

# Enhancing Flood Prediction in Urban Areas: A Machine Learning Approach for Makassar City

**H. Muh Rizal**

Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Indonesia  
7022211021@student.its.ac.id

**Mochamad Hariadi**

Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia |  
Department of Computer Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia  
mochar@ee.its.ac.id (corresponding author)

**Yunifa Miftachul Arif**

Department of Informatics Engineering, Universitas Islam Negeri Maulana Malik Ibrahim, Indonesia  
yunif4@ti.uin-malang.ac.id

**Elly Warni**

Department of Informatics Engineering, Universitas Hasanuddin, Indonesia  
elly@unhas.ac.id

*Received: 25 October 2024 | Revised: 04 December 2024 and 5 January 2025 | Accepted: 24 January 2025*

*Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.9402>*

## ABSTRACT

**Accurate and rapid predictions regarding urban flooding, are essential in supporting risk mitigation efforts. Flood phenomena have the potential to cause extensive damage and disrupt the functions of economic and governmental sectors. However, these impacts can be minimized through comprehensive planning and preparation to reduce potential losses. Machine learning techniques have emerged as a promising method for predicting complex hydrological processes. This research develops a flood prediction model by comparing seven machine learning algorithms, namely Logistic Regression, Linear Discriminant Analysis, k-Nearest Neighbors, Gaussian Naive Bayes, Support Vector Machine, AdaBoost, and Random Forest. The results show that Random Forest has the highest performance, demonstrating the reliability of Random Forest in processing complex urban flood datasets. This model is expected to enhance disaster preparation and contribute significantly to flood risk management in urban areas.**

*Keywords-flood prediction; early warning system; machine learning; random forest; Makassar City*

## I. INTRODUCTION

The escalating impact of climate change and rapid urbanization has dramatically increased the frequency and severity of urban flooding, making accurate flood prediction increasingly crucial. Urban regions are particularly vulnerable, experiencing significant challenges from increased flood occurrences due to intensified rainfall patterns and inadequate drainage infrastructure [1, 2]. These floods can cause substantial damage to infrastructure, disrupt economic activities, and severely impact the daily lives of the residents, highlighting the need for improved prediction methods. Previous research has demonstrated various approaches to flood prediction, with Machine Learning (ML) emerging as a promising solution. Recent studies have shown that ML algorithms can effectively process historical and real-time data to generate accurate flood predictions [3-5]. Specifically,

Random Forest (RF) and Support Vector Machine (SVM) have demonstrated superior capabilities in predicting flood risk with high precision, particularly when integrating big data for early warning systems [6, 7]. The application of ML in flood prediction has evolved significantly, with several algorithms showing promising results in different contexts. Logistic Regression (LR) has proven effective in identifying vulnerable areas through historical data analysis [8, 9] though it struggles with non-linear relationships. RF has shown superior performance in handling complex, unstructured datasets through its ensemble approach [10, 11], while SVM has demonstrated strength in managing unbalanced data and noise reduction [12, 13]. Additional algorithms such as AdaBoost and k-Nearest Neighbors (kNN) have shown varying degrees of success, with RF generally maintaining superior accuracy and stability [14, 15]. Despite these advances, significant challenges remain in flood prediction. Most ML models require

extensive historical data for accurate predictions, which are often unavailable in many regions [16, 17]. Furthermore, existing models face limitations in generalizability, particularly in urban areas with diverse geographical characteristics [18]. The integration of real-time data and physical hydrological models remains underdeveloped, creating a significant gap in current flood prediction capabilities.

Makassar City, located on Sulawesi's southwest coast, presents a compelling case study for urban flood prediction. The city faces complex flooding challenges due to a combination of natural factors, including high rainfall, lowland topography, and tidal influences, as well as human activities such as rapid urbanization, reduced water catchment areas, and inadequate drainage. These challenges are further compounded by infrastructure limitations, particularly in real-time data collection and monitoring systems [19]. The unique combination of these factors makes Makassar an ideal location for developing and testing advanced flood prediction methods.

ML offers various classification algorithms, each with distinct advantages for different scenarios. In the realm of binary classification, LR stands out as a fundamental statistical approach that evaluates the probability of binary outcomes using predictor variables. While this method performs optimally when predictors exhibit low correlation, researchers can address multicollinearity challenges through ridge regression techniques [20-22]. Linear Discriminant Analysis (LDA) presents an effective solution for scenarios involving normally distributed classes with uniform covariance matrices. Its utility extends beyond classification, as it excels in reducing dimensionality before undertaking classification tasks [23, 24]. When considering non-parametric approaches, kNN emerges as a straightforward yet powerful method that determines classification based on neighboring data points. While particularly effective in applications like text classification and fraud detection, its success heavily relies on appropriate parameter selection, including the number of neighbors and distance metrics [25-27]. In handling high-dimensional datasets, Gaussian Naive Bayes (GNB) provides computational efficiency through its probabilistic approach based on Bayes' theorem. Though it assumes feature independence, which may not always hold true in real-world applications, it remains valuable for many classification tasks [28, 29]. SVM distinguishes itself through its ability to identify optimal separating hyperplanes in feature space, particularly excelling in high-dimensional environments with clear class separation [30, 31]. The field of ensemble methods introduces powerful alternatives like AdaBoost, which enhances classification accuracy by iteratively combining weak classifiers while focusing on misclassified instances [32]. Similarly, RF leverages multiple decision trees to generate robust predictions. This method has demonstrated superior performance compared to LR [33, 34]. Selecting the most appropriate classification algorithm requires careful consideration of the specific problem context, including data characteristics, computational resources, and performance requirements. Each method offers unique strengths that make it particularly suitable for certain applications while potentially less optimal for others.

This study aims to enhance flood prediction accuracy in urban environments through the development and implementation of advanced ML algorithms, integration of historical and real-time data from BMKG and BBWS Pompengan Jeneberang, and comparative analysis of seven ML algorithms to identify the most effective approach. The novelty of this research lies in creating an integrated framework that combines multiple data sources with sophisticated ML algorithms, specifically tailored for urban flood prediction.

The research makes significant contributions to flood mitigation through three key aspects: First, it develops an urban-specific flood prediction model using ML algorithms, with a particular focus on Makassar City's unique characteristics. Second, it provides a comprehensive comparative analysis of seven ML approaches for flood prediction, including LR, LDA, kNN, GNB, SVM, AdaBoost, and RF. Finally, it creates an adaptive framework that can be applied to similar urban environments, thereby enhancing disaster preparation and mitigation efforts. The outcomes of this research will significantly enhance flood preparation and mitigation efforts in urban areas, potentially reducing the impact of flooding on infrastructure and communities.

## II. METHODOLOGY

### A. Study Area and Dataset

Makassar City, located at 5.1616° S and 119.4359° E, is on the southwest coast of Sulawesi Island and directly facing the Makassar Strait. The datasets used for flood prediction testing were obtained from two main sources, namely the Balai Besar Wilayah Sungai (BBWS) and the Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) [39]. BBWS Pompengan Jeneberang provides information related to river conditions and behavior, including water level, water discharge, and tides. Meanwhile, BMKG provides meteorological data on rainfall intensity, temperature, and humidity. This dataset covers the year range from January 1, 2018 to December 31, 2022. This period provides data for five full years, which includes seasonal and annual variations relevant to flood predictions in Makassar City. This wide time range ensures sufficient datasets to train and test machine learning models with high accuracy. The dataset consisting of 1,826 records was then divided into two parts, 80% for training and 20% for testing. The division ensures that most of the data are used to train the model to learn patterns effectively, while the rest are used to evaluate the accuracy of the model's predictions. To overcome the class imbalance especially in the context of the classification of flood potential, the SMOTE (Synthetic Minority Over-sampling Technique) technique is used. SMOTE works by synthesizing new data on minority classes, thereby reducing the risk of analysis errors due to unbalanced data distribution [35]. Through the application of this method, the performance and reliability of the flood prediction model in Makassar City can be significantly improved. Table I lists the parameters considered in this study.

### B. Data Preprocessing

Data Preprocessing on the dataset of 1826 rows and 6 columns begins with the data cleaning and transformation stages. This dataset includes temporal variables in the form of

date, meteorological variables consisting of rainfall, temperature, and humidity, as well as hydrological variables including water level height, water discharge, and tides. The initial stage of preprocessing focuses on cleaning the data from missing or invalid values, especially in the Float Precipitation column. the water discharge column, which was originally a string type, was converted to numeric format and cleaned of error characters. in addition, outlier values or anomalies were checked in other numerical columns such as water level height, tide, temperature, and humidity.

Figure 1 presents a correlation matrix in the form of a heatmap, illustrating the relationships among the variables Rainfall, Water Level, Water discharge, Tides, Temperature, and Humidity. In this heatmap, red indicates a strong positive correlation, while blue indicates a strong negative correlation. The closer the color is to white, the weaker or closer to zero the correlation becomes. The value of 1.00 on the main diagonal signifies a perfect correlation of each variable with itself. From

the results, Water Level and Water discharge show a high positive correlation (0.90), suggesting a strong link between them. Meanwhile, Temperature and Humidity exhibit a notably negative correlation (-0.66), indicating that as temperature increases, humidity tends to decrease, and vice versa.

TABLE I. PARAMETER LIST

Parameter	Unit	Description
Rainfall	mm	Measurement of rainfall intensity using a rain gauge
Water level	m	Height of the water surface measured from a reference point
Water discharge	m <sup>3</sup> /s	Volume of water flowing per unit time through a given cross-section
Tides	m	Height of sea water level influenced by tidal phenomena
Temperature	°C	Ambient air temperature at the time of observation
Humidity	%	Relative humidity of the air at the time of observation

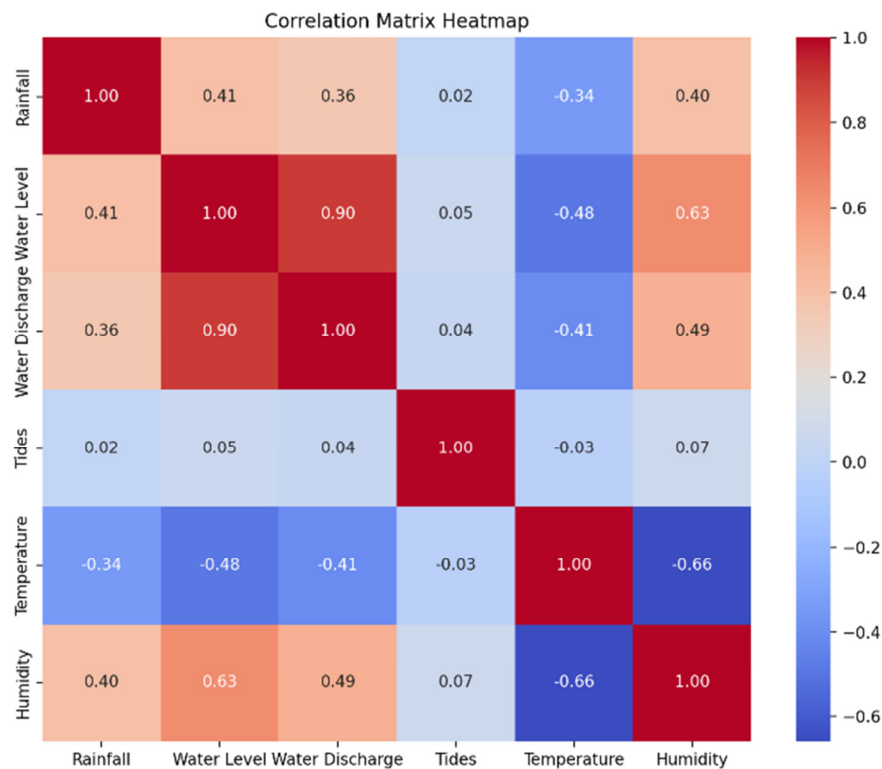


Fig. 1. Correlation matrix of the dataset.

C. Model Selection and Evaluation

This section evaluates several ML models to identify the one that achieves the highest accuracy in flood prediction, including Logistic Regression (LR), Linear Discriminant Analysis (LDA), k-Nearest Neighbors (kNN), Gaussian Naive Bayes (NB), Support Vector Machine (SVM), AdaBoost, and Random Forest (RF). Figure 2 shows the proposed Flood Predictive System (FPS), emphasizing the systematic approach employed for accurate flood prediction.

In this study, the RF classifier model was used as the main classification algorithm, optimized using the GridSearchCV method with 5-fold cross-validation. The RF model training process began with initializing the model parameters. The number of decision trees was set to 100. Each decision tree was restricted to a maximum depth of 4 and the maximum number of features was set to 2. The minimum number of samples required to split a node was set to 15. Additionally, the random state was set to 42 to control the randomization aspects of the algorithm, ensuring reproducible results every time the code is run, which is crucial for repeatable experiments.

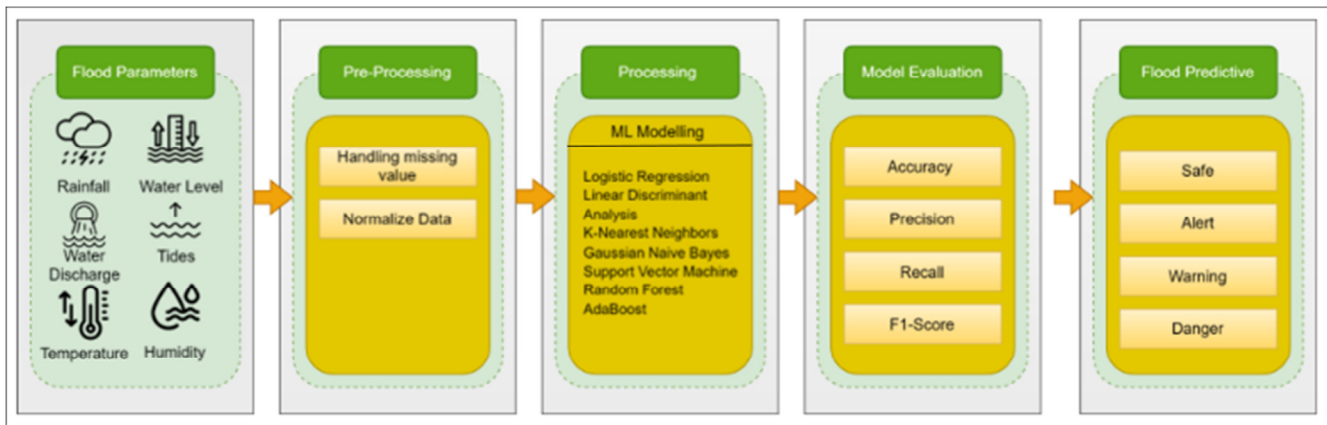


Fig. 2. Proposed flood predictive system model.

Classifier performance can be evaluated using multiple metrics. Choosing the right performance metrics is essential, as they must align with the particular needs of the application in question. Classifiers can outperform in one evaluation criterion but struggle in another. A key goal of model evaluation is to get the most out of these performance metrics [2-5]. The formulas for calculating accuracy, precision, recall, and F1 Score using a confusion matrix is given by:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100 \quad (1)$$

$$\text{Precision} = \frac{TP}{(TP+FN)} \times 100 \quad (2)$$

$$\text{Recall} = \frac{TN}{(TN+FP)} \times 100 \quad (3)$$

$$\text{F1 Score} = 2 \times \frac{P \times R}{P+R} \times 100 \quad (4)$$

where  $TP$  represents True Positives,  $TN$  represents True Negatives,  $FP$  represents False Positives, and  $FN$  represents False Negatives.

### III. RESULTS AND DISCUSSION

This study evaluated the performance of six algorithms (LR, LDA, kNN, NB, SVM, AdaBoost, and RF), by assessing their prediction accuracy. The model tuning process in this study was carried out carefully by applying cross-validation using the k-fold method to avoid overfitting and ensure the stability of performance metrics on various data partitions. The k-fold cross-validation, with a k-value set at 5, divided the dataset into five subsets (folds), where each fold was used as one-time test data, while the rest became training data. This strategy helps ensure that the model can generalize well on new data, especially when working with limited datasets, such as rainfall, water discharge, and humidity data in the Makassar region. The results show that the RF algorithm predicts Makassar City floods better than the other ones. RF can anticipate floods, improving disaster preparedness and management.

As shown in Table I and regarding the testing data comprised of 20% of the initial dataset, the RF algorithm demonstrates the best performance in flood prediction, achieving an accuracy of 99.18% and an F1 score of 94.20%.

Among all the algorithms tested, there is noticeable variation in evaluation metrics, particularly between accuracy and precision. The produced confusion matrices can be seen in Figure 3.

TABLE II. PERFORMANCE COMPARISON OF ML ALGORITHMS IN FLOOD PREDICTION

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
LR	97.27	85.40	91.60	86.80
LDA	91.26	78.40	85.40	78.50
KNN	96.72	81.10	82.80	81.60
NB	94.81	81.00	91.90	85.80
SVM	98.91	95.30	95.50	95.30
AdaBoost	89.62	55.20	75.00	58.60
RF	99.18	92.70	96.20	94.20

Figure 4 illustrates a comparison of the performance metrics Accuracy, Precision, Recall, and F1-Score for the considered classification algorithms. It can be concluded that the RF model surpassed the other considered models. Classification errors occur relatively rarely, indicating that the model is very reliable in identifying different categories of flood risk. These results confirm the effectiveness of RF as an accurate and reliable flood prediction model.

As shown in Figure 5, the RF algorithm demonstrates strong performance in predicting water levels, a key indicator of flooding. The graph illustrates that RF's predicted values closely align with the actual values, with minimal differences between the two. This minimal deviation highlights the algorithm's robust predictive capabilities in flood prediction scenarios. The RF's accuracy of 99.18% (Table I) further reinforces these findings, indicating that its predictions closely reflect real-world field conditions.

RF excels in flood prediction due to its ability to handle complex datasets and integrate multiple predictor variables effectively. This ML algorithm demonstrates robustness across various geospatial data sources, such as remote sensing imagery and Digital Elevation Models (DEMs), which are crucial for accurate flood mapping and risk assessment [6]. RF's capacity to identify significant predictors enhances its predictive accuracy, as evidenced by studies showing precision rates between 83% and 94% in fluvial hazard predictions [36].

Additionally, RF models exhibit lower overfitting tendencies compared to other algorithms, ensuring reliable performance even with diverse training parameters [37]. The integration of RF with other statistical methods, such as binary statistics,

further improves its predictive capabilities, capturing a high percentage of historical disaster points [38]. Overall, RF's versatility, accuracy, and robustness make it a preferred choice for flood prediction in various contexts.

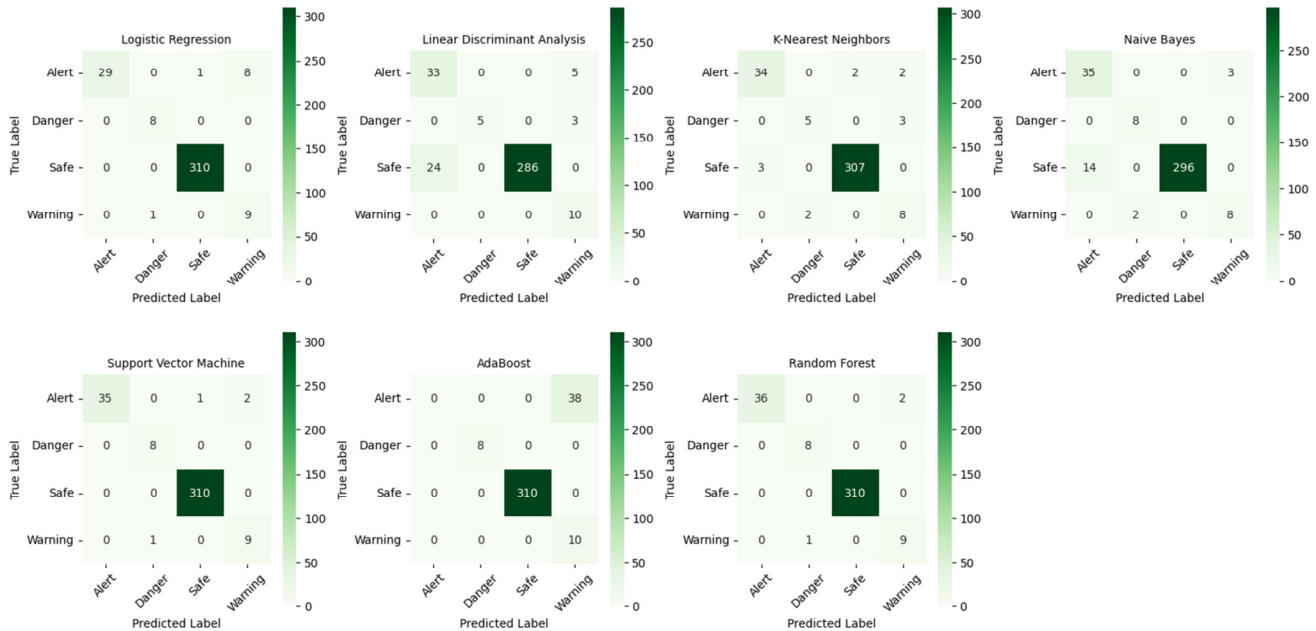


Fig. 3. Confusion matrices of the considered ML algorithms.

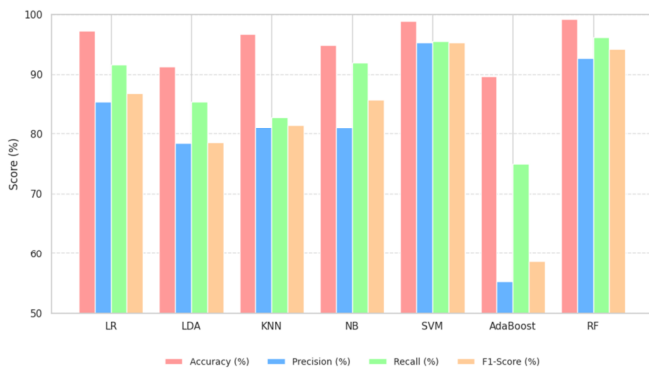


Fig. 4. Comparison of ML algorithm performance.

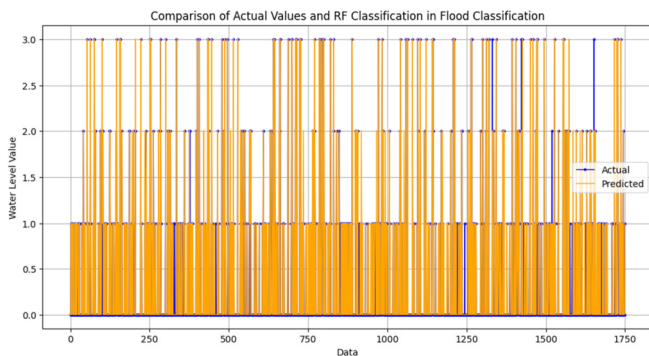


Fig. 5. Comparison of actual values and RF classification in flood prediction.

#### IV. CONCLUSION

Our research investigated the potential of Machine Learning (ML) approaches for improving urban flood forecasting, specifically examining their application in Makassar City. The investigation encompassed a thorough evaluation of seven distinct ML methods: Logistic Regression (LR), Linear Discriminant Analysis (LDA), k-Nearest Neighbors (kNN), Gaussian Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), and AdaBoost. The analysis revealed that the RF algorithm demonstrated superior performance compared to other methods, achieving exceptional metrics, highlighting its capability to effectively process complex flood-related data in urban environments.

A key factor in achieving these results was the successful incorporation of the datasets from BMKG and BBWS Pompengan Jeneberang, which significantly enhanced the prediction system's accuracy and response time. The research findings indicate that this approach not only improves prediction accuracy but also provides a foundation for developing effective flood early warning systems in urban settings.

The study's outcomes emphasize how ML techniques, particularly the RF algorithm, can effectively address urban flood prediction challenges. This methodological framework shows promise for adaptation to various urban contexts, potentially contributing to improved disaster preparedness globally. Future research directions could explore additional variables, such as urban development patterns and

infrastructure modifications, to further enhance prediction models and strengthen cities' flood resilience capabilities.

#### ACKNOWLEDGMENT

This research was carried out in two laboratories: Institut Teknologi Sepuluh Nopember Surabaya's visicom engineering laboratory and the Universitas Teknologi Akba Makassar's multimedia laboratory.

#### REFERENCES

- [1] T. Panfilova *et al.*, "Flood Susceptibility Assessment in Urban Areas via Deep Neural Network Approach," *Sustainability*, vol. 16, no. 17, Jan. 2024, Art. no. 7489, <https://doi.org/10.3390/su16177489>.
- [2] R. Kondo, B. Du, Y. Narusue, and H. Morikawa, "Machine Learning Framework Supervised by Hydraulic Mechanical Models for Real-time Pluvial Flood Prediction," *Journal of Information Processing*, vol. 31, pp. 256–264, 2023, <https://doi.org/10.2197/ipsjip.31.256>.
- [3] L. Schmidt, F. Heße, S. Attinger, and R. Kumar, "Challenges in Applying Machine Learning Models for Hydrological Inference: A Case Study for Flooding Events Across Germany," *Water Resources Research*, vol. 56, no. 5, 2020, Art. no. e2019WR025924, <https://doi.org/10.1029/2019WR025924>.
- [4] S. S. Mallela and S. K. Jonnalagadda, "Rainfall Prediction Based on Spatial Attention Layer: A Case Study Analysis," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 3, pp. 49–58, 2022, <https://doi.org/10.22266/ijies.2022.0630.05>.
- [5] W. M. Ridwan, M. Sapitang, A. Aziz, K. F. Kushiar, A. N. Ahmed, and A. El-Shafie, "Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1651–1663, Jun. 2021, <https://doi.org/10.1016/j.asej.2020.09.011>.
- [6] C. Albertini, A. Gioia, V. Iacobellis, G. P. Petropoulos, and S. Manfreda, "Assessing multi-source random forest classification and robustness of predictor variables in flooded areas mapping," *Remote Sensing Applications: Society and Environment*, vol. 35, Aug. 2024, Art. no. 101239, <https://doi.org/10.1016/j.rsase.2024.101239>.
- [7] H. Yu, Z. Luo, L. Wang, X. Ding, and S. Wang, "Improving the Accuracy of Flood Susceptibility Prediction by Combining Machine Learning Models and the Expanded Flood Inventory Data," *Remote Sensing*, vol. 15, no. 14, Jan. 2023, Art. no. 3601, <https://doi.org/10.3390/rs15143601>.
- [8] E. H. Ighile, H. Shirakawa, and H. Tanikawa, "Application of GIS and Machine Learning to Predict Flood Areas in Nigeria," *Sustainability*, vol. 14, no. 9, Jan. 2022, Art. no. 5039, <https://doi.org/10.3390/su14095039>.
- [9] Hanshika. V. G, Preethi. S, Sreemathy. S. P, and E. Freeda. S, "Flood Prediction Using Logistic Regression," in *2023 International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Dec. 2023, pp. 1174–1179, <https://doi.org/10.1109/ICCPCT58313.2023.10245832>.
- [10] C. Vlachas *et al.*, "Random forest classification algorithm for medical industry data," *SHS Web of Conferences*, vol. 139, 2022, Art. no. 03008, <https://doi.org/10.1051/shsconf/202213903008>.
- [11] Z. Zhu and Y. Zhang, "Flood disaster risk assessment based on random forest algorithm," *Neural Computing and Applications*, vol. 34, no. 5, pp. 3443–3455, Mar. 2022, <https://doi.org/10.1007/s00521-021-05757-6>.
- [12] H. Sato, Y. Kimura, M. Ohba, Y. Ara, S. Wakabayashi, and H. Watanabe, "Prediction of Prednisolone Dose Correction Using Machine Learning," *Journal of Healthcare Informatics Research*, vol. 7, no. 1, pp. 84–103, Mar. 2023, <https://doi.org/10.1007/s41666-023-00128-3>.
- [13] Y. Qian *et al.*, "Classification of Imbalanced Travel Mode Choice to Work Data Using Adjustable SVM Model," *Applied Sciences*, vol. 11, no. 24, Jan. 2021, Art. no. 11916, <https://doi.org/10.3390/app112411916>.
- [14] P. Deivendran, S. Selvakanmani, G. Malathi, and V. Vinoth Kumar, "Machine Learning based IoT Flood Prediction Using Data Modeling and Decision Support System," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 9, pp. 2489–2499, Nov. 2023, <https://doi.org/10.17762/ijritcc.v11i9.9318>.
- [15] J. Liang, Q. Liu, N. Nie, B. Zeng, and Z. Zhang, "An Improved Algorithm based on KNN and Random Forest," in *3rd International Conference on Computer Science and Application Engineering*, Sanya, China, Oct. 2019, pp. 1–6, <https://doi.org/10.1145/3331453.3360963>.
- [16] Y. Xu, J. Zhao, B. Wan, J. Cai, and J. Wan, "Flood Forecasting Method and Application Based on Informer Model," *Water*, vol. 16, no. 5, Jan. 2024, Art. no. 765, <https://doi.org/10.3390/w16050765>.
- [17] W. Dai, Y. Tang, Z. Zhang, and Z. Cai, "Ensemble Learning Technology for Coastal Flood Forecasting in Internet-of-Things-Enabled Smart City," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, Sep. 2021, Art. no. 166, <https://doi.org/10.1007/s44196-021-00023-y>.
- [18] S. Law, P. Jeszenszky, and K. Yano, "Examining geographical generalisation of machine learning models in urban analytics through street frontage classification and house price regression," in *GIScience 2021 Short Paper Proceedings*, Sep. 2021, <https://doi.org/10.25436/E2VC71>.
- [19] E. Wolff and D. Ramirez-Lovering, "Living with Floods in Informal Settlements: Compounding and Cascading Risks in Makassar, Indonesia," in *Complex Disasters: Compounding, Cascading, and Protracted*, A. Lukaszewicz and T. O'Donnell, Eds. New York, NY, USA: Springer, 2022, pp. 161–181.
- [20] A. B. Amjoud and M. Amrouch, "Transfer Learning for Automatic Image Orientation Detection Using Deep Learning and Logistic Regression," *IEEE Access*, vol. 10, pp. 128543–128553, Jan. 2022, <https://doi.org/10.1109/ACCESS.2022.3225455>.
- [21] H. P. Singh and H. N. Alhulail, "Predicting Student-Teachers Dropout Risk and Early Identification: A Four-Step Logistic Regression Approach," *IEEE Access*, vol. 10, pp. 6470–6482, Jan. 2022, <https://doi.org/10.1109/ACCESS.2022.3141992>.
- [22] R. Madhuri, S. Sista, and K. Srinivasa Raju, "Application of machine learning algorithms for flood susceptibility assessment and risk management," *Journal of Water and Climate Change*, vol. 12, no. 6, pp. 2608–2623, Apr. 2021, <https://doi.org/10.2166/wcc.2021.051>.
- [23] C.-N. Li, Y.-H. Shao, Z. Wang, N.-Y. Deng, and Z.-M. Yang, "Robust Bhattacharyya bound linear discriminant analysis through an adaptive algorithm," *Knowledge-Based Systems*, vol. 183, Nov. 2019, Art. no. 104858, <https://doi.org/10.1016/j.knosys.2019.07.029>.
- [24] C.-N. Li, Y. Li, Y.-H. Meng, P.-W. Ren, and Y.-H. Shao, "L<sub>2,1</sub>-Norm Regularized Robust and Sparse Linear Discriminant Analysis via an Alternating Direction Method of Multipliers," *IEEE Access*, vol. 11, pp. 34250–34259, Jan. 2023, <https://doi.org/10.1109/ACCESS.2023.3264688>.
- [25] A. Dhurandhar and A. Dobra, "Probabilistic characterization of nearest neighbor classifier," *International Journal of Machine Learning and Cybernetics*, vol. 4, no. 4, pp. 259–272, Aug. 2013, <https://doi.org/10.1007/s13042-012-0091-y>.
- [26] T. Liao, Z. Lei, T. Zhu, S. Zeng, Y. Li, and C. Yuan, "Deep Metric Learning for K Nearest Neighbor Classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 1, pp. 264–275, Jan. 2023, <https://doi.org/10.1109/TKDE.2021.3090275>.
- [27] K. Taunk, S. De, S. Verma, and A. Swetapadma, "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification," in *International Conference on Intelligent Computing and Control Systems*, Madurai, India, Dec. 2019, pp. 1255–1260, <https://doi.org/10.1109/ICCS45141.2019.9065747>.
- [28] A. H. Jahromi and M. Taheri, "A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features," in *Artificial Intelligence and Signal Processing Conference*, Shiraz, Iran, Oct. 2017, pp. 209–212, <https://doi.org/10.1109/AISP.2017.8324083>.
- [29] Y. Dani and M. A. Ginting, "Comparison of Iris dataset classification with Gaussian naive Bayes and decision tree algorithms," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, pp. 1959–1968, Apr. 2024, <https://doi.org/10.11591/ijece.v14i2.pp1959-1968>.

- [30] N. A. F. Sulaiman, S. M. Shaharudin, S. Ismail, N. H. Zainuddin, M. L. Tan, and Y. Abd Jalil, "Predictive Modelling of Statistical Downscaling Based on Hybrid Machine Learning Model for Daily Rainfall in East-Coast Peninsular Malaysia," *Symmetry*, vol. 14, no. 5, May 2022, Art. no. 927, <https://doi.org/10.3390/sym14050927>.
- [31] P. Katta and N. Hegde, "A Hybrid Adaptive Neuro-Fuzzy Interface and Support Vector Machine Based Sentiment Analysis on Political Twitter Data," *International Journal of Intelligent Engineering and Systems*, vol. 12, no. 1, pp. 165–173, Feb. 2019, <https://doi.org/10.22266/ijies2019.0228.17>.
- [32] Y.-S. Jeon, D.-H. Yang, and D.-J. Lim, "FlexBoost: A Flexible Boosting Algorithm With Adaptive Loss Functions," *IEEE Access*, vol. 7, pp. 125054–125061, Jan. 2019, <https://doi.org/10.1109/ACCESS.2019.2938356>.
- [33] M. Liao, H. Wen, L. Yang, G. Wang, X. Xiang, and X. Liang, "Improving the model robustness of flood hazard mapping based on hyperparameter optimization of random forest," *Expert Systems with Applications*, vol. 241, May 2024, Art. no. 122682, <https://doi.org/10.1016/j.eswa.2023.122682>.
- [34] X. Guan, F. Yu, H. Xu, C. Li, and Y. Guan, "Flood risk assessment of urban metro system using random forest algorithm and triangular fuzzy number based analytical hierarchy process approach," *Sustainable Cities and Society*, vol. 109, Aug. 2024, Art. no. 105546, <https://doi.org/10.1016/j.scs.2024.105546>.
- [35] H. Hairani, T. Widiyaningtyas, and D. D. Prasetya, "Addressing Class Imbalance of Health Data: A Systematic Literature Review on Modified Synthetic Minority Oversampling Technique (SMOTE) Strategies," *JOIV: International Journal on Informatics Visualization*, vol. 8, no. 3, pp. 1310–1318, Sep. 2024, <https://doi.org/10.62527/joiv.8.3.2283>.
- [36] M. Gava, P. M. Biron, and T. Buffin-Belanger, "A random forest machine learning model to detect fluvial hazards," *River Research and Applications*, vol. 40, no. 10, pp. 1837–1854, 2024, <https://doi.org/10.1002/rra.4353>.
- [37] H. Zhang, A. Zhou, and H. Zhang, "An Evolutionary Forest for Regression," *IEEE Transactions on Evolutionary Computation*, vol. 26, no. 4, pp. 735–749, Dec. 2022, <https://doi.org/10.1109/TEVC.2021.3136667>.
- [38] X. Fang, M. Liao, L. Yang, X. Jiang, J. Jin, and Q. Zhu, "Assessment of Flash Flood Risk Using Random Forest Regression Model Integrated with Binary Statistics," in *IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium*, Athens, Greece, Jul. 2024, pp. 3952–3955, <https://doi.org/10.1109/IGARSS53475.2024.10640642>.
- [39] "BMKG Dataonline," *BMKG Dataonline*. <https://dataonline.bmkg.go.id>.