in the history. Major Reservoirs in Kerala are listed in Table-2 only 7 reservoirs are having a live storage capacity that constitute 74% of the total live storage in Kerala. The rains resulted in landslides in hilly areas after torrents of water loosened soils from hill slopes. These slurries of water, soil, rock, and vegetation overwhelmed villages, downed power lines, and cut some communities off from receiving immediate aid. About 341 landslides were reported from 10 districts (**The Indian Express, 23 August 2019**).

Sl.No.	Name of Reservoir	Live Storage Capacity (MCM)
1.	Idukki	1460
2.	Idamalayar	1018
3.	Kallada	488
4.	Kakki	447
5.	Parambikulam (for use of TN)	380
6.	Mullaperiyar (for use of TN)	271
7.	Malampuzha	227

Table-2: Major Reservoirs in Kerala

The floods highlighted a number of structural constraints that left Kerala unprepared for major disasters caused by natural hazard or climate change shocks. Due to these systemic weaknesses, Kerala was at the mercy of the 2018 floods and landslides and suffered major socio-economic losses.



February 6, 2018 August 22, 2018

Figure 2: The before and after images released by the NASA, 2018 **[Source]**

The disaster resulted in loss of lives, livestock and agriculture, damaged houses and crops, destroyed roads bridges, school etc. The Cochin International Airport got flooded and had to hold back its operations from 15 to 29 August, 2018. It is to be noted that Cochin International Airport is one of the busiest international airport in India. The cumulative loss and damage from the preliminary and additional memorandum are discussed below:

a. **Human Fatalities:** The disaster of floods and landslides resulted in 433 fatalities; 268 men, 98 women and 67 children up to 22nd May–29th August, 2018 **(UNDP PDNA report on Kerala floods, 2018)**. All 14 districts and 1260 out of 1664 villages were affected. 687 km square of land was flooded. The floods were accompanied by 341 landslides. Landslides occurred inland from the rivers and occurred independently of the high flood levels in the river. It happened mainly due to soaking of soils, soil piping, and human activities such as road construction and housing. A large number of houses were completely or severely damaged. The cyclonic storms, wind and rainfall caused severe damage to the fisheries sector of the state.

b. **Agriculture:** The devastating floods damaged the state's agriculture production mainly the plantation and spice crops which are the backbone of the state's agriculture. Kerala cultivates around 1,62,660 ha of spice crops across the state with a production of 140,000 tonnes per annum. Idukki and Wayanad together are contributors of nearly 62 per cent of the total area under spices in the state.

c. **Fisheries:** The floods resulted in the aggregate loss of ` 10,304 lakh in aquaculture and inland capture fisheries. As many as 235 boats were fully damaged out of which 96 boats were from Ernakulam district. Out of the 1002 boats that were partially damaged, 818 boats were solely from the Kottayam district. A total of 1748 nets were fully damaged while 1620 nets were partially damaged during this disaster in Kerala **(Government of Kerala, 2018).**

d. **Animal Husbandry:** The unprecedented rainfall which triggered flooding in the state resulted in the death of scores of cattle, buffaloes, goats and poultry. Alappuzha was the worst affected district with regard to this sector. A total of 7146 cattle died, including 650 cows and buffalo, 2994 sheep and 3502 calves. Almost 500792 poultry died in these flash floods **(Government of Kerala, 2018)**.

e. **Damaged Houses:** Around 14 lakh people were shifted to relief camps during the floods as their houses were inundated with flood water. The floods and landslides caused massive damage to houses, infrastructure like roads, railways, bridges, power supplies and communications networks. The floods washed away crops and livestock thereby impacting the lives and livelihoods of people. A total of 17,316 houses were completely destroyed or damaged as per the data compiled on 4 October, 2018. The total damage (in monetary terms) to education and child protection sector was estimated at ₹179.48 crore, A total of ₹214 crore was estimated to be the recovery and reconstruction needs for the education sector for the next 3–5 years (**UNDP, 2018**)

f. **Kochi Airport:** The flood had damaged Kochi International Airport also. The total damage caused to the Kochi International Airport during Kerala floods was estimated to be between ₹200 to ₹250 crore. Kochi was the busiest airport in Kerala and receives bulk of its international passengers from the Gulf countries. All the operations in Kochi airport were cancelled from 15 August, 2018 after floodwater crossed the periphery walls and flooded the runway, making it unfit for use. Only four out of the eight power storage plants were functional. The cost of repairing and replacing solar panels was estimated to be around ₹10 crore (**India Today, 2018**).

g. **Road Transport:** Roads are the principal mode of transport in Kerala that share about 75 percent of freight and 85 percent of passenger load. Kerala has a dense road network which is about three times the national average. Roads were fully damaged and would need complete depth pavement reconstruction, considerable repair/reconstruction of drainage, cross drainage and slope protection works and limited road raising, and new cross drainage works **(Government of Kerala, 2018**).

Causes of Floods: There were several causative factors which contributed to the immense rainfall in Kerala getting converted into a disaster. The natural factor of torrential rains was augmented with several human factors, which resulted in loss of lives, infrastructure and livelihoods.

1. High Rainfall
2. Dam Management
3. Overflow of Rivers and Blockage of Water Bodies
4. Poor Resource Management
5. Lack of Awareness
6. Poor Discharge Capacities of Water Bodies
7. Unplanned Urbanization

The proposed study can forecast rainfall and predict in Kerala for any season.

III. Proposed Methodology

A. DATASET DESCRIPTION

This proposed research aims to determine whether, by utilising machine learning algorithms, a higher accuracy rate can be attained while also reducing error. The dataset includes information on Kerala's monthly and yearly rainfall (1901 to 2018). Sourced from the Kerala – India Meteorological Department, this dataset will be used as input to make accurate predictions. India has established a comprehensive network of weather monitoring stations across Kerala to enhance meteorological observations and forecasting capabilities. That includes 109 Automatic weather Stations, Automatic Rain Gauges of 30 operational station in Kerala.

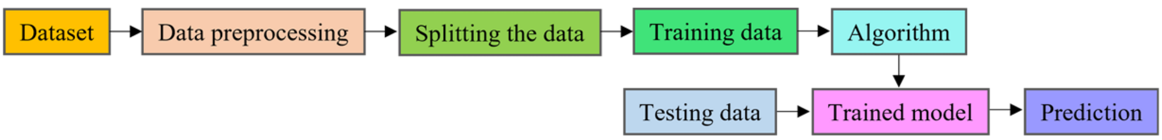


Figure 3: The overview of the methodology of the proposed work

For instance, the target variable to forecast is whether or not it will rain tomorrow, indicated by a binary value of "yes" or "no." In this context, "yes" signifies that it will rain the following day if the rainfall for that day is recorded as 1mm or more.

B. DATA PREPROCESSING

Data Preprocessing is often used in the field of machine learning to describe the steps taken to clean, organize, and prepare raw data before it is used to construct machine learning models [9]. Preprocessing methods may be used to get rid of certain abnormalities while keeping others untouched [10]. The goal of data preprocessing is to improve the quality of the data and to make it more suitable for the specific data mining task. Some common steps in data preprocessing include: Data Cleaning, Data Integration, Data Transformation, Data Reduction, Data Normalization.

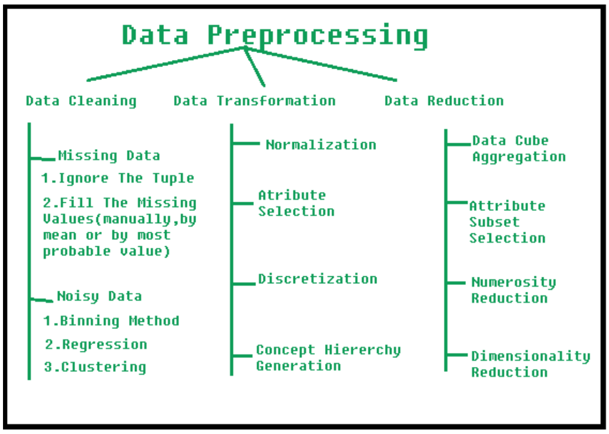


Figure 4: Data Preprocessing

C. OUTLIER

An outlier is a value in a random sample of a population that deviates abnormally from the other values [11]. An outlier may occur due to the variability in the data, or due to experimental error/human error. If we have a huge dataset, we can identify using visualization and mathematical techniques such as Boxplots, Z-score, Inter Quantile Range (IQR). Outliers for each attribute are presented using the boxplot function. These outliers needed to be processed to ensure the model's correct performance since they represent discrepancies in the data instances [5].

D. NORMALIZATION

A normalization is an approach to reducing the number of inserts, deletes, and changes that happen in a database because of duplicate data that can cause problems [12]. The process of normalization can improve data integrity and reduce dataset redundancy [13]. Normalization gives Improved performance of machine learning algorithms, Improved interpretability of results and improve the generalization of a model, by reducing the impact of outliers and by making the model less sensitive to the scale of the inputs. The equation for data normalisation using the min–max scaling technique is as follows:

$$x_{normalized} = \frac{x - \min(x)}{\max(x) - \min(x)}$$

E. ***CLASSIFICATION MODELS***

Classification is a supervising technique that categorizes the data into the desired number of classes. Multiple factors make intelligible classification models important. Users must understand a computer-induced model to trust and follow its predictions [14]. We have employed six classifiers: Decision Tree, SVM, Random Forest ad Logistic Regression. Finally, we compared their performance based on different model evaluation metrics and found the best-fitting algorithm for this problem [5].

Figure 5: The figure illustrates various factors that influence rainfall.

F. MACHINE LEARNING MODELS:

1. **Binary Logistic Regression:** The logistic regression model employed in this flood mapping study is designed to predict the probability of flood occurrence in a given area based on a set of predictor variables. Logistic regression is a statistical method used for binary classification problems, where the outcome variable is categorical and can take one of two possible out-comes: flood or no flood. A classical linear model can be denoted in the following manner:

$$Y = \alpha + \beta X + \varepsilon$$

   Where Y is the dependent variable, α is the Y intercept when X is equal to zero, X is the independent variable, β is the regression coefficient representing the variation in Y due to change in values of X and ε is the error of the model.

2. **Support Vector Classifier:** The Support Vector Classifier (SVC) is a supervised machine learning algorithm (Wan and Lei, 2009) that uses both regression, classification and outliers' detection. It works especially well when handling complicated datasets and is commonly used in various domains, including flood prediction [18]. Historical data related to floods, including features such as rainfall patterns, river levels, soil moisture, and topography, is collected and pre-processed [19]. The SVM algorithm is applied to the training dataset, using flood occurrence as the target variable. SVM searches for the optimal hyperplane that can separate flood and non-flood instances with the maximum margin, or in the case of non-linear data, it is mapped into a higher-dimensional space using kernel functions [20]. The predicted outcomes can be binary (flood or non-flood) or continuous (indicating the severity or probability of flooding).

3. **Decision Tree:** A decision tree is a flowchart-like structure used to make decisions or predictions. It consists of nodes representing decisions or tests on attributes, branches representing the outcome of these decisions, and leaf nodes representing final outcomes or predictions. It is used in flood prediction by analysing historical data and relevant features to make predictions about the occurrence or severity of floods. In a dataset, decision trees can be used to determine which variables or characteristics are most significant [21]. Decision trees can be used for exploratory data analysis and offer a visual picture of the decision-making process. They allow users to understand the relationships between features and their impact on the outcome or class prediction [22]. Decision trees can help uncover patterns, interactions, and decision rules within the data.

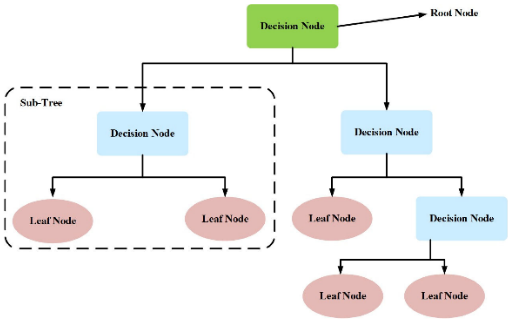


Figure 6: The working flow of decision tree.

4. **Random Forest:** RF is one of the supervised machine learning algorithms in the field of regression and classification which was introduced by Breiman (2001). The technique used to reduce the estimated variance is called bagging. Bagging seems to work especially well for high variance, and low bias procedures such as trees in decision tree models. RF is a basic modification of bagging, which is a large collection of trees (Hastie, 2009). In other words, a RF model is a collection of decision trees, as the building block of a RF model is a decision tree (Caigny et al., 2018). Each tree is trained on a sample of training data. One of the key advantages of random forests is their ability to mitigate the overfitting tendency of decision trees. By aggregating the predictions of multiple trees, random forests provide a more robust and accurate prediction [5]. Then, if the goal is classification; prediction is undertaken by majority vote of trees. By satisfying these conditions, random forests can effectively capture diverse patterns and make accurate predictions by leveraging the collective knowledge of the ensemble.

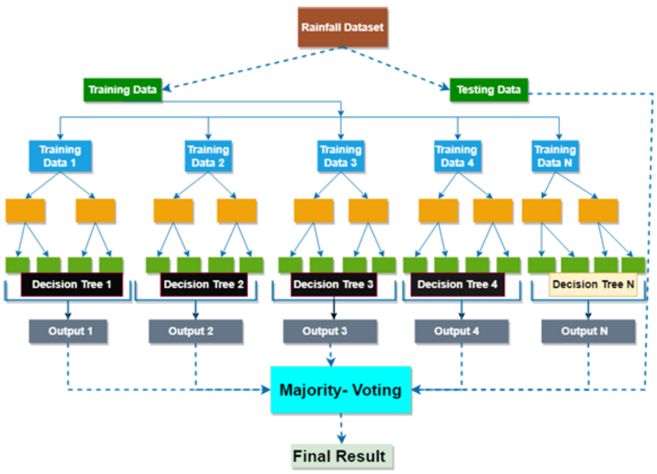


Figure 7: Random forest classifier architecture

IV. RESULT AND ANALYSIS

By completing the training dataset, SVM and LR models have been implemented by using the Python programming language. Any prediction model will work best if the data used to train it is accurate and healthy. Weak classifiers are frequently the result of incorrect data values. This step is critical for ensuring valuable data. We count missing values for each attribute except date and location in our preprocessing effort [5].

A. Table 1901 to 2018

Machine Learning Models	Accuracy Score	Recall Score	ROC Curve Score
K-Nearest Neighbors (KNN)	0.8750	0.80	90.00
Logistic Regression (LR)	0.9583	0.86	82.00
Support Vector Machine (SVM)	0.9167	0.86	82.00
Decision Tree Classifier (DTC)	0.7500	0.72	82.22
Random Forest	0.8333	0.79	76.67

From the table, Binary Logistic Regression has the highest accuracy rate of 0.9583 with ROC score and recall score of 0.82 and 0.86 respectively.

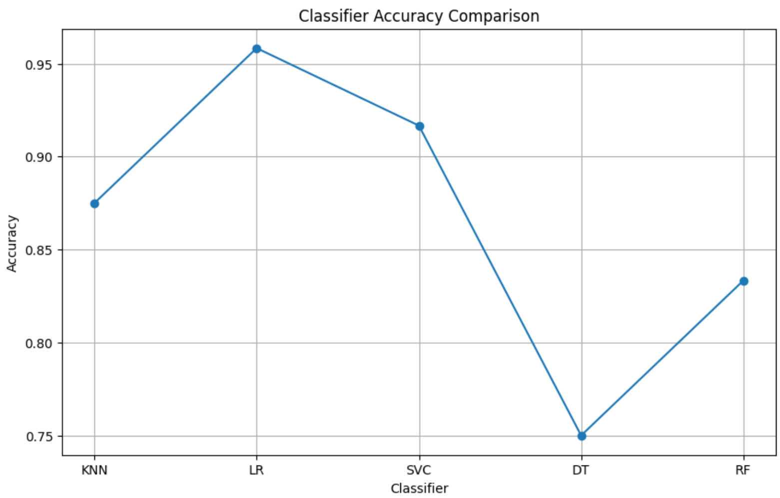


Figure 8: Accuracy comparison of models using a Line Graph

After that, the Support Vector Classifier (SVC) has accuracy of 0.9167 with a ROC score of 0.82 and recall score of 0.82.

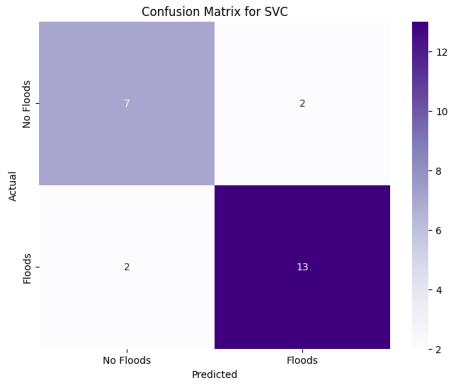


Figure 9: Confusion matrix of SVC

V. CONCLUSION AND FUTURE DIRECTIONS

This study proposes a real-time flood extent prediction method using logistic regression. This thesis presents a systematic approach to developing a robust classification system for this task. Various machine learning classification techniques are investigated and evaluated at different stages of the research [5]. Furthermore, this study envisions the potential for using different machine learning methods to predict various outcomes in the future. The research can be extended to address real world data challenges and enhance automation in analysis by incorporating alternative machine learning algorithms. Future extensions of this research could explore other high performing classification models and conduct more in-depth descriptive analysis to gain further insights and determine the need for factor analysis [5]. The survey represents the performance analysis and investigation of more than 20 articles. As a result, in order to develop a machine learning model to produce a flood risk map, it is necessary to pay attention to the amount and characteristics of the training data available in the target area [24]. Expanding and refining the work to include a range of machine learning methods and real-world applications of artificial intelligence would enhance analytical automation. Again, learning more about the characteristics that are associated with rainfall in the future can lead to more advanced technology. Rainfall may be predicted using advanced machine learning and deep learning models, and even based on this, one may draw solid, data driven conclusions that are more effective in determining whether or not it will rain tomorrow [5]. In the future, we are planning to work on big datasets, and we will engage in federated learning to improve our application and model performance.

REFERENCES

[1] Syeed, M.M.A.; Farzana, M.; Namir, I.; Ishrar, I.; Nushra, M.H.; Rahman, T. Flood prediction using machine learning models. In Proceedings of the 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 9–11 June 2022; IEEE: New York, NY, USA, 2022.

[2] Kumar, V.; Azamathulla, H.M.; Sharma, K.V.; Mehta, D.J.; Maharaj, K.T. The state of the art in deep learning applications, challenges, and future prospects: A comprehensive review of flood forecasting and management. Sustainability 2023, 15, 10543. [CrossRef]

[3] P.K. Srivastava, A. Mehta, M. Gupta, S.K. Singh, andT.Islam, "Assessing impact of climate change on mundra mangrove forest ecosystem, gulf of kutch, western coast of india: a synergistic evaluation using remote sensing," Theoretical and Applied Climatology, vol. 120, no. 3, pp. 685 700, 2015.

[4] G.K Today. https://www.gktoday.in/climate-of-india-during-2021-report/?form=MG0AV3

[5] MD. MEHEDI HASSAN (Member, IEEE), MOHAMMAD ABU TARAQ RONY, MD. ASIF RAKIB KHAN, MD. MAHEDI HASSAN, FARHANA YASMIN, ANINDYA NAG1 (Member, IEEE), TAZRIA HELAL ZARIN, ANUPAM KUMAR BAIRAGI (Senior Member, IEEE), SAMAH ALSHATHRI, WALID EL-SHAFAI. Machine Learning-Based Rainfall Prediction: Unveiling Insights and Forecasting for Improved Preparedness; https://ieeexplore.ieee.org/ielx7/6287639/6514899/10320349.pdf

[6] Adel Rajab, Hira Farman, Noman Islam, Darakhshan Syed, M. A. Elmagzoub, Asadullah Shaikh, Muhammad Akram, and Mesfer Alrizq; Flood Forecasting by Using Machine Learning: A Study Leveraging Historic Climatic Records of Bangladesh, https://www.mdpi.com/2073-4441/15/22/3970

[7] STUDY REPORT KERALA FLOODS OF AUGUST 2018 by Government of India Central Water Commission Hydrological Studies Organisation Hydrology (S) Directorate, https://sdma.kerala.gov.in/wp-content/uploads/2020/08/CWC-Report-on-Kerala-Floods.pdf?utm_source=chatgpt.com

[8] NATURAL DISASTERS AND ECONOMIC DYNAMICS: EVIDENCE FROM THE KERALA FLOODS, Robert C. M. Beyer, Abhinav Narayanan, Gogol Mitra Thakur; https://cds.edu/wp-content/uploads/WP508-Dr.Gogol_.pdf?utm_source=chatgpt.com

[9] S. Al Azwari, "Predicting myocardial rupture after acute myocardial in farction in hospitalized patients using machine learning," in 2021 National Computing Colleges Conference (NCCC), pp. 1–6, IEEE, 2021.

[10] P. Mishra, A. Biancolillo, J. M. Roger, F. Marini, and D. N. Rutledge, "New data preprocessing trends based on ensemble of multiple prepro cessing techniques," TrAC Trends in Analytical Chemistry, vol. 132, p. 116045, 2020.

[11] L. Klebanov and I. Volchenkova, "Outliers and the ostensibly heavy tails," Mathematical Methods of Statistics, vol. 28, pp. 74–81, 2019.

[12] G. Agapito, C. Zucco, and M. Cannataro, "Covid-warehouse: A data warehouse of italian covid-19, pollution, and climate data," International Journal of Environmental Research and Public Health, vol. 17, no. 15, p. 5596, 2020.

[13] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," Applied Soft Computing, vol. 97, p. 105524, 2020.

[14] A. Freitas, "Comprehensible classification models: A position paper," ACMSIGKDD Explorations Newsletter, vol. 15, pp. 1–10, 03 2014.

[15] Kerala Floods – 2018 by State Relief Commissioner, Disaster Management (Additional Chief Secretary) Government of Kerala; https://sdma.kerala.gov.in/wp-content/uploads/2019/08/Memorandum2-Floods-2018.pdf?utm_source=chatgpt.com
[16] Kerala Floods, 2018 by National Institute of Disaster Management (Ministry of Home Affairs, Govt. of India) https://nidm.gov.in/PDF/pubs/KeralaFlood_18.pdf?utm_source=chatgpt.com
[17] Flood Prediction Using Machine Learning Models; https://arxiv.org/pdf/2208.01234
[18] Hussein. E, M. Ghaziasgar, C. Thron : "Regional Rainfall Prediction Using Support Vector Machine Classification of Large-Scale Precipitation Maps", IEEE 23rd International Conference on Information Fusion (FUSION), 2020. https://doi.org/10.23919/FUSION45008.2020.9190285
[19] Samantaraya. S, A. Sahoo, A. Agnihotri : "Prediction of Flood Discharge Using Hybrid PSO-SVM Algorithm in Barak River Basin", MethodsX Volume 10, 2023. https://doi.org/10.1016/j.mex.2023.102060
[20] Nadia Zehra : "Prediction Analysis of Floods Using Machine Learning Algorithms (NARX & SVM)", International Journal of Sciences: Basic and Applied Research (IJSBAR) (2020) Volume 49, No 2, pp 24-34, https://core.ac.uk/download/pdf/287366682.pdf
[21] Naveed Ahamed, S.Asha : "Flood prediction forecasting using machine Learning Algorithms", International Journal of Scientific & Engineering Research Volume 11, Issue 12, December-2020
[22] Mr. B. Samuel John Peter, Mr. N. Thilak Chandhra, Mr. Shaik.Afrid, Mr. N. Prasanna Kumar : "MACHINE LEARNING BASED FLOOD PREDICTION", International Journal of Research in Engineering, IT and Social Sciences, Volume 12 Issue 12, December 2022
[23] Rainfall Based Flood Prediction in Kerala Using Machine Learning, https://ijisae.org/index.php/IJISAE/article/view/4800
[24] FLOOD RISK MAPPING USING RANDOM FOREST AND SUPPORT VECTOR MACHINE by M.Ganjirad, M.R.Delavar; https://isprs-annals.copernicus.org/articles/X-4-W1-2022/201/2023/isprs-annals-X-4-W1-2022-201-2023.pdf?utm_source=chatgpt.com
[25] Feature scaling | Standardization vs Normalization. (2020, April 3). Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/04/feature-scaling-machi ne-learning-normalization-standardization/
[26] 1. 10. Decision trees. (n.d.). Scikit-Learn. Retrieved January 16, 2022, from https://scikit-learn/stable/modules/tree.html
[27] Decision Tree Algorithm for a Predictive Model. (n.d.). TechLeer. Retrieved January 16, 2022, from https://www.techleer.com/articles/120-decision-tree-algorithm-for-a-p redictive-model/