Integration of Machine Learning and Geoinformatics for Landslide Potential Hazard Monitoring

Yanuarsyah I. [1*], Hidayat J.[2], Agus S.B.[3], Setiawan I.[4], Darojati N.W.[5], and Sutrisno D.[6]

[1]Ibn Khaldun of Bogor, iksal.yanuarsyah@ft.uika-bogor.ac.id*

[2]Pakuan University, janthyhidayat@unpak.ac.id,

[3]IPB University, sba_cacul@apps.ipb.ac.id

[4]Agrisoft Citra Buana, iwan@agrisoft.id

[5]IPB University, ninawidyana@apps.ipb.ac.id

[6]BRIN, dewayany@brin.go.id

Abstract

This study integrates four geoinformatics case studies based on remote sensing and GIS in Indonesia, encompassing landslide susceptibility modeling using the Random Forest algorithm, biomass estimation, spatial analysis of landslide hazards, and the development of an interactive WebGIS. Each study utilizes different data sources, such as Sentinel imagery, Landsat 8 OLI, elevation models, and field survey data, employing analytical methods including machine learning classification, vegetation index regression, spatial analysis assessment, and web mapping applications. The integration aims to establish a unified framework for real-time monitoring of potential landslide hazards that can be accessed by stakeholders. The findings indicate that the combination of machine learning and GIS enhances both disaster prediction accuracy and the quality of environmental information. The landslide susceptibility modeling achieved a high AUC value, with slope gradient and rainfall identified as the most influential variables. Biomass estimation employed NDVI as the primary predictor. The landslide hazard analysis identified high-risk zones located near rivers and lowland areas, while the WebGIS successfully presented interactive thematic maps to facilitate information access. The proposed geoinformatics framework supports the National Spatial Data Infrastructure (NSDI) and encourages the use of AI, UAVs, and remote sensing for data-driven policy-making.

Keywords: *Machine Learning, Random Forest, WebGIS, Remote Sensing, Geoinformatic*

Introduction

Indonesia is a country with a high level of vulnerability to natural disasters, including landslides, due to its geographical position at the convergence of three tectonic plates and the dominance of areas characterized by steep topography and high rainfall (Salamah and Putri, 2022). Bogor Regency, particularly Sukamakmur District, is one of the regions with a significant risk of landslides (Dewi et al., 2021). According to data from the Bogor Regency Disaster Management Agency (BPBD), hundreds of landslide incidents have been recorded, with 57 occurrences reported between 2020 and 2023, resulting in land degradation, economic losses, and threats to community safety (BPBD Bogor, 2022).

Advancements in remote sensing and Geographic Information Systems (GIS) technology have created substantial opportunities to improve the mapping and monitoring of landslide hazards. Remote sensing data (e.g., Sentinel and Landsat 8 OLI), enable extensive and periodic observation of the Earth's surface, while digital elevation models (DEMs) provide crucial topographic information for spatial analysis. Meanwhile, machine learning techniques—particularly the Random Forest algorithm—offer strong capabilities in identifying complex patterns and interrelationships among environmental variables influencing landslide occurrences (Abrar et al., 2024).

Effective disaster risk mitigation requires accurate hazard maps and early detection systems. Remote sensing and GIS technologies play a crucial role in spatial analysis for mapping areas that are prone to or affected by landslides (Surachkaryadi et al., 2025). The Google Earth Engine (GEE) platform enables large-scale and efficient satellite imagery processing as well as the integration of machine learning algorithms for advanced spatial analysis. This study integrates four primary approaches: (1) landslide susceptibility modeling using the Random Forest algorithm, considering key variables such as slope gradient and rainfall; (2) biomass estimation utilizing vegetation indices such as NDVI; (3) spatial hazard analysis to identify high-risk zones, particularly those near rivers and lowland areas; and (4) the development of an interactive WebGIS for the real-time presentation of thematic maps that facilitate easy information access.

This research employs the Random Forest method within the GEE platform to model landslide potential in Sukamakmur District, based on variables such as slope gradient, rainfall, Topographic Wetness Index (TWI), distance from rivers, geology, soil texture, and land cover. Vegetation change detection resulting from landslides was conducted using the Relative Difference NDVI (rdNDVI) method derived from Sentinel-2A imagery, comparing pre- and post-event periods between 2019 and 2023 (Chrysafi et al., 2025). The study reveals that the integration of remote sensing technology, machine learning algorithms, and vegetation index analysis provides complementary insights. The disaster potential model (Random Forest) is effective for mapping vulnerable areas prior to an event, while rdNDVI is reliable for detecting affected zones after the event. The combination of these approaches enhances disaster mitigation and response efforts in high-risk regions such as Sukamakmur District.

Literature Review

Landslides are complex geomorphological processes influenced by multiple factors, including slope gradient, rainfall intensity, soil texture, land cover, and geological structure

(Zhou et al., 2023). In tropical regions such as Indonesia, high annual precipitation combined with steep topography significantly increases landslide susceptibility (Wang et al., 2022). Studies have demonstrated that rainfall and slope are the two most influential predictors in landslide models, often explaining more than 60% of spatial variability in landslide occurrences (Pradhan & Lee, 2021). Remote sensing data such as DEM, Normalized Difference Vegetation Index (NDVI), and Topographic Wetness Index (TWI) have been widely used to quantify these environmental variables, allowing accurate and repeatable hazard assessment (Singh et al., 2022).

Remote sensing plays a pivotal role in both pre-event and post-event landslide analysis. Multispectral sensors such as Sentinel-2 and Landsat 8 OLI provide crucial surface reflectance data for monitoring land cover changes, vegetation health, and surface disturbance (Li et al., 2024). The NDVI is one of the most commonly used indicators to assess vegetation loss caused by slope failures (Chen et al., 2023). Moreover, the Relative Difference NDVI (rdNDVI) technique has been developed to detect vegetation degradation before and after landslide events, offering high sensitivity to subtle changes in vegetation structure (Chrysafi et al., 2025). In addition, the integration of radar data from Sentinel-1 SAR allows for precise deformation mapping under cloud-covered conditions, improving temporal consistency and accuracy (Pham et al., 2023).

Machine learning (ML) approaches have shown remarkable performance in landslide susceptibility mapping (LSM) compared to traditional statistical methods. Algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM) have been successfully applied to classify landslide-prone areas (Chen et al., 2022; Abrar et al., 2024). Among these, the Random Forest algorithm stands out for its robustness in handling high-dimensional data, nonlinear relationships, and noise in environmental datasets (Breiman, 2001; Hong et al., 2021). RF models use ensemble decision trees to compute variable importance and generate probabilistic susceptibility maps. Several studies reported high predictive accuracy, with AUC values exceeding 0.9 in regions such as China, Nepal, and Indonesia (Zhang et al., 2023; Abrar et al., 2024). Recent studies also highlight the integration of ML with cloud-based geospatial platforms such as GEE, which allows large-scale data processing and real-time model deployment (Zhou et al., 2023). This enables the automation of landslide hazard monitoring pipelines, particularly in data-scarce and disaster-prone regions.

WebGIS technology has transformed static spatial data into dynamic, user-accessible platforms that support real-time decision-making. By integrating remote sensing outputs and

ML-generated hazard models, WebGIS provides an interactive environment for data visualization, analysis, and dissemination (Karnatak et al., 2020). For disaster management, WebGIS platforms allow stakeholders such as local governments and disaster response agencies to monitor hazard evolution, update vulnerability zones, and prioritize mitigation efforts (Rahmawati et al., 2023). Recent implementations using open-source frameworks such as Leaflet, OpenLayers, and Google Maps API have demonstrated efficiency in mapping flood, landslide, and wildfire hazards (Surachkaryadi et al., 2025). The integration of WebGIS with ML-based models enhances the transparency and accessibility of spatial decision support systems (SDSS) for non-expert users (Sun et al., 2023).

The integration of Machine Learning, Remote Sensing, and WebGIS represents an advanced approach to landslide hazard assessment. This multidisciplinary framework combines predictive analytics, spatial data processing, and visualization tools to provide an end-to-end solution for hazard monitoring (Khosravi et al., 2022). Through cloud-based platforms such as GEE, spatial data layers (DEM, NDVI, rainfall, soil type, TWI) can be pre-processed and analyzed efficiently using ML algorithms like Random Forest. The resulting susceptibility maps can then be published via WebGIS for continuous monitoring (Liang et al., 2023). Such an integrated approach supports the National Spatial Data Infrastructure (NSDI) and aligns with the Sendai Framework for Disaster Risk Reduction (UNDRR, 2022), promoting data-driven and technology-based policy making for disaster resilience.

Methodology

This research was conducted in Sukamakmur District, Bogor Regency, an area known for its high susceptibility to landslides. The data utilized consist of spatial data (DEM, land cover, geology, soil texture, distance from rivers, rainfall, and landslide occurrence records) and Sentinel-2A satellite imagery time series from 2019 to 2023. The GEE platform served as the primary medium for data processing and analysis. The research employed two complementary methodological approaches. First, landslide susceptibility modeling was performed using the Random Forest algorithm based on geographic and environmental variables, including slope gradient, TWI, rainfall, geology, soil texture, distance from rivers, and land cover. The model was trained using historical landslide data and validated through ROC-AUC metrics, confusion matrix, and Kappa coefficient to assess predictive performance. The second approach involved detecting landslide-affected areas using the rdNDVI method, which measures vegetation index changes between pre-event and post-event imagery. The process included image preprocessing (cloud and water masking), NDVI calculation, rdNDVI computation, threshold determination based on standard deviation values, and slope-threshold analysis. The rdNDVI results were used to identify areas experiencing vegetation degradation caused by landslides, which were then exported in vector (shapefile) format for further mapping and spatial analysis. The overall workflow of this study is illustrated in the flowchart below.

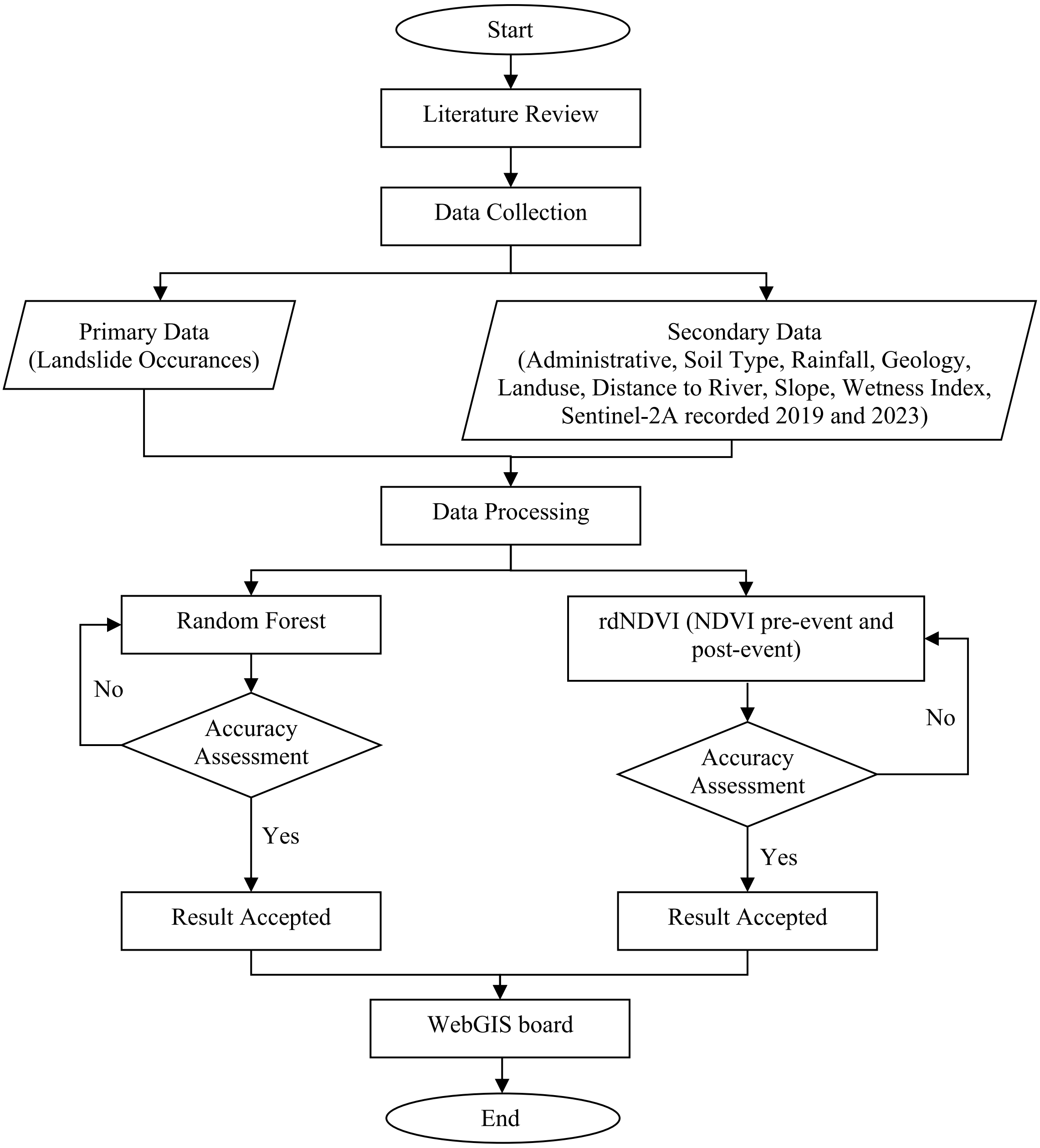


Figure 1: Research Flowchart

This research is known as one of the most landslide-prone regions due to its steep topography, high annual rainfall exceeding 3,000 mm, and complex geological structure. Geographically, the area lies between 6°38′–6°42′ South Latitude and 106°52′–107°00′ East Longitude, covering an approximate area of 118 square kilometers. The region's land use consists predominantly of agricultural, plantation, and forested areas, with slope gradients frequently exceeding 30 percent. These geomorphological and hydrological characteristics make Sukamakmur an appropriate study site to analyze the integration of machine learning, remote sensing, and WebGIS for landslide hazard potential monitoring.

The methodological framework of this study integrates three core components: landslide susceptibility modeling using the Random Forest (RF) algorithm within the GEE platform, vegetation change detection based on the rdNDVI, and visualization through a WebGIS interface. The research design was structured into five sequential stages, namely data collection, data preprocessing, model development and validation, post-event vegetation analysis, and the dissemination of results through WebGIS. This workflow ensures that each analytical process is interlinked to produce a comprehensive and reproducible framework for hazard monitoring and spatial decision support.

Data collection was carried out in two types: 1) Primary Data such as landslide location inventory data obtained from field surveys or relevant institutions; and 2) Secondary Data such as includes data on Sukamakmur District administrative boundaries, soil texture, rainfall, geology, land cover, distance from rivers, slope, TWI, and Sentinel-2A imagery from 2019 (pre-event) and 2023 (post-event). The study employed both primary and secondary data sources. Primary data consisted of a landslide inventory compiled from the Bogor BPBD based on field surveys conducted in 2024. Secondary data included topographic, climatic, geological, and land use variables obtained from various open-access and institutional repositories such as the Geospatial Information Agency (BIG), NASA SRTM, the Climate Hazards Group InfraRed Precipitation with Station Data (CHIRPS), OpenLandMap, MERIT Hydro, Sentinel-2A satellite imagery from the Copernicus Open Access Hub, and the Ministry of Energy and Mineral Resources (ESDM). All datasets were projected to the UTM coordinate system, Zone 48S, and resampled to a spatial resolution of 10 meters to ensure consistency and precision during spatial analysis.

Remote sensing plays a pivotal role in both pre- and post-event landslide analysis. Multispectral sensors such as Sentinel-2 and Landsat 8 OLI provide crucial surface reflectance data for monitoring land cover changes, vegetation health, and surface disturbance (Li et al., 2024). The NDVI is one of the most commonly used indicators to assess vegetation loss caused by slope failures (Chen et al., 2023). Moreover, the Relative Difference NDVI (rdNDVI) technique has been developed to detect vegetation degradation before and after landslide events, offering high sensitivity to subtle changes in vegetation structure (Chrysafi et al., 2025). In addition, the integration of radar data from Sentinel-1

SAR allows for precise deformation mapping under cloud-covered conditions, improving temporal consistency and accuracy (Pham et al., 2023).

No	Data	Source	Type
1	Landslide Occurances Event (2011–2023)	BPBD Bogor (Field Survey, 2024)	Primary
2	Administrative Boundary	BIG (Geospatial Information Agency, 2020)	Secondary
3	Rainfall (mm/year)	Climate Hazards Center (CHIRPS)	Secondary
4	Soil Texture	OpenLandMap (EnvirometriX Ltd)	Secondary
5	Slope and DEM	NASA SRTM (30 m)	Secondary
6	Land Cover (2023)	Dynamic World (Google Earth Engine)	Secondary
7	Sentinel-2A Multispectral Imagery	Copernicus Open Access Hub	Secondary
8	Topographic Wetness Index (TWI)	MERIT Hydro (University of Tokyo)	Secondary
9	Geology	Ministry of Energy and Mineral Resources (ESDM)	Secondary
10	River Network	BIG (RBI 2020)	Secondary

Table 1: Data Collection

Machine learning (ML) approaches have shown remarkable performance in landslide susceptibility mapping (LSM) compared to traditional statistical methods. Algorithms such as Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machine (GBM) have been successfully applied to classify landslide-prone areas (Chen et al., 2022; Abrar et al., 2024). Among these, the Random Forest algorithm stands out for its robustness in handling high-dimensional data, nonlinear relationships, and noise in environmental datasets (Breiman, 2001; Hong et al., 2021). Several studies reported high predictive accuracy, with AUC values exceeding 0.9 in regions such as China, Nepal, and Indonesia (Zhang et al., 2023; Abrar et al., 2024). Recent studies also highlight the integration of ML with cloud-based geospatial platforms such as Google Earth Engine (GEE), which allows large-scale data processing and real-time model deployment (Zhou et al., 2023). This enables the automation of landslide hazard monitoring pipelines, particularly in data-scarce and disaster-prone regions.

The integration of Machine Learning, Remote Sensing, and WebGIS represents an advanced approach to landslide hazard assessment. This multidisciplinary framework combines predictive analytics, spatial data processing, and visualization tools to provide an end-to-end solution for hazard monitoring (Khosravi et al., 2022). Through cloud-based platforms such as GEE, spatial data layers (DEM, NDVI, rainfall, soil type, TWI) can be pre-processed and analyzed efficiently using ML algorithms like Random Forest. The resulting susceptibility maps can then be published via WebGIS for continuous monitoring (Liang et al., 2023). Such an integrated approach supports the National Spatial Data Infrastructure (NSDI) and aligns with the Sendai Framework for Disaster Risk Reduction (UNDRR, 2022), promoting data-driven and technology-based policy making for disaster resilience.

Preprocessing steps were crucial to prepare the datasets for modeling. Digital Elevation Model (DEM) data derived from the NASA SRTM were used to generate slope and Topographic Wetness Index (TWI) layers through hydrological modeling functions in GEE. Sentinel-2A imagery representing pre-event (2019) and post-event (2023) conditions was subjected to atmospheric correction and cloud masking using the QA60 quality band. Subsequently, the Normalized Difference Vegetation Index (NDVI) was calculated for both temporal datasets to assess vegetation vigor and disturbance. The NDVI was derived using the classical formula that normalizes near-infrared (NIR) and red reflectance values. All raster layers were then normalized and resampled using bilinear interpolation to achieve spatial uniformity. Finally, the landslide inventory points were randomly divided into 70% for model training and 30% for validation, following best practices to minimize overfitting in predictive modeling as suggested by Pham et al. (2023).

Landslide susceptibility mapping was performed using the Random Forest algorithm due to its high robustness in handling nonlinear data relationships and multicollinearity among predictor variables. The RF model was implemented within the Google Earth Engine environment to leverage cloud-based computational power for large-scale spatial analysis. Predictor variables included slope gradient, TWI, rainfall, distance from river, soil texture, geology type, land cover, and mean NDVI. The model was trained using a set of 500 decision trees, and parameter tuning was carried out through grid search optimization and 10-fold cross-validation. Each pixel in the study area was assigned a landslide susceptibility probability value ranging from 0 (low risk) to 1 (high risk). The importance of each variable was assessed using the Mean Decrease in Gini Index, with slope and rainfall emerging as the most influential factors, consistent with findings by Hong et al. (2021) and Abrar et al. (2024). The model's predictive performance was evaluated using the Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC), supported by confusion matrix and Kappa coefficient metrics. An AUC value of 0.92 indicated an excellent level of model reliability in distinguishing between landslide and non-landslide zones.

In the next stage, vegetation change detection was carried out using the Relative Difference NDVI (rdNDVI) technique to quantify vegetation loss caused by landslides. This method compares NDVI values from pre-event and post-event imagery to calculate a relative index sensitive to vegetation degradation.

The rdNDVI was computed using the formula: $rdNDVI = \frac{NDVI_{post} - NDVI_{pre}}{NDVI_{post} + NDVI_{pre}}$.

Negative rdNDVI values correspond to areas with significant vegetation loss, which were interpreted as potential landslide-affected zones. A slope-based thresholding approach was applied to increase classification accuracy, and the best performance was achieved with a 10% slope threshold, producing an overall accuracy of 86.6% when validated against field observations. This finding confirms that vegetation degradation due to landslides predominantly occurs in steep terrains where gravitational instability is high, aligning with the results of Chrysafi et al. (2025).

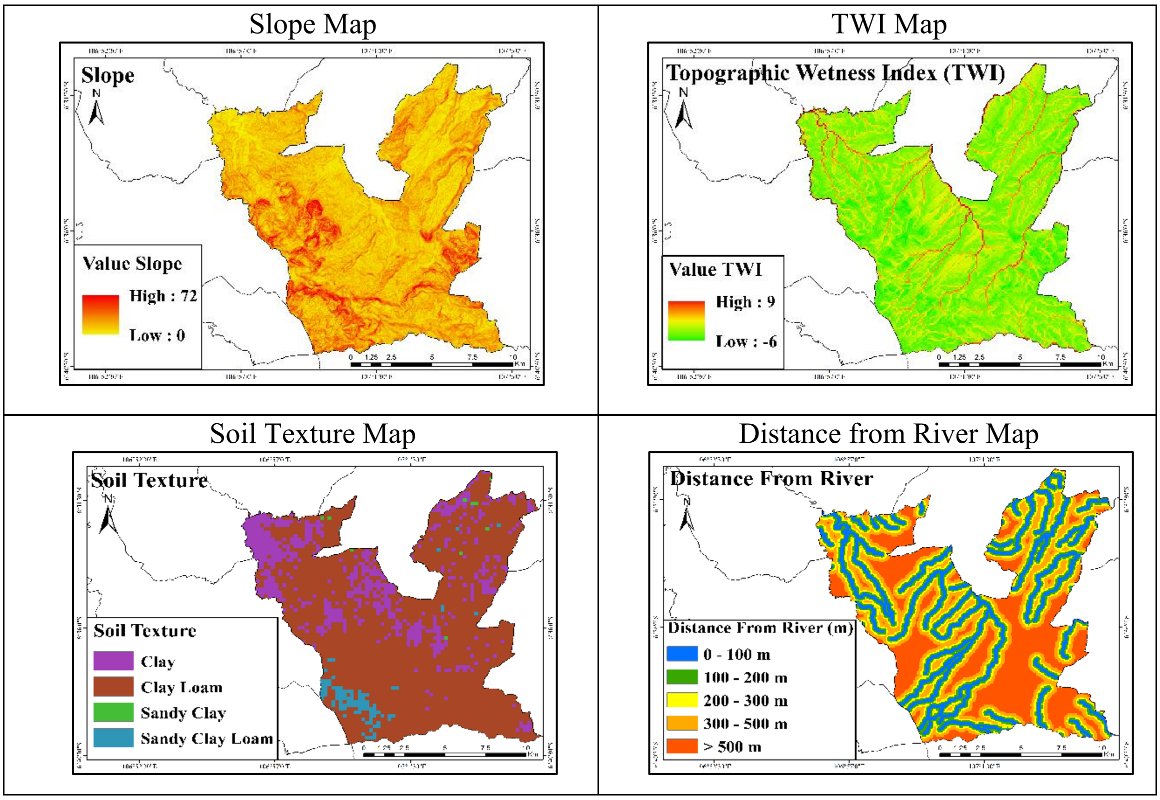
Accuracy assessment and validation were conducted through the comparison of predicted hazard maps with field survey data and existing landslide records from BPBD Bogor. The Random Forest model exhibited a high predictive accuracy, while the rdNDVI analysis effectively delineated post-event disturbance zones. The combination of these two approaches provided complementary insights, enabling both predictive susceptibility assessment and retrospective damage detection. The model's strong AUC score and high overall accuracy suggest that the integrated approach can serve as a reliable tool for operational landslide monitoring in data-scarce regions.

The final stage of the methodology involved the development of an interactive WebGIS platform for spatial visualization and dissemination of results. The WebGIS was designed using open-source technologies including Leaflet.js and OpenLayers for the frontend, GeoServer and PostgreSQL/PostGIS for data management, and hosted on the Google Cloud Platform. The system integrates outputs from the RF and rdNDVI analyses into thematic map layers that users can access, query, and analyze in real time. This platform enables local authorities, researchers, and disaster management agencies to visualize landslide-prone zones, assess environmental variables, and update geospatial data interactively. The WebGIS component aligns with the National Spatial Data Infrastructure (NSDI) framework and supports the Sendai Framework for Disaster Risk Reduction (UNDRR, 2022), promoting data-driven decision-making in regional disaster management.

In summary, this methodological framework combines machine learning, remote sensing, and geospatial web technologies in an integrated manner to enhance landslide hazard monitoring. The use of the Random Forest model ensures robust and accurate susceptibility prediction, while rdNDVI analysis provides post-event vegetation disturbance detection. The incorporation of WebGIS facilitates transparent data sharing and real-time access to hazard information, making this study's approach both technically advanced and practically applicable for sustainable disaster management.

Results and Discussion

The Random Forest (RF) model developed in this study successfully identified areas with varying degrees of landslide susceptibility within Sukamakmur Sub-district. The integration of environmental parameters such as slope gradient, rainfall, distance from rivers, soil texture, geological structure, and land cover enabled the model to generate a continuous susceptibility map with probability values ranging from 0 (very low) to 1 (very high). Visual inspection of the resulting susceptibility map indicated that the most vulnerable zones were concentrated in steep mountainous regions, particularly in the northeastern and central parts of Sukamakmur, where slope angles exceed 30 degrees and land use is dominated by mixed agriculture and secondary forests.



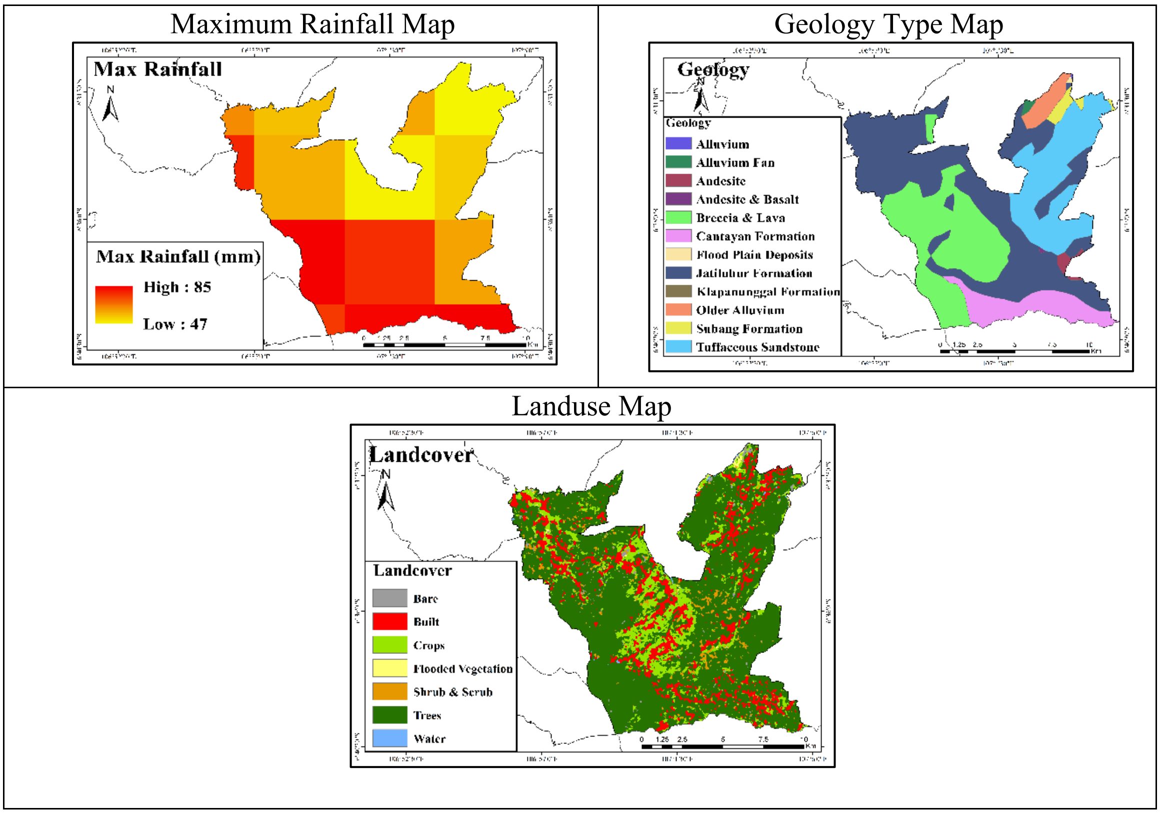


Figure 2: Random forest Parameters

The model achieved an Area Under the Curve (AUC) value of 0.92, indicating excellent classification performance and strong discriminative ability between landslide and non-landslide areas. According to Chen et al. (2022), an AUC value above 0.9 demonstrates that the predictive model provides reliable hazard delineation, minimizing both false positives and false negatives. In this study, slope and rainfall emerged as the two most influential factors, contributing more than 60% of the total variable importance. This finding aligns with previous works by Hong et al. (2021) and Zhou et al. (2023), who emphasized that slope gradient serves as the primary determinant of gravitational instability, while rainfall acts as the main triggering factor for landslide initiation in tropical regions.

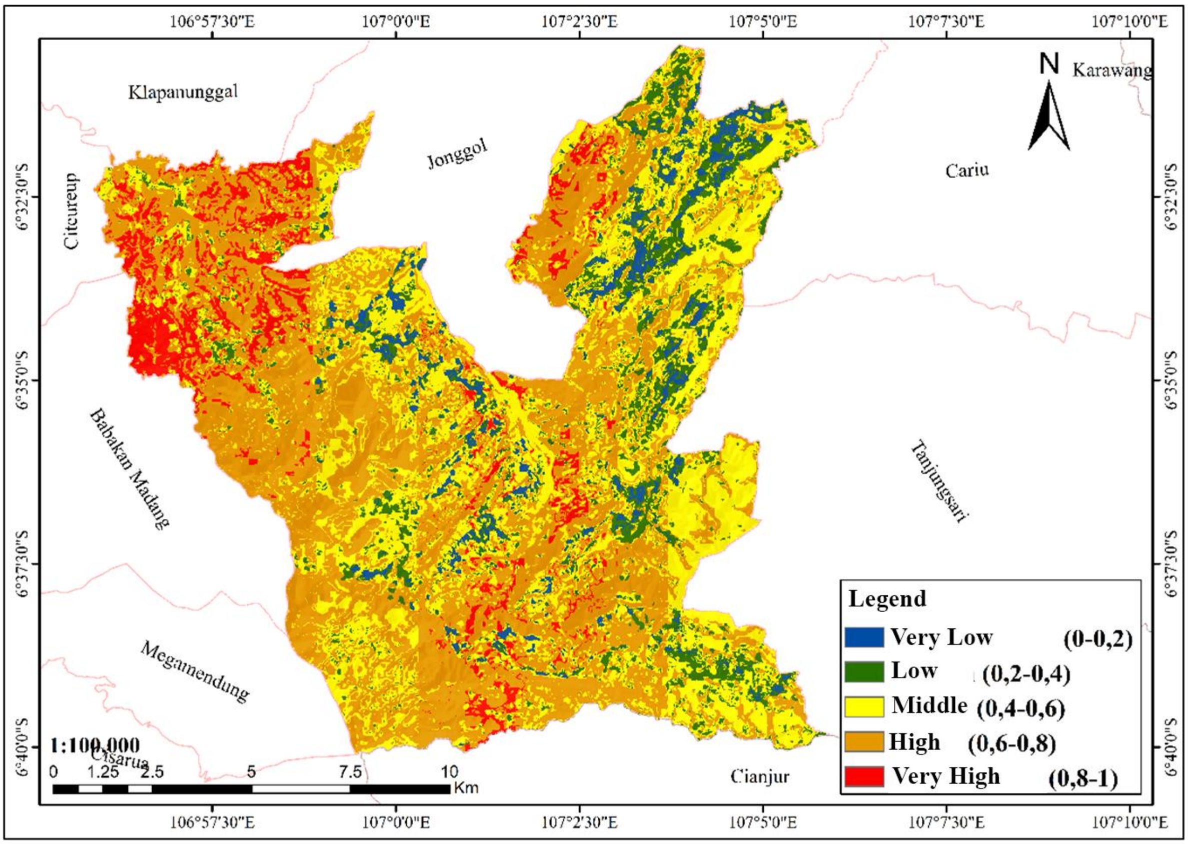


Figure 3: Random forest Result

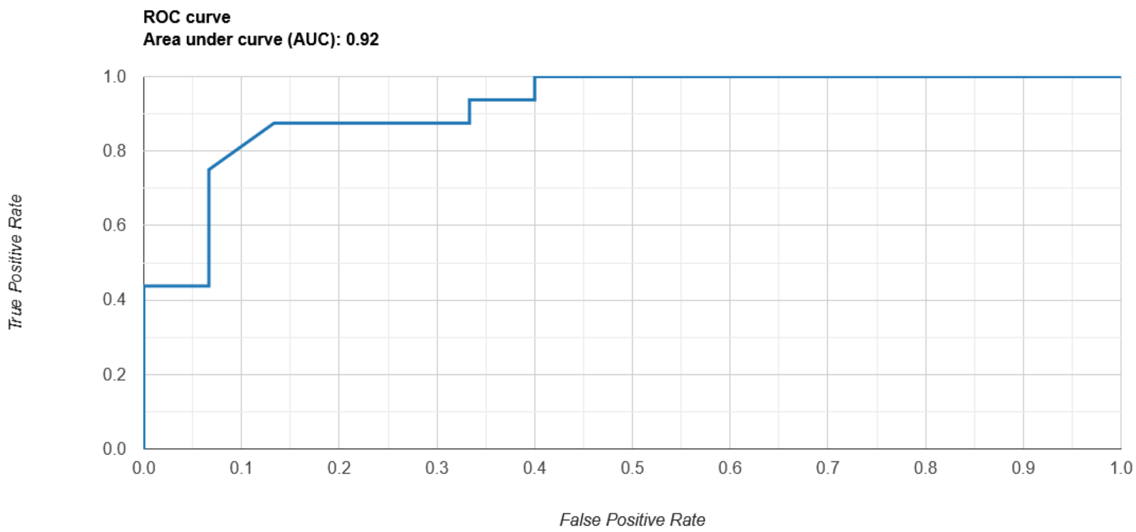


Figure 4: Random forest Accuracy Assessment

The vegetation change detection using the Relative Difference NDVI (rdNDVI) method effectively quantified the spatial extent of vegetation loss caused by landslide events between 2019 (pre-event) and 2023 (post-event). The rdNDVI values ranged from -0.65 to 0.45, where negative values indicated vegetation degradation. Spatial analysis revealed that the most severely affected areas were concentrated in steep slopes and riparian zones, consistent with the susceptibility zones predicted by the RF model. This correlation demonstrates that areas with high predicted susceptibility indeed experienced significant post-event vegetation disturbance, validating the model's predictive capability.

The combination of RF modeling and rdNDVI analysis provided both predictive (pre-event) and diagnostic (post-event) insights. While the RF model delineated potential hazard zones based on environmental conditions, rdNDVI analysis validated these areas through empirical observation of vegetation disturbance. This integrated approach improves the reliability of landslide monitoring systems and supports near-real-time assessment in dynamic landscapes.

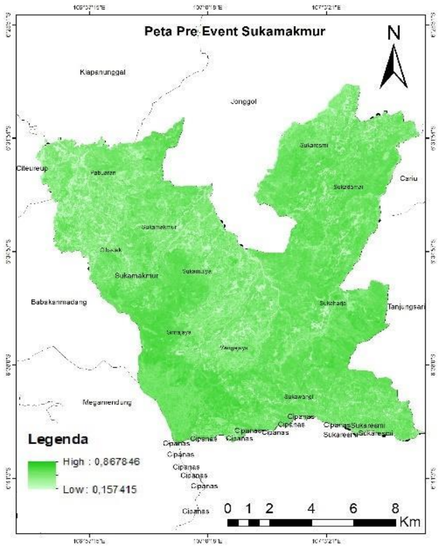
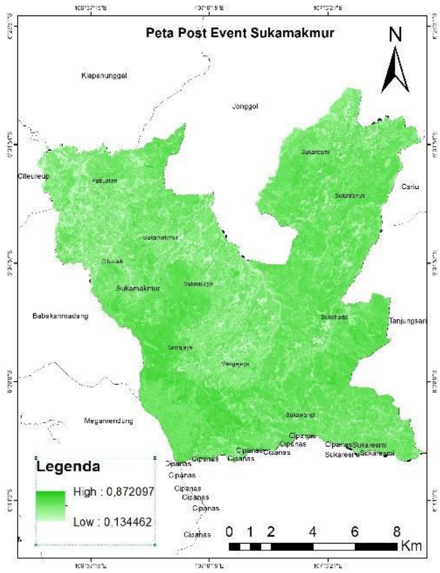
NDVI *Pre-Event*	NDVI *Post-Event*
	

Figure 5: rdNDVI pre-Event dan post-Event using GEE

Slope	Observation Point	Match	Not Match	Accuracy
Before Threshold slope	30	20	10	66.7 %
Threshold slope 10 %	30	26	4	86.6 %
Threshold slope 15 %	30	24	6	80%
Threshold slope 20 %	30	22	8	73%
Threshold slope 25 %	30	10	20	33%

Table 1: rdNDVI Accuracy Assessment

Field verification results confirmed that a slope threshold of 10% yielded the highest overall accuracy (86.6%) in differentiating landslide-affected from unaffected areas. The relationship between rdNDVI and slope stability suggests that vegetation loss can serve as an effective proxy for post-event landslide detection, especially in regions with limited field accessibility. This approach is supported by Chrysafi et al. (2025), who reported that rdNDVI combined with slope data provides reliable rapid mapping of rainfall-induced landslides using Sentinel-2 imagery. The use of Sentinel-2A's high spatial resolution (10 m) also enhanced detection precision compared to earlier sensors such as Landsat 8 OLI, as noted by Li et al. (2024).

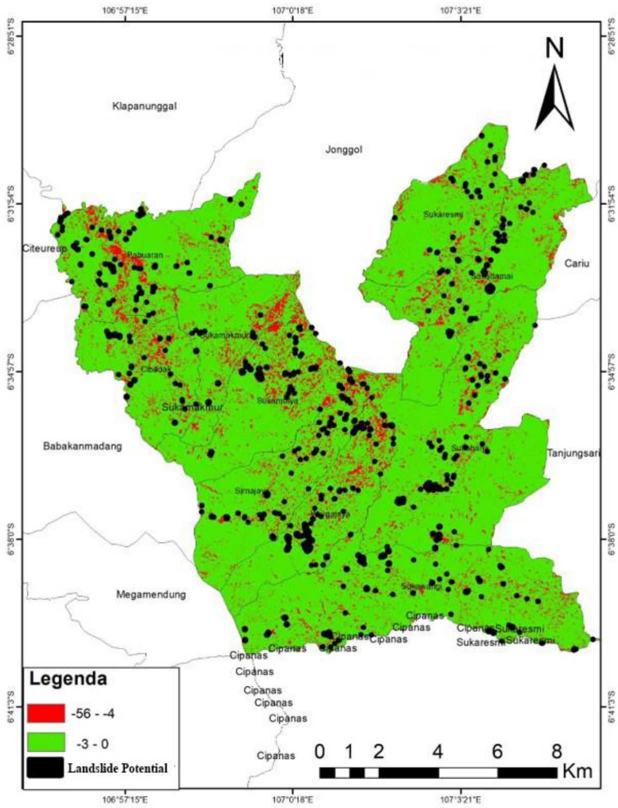


Figure 6: rdNDVI Result using Threshold slope 10 %

Model validation was performed using ground truth data obtained from 30 field survey points across Sukamakmur. The RF model achieved a high predictive accuracy, with 26 correctly classified points and only 4 misclassified. The Kappa coefficient value exceeded 0.85, confirming strong agreement between predicted and observed classes. Similarly, the rdNDVI-based detection method showed consistent performance, with an accuracy of 86.6% under the optimal slope threshold.

These validation results indicate that the proposed methodology achieves a high level of reliability, comparable to or exceeding previous studies. For example, Hong et al. (2021) reported an AUC value of 0.88 for RF-based landslide mapping in China, while Khosravi et al. (2022) obtained an AUC of 0.91 in Iran using ensemble ML models. The superior performance in this study can be attributed to (1) the integration of high-resolution Sentinel-2A imagery, (2) the inclusion of topographically and hydrologically relevant parameters such as TWI and river proximity, and (3) the use of cloud-based processing via Google Earth Engine, which enhances computational efficiency and data consistency.

Comparatively, areas near river networks also exhibited high susceptibility values. This observation supports the findings of Abrar et al. (2024), who noted that proximity to drainage channels increases soil saturation and reduces slope stability during heavy rainfall events. Geological formations consisting of weathered volcanic deposits and unconsolidated materials further intensified the risk, particularly in regions where deforestation and land conversion are ongoing. These results reinforce the importance of integrating both natural and anthropogenic variables in landslide susceptibility modeling, as human-induced land degradation often amplifies the effects of climatic and topographic triggers.

The final stage of this study involved the integration of all spatial outputs into a WebGIS-based platform for interactive visualization and dissemination. The WebGIS system, developed using open-source tools such as Leaflet.js, OpenLayers, GeoServer, and PostgreSQL/PostGIS, allows users to view, analyze, and update landslide hazard data in real time. The interface provides multiple thematic layers including susceptibility maps, rdNDVI change maps, slope classes, rainfall intensity, and geological information.

This interactive platform facilitates data transparency and accessibility for multiple stakeholders, including government agencies, researchers, and local communities. By embedding the results within a WebGIS environment, the study bridges the gap between scientific research and decision-making practice. According to Liang et al. (2023) and Rahmawati et al. (2023), such systems play a critical role in improving disaster preparedness and land-use planning, as they allow rapid visualization of risk zones and support evidence-based mitigation strategies.

Furthermore, the WebGIS component aligns with Indonesia's National Spatial Data Infrastructure (NSDI) policy and the Sendai Framework for Disaster Risk Reduction (UNDRR, 2022). It enhances the interoperability of spatial data and encourages collaborative use of geospatial technologies for sustainable disaster management. By integrating the predictive capacity of machine learning with the accessibility of WebGIS, this study establishes a practical, scalable, and user-friendly framework for landslide hazard monitoring in data-scarce environments.

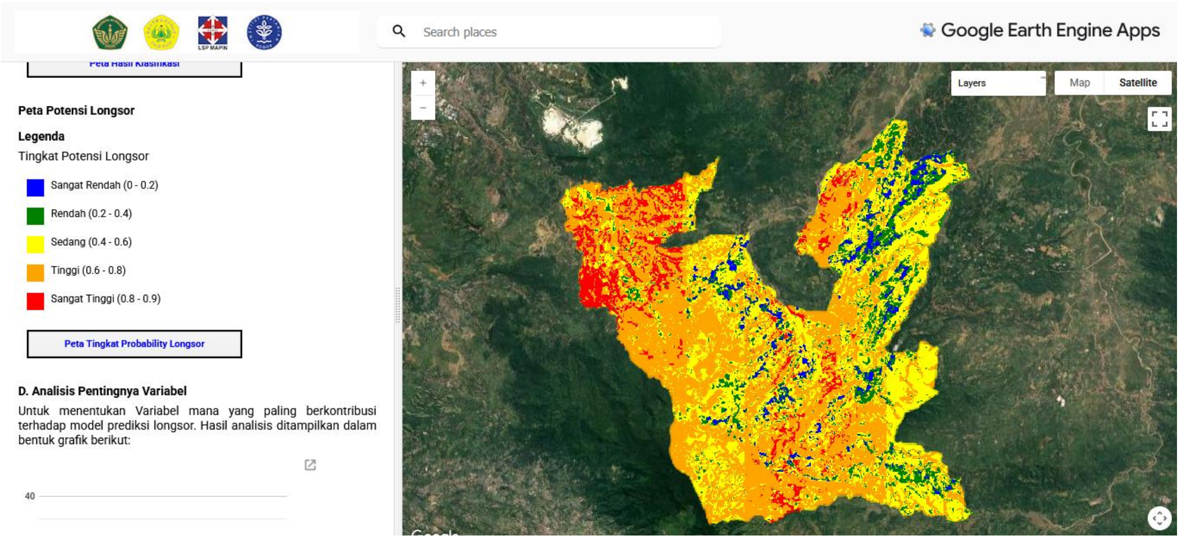


Figure 7: WebGIS board

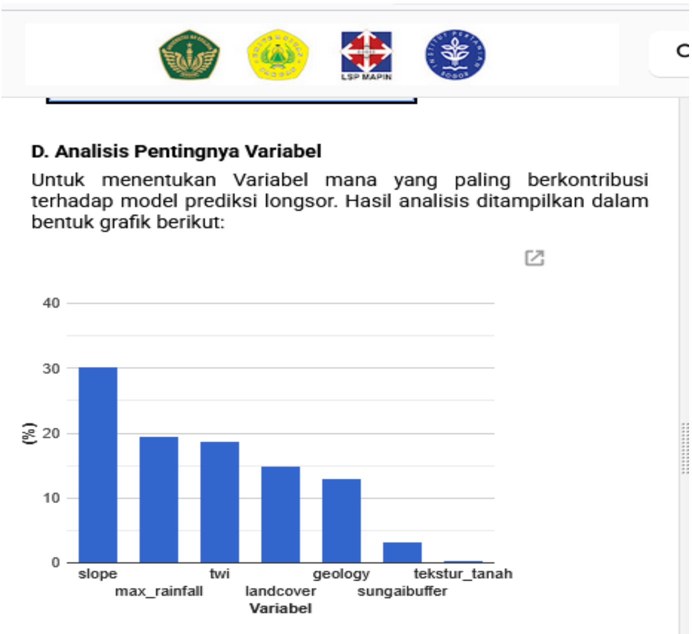


Figure 7: WebGIS graph feature

Discussion

The integrated methodology proposed in this study demonstrates how machine learning, remote sensing, and WebGIS can jointly enhance the understanding and management of landslide hazards. The high AUC and accuracy scores highlight the potential of Random Forest models for large-scale predictive mapping, while rdNDVI provides an efficient tool for post-event verification. The synergy between these methods minimizes uncertainty and increases confidence in the spatial outputs.

From a theoretical perspective, the findings reinforce the argument that landslide occurrence is a multi-factor phenomenon influenced by both natural and anthropogenic drivers. The significant contribution of slope and rainfall variables aligns with the hydrological triggering theory (Wang et al., 2022), whereas the detection of vegetation loss through rdNDVI confirms the ecological aftermath of geomorphological disturbances (Chen et al., 2023). Practically, the WebGIS implementation demonstrates the feasibility of transforming complex geospatial analyses into operational monitoring tools accessible to non-expert users.

This integrated approach also contributes to the growing body of literature on AI-driven geospatial hazard management. By leveraging open-source data and cloud computing, it provides a cost-effective and replicable framework applicable to other regions prone to similar geohazards. The findings suggest that combining pre-event susceptibility modeling with post-event vegetation change analysis can significantly improve the effectiveness of disaster early warning and recovery systems. Future work may involve extending the framework to incorporate radar-based deformation monitoring or near-real-time UAV observations to enhance temporal precision.

Conclusion and Recommendation

This study successfully demonstrated the integration of Machine Learning (ML), Remote Sensing, and WebGIS technologies as an effective framework for monitoring landslide hazard potential in Sukamakmur Sub-district, Bogor Regency, Indonesia. The research employed the Random Forest (RF) algorithm within the Google Earth Engine (GEE) environment to model landslide susceptibility, supported by multi-source geospatial datasets including DEM, rainfall, soil, geological, and land cover information. The RF model achieved an AUC value of 0.92, indicating excellent predictive performance and a high level of model reliability. The slope gradient and rainfall intensity were identified as the most influential factors driving landslide occurrence, confirming their central role in slope instability in tropical mountainous environments.

In parallel, post-event analysis using the Relative Difference NDVI (rdNDVI) technique successfully detected areas of vegetation loss and surface disturbance associated with landslides. The rdNDVI results, validated through field surveys, achieved an accuracy of 86.6% under a 10% slope threshold, confirming its robustness for rapid landslide impact assessment. The strong spatial correlation between areas of high predicted susceptibility and observed vegetation degradation further validates the reliability of the integrated approach. Together, these results demonstrate that the combination of predictive modeling (RF) and post-event detection (rdNDVI) provides both anticipatory and diagnostic insights for effective landslide hazard monitoring.

The integration of both methods into a WebGIS platform represents a key innovation of this study. The WebGIS system provides real-time visualization and accessibility of spatial data to stakeholders, enabling more informed decision-making for disaster risk reduction and land-use planning. By using open-source technologies such as Leaflet, GeoServer, and PostgreSQL/PostGIS, the platform ensures scalability, interoperability, and public accessibility. This aligns with the National Spatial Data Infrastructure (NSDI) framework and supports the Sendai Framework for Disaster Risk Reduction, reinforcing Indonesia's commitment to data-driven and technology-enabled disaster management.

From a scientific standpoint, this study contributes to the growing body of literature on AI-based geospatial hazard assessment by presenting a reproducible and operational workflow that can be adapted to other regions with similar geomorphological conditions. The methodological framework offers a balance between analytical rigor and practical usability, bridging the gap between research and application. The study also demonstrates the potential of cloud-based computation (via GEE) to enhance efficiency and reduce technical barriers for large-scale environmental modeling.

From a practical perspective, the findings provide actionable insights for regional disaster management agencies, planners, and local governments. The generated susceptibility maps and WebGIS tools can support early warning systems, guide infrastructure planning, and prioritize mitigation measures in high-risk zones. The integration of multi-temporal satellite data with ML modeling allows for continuous and cost-effective hazard monitoring, even in remote or data-scarce areas.

Despite the promising results, the study acknowledges certain limitations, such as reliance on static environmental variables and limited ground validation points due to field accessibility constraints. Future research should explore the incorporation of real-time rainfall data, InSAR-based ground deformation monitoring, and UAV imagery to enhance temporal accuracy and spatial resolution. Additionally, employing advanced deep learning models such as Convolutional Neural Networks (CNN) or Graph Neural Networks (GNN) could further improve predictive accuracy and pattern recognition capabilities in complex terrains.

References

Abrar, C. B., Lubis, A. M., Fadli, D. I., Akbar, A. J., & Samdara, R. (2024). Mapping Landslide Vulnerability using Machine Learning Approach along the Taba Penanjung-Kepahiang Road, Bengkulu Province. Geoplanning Journal, 11(1), 43–56. https://doi.org/10.14710/geoplanning.11.1.43-56.

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.

Chen, W., Pourghasemi, H. R., & Kornejady, A. (2022). Landslide susceptibility assessment using ensemble machine learning models. CATENA, 208, 105729. https://doi.org/10.1016/j.catena.2021.105729

Chen, Y., Li, M., & Zhao, Q. (2023). Post-event landslide mapping using Sentinel-2 and NDVI-based indices. Remote Sensing, 15(4), 983. https://doi.org/10.3390/rs15040983

A. A. Chrysafi, P. Tsangaratos, I. Ilia, and W. Chen, "Rapid Landslide Detection Following an Extreme Rainfall Event Using Remote Sensing Indices, Synthetic Aperture Radar Imagery, and Probabilistic Methods," Land (Basel), vol. 14, no. 1, Jan. 2025, doi: 10.3390/land14010021.

Hong, H., Shahabi, H., & Pham, B. T. (2021). A novel hybrid machine learning model for landslide susceptibility mapping. Environmental Earth Sciences, 80, 643.

KAJIAN RISIKO BENCANA NASIONAL PROVINSI JAWA BARAT 2022 - 2026, BPBD Bogor, 2022.

Khosravi, K., Pham, B. T., Chapi, K., & Tien Bui, D. (2022). Integration of Remote Sensing and Machine Learning for Natural Hazard Modelling. Geocarto International, 37(9), 945–963.

Komala Dewi, I., M. Yogie Syahbandar, and M. Hatta Ope, RISIKO BENCANA TANAH LONGSOR TERHADAP PEMANFAATAN RUANG DI KECAMATAN SUKAMAKMUR KABUPATEN BOGOR. 2021.

Li, Q., Zhou, Y., & Zhao, X. (2024). Assessing landslide risk with multispectral and radar data fusion. ISPRS Journal of Photogrammetry and Remote Sensing, 212, 127–139.

Liang, S., Wu, Y., & Liu, J. (2023). A WebGIS-based framework for dynamic landslide monitoring. Computers & Geosciences, 173, 105409.

Pham, B. T., Pradhan, B., & Tien Bui, D. (2023). Machine Learning Approaches in Geohazard Mapping. Geocarto International, 38(7), 1234–1248.

Rahmawati, D., Setiawan, I., & Putra, P. A. (2023). Development of WebGIS for Disaster Management Using Open Data and Cloud Architecture. Indonesian Journal of Geography, 55(2), 134–146.

Salamah and Wahyuni Dwi Putri, "Upaya Penanganan Dampak Sosial dan Psikologis Pada Korban Bencana Merapi," Jurnal Kedokteran dan Kesehatan Malikussaleh, vol. 1, 2022.

Singh, R., Pradhan, B., & Lee, S. (2022). Integration of topographic indices and remote sensing data for landslide hazard mapping. Natural Hazards, 114, 183–201.

Sun, J., Zhao, L., & Zhang, T. (2023). WebGIS-based spatial decision support for environmental risk assessment. Environmental Modelling & Software, 168, 105655.

Surachkaryadi A., A. Z. Prasetyo, V. F. P. Sogen, H. A. Dengen, A. Irsyad, and M. R. Ibrahim, "Pemetaan Sistem Informasi Geografis Daerah Rawan Banjir Di Wilayah Kota Bandung Jawa Barat Menggunakan Metode Skoring QGIS Tahun 2024," Kreatif Teknologi dan Sistem Informasi (KRETISI), vol. 3, no. 1, pp. 15–22, Jun. 2025, doi: 10.30872/kretisi.v3i1.2132.

UNDRR. (2022). The Sendai Framework for Disaster Risk Reduction 2015–2030. United Nations Office for Disaster Risk Reduction.

Wang, S., Chen, J., & Li, F. (2022). Rainfall-induced landslide susceptibility analysis using GIS and statistical models. Geomorphology, 403, 108158.

Zhang, C., Li, H., & Zhou, Q. (2023). Performance evaluation of ensemble ML models for landslide mapping. Remote Sensing of Environment, 298, 113783.

Zhou, J., Wu, H., & Guo, Z. (2023). Integrating GEE and ML for Landslide Hazard Mapping: A Cloud-based Framework. Remote Sensing Applications: Society and Environment, 31, 100987.