Predictive Analysis of Flood Risk Factors Based on a Machine Learning Approach: Comparative Study of SVM and XGBoost Algorithms

Surya Darma[1], Ahmad Jihad Al Fayed[2], Surya Maruli P Pardede[3], Muhammad Hizbul Aqsha[4], Muhammad Syahputra Novelan[5]
[1]Department of Film and Television, Faculty of Arts and Design, University of Potensi Utama, Medan, Indonesia
[2,3,4,5]Department of Information Technology, Faculty of Graduate School, University of Pembangunan Panca Budi, Medan, Indonesia
[1]suryadarma766@gmail.com, [2]jihadahmad000@gmail.com, [3]sp.trainer.belajar.id@gmail.com, [4]hizbul2000@gmail.com, [5]putranovelan@dosen.pancabudi.ac.id

Article Info

Article history:

Received January 08, 2026
Revised February 02, 2026
Accepted February 09, 2026

Keywords:

Flood risk
Machine learning
Prediction
Support vector machine
Xgboost

ABSTRACT

Flood events in Indonesia continue to increase in frequency and impact due to high rainfall variability, land-use change, and complex hydrological conditions. Accurate predictive modeling is therefore essential to support flood risk assessment and mitigation planning. This study evaluates the predictive performance of two supervised machine learning algorithms, Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost), for flood risk classification. The analysis is conducted using a publicly available dataset comprising 500 samples that represent multiple environmental and spatial factors related to flood occurrence. Data preprocessing includes cleaning, normalization, and feature consistency adjustment prior to model implementation. Both algorithms are trained and tested using the same dataset configuration to ensure objective comparison. Model performance is assessed using accuracy, precision, recall, and F1-score metrics. Experimental results indicate that XGBoost achieves higher accuracy and precision, demonstrating stronger capability in reducing false-positive predictions, while SVM shows relatively higher recall, reflecting better sensitivity in identifying flood-prone cases. Overall, XGBoost provides more reliable predictive performance for flood risk modeling on the dataset used. The findings confirm the effectiveness of machine learning-based approaches for flood risk prediction and highlight the importance of algorithm selection in disaster risk analysis.

Corresponding Author:

Surya Darma
Universitas Potensi Utama
Email: suryadarma766@gmail.com

1. INTRODUCTION

Flooding is one of the most frequent natural disasters in Indonesia and has a significant impact on people's lives, urban and rural infrastructure, and environmental sustainability. Geographically, Indonesia has diverse topography, high tropical rainfall, and many areas located in large river basins, which increases its vulnerability to flooding.[1],[2],[3] Floods caused by high rainfall can result in significant loss of life and damage to infrastructure, especially in flood-prone areas. [4],[5],[6] In developing countries, the lack of real-time water level monitoring and manual operation of floodgates increases the risk of channel overflow and flash floods. [7]

This situation is exacerbated by climate change, which triggers an increase in the intensity of extreme rainfall, as well as the uncontrolled pace of urbanization and land use change, particularly in urban areas and river buffer zones. As a result, flooding in Indonesia has not only increased in frequency, but has also spread from urban areas to peri-urban and rural areas [8], [9], [10]. Therefore, accurate flood risk modeling and analysis is urgently needed as a basis for sustainable disaster risk mitigation and management planning in Indonesia. [11], [12], [13].

In recent years, the use of spatial data, remote sensing imagery, and hydrometeorological data available in Indonesia, such as rainfall data, slope, land use, soil type, and distance from rivers, has opened up great opportunities for the application of machine learning (ML) approaches in flood risk modeling [14], [15]. This approach is considered capable of accommodating the complexity of Indonesia's non-linear and heterogeneous environmental conditions. [16]

Various machine learning algorithms, including Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost), have been used to identify and analyze the main factors that influence flood vulnerability. SVM is widely used due to its ability to stably model non-linear relationships between variables, while XGBoost is known for its computational efficiency and high accuracy, especially in processing large data sets commonly found in spatial-based disaster studies. [17], [18]

A number of previous studies have shown that machine learning models are effective in flood risk assessment, including in regions with tropical climates and complex geographical conditions such as Indonesia. Research by Nguyen et al. proves that machine learning algorithms can be used in multi-hazard assessment with good predictive performance, which is relevant to the context of Indonesia as a country with a high level of disaster risk. [19]

In addition, a study by Janizadeh et al. shows that machine learning-based flash flood models produce high AUC values, indicating a reliable level of accuracy in mapping flood-prone areas, especially in areas with high rainfall intensity and rapid hydrological response [20] Furthermore, research by Shirmohammadi et al. emphasizes the importance of selecting and combining environmental parameters to improve the accuracy of machine learning models for flood vulnerability assessment. These findings are particularly relevant to Indonesia, which has a wide variety of land uses and watershed characteristics. [21]

Research by Yaseen et al., which applied an ensemble machine learning approach, also showed that combined models were able to provide better prediction results than single models, including SVM and XGBoost, in mapping flood vulnerability in certain study areas [22]. This indicates that comparative evaluation of various ML algorithms is an important step before their application to flood cases in Indonesia. Therefore, this study focuses on an in-depth exploration of the SVM and XGBoost algorithms to understand their performance and suitability in predictive analysis of flood risk in Indonesia. [23]

Both algorithms have different characteristics and advantages. SVM is known to be effective in handling high-dimensional datasets and limited sample sizes, conditions that are often encountered in disaster studies in Indonesia due to the limited availability of consistent historical data [24], [25]. On the other hand, XGBoost excels at overcoming overfitting issues and is capable of optimizing model performance through boosting and regularization techniques, making it highly suitable for complex and large-scale data analysis, such as national or regional spatial data. [22]

2. METHOD

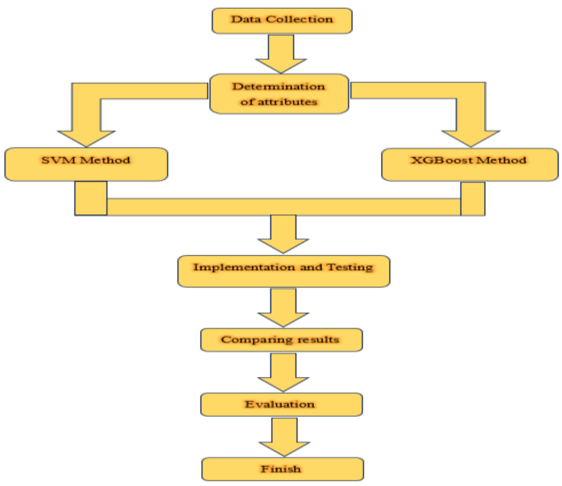

Figure 1. Research Methods and Stages

In developing this research method, the process was carried out through several systematic stages aimed at building a reliable predictive model for analyzing flood risk factors. These stages include :

2.1 Data Collection

The data used in this study was sourced from a public dataset available on the Kaggle platform. The dataset used consisted of 500 data samples representing various risk factors for flooding. The variables contained in the dataset included relevant environmental and geographical parameters, such as rainfall, regional elevation, slope gradient, distance to rivers, and a number of other supporting attributes that contributed to the analysis of flooding events.

Before modeling, data undergoes pre-processing, which includes data cleaning, duplicate data removal, and missing value handling. In addition, feature selection is performed to eliminate irrelevant attributes or those with low data quality, with the aim of improving data structure consistency and model performance. This step also aims to ensure consistency between the training data and test data used in the study.

In the modeling stage, two machine learning algorithms, namely Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost), were applied to perform predictive analysis of flood risk. Both models were trained and evaluated using the same dataset to obtain an objective performance comparison. Model evaluation was conducted using relevant performance metrics to assess each algorithm's predictive ability in accurately classifying flood risk.

The use of public datasets in this study did not involve personal data or direct intervention with specific subjects. Therefore, this study did not raise any significant ethical implications and was conducted in accordance with applicable scientific research ethics principles.

Table 1. Dataset

1	curah_hujan	elevasi	lereng	jarak_sungai	tata_guna_lahan	kepadatan_tinggi	penyerapan_tanah	label_banjir
2	718	1101	2386	1499	1	3007	25	1
3	448	3416	2782	2541	2	3018	81	0
4	415	381	704	4419	2	3039	91	0
5	415	4256	1198	165	3	2897	69	0
6	1395	2476	457	2630	2	3095	92	1
7	860	2403	2977	415	3	3006	31	1
8	339	2962	2781	3906	2	2979	56	0
9	741	4123	162	3419	1	3027	65	1
10	600	1739	2526	1557	1	2993	64	1
11	65	3390	1563	2477	2	2958	5	0
12	201	2829	1871	755	2	2989	45	0
13	632	1335	267	1274	1	3024	3	1
14	1310	4393	2266	3448	1	2893	49	1
15	418	3987	383	1230	3	2966	32	0
491	315	2811	2901	1138	1	2984	21	1
492	787	4788	1253	4815	2	2993	78	0
493	1371	877	2952	947	2	3106	55	1
494	1568	3450	2004	6653	2	2926	94	1
495	298	1005	1904	187	2	2891	56	0
496	364	2679	498	2323	1	3001	94	0
497	619	483	2646	1362	3	2992	77	1
498	641	2252	1282	981	3	2984	39	1
499	435	3781	487	2003	3	3076	6	0
500	250	1738	38	2434	3	2982	94	0
501	487	3325	1679	9	3	2991	84	0

2.2. Feature Selection

During the data pre-processing stage, adjustments and feature selection were carried out to ensure that the dataset used met the requirements for predictive flood risk modeling. The research dataset consisted of a number of variable attributes representing behavioral conditions, activities, and environmental pressures that could potentially contribute to flood risk. The attributes used in the model training process included : *Sleep Quality, Academic Performance, Study Load, Headache Frequency, Extracurricular Activity, Social Support Level, Screen Time Hours, Physical Activity Min, Commute Time Min, Financial Pressure, Academic Workload, Part Time Job, dan Stress Level.* All of these features are retained because they have complete data availability and a consistent structure across all data used.

Next, the data undergoes a cleaning process that includes removing invalid values, handling outliers, and normalizing numerical values to improve the stability and performance of the machine learning algorithm. The processed dataset is then used in the model training and testing stages using the Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) algorithms. This approach aims to ensure that the comparison of the two algorithms' performance is conducted objectively and based on data that has undergone uniform pre-processing stages.

The description of each variable attribute and value range is explained in Table 2 below.

Table 2. Variable Features in Datasets

No	Features	Description
1	Curah Hujan	Intensitas hujan di wilayah pengamatan (mm/bulan)
2	Elevasi	Ketinggian wilayah dari permukaan laut (meter (mdpl))
3	Lereng	Kemiringan permukaan tanah (derajat (°))
4	Jarak Sungai	Jarak lokasi ke sungai terdekat (meter)
5	Tata Guna Lahan	Jenis penggunaan lahan (1 = Permukiman, 2 = Sawah / pertanian, 3 = Hutan)
6	Kepadatan Tinggi	Kepadatan bangunan atau penduduk (jiwa/km²)
7	Penyerapan Tanah	Kemampuan tanah menyerap air (indeks (0–1))
8	Label Banjir	Status kejadian banjir (target/label, 0 = Tidak banjir, 1 = Banjir)

2.3. Modeling

The next stage in this research is the modeling process, which aims to determine the most suitable machine learning algorithm for performing predictive analysis of flood risk factors. At this stage, the model is understood as a mathematical representation implemented in the form of a computational algorithm to study patterns from data and generate predictions based on the given input characteristics.

In line with the research focus, the algorithms used in the modeling process are Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost). These two algorithms were chosen because they have different characteristics in handling nonlinear data and good generalization capabilities. The modeling process was carried out by training each algorithm using a dataset that had undergone pre-processing, then evaluating the performance of both to obtain an objective comparison in predicting flood risk.

a. Support Vector Machine (SVM)

Support Vector Machine (SVM) is an algorithm *supervised learning* which works on the principle of optimal hyperplane search to separate data with maximum margin. This algorithm was first introduced [25], and is known to perform well on small to medium-sized datasets and high-dimensional data.

In the context of flood prediction, SVM has been widely used as a classification and regression model. One of the main advantages of SVM is the use of kernel functions (kernel trick) that enable the mapping of non-linear data to higher dimensions so that it can be separated linearly. Some commonly used kernels include *linear kernel*, *polynomial kernel*, *radial basis function (RBF)*, dan *sigmoid kernel*. Schölkopf dan Smola (2002) explains that kernel selection greatly affects the performance of SVM models. The RBF kernel, for example, is often used because of its ability to handle non-linear relationships well [26]. However, inappropriate kernel selection can reduce model performance and increase computational complexity.

In its implementation, SVM is divided into Support Vector Classification (SVC) and Support Vector Regression (SVR). SVC is used to separate data into specific classes, while SVR aims to predict continuous values with a certain error tolerance (ε-insensitive loss). Smola and Schölkopf (2004) state that SVR has advantages in modeling non-linear relationships in time series data and environmental data. Therefore, SVM is widely used in the field of forecasting, including hydrology, weather, and disasters. [27]

Dibike et al. (2001) was one of the early researchers to apply SVM to flood forecasting and found that SVM provided more stable results than artificial neural networks (ANN) in limited data conditions [28]. Research conducted by Shafizadeh-Moghadam et al. (2018) shows that SVM is capable of producing flood vulnerability maps with a high degree of accuracy when combined with topographical and hydrological factors [29]. However, the performance of SVM is highly dependent on kernel and parameter selection, which, if not optimized, can cause overfitting [30]. The mathematical representation is as follows :

$$\text{Si } Yi = +1;\ wxi + b \geq 1 \quad (1)$$
$$\text{Si } Yi = -1;\ wxi + b \leq 1 \quad (2)$$
$$\text{i; } Yi(wxi + b) \geq 1 \quad (3)$$

Equations (1) and (2) show the condition of data separation by a hyperplane defined by the equation $wxi + b = 0$, where :

- w is a (weight vector),
- xi is the feature vector of the i-th data point,
- b is a biased value.

Data labeled +1 is expected to be on one side of the hyperplane with a certain minimum distance, while data labeled –1 is on the opposite side. The ±1 threshold is used to determine the margin, which is the distance between the hyperplane and the nearest data point of each class.
Equation (3) is a combination of equations (1) and (2) that simplifies the SVM classification requirements. This equation states that each data point must be classified correctly and lie outside or exactly on the margin boundary. If the value of Y_i (wx_i+b) is greater than or equal to 1, then the data is classified correctly by the model.

b. Extreme Gradient Boosting (XGBoost)

XGBoost is an extension of the gradient boosting algorithm introduced by Chen and Guestrin (2016). This algorithm uses an ensemble learning approach by gradually combining multiple decision trees and is equipped with a regularization mechanism to avoid overfitting. XGBoost uses an objective function consisting of two main components, namely training loss and regularization term. This regularization includes penalties for model complexity, such as the number of leaves and tree weights, so that the resulting model is simpler and has better generalization capabilities [31]. In addition, XGBoost implements shrinkage (learning rate), column subsampling, and row sampling techniques that effectively reduce model variance. The boosting approach optimized with regularization has proven to be superior in modeling complex non-linear relationships. [32]

XGBoost can be used for various supervised learning problems, including classification and regression. This algorithm supports various loss functions, such as logistic loss for classification and squared error for regression. This flexibility makes XGBoost widely used in various research domains. State that tree-based ensemble methods such as gradient boosting have a high ability to capture interactions between variables without requiring explicit feature transformations. Therefore, XGBoost often produces better performance than linear algorithms or single tree models. [33]

In the field of disaster management, particularly flood risk analysis, XGBoost has been widely used due to its ability to process complex spatial and environmental data. Wang et al. (2020) showed that XGBoost produced higher accuracy and Area Under Curve (AUC) compared to SVM and Random Forest in flood vulnerability mapping. Another study by Zhao et al. (2021) also reported that XGBoost is able to model the interaction between rainfall, elevation, slope, and land use factors more effectively, resulting in more accurate flood risk maps. These results confirm the superiority of XGBoost in the context of hydrometeorological disaster prediction and mitigation. [34]

The application of XGBoost combined with data balancing techniques, such as Synthetic Minority Over-sampling Technique (SMOTE), has been reported to improve classification performance in datasets with class imbalance. This approach is relevant in the context of flood research, given that flood events generally have a smaller proportion than non-flood conditions in the dataset.

Neural Network-based methods were not applied in this study due to data limitations and relatively low interpretability. Model interpretability is an important aspect in flood risk analysis, especially to support understanding and decision-making by relevant stakeholders.
All models used, namely Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost), were run using the default hyperparameter settings from the scikit-learn and XGBoost libraries. This approach was applied to maintain objectivity in comparing the performance of the two algorithms and to minimize the risk of overfitting, given the limited amount of data used in the study.

2.4. Implementation and Testing

During the implementation and testing stages, the research dataset was processed using the Python programming language with the support of the Google Colaboratory platform as the computing environment. This process included data loading, pre-processing, and dividing the dataset into training and testing data to ensure that model evaluation was conducted objectively.

Next, the Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) algorithms were applied to perform predictive analysis of flood risk factors. The performance of each model was evaluated using several evaluation metrics, namely accuracy, precision, recall, F1-score, and Area Under the Curve (AUC). The use of these various metrics aims to provide a more comprehensive picture of the model's ability to perform classification, particularly in dealing with class imbalances between flood and non-flood events.

The test results from each evaluation metric were then compared to assess the relative advantages between the SVM and XGBoost algorithms. This comparison was used as the basis for determining the algorithm with the most optimal predictive performance in modeling flood risk factors based on the dataset used in the study.

2.5. Comparison and Interpretation of Results

At this stage, an evaluation and comparison of performance between the Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost) algorithms was conducted to predict flood risk factors. The comparison process begins with the application of each algorithm to a dataset that has undergone pre-processing, including data cleaning, handling extreme values, and dividing the dataset into a training set and a testing set. The algorithms are implemented using the Python programming language through the Google Colaboratory platform, which provides an efficient computing environment and supports the use of the scikit-learn and XGBoost libraries.

The performance of each algorithm is evaluated using several metrics, including :

1. **Accuracy** to measure the proportion of correct predictions from the entire sample.
2. **Precision**, to assess the model's ability to generate correct positive predictions, particularly for the flood event class.
3. **Recall**, to assess the model's ability to detect all positive samples completely.
4. **F1-score**, as a measure of the balance between precision and recall, providing an overview of the balance between the two.
5. **Area Under the Curve (AUC)**, to assess the model's ability to distinguish between flood and non-flood classes at various prediction thresholds.

The use of these various metrics is important because flood datasets typically have an imbalanced class distribution, where flood events are rarer than non-flood conditions. Therefore, relying solely on accuracy can lead to misleading conclusions. By considering precision, recall, F1-score, and AUC, performance evaluation becomes more comprehensive and reflects the algorithm's ability to handle data imbalance.

The evaluation results from all metrics are then systematically compared to determine the algorithm with the most optimal predictive performance. The superior algorithm will be the primary recommendation for the implementation of a machine learning-based flood risk prediction system, while also providing a basis for decision-making in risk mitigation and policy planning related to flood disasters.

3. RESULTS AND DISCUSSION

This study compares the performance of Support Vector Machine (SVM) and XGBoost in predicting flood risk using Python. The dataset used was obtained from Kaggle, consisting of 500 entries with 8 relevant attributes, and divided into 20% training data and 80% test data. The research process includes data cleaning, feature normalization, model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The results of this study are expected to show a more effective classification method for predicting floods, while also providing an overview of the application of machine learning for disaster risk mitigation.

3.1. Algorithm Testing *Support Vector Machine* (SVM)

Testing of the SVM model on the flood prediction dataset was conducted by dividing 80% of the data for testing and 20% for training using Python in Google Colab. The evaluation results show that this model performs quite well, with an accuracy of 89%, precision of 90%, recall of 88%, and an F1-Score of 89%. These values indicate that the SVM model is capable of predicting flood events with a high degree of accuracy and a balance between the ability to recognize positive and negative cases. The results of these calculations are visualized in the following figure, which provides an illustration of the overall performance of the model.

```
==== SVM ====
Accuracy  : 0.8900
Precision : 0.8980
Recall    : 0.8800
F1-Score  : 0.8889

Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.90      0.89        50
           1       0.90      0.88      0.89        50

    accuracy                           0.89       100
   macro avg       0.89      0.89      0.89       100
weighted avg       0.89      0.89      0.89       100
```

Figure 2. Algorithm Testing Results Support Vector Machine (SVM)

3.2. Algorithm Testing *Extreme Gradient Boosting* (XGBoost)

Testing of the XGBoost model on the flood prediction dataset was conducted by dividing 80% of the data for testing and 20% for training using Python in Google Colab. The evaluation results show that this model performs very well, with an accuracy of 90%, precision of 95%, recall of 84%, and an F1-Score of 89%. The high precision value indicates that the model is able to accurately predict flood events, although the recall is slightly lower, meaning that some positive cases are not detected. A visualization of the model's performance can be seen in the following figure, which shows XGBoost's ability to capture flood data patterns comprehensively.

```
==== XGBoost ====
Accuracy  : 0.9000
Precision : 0.9545
Recall    : 0.8400
F1-Score  : 0.8936

Classification Report:
              precision    recall  f1-score   support

           0       0.86      0.96      0.91        50
           1       0.95      0.84      0.89        50

    accuracy                           0.90       100
   macro avg       0.91      0.90      0.90       100
weighted avg       0.91      0.90      0.90       100
```

Figure 3. Algorithm Testing Results Extreme Gradient Boosting (XGBoost)

Based on the test results, a comparison of the performance of SVM and XGBoost was conducted using accuracy, precision, recall, and F1-Score metrics to determine the most suitable algorithm for the flood prediction dataset. The results show that XGBoost has higher precision, while SVM excels slightly in recall, while the accuracy and F1-Score of both algorithms are relatively comparable. A summary of this comparison is presented in Table 3, which provides an overview of the effectiveness of each method in predicting flood events.

Table 3. Model Performance Comparison Table

	Model	Accuracy	Precision	Recall	F1-Score
1	SVM	0.89	0.897959	0.88	0.888889
2	XGBoost	0.90	0.954545	0.84	0.893617

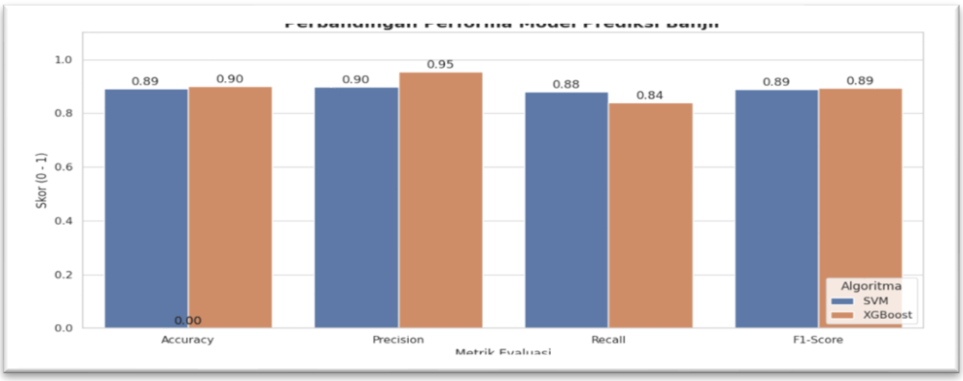

Figure 4. Comparison of Flood Prediction Model Performance

Based on Table 3 and Figure 4, it can be seen that the SVM model achieved an accuracy of 89%, precision of 90%, recall of 88%, and an F1-Score of 89%. These values indicate that SVM is capable of predicting flood events with a high degree of accuracy and a good balance between the ability to recognize positive and negative cases. Meanwhile, the XGBoost model shows an accuracy of 90%, precision of 95%, recall of 84%, and an F1-Score of 89%. These results indicate that XGBoost is superior in terms of precision, making it more capable of minimizing false positive prediction errors, even though its recall value is slightly lower than SVM, which means that some flood cases are not detected. From this comparison, it can be concluded that both algorithms perform well, but XGBoost tends to be more effective in providing accurate predictions, while SVM is more balanced in detecting all flood events in the dataset used.

4. CONCLUSION

Based on the research results, it can be concluded that the application of the XGBoost method on the flood prediction dataset provides the best performance on 20% of the training data, with an accuracy of 90%, precision of 95%, recall of 84%, and an F1-Score of 89%. Meanwhile, the application of the SVM method on the same training data resulted in an accuracy of 89%, precision of 90%, recall of 88%, and an F1-Score of 89%. From a comparison of the performance of the two methods, it can be seen that XGBoost shows higher effectiveness and accuracy in predicting flood risk factors than SVM, especially in terms of precision and overall accuracy. This indicates that XGBoost is superior in producing accurate and reliable predictions for the dataset used, although SVM still shows balanced performance in terms of recall and F1-Score. Thus, XGBoost can be considered a more optimal method to apply in flood risk prediction.

ACKNOWLEDGEMENTS

The author expresses sincere appreciation and gratitude for the various forms of support and guidance received throughout the research process. Special acknowledgment is given to **Assoc. Prof. Dr. Muhammad Syahputra Novelan, S.Kom., M.Kom.** for his valuable supervision, constructive input, and meaningful direction. The author also extends thanks to colleagues and all parties who contributed and provided support during the course of this study.

REFERENCES

[1] S. Choosumrong, K. Piyathamrongchai, R. Hataitara, U. Soteyome, N. Konkong, R. Chalongsuppunyooet al., "Development of an IoT-Based Flood Monitoring System Integrated with GIS for Lowland Agricultural Areas",, 2025. https://doi.org/10.20944/preprints202506.1403.v1

[2] M. Damayanti, S. Wahyuningtyas, D. Balistyadhana, A. Hapsari, A. Shammah, N. Yusaet al., "The IoT-Based Early Warning System for Detecting High Tide Floods (ROB-EWS) in Tambak Lorok, Semarang Indonesia", Journal of Geography Environment and Earth Science International, vol. 28, no. 10, p. 13-24, 2024. https://doi.org/10.9734/jgeesi/2024/v28i10822

[3] H. Yuliandoko, S. Subono, V. Wardhani, S. Pramono, & P. Suwindarto, "Design of Flood Warning System Based IoT and Water Characteristics", Telkomnika (Telecommunication Computing Electronics and Control), vol. 16, no. 5, p. 2101, 2018. https://doi.org/10.12928/telkomnika.v16i5.7636

[4] M. Farabi and A. Sintawati, "Flood early warning system at Jakarta dam using internet of things (IoT)-based real-time fishbone method to support industrial revolution 4.0", Journal of Soft Computing Exploration, vol. 5, no. 2, p. 99-106, 2024. https://doi.org/10.52465/joscex.v5i2.293

[5] S. Choosumrong, K. Piyathamrongchai, R. Hataitara, U. Soteyome, N. Konkong, R. Chalongsuppunyooet al., "Development of an IoT-Based Flood Monitoring System Integrated with GIS for Lowland Agricultural Areas",, 2025. https://doi.org/10.20944/preprints202506.1403.v1

[6] D. Perera, J. Agnihotri, O. Seidou, & R. Djalante, "Identifying societal challenges in flood early warning systems", International Journal of Disaster Risk Reduction, vol. 51, p. 101794, 2020. https://doi.org/10.1016/j.ijdrr.2020.101794

[7] F. Razak, M. S. Novelan, dan Arpan, "Perancangan sistem pendeteksi banjir dan sistem peringatan dini berbasis bot Telegram," *Ensiklopedia Education Review*, vol. 4, no. 1, Apr. 2022. [Online]. Tersedia: http://jurnal.ensiklopediaku.org

[8] X. Wang and H. Xie, "A Review on Applications of Remote Sensing and Geographic Information Systems (GIS) in Water Resources and Flood Risk Management", Water, vol. 10, no. 5, p. 608, 2018. https://doi.org/10.3390/w10050608

[9] N. Rehman, U. Zada, & K. Haleem, "Quantifying the Influences of Land Use and Rainfall Dynamics on Probable Flood Hazard Zoning", Nust Journal of Engineering Sciences, vol. 16, no. 1, p. 54-68, 2023. https://doi.org/10.24949/njes.v16i1.739

[10] M. Afifi, R. Wahaab, A. Khalifa, I. Moukhtar, & E. Elalfy, "Assessment of Anticipatory Approach Using the Integration of GIS, and Remote Sensing Techniques for Flood Management in Alexandria City, Egypt", Civil Engineering and Architecture, vol. 11, no. 3, p. 1439-1453, 2023. https://doi.org/10.13189/cea.2023.110326

[11] A. Elaji and W. Ji, "Urban Runoff Simulation: How Do Land Use/Cover Change Patterning and Geospatial Data Quality Impact Model Outcome?", Water, vol. 12, no. 10, p. 2715, 2020. https://doi.org/10.3390/w12102715

[12] M. Jayantari, A. Yesaya, I. P., & G. Eryani, "Flood Hazzard Assessmentand Mapping in Yeh Ho Watershed Using Geographic Information System (GIS)", The International Journal of Engineering and Science, vol. 13, no. 11, p. 75-83, 2024. https://doi.org/10.9790/1813-13117583

[13] H. Rizeei, B. Pradhan, & M. Saharkhiz, "An integrated fluvial and flash pluvial model using 2D high-resolution sub-grid and particle swarm optimization-based random forest approaches in GIS", Complex & Intelligent Systems, vol. 5, no. 3, p. 283-302, 2018. https://doi.org/10.1007/s40747-018-0078-8

[14] I. Safitri, Y. Ruhiat, & A. Saefullah, "Analysis of Flood Vulnerability Levels Using Overlay Method With System-Based Scoring Geographical Information (Case Study: District Tangerang)", Newton-Maxwell Journal of Physics, vol. 4, no. 2, p. 56-64, 2023. https://doi.org/10.33369/nmj.v4i2.30354

[15] W. Widyatmanti, A. Puti, G. Priatmadja, H. Raihan, S. Sholihah, R. Hanifet al., "Flood Hazard Mapping in Cimandiri Watershed using Fuzzy Logic Method Based on Multi-Source Geospatial Datasets", Iop Conference Series Earth and Environmental Science, vol. 1406, no. 1, p. 012019, 2024. https://doi.org/10.1088/1755-1315/1406/1/012019

[16] A. Chinedu, N. Ezebube, S. Uchegbu, & V. Ozorme, "Integrated assessment of flood susceptibility and exposure rate in the lower Niger Basin, Onitsha, Southeastern Nigeria", Frontiers in Earth Science, vol. 12, 2024. https://doi.org/10.3389/feart.2024.1394256

[17] C. Meng and H. Jin, "A Comparison of Machine Learning Models for Predicting Flood Susceptibility Based on the Enhanced NHAND Method", Sustainability, vol. 15, no. 20, p. 14928, 2023. https://doi.org/10.3390/su152014928

[18] Z. Li, J. Tian, Y. Zhu, D. Chen, Q. Ji, & D. Sun, "A Study on Flood Susceptibility Mapping in the Poyang Lake Basin Based on Machine Learning Model Comparison and SHapley Additive exPlanations Interpretation", Water, vol. 17, no. 20, p. 2955, 2025. https://doi.org/10.3390/w17202955

[19] H. D. Nguyen, D. K. Dang, Q. T. Bui, and A. I. Petrisor, "Multi-hazard assessment using machine learning and remote sensing in the North Central region of Vietnam," *Transactions in GIS*, vol. 27, no. 5, pp. 1614–1640, 2023, doi: 10.1111/tgis.13091.

[20] S. Janizadeh, M. Avand, A. Jaafari, T. V. Phong, M. Bayat, E. Ahmadisharaf, I. Prakash, B. T. Pham, and S. Lee, "Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran," *Sustainability*, vol. 11, no. 19, p. 5426, Sep. 2019, doi: 10.3390/su11195426.

[21] M. Shirmohammadi, S. Pirasteh, W. Li, and D. Mafi-Gholami, "Flood risk mapping and performance efficiency evaluation of machine learning algorithms: Best practice in northern Iran," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-G, pp. 1347–1352, 2025, doi: 10.5194/isprs-archives-XLVIII-G-2025-1347-2025.

[22] A. Yaseen, J. Lu, and X. Chen, "Flood susceptibility mapping in an arid region of Pakistan through ensemble machine learning model," *Stoch. Environ. Res. Risk Assess.*, vol. 36, pp. 3041–3061, 2022, doi: 10.1007/s00477-022-02179-1.

[23] S. T. Seydi, Y. Kanani-Sadat, M. Hasanlou, R. Sahraei, J. Chanussot, and M. Amani, "Comparison of machine learning algorithms for flood susceptibility mapping," *Remote Sensing*, vol. 15, no. 1, p. 192, Dec. 2022, doi: 10.3390/rs15010192.

[24] A. Salvati, A. M. Nia, A. Salajegheh, K. Ghaderi, D. T. Asl, N. Al-Ansari, F. Solaimani, and J. J. Clague, "Flood susceptibility mapping using support vector regression and hyper-parameter optimization," *J. Flood Risk Management*, vol. 16, no. 4, p. e12920, 2023, doi:10.1111/jfr3.12920.

[25] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/BF00994018.

[26] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.
[27] A. J. Smola and B. Schölkopf, “A tutorial on support vector regression,” *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: 10.1023/B:STCO.0000035301.49549.88.
[28] Y. B. Dibike, S. Velickov, D. P. Solomatine, and M. B. Abbott, “Model induction with support vector machines: Introduction and applications,” *Journal of Computing in Civil Engineering*, vol. 15, no. 3, pp. 208–216, 2001, doi: 10.1061/(ASCE)0887-3801(2001)15:3(208).
[29] H. Shafizadeh-Moghadam, A. Valavi, M. Shahabi, T. Chapi, and B. Pradhan, “Novel forecasting approaches using machine learning and statistical models for flood susceptibility mapping,” *Journal of Hydrology*, vol. 556, pp. 191–206, 2018, doi: 10.1016/j.jhydrol.2017.11.022.
[30] M. S. Tehrany, B. Pradhan, and M. N. Jebur, “Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS,” *Journal of Hydrology*, vol. 512, pp. 332–343, 2014, doi: 10.1016/j.jhydrol.2014.02.033.
[31] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[32] A. Natekin and A. Knoll, “Gradient boosting machines, a tutorial,” *Frontiers in Neurorobotics*, vol. 7, p. 21, 2013, doi: 10.3389/fnbot.2013.00021.
[33] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009.
[34] H. Ren, B. Pang, P. Bai, G. Zhao, S. Liu, Y. Liu, and M. Li, “Flood susceptibility assessment with random sampling strategy in ensemble learning: Random Forest and XGBoost,” *Remote Sens.*, vol. 16, no. 2, p. 320, Jan. 2024, doi: 10.3390/rs16020320.