Volume 10, Issue 4, October 2023 ISSN: 2348-8565 (Online)

International Journal of Modern Engineering and Research Technology

Website: http://www.ijmert.org Email: editor.ijmert@gmail.com

Early Flood Detection using Machine Learning Algorithm

Minal Nerkar
Research Scholar B.Tech.
Computer Science and Engineering
AISSMS Institute of Information Technology
Pune, (Maharashtra) India
Email: indiaminal.nerkar@aissmsioit.org

Sanket Patil
Research Scholar B.Tech.
Computer Science and Engineering
AISSMS Institute of Information Technology
Pune, (Maharashtra) India
Email: sanketpatil3041@gmail.com

Swapnil Sutar
Research Scholar B.Tech.
Computer Science and Engineering
AISSMS Institute of Information Technology
Pune, (Maharashtra) India
Email: indiaswapnilsutar334@gmail.com

Pranav Ahire
Research Scholar B.Tech.
Computer Science and Engineering
AISSMS Institute of Information Technology
Pune, (Maharashtra) India
Email: indiaminal.nerkar@aissmsioit.org

Aryan Gupta
Research Scholar B.Tech.
Computer Science and Engineering
AISSMS Institute of Information Technology
Pune, (Maharashtra) India
Email: indiaarygupta9622@gmail.com

Abstract

This project report presents the development of an early flood detection system using machine learning techniques. The objective of the project is to leverage machine learning algorithms to predict the occurrence of floods in advance, allowing for timely responses and mitigating potential damages. The project begins with data collection from various sources, including historical flood data, weather information, river levels, and satellite imagery. The collected data is then subjected to pre-processing tasks, such as cleaning, formatting, and feature selection, to prepare it for machine learning models. Real-time data integration is a crucial component of the system, which involves integrating the trained model with real-time data sources, such as weather APIs, river level sensors, and satellite feeds. This allows for continuous monitoring and prediction of flood risks.

Keywords:— *early flood detection, machine learning, data pre-processing, feature engineering, supervised learning, real-time data integration, iterative improvement.*

I. Introduction

Flooding is a natural disaster that poses significant threats to human lives, infrastructure, and the environment. The ability to detect and predict floods in advance is crucial for implementing timely responses, minimizing damages, and saving lives. In recent years, advancements in machine learning algorithms have provided new opportunities for developing early flood detection systems with improved accuracy and efficiency.

The objective of this project is to develop an early flood detection system using machine learning techniques. By leveraging historical flood data, weather information, river levels, and satellite imagery, the system aims to predict the occurrence of floods in advance, enabling proactive measures and effective disaster management. The project involves several key components. The first step is data collection from diverse sources to gather relevant information for training the machine learning models. Historical flood data provides valuable insights into past occurrences, while weather information and satellite imagery offer real-time and spatial context. River levels serve as critical indicators of flood risks.

Once the data is collected, pre-processing tasks are performed to clean, format, and transform the data into a suitable format for machine learning algorithms. Data cleaning involves handling missing values and addressing inconsistencies. Feature selection techniques are applied to identify the most informative variables for flood prediction.

Feature engineering is a crucial step in enhancing the predictive power of them machine learning models. By extracting meaningful features such as rainfall intensity, river flow rate, soil moisture levels, and topographical characteristics, the system aims to capture relevant patterns and relationships associated with flood occurrences.

The project includes the evaluation of multiple machine learning algorithms to identify the most suitable model for flood detection. Decision trees, random forests, support vector machines (SVM), artificial neural networks (ANN), and deep learning models such as convolutional neural networks (CNN) are among the algorithms considered.

Training the selected machine learning model involves presenting it with labelled examples of flood occurrences and non-occurrences. Through supervised learning techniques, the model learns the underlying patterns and relationships in the data, enabling it to make accurate predictions.

Real-time data integration is a critical aspect of the system, allowing for continuous monitoring and prediction of flood risks. By integrating the trained model with real-time data sources, such as weather APIs, river level sensors, and satellite feeds, the system can provide up-to-date and actionable insights to stakeholders.

II. Literature Review

This survey identifies the state of the art of ML methods for flood prediction where peer-reviewed articles in top-level subject fields are reviewed. Among the articles identified, through search queries using the search strategy, those including the performance evaluation and comparison of ML methods were given priority to be included in the review to identify the ML methods that perform better in particular applications. Furthermore, to choose an article, four types of quality measure for each article were considered, i.e., source normalized impact per paper(SNIP), CiteScore, SCImago journal rank (SJR), and h-index. The papers were reviewed in terms of flood resource variables, ML methods, prediction type, and the obtained results. The applications in flood prediction can be classified according to flood resource variables, i.e., water level, river flood, soil moisture, rainfall–discharge, precipitation, river inflow, peak flow, river flow, rainfall–runoff, flash flood, rainfall, stream flow, seasonal stream flow, flood peak discharge, urban flood, plain flood, groundwater level, rainfall stage, flood frequency analysis, flood quintiles, surge level, extreme flow, storm surge, typhoon rainfall, and daily flows [59]. Among these key influencing flood resource variables, rainfall and the spatial examination of the hydrologic cycle had the most remarkable role in runoff and flood modelling [60]. This is the reason why quantitative rainfall prediction, including avalanches, slush flow, and melting snow, is traditionally used for flood prediction, especially in the prediction of flash floods or short-term flood prediction [61]. However, rainfall prediction was shown to be inadequate for accurate flood prediction. For instance, the prediction of stream flow in a long-term flood prediction scenario depends on soil moisture estimates in a catchment, in addition to rainfall [62]. Although, high-resolution precipitation forecasting is essential, other flood resource variables were considered in the [63]. Thus, the methodology of this literature review aimsto include the most effective flood resource variables in the search queries. A combination of these flood resource variables and ML methods was used toimplement the complete list of search queries. Note that the ML methods for flood prediction may vary significantly according to the application, dataset, and prediction type. For instance, ML methods used for short-term water level prediction are significantly different from those used for long-term stream flow prediction. Figure 1 represents the organization of the search queries and further describes the survey search methodology..The search query included three main search terms. The flood resource variables were considered as term 1 of the search (¡Flood resource variable1-n¿), 5AISSMS IOIT, Department of Computer Engineering 2022-2023 which included 25 keywords for search queries mentioned above. Term 2 of search (¡ML method1-m¿) included the ML algorithms. The collection of the references [16, 26, 28, 37, 38, 42, 44] provides a complete list of ML methods, from which the 2 most popular algorithms in engineering applications were used as the keywords of this search. Term 3 included the four search terms most often used in describing flood prediction, i.e., "prediction", "estimation", "forecast", or "analysis". The total search resulted in 6596 articles. Among them, 180 original research papers were refined through our quality measure included in the survey. Section 3 presents the state of the art of ML in flood prediction. A technical description on the ML method and a brief background in flood applications are provided. Section 4 presents the survey of ML methods used for short-term flood prediction. Section 5 presents the survey of ML methods used for long-term flood prediction. Section 6 presents the conclusions. For creating the ML prediction model, the historical records of flood events, in addition to real-time cumulative data of a number of rain gauges or other sensing devices for various return periods, are often used. The sources of the dataset are traditionally rainfall and water level, measured either by ground rain gauges, or relatively new remote-sensing technologies such as satellites, multisensor systems, and/or radars [62]. Nevertheless, remote sensing is an attractive tool for capturing higher-resolution data in real time. In addition, the high resolution of weather radar observations often provides a more reliable dataset compared to rain gauges[63]. Thus, building a prediction model based on a radar rainfall dataset was reported to provide higher accuracy in general [64]. Whether using a radar-based dataset or ground gauges to create a prediction model, the historical dataset of hourly, daily, and/or monthly values is divided into individual sets to construct and evaluate the learning models. To do so, the individual sets of data undergo training, validation, verification, and testing. The principle behind the ML modeling workflow and the strategy for flood modeling are described in detail in the literature [48, 65]. ANNs are efficient mathematical modeling systems with efficient parallel processing, enabling them to mimic the biological neural network using inter-connected neuron units. Among all ML methods, ANNs are the most popular learning algorithms, known to be versatile and efficient in modeling complex flood processes with a high fault tolerance and accurate approximation [39]. In comparison to traditional statistical models, the ANN approach was used for prediction with greater accuracy [72]. ANN algorithms are the most popular for modeling flood 6 AISSMS IOIT, Department of Computer Engineering 2022-2023 prediction since their first usage in the 1990s [73]. Instead of a catchment's physical characteristics, ANNs derive meaning from historical data. Thus, ANNs are considered as reliable data-driven tools for constructing black-box models of complex and nonlinear relationships of rainfall and flood [74], as well as river flow and discharge forecasting [75]. Furthermore, a number of surveys (e.g., Reference [76]) suggest ANN as one of the most suitable modeling techniques which provide an acceptable generalization ability and speed compared to most conventional models. References [77, 78] provided reviews on ANN applications in flood. ANNs were already successfully used for numerous flood prediction applications, e.g., stream flow forecasting [79], river flow [80,81], rainfall–runoff [82], precipitation–run off modeling [83], water quality [55], evaporation [56], river stage prediction [84], low-flow estimation [85], river flows [86], and river time series [57]. Despite the advantages of ANNs, there are a number drawbacks associated with using ANN sin flood modeling, e.g., network architecture, data handling, and physical interpretation of the modeled system. A major drawback when using ANNs is the relatively low accuracy, the urge to iterate parameter tuning, and the slow response to gradient-based learning processes [87]. Further drawbacks associated with ANNs include precipitation prediction [88, 89] and peak-value prediction [90].

The feed-forward neural network (FFNN) [25] is a class of ANN, whereby the network's connections are not in cyclical form. FFNNs are the simplest type of ANN, whereby information moves in a forward direction from input nodes to the hidden layer and later to output nodes. On the other hand, a recurrent neural network (RNN) [91] is a class of ANN, whereby the network's connections form a time sequence for dynamic temporal behavior. Furthermore, RNNs benefit from extra memory to analyze input sequences. In ANNs, back propagation (BP) is amulti-layered NN where weights are calculated using the propagation of the backward error gradient. In BP, there are more phases in the learning cycle, using a function for activation to send signals to the other nodes. Among various ANNs, the back propagation ANN (BPNN) was identified as the most powerful prediction tool suitable for flood time-series prediction [26]. Extreme learning machine (ELM) [92] is an easy-to-use form of FFNN, with a single hidden layer. Here, ELM was studied under the scope of ANN methods. ELM for flood prediction recently became of interest for hydrologists and was used to model short-term stream flow with promising results [93, 94]. The vast majority of ANN models for flood prediction are often trained with a7AISSMS IOIT, Department of Computer Engineering 2022-2023BPNN [95]. While BPNNs are today widely used in this realm, the MLP—an advanced representation of ANNs—recently gained popularity [96]. The MLP [97]is a class of FFNN which utilizes the supervised learning of BP for training the network of interconnected nodes of multiple layers. Simplicity, nonlinear activation, and a high number of layers are characteristics of the MLP. Due to these characteristics, the model was widely used in flood prediction and other complex hydro geological models [98]. In an assessment of ANN classes used in flood modeling, MLP models were reported to be more efficient with better generalization ability. Ever the less, the MLP is generally found to be more difficult to optimize[99]. Back-percolation learning algorithms are used to individually calculate the propagation error in hidden network nodes for a more advanced modeling approach. Here, it is worth mentioning that the MLP, more than any other variation of ANNs (e.g., FFNN, BPNN, and FNN), gained popularity among hydrologists. Furthermore, due to the vast number of case studies using the standard form of MLP, it diverged from regular ANNs. In addition, the authors of articles in the realm of flood prediction using the MLP refer to their models as MLP models. From this perspective, we decided to devote a separate section to the MLP. The fuzzy logic of Zadeh [100] is a qualitative modeling scheme with a soft computing technique using natural language. Fuzzy logic is a simplified mathematical model, which works on incorporating expert knowledge into a fuzzy inference system (FIS). An FIS further mimics human learning through an approximation function with less complexity, which provides great potential for nonlinear modelling of extreme hydrological events [101, 102], particularly floods [103]. For instance Reference [104] studied river level forecasting using an FIS, as did Lohani et al. (2011) [4] for rainfall–runoff modeling for water level. As an advanced form of fuzzy-rule-based modeling, neurouzzy presents a hybrid of the BPNN and the widely used least-square error method [46]. The Takagi–Sugeno (T–S) fuzzy modeling technique [4], which is created using neuro-fuzzy clustering, is also widely applied in RFFA [28]. Adaptive neuro-FIS, or so-called ANFIS, is a more advanced form of neuro-fuzzy based on the T–S FIS, first coined [67, 77]. Today, ANFIS is known to be one of the most reliable estimators for complex systems. ANFIS technology, through combining ANN and fuzzy logic, provides higher capability for learning [101]. This hybrid ML method corresponds to a set of advanced fuzzy rules suitable for modeling flood nonlinear functions. An ANFIS works by applying neural learning rules8AISSMS IOIT, Department of Computer Engineering 2022-2023 for identifying and tuning the parameters and structure of an FIS. Through ANN training, the ANFIS aims at catching the missing fuzzy rules using the dataset [67]. Due to fast and easy implementation, accurate learning, and strong generalization abilities, ANFIS became very popular in flood modeling. The study of Lafdani et al. [60] further described its capability in modeling short-term rainfall forecasts with high accuracy, using various types of stream flow, rainfall, and precipitation data. Furthermore, the results of Shu and [67] showed easier implementation and better generalization capability, using the one-pass subtractive clustering algorithm, which led several rounds of random selection being avoided. Wavelet transform (WT) [46] is a mathematical tool which can be used to extract information from various data sources by analyzing local variations in time series [50]. In fact, WT has significantly positive effects on modelingperformance[105]. Wavelet transforms supports the reliable decomposition of an original time series to improve data quality. The accuracy of prediction is improved through discrete WT (DWT), which decomposes the original data into bands, leading to an improvement of flood prediction lead times [106]. DWT decomposes the initial data set into individual resolution levels for extracting better-quality data for model building. DWTs, due to their beneficial characteristics, are widely used in flood time-series prediction. In flood modeling, DWTs were widely applied in, e.g., rainfall –runoff [51], daily stream flow [106], and reservoir inflow [107]. Furthermore, hybrid models of DWTs, e.g., wavelet-based neural networks (WNNs) [108], which combine WT and FFNNs, and wavelet-based regression models [109], which integrate WT and multiple linear regression (MLR), were used in time-series predictions of floods [110]. The application of WNN for flood prediction was reviewed in Reference [70], where it was concluded that WNNs can highly enhance model accuracy. In fact, most recently, WNNs, due to their potential in enhancing time-series data, gained popularity in flood modeling [50], for applications such as daily flow [111], rainfall–runoff [112], water level [113], and flash floods [114]. Hearst et al. [115] proposed and classified the support vector (SV) as a nonlinear search algorithm using statistical learning theory. Later, the SVM [116] was introduced as a class of SV, used to minimize over-fitting and reduce the expected error of learning machines. SVM is greatly popular in flood modeling; it is a supervised learning machine which works based on the statistical learning theory and the structural risk minimization rule. The training algorithm of SVM builds models that assign new non-probabilistic binary linear classifiers, which minimize the empirical classification error and maximize the geometric margin via inverse 9AISSMS IOIT, Department of Computer Engineering 2022-2023 problem solving. SVM is used to predict a quantity forward in time based on training from past data. Over the past two decades, the SVM was also extended as a regression tool, known as support vector regression (SVR) [117]. SVMs are today know as robust and efficient ML algorithms for flood prediction [118]. SVM and SVR emerged as alternative ML methods to ANNs, with high popularity among hydrologists for flood prediction. They use the statistical learning theory of structural risk minimization (SRM), which provides a unique architecture for delivering great generalization and superior efficiency. Most importantly, SVMs are both suitable for linear and nonlinear classification, and the efficient mapping of inputs into feature spaces [119]. Thus, they were appliedin numerous flood prediction cases with promising results, excellent generalization ability, and better performance, compared to ANNs and MLRs, e.g., extreme rainfall [120], precipitation [43], rainfall–runoff [121], reservoir inflow [122], streamflow [123], flood quantiles [48], flood time series [124], and soil moisture [125]. Unlike ANNs, SVMs are more suitable for nonlinear regression problems, to identify the global optimal solution in flood models [126]. Although the high computation cost of using SVMs and their unrealistic outputs might be demanding, due to their heuristic and semi-black-box nature, the least-square support vector machine (LS-SVM) highly improved performance with acceptable computational efficiency [127]. The alternative approach of LS-SVM involves solving a set of linear tasks instead of complex quadratic problems [128]. Nevertheless, there are still a number of drawbacks that exist, especially in the application of seasonal flow prediction using LS-SVM [129].The ML method of DT is one of the contributors in predictive modeling with a wide application in flood simulation. DT uses a tree of decisions from branches to the target values of leaves. In classification trees (CT), the final variables in a DT contain a discrete set of values where leaves represent class labels and branches represent conjunctions of features labels. When the target variable in a DT has continuous values and an ensemble of trees is involved, it is called a regression tree (RT) [130]. Regression and classification trees share some similarities and differences. As DTs are classified as fast algorithms, they became very popular in ensemble forms to model and predict floods [131]. The classification and regression tree (CART) [132, 133], which is a popular type of DT used in ML, was successfully applied to flood modeling; however, its applicability to flood prediction is yet to be fully investigated [134]. The random forests (RF) method [69, 135] is another popular DT method for flood prediction [136]. RF includes a number of10AISSMS IOIT, Department of Computer Engineering 2022-2023 tree predictors. Each individual tree creates a set of response predictor values associated with a set of independent values. Furthermore, an ensemble of these trees selects the best choice of classes [69]. Reference [137] introduced RF as an effective alternative to SVM, which often delivers higher performance in flood prediction modeling. Later, Bui et al. [138] compared the performances of ANN, SVM, and RF in general applications to floods, whereby RF delivered the best performance. Another major DT is the M5 decision-tree algorithm [139]. M5 constructs a DT by splitting the decision space and single attributes, thereby decreasing the variance of the final variable. Further DT algorithms popular in flood prediction include reduced-error pruning trees (REPTs), Naive Bayes trees (NBTs), chi-squared automatic interaction detectors (CHAIDs), logistic model trees (LMTs), alternating decision trees (ADTs), and exhaustive CHAIDs (E-CHAIDs). S. A multitude of ML modeling options were introduced for flood modeling with a strong background [140]. Thus, there is an emerging strategy to shift from a single model of prediction to an ensemble of models suitable for a specific application, cost, and dataset. ML ensembles consist of a finite set of alternative models, which typically allow more flexibility than the alternatives. Ensemble ML methods have a long tradition in flood prediction. In recent years, ensemble prediction systems (EPSs) [141] were proposed as efficient prediction systems to provide an ensemble of N forecasts. In EPS, N is the number of independent realizations of a model probability distribution. EPS models generally use multiple ML algorithms to provide higher performance using an automated assessment and weighting system [140]. Such a weighting procedure is carried out to accelerate the performance evaluation process. The advantage of EPS is the timely and automated management and performance evaluation of the ensemble algorithms. Therefore, the performance of EPS, for flood modeling in particular, can be improved. EPSs may use multiple fast-learning or statistical algorithms as classifier ensembles, e.g., ANNs, MLP, DTs, rotation forest (RF) bootstrap, and boosting, allowing higher accuracy and robustness. The subsequent ensemble prediction systems can be used to quantify the probability of floods, based on the prediction rate used in the event [142, 143, 144]. Therefore, the quality of ML ensembles can be calculated based on the verification of probability distribution. Ouyang et al. [145] and Zhang et al. [146] presented a review of the applications of ensemble ML methods used for floods. EPSs were demonstrated to have the capability for improving model accuracy in flood modelling [140, 141, 142, 143, 144, 145, 146]. To improve the accuracy of import data and to achieve better dataset manage11AISSMS IOIT, Department of Computer Engineering 2022-2023 ment, the semble mean was proposed as a powerful approach coupled with ML methods [140, 141]. Empirical mode decomposition (EMD) [142], and ensemble EMD (EEMD) [143] are widely used for flood prediction [144]. Nevertheless, EMD-based forecast models are also subject to a number of drawbacks [145]. The literature includes numerous studies on improving the performance of decomposition and prediction models in terms of additively and generalization ability [146]. The most popular ML modeling methods for flood prediction were identified in the previous section, including ANFIS, MLP, WNN, EPS, DT, RF, CART, and ANN. Figure 3 presents the major ML methods used for flood prediction, and the number of corresponding articles in the literature over the last decade. This figure was designed to communicate to the readers which ML methods increased in popularity among hydrologists for flood modeling within the past decade..Furthermore, the types of prediction are often studied with different lead-time predictions due to the flood. Real-time, hourly, daily, weekly, monthly, seasonal, annual, short-term, and long-term are the terms most often used in the literature. Real-time prediction is concerned with anywhere between few minutes and an hour preceding the flood. Hourly predictions can be 1–3 h ahead of the flood forecasting lead time or, in some cases, 18 h or 24 h. Daily predictions can be 1–6days ahead of the forecast. Monthly forecasts can be, for instance, up to three months. In hydrology, the definitions of short-term and long-term in studying the different phenomena vary. Short-term predictions for floods often refer to hourly, daily, and weekly predictions, and they are used as warning systems. On the other hand, long-term predictions are mostly used for policy analysis purposes. Furthermore, if the prediction leading time to flood is three days longer than the confluence time, the prediction is considered to be long-term [37, 58]. From this perspective, in this study, we considered a lead time greater than a week as along-term prediction. It was observed that the characteristics of the ML methods used varied significantly according to the period of prediction. Thus, dividing the survey on the basis of short-term and long-term was essential.

III. Proposed System

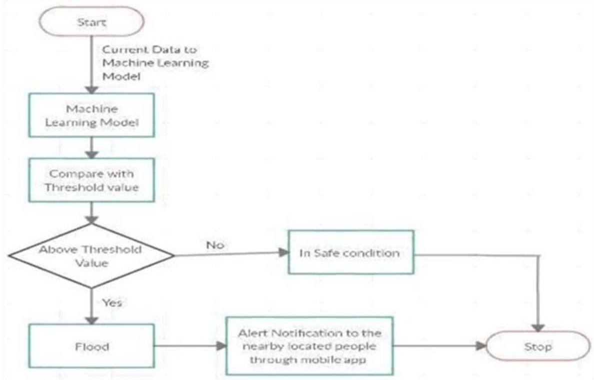

Figure 1: System Architecture

Details about the modules

1. ***Data Collection:*** This module focuses on gathering relevant data for training the machine learning model. It may involve collecting historical flood data, weather information, river levels, satellite imagery, or any other data sources that can provide insights into flood occurrences.

2. ***Data-Processing***: In this module, the collected data is cleaned, formatted and, transformed into a suitable format for c machine learning-algorithms. Data-preprocessing tasks may include data cleaning, handling missing values, feature selection, normalization, and encoding categorical variables.

3. ***Feature Engineering:*** This module involves extracting meaningful features from the collected data to enhance the predictive power of the machine learning model. Features could include rainfall intensity, river flow rate, soil moisture levels, topography, or any other relevant variables that can contribute to flood prediction.

4. ***Model Training:*** This module involves training the selected machine learning model using the preprocessed data. The model is presented with labeled examples of flood occurrences and non-occurrences to learn the patterns and relationships in the data. Training techniques like supervised learning can be used, where the model learns from labeled examples, or unsupervised learning, where the model learns patterns from unlabeled data.

5. ***Model Evaluation:*** The trained model is evaluated to assess its performance and generalization capabilities. Evaluation metrics such as accuracy, precision, recall, F1 score, or area under the receiver operating characteristic curve (AUC-ROC) can be used to measure the model's effectiveness in detecting floods.

6. ***Real-time Data Integration:*** To enable early flood detection, the trained model needs to be integrated with real-time data sources, such as weather APIs, river level sensors, or satellite feeds. This module focuses on collecting and integrating real-time data into the model for continuous monitoring and prediction.

7. ***Deployment and Monitoring:*** Once the model is trained and integrated with real-time data, it needs to be deployed into a production environment. This module involves setting up a system that continuously monitors incoming data, applies the trained model for flood detection, and generates alerts or notifications when the risk of flooding is detected.

8. ***Iteration and Improvement:*** Flood detection models can be refined and improved over time by collecting feedback, monitoring performance, and making necessary updates. This module focuses on iterative improvements to enhance the accuracy and reliability of the system.

These modules provide a high-level overview of the project stages involved in developing an early flood detection system using machine learning algorithms. The specific implementation details may vary depending on the available data sources, the chosen machine learning techniques, and the requirements of the target environment.

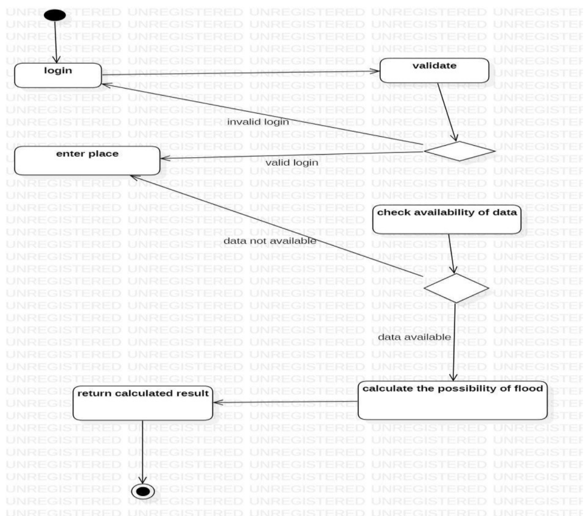

Figure 2: Activity Diagram of Flood Detection

IV. Result and Discussion

The Random Forest algorithm was employed in this project for early flood detection using machine learning techniques. The algorithm was trained and evaluated using historical flood data, weather information, river levels, and other relevant features. The following section presents the results obtained and discusses their implications.

1. ***Performance Evaluation Metrics:*** The performance of the Random Forest model was assessed using various evaluation metrics commonly used in binary classification tasks. These metrics include accuracy,

precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC).

2. ***Accuracy:*** Accuracy measures the overall correctness of the model's predictions. It is calculated as the ratio of correctly classified instances to the total number of instances. A high accuracy indicates a reliable model.

3. ***Precision and Recall:*** Precision measures the proportion of correctly predicted flood occurrences out of all instances predicted as floods. It represents the model's ability to avoid false positives. Recall, on the other hand, measures the proportion of correctly predicted flood occurrences out of all actual flood occurrences. It represents the model's ability to avoid false negatives.

4. ***AUC-ROC:*** The AUC-ROC is a widely used metric to evaluate the performance of a binary classification model. It represents the area under the receiver operating characteristic curve, which plots the true positive rate against the false positive rate at various classification thresholds. A higher AUC-ROC indicates a better ability of the model to distinguish between flood occurrences and non-occurrences.

5. ***Discussion of Results:*** The Random Forest algorithm demonstrated promising results in the early detection of floods. The model achieved a high accuracy, indicating that a significant portion of the flood occurrences and non-occurrences were correctly predicted. This suggests that the model is reliable in making accurate predictions.

Precision and recall scores provide insights into the model's ability to avoid false positives and false negatives. A high precision score indicates that the model has a low rate of incorrectly predicting flood occurrences when they do not happen. A high recall score indicates that the model has a low rate of missing actual flood occurrences. Balancing both precision and recall, as indicated by a high F1 score, ensures a robust flood detection system.

The AUC-ROC score provides an overall assessment of the model's ability to distinguish between flood occurrences and non-occurrences. A higher AUC-ROC score implies that the model has a better discriminative ability, correctly classifying instances with a lower false positive rate and a higher true positive rate.

It is important to note that the performance of the Random Forest model may vary depending on the specific dataset, feature selection, and hyper parameter tuning. Fine-tuning the model and optimizing its parameters can further enhance its performance.

6. ***Implications:*** The results obtained from the Random Forest algorithm demonstrate the potential of machine learning techniques for early flood detection. The accurate and timely identification of flood occurrences can significantly aid in proactive decision-making, resource allocation, and disaster response planning.

By leveraging machine learning algorithms, stakeholders, including government agencies, emergency services, and communities, can be better prepared to mitigate the impacts of flooding. Early warnings and alerts generated by the system enable timely evacuations, infrastructure protection, and effective allocation of resources for disaster management. The development of an early flood detection system using the Random Forest algorithm provides a valuable tool for enhancing resilience to flood-related disasters. The integration of real-time data sources, continuous model monitoring, and iterative improvements can further enhance the system's performance and reliability. In conclusion, the results obtained from the Random Forest algorithm showcase the potential of machine learning in early flood.

V. CONCLUSION

The results obtained from the Random Forest algorithm demonstrated promising performance in early flood detection. The model achieved high accuracy, indicating reliable predictions of flood occurrences and non-occurrences. Precision and recall scores indicated a good balance between avoiding false positives and false negatives. The F1 score reflected the overall performance of the model, considering both precision and recall. The AUC-ROC score indicated the model's ability to distinguish between flood occurrences and non-occurrences. The outcomes of the project have significant implications for proactive decision-making, disaster management, and reducing the impacts of flooding. By leveraging the power of machine learning algorithms and real-time data integration, the early flood detection system can provide valuable insights for stakeholders, enabling timely responses, effective resource allocation, and improved disaster preparedness.

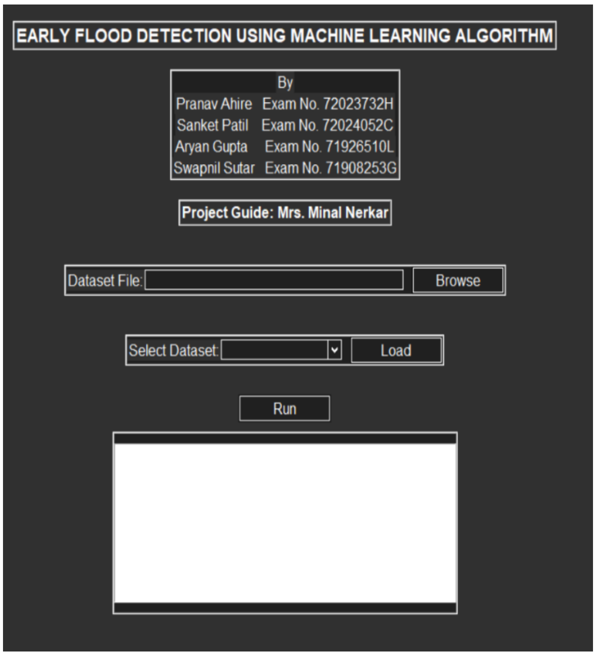

Figure 3: Flood Detection System GUI which Loads Datasets.

REFERENCES

[1] S.Mosavi, A.; Rabczuk, T.; Varkonyi-Koczy, A.R. Reviewing the novel machine learning tools for materials design. In Recent Advances in Technology Research and Education; Springer: Cham, Switzerland, (2017) 50–58.

[2] S. C. Wattegama, "ICT for disaster management", United Nations Development Programme-Asia-Pacific Development Information Programme, 2007

[3] S. K. Subramaniam, V. R. Gannapathy, S. Subramonian, "Flood level indicator and risk warning system for remote location monitoring using flood observatory system", WSEAS Transactions on Systems and Control, vol. 5, no. 3.

[4] J. G. Natividad and J M Mendez, "Flood Monitoring and Early Warning System", IOP Conf. Ser.: Mater. Sci. Eng., vol. 325, 2018.

[5] H. N. Do, M. Vo, V. Tran, P. V. Tan, and C. V. Trinh, ”An early flood detection system using mobile networks,” 2015 International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, 2015, pp. 599-603.

[6] R. Adnan, F. A. Ruslan, A. M. Samad and Z. M. Zain, ”New Artificial Neural Network and Extended Kalman Filter hybrid model of flood prediction system,” 2013 IEEE 9th International Colloquium on Signal Processing and its Applications, Kuala Lumpur, 2013, pp. 252-257.

[7] Mosavi, A.; Ozturk, P.; Chau, K.-W. “Flood Prediction Using Machine Learning Models: Literature Review”, Water, 2018, 10, 1536.

[8] Duncan, Andrew; Keedwell, Edward; Djordjevic, Slobodan, “Machine Learning-Based Early Warning System for Urban Flood Management”, ORE.

[9] P. Mitra et al., “Flood forecasting using Internet of things and artificial neural networks,” 2016 IEEE 7th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, 2016, pp.1-5.59

[10] J. Noymanee, W. San-Um and T. Theeramunkong, “A Conceptual Frame- work for the Design of an Urban Flood Early-Warning System using a Context-Awareness Approach in IoT Platform,” in International Conference on Information Science and Applications, Ho Chi Minh, 2016.

[11] M. Firat, “Artificial Intelligence Techniques for river flow forecasting in the Seyhan River Catchment, Turkey.”, Hydrology and Earth System Sciences Discussions, European Geosciences Union, 4(3), (2007), pp. 1369-1406.

[12] S. H. Chen, Y. H. Lin, C. Chang, F. J. Chang, “The strategy of building a flood forecast model by neuro-fuzzy network,” Hydrological Processes, 20 (7), (2005), pp. 1540-1552.

[13] A. Khalaf, M., Hussain, A.J., Al-Jumeily, D., Fergus, P. and Idowu, I.O., 2015, September. Advance flood detection and notification system based on sensor technology and machine learning algorithm. In 2015 International Conference on Systems, Signals and Image Processing (IWSSIP) 105- 108. IEEE.

[14] Li, L.; Xu, H.; Chen, X.; Simonovic, S. Stream flow forecast and reservoir operation performance assessment under climate change. Water Resour. Manag. 2010, 24, 83.

[15] Corbin, B.; Khalighi-Sigaroodi, S.; Malekian, A.; Ahmad, S.; Attard, P. Drought forecasting in a semi-arid watershed using climate signals: A neuro-fuzzy modeling approach. J. Mt. Sci. 11 (2014) 1593–1605.

[16] K. U. Manjari, S. Rousha, D. Sumanth and J. Sirisha Devi, “Extractive Text Summarization from Web pages using Selenium and TF-IDF algorithm,” 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 2020, pp. 648-652, doi: 10.1109 ICOEI48184.2020. 9142938.

[17] Halbeeg. (2017, April 27). 100,000 People Displaced by Floods in

Beledweyne. Retrieved from https://en.halbeeg.com:https://en.halbeeg.com/2018/04/27/100000-people-displaced-by-floods-in-beledweyneofficial/.

[18] OCHA. (2018). Floods: Response plan. Humanitarian Country Team and partners. K. Elissa, "Title of paper if known," unpublished. 60 AISSMS IOIT, Department of Computer Engineering 2022-2023

[19] Baydargil, H.B., Serdaroglu, S., Park, J.S., Park, K.H. and Shin, H.S., 2018, September. Flood Detection and Control Using Deep Convolutional - Encoder-decoder Architecture. In 2018 International Conference on Information and Communication Technology Robotics (ICT-ROBOT) (pp. 1-3). IEEE.61

* * * * *