Machine Learning Application of Flood Detecting and Mapping in Urban Area with Geographic Information Data

Leni Novianti[1], Ermatita[2], Abdiansah[3], Ade Silvia Handyani[4], Novie Rahmadani[5], Nyayu Latifah Husni[6]

[1] Doctoral Program in Engineering Science, Universitas Sriwijaya, Indonesia
[2-3] Faculty of Computer Science Universitas Sriwijaya, Indonesia
[4-6] Departement of Electrical Engineering, Politeknik Negeri Sriwijaya, Indonesia
ade_silvia@polsri.ac.id

Abstract. Flooding is a prevalent natural calamity in urban areas, particularly in tropical regions like Indonesia. Some places in Palembang are more likely to flood than others because of how people act and how nature acts. Human activities that contribute to flooding in Palembang include the construction of illicit structures on water channels, littering, and disregarding the priority of the surrounding water flow. In this study, the application of Machine Learning (ML) to flood mapping and detection is examined through the use of Geographic Information System (GIS) data. Sentinel-1 satellite images were preprocessed by the Sentinel Application Platform (SNAP) to remove thermal noise, calibrate the radiometric data, rectify the geometry, and filter out speckles. This enhanced the quality of spatial data. We employed machine learning algorithms, such as linear regression for precipitation forecasting and Convolutional Neural Networks (CNN) for spatial feature extraction, to enhance flood risk evaluation. The results suggested that regions like Ilir Timur II and Kemuning, which are particularly vulnerable to inundation due to their low elevation and proximity to river courses, are particularly susceptible. The integration of GIS and ML demonstrated considerable accuracy in predicting flood-prone regions and precipitation trends, producing precise and actionable flood risk maps. These findings provide critical information for policymaking, urban planning, and disaster management, improving preparedness and resilience in areas vulnerable to urban flooding.

Keywords: Flood Detecting, Geographic Information System (GIS), Machine Learning (ML)

1 Introduction

Floods are among the most prevalent natural disasters in urban environments, particularly in tropical places such as Indonesia. This occurrence results in considerable economic costs, including damage to infrastructure, property, and interruption of daily activities, while also adversely affecting social life and public

N. L. Husni et al. (eds.), *Proceedings of the 8th FIRST 2024 International Conference on Global Innovations (FIRST-ESCSI 2024)*, Advances in Engineering Research 261,
https://doi.org/10.2991/978-94-6463-678-9_36

health [1][2]. In Palembang, a prominent city in Indonesia, the heightened intensity of rainfall and inadequate spatial planning are the primary factors contributing to the elevated flood risk in the region [3]. Uncontrolled urbanization, lacking sufficient planning, has intensified this issue by diminishing water absorption zones through their transformation into residential or industrial developments [4][5]. Recent study indicates that factors including terrain, excessive rainfall, land use, and insufficient drainage infrastructure are primary contributors to elevated flood risk [6][7].

To mitigate the effects of flood disasters, Geographic Information System (GIS) technology has been extensively employed to delineate flood-prone regions through the analysis of spatial data, including elevation, land use, and drainage patterns. Studies indicate that GIS is proficient in mapping and analyzing flood susceptibility with considerable precision [3][8]. Nonetheless, GIS exhibits constraints in the analysis of intricate and dynamic data, particularly regarding the non-linear interrelationships among diverse components influencing floods. The amalgamation of GIS with Machine Learning (ML) technologies provides a more advanced answer to these restrictions. Machine learning can identify intricate patterns in spatial and temporal data that conventional GIS struggles to detect [9][10].

Various machine learning techniques have been employed in studies to enhance the precision of flood risk predictions. Random Forest (RF) is a widely utilized approach that effectively manages high-dimensional data and intricate patterns [11][12]. In [13] demonstrate that Random Forest (RF) can get an accuracy of 87% in delineating flood risk in urban environments by integrating elevation data, land use, and drainage patterns. An additional option is the Support Vector Machine (SVM), which is highly effective for datasets with skewed distributions.

Research has been conducted [14] to develop an early flood warning system that accurately predicts the depth and area of inundation by integrating hydrodynamic models, k-means clustering algorithms, and support vector machines (SVM) to detect typhoon flood events [15]. In flood detection, Convolution Neural Networks (CNNs) are utilized. This framework is one of the most classic and prevalent deep learning frameworks, and it is a multilayer neural network [16]. Classic machine learning models like SVM have successfully accomplished flood detection; however, the model's complexity noticeably increases as the size of the training dataset increases. SVM must also be adjusted in order to determine the best kernel function for training. Enhancing preparedness and resilience in urban areas susceptible to inundation can enhance disaster management.

Deep Learning provides enhanced analytical skills for geographical and temporal data [17]. Deep learning approaches significantly reduce the dimensionality of parameters in the neural network during the training phase, which leads to more dependable disaster risk maps. In deep learning, neural networks are among the most widely used models. This is due to the fact that urban watersheds differ greatly from rural, forest, and alpine watersheds in their hydrological functions and structural features, which are dominated by pavements, roads, buildings, and parking lots. However, there are typically no hydrometric stations in metropolitan areas, thus there is little hydrological data available. These regions are referred to as data-deficient areas.

In Palembang City, the amalgamation of GIS and ML facilitates the identification of flood-prone regions and enhances the precision of risk assessments to aid in spatial planning and disaster prevention. This methodology facilitates the creation of more comprehensive and predictive flood risk maps, offering a robust scientific foundation for decision-makers to enact data-driven mitigation strategies. The integration of GIS and machine learning algorithms, including Random Forest and Deep Learning, is anticipated to yield comprehensive and dependable solutions for future flood concerns.

2. Materials and Methodology

2.1 Study Area

Palembang serves as the capital of South Sumatra province, with an area of roughly 400.61 square kilometers, in Figure 1. Palembang is situated between 2°59' - 3°15' South Latitude and 104°45' - 104°59' East Longitude. It is bordered by Banyuasin Regency to the north, east, and west, and by Ogan Ilir Regency to the south. Palembang is renowned for the Musi River, which bifurcates the city into two sections: Seberang Ulu and Seberang Ilir. The city of Palembang is administratively divided into 16 districts: Alang-Alang Lebar, Bukit Kecil, Gandus, Ilir Barat I, Ilir Barat II, Ilir Timur I, Ilir Timur II, Kalidoni, Kemuning, Kertapati, Plaju, Sako, Seberang Ulu I, Seberang Ulu II, Sematang Borang, and Sukarami.

Fig 1. The study area's location

From 2019 to 2022, Palembang City has experienced numerous flood disasters. The flood disaster in Palembang City, for instance, occurred on (i) December 24–25, 2021; (ii) October 5–6, 2022; and (iii) December 8–9, 2022. The city was abruptly overrun by high water due to a combination of rising river levels and high rainfall levels. The flood caused damage to health facilities, places of worship, and many educational facilities. In addition, Palembang City and its surroundings are home to a number of slums, especially those close to river areas.

2.2 Data and Methods

The dataset comprises spatial and non-spatial information related to flood potential metrics. Subsequently, the data is subjected to a classification procedure to identify flood-prone locations and derive flood potential objects (data set). Image capture utilizes spatial data obtained from the Google Map API functionality. Segmenting the image to delineate sections of the city of Palembang that have been categorized based on flood susceptibility criteria. The dataset is derived from images processed using CNN, resulting in the separation of images and labels, with the partitioned data (training folder and testing folder) organized within the dataset folder. The Dataset is divided into two components: the Training and Validation Dataset, and the Testing Dataset. The training procedure initiates with the convolution operation utilizing CNN, followed by feature extraction to identify regions of interest that align with the characteristics of prospective flood-prone places. The model is constructed and prepared for utilization based on the acquired features. The initial testing procedure for the data entails pre-processing (normalization or data modification) to verify compatibility with the training data. Testing Procedure (testing data will be inputted into the model for evaluation). The model will endeavor to identify patterns or characteristics in the new data. The third stage involves identifying things in flood-prone areas, conducted subsequent to testing. The model will generate an output delineating a zone of interest based on objects in flood-prone areas. The model evaluation procedure occurs post-testing to assess the model's performance through accuracy, precision, recall, and confusion matrix metrics.

2.3 Data Preprocessing for Machine Learning Applications in Flood Detection and Mapping

This work employs Sentinel-1 picture data as the principal resource for the detection and mapping of urban floods. The preliminary phase of data processing is conducted utilizing the SNAP (Sentinel Application Platform) created by the European Space Agency (ESA). The objective of this method is to enhance the quality of geographic data prior to its utilization in machine learning models.

1. Apply to Orbit

The initial stage is Apply to Orbit, during which geometric correction is implemented to rectify the geographical positioning of the raw data. This procedure is crucial for ensuring the precision of the location required in machine learning analysis. The discrepancies in geographical positioning due to orbital fluctuations of the satellites are rectified using sophisticated algorithms supplied by SNAP. This phase generates photos with accurate geographic coordinates, establishing a dependable spatial database.

2. Radiometric Calibration

The subsequent phase is Radiometric Calibration, which transforms the raw pixel values in the image into values with physical significance. In the case of radar data, pixel values are calibrated to backscatter, facilitating the detection of surface features like water puddles or alterations in soil texture. The outcome of this calibration serves as a crucial input in feature extraction for machine learning algorithms.

3. Mitigation of Thermal Noise

The thermal noise removal procedure is conducted to eradicate thermal interferences that may compromise the quality of the radar image. The interferences originate from the heat radiation emitted by the satellite sensor and its surroundings. Cleaner data enhances the precision and reliability of flood detection interpretations.

4. Speckle Mitigation

The speckle filtering phase aims to diminish the speckle noise present in radar images, which arises from the coherent interference of radar waves. Speckle noise can conceal critical information in the image and compromise the precision of machine learning predictions. Employing speckle filters enhances image clarity and augments data quality for further processing.

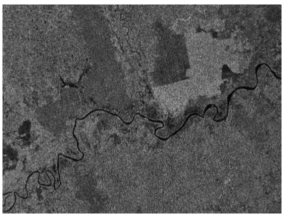
3 Result and Discussion

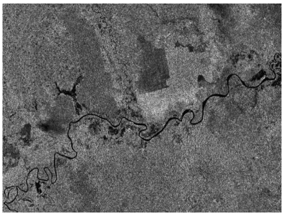
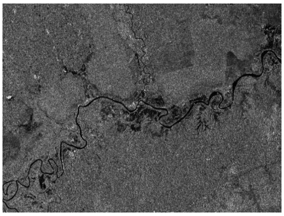
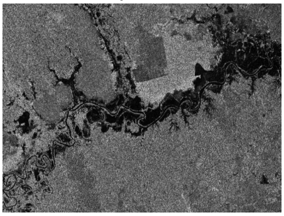
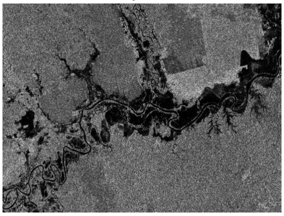
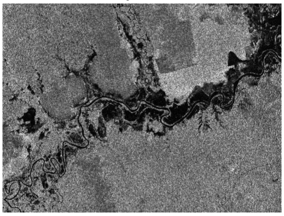
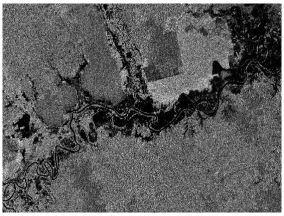
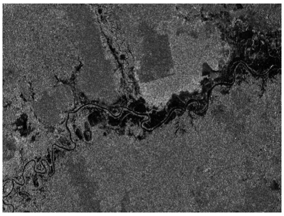
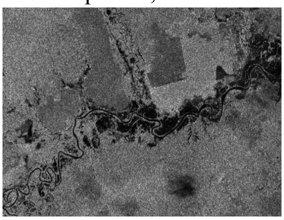
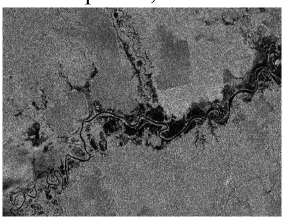
3.1 Integration with Machine Learning

The outcomes of data preprocessing utilizing SNAP are amalgamated with machine learning methods, including Random Forest, Support Vector Machine (SVM), or Deep Learning, to identify and delineate floods in metropolitan regions. Data that has undergone geometric correction, radiometric calibration, and noise reduction is utilized for training machine learning models. The images obtained from preprocessing yield precise and pertinent spatial data regarding significant features such as flood zones, ground surface elevation, or alterations in surface patterns.

The culmination of this research is a flood risk map generated by the amalgamation of Sentinel-1 spatial data and machine learning analysis. This map facilitates disaster mitigation and urban spatial planning.

Table 1. The Flood's Distribution Pattern through Low and High Precipiation in Palembang

Low Precipitation	High Precipitation
November 11, 2023 	November 23, 2023 
December 1, 2023	December 29, 2023

	
January 10, 2024 	January 22, 2024 
February 3, 2024 	February 15, 2024 
March 5, 2024 	March 22, 2024 
April 15, 2024 	April 27, 2024 

According to the table 1 above, the floods in Palembang from November 2023 until April 2024 has the distribution patterns of flooding that affected by precipitation, land utilization, and the state of drainage infrastructure. Geographic data and satellite imagery reveal that numerous regions encountered heightened flood intensity during that period, indicating the direct influence of climatic and topographical factors.

During November 2023 and January 2024, floods do not seem to exhibit a substantial increase, particularly during periods of low precipitation. Between February and April 2024, floods commenced in many places surrounding the city of Palembang, particularly in regions next to the Sekanak River.

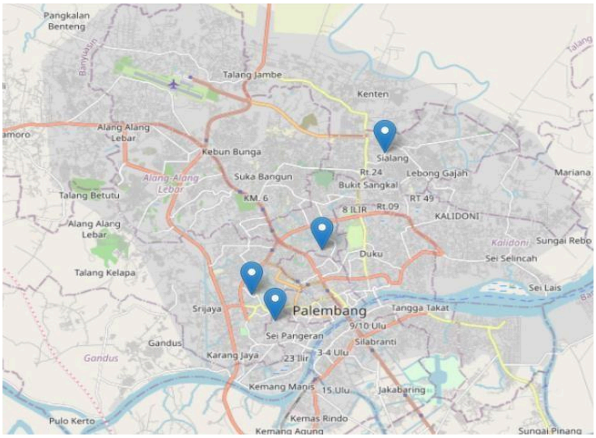


Fig 2. Flood Distribution Map in November 2023

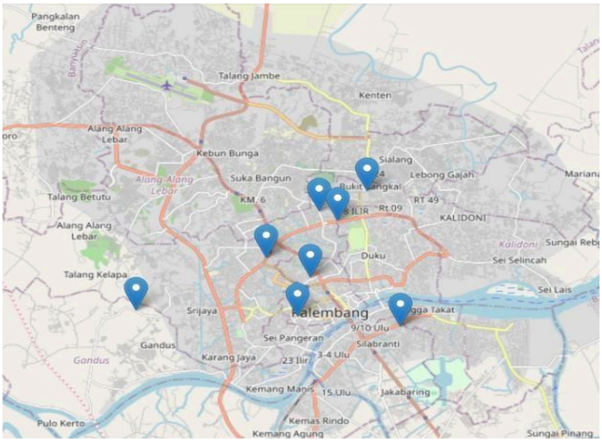


Fig 3. Flood Distribution Map in December 2023

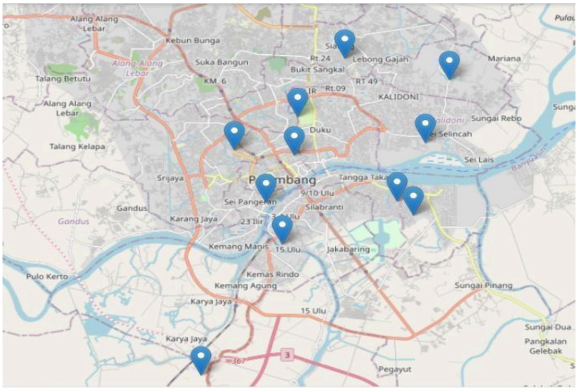


Fig 4. Flood Distribution Map in January 2024

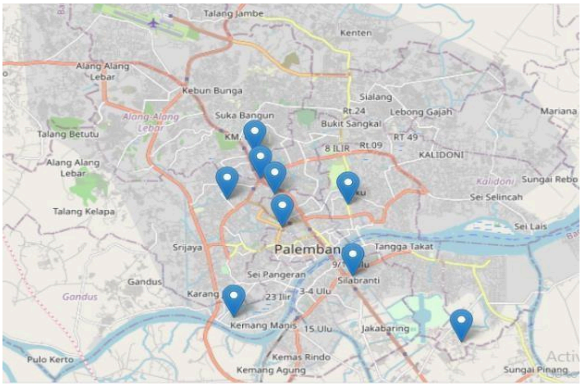


Fig 5. Flood Distribution Map in February 2024

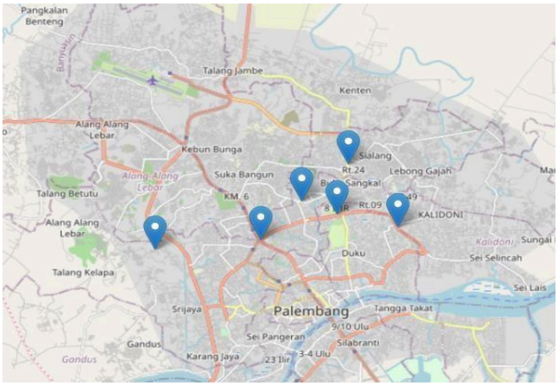


Fig 6. Flood Distribution Map in March 2024

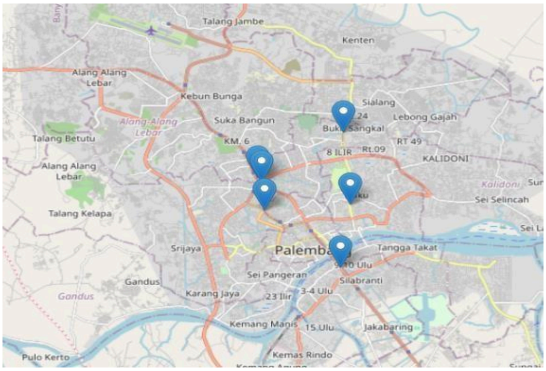


Fig 7. Flood Distribution Map in April 2024

3.2 Linear Regretion on Rainfall Prediction

The last phase involves employing a linear regression model for rainfall prediction to analyze the correlation between precipitation and the possible incidence of floods in Palembang. Linear regression is employed in the Machine Learning study to model and forecast rainfall levels utilizing the obtained historical data. This approach facilitates the detection of precipitation patterns and the assessment of regions likely to experience flooding based on the produced rainfall forecasts.

The linear regression approach in Machine Learning is employed to assess and forecast rainfall that may impact flood-prone regions in the city of Palembang. This linear regression model is developed from historical precipitation data and prior observations. Utilizing a dataset of weather factors, including rainfall intensity and temperature, the model may deliver a more precise estimation of forthcoming rainfall variations.

4 Conclusions

This study concludes that the application of Geographic Information System (GIS) in Machine Learning is highly effective for identifying and mapping flood-prone areas in Palembang. Geographic Information Systems (GIS) are crucial for visualizing flood-prone regions, particularly through the integration of geographic data and satellite imagery to pinpoint areas vulnerable to waterlogging. The investigation revealed that regions regularly inundated with water are typically situated along river courses and in residential areas characterized by low elevation. This signifies a notable pattern concerning the proliferation of floods linked to geographical conditions and elevated rainfall distribution, which precipitates waterlogging.

The interval from February 2024 to April 2024 is anticipated to exhibit exceptionally high rainfall levels, which may precipitate flooding, particularly in the Ilir Timur II and Kemuning areas. The Ilir Timur II District, situated along the Sekanak River, and the highly populated district of Kemuning exhibit a heightened susceptibility to flooding. Employing Machine Learning techniques, particularly linear regression and Random Forest, enables more precise predictions of flood potential, facilitating the identification of specific high-risk zones.

This research offers profound insights into the correlation between precipitation data, topographical features, and river flow in forecasting and delineating flood-prone regions. The amalgamation of GIS and Machine Learning facilitates the development of more precise flood risk maps, serving as a scientific foundation for devising enhanced catastrophe mitigation measures and improving data-driven spatial planning in the future. Consequently, this application possesses the capacity to optimize flood risk management and improve community preparedness for disasters in urban environments.

References

[1] M. S. Rahman *et al.*, "Climate Disasters and Subjective Well-Being among Urban and Rural Residents in Indonesia," pp. 1–14, 2022.

[2] M. Irfansyah, L. Id, M. Linkie, J. Ser, and H. Lee, "Tropical forest cover , oil palm plantations , and precipitation drive flooding events in Aceh , Indonesia , and hit the poorest people hardest," pp. 1–19, 2024, doi: 10.1371/journal.pone.0311759.

[3] M. Rendana *et al.*, "Flood risk and shelter suitability mapping using geospatial technique for sustainable urban flood management : a case study in Palembang city , South Sumatera , Indonesia," *Geol. Ecol. Landscapes*, vol. 00, no. 00, pp. 1–11, 2023, doi: 10.1080/24749508.2023.2205717.

[4] D. Santoso, A. Suroso, B. Kombaitan, and B. Setiawan, "Exploring the use of risk assessment approach for climate change adaptation in Indonesia: Case study of flood risk and adaptation assessment in the South Sumatra province," *Procedia Environ. Sci.*, vol. 17, pp. 372–381, 2013, doi: 10.1016/j.proenv.2013.02.050.

[5] E. Science, "Identifying environmental variables in potential flood hazard areas using machine learning approach at Musi Banyuasin Regency , South Sumatra Identifying environmental variables in potential flood hazard areas using machine learning approach at Musi Banyuasin Regency , South Sumatra," doi: 10.1088/1755-1315/1201/1/012037.

[6] N. W. Arnell and S. N. Gosling, "The impacts of climate change on river flood risk at the global scale," pp. 387–401, 2016, doi: 10.1007/s10584-014-1084-5.

[7] F. Management, "Urbanization and Increasing Flood Risk in the Northern Coast of Central Java — Indonesia : An Assessment towards Better Land Use Policy and," 2020.

[8] F. R. Management, "A Review on Applications of Remote Sensing and Geographic Information Systems (GIS) in Water Resources and Flood Risk Management," no. Table 1, pp. 1–11, 2018, doi: 10.3390/w10050608.

[9] M. Motta, M. D. C. Neto, and P. Sarmento, "A mixed approach for urban flood prediction using Machine Learning and GIS published by Elsevier :," no. February, 2021.

[10] N. Khoirunisa and C. Ku, "A GIS-Based Artificial Neural Network Model for

Flood Susceptibility Assessment," 2021.
[11] F. Technique, "Flood Risk Mapping by Remote Sensing Data and Random," 2021.
[12] L. Models, "Urban Flood Risk Assessment through the Integration of Natural and Human Resilience Based on Machine Learning Models," 2023.
[13] B. T. Pham, A. Jaafari, T. Nguyen-thoi, and T. Van Phong, "Ensemble machine learning models based on Reduced Error Pruning Tree for prediction of rainfall-induced landslides," 2021, doi: 10.1080/17538947.2020.1860145.
[14] M. Chang, I. Huang, C. Hsu, S. Wu, J. Lai, and G. Lin, "Long-Term Flooding Maps Forecasting System Using Series Machine Learning and Numerical Weather Prediction System," 2022.
[15] C. Castillo *et al.*, "Combining Human Computing and Machine Learning to Make Sense of Big (Aerial) Data for Disaster Response," vol. 00, no. 00, pp. 1–13, 2016, doi: 10.1089/big.2014.0064.
[16] N. G. Trong, P. N. Quang, N. Van Cuong, and H. A. Le, "Spatial Prediction of Fluvial Flood in High-Frequency Tropical Cyclone Area Using TensorFlow 1D-Convolution Neural Networks and Geospatial Data," 2023.
[17] D. D. Nguyen *et al.*, "Enhanced Landslide Spatial Prediction Using Hybrid Deep Learning Model and SHAP Analysis : A Case Study of the Tuyen Quang ‑ Ha Giang Expressway , Vietnam," *J. Indian Soc. Remote Sens.*, no. December, 2024, doi: 10.1007/s12524-024-02055-1.