Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 23 | No. 4 | pp. 1885-1900 | 2024

Original Research Paper

https://doi.org/10.46488/NEPT.2024.v23i04.002

Open Access Journal

Dynamic Impact-Based Heavy Rainfall Warning with Multi-classification Machine Learning Approaches

Anand Shankar[1, 2†]

[1]Department of Electronics & Communication Engineering, National Institute of Technology, Patna, India

[2]Indian Meteorological Department, Ministry of Earth Sciences, Govt. of India, Patna, India

†Corresponding author: Anand Shankar; anand.shankar@imd.gov.in

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

Received: 02-02-2024
Revised: 14-03-2024
Accepted: 29-03-2024

Key Words:
Impact-based heavy rainfall warning
Multi-classification machine learning
Impacts of floods
Flood assessment
Cascading impact

ABSTRACT

The majority of flood assessment and warning systems primarily focus on the occurrence of floods caused by river overflow, taking into account factors such as intense precipitation. Improving flood resilience, on the other hand, requires a deeper understanding of how these factors affect each other and how specific local conditions can have an impact. This study offers impartial tools for estimating the severity of the effects brought on by heavy rainfall to facilitate the prompt communication of effective measures, such as the evacuation of livestock and human settlements and the provision of medical assistance. These tools take into account the cascading effects of various causative factors contributing to heavy rainfall. This article aims to assess the various factors that contribute to the impacts of heavy rainfall, including the timestamp (indicating soil saturation and moisture levels), river gauges (determining water congestion in canal systems), average aerial precipitation (indicating runoff), and the rainfall itself, taking into account both in situ and ex-situ impacts. Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbour (KNN), and Naive Bayes are some of the machine learning methods used in the study to find out how dynamically vulnerable affected districts are to flooding in different compound scenarios. This analysis is conducted by leveraging historical observed datasets. The results demonstrate the feasibility of mitigating the issue of excessive and insufficient flood warnings resulting from the cumulative effects of intense precipitation. By implementing a categorization system that divides the affected areas into various portions, or districts, according to the main factors contributing to flooding, namely rainfall, river discharge, and runoff, The suggested model presents novel insights into the sequential consequences of intense precipitation in the regularly inundated regions of North Bihar, India. Innovative tools can serve as valuable resources for flood forecasters and catastrophe managers to comprehend the extent of flooding and the consequential effects of intense precipitation.

INTRODUCTION

Several studies conducted by (Goswami et al. 2006), (Zou & Ren 2015), and (Abbas et al. 2023) have shown an increase in the frequency of heavy or extreme rainfall events in various regions, indicating a pan-Asian phenomenon. Similarly, global research conducted by (Brunetti et al. 2004), (Groisman et al. 2005), and (De Luis et al. 2011) has also observed a similar trend of extreme rainfall occurrences across different parts of the world. A higher concentration of greenhouse gases in the atmosphere may make it more likely for heavy rain to occur (Easterling et al. 2000, Groisman et al. 2005). This is supported by many studies and real-world evidence from observations or model predictions. Increasing concentrations of greenhouse gases are blamed for the shift in the frequency of severe rainfall (Meehl & Tebaldi 2004).

Regional land-use and land-cover changes (LULC) have been observed to have an impact on mesoscale convection, as noted by (Pielke et al. 2011) and (Niyogi et al. 2017). The frequency of floods has exhibited a notable upward trend during the past three decades, as documented by (Najibi & Devineni 2018). The susceptibility of different regions within the country to flooding is attributed to the unanticipated precipitation patterns resulting from the geographical and hydrological characteristics of the subcontinent. The primary factors contributing to floods, which frequently lead to the loss of human lives and property, include heavy precipitation, obstruction in river outflow and associated canal systems, uncontrolled urban development, and alterations in land use and land cover. The adverse impacts of floods, resulting from the overflow of water from various sources such as riverine flooding, runoff, in-situ rainfall, and groundwater status, can be attributed to factors such as exposure, lack of resilience, and inadequate early warning systems. These negative effects encompass a wide range of areas, including infrastructure, human health, economic activity, and the environment (Beevers et al. 2012). According to (Mohapatra et al. 2021), there has been an improvement in the precision of predicting extreme precipitation within a short- to medium-term timeframe of up to five days. Nevertheless, the current level of proficiency in predicting and alerting to heavy precipitation is inadequate for mitigating the risks associated with potential casualties and property damage. To safeguard both human lives and economic stability, it is imperative to expand the scope of weather forecasting and warning systems beyond solitary predictions of severe weather events. This entails including impact-based forecasting systems, which utilize impact modeling techniques, and subsequently integrating impact estimation systems that employ impact-based modeling. This comprehensive approach is crucial for facilitating appropriate response actions.

According to reports, the flooding in North Bihar affected a significant proportion of the population, specifically 76%. Each year, a significant number of individuals, along with their domesticated animals, become displaced as a result of the catastrophic floods occurring in the Indian state of Bihar. The catchment areas of the rivers that comprise the primary river systems in northern Bihar, namely the Gandak, the Bagmati/Adhawara, and the Kosi/Mahananda, are located within the mountainous region of Nepal. Originating in Nepal, these rivers traverse densely populated regions of North Bihar, including Supaul, Araria, west Champran, east Champaran, Sitamarhi, and other locations. The slope of these rivers exhibits a gradual decline, transitioning from a rate of 6 meters per kilometer to 6 centimeters per kilometer near the Gangetic floodplain. During the monsoon season's intense precipitation, the lower catchment areas of the river basin experience a significant influx of run-off. The occurrence of intense precipitation in the vicinity of the Himalayan foothills, particularly near Nepal and other geographical regions, resulted in a sudden and significant rise in water levels, commonly catalyzing the onset of flooding events. Hence, the considerable inundation observed in these river systems can be attributed to the interplay between precipitation in the catchment areas and precipitation in the higher tributary regions. Hence, the comprehensive inundation of settlements situated on either side of the embankment of these rivers is a multifaceted interaction between the water level at a specific moment in a specific river section, the actual precipitation in the associated catchments, and the precipitation in the downstream catchment area. The understanding of both the individual and combined effects of these flooding mechanisms is often limited, resulting in either an underestimation or overestimation of the consequences of such an event, depending on the information source.

Meteorological systems that range in scale from the synoptic to the local scale have an impact on the occurrence of flooding(Reddy et al. 2008, Pandit 2009, Ranalkar et al. 2016, Kumar et al. 2021, Prasad et al. 2021). These systems contribute to the process of flooding by facilitating the transfer

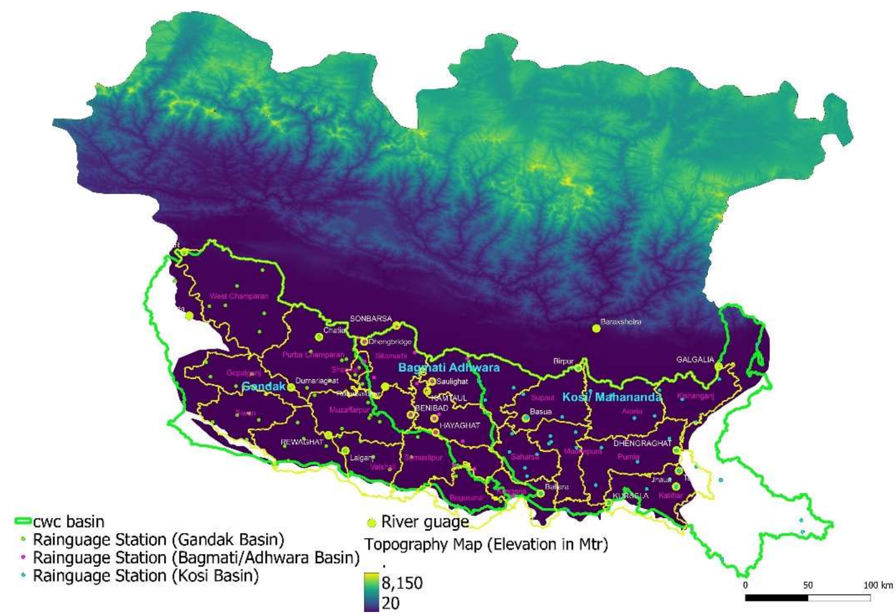

Fig. 1: The topographic map of the district of North Bihar, where the orography makes the location vulnerable to flooding.

of rainfall to river systems through numerous tributaries. This process occurs across a wide geographical area characterized by intricate orography, as illustrated in Fig. 1 The inundation of the Ganga River's catchment area, predominantly located in the northern region of Bihar, resulted in widespread flooding across the state. This occurrence was primarily attributed to the excessive precipitation experienced during the summer monsoon season, leading to the river's water exceeding its normal capacity and inundating floodplains and canals. The objective assessment conducted by (Shankar et al. 2022) examines the cascading impacts of substantial river releases and the accompanying high rainfall in North Bihar, India, which has led to recurrent flooding incidents. There isn't a lot of written material that fully talks about the dynamic vulnerability assessment of the cascading effects of in situ rainfall and how it affects river levels, as well as the right order of events over time. The objective of this study is to address the existing knowledge gap by implementing the techniques utilized by operational forecasters and disaster managers to do precise risk and vulnerability evaluations. This would consequently diminish the probability of both excessive and insufficient warnings.

In the realm of natural hazards, the most often employed vulnerability assessment approaches encompass historically derived impact data, evaluation indicator systems, hazard loss curves, and machine learning techniques. The presence of localized historical hazard data for every research site is crucial for the efficacy of the historical hazard-oriented approach. Although this methodology proves highly effective in facilitating cross-regional comparisons, its predominant use lies across expansive study domains, rendering the evaluation of specific phenomena challenging. Nevertheless, (Goyal et al. 2021) highlighted the fact that the assessment indicator system's methodology is highly subjective, primarily as a result of the arbitrary nature of weight computation. The use of expert scoring is a common method for assigning weight, as (Moghadas et al. 2019) and (Xu et al. 2023) demonstrate. However, it is important to note that this method is inherently subjective and has the potential for substantial errors in the ultimate evaluation outcomes. The use of a hazard damage curve, alternatively referred to as a vulnerability curve, facilitates the evaluation of the correlation between the magnitude of a cause and the resultant impact incurred by the entities that are most susceptible to its impact. The present methodology can assess vulnerability outcomes for a given geographical area through the measurement of the extent of harm incurred by distinct characteristics. The susceptibility of social and economic structures is an often-discussed subject when examining the impacts of catastrophic occurrences. To evaluate the susceptibility to disasters, scholars have conventionally placed significant reliance on the direct use of machine learning methodologies. While the use of extensive data sets for training purposes can lead to precise vulnerability forecasts, the application of vulnerability mapping specifically for flood assessments remains unexplored and is considered a nascent topic.

This research provided a comparative analysis of several multiclassification methods, including Decision Tree (DT), Naive Bayes, Support Vector Machine (SVM), and K-Nearest Neighbour (KNN), in the context of dynamic impact-based forecasting utilizing machine learning techniques. This paper presents an analysis of the impacts resulting from rainfall and associated systems, utilizing the most up-to-date historical data sets of impacts specifically about the districts of north Bihar. This research article discusses the investigation of the impacts in three districts, specifically West Champaran, Darbhanga, and Vaishali. This work aims to address a research gap in the quantitative analysis of dynamic impact-based forecasting and its corresponding vulnerability assessment. The article's organizational structure is outlined as follows: the discussion of the datasets and their corresponding research areas is presented in Section 2. Following this, Section 3 explains the suggested strategy, and Section 4 presents the findings. Sections 5 and 6 then provide a thorough analysis and conclusion, respectively.

STUDY AREAS AND DATASET

Study Areas

The scope of this study is the prediction of dynamic impacts and the assessment of dynamic vulnerability related to rainfall and accompanying floods in three selected districts of North Bihar, namely West Champaran, Darbhanga, and Vaishali, as depicted in Fig. 2 The districts chosen for this study were based on a combination of different geographical settings and various causative factors for the impacts of floods. West Champaran, the largest district, was included due to its significant size and influence in the region and its proximity to the central region of Nepal. Additionally, the topography of the Himalayan mountain range has a significant impact on the rainfall patterns in this region. The second district, Vaishali, exemplifies the characteristics of flat lands located at a distance from the Himalayan foothills. The selection of the third district, Darbhanga, is based on its geographical location within the catchment areas of the Bagmati and Adhawara rivers (see Fig. 2). The geographical extent of the region under consideration encompasses the longitudes ranging from 84° E to 88.5° E, together with the latitudes spanning from 25° N to 27.6° N. The Upper Ganga basin has a multitude of rivers that traverse the northern region of Bihar. The Gandak River basin spans an area of

58,800 square kilometers, while the Bagmati/Adhawara River covers 18,845 square kilometers. Additionally, the Kosi/Mahananda river basin extends over an area of 95,552 square kilometers (see Fig. 2). The governments of Bihar (India) and Nepal have implemented several remedial steps to establish regular monitoring systems for floods and safeguard the low-lying regions of North Bihar (Sinha et al. 2008). Subtropical monsoons have a significant impact on the state of Bihar in India. The hydrological regime of this particular state is significantly impacted by its climate, thereby influencing a diverse range of geographical features. Approximately 84.8% of the state's yearly precipitation is derived from the monsoons, a meteorological phenomenon resulting from the interaction between easterly and westerly monsoonal winds, owing to the state's particular geographical positioning. The economic development of the local area is contingent upon the sectors of agriculture, aquaculture, horticulture, and tourism, all of which are susceptible to fluctuations in climatic conditions. Consequently, the effective control of floods in the rivers is of paramount importance.

Dataset

The details of the datasets utilized for the creation of the dynamic impact-based rainfall forecasting system are presented in Table 1.

Therefore, the input characteristics are obtained from datasets containing information on river levels, average aerial precipitation, and district rainfall. The corresponding impact datasets are also collected, and the level of impacts is measured using the methods outlined in Fig. 3 and 4. The dataset comprised 70% of the total data, which was used for training the models. The remaining 30% of the dataset was utilized for evaluating the performance of the trained models. The time steps taken throughout the training process were also recorded. The impact datasets before quantification are displayed in Table 2.

MATERIALS AND METHODS

AI/ML-Based Impact Modeling

This approach involves the development of a model that

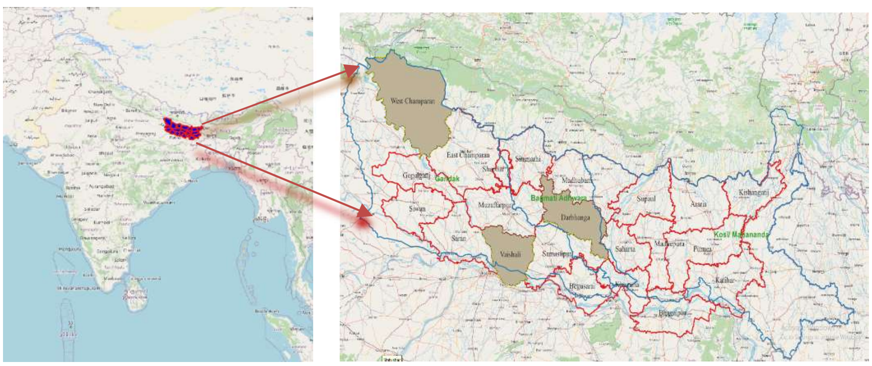

Fig. 2: Map of North Bihar (Districts, Catchments, Water Body) and the Sample Districts (West Cahmparan, Darbhanga, and Vaishali).

Table 1: The details of the datasets used in the dynamic impact-based rainfall prediction system.

Type of the Datasets	Name	Period	Unit	Source
River levels	River gauge of associated districts	June to October(2020-2022)	In metre	Central Water Commission, Govt. of India.
Rainfall	Point Rainfall		In mm	India Meteorological Department, Govt. of India
Overall Aerial Precipitation of the Catchement	Interpolated Rainfall(Storm Analysis)		In mm	Flood Meteorological Office, IMD, Govt. of India
Associated impacts datasets	Loss of Life(population)		quantification	Disaster Management Department, Govt. of Bihar, India
	Evacuation of population		quantification	
	Crop damages		quantification	
	Dwelling Unit damages		quantification	

Table 2: Sample of impact Dataset prepared by the executive agency and Compiled by the Disaster Management Department, Govt. of Bihar.

Date	Name of District	No. of affected: No. of Blocks	No. of affected: Panchayats Fully	No. of affected: Panchayats Partly	No. of affected: Village	No. of affected: Persons lac	No. of affected: Animals lac	No. of affected: Affected area (lac ha) Agric	Estd crop damage (Rs lac)	Houses damaged: Pucca Fully	Houses damaged: Pucca Partly	Houses damaged: Kucha Fully	Houses damaged: Kucha Partly	Houses damaged: Huts	Houses damaged: Total	Houses damaged: No. of Damage Cattle Shade	Public property damaged (Rs lac)	Lives lost: Human	Lives lost: Animal	Population evacuated	No. of villages surrounded by water: Population	Population on roads/rail tracks/rooftops etc.
15-Aug-21	Vaishali	7	20	19	65	1.9	0.0	0.2	0	0	0	0	0	0	0	0	0	0	0	608	608	65
31-Aug-21		9	21	49	152	3.0	0.0	0.2	0	3	7	3	7	11	31	13	10.5	4	0	830	0	152
01-Sep-21		10	21	61	176	3.4	0.0	0.2	0	3	7	3	7	11	31	13	10.5	4	0	830	0	176
15-Sep-21		16	200	90	366	4.3	0.1	0.6	8286	0	0	0	0	0	0	0	0	6	5	830	0	366
30-Sep-21		16	203	91	372	4.6	0.1	0.6	8286	0	0	0	0	0	0	0	0	6	5	830	0	372
01-Oct-21		16	203	91	372	4.6	0.1	0.6	8286	0	0	0	0	0	0	0	0	6	5	830	0	372
15-Oct-21		16	204	90	398	5.0	0.1	0.6	8286	0	0	0	0	0	0	0	0	16	10	830	0	392

integrates impact magnitude(quantified) or vulnerability, exposure data sets, i.e., the causative factors of the floods (both in-situ and ex-situ), and the current states of the districts in terms of timeframe. Fig. 3 shows the four levels of impact quantification based on the combined impacts of the factors that caused the floods. These levels are green, yellow, orange, and red. The historical impact data is confirmed based on the flow chart outlined in Fig. 5. In cases of severe weather, the causative factor, which is based on its ability to have an impact, is largely constant (the five factors outlined in Fig. 4). Conversely, the vulnerability of the affected population is contingent upon the extent of their exposure. Graded impact warning systems play a crucial role in the management of population vulnerability and exposure by providing a means to estimate the amount of impact. The causative factor, which is based on the potential for impact, is relatively constant in cases of severe weather events. Conversely, the vulnerability of the affected population is contingent upon the degree of exposure. The implementation of impact-graded warning systems plays a crucial role in effectively managing and mitigating people's exposure to various causative factors. Therefore, comprehending the detrimental effects of various meteorological phenomena is crucial for effectively mitigating natural disasters like floods. Based on the matrix presented in Fig. 3, the quantification of the graded impacts (0-green, 1-yellow, 2-orange, and 3-red) at four levels was carried out. by following the standard operating procedure of the India Meteorological Department, India (IMD, Ministry of Earth Sciences 2021).

The current work uses machine learning techniques to create a decision tree model whose goal is to figure out the best level of graded warning issue when there is heavy rain and the impact systems that go with it. The local administrative body has used the precise graded warning to

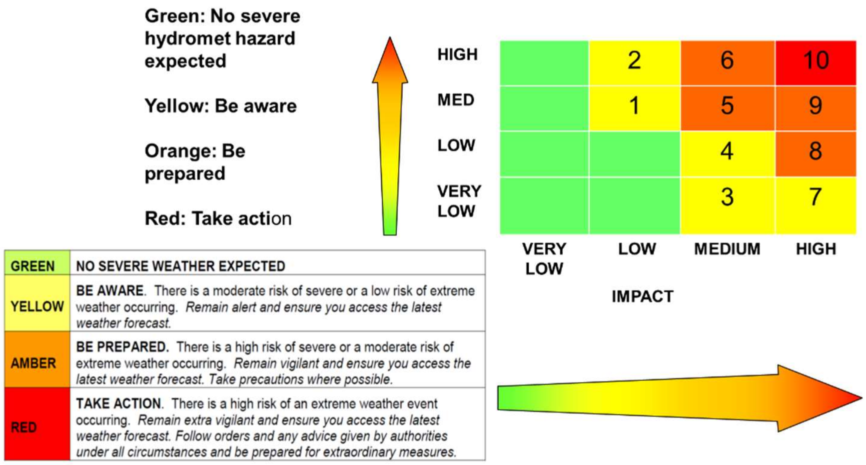

Fig. 3: Level of impacts and associated matrix (quantification of the graded impacts: 0-green, 1-yellow, 2-orange, and 3-red).

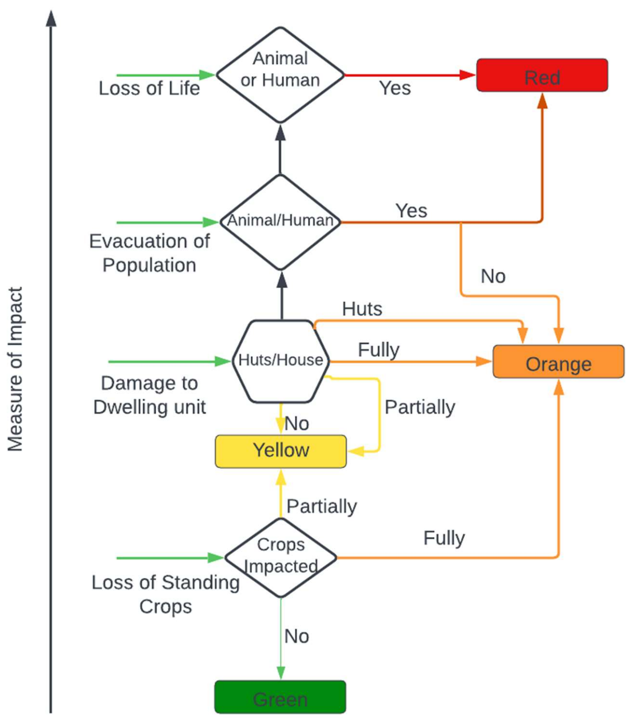

Fig. 4: Objective criteria (methodology) used for the quantification of the graded warning system (0-green, 1-yellow, 2-orange, and 3-red) for the studied areas.

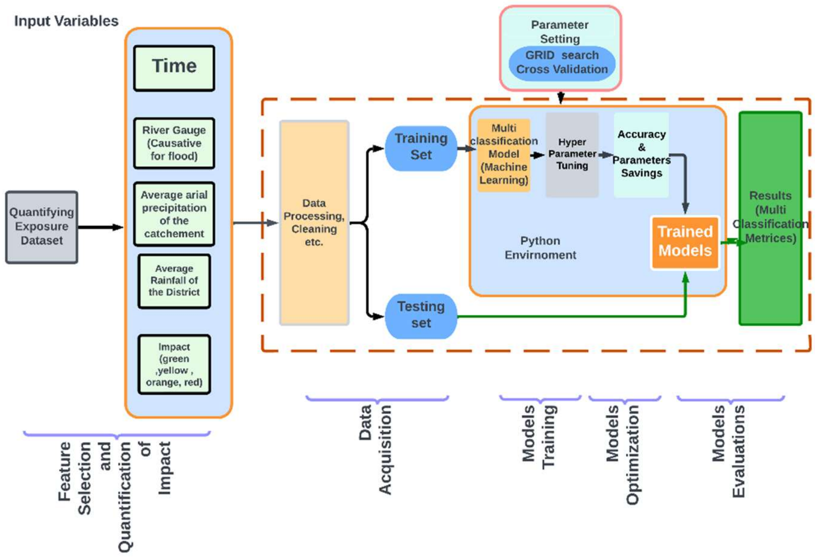

Fig. 5: Process block diagram of dynamic impact-based prediction system.

evaluate the extent of damage and potential threats resulting from the occurrence of heavy rainfall. The levels of impacts of flood assessment include the examination of existing crops, the assessment of damage to residential structures, the appraisal of population evacuation measures, and the analysis of casualties (presented in Fig. 4).

The proposed model exhibits considerable potential for utilization and possesses substantial value in terms of its practicality. Fig. 5: is a graph that shows the steps that were used to make the machine learning-based multiclassification algorithm that was used in the dynamic impact-based heavy rainfall prediction system. The steps included finding historical impact data, quantifying the exposure data, training the multiclassification algorithms, etc.

Multi-Classification Algorithm(M\L)

Machine learning, a branch within the broader domain of artificial intelligence, has become a focal point for addressing the challenges of digitalization, attracting significant attention within the digital realm. In this study, the multiclassification of the assessment and prediction of the impacts of heavy rainfall was conducted using four machine learning models: Support Vector Machines (SVM), Decision Trees (DT), k-Nearest Neighbors (KNN), and Naive Bayes (NB).

Support Vector Machine

Support Vector Machines (SVM) are capable of handling difficulties including classification as well as regression. The decision boundary in this approach is a hyperplane, which must be determined. A decision plane is necessary to divide a set of objects into their respective classes when there are several classes represented. Separating items into their respective classes may or may not need complex mathematical functions known as kernels if the objects are not linearly separable. SVM attempts to correctly classify objects based on examples in the training data set. The benefits of support vector machines (SVM) include: They work well with semi-structured and structured data, and they can even deal with complex functions if the right kernel function is generated. Overfitting is reduced by SVM's reliance on generalization.

This method works well with high-dimensional data and scales well. Local minima are not a problem for it. The longer it takes to train an SVM, the less well it performs with more data. An adequate kernel function will be hard to locate. When the dataset is noisy, SVM performs poorly. If there are a lot of features and observations, then SVM is worth a go (Ray 2019).

Decision Tree

Decision trees are supervised machine learning that can be used to resolve classification and regression issues by continually separating data based on certain parameters. The leaves are where the decisions are made, whereas the nodes are where the data is divided up. In multiclassification decision variable is graded. The decision in the leaves and the data split in the nodes. The benefits of using a Decision Tree include its adaptability to both regression and classification problems, its straightforward interpretation, its ability to handle both quantitative and qualitative values, its capacity to fill missing values in attributes with the most likely value, and its high-performance thanks to the efficiency of its treetraversal algorithm. Over-fitting is an issue that can arise with Decision trees, but it can be remedied with the help of the ensemble modeling approach that Random Forest employs. it can be unstable, it's not always easy to regulate the tree's size, it's vulnerable to sampling error, and it only provides a locally optimal solution, not the best possible one.

Naïve Bayes

This algorithm relies on conditional probability and is therefore easy to implement. In this approach, training data is used to define a probability table that serves as the model. The "probability table" uses feature values to anticipate fresh observations by looking up class probabilities. Predicting a new observation requires consulting a "probability table" whose columns include the feature values and whose rows contain the class probabilities. The term "naive" refers to the underlying premise of conditional independence. Taking into account all input features as though they were unrelated to one another is unrealistic in practice.

The benefits of Naive Bayes (NB) include straightforward implementation; high performance; use of a smaller sample of training data; linear scalability in terms of predictors and data points; the ability to deal with both continuous and discrete data; the capability to deal with multi-class classification problems; and the ability to make probabilistic predictions. Both continuous and discrete data types are supported. It has a low sensitivity to non-essential details. Properly trained and optimized models typically outperform NB models because NB models are overly simplistic. It is challenging to directly apply Naive Bayes if one of the features must be a "continuous variable," such as time. While "buckets" can be created for "continuous variables," they are not always accurate.

Because there is no genuine online form of Naive Bayes, all data must be retained for retraining the model. For a large enough number of classes, say above 100,000, it will not scale. It requires more memory at runtime for prediction than support vector machines or standard logistic regression. High-end central processing unit (CPU), especially for complex models with lots of variables.

K Nearest Neighbour Algorithm (KNN)

It is an example of a classification method. The algorithm(KNN) uses a database partitioned into classes into which it must place a single sample data point and solve the ensuing classification problem. It is said that KNN is non-parametric since it does not make any assumptions about the underlying data distribution. It's a straightforward method with easy implementation. It's an adaptable system that works well with multi-modal classification. Several category tags appear in the record set. The rate of error is no more than twice the Bayes error rate. In some cases, this is the most effective technique.

Multi Classification Performances Metrices

Classification is a commonly used technique in data analysis that involves the categorization of data into more than two categories. While the traditional approach involves the separation of data into two groups, known as binary classification, it is also possible to extend this method to encompass more than two groups, which is known as multi-class classification. From an algorithmic perspective, (Mohandes et al. 2018) discuss how the prediction process is dependent on advanced mathematical techniques. These methodologies employ the provided input data (represented by the x variables) to provide precise predictions for the outcome variable y. Given that y is a variable that can take on values from 1 to K, where each value represents a unique class, it is possible to see both the response variable y and the prediction (y-hat) as discrete random variables in the context of multi-class classification. The algorithm calculates the probability that a given unit is a member of a particular class and subsequently utilizes a classification rule to allocate each unit to one of these classes. In general, the rule is characterized by its simplicity: an object is allocated to the category that exhibits the greatest probability. In the context of employing a classification model, it is possible to make estimations regarding the probability of membership for each potential unit inside a given class. In the context of a binary classification task, it is customary to use a threshold value to determine the appropriate class prediction for each instance while taking the model's probability output into account (Grandini et al. 2020).

Confusion Matrix: The confusion matrix is a cross table that keeps track of the frequency with which the true/actual classification differs from the expected classification(Fig. 6).

The details of the performance metrics used in this research article are presented in Table 3.

RESULTS AND DISCUSSION

This section focuses on assessing the efficacy of the quantification of causative factors contributing to floods in North Bihar. Additionally, it examines the effectiveness of a proposed multi-classification algorithm in categorizing the impact of floods in the districts of West Champaran, Darbhanga, and Vaishali in the studied areas. The selection of these districts was made by considering the geographical settings and the diverse factors associated with floods in these areas. The implementation code was developed using Python 3.10. A machine learning model for simple

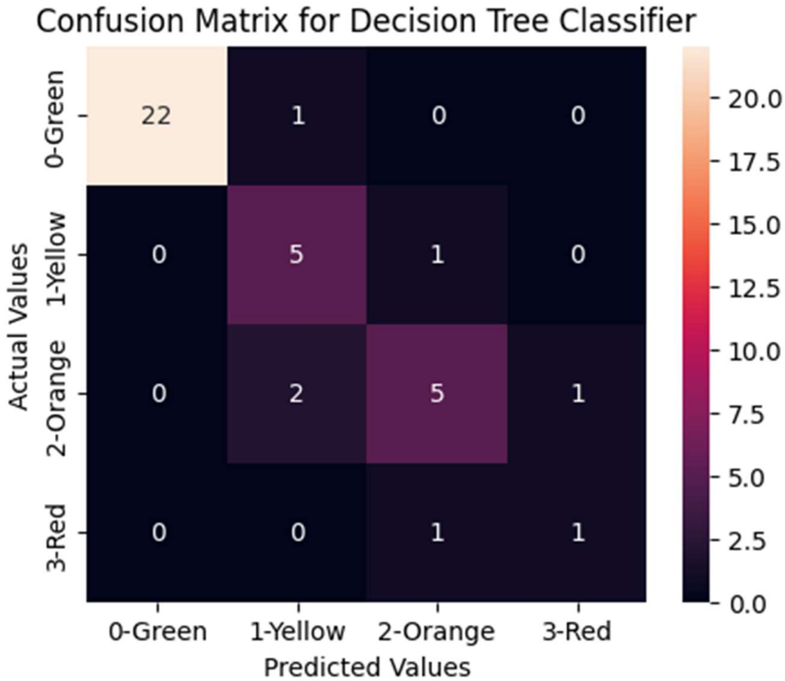

Fig. 6: Schematic of Confusion Matrix on the test datasets.

Table 3: Details of the performance matrices used in the assessment of multiclassification ML approaches.

Performance Metrics	Remarks
$Balanced\ Accuracy = \frac{\frac{TP}{Total_{row1}}+\frac{TN}{Total_{row2}}}{2}$ (1)	It is the average value of recall of each class. How likely is it that an individual of that class will be classified correctly? The recall value for each class response. Therefore, balanced accuracy provides a class-general mean measure of this idea.
$Macro\ F1 - Score = 2 * (\frac{Macro\ Average\ Precision * marcro\ Average\ Recall}{macro\ AveragePrecsion^{-1} + Macro\ Average\ Recall^{-1}}$ (2)	Since the numerators of Macro Average Precision and Macro Average Recall are composed of values in the range [0, 1], macro-average methods often compute an overall mean of several metrics. There is no correlation between class size and the denominator because classes of varying sizes are counted the same. This means that the influence of the largest classes is just as significant as that of the smallest. High Macro-F1 values suggest that the method performs well across all classes, while low Macro-F1 values show classes that are poorly predicted by the system.
Micro-Average F1 Score=$\frac{\sum_{k=1}^{K} TP_k}{Grand\ Total}$ (3)	Macro F1-Score is an average measurement of the classes' average precision and average recall. This metric is calculated at the class level so that each class receives equal weight. Small classes are equivalent to large ones, and algorithm performance regardless of class size is of equal importance.
Cross EntropyH(p(y_i),p($\hat{y}_i$))=-$\sum_{k=1}^{K} \sum p(Y = k \mid X_i) logp(\hat{Y}_i = k \mid X_i)$ (4)	Cross-entropy values of the individual units to derive a measure of agreement for the entire dataset. Cross-entropy exploits only the value of p(Y^ i = k|Xi) for the k value representing the true class.

multiclassification was implemented on a laptop running the Windows 11 operating system. The laptop is equipped with an Intel (R) Core (TM) i5-1035G1 processor working at a frequency of 1.00 GHz and has 8 GB of memory. The specifics regarding the datasets are outlined in Section 2. The process of the proposed methodologies is outlined in subsection 3.1, and the subsequent optimal hyperparameters are presented in Table 4.

Quantification of Causative Factors of The Impact of Floods

The studied areas, i.e., districts of North Bihar, lie in the basins of the Gandak, Bagmati/Adhawara, and Kosi/Mahananda groups of rivers, which are prone to flooding. The catchments of the Gandak, Bagmati/Adhawara, and Koshi/Mahananda rivers are located in the mountainous central and eastern regions of Nepal (presented in Fig. 1). Consequently, any amount of rainfall in adjoining Nepal, even if it is of moderate quantity, leads to an increase in the water level of these river systems. Depending on the varying capacity of the rivers, floods occur in the downstream areas. The river system mentioned in this study causes flooding in the lowlands of northern Bihar, India, when it traverses the border from Nepal, where the terrain is steeper (Shankar et al. 2022). The steep topography of the region leads to sudden or flash floods in the downstream districts of North Bihar during periods of excessive rainfall in neighboring Nepal. Therefore, it is seen that floods can occur in the lowlands of Bihar even in the absence of significant in-situ rainfall. To establish a comprehensive warning system for heavy rainfall-induced floods, our objective is to uncover the underlying elements that contribute to these impacts. The assessment of impacts has been derived from historical data about several factors, including the loss of human lives, displacement of communities, destruction of residential structures, and losses to crops. This technique is derived from the discussion in Fig. 4, which focuses on the quantification of the graded warning system. The objective of this approach is to verify that the graded warning aligns with the probability of potential impacts. The inputs for this study include various causative factors, such as the river gauges at different locations, which may be influenced by rainfall in neighboring regions of Nepal. Additionally, the average aerial precipitation response for water runoff and district rainfall are also considered. As shown in Fig. 5, these inputs come with the corresponding time data. The details of these elements have been assessed for each district. This study presents an assessment of three districts from the perspective of rational representation. A comprehensive examination of three districts, namely West Champaran, Vaishali, and Darbhanga, is presented in the following subsection. West Champaran is situated in the higher catchment area of the river, while Vaishali encompasses both the lower catchment area of the river and the region affected by the Ganga River. Lastly, Darbhanga is a district prone to flooding. This section focuses on the evaluation of the quantification of causative factors contributing to floods in North Bihar. It also checks how well a suggested multi-classification algorithm sorts the graded impact-based flood warning system or the dynamic impact-based vulnerability assessment works in the districts of Darbhanga, Vaishali, and West Champaran. The selection of these districts took into account their geographical settings

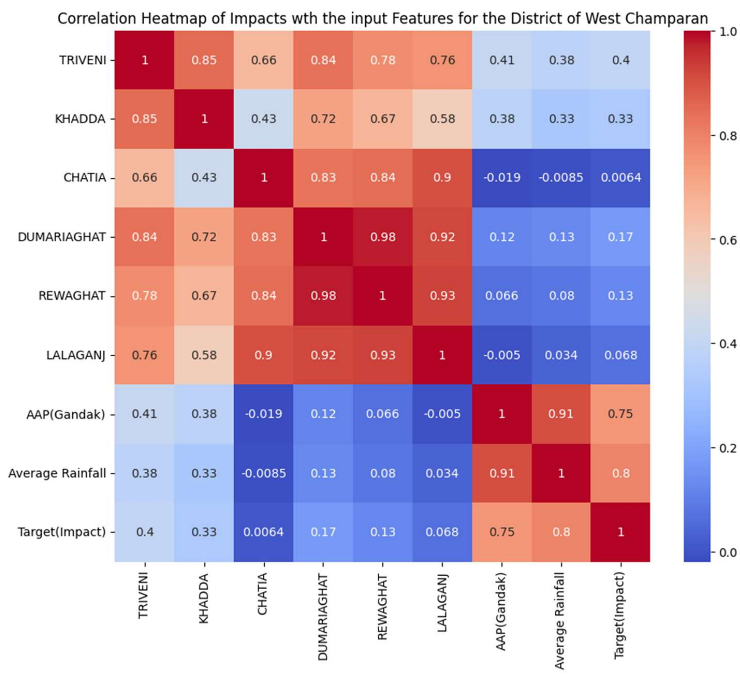

Fig. 7: The correlation matrix of the impacts of heavy rainfall with its causative feature in the districts of West Champaran, Bihar, India.

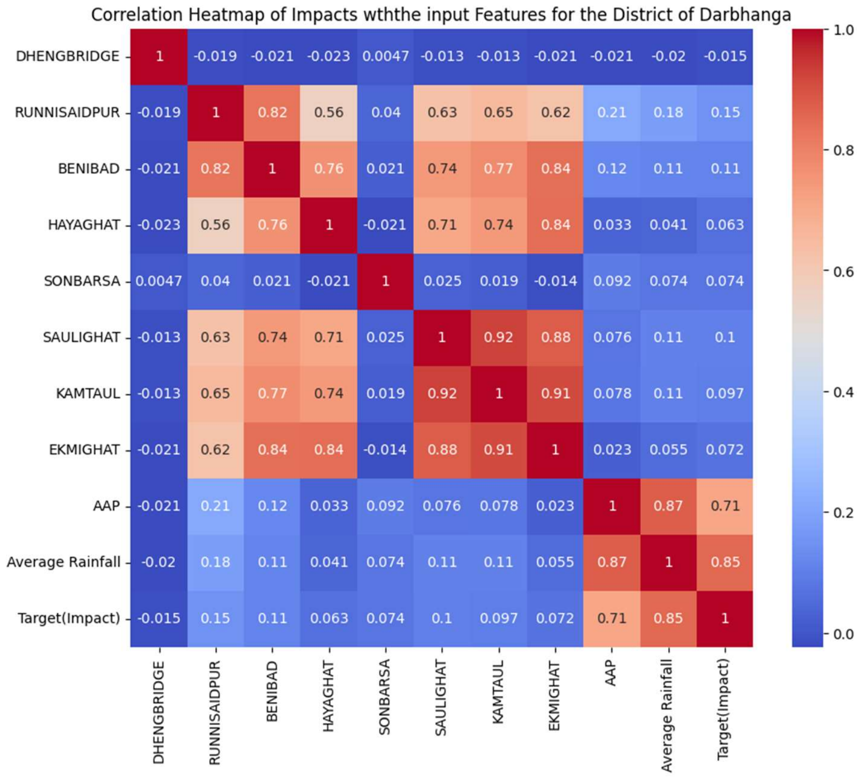

Fig. 8: The correlation matrix of the impacts of heavy rainfall with its causative feature in the districts of Darbhanga, Bihar, India.

and the diverse factors associated with floods in these areas.

West Champaran

The districts of West Champaran hold the distinction of being the largest in Bihar in terms of their geographical expanse. It is noteworthy that flash floods frequently affect this area. The primary cause of these floods may be attributed to the substantial precipitation in the Soemshwar Doon region. According to (Jha & Gundimeda 2019), the district experiences floods very frequently, with a frequency of 0.8 to 1. The districts exhibit a north-to-south slope, whereas the canal runs in an east-to-west direction. This configuration leads to congestion in the canal's capacity at peak events. Nevertheless, due to the topography, the floodwaters dissipated swiftly. Fig. 7 displays the correlation matrix between the input attributes and the effects. The correlation coefficients of 0.8 for rainfall and 0.75 for AAP clearly show a strong correlation between heavy rainfall and its effects. This relationship is further supported by the data collected from the river gauges located at Triveni and Khadda, which are situated close to the border with Nepal.

Darbhanga

According to (Jha & Gundimeda 2019), the Darbhanga districts are classified as having a high frequency of floods, with a range of 0.6 to 0.8. The river group Bagmati/Adhawra is responsible for the floods in these two districts. Blocks such as Kusheshwar Asthan, Hayafghat, Jale, and others are susceptible to flooding. The duration of the flood is characterized as mild since the water recedes within a timeframe of around one to two weeks. The study conducted by (Kumar et al. 2016) examines the progress made in the regions affected by the significant flooding incidents. Fig. 8: displays the correlation matrix, illustrating the relationships between the impacts of the districts and the input attributes. The data indicates a strong correlation between severe rainfall impacts and the average rainfall (0.85) in the districts, as well as the average annual precipitation (AAP) (0.75) in the adjacent catchments, in conjunction with the river gauges at Benibad.

Vaishali

Flood-prone areas are often located in the downstream regions of the Ganga River, and the Vaishali district is categorized

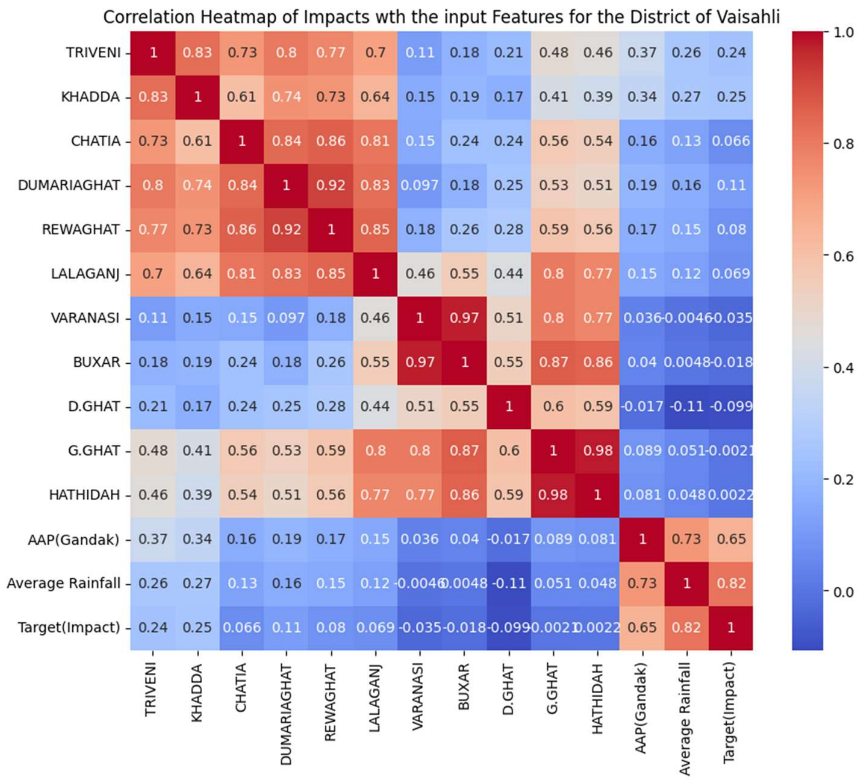

Fig. 9: The correlation matrix of the impacts of heavy rainfall with its causative feature in the districts of Vaishali, Bihar, India.

as one such area where flooding is mostly attributed to the combined influence of two river systems, namely the Ganga and the Gandak. Typically, amid the monsoon season, the lower regions known as Diyara have regular occurrences of flooding. According to (Jha & Gundimeda 2019), the district is within the range of medium flood frequency, namely between 0.4 and 0.6. The drainage process of floods in this district is often characterized by a prolonged duration. Fig. 9: displays the correlation matrix, illustrating the relationships between the districts' impacts and the input attributes. The data indicates a strong correlation between severe rainfall impacts and the average rainfall (0.82) in the districts, as well as the average annual precipitation (AAP) (0.65) in the connected catchments, in addition to the river gauges at Khadda (0.25) and Triveni (0.24).

Assessment of the Proposed M/L Multi-classification Models

The performance indicators outlined in paragraph 3.3 were employed to assess the effectiveness of the proposed machine learning multi-classification models. The models were evaluated for four consecutive days: day 0, day 1, day 2, and day 3. The inputs used for each evaluation were based on the characteristics of the corresponding day or the preceding days. The grid search methodology is employed to refine and optimize the selection of hyperparameters. Table 4 displays the ideal hyperparameters for the Decision Tree (DT), K-Nearest Neighbors (KNN), and Support Vector Machine (SVM) models that were trained in the districts of West Champaran, Darbhanga, and Vaishali. These districts are representative of the northern region of Bihar. The change of hyperparameters is deemed unnecessary for the Naïve Bayes multi-classification approach. The observed variation in the quantity of hyper parameterized parameters throughout the districts of northern Bihar can be ascribed to disparities in the number of input features and the heterogeneous attributes of floods.

The performance metrics from subsection 3.3 were used to rate the trained models. The results can be seen in Table 5 (Decision Tree and Support Vector Machine models) and Table 6 (KNN and Naïve Bays models).

To build an efficient early warning system and a thorough vulnerability assessment that considers the diverse causes of floods (Zhang et al. 2022), This methodology employs an innovative machine learning technique for multi-classification, which effectively combines diverse sources

Table 4: Best Hyperparameter of the Decision Tree, SVM, and KNN Multi classification algorithms for the proposed dynamic impact-based multi-classification algorithms for the districts of West Chamapran, Darbhanga, and Vaishali.

Algorithm/ Best Parameter	Decision Tree			SVM					KNN		
	max_ depth	min_ samples_leaf	min_samples_ split	C	coef0	degree	gamma	kernel	n_ neighbors	p	weights
West Champaran	3	2	2	0.1	-1	1	0.01	poly	3	1	distance
Darbhanga	3	1	5	10	-1	5	0.0001	poly	9	2	uniform
Vaishali	10	1	2	10	1	5	scale	poly	7	2	distance

Table 5: Presentation of performance metrics for the day (0–3) for DT and SVM.

Decision Tree Multi Classifier												
District(Days)/ Performance Parameter	West Champaran				Darbhanga				Vaishali			
	Day0	Day1	Day2	Day3	Day0	Day1	Day2	Day3	Day0	Day1	Day2	Day3
Balanced Accuracy	0.7287	0.7287	0.7287	0.7287	0.8068	0.8068	0.8068	0.7840	0.7174	0.6787	0.6816	0.6758
Binary Cross Entropy	2.1128	2.1128	2.1128	2.1128	0.0688	0.0688	0.0688	0.0741	3.656	3.9177	3.6566	4.1789
Macro F1 Score	0.71	0.71	0.71	0.71	0.80	0.80	0.80	0.77	0.58	0.56	0.57	0.54
Micro Weighted F1 Score	0.85	0.85	0.85	0.85	0.96	0.96	0.96	0.96	0.90	0.89	0.89	0.88
Support Vector Machine												
Balanced Accuracy	0.5729	0.5729	0.5729	0.5729	0.6717	0.6717	0.6717	0.6717	0.7412	0.7412	0.7412	0.5729
Binary Cross Entropy	0.3973	0.3782	0.3880	0.3847	0.5910	0.6063	0.6049	0.6067	0.2364	0.2522	0.2516	0.3847
Macro F1 Score	0.61	0.61	0.61	0.61	0.64	0.64	0.64	0.64	0.60	0.60	0.60	0.61
Micro Weighted F1 Score	0.74	0.74	0.74	0.74	0.94	0.94	0.94	0.94	0.91	0.91	0.91	0.74

Table 6: Presentation of performance metrics for the day (0–3) for KNN and Naïve Bays.

KNN												
District(Days)/ Performance Parameter	West Champaran				Darbhanga				Vaishali			
	Day0	Day1	Day2	Day3	Day0	Day1	Day2	Day3	Day0	Day1	Day2	Day3
Balanced Accuracy	0.6875	0.6875	0.6875	0.6875	0.6489	0.6489	0.6489	0.6489	0.8662	0.8662	0.8662	0.8662
Binary Cross Entropy	2.9199	2.9199	2.9199	2.9199	0.1380	0.1380	0.1380	0.1380	0.1662	0.1662	0.1662	0.1662
Macro F1 Score	0.65	0.65	0.65	0.65	0.63	0.63	0.63	0.63	0.87	0.87	0.87	0.87
Micro Weighted F1 Score	0.85	0.88	0.88	0.88	0.93	0.93	0.93	0.93	0.94	0.94	0.94	0.94
Naïve Bays Classifier												
Balanced Accuracy	0.8849	0.8849	0.8849	0.8849	0.7030	0.7030	0.7030	0.7030	0.5478	0.5478	0.5478	0.5478
Binary Cross Entropy	0.4672	0.4672	0.4672	0.4672	0.7065	0.7065	0.7065	0.7065	0.334	0.334	0.334	0.334
Macro F1 Score	0.84	0.84	0.84	0.84	0.60	0.60	0.60	0.60	0.55	0.55	0.55	0.55
Micro Weighted F1 Score	0.90	0.90	0.90	0.90	0.84	0.84	0.84	0.84	0.84	0.84	0.84	0.84

of both heterogeneous and homogeneous information. The implementation of this improvement resulted in the enhancement of the current early warning system while also enabling the retention of a greater quantity of information during the fusion process. The individual who bears the responsibility for decision-making holds the capacity to evaluate the sources and extent of consequences, which might serve as important considerations for those tasked with disaster management. The Naïve Bayes algorithms exhibit noteworthy performance in the districts of West Champaran, achieving a balanced accuracy of 0.88, which closely approaches the ideal value of 1. Furthermore, it can be observed that the decision tree multi-classification algorithms exhibit a high level of effectiveness, ranking as the second most efficient models. One significant advantage of utilizing balanced accuracy is its fair assessment of all classes, which is especially important when dealing with imbalanced datasets that have significant differences in sample numbers between classes. Furthermore, the macro F1 score demonstrates a significant degree of efficacy when employed in conjunction with the naive Bayes multi-classifier. The ability to punish the model severely when it shows excessive confidence in an incorrect class is one of the benefits of using binary cross entropy. As a result, this contributes to improving the accuracy of the model. In the context of the Naive Bayes classifier, when the value tends towards 1, it signifies a desirable level of performance. In the districts of West Champaran, it is evident that the algorithms operate in the following sequence: The performance of Naive Bayes is seen to be the greatest, followed by Decision Tree, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM). In the context of the Darbhanga districts, the performance of multiclassification algorithms may be rated as follows: The performance of the Decision Tree surpasses that of the Support Vector Machine (SVM), which in turn surpasses that of K-Nearest Neighbors (KNN), while Naive Bayes exhibits the least effective performance. In the Vaishali areas, where the possible ramifications of floods may be ascribed to the combined impacts of the Gandak and Ganga river systems, the algorithms can be hierarchically ordered based on their efficacy as follows: The K-Nearest Neighbors (KNN) algorithm has the highest level of effectiveness, followed by Support Vector Machines (SVM), Decision Trees (DT), and Naive Bayes (NB). Consequently, a thorough assessment was undertaken to evaluate the performance of several machine learning algorithms for multi-classification, to examine their effectiveness concerning the proposed methodologies. The comprehensive assessment of available sources is crucial to effectively enable the actual application of models in the field. The present study investigates the possible use of big data technologies in the evaluation of flood risk, as stated in the work of (Monrat et al. 2019). Furthermore, this study examines the practical use of graded impact-based warning systems in the context of severe rainfall events. The study also examines the issues that require attention and the tactics that must be implemented to effectively harness the capabilities of these technologies. The performance metrics outlined in paragraph 3.3 were employed to assess the effectiveness of the proposed machine learning multiclassification models. The models underwent evaluation for four consecutive days: day 0, day 1, day 2, and day 3. The evaluation process involved using input characteristics from the same day for day 0, while for subsequent days (day 1, day 2, and day 3), inputs from prior days were used. The grid search method is employed to refine the optimal hyperparameterization. Table 4 displays the best hyperparameters for the Decision Tree (DT), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) models that were trained using data from the districts of West Champaran, Darbhanga, and Vaishali. There is no need for hyperparameter changes in the Naïve Bayes multiclassification approach. The observed heterogeneity in the quantity of hyperparametrized parameters throughout the districts of northern Bihar may be ascribed to the disparities in the number of input features and the distinct attributes of floods in these regions.

Different multiclassification methods behave in different ways, especially when they are given datasets that aren't balanced and have a big difference in the number of cases in each class. The aforementioned disparity has the potential to result in models that exhibit bias and demonstrate inferior performance when it comes to underrepresented classes. To solve these challenges, a range of strategies are utilized. In this article, the Synthetic Minority Oversampling Technique (SMOTE) is employed to handle unbalanced datasets by producing synthetic samples for the minority class. The generation of synthetic examples is achieved by the process of interpolating between pre-existing cases, effectively equalizing the distribution of classes. Also, stratified cross-validation is employed to guarantee that every fold inside the cross-validation procedure maintains an equitable distribution of all classes. This methodology facilitates the assessment of the model's efficacy by ensuring that each class is equally represented. In our forthcoming stages, we intend to augment the efficacy of our models by capitalizing on the benefits of sophisticated methodologies, such as Easy-Ensemble or Balanced Random Forest. The efficacy of the presented models is contingent upon the caliber of previous flood impact data and the fundamental causal elements. The success of multiclassification models relies heavily on the presence of high-quality data. The incorporation of continuous input from real-time datasets is crucial to maintaining the real-time correctness of these models. Moreover, the precision and dependability of historical data play a crucial role in efficiently training the models. The outcomes derived from these models offer significant insights into the quantification of the economic implications associated with varying degrees of warning accuracy. This evaluation encompasses the estimation of possible cost savings that may arise as a consequence of less damage and improved procedures for disaster preparedness.

CONCLUSION

The primary aim of this research is to improve our understanding of flood risk categorization and the impact of individual or combined causative elements through the development of innovative ML approaches. The study sites were chosen from places located in northern Bihar, India, which are susceptible to frequent occurrences of flooding. The geography of the North Bihar region and its adjacent Tarai territories is characterized by a significant incline. Consequently, the development of an impact-based flood warning system in these regions necessitates consideration of the combined impacts of in-situ and runoff rainfall, as well as the overflow of rivers and associated canal systems. A novel machine learning approach has been devised to improve comprehension and communication regarding the diverse degrees of impacts of flood consequences arising from rainfall, average aerial precipitation in nearby catchment areas, and rises in river gauges due to rainfall in neighboring regions of Nepal. To accomplish this goal, we utilized a modeling methodology to evaluate the quantification of floods resulting from several causative factors. The utilization of multi-classification ML classifiers in tandem enhances the decision-making process by capitalizing on the collective proficiency of the machine learning algorithms. The tools that have been suggested aim to improve decision-making in the context of operational impact-based forecasting. Furthermore, this approach maximizes the likelihood of both over- and under-warnings. The ML models under consideration are evaluated in terms of their performance relative to the base models, to identify the optimal model that exhibits robustness in the presence of variations. The main objective of these proposed tools is to assist operational forecasters by generating a classification of flood impacts as a result. Hence, the executive line agency operates to the specific requirements of the situation, allocating resources as necessary to minimize the effects of flooding. This paper presents a conceptual framework for evaluating the dynamic consequences of intense precipitation, incorporating several additional contributing components. The suggested framework has the potential to be implemented in many geographical areas to improve flood management tactics. The presented model will be employed in further investigations to examine several facets of compound runoff and rainfall-induced flooding. While this study acknowledges the important influence of other contributing factors, such as land use and land cover, on impact-based flooding occurrences within the relevant period, it does not quantify their effects. Future research endeavors should prioritize the exploration of the integration of these components. However, the potential for intensified rainfall as a result of climate change might significantly amplify the likelihood of flooding in the area under study. Future evaluations employing this model aim to quantitatively evaluate the influence of altering climatic conditions on the susceptibility to flooding.

ACKNOWLEDGMENTS

The authors are grateful to the Ministry of Earth Sciences, Govt. of India for implementing the Impact heavy rainfall warning system in the country.

ABBREVIATIONS

This manuscript employs the following abbreviations:

AI/ML	Artificial Intelligence/Machine Learning
SVM	Support Vector Machine
DT	Decision Tree
KNN	k-Nearest Neighbour
NB	Naïve Bays
LU/LC	Land Use/Land Cover

REFERENCES

Abbas, S., Waseem, M., Yaseen, M., Latif, Y., Leta, M.K., Khan, T.H. and Muhammad, S., 2023. Spatial-temporal seasonal variability of extreme precipitation under warming climate in Pakistan. *Atmosphere (Basel)*, 14, 210. https://doi.org/10.3390/atmos14020210.

Beevers, L., Douven, W., Lazuardi, H. and Verheij, H., 2012. Cumulative impacts of road developments in floodplains. *Transportation Research Part D: Transport and Environment*, 17, pp.398–404. https://doi.org/10.1016/j.trd.2012.02.005.

Brunetti, M., Buffoni, L., Mangianti, F., Maugeri, M. and Nanni, T., 2004. Temperature, precipitation and extreme events during the last century in Italy. *Global and Planetary Change*, 40, pp.141–149. https://doi.org/10.1016/S0921-8181(03)00104-8.

De Luis, M., González-Hidalgo, J.C., Brunetti, M. and Longares, L.A., 2011. Precipitation concentration changes in Spain 1946-2005. *Natural Hazards and Earth System Sciences*, 11, pp.1259–1265. https://doi.org/10.5194/nhess-11-1259-2011.

Easterling, D.R., Evans, J.L., Groisman, P.Y., Karl, T.R., Kunkel, K.E. and Ambenje, P., 2000. Observed variability and trends in extreme climate events: a brief review. *Bulletin of the American Meteorological Society*, 81(3), pp.417-426.

Goswami, B.N., Venugopal, V., Sangupta, D., Madhusoodanan, M.S. and Xavier, P.K., 2006. Increasing trend of extreme rain events over India in a warming environment. *Science*, 314, pp.1442–1445. https://doi.org/10.1126/science.1132027.

Goyal, D., Haritash, A.K. and Singh, S.K., 2021. A comprehensive review of groundwater vulnerability assessment using index-based, modelling and coupling methods. *Journal of Environmental Management*, 296, 113161. https://doi.org/10.1016/j.jenvman.2021.113161.

Grandini, M., Bagli, E. and Visani, G., 2020. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*.

Groisman, P.Y., Knight, R.W., Easterling, D.R., Karl, T.R., Hegerl, G.C. and Razuvaev, V.N., 2005. Trends in intense precipitation in the climate record. *Journal of Climate*, 18, pp.1326–1350. https://doi.org/10.1175/JCLI3339.1.

IMD, Ministry of Earth Sciences, 2021. Standard Operation Procedure - Weather Forecasting and Warning Services.

Jha, R.K. and Gundimeda, H., 2019. An integrated assessment of vulnerability to floods using composite index – A district level analysis for Bihar, India. *International Journal of Disaster Risk Reduction*, 35, 101074. https://doi.org/10.1016/j.ijdrr.2019.101074.

Kumar, A., Sarthi, P.P., Kumari, A. and Sinha, A.K., 2021. Observed characteristics of rainfall indices and outgoing longwave radiation over the Gangetic Plain of India. *Pure and Applied Geophysics*, 178, pp.619–631. https://doi.org/10.1007/s00024-021-02666-6.

Kumar, V., Yin, C.C.S., Kumar Singh, A., Yin, C.S. and Kumar, A., 2016. Impact of flood on rural population and strategies for mitigation: A case study of Darbhanga district, Bihar state, India. *Contemporary Rural Social Work Journal*, 8.

Meehl, G.A. and Tebaldi, C., 2004. More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, 305, pp.994–998.

Moghadas, M., Asadzadeh, A., Vafeidis, A., Fekete, A. and Kötter, T., 2019. A multi-criteria approach for assessing urban flood resilience in Tehran, Iran. *International Journal of Disaster Risk Reduction*, 35. https://doi.org/10.1016/j.ijdrr.2019.101069.

Mohandes, M., Deriche, M. and Aliyu, S.O., 2018. Classifiers Combination Techniques: A Comprehensive Review. *IEEE Access*, 6, pp.19626–19639. https://doi.org/10.1109/ACCESS.2018.2813079.

Mohapatra, M., Kumar, N., Mishra, K. and Devi, S., 2021. Evaluation of heavy rainfall warnings of India National Weather Forecasting Service for monsoon season (2002–2018). *Journal of Earth System Science*, 130. https://doi.org/10.1007/s12040-020-01549-z.

Monrat, A.A., Ul Islam, R., Hossain, M.S. and Andersson, K., 2019. A belief rule based flood risk assessment expert system using real time sensor data streaming. *Proceedings of the 43rd Annual IEEE Conference on Local Computer Networks, LCN Workshops 2018*, pp.38–45. https://doi.org/10.1109/LCNW.2018.8628607.

Najibi, N. and Devineni, N., 2018. Recent trends in the frequency and duration of global floods. *Earth System Dynamics*, 9, pp.757–783. https://doi.org/10.5194/esd-9-757-2018.

Niyogi, D., Lei, M., Kishtawal, C., Schmid, P. and Shepherd, M., 2017. Urbanization impacts on the summer heavy rainfall climatology over the eastern United States. *Earth Interactions*, 21, pp.1–17. https://doi.org/10.1175/EI-D-15-0045.1.

Pandit, C., 2009. Some common fallacies about floods and flood management. *Current Science*, 97, pp.991–993.

Pielke, R.A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., Goldewijk, K.K., Nair, U., Betts, R., Fall, S., Reichstein, M., Kabat, P. and de Noblet, N., 2011. Land use/land cover changes and climate: Modeling analysis and observational evidence. *Wiley Interdisciplinary Reviews: Climate Change*, 2, pp.828–850. https://doi.org/10.1002/wcc.144.

Prasad, K., Afroz, R., Sarker, M.A. and Rahman, M., 2021. A diagnostic study of some flood producing rainfall events in Bangladesh with a limited area analysis-forecast system. *Mausam*, 57, pp.475–488. https://doi.org/10.54302/mausam.v57i3.492.

Ranalkar, M.R., Chaudhari, H.S., Hazra, A., Sawaisarje, G.K. and Pokhrel, S., 2016. Dynamical features of incessant heavy rainfall event of June 2013 over Uttarakhand, India. *Natural Hazards*, 80, pp.1579–1601. https://doi.org/10.1007/s11069-015-2040-z.

Ray, S., 2019. Introduction to machine learning and different types of machine learning algorithms. *Proceedings of the International Conference on Machine Learning, Big Data, Cloud and Parallel Computing: Trends, Perspectives and Prospects*, pp.35–39.

Reddy, D.V., Kumar, D., Saha, D. and Mandal, M.K., 2008. The 18 August 2008 Kosi river breach: an evaluation. *Current Science*, 95, pp.1668–1669.

Shankar, A., Kumar, A., Sahana, B.C. and Sinha, V., 2022. A Case study of heavy rainfall events and resultant flooding during the summer monsoon season 2020 over the river catchments of north Bihar, India. *VayuMandal*, 48(2), pp.17–28.

Sinha, R., Bapalu, G.V., Singh, L.K. and Rath, B., 2008. Flood risk analysis in the Kosi river basin, north Bihar using multi-parametric approach of analytical hierarchy process (AHP). *Journal of the Indian Society of Remote Sensing*, 36, pp.335–349. https://doi.org/10.1007/s12524-008-0034-y.

Xu, C., Gong, A., Liang, L., Song, X. and Wang, Y., 2023. Vulnerability assessment method for immovable cultural relics based on artificial neural networks—an example of a heavy rainfall event in Henan Province. *International Journal of Disaster Risk Science*, 14, pp.41–51. https://doi.org/10.1007/s13753-022-00461-y.

Zhang, Z.X., Wang, L., Duan, J.L. and Wang, Y.M., 2022. An early warning method based on fuzzy evidential reasoning considering heterogeneous information. *International Journal of Disaster Risk Reduction*, 82, 103356. https://doi.org/10.1016/j.ijdrr.2022.103356.

Zou, X. and Ren, F., 2015. Changes in regional heavy rainfall events in China during 1961–2012. *Advances in Atmospheric Sciences*, 32, pp.704–714. https://doi.org/10.1007/s00376-014-4127-y.

ORCID DETAILS OF THE AUTHORS

Anand Shankar: https://orcid.org/0000-0001-8141-6400