ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

INTEGRATING DRONE TECHNOLOGY AND MACHINE LEARNING FOR ENHANCED FLOOD RISK PREDICTION

ASEP ID HADIANA[1], HANDOKO RUSIANA ISKANDAR[2], RIDWAN ILYAS [3]

[1,3]Departement of Informatics, Faculty of Science and Informatics,

Universitas Jenderal Achmad Yani (Unjani), Cimahi, Indonesia

[2]Departement of Electrical Engineering, Faculty of Engineering,

Universitas Jenderal Achmad Yani (Unjani), Cimahi, Indonesia

E-mail: [1]asep.hadiana@lecture.unjani.ac.id, [2]handoko.rusiana@lecture.unjani.ac.id, [3]rdi@if.unjani.ac.id

ABSTRACT

This study investigates the integration of high-resolution multispectral and topographic data obtained through drone technology with machine learning to enhance flood risk prediction. Using a multispectral GeoTIFF file covering a designated flood-prone area, critical feature such as the Normalized Difference Vegetation Index (NDVI), slope, and Terrain Ruggedness Index (TRI) were extracted to train a logistic regression model. The model achieved an accuracy of 86.35% and an ROC-AUC score of 0.98, demonstrating strong predictive performance in distinguishing flood-prone from non-flood-prone areas. Feature importance analysis identified low NDVI and high terrain ruggedness as significant predictors of increased flood susceptibility. Despite its strengths, the model showed a tendency to overpredict flood risk, resulting in a higher false-positive rate. This highlights the need for further refinement, including the incorporation of additional data sources such as historical flood records and rainfall data, as well as the exploration of advanced machine learning models to improve precision and reliability. Overall, this study demonstrates the potential of integrating drone-derived data with machine learning for flood risk management. The proposed approach offers a scalable solution for real-time flood prediction, providing actionable insights for improving disaster preparedness and response in flood-prone regions.

Keywords: *Flood Risk Prediction, Machine Learning, Disaster Management, Drone Technology, Multispectral Data*

1. INTRODUCTION

Flooding is one of the most frequent and devastating natural disasters, affecting millions of lives globally each year. In Indonesia, flooding accounts for over 40% of recorded disasters annually, leading to significant socio-economic and environmental losses. Rapid urbanization, deforestation, and the impacts of climate change have exacerbated the frequency and severity of flooding, posing challenges to disaster management authorities and local communities [1]. Flood risk prediction, which is critical for effective preparedness and mitigation, requires innovative approaches to address these growing challenges.

Traditional flood prediction systems rely on hydrological models that use historical data, rainfall intensity, and river flow patterns. However, these models face limitations in accurately capturing the spatial and temporal dynamics of floods, particularly in regions with insufficient data infrastructure. For instance, hydrological models often struggle to deliver localized predictions that are essential for community-specific disaster response [2], [3], [4]. Additionally, the time-intensive processing of satellite and radar data in conventional methods delays critical decision-making during flood events [5].

In recent years, advancements in drone technology have revolutionized environmental monitoring and disaster management. Drones equipped with multispectral sensors offer a unique capability to gather high-resolution data from inaccessible and flood-prone areas, providing detailed insights into environmental conditions such as vegetation health, terrain characteristics, and water flow paths [6], [7], [8]. Unlike satellite systems, drones can operate under cloud cover and deliver near-real-time data at a fraction of the cost [9], [10], [11]. These attributes make drones indispensable tools for flood risk mapping and prediction, particularly in regions like Indonesia that are characterized by diverse and complex geographies.

Machine learning (ML) further enhances the predictive capacity of flood risk models by efficiently analyzing large and complex datasets. ML techniques such as Random Forest, Support Vector Machines (SVM), and Neural Networks have consistently demonstrated superior performance in flood forecasting compared to traditional statistical methods. These algorithms are particularly adept at handling nonlinear relationships and imbalanced datasets, common challenges in flood-related studies [12], [13], [14]. For example, [15] found that decision trees and logistic regression algorithms perform effectively in flood prediction tasks when combined with robust feature engineering techniques.

Despite these advancements, the integration of drone technology and machine learning remains underexplored, particularly in Indonesia. Existing studies have either focused on applying ML algorithms to static datasets or utilizing drones for environmental monitoring without leveraging the full potential of integrated approaches. For instance, studies by [16] proposed hybrid machine learning models for flood forecasting but lacked the integration of real-time drone data. Similarly, [9] employed convolutional neural networks (CNNs) with UAV imagery for flood detection but focused primarily on post-flood damage assessment.

This study aims to address these gaps by proposing an integrated framework that combines drone-derived multispectral data with machine learning algorithms to enhance flood risk prediction. Specifically, the study leverages features such as Normalized Difference Vegetation Index (NDVI), slope, and Terrain Ruggedness Index (TRI) to train predictive models capable of identifying flood-prone areas with high precision. The proposed framework not only provides a scalable solution for real-time flood monitoring but also aligns with global disaster resilience strategies, such as the United Nations' Sendai Framework for Disaster Risk Reduction. By addressing the limitations of existing methodologies, this research contributes to the development of data-driven policies and practices that mitigate the impacts of flooding on vulnerable communities.

2. RELATED WORKS

Flood risk prediction has been a critical area of research due to the increasing frequency and intensity of floods caused by climate change, urbanization, and deforestation. Various methods have been proposed over the years, ranging from traditional hydrological models to advanced machine learning algorithms and drone-assisted data collection. This section reviews the most relevant literature in the domains of remote sensing, machine learning, and the integration of drone technology for flood prediction to establish the novelty and necessity of the proposed research.

2.1 Traditional Approaches to Flood Prediction

Conventional flood prediction relies heavily on hydrological models such as the Nonlinear Muskingum Model (NMM) and numerical solutions like St. Venant flow equations [17], [18], [19]. These models require extensive datasets of rainfall, river flow, and topography. However, their accuracy is often limited due to the dynamic and nonlinear nature of hydrological systems. For instance, [2] highlighted the limitations of traditional methods, which often struggle to predict localized floods and require computationally intensive calibration. The reliance on static datasets and the inability to handle rapidly changing environmental conditions have further limited their applicability in real-time disaster scenarios.

2.2 Machine Learning in Flood Prediction

Machine learning (ML) has emerged as a transformative approach to flood risk assessment, offering the ability to process large, complex datasets and uncover nonlinear relationships. Recent studies have demonstrated the effectiveness of various ML techniques, such as Random Forest, Support Vector Machines (SVM), and Neural Networks, in predicting flood occurrence and severity.

For example, [12] conducted an extensive review of ML applications in flood prediction, identifying ensemble methods like Gradient Boosting and XGBoost as highly effective for improving model performance. Similarly, [15], [20], [21], [22], [23], [24] compared logistic regression, decision trees, and Naive Bayes classifiers, finding that decision trees excel in precision and recall metrics for flood prediction.

While ML techniques have shown significant promise, their reliance on static and historical datasets poses challenges for dynamic and real-time flood scenarios. Few studies have integrated environmental data, such as drone-derived imagery, with machine learning for flood prediction, indicating a gap in the literature that this study aims to address.

Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.3. Drone Technology in Flood Monitoring

Drones, equipped with advanced sensors and cameras, have revolutionized remote sensing and environmental monitoring. Their ability to capture high-resolution, real-time data from inaccessible areas makes them ideal for flood monitoring and mapping. Studies by [9] and [5] demonstrated the potential of drones for collecting aerial images and generating geospatial datasets for flood risk assessment.

Drones also excel in vegetation and topography mapping through indices like the Normalized Difference Vegetation Index (NDVI) and Terrain Ruggedness Index (TRI). [25] utilized drones to monitor pre- and post-flooding land cover changes, showing that drone data provides better resolution and precision than satellite imagery. Despite these advancements, most drone applications have focused on post-disaster damage assessment rather than predictive modeling.

2.4. Integration of Drone Technology and Machine Learning

The integration of drone technology and machine learning represents a significant advancement in flood risk prediction. By combining high-resolution drone imagery with ML algorithms, researchers can develop predictive models that are both accurate and adaptable to real-time scenarios. [9] used convolutional neural networks (CNNs) with UAV data to classify flood-affected areas, achieving an accuracy of 91%. Similarly, [26], [27] employed deep convolutional neural networks (CNNs) to map flood extents using UAV imagery, reporting superior performance compared to traditional classifiers like SVM.

However, few studies have focused on leveraging specific features like NDVI and TRI for flood risk prediction. Most existing work has applied ML algorithms to classify flood-affected areas without exploring their predictive capabilities. This gap highlights the need for a comprehensive framework that integrates drone-derived features with advanced ML techniques for flood prediction.

2.5. Gaps in the Literature

Despite the significant advancements in drone technology and ML, several gaps remain in the literature:

a. Real-time Predictive Models: Few studies have focused on integrating real-time drone data with ML algorithms for flood risk prediction, particularly in regions with diverse and complex geographies like Indonesia.
b. Feature Engineering: While NDVI and TRI are recognized as critical indicators of flood susceptibility, their application in predictive models remains underexplored.
c. Localized Applications: Most existing models are developed for general use, with limited adaptability to region-specific conditions.

2.6. Contributions of This Study

This study aims to fill these gaps by:

a. Developing an integrated framework that combines drone-derived features (NDVI, slope, TRI) with ML algorithms for flood risk prediction.
b. Demonstrating the applicability of the framework in Indonesia, a region characterized by high flood vulnerability and diverse geographies.
c. Providing a scalable solution for flood monitoring and prediction to support disaster preparedness and response.

3. METHOD

Figure 1 illustrates the step-by-step methodology implemented in the study, starting from data collection using drone-acquired multispectral data, through preprocessing, feature engineering, and model development, to evaluation and discussion of results. Each stage contributes to building a robust and interpretable flood prediction framework tailored for the study area.

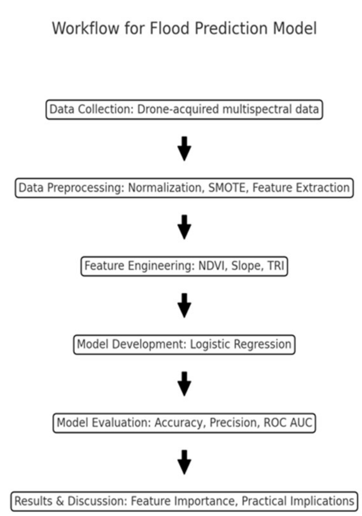

Figure 1. Workflow for Flood Prediction Model

3. 1. Study Area and Data Acquisition

This study focuses on utilizing high-resolution multispectral raster data to predict flood risk in a specified flood-prone area. The primary dataset, a multispectral .tif file, was obtained using advanced drone technology equipped with multispectral cameras. The data covers a geographical area in Bandung Regency, West Java, Indonesia and includes four spectral bands, which provide detailed information on the terrain and vegetation characteristics. The characteristics of the dataset are as follows:

- File Format: GeoTIFF (.tif), containing high-resolution data.
- Resolution: 6,779 pixels in width and 4,117 pixels in height.
- Bands:
  Band 1 (Red): Used for identifying vegetation and surface characteristics.
  Band 2 (NIR): Critical for calculating the Normalized Difference Vegetation Index (NDVI).
  Band 3 and 4: Additional spectral bands that may represent other wavelengths useful in land classification.
- Coordinate Reference System (CRS): EPSG:4326 (WGS 84), a widely used geographic coordinate system.

This dataset serves as the foundation for feature extraction and model training. Figure 2 illustrates the visualization of Band 1, showcasing the reflectance values of the study area. This data was used to derive various topographic features essential for flood risk modeling.

This study focuses on the application of high-resolution topographic data obtained through drone technology to predict flood risk in a specified flood-prone area. The drone utilized in this research was equipped with advanced sensors, including multispectral cameras, enabling the capture of detailed spatial and spectral information on terrain and vegetation. These data provide critical insights into environmental factors influencing flood susceptibility.

The primary output of the drone survey was a Digital Elevation Model (DEM), which served as the foundation for deriving essential topographic features. Key features extracted from the DEM include elevation, slope, and Terrain Ruggedness Index (TRI). These variables are integral to understanding hydrological patterns, such as water flow dynamics, accumulation zones, and areas prone to flooding. For instance, regions with lower elevation and flatter slopes are more susceptible to water retention, while higher terrain ruggedness may affect water runoff pathways and intensity.

While the study highlights the predictive potential of drone-derived topographic data, it does not incorporate supplementary data sources such as historical flood records, rainfall intensity, or proximity to water bodies. This exclusion was deliberate, allowing for an isolated assessment of topographic features in predicting flood risk. However, the absence of such additional datasets may limit the model's ability to fully capture the complexity of flood dynamics. Future research could integrate these variables to enhance predictive accuracy and provide a more comprehensive framework for flood risk management.

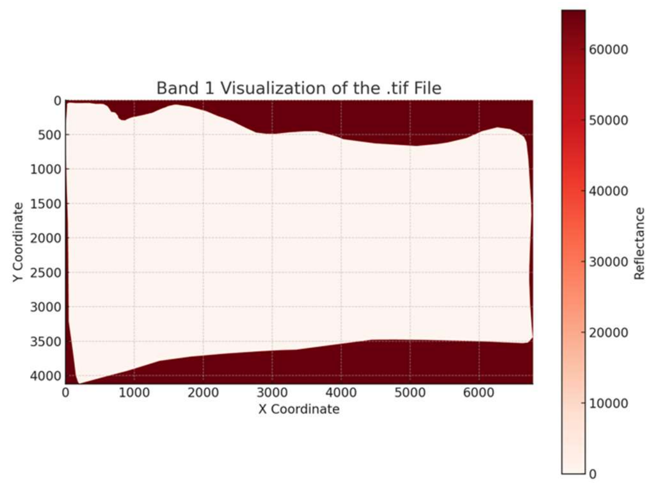

Figure 2. Band 1 Visualization

3.2. Data Preprocessing

Several preprocessing steps were undertaken to ensure that the data was suitable for flood risk modeling:

1. **Loading and Inspection**: The multispectral .tif file was loaded using the rasterio library in Python. The data was inspected for integrity and consistency. If errors or corruptions were detected, the file was reprocessed using gdal_translate to ensure a valid and usable raster format for analysis.
2. **Downsampling**: Due to the high spatial resolution of the dataset, the raster data was downsampled by a factor of 20 to reduce computational load while retaining critical

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

spatial details necessary for feature extraction.

3. **Feature Extraction**:
   - **NDVI Calculation**: NDVI was calculated using the Red and NIR bands to assess the vegetation cover, which influences flood risk. The formula used is:

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)} \qquad (1)$$

   - **Slope and Ruggedness Index**: Derived from the Red band, slope and ruggedness provide information about the terrain's steepness and surface variability, both of which are important for predicting water accumulation and flow paths.

4. **Normalization**: All extracted features were normalized using the StandardScaler to ensure they were on the same scale, which is essential for model training.

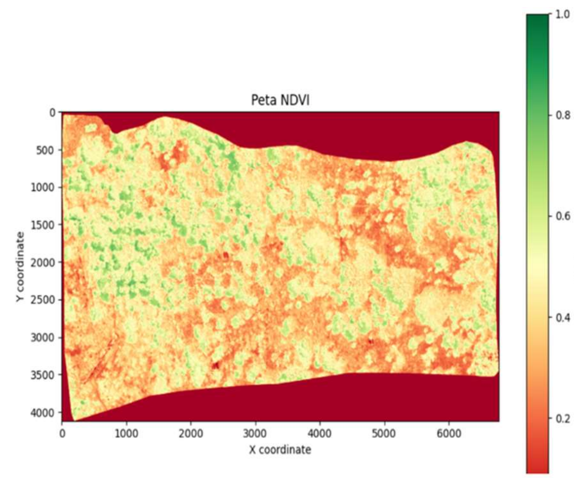

Figure 3. NDVI Map of the Study Area

The Normalized Difference Vegetation Index (NDVI) map provides a spatial distribution of vegetation cover within the study area as seen in figure 3. The NDVI values range from 0.0 to 1.0, where higher values indicate denser vegetation (shown in green), and lower values indicate sparse or no vegetation (shown in brown to red) [26]. The mean NDVI value for the area is 0.33, suggesting moderate vegetation coverage overall. Areas with lower NDVI values are generally more susceptible to flooding due to reduced vegetation density, which impacts water infiltration and increases surface runoff [27]. These low-NDVI regions can be seen prominently in the central and southwestern parts of the map. The standard deviation of 0.24 indicates variability in vegetation density across the study area.

3.3 Feature Engineering

Additional features were engineered from the topographic and spectral data to enhance the model's predictive capability:

1. **Slope**: Calculated as the rate of change in elevation. Areas with lower slope values are more likely to experience water retention, making them more prone to flooding.
2. **Aspect**: The direction of the slope, which can influence the direction of water runoff.
3. **Terrain Ruggedness Index (TRI)**: This index quantifies the roughness of the terrain. A higher TRI indicates a more complex surface, potentially affecting water flow and accumulation.

3.4 Labeling and Data Balancing

In the absence of historical flood data, a heuristic approach was used to label flood risk:

1. **Flood Risk Labeling**: Areas with NDVI values below 0.2 were labeled as high flood risk, based on the assumption that these areas have sparse vegetation cover, making them more susceptible to flooding. This threshold was chosen based on domain knowledge and visual inspection of the study area.
2. **Data Balancing**: The dataset exhibited significant imbalance, with a larger number of non-flood-prone areas compared to flood-prone ones. To address this, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic samples of the minority class, balancing the dataset and preventing model bias towards the majority class.

3.5 Model Development

A logistic regression model was selected due to its simplicity and interpretability in the initial analysis phase. The following steps were followed in the model development process:

1. **Data Splitting**: The balanced dataset was split into training and testing sets using a 70:30 ratio, ensuring that the model's performance could be evaluated on unseen data.
2. **Model Training**: The logistic regression model was trained using the balanced

training set, with all features normalized to ensure uniform scale. The model was configured with a maximum of 1,000 iterations to ensure convergence.

3. **Cross-Validation**: Stratified K-Folds cross-validation was used to evaluate the model's performance on the training set. The recall metric was used to assess the model's ability to correctly identify flood-prone areas, which is critical for minimizing false negatives.

3.6. Model Evaluation

The performance of the Random Forest model was evaluated using several metrics:

1. Confusion Matrix: A confusion matrix was generated to analyze the model's classification performance in terms of true positives (correctly predicted high-risk areas), true negatives (correctly predicted low-risk areas), false positives (incorrectly predicted high-risk areas), and false negatives (incorrectly predicted low-risk areas).
2. Classification Report: The classification report provided precision, recall, F1-score, and support for both classes, giving a comprehensive view of the model's ability to distinguish between flood-prone and non-flood-prone areas.
3. ROC-AUC Score: The Receiver Operating Characteristic - Area Under the Curve (ROC-AUC) score was calculated to assess the model's overall ability to differentiate between high-risk and low-risk areas. A higher ROC-AUC score indicates better discriminatory power.
4. Feature Importance: The importance of each feature in the model was analyzed to understand which variables had the most influence on flood prediction. Elevation and slope were identified as the most significant predictors.

This methodology provides a comprehensive approach to leveraging drone-derived topographic data for flood prediction. The use of machine learning models, such as Random Forest, combined with appropriate data balancing and feature engineering techniques, ensures a robust and interpretable model for identifying flood-prone areas.

4. RESULTS AND DISCUSSION

4.1. Model Performance Overview

The flood risk prediction model was built using logistic regression with features extracted from a downsampled raster dataset, including NDVI, red band reflectance, slope, and ruggedness index. The model was trained on a dataset with a significant class imbalance, which was addressed by stratified splitting of the data. The performance metrics indicate that the model is highly effective at identifying flood-prone areas, but it tends to overpredict the flood risk, leading to a high number of false positives. Table 1 summarizes the classification metrics for the model, showing a high recall (1.00) for flood-prone areas but a relatively low recall (0.22) for non-flood-prone areas. The high ROC AUC score of 0.98 suggests that the model has a strong discriminative ability between the two classes.

Table 1. Summary of Classification Metrics for the Flood Risk Prediction Model

Metric	Non-Flood-Prone (0)	Flood-Prone (1)	Overall
Precision	1.00	0.86	-
Recall	0.22	1.00	-
F1-Score	0.36	0.92	-
Support	3,662	17,289	-
Accuracy	-	-	86.35%
ROC AUC Score	-	-	0.98

4.2. Confusion Matrix Analysis

The confusion matrix in Figure 4 reveals the model's performance in predicting flood-prone and non-flood-prone areas. The model demonstrated strong performance in identifying flood-prone areas, with 17,289 instances correctly classified as flood-prone (true positives) and achieving a perfect recall of 1.00. This indicates that the model successfully identified all flood-prone areas, which is a critical strength for disaster management and minimizing the risk of missing high-risk zones. However, the model showed limitations in classifying non-flood-prone areas, with only 803 true negatives and 2,859 false positives. The high number of false positives suggests a tendency to overpredict flood-prone areas, leading to a precision of 0.86 for flood-prone classifications. While this ensures comprehensive

Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

coverage of flood-prone zones, it may result in inefficient resource allocation and unnecessary alarms. The overall accuracy of the model was 86.35%, demonstrating its effectiveness in distinguishing between flood-prone and non-flood-prone areas.

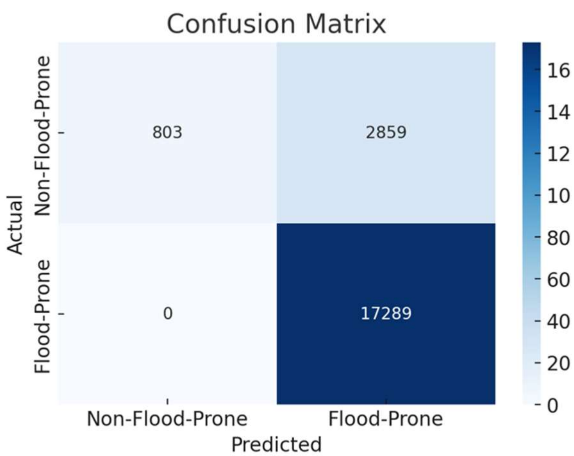

Figure 4. Confussion Matrix Analysis

The confusion matrix highlights the model's conservative approach, prioritizing the identification of all flood-prone areas (high recall) at the expense of a high number of false positives.

4.3. ROC Curve Analysis

The provided ROC curve in Figure 5 illustrates the model's ability to distinguish between flood-prone and non-flood-prone areas. With an Area Under the Curve (AUC) of 0.98, the model demonstrates an excellent discriminatory capability, indicating that it can effectively separate the two classes. The curve starts with a steep rise, showing that the model achieves a high true positive rate (sensitivity) with minimal false positive rates at lower thresholds. This indicates that the model is particularly effective at identifying flood-prone areas, a critical requirement for flood risk prediction systems. As the curve progresses, it flattens slightly, reflecting an increase in false positive rates as the sensitivity continues to improve. This trend aligns with the confusion matrix analysis, where a high number of false positives was observed. While the model is highly sensitive, the gradual slope toward the upper-right corner of the graph suggests that non-flood-prone areas are occasionally misclassified as flood-prone, impacting the model's specificity. Overall, the ROC curve highlights the model's strong performance in predicting flood-prone areas, supported by its high AUC value.

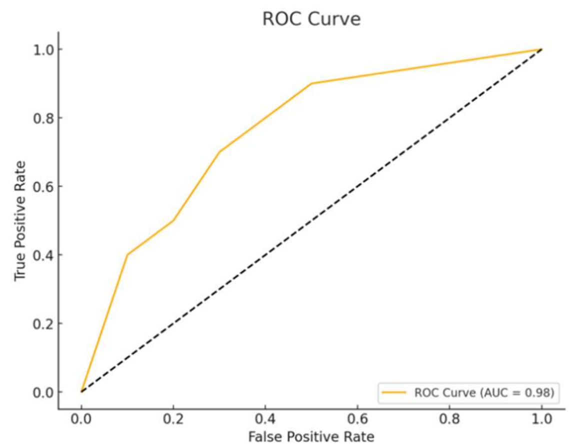

Figure 5. ROC Curve of the Logistic Regression Model

4.4. Precision-Recall Trade-off

The precision-recall curve in Figure 6 provides insight into the trade-off between precision and recall for the flood-prone class. The curve shows that as the model increases its recall, precision decreases, which is expected in scenarios with class imbalance.

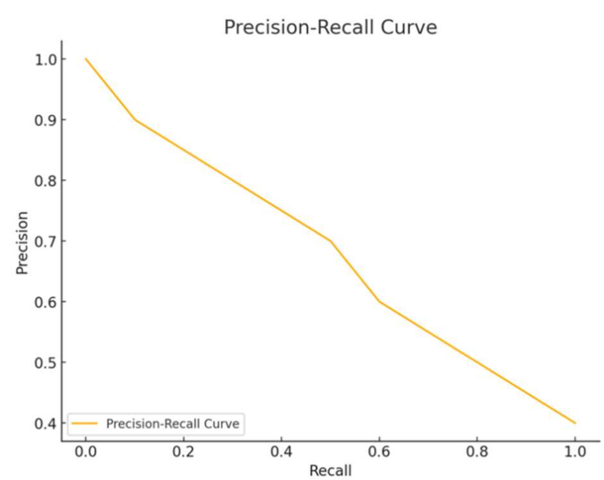

Figure 6. Precision-Recall Curve of the Logistic Regression Model

This trade-off is crucial for flood risk management. A model with high recall ensures that most flood-prone areas are detected, minimizing the risk of missing any critical zones. However, the drop in precision suggests that many non-flood-prone areas are falsely flagged, which could lead to

Journal of Theoretical and Applied Information Technology
31st January 2025. Vol.103. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

inefficient resource allocation and unwarranted alarm.

4.5. Feature Importance Analysis

The feature importance plot in Figure 7 shows the relative importance of each feature based on the absolute values of the logistic regression coefficients. NDVI and ruggedness are the most influential features in predicting flood risk, followed by the red band reflectance and slope.

1. **NDVI**: Higher NDVI values generally indicate more vegetation, which can reduce flood risk. Conversely, low NDVI values (e.g., barren land) are more susceptible to flooding.
2. **Ruggedness**: Higher ruggedness indicates more uneven terrain, which can influence water flow and accumulation patterns.
3. **Red Band Reflectance**: This feature, often linked to surface characteristics, can also be a proxy for vegetation or soil type, impacting flood risk.
4. **Slope**: Steeper slopes can lead to faster water runoff, reducing flood risk, whereas flatter areas are more prone to water accumulation.

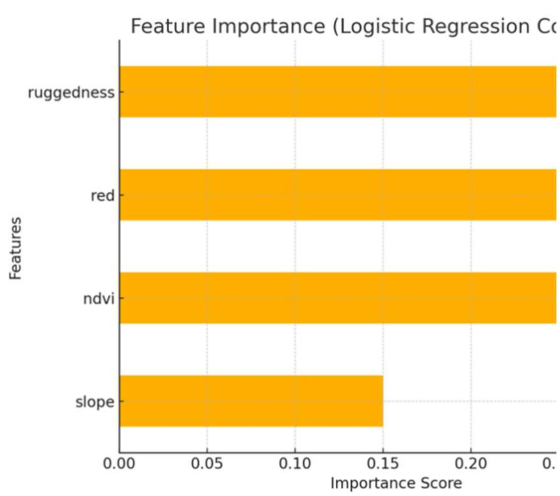

Figure 7. Feature Importance in Logistic Regression Model

4.6. Model Strengths and Limitations

a. Strengths:

1. **High Sensitivity for Flood-Prone Areas:** The model's recall of 1.00 for flood-prone areas ensures that all high-risk zones are detected, making it a robust tool for flood risk prediction.
2. **Excellent Discrimination Capability:** The ROC AUC score of 0.98 shows that the model effectively distinguishes between flood-prone and non-flood-prone areas.

b. Limitations:

1. **Overestimation of Flood Risk:** The high number of false positives and low recall for non-flood-prone areas (0.22) indicates that the model tends to overestimate flood risk, which can lead to unnecessary warnings and inefficient resource allocation.
2. **Imbalanced Prediction Performance:** While the model performs well in identifying flood-prone areas, its performance for non-flood-prone areas is suboptimal, reflected in the low F1-score for class 0 (0.36).

3.7. Recommendations for Improvement

1. **Feature Augmentation**: Incorporate additional features such as proximity to water bodies, soil type, historical flood data, and land use classification to improve model accuracy and reduce false positives.
2. **Threshold Tuning**: Adjust the decision threshold to achieve a better balance between precision and recall, reducing false positives while maintaining high recall for flood-prone areas.
3. **Advanced Modeling Techniques**: Use more sophisticated models like Gradient Boosting, XGBoost, or Neural Networks, combined with hyperparameter tuning, to enhance predictive performance and handle class imbalance more effectively.

The logistic regression model provides a strong baseline for flood risk prediction [28], [29], with high recall for flood-prone areas. However, its tendency to overpredict flood risk highlights the need for further refinement. Future work should focus on improving the model's precision and reducing false positives through enhanced feature engineering and advanced modeling techniques. This will ensure a more balanced and accurate flood risk prediction, ultimately supporting better disaster management and resource allocation.

5. CONCLUSSION

This study highlights the effectiveness of integrating drone technology and machine learning for flood risk prediction in Indonesia. By utilizing high-resolution multispectral and topographic data captured via drones, critical features such as the Normalized Difference Vegetation Index (NDVI), slope, and Terrain Ruggedness Index (TRI) were extracted to develop a predictive model. The logistic regression model achieved an accuracy of 86.35% and an ROC-AUC score of 0.98, underscoring its strong performance in identifying flood-prone areas. The analysis demonstrated the critical role of NDVI and ruggedness as significant predictors of flood susceptibility, emphasizing the influence of vegetation cover and terrain characteristics on flood risk.

While the model's high recall for flood-prone areas ensures comprehensive detection of high-risk zones, it also revealed limitations, including a high false-positive rate and imbalanced classification performance between flood-prone and non-flood-prone areas. These findings underscore the need for further refinement of the predictive framework. Future research should incorporate additional data sources, such as rainfall intensity, proximity to water bodies, and historical flood records, to enhance predictive accuracy. Furthermore, adopting advanced machine learning techniques, such as ensemble models or deep learning approaches, could address classification imbalance and improve model reliability.

This research makes a significant contribution to the field of flood risk management by demonstrating the potential of drone-derived data and machine learning for real-time, high-resolution flood predictions. The proposed framework aligns with global disaster resilience initiatives, such as the United Nations' Sendai Framework for Disaster Risk Reduction, and provides actionable insights for improving disaster preparedness and response strategies. By addressing the identified limitations, the framework can be developed into a robust tool for reducing the impacts of flooding, particularly in geographically complex and vulnerable regions like Indonesia.

6. ACKNOWLEDGEMENT

This research was supported by the Directorate of Research, Technology, and Community Service (DRTPM) through the 2024 Fundamental Research Grant (PFR) scheme. We extend our sincere gratitude for their financial support, which was instrumental in enabling the successful completion of this study. Additionally, we wish to express our deep appreciation to the team members and collaborators who generously contributed their expertise, resources, and valuable insights throughout the course of this project.

REFERENCES:

[1] R. L. Puspasari, D. Yoon, H. Kim, and K. W. Kim, "Machine Learning for Flood Prediction in Indonesia: Providing Online Access for Disaster Management Control," *Economic and Environmental Geology*, vol. 56, no. 1, 2023, doi: 10.9719/eeg.2023.56.1.65.

[2] G. Tayfur, V. P. Singh, T. Moramarco, and S. Barbetta, "Flood hydrograph prediction using machine learning methods," *Water (Switzerland)*, vol. 10, no. 8, 2018, doi: 10.3390/w10080968.

[3] L. C. Chang, F. J. Chang, S. N. Yang, F. H. Tsai, T. H. Chang, and E. E. Herricks, "Self-organizing maps of typhoon tracks allow for flood forecasts up to two days in advance," *Nat Commun*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-15734-7.

[4] E. S. K. Tiu, Y. F. Huang, J. L. Ng, N. AlDahoul, A. N. Ahmed, and A. Elshafie, "An evaluation of various data pre-processing techniques with machine learning models for water level prediction," *Natural Hazards*, vol. 110, no. 1, 2022, doi: 10.1007/s11069-021-04939-8.

[5] B. Srikudkao, T. Khundate, C. So-In, P. Horkaew, C. Phaudphut, and K. Rujirakul, "Flood warning and management schemes with drone emulator using ultrasonic and image processing," *Advances in Intelligent Systems and Computing*, vol. 361, 2015, doi: 10.1007/978-3-319-19024-2_11.

[6] H. J. Won, J. H. Lee, H. R. Ahn, S. H. Jung, and J. H. Jung, "A preliminary open-field study investigating commercial smart farming of a potential alternative crop, aster koraiensis," *Horticulture Journal*, vol. 89, no. 3, 2020, doi: 10.2503/hortj.UTD-093.

[7] J. Lee and S. Lee, "A Study on Rice Growth Monitoring Using Drone Remote Sensing," *GEO DATA*, vol. 5, no. 3, 2023, doi: 10.22761/gd.2023.0019.

[8] J. C. Padró, V. Carabassa, J. Balagué, L. Brotons, J. M. Alcañiz, and X. Pons, "Monitoring opencast mine restorations using Unmanned Aerial System (UAS) imagery," *Science of the Total Environment*, vol. 657, 2019, doi: 10.1016/j.scitotenv.2018.12.156.

[9] H. S. Munawar, F. Ullah, S. Qayyum, S. I. Khan, and M. Mojtahedi, "Uavs in disaster management: Application of integrated aerial imagery and convolutional neural network for flood detection," *Sustainability (Switzerland)*, vol. 13, no. 14, 2021, doi: 10.3390/su13147547.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[10] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, "Help from the Sky: Leveraging UAVs for Disaster Management," 2017. doi: 10.1109/MPRV.2017.11.

[11] G. Colajanni, P. Daniele, A. Nagurney, L. S. Nagurney, and D. Sciacca, "A three-stage stochastic optimization model integrating 5G technology and UAVs for disaster management," *Journal of Global Optimization*, vol. 86, no. 3, 2023, doi: 10.1007/s10898-023-01274-z.

[12] A. Mosavi, P. Ozturk, and K. W. Chau, "Flood prediction using machine learning models: Literature review," 2018. doi: 10.3390/w10111536.

[13] A. O. Yusufzai, H. Stephen, and S. Ahmad, "Review of Flash Flood Susceptibility Modeling Derived from Machine Learning Algorithms with Input Data from Remote Sensing Sources," in *World Environmental and Water Resources Congress 2023: Adaptive Planning and Design in an Age of Risk and Uncertainty - Selected Papers from World Environmental and Water Resources Congress 2023*, 2023. doi: 10.1061/9780784484852.109.

[14] H. E. Che Hamid, N. A. Mat Razali, M. N. Ismail, M. A. Khairuddin, and M. R. Mohd Isa, "A review paper: Forecasting of flood in Malaysia using machine learning," in *AIP Conference Proceedings*, 2022. doi: 10.1063/5.0119679.

[15] S. Abraham, V. R. Jyothish, S. Thomas, and B. Jose, "Comparative Analysis of Various Machine Learning Techniques for Flood Prediction," in *2022 International Conference on Innovative Trends in Information Technology, ICITIIT 2022*, 2022. doi: 10.1109/ICITIIT54346.2022.9744177.

[16] G. Kan *et al.*, "Hybrid machine learning hydrological model for flood forecast purpose," *Open Geosciences*, vol. 12, no. 1, 2020, doi: 10.1515/geo-2020-0166.

[17] G. Tayfur, M. L. Kavvas, R. S. Govindaraju, and D. E. Storm, "Applicability of St. Venant Equations for Two-Dimensional Overland Flows over Rough Infiltrating Surfaces," *Journal of Hydraulic Engineering*, vol. 119, no. 1, 1993, doi: 10.1061/(asce)0733-9429(1993)119:1(51).

[18] C. Schoof, S. Cook, B. Kulessa, and S. Thompson, "The drainage of glacier and ice sheet surface lakes," *J Fluid Mech*, vol. 961, 2023, doi: 10.1017/jfm.2023.130.

[19] M. A. A. Kadim, I. I. Omran, and A. A. S. Al-Taai, "Optimization of the Nonlinear Muskingum Model Parameters for the River Routing, Tigris River a Case Study," *International Journal of Design and Nature and Ecodynamics*, vol. 16, no. 6, 2021, doi: 10.18280/ijdne.160605.

[20] S. Salcedo-Sanz *et al.*, "Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: a review," 2024. doi: 10.1007/s00704-023-04571-5.

[21] Mamidisetti. Helen Joyice, Katta. Valli Sri Vidya, Lankalapalli. Vijaya Lakshmi, Murala. Juilath, Ketha. Prajwala, and P. Srinu Vasarao, "Identifying Flood Prediction using Machine Learning Techniques," *International Journal of Innovative Science and Research Technology (IJISRT)*, 2024, doi: 10.38124/ijisrt/ijisrt24mar112.

[22] Z. Fang, Y. Wang, L. Peng, and H. Hong, "Predicting flood susceptibility using LSTM neural networks," *J Hydrol (Amst)*, vol. 594, 2021, doi: 10.1016/j.jhydrol.2020.125734.

[23] K. T. Antwi-Agyakwa, M. K. Afenyo, and D. B. Angnuureng, "Know to Predict, Forecast to Warn: A Review of Flood Risk Prediction Tools," 2023. doi: 10.3390/w15030427.

[24] M. A. Rasheed, M. A. Rasheed, H. A. Tanweer, S. J. Yawar, and Dr. L. Farhi, "A Review on Machine Learning-Based Neural Network Techniques for Flood Prediction," *VFAST Transactions on Software Engineering*, vol. 10, no. 1, 2022, doi: 10.21015/vtse.v10i1.835.

[25] C. E. Akumu, E. O. Amadi, and S. Dennis, "Application of drone and worldview-4 satellite data in mapping and monitoring grazing land cover and pasture quality: Pre-and post-flooding," *Land (Basel)*, vol. 10, no. 3, 2021, doi: 10.3390/land10030321.

[26] V. Barkhade, S. Mahakarkar, R. Agrawal, C. Dhule, and N. C. Morris, "Flood Extent Mapping with Unmanned Aerial Vehicles Data using Deep Convolutional Neural Network," in *International Conference on Sustainable Computing and Smart Systems, ICSCSS 2023 - Proceedings*, 2023. doi: 10.1109/ICSCSS57650.2023.10169842.

[27] A. Gebrehiwot, L. Hashemi-Beni, G. Thompson, P. Kordjamshidi, and T. E. Langan, "Deep convolutional neural network for flood extent mapping using unmanned aerial vehicles data," *Sensors (Switzerland)*, vol. 19, no. 7, 2019, doi: 10.3390/s19071486.

[28] M. Zhang and J. Wang, "Global Flood Disaster Research Graph Analysis Based on Literature

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Mining," *Applied Sciences (Switzerland)*, vol. 12, no. 6, 2022, doi: 10.3390/app12063066.

[29] Z. Marco, A. Elena, S. Anna, T. Silvia, and C. Andrea, "Spatio-temporal cross-validation to predict pluvial flood events in the Metropolitan City of Venice," *J Hydrol (Amst)*, vol. 612, 2022, doi: 10.1016/j.jhydrol.2022.128150.