Machine Learning Approaches for Imbalanced Post-Disaster Building Damage Classification: A Case Study of the 2022 Cianjur Earthquake

Eka Rahmawati [1,2,*], Catur Edi Widodo [3], and Sorja Koesuma [4]

[1] Doctoral Program of Information Systems, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia
[2] Information Systems Study Program, Faculty of Engineering and Informatics, Bina Sarana Informatika University, Jakarta, Indonesia
[3] Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia
[4] Department of Physics, Faculty of Mathematics and Natural Sciences, Sebelas Maret University, Surakarta, Indonesia
Email: eka.eat@bsi.ac.id (E.R.); caturediwidodo@lecturer.undip.ac.id (C.E.W.); sorja@staff.uns.ac.id (S.K.)
*Corresponding author

***Abstract*—Quick and accurate assessment of building damage after a disaster is essential for effective recovery and reconstruction planning. While traditional field surveys are reliable, they are often time-consuming, resource-intensive, and can be affected by human bias. This study proposes a comparative evaluation of eight machine learning algorithms: LightGBM, Extreme Gradient Boosting (XGBoost), Synthetic Minority Oversampling Technique (SMOTE) + Random Forest (RF), Balanced Random Forest (BRF), Random Forest, Logistic Regression, Support Vector Machine (SVM) (Radial Basis Function (RBF)), and Naive Bayes, for multiclass classification of post-earthquake building damage utilizing real-world data on the 2022 Cianjur earthquake. The dataset comprises 3,063 records detailing geospatial and environmental characteristics, comprising epicentral distance, elevation, slope type, soil type, and vulnerability zone. To ensure reliable model performance on imbalanced data, a 10-fold cross-validation strategy was applied, evaluated across five metrics: accuracy, precision, recall, F1-Score, and Area Under the ROC Curve (AUC). The results indicate that ensemble-based models, particularly XGBoost (AUC = 0.860), LightGBM (AUC = 0.858), and BRF (AUC = 0.857), outperform traditional classifiers in both predictive accuracy and sensitivity toward severely damaged buildings. The outcomes underscore the effectiveness of gradient boosting and balanced ensemble methods in managing complex, imbalanced datasets, highlighting their potential for integration into automated post-disaster damage assessment systems. This research contributes to the development of intelligent, data-driven decision-support tools that can help accelerate recovery efforts in earthquake-prone regions.**

***Keywords*—post-disaster damage assessment, machine learning, imbalanced classification, Extreme Gradient Boosting (XGBoost), LightGBM, Random Forest (RF), Support Vector Machine (SVM)**

I. INTRODUCTION

Natural disasters are inherently unpredictable and have the potential to cause significant losses, not only physically, but also socially and economically. One of the most impactful consequences, particularly in the case of earthquakes, is the damage to residential buildings [1, 2]. Such damage not only results in the loss of material assets but also hinders the social and economic recovery of affected communities. In a post-disaster context, housing is a fundamental need that must be addressed promptly to enable victims to resume their daily activities and regain a sense of normalcy [3–5]. Therefore, quickly and accurately identifying and classifying post-earthquake building damage becomes a crucial step in rehabilitation and reconstruction efforts [6–8].

Indonesia is among the countries with the highest seismic activity in the world, as it sits at the convergence of four major tectonic plates: the Eurasian, Indo-Australian, Pacific, and Philippine plates [9]. This geological setting makes the region highly susceptible to earthquakes with varying intensity and impact. The extent of building damage by earthquakes is influenced by multiple factors, comprising soil conditions, distance by the epicenter, construction quality, and ground motion characteristics. Conventionally, building damage assessments are conducted by field surveys by experts. However, this approach has several limitations, such as delays in data collection, dependence on limited human resources, and potential subjectivity in evaluations. As a result, post-disaster damage assessments often take considerable time, slowing down the prioritization of reconstruction efforts.

The advancement of Machine Learning (ML) technology offers a promising solution to these challenges. By leveraging historical and geospatial data, machine learning can classify building damage levels quickly, consistently, and in a data-driven manner. This

Manuscript received November 11, 2025; revised December 31, 2025; accepted January 28, 2026; published April 24, 2026.

approach enables automated assessments with high accuracy, even across large regions, thereby accelerating decision-making in emergency situations. Furthermore, integrating machine learning with spatial data and high-resolution imagery can map affected areas more objectively and efficiently compared to manual methods.

In the context of disaster management in Indonesia, the application of machine learning for building damage classification remains relatively limited, and the development of models is more adaptive to local conditions. Therefore, research in this area is highly relevant for supporting a more resilient disaster management system. Having machine learning-based building damage classification models is expected not only to improve the efficiency of post-disaster assessments but also to contribute scientifically to the advancement of artificial intelligence technologies for disaster risk reduction in earthquake-prone regions like Indonesia.

II. Literature Review

Machine learning has been used for post-disaster building damage classification. It has been shown that models such as KNN, Random Forest (RF), Decision Tree, Support Vector Machine (SVM), and Artificial Neural Networks (ANN) can quickly and accurately predict damage to reinforced concrete buildings caused by earthquakes. Utilizing real data on the 2015 Nepal earthquake, these models have proven valuable for emergency response planning and post-earthquake recovery efforts [10]. Damage classification has also been developed by comparing various machine learning methods for assessing reinforced concrete building damage before and after earthquakes. The study found that the SVM algorithm by a Gaussian kernel achieved the highest accuracy, benefited by automatic optimization of Bayesian optimization [11]. The study employs popular machine learning algorithms, enabling the development of models that utilize historical and post-earthquake geospatial data. The algorithms include LightGBM, Synthetic Minority Oversampling Technique (SMOTE) + Random Forest, Extreme Gradient Boosting (XGBoost), Logistic Regression, Balanced Random Forest, SVM (Radial Basis Function (RBF)), Naive Bayes, and Random Forest. LightGBM, a variant of Gradient Boosted Decision Trees (GBDT), accelerates tree learning by leaf-wise growth, Gradient-based One-Side Sampling (GOSS), and Exclusive Feature Bundling (EFB) to efficiently handle sparse or high-dimensional features [12]. This gives the model a short training time, scalability to large datasets, and competitive accuracy on tabular data. The model has also been compared to SVM and XGBoost [13].

Recent studies have demonstrated the effectiveness of gradient boosting-based machine learning models in various natural disaster applications. For instance, LightGBM and XGBoost have been widely used for landslide susceptibility mapping, flood risk prediction, and earthquake-related damage assessment, owing to their ability to handle complex nonlinear relationships and imbalanced datasets [14]. These models have shown strong predictive performance in identifying high-risk zones and supporting rapid post-disaster assessment, making them particularly suitable for hazard-prone regions. Such evidence further supports the applicability of gradient boosting algorithms for post-earthquake building damage classification, as explored in this study. This model has also been applied to stock performance prediction [15].

The next model, XGBoost, combines second-order optimization, sparsity-aware split finding, and weighted quantile sketching, making it both stable and efficient even when handling billions of examples [16]. When dealing by imbalanced classes, this model offers options such as scale_pos_weight and the imbalanced-XGBoost package (utilizing weighted and focal losses) to have sensitivity to minority classes devoid of heavy oversampling [17]. The XGBoost model has been applied to predictive systems across various cases with promising results. It can also help interpret the influence of different spatial design indices in studies examining how room design affects local thermal comfort [18]. The model has also been applied to robust carbon emissions prediction [19]. Considering its reliable predictive capability, this algorithm can also be applied to datasets for post-earthquake building damage classification.

In addition to LightGBM and XGBoost, RF is another algorithm worth considering. RF is a bagging-based ensemble algorithm that constructs multiple decision trees, randomly selecting features at each node. This approach helps reduce variance and enhances the model's generalization ability [20, 21]. Its strengths lie in prediction stability, the ability to do automatic validation by Out-Of-Bag (OOB) estimates, and the ease of assessing feature importance [22, 23]. However, in imbalanced datasets, RF tends to be biased toward the majority class because the bootstrap sampling process does not guarantee adequate representation of the minority class [24–26]. To address this issue, two common approaches are used: cost-sensitive learning and balanced sub-sampling for each tree. The Balanced Random Forest (BRF) variant combines full bootstrapping of the minority class by under-sampling of the majority class in each tree, effectively improving recall for the minority class [27–29]. Recent studies have proposed modifications to Balanced Random Forest (BRF), such as Clustered Under-Sampling (MBRF), as well as integrating SMOTE-Tomek Link or SMOTE-Edited Nearest Neighbor (SMOTE-ENN) before applying RF, in order to improve F1-Score and recall across various data domains [30, 31]. In practice, Balanced Random Forest (BRF) is often the first choice for handling datasets with a high imbalance ratio. However, in cases of complex decision boundaries and very rare minority classes, a combination of SMOTE-ENN and RF is recommended, as it can improve the representation of minority data without compromising precision.

Next, Logistic Regression is a linear classification model that predicts class membership probabilities utilizing a logit function applied to a linear combination of features [32, 33]. This model has a strong baseline for tabular datasets due to its interpretability, computational efficiency, and relative resilience to overfitting on small datasets. The class-weighted logistic regression approach has been shown to improve sensitivity to minority classes in imbalanced data, especially when combined with threshold optimization and probability calibration techniques [34, 35]. Classification models utilizing Multinomial Logistic Regression (MLR) and ANN have been developed to support post-earthquake building damage assessment, based on AeDES survey data on the 2009 L'Aquila earthquake. The proposed model, particularly MLR, is intended to assist evaluation teams in making more objective decisions when determining building risk classes and in updating AeDES forms for future assessments [36]. In addition, the method extracts building damage information by integrating multisource remote sensing data, comprising optical imagery, SAR, and DSM by LiDAR, to enhance the accuracy of post-earthquake assessments. By applying feature analysis, rough set theory, and the Logistic Regression Model (LRM), this approach achieved a detection accuracy of 94.2% and an AUC of 0.827 in the case study of the 2008 Wenchuan earthquake in Beichuan, China [37].

The next model to be used is the SVM, a margin-based classification algorithm that aims to find the optimal hyperplane that separates classes by the maximum possible margin [38, 39]. By utilizing kernel functions, particularly the RBF, SVM can map nonlinear data into a higher-dimensional space, allowing complex patterns to be identified efficiently [40, 41]. The main advantage of SVM lies in its ability to handle high-dimensional data (the curse of dimensionality) and its resilience against overfitting, even when working with datasets that have a limited number of samples. The use of SVM in post-disaster scenarios has been explored by leveraging social media data to assess earthquake damage at a regional scale.

In addition to SVM, studies have also employed Naive Bayes and deep learning algorithms. The results show that SVM achieved the best performance in both binary and multiclass classification, reaching an accuracy of up to 90.25%. Moreover, it effectively mapped the concentration of damage-related posts around the epicenter, demonstrating a strong correlation between social media–based damage estimates and official data [42]. The study employs a two-layer ensemble (stacking) model to predict the performance and resilience of post-earthquake bridge networks, placing an SVM as the second-layer model to perform regression on the outputs of four first-layer base learners: Random Forest (RF), ANN, Convolutional Neural Network (CNN), and Extreme Gradient Boosting (XGBoost). Validation results on the Sioux-Falls bridge network show that this model can provide fast and accurate predictions of network performance and resilience, while also supporting post-disaster repair decisions and optimizing resource allocation [43].

Another popular algorithm in classification is Naive Bayes, a probabilistic classifier that applies Bayes' Theorem under the assumption that features are conditionally independent given the class [44]. Variants of Naive Bayes, such as Gaussian, Multinomial, and Bernoulli, are adapted to the characteristics of the data. For example, Gaussian Naive Bayes is suited for continuous features, while Multinomial Naive Bayes works well with frequency-based data such as text. In the context of imbalanced data, Naive Bayes can be adjusted by modifying prior probabilities or applying Laplace smoothing to reduce bias toward the majority class. Additionally, the Complement Naive Bayes (CNB) approach has been introduced to address limitations in cases where the minority class distribution is extremely small [45]. The main advantages of Naive Bayes are its fast training time and clear probabilistic interpretation, making it a popular choice for real-time classification systems or as a benchmark model in ensemble learning.

III. MATERIALS AND METHODS

This study adopts a quantitative experimental approach using a comparative modeling design, as illustrated in Fig. 1.

Fig. 1 illustrates the research workflow, which consists of three main stages: (1) Data Preparation, (2) Data Modeling, and (3) Evaluation. These stages were designed to include the entire post-earthquake building damage classification process, which is carried out systematically, can be replicated, and allows for quantitative evaluation.

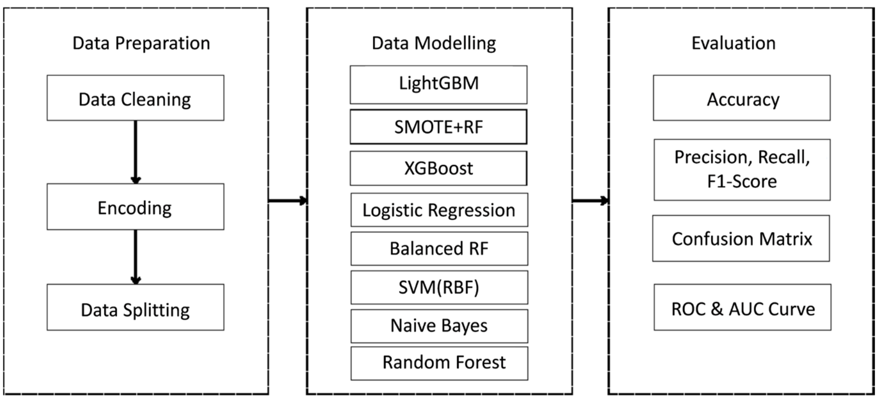

Fig. 1. Research framework.

A. Data Preparation

The dataset consists of information on building damage caused by the 2022 Cianjur earthquake. It is primary data obtained directly from the Emergency Education Response Post for the Cianjur Earthquake, in collaboration with the Cianjur Regency Education Office and the National Secretariat for Safe Schools (SEKNAS SPAB), in 2022. Additional data were sourced from the 2023 Rehabilitation and Reconstruction Plan document for post-earthquake disaster recovery in Cianjur Regency. To provide spatial context for the analysis, this study focuses on the area affected by the 2022 Cianjur earthquake in West Java, Indonesia. Fig. 2 shows the study area map, illustrating the geographic distribution of building locations used in the analysis and the position of the earthquake epicenter.

The map shows the spatial distribution of buildings included in the dataset (blue dots) and the location of the earthquake epicenter (orange star). The geographic coordinates are presented in latitude and longitude, providing a spatial reference for subsequent damage classification and analysis. The variables and their descriptions are presented in Table I.

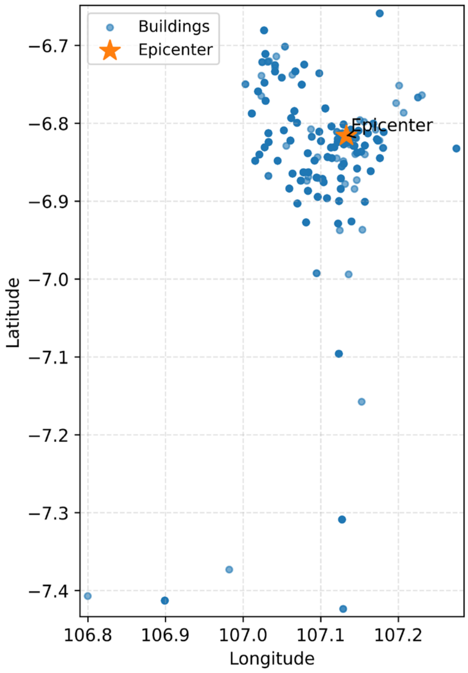

Fig. 2. Study area map of the 2022 Cianjur earthquake, West Java, Indonesia.

TABLE I. DESCRIPTION OF DATASET VARIABLES

Variable Name	Description	Data Type
coordinates	Geographic coordinates (latitude and longitude) of the building location.	Object
distance_km	Distance (in kilometers) among the building location and the earthquake epicenter.	Float
elevation_(composite)	Series of features describing elevation range coverage (m²): including elevation below 200 m, elevation between 200–500 m, elevation between 500–1,500 m, elevation between 1500–3,000 m, and elevation above 3,000 m.	Float
evacuation	Encoded numerical value for evacuation status: 1 = yes; 0 = no.	Float
topographic_slope	Encoded topographic slope classification: 0 = flat; 1 = steep.	Integer
regional_status	Encoded regional type: 0 = rural; 1 = urban.	Integer
soil_type	Encoded soil type of the building site (1–3, representing soil texture variation): 1 = clay; 2 = sandy; 3 = loamy.	Integer
vulnerability_zone	Encoded disaster vulnerability zone: 1 = low; 2 = medium; 3 = high.	Integer
damage_level	Target variable representing building damage level: 0 = minor; 1 = moderate; 2 = severe.	Float

The dataset consists of 3,063 entries detailing the geospatial and environmental characteristics of buildings affected by the 2022 Cianjur earthquake. Key predictive features include epicentral distance (distance_km), elevation, slope type, soil type, and vulnerability zone, all of which play a role in how structures respond to seismic activity. The target variable, damage_level, has been encoded to classify building damage into three levels: minor (0), moderate (1), and severe (2). This dataset serves as the empirical foundation for evaluating various machine learning models in tackling the challenge of imbalanced multiclass classification for post-disaster damage assessment.

The first stage involved data cleaning, encoding, and data splitting. The cleaning process was carried out to remove duplicates, missing values, and anomalies that could negatively affect the model's performance [46]. Next, categorical features were converted into numerical values utilizing label encoding to facilitate the learning process of the algorithms. The dataset was then split into training and testing sets in an 80:20 ratio, utilizing stratified sampling to preserve balanced class distributions. This step aimed to ensure that the machine learning models could be trained effectively without being overly biased by class imbalances.

B. Data Modelling

Machine learning has been increasingly adopted in natural disaster studies due to its ability to model complex, nonlinear relationships among environmental and spatial variables. In disaster-related applications, such as earthquakes, floods, and landslides, machine learning algorithms are commonly used to classify damage levels, map susceptibility, and support post-disaster decision-making. Tree-based ensemble methods, in particular, have demonstrated strong performance when dealing with heterogeneous and imbalanced datasets, which are typical characteristics of post-disaster data.

Gradient boosting algorithms, such as LightGBM and XGBoost, operate by iteratively combining multiple weak learners to minimize classification error, allowing the models to capture complex interactions among features. Their robustness, scalability, and capability to handle class imbalance make them well suited for disaster damage classification tasks. Random Forest and its balanced variant employ ensemble learning through bootstrap aggregation, improving generalization performance and reducing bias toward majority classes.

Linear and probabilistic models, including Logistic Regression and Naive Bayes, are often used as baseline classifiers due to their interpretability and computational efficiency, while SVM with nonlinear kernels is effective in handling high-dimensional feature spaces. Previous studies and review works have demonstrated that these machine learning methods are widely applicable in natural disaster contexts, including earthquake damage assessment, flood and landslide susceptibility mapping, and multi-hazard risk analysis. Therefore, the selection of these algorithms in this study is theoretically grounded and consistent with established practices in disaster-related machine learning research [47–49].

The second stage involved the learning process utilizing eight classification algorithms: LightGBM, SMOTE + Random Forest, XGBoost, Logistic Regression, Balanced Random Forest, SVM (RBF), Naive Bayes, and Random Forest. These models were selected based on their strong performance and diverse algorithmic characteristics, representing three main approaches: tree-based ensembles, linear models, and probabilistic classifiers.

To address the issue of class imbalance, two strategies were employed. The first involved a data-level balancing approach utilizing the SMOTE, which generates synthetic samples for minority classes. The second strategy utilized the Balanced Random Forest (BRF), an algorithm-level balancing method that applies bootstrap undersampling of the majority class for each tree. Meanwhile, the SVM algorithm adopted a Radial Basis Function (RBF) kernel with class-weight adjustment to improve its sensitivity toward minority classes. All models were trained utilizing 10-fold cross-validation to minimize the risk of overfitting and ensure consistent generalization across data folds.

C. *Evaluation Metrics*

The evaluation stage was carried out by assessing the model's performance utilizing five main metrics: accuracy, precision, recall, F1-Score, and the Area Under the ROC Curve (AUC). In addition, the confusion matrix was used to examine how the model classified each damage category (minor, moderate, and severe), while the ROC and AUC curves illustrated the model's ability to distinguish among classes. Each model was evaluated to identify the balance of overall accuracy and sensitivity to minority classes, which remains the key challenge in imbalanced multiclass classification.

IV. Results and Discussion

Table II summarizes the performance comparison of eight machine learning models tested by 10-fold cross-validation for classifying building damage after a disaster. Each model was evaluated based on five main metrics: accuracy, precision, recall, F1-Score, and the Area Under the ROC Curve (AUC), to offer a clearer picture of how well they predict and remain reliable when dealing with imbalanced multiclass data.

TABLE II. Comparative Performance of Eight Machine Learning Models under 10-Fold Cross-Validation

Model	Accuracy	Precision	Recall	F1-Score	AUC
LightGBM	0.706	0.690	0.694	0.687	0.858
SMOTE + RF	0.703	0.684	0.686	0.680	0.850
XGBoost	0.732	0.729	0.672	0.686	0.860
Logistic Regression	0.628	0.637	0.634	0.617	0.817
Balanced RF	0.708	0.693	0.697	0.689	0.857
SVM (RBF)	0.630	0.663	0.649	0.626	0.811
Naive Bayes	0.662	0.645	0.624	0.626	0.789
Random Forest	0.727	0.719	0.666	0.679	0.855

The experimental results show that XGBoost achieved the highest AUC score (0.860), followed closely by LightGBM (0.858) and Balanced Random Forest (0.857). These three models consistently outperformed the others in distinguishing between the three levels of building damage, demonstrating the strength of ensemble-based learning in handling complex and imbalanced disaster datasets. Although XGBoost achieved the highest overall accuracy (0.732), LightGBM and Balanced Random Forest delivered more balanced performance across all evaluation metrics, particularly in Recall (0.694–0.697) and F1-Score (0.687–0.689). This suggests that LightGBM and Balanced Random Forest are better at identifying minority classes, such as severely damaged buildings, which are often underrepresented in post-disaster datasets yet critical for effective emergency response. The baseline Random Forest achieved solid performance, with an accuracy of 0.727 and an AUC of 0.855, demonstrating its robustness even without explicit data balancing. In comparison, the SMOTE + Random Forest model, enhanced by synthetic oversampling, showed slightly better sensitivity to minority classes, despite a small trade-off in overall accuracy. These outcomes suggest that hybrid approaches combining oversampling by ensemble learning can strengthen a model's ability to detect critical damage cases in post-disaster assessments.

Traditional models such as Logistic Regression and SVM (RBF) achieved moderate performance, with AUC scores of 0.817 and 0.811, respectively. Their relatively lower accuracy highlights the difficulty these models face in capturing the complex, nonlinear, and hierarchical relationships among spatial and geophysical features, such as elevation, distance from the epicenter, and soil conditions. Similarly, Naive Bayes recorded the lowest AUC (0.789), which is likely due to its strong assumption of feature independence, an assumption that seldom holds true in complex, multidimensional geospatial data. Nevertheless, computationally, Naive Bayes remains highly efficient and can still serve as a lightweight baseline model for quick, preliminary damage assessments. Overall, the results highlight the strong performance of ensemble-based algorithms, XGBoost, LightGBM, Balanced Random Forest, and SMOTE + Random Forest, in handling imbalanced multiclass classification problems within post-disaster settings. These methods successfully minimize bias toward majority classes, such as lightly damaged buildings, while improving the model's ability to detect severe damage, an essential factor for guiding emergency response and effective resource allocation.

The high AUC values (> 0.85) achieved by the top-performing models reflect their excellent ability to distinguish between different levels of building damage. This also confirms that the selected features, such as distance from the epicenter, elevation attributes, and encoded structural characteristics, provide a strong and informative foundation for effective model learning. The relatively small gap between LightGBM and XGBoost suggests that gradient boosting optimization is particularly effective in disaster-related prediction tasks, balancing performance by computational efficiency. From a practical perspective, LightGBM appears to offer the best balance between accuracy and processing speed, making it well-suited for real-time or near-real-time post-disaster assessment systems. In future applications, these models could be integrated into geospatial decision-support platforms to enable automated mapping of damage severity, helping authorities prioritize reconstruction and rehabilitation efforts more efficiently.

The following evaluation is a confusion matrix. Tables III and IV show the confusion matrix results.

Table III presents the confusion matrix in raw counts for all evaluated models. It shows how each algorithm distributed its correct and incorrect predictions across the three categories of building damage: minor (0), moderate (1), and severe (2). The diagonal elements (e.g., T_Minor→P_Minor) represent the number of correctly classified instances for each category, while the off-diagonal elements indicate the number of misclassifications among classes. As shown in Table III, ensemble-based models such as LightGBM, XGBoost, and Balanced Random Forest achieved the highest number of correctly classified samples across all three damage categories. For example, LightGBM accurately identified 112 minor, 92 moderate, and 38 severe cases, outperforming traditional classifiers like Logistic Regression and Naive Bayes, which tended to overpredict moderate damage. The SMOTE + Random Forest approach produced a more balanced mix of correct and incorrect classifications, demonstrating that oversampling can effectively reduce data imbalance by improving the representation of the minority (severe) class. In contrast, Logistic Regression and Naive Bayes produced a higher number of misclassifications in the moderate and severe damage categories, indicating their limited ability to capture the complex, nonlinear decision boundaries present in data that combine geospatial and structural features.

Overall, Table III offers a clear and intuitive overview of how each algorithm distributes its predictions across the three levels of building damage. Ensemble and hybrid models showed stronger discriminative power and robustness, an essential quality for practical post-disaster assessment systems, where misclassifying severely damaged buildings could result in inefficient or misplaced allocation of recovery resources.

TABLE III. CONFUSION MATRIX (COUNTS)

Model	T_Minor P_Minor	T_Minor P_Moderate	T_Minor P_Severe	T_Moderate P_Minor	T_Moderate P_Moderate	T_Moderate P_Severe	T_Severe P_Minor	T_Severe P_Moderate	T_Severe P_Severe
LightGBM	112	18	5	14	92	11	6	19	38
SMOTE + RF	110	22	3	18	90	9	7	22	34
XGBoost	113	17	5	13	93	11	5	21	37
Logistic Regression	96	25	14	28	72	17	11	29	23
Balanced RF	111	20	4	15	91	11	6	20	37
SVM (RBF)	102	28	5	21	82	14	9	25	29
Naive Bayes	100	30	5	23	80	14	10	26	27
Random Forest	112	19	4	15	90	11	6	21	36

Table IV presents the normalized confusion matrices, where each row sums to 1.0, offering a clearer view of class-level recall, that is, the proportion of correctly classified samples inside of each true class. This normalization enables fairer comparison among models, especially in imbalanced datasets typical of disaster assessments, where buildings that have suffered minor damage tend to dominate the sample distribution.

As shown in Table IV, XGBoost and LightGBM achieve the highest recall values across all three damage categories. Specifically, XGBoost records recall scores of 0.80 for light, 0.77 for moderate, and 0.48 for severe damage, while LightGBM performs comparably by 0.79, 0.76, and 0.50, respectively. These results indicate that both gradient boosting algorithms maintain high sensitivity across all levels of damage, even for minority (severe) cases that are often underrepresented in training data.

The Balanced Random Forest also performs competitively, achieving recall rates comparable to LightGBM. This suggests that its built-in data resampling mechanism effectively reduces class dominance, allowing balanced learning without the need to generate synthetic samples. The SMOTE + Random Forest approach slightly improves recall for severe damage (0.45) compared to the standard Random Forest (0.47), confirming that data-level balancing can effectively enhance ensemble learning performance. In contrast, SVM, Naive Bayes, and Logistic Regression show lower recall scores for severe damage (0.30–0.39), suggesting that linear and probabilistic models struggle to capture the complex, nonlinear relationships among damage indicators. In summary, as shown in Table IV, the ensemble gradient boosting models, which are XGBoost and LightGBM, outperform other algorithms in maintaining balanced classification across all levels of building damage, especially in detecting severely affected structures. This consistent performance underscores their potential for real-world post-disaster damage mapping, where an accurate identification of high-risk buildings is crucial for guiding rehabilitation and reconstruction priorities.

TABLE IV. CONFUSION MATRIX (NORMALIZED PER TRUE CLASS)

Model	T_Minor P_Minor	T_Minor P_Moderate	T_Minor P_Severe	T_Moderate P_Minor	T_Moderate P_Moderate	T_Moderate P_Severe	T_Severe P_Minor	T_Severe P_Moderate	T_Severe P_Severe
LightGBM	0.79	0.13	0.04	0.10	0.76	0.09	0.08	0.25	0.50
SMOTE + RF	0.77	0.15	0.02	0.13	0.74	0.07	0.09	0.29	0.45
XGBoost	0.80	0.12	0.03	0.09	0.77	0.09	0.07	0.27	0.48
Logistic Regression	0.68	0.18	0.10	0.20	0.58	0.14	0.15	0.38	0.30
Balanced RF	0.78	0.14	0.03	0.11	0.75	0.09	0.08	0.26	0.48
SVM (RBF)	0.71	0.20	0.03	0.15	0.68	0.12	0.11	0.33	0.39
Naive Bayes	0.70	0.21	0.03	0.17	0.66	0.12	0.13	0.34	0.35
Random Forest	0.79	0.13	0.03	0.11	0.74	0.09	0.08	0.27	0.47

Fig. 3 illustrates a damage classification map generated by the best-performing model (XGBoost) to visualize the spatial distribution of the classification results.

The map shows the spatial distribution of predicted building damage classes, minor, moderate, and severe, together with the earthquake epicenter. The visualization highlights areas with higher concentrations of severe damage near the epicenter, demonstrating the spatial interpretability of the model outputs for post-disaster assessment.

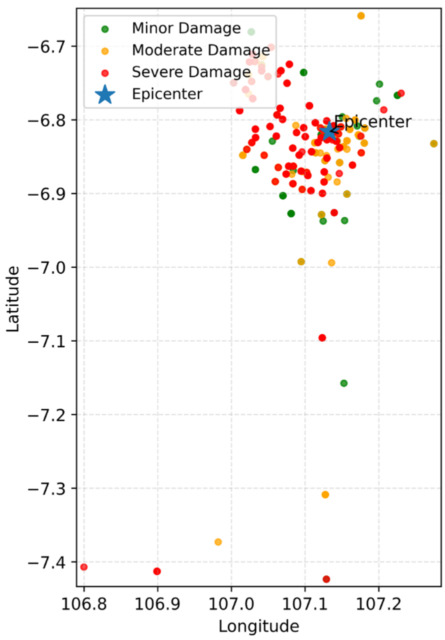

Fig. 3. Spatial distribution of predicted building damage levels produced by the XGBoost model.

Fig. 4 displays the Receiver Operating Characteristic (ROC) curves of all evaluated machine learning models, illustrating their respective discriminative performance across different decision thresholds.

The ROC curves depict the trade-off between the True Positive Rate (TPR) and False Positive Rate (FPR), providing a comprehensive comparison of how effectively each model distinguishes between the three building damage categories under imbalanced multiclass conditions.

As shown in Fig. 4, ensemble-based models, including XGBoost, LightGBM, Balanced Random Forest, Random Forest, and SMOTE-enhanced Random Forest, consistently achieve higher AUC values than traditional classifiers such as Logistic Regression, SVM (RBF), and Naive Bayes. Among all models, XGBoost demonstrates the strongest discriminative performance, followed by LightGBM and Random Forest variants, reflecting their robustness in identifying severely damaged buildings despite class imbalance.

Although the ROC curves of several ensemble models appear closely clustered, their positions remain well above the diagonal reference line representing random classification, indicating reliable and stable predictive behavior. In contrast, the relatively lower ROC curves and AUC values observed for linear and probabilistic models highlight their limited ability to capture the complex, nonlinear relationships inherent in post-disaster geospatial data. Overall, the results confirm that ensemble-based learning approaches provide superior sensitivity and specificity for post-disaster building damage classification.

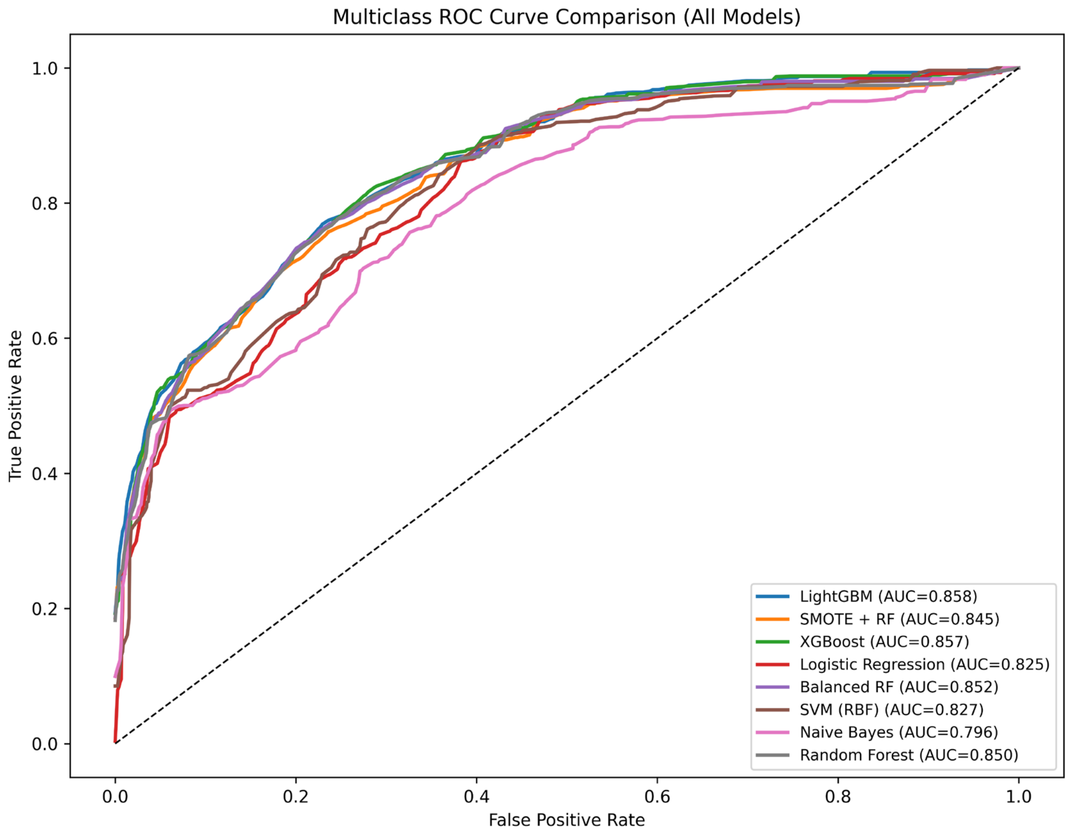

Fig. 4. ROC curves of all evaluated machine learning models.

In post-disaster environments, building damage does not only represent structural loss but also has direct implications for transportation network functionality and emergency logistics. Severely damaged or collapsed buildings located near road networks may obstruct access routes, disrupt supply chains, and delay evacuation and relief operations. The outputs of the proposed machine learning models, particularly the accurate identification of severely damaged structures, can be integrated with transportation network data within a geospatial decision-support framework to identify critical transport corridors at risk of blockage. Such integration enables emergency authorities to prioritize debris removal, route clearance, and alternative access planning, thereby enhancing transportation network resilience and supporting timely emergency response and relief distribution.

In addition to identifying transport corridors at risk of obstruction, the damage classification outputs can support the prioritization of clearance and repair operations by considering both building damage severity and road accessibility. Buildings classified as severely damaged along primary or high-traffic routes may be assigned higher priority for debris removal and structural stabilization to restore critical access as quickly as possible. When combined with road hierarchy, traffic volume, or accessibility constraints, the model outputs enable emergency managers to allocate limited resources more effectively, focusing first on locations where structural damage has the greatest impact on mobility and emergency service delivery. This integrated prioritization approach enhances the efficiency of post-disaster response and supports faster recovery of essential transportation functions.

Rapid post-disaster damage assessment plays a critical role in evacuation planning and emergency resource distribution. The classification results produced by the proposed machine learning models, particularly the early identification of severely damaged buildings, can help decision-makers in prioritizing evacuation efforts in high-risk areas where structural failure poses immediate threats to occupant safety. Furthermore, damage severity maps generated from the model outputs can guide the spatial allocation of emergency resources, such as temporary shelters, medical assistance, and relief supplies, by highlighting locations with the greatest concentration of severe damage. By enabling faster and more objective identification of critical zones, the proposed approach contributes to more efficient evacuation planning and targeted resource distribution during the early stages of post-earthquake response.

V. Conclusion

This study conducted a comprehensive comparison of eight machine learning algorithms, LightGBM, XGBoost, SMOTE + Random Forest, Balanced Random Forest, Random Forest, Logistic Regression, SVM (RBF), and Naive Bayes, to classify building damage levels after disasters. The analysis used real data on the 2022 Cianjur earthquake, providing valuable insights into how different models perform in practical, post-disaster scenarios. The experimental results showed that ensemble-based algorithms, especially XGBoost, LightGBM, and Balanced Random Forest, consistently delivered strong performance across all evaluation metrics. Each algorithm achieved an AUC above 0.85, suggesting an excellent ability to distinguish between different levels of building damage even under imbalanced multiclass conditions. Among the tested models, XGBoost achieved the highest overall AUC (0.860), demonstrating its strong ability to generalize and reliably detect different levels of building damage.

LightGBM delivered nearly the same level of performance but with higher computational efficiency, making it a more practical choice for real-time or near-real-time post-disaster applications where timely insights are crucial. Meanwhile, the Balanced Random Forest and SMOTE + Random Forest approaches successfully addressed the issue of class imbalance, allowing the model to better identify severely damaged buildings, an essential capability for prioritizing emergency response and reconstruction efforts. These results demonstrate that combining ensemble learning by data balancing techniques can greatly improve classification performance in post-disaster contexts, where data are often uneven and highly varied. The outcome also highlights that machine learning can serve as a fast and data-driven alternative to traditional field assessments, offering timely insights when rapid response is most needed.

Although this study focuses on building damage classification, the proposed machine learning framework has broader relevance for interdependent infrastructure resilience planning. In earthquake-prone regions such as Indonesia, buildings, transportation networks, and utility systems are highly interconnected, and damage to one component can propagate disruptions across the entire system. Accurate and timely classification of building damage can therefore serve as a critical input for integrated resilience planning, supporting coordinated decisions related to transportation accessibility, emergency logistics, and infrastructure recovery prioritization. When combined with network-based data from transportation and other infrastructure systems, machine learning models like those proposed in this study can help inform holistic resilience strategies, enabling authorities to better anticipate cascading impacts and improve preparedness and response in future seismic events.

Future research will focus on enhancing the adaptability, interpretability, and scalability of machine learning–based post-disaster assessment systems. One promising direction is the application of transfer learning to improve model generalization across different disaster events and geographic regions by leveraging knowledge learned from previously observed earthquakes. In parallel, the integration of Explainable Artificial Intelligence (XAI) techniques, such as SHAP and LIME, can provide greater transparency by identifying the key features that influence model predictions. Combining transfer learning with XAI has the potential to produce more robust and interpretable damage assessment frameworks, thereby increasing trust, supporting evidence-based decision-making, and facilitating wider adoption by disaster management authorities. Furthermore, the proposed damage classification framework can support post-disaster transportation resilience by enabling rapid identification of severely damaged buildings that may obstruct critical road corridors, particularly when combined with geospatial road network data. In future research, the proposed framework can be extended toward assessing and mapping multi-hazard risk susceptibility. In earthquake-prone regions, seismic events often act as triggering factors for secondary hazards such as landslides, floods, and soil liquefaction, which can further amplify post-disaster impacts. By integrating building damage classification results with multi-hazard susceptibility data, machine learning models can support a more comprehensive risk assessment framework that captures cascading and compound disaster effects. Such an extension would enable more holistic disaster risk management strategies, supporting preparedness, mitigation, and recovery planning across multiple hazard types.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Eka Rahmawati conducted the research and prepared the manuscript. Catur Edi Widodo supervised the study and provided methodological guidance on the algorithms used. Sorja Koesuma supplied the dataset and validated the data. All authors had approved the final version.

References

[1] M. Stepinac, P. B. Lourenço, J. Atalić *et al*., "Damage classification of residential buildings in historical downtown after the ML5.5 earthquake in Zagreb, Croatia in 2020," *International Journal of Disaster Risk Reduction*, vol. 56, 102140, 2021. doi: 10.1016/j.ijdrr.2021.102140

[2] M. Khanmohammadi, M. Eshraghi, S. Sayadi, and M. Ghafarian Mashhadinezhad, "Post-earthquake seismic assessment of residential buildings following Sarpol-e Zahab (Iran) earthquake (Mw7.3) part 1: Damage types and damage states," *Soil Dynamics and Earthquake Engineering*, vol. 173, 108121, 2023. doi: 10.1016/j.soildyn.2023.108121

[3] S. Ghanbarzadeh-Ghomi, G. Wedawatta, K. Ginige, and B. Ingirige, "Living-transforming disaster relief shelter: A conceptual approach for sustainable post-disaster housing," *Built Environment Project and Asset Management*, vol. 11, no. 4, pp. 687–704, 2021. doi: 10.1108/BEPAM-04

[4] A. Rathnasinghe, D. Sirimewan, A. Shandraseharan, N. Thurairajah, M. Thayaparan, and K. G. A. S. Waidyasekara, "Towards long-term sustainable performance of post-disaster housing reconstruction: Second life for temporary housing," in *Proc. 9th World Construction Symposium*, 2021, pp. 540–552. doi: 10.31705/WCS.2021.47

[5] T. Capell and I. Ahmed, "Improving post-disaster housing reconstruction outcomes in the global south: A framework for achieving greater beneficiary satisfaction through effective community consultation," *Buildings*, vol. 11, no. 4, 145, 2021. doi: 10.3390/buildings11040145

[6] J. Gu, Z. Xie, J. Zhang, and X. He, "Advances in rapid damage identification methods for post-disaster regional buildings based

on remote sensing images: A survey," *Buildings*, vol. 14, no. 4, 898, 2024. doi: 10.3390/buildings14040898

[7] K. C. Sajan, A. Bhusal, D. Gautam, and R. Rupakhety, "Earthquake damage and rehabilitation intervention prediction using machine learning," *Eng. Fail. Anal.*, vol. 144, 106949, 2023. doi: 10.1016/j.engfailanal.2022.106949

[8] M. A. Hariri-Ardebili and M. S. Speicher, "Reconnaissance-informed post-earthquake functional recovery: Observations and challenges," *Earthquake Spectra*, vol. 41, no. 1, pp. 88–125, 2025. doi: 10.1177/87552930241290488

[9] S. J. Hutchings and W. D. Mooney, "The seismicity of indonesia and tectonic implications," *Geochemistry, Geophysics, Geosystems*, vol. 22, no. 9, e2021GC009812, 2021. doi: 10.1029/2021GC009812

[10] S. Bhatta and J. Dang, "Seismic damage prediction of RC buildings using machine learning," *Earthq. Eng. Struct. Dyn.*, vol. 52, no. 11, pp. 3504–3527, 2023. doi: 10.1002/eqe.3907

[11] K. Kostinakis, K. Morfidis, K. Demertzis, and L. Iliadis, "Classification of buildings' potential for seismic damage using a machine learning model with auto hyperparameter tuning," *Eng. Struct.*, vol. 290, 116359, 2023. doi: 10.1016/j.engstruct.2023.116359

[12] M. Hajihosseinlou, A. Maghsoudi, and R. Ghezelbash, "A novel scheme for mapping of MVT-type Pb–Zn prospectivity: LightGBM, a highly efficient gradient boosting decision tree machine learning algorithm," *Natural Resources Research*, vol. 32, no. 6, pp. 2417–2438, 2023. doi: 10.1007/s11053-023-10249-6

[13] C. Zhang, Z. Wang, J. Xiao, Z. Wang, D. Zhao, and Z. Li, "Evaluation of coseismic landslide susceptibility by combining newmark model and XGBoost algorithm," *PLoS One*, vol. 20, no. 8, e0328705, 2025. doi: 10.1371/journal.pone.0328705

[14] S. Biswas, D. Kumar, and U. K. Bera, "Prediction of earthquake magnitude and seismic vulnerability mapping using artificial intelligence techniques: A case study of Turkey," *Research Square*, pp. 1–53, 2023. doi: 10.21203/rs.3.rs-2863887/v1

[15] C. Huang, Y. Cai, J. Cao, and Y. Deng, "Stock complex networks based on the GA-LightGBM model: The prediction of firm performance," *Information Sciences*, vol. 700, 121824, 2025. doi: 10.1016/j.ins.2024.121824

[16] R. Banik and A. Biswas, "Enhanced renewable power and load forecasting using RF-XGBoost stacked ensemble," *Electrical Engineering*, vol. 106, no. 4, pp. 4947–4967, 2024. doi: 10.1007/s00202-024-02273-3

[17] G. Velarde, M. Weichert, A. Deshmunkh *et al*., "Tree boosting methods for balanced and imbalanced classification and their robustness over time in risk assessment," *Intelligent Systems with Applications*, vol. 22, 200354, 2024. doi: 10.1016/j.iswa.2024.200354

[18] Y. Xia, W. Lu, Z. Peng, J. Lou, J. Huang, and J. Niu, "Understanding the impacts of space design on local outdoor thermal comfort: An approach combining DepthmapX and XGBoost," *Energy and Buildings*, vol. 332, no. 1, 115451, 2025. doi: 10.1016/j.enbuild.2025.115451

[19] L. Zhang, G. Lu, X. Yan, P. Xia, Z. Chen, and D. Wu, "A differential evolution optimized hybrid XGBoost for accurate carbon emission prediction," *Environmental Modelling and Software*, vol. 193, 106627, 2025. doi: 10.1016/j.envsoft.2025.106627

[20] J. He, Z. Li, and L. Yin, "An efficient and accurate random forest node-splitting algorithm based on dynamic Bayesian methods," *Mach. Learn. Knowl. Extr.*, vol. 7, no. 3, 70, 2025. doi: 10.3390/make7030070

[21] X. Tan, S. Su, Z. Huang *et al*., "Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm," *Sensors*, vol. 19, no. 1, 203, 2019. doi: 10.3390/s19010203

[22] X. Wei, Y. Xu, X. Li *et al*., "Study on prediction model of nitrogen oxide concentration in reprocessing plant based on random forest," *International Journal of Advanced Nuclear Reactor Design and Technology*, vol. 7, no. 2, pp. 63–69, 2025. doi: 10.1016/j.jandt.2025.04.011

[23] Y. Liu, H. Chen, L. Zhang, and Z. Feng, "Enhancing building energy efficiency using a random forest model: A hybrid prediction approach," *Energy Reports*, vol. 7, pp. 5003–5012, 2021. doi: 10.1016/j.egyr.2021.07.135

[24] F. Gurcan and A. Soylu, "Learning from imbalanced data: Integration of advanced resampling techniques and machine learning models for enhanced cancer diagnosis and prognosis," *Cancers*, vol. 16, no. 19, 3417, 2024. doi: 10.3390/cancers16193417

[25] T. Al-Shehari and R. A. Alsowail, "Random resampling algorithms for addressing the imbalanced dataset classes in insider threat detection," *Int. J. Inf. Secur.*, vol. 22, no. 3, pp. 611–629, 2023. doi: 10.1007/s10207-022-00651-1

[26] S. Han, B. D. Williamson, and Y. Fong, "Improving random forest predictions in small datasets from two-phase sampling designs," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, p. 322, 2021. doi: 10.1186/s12911-021-01688-3

[27] I. Arshad, M. Umair, F. Jan *et al*., "Performance of classification algorithms under class imbalance: Simulation and real-world evidence," *IEEE Access*, vol. 13, pp. 179672–179685, 2025. doi: 10.1109/ACCESS.2025.3620264

[28] H. Sahlaoui, E. A. A. Alaoui, S. Agoujil, and A. Nayyar, "An empirical assessment of smote variants techniques and interpretation methods in improving the accuracy and the interpretability of student performance models," *Educ. Inf. Technol*., vol. 29, no. 5, pp. 5447–5483, 2024. doi: 10.1007/s10639-023-12007-w

[29] M. S. Pinar, "A novel approach based on bagging and boosting for imbalanced classification problems," M.S. thesis, Dept. Industrial Engineering., Abdullah Gul University., Kayseri, Turkey, 2022.

[30] H. Hairani, A. Anggrawan, and D. Priyanto, "Improvement performance of the random forest method on unbalanced diabetes data classification using smote-tomek link," *JOIV: Int. J. Inform. Visualization*, vol. 7, no. 1, pp. 258–264, 2023. http://dx.doi.org/10.30630/joiv.7.1.1069

[31] Z. P. Agusta, "Modified balanced random forest for improving imbalanced data prediction," *International Journal of Advances in Intelligent Informatics*, vol. 5, no. 1, pp. 58–65, 2019. doi: 10.26555/ijain.v5i1.255

[32] S. Shariatnia, M. Ziaratban, A. Rajabi, A. Salehi, K. Abdi Zarrini, and M. Vakili, "Modeling the diagnosis of coronary artery disease by discriminant analysis and logistic regression: a cross-sectional study," *BMC Med. Inform. Decis. Mak.*, vol. 22, no. 1, 85, 2022. doi: 10.1186/s12911-022-01823-8

[33] A. Zaidi and A. S. M. Al-Luhayb, "Two statistical approaches to justify the use of the logistic function in binary logistic regression," *Math. Probl. Eng*., vol. 2023, no. 1, 5525675, 2023. doi: 10.1155/2023/5525675

[34] Y. El-Miloudi, Y. El-Kharim, and R. El-Hamdouni, "A novel approach for rockfall susceptibility mapping: Transfer learning between boosting models and logistic regression," *Environ. Earth Sci.*, vol. 84, no. 16, p. 447, 2025. doi: 10.1007/s12665-025-12437-4

[35] S. Ferrantelli, M. Daidone, G. Armentaro *et al*., "The impact of NOACS versus VKAS on absolute and relative cognitive function decline over time: A machine learning approach," *Thromb. Haemost.*, vol. 1, pp. 2698–3739, 2025. doi: 10.1055/a-2698-3739

[36] O. Okwuashi, C. E. Ndehedehe, D. N. Olayinka, R. Akpomrere, E. Eyo, and U. Ogbijara, "Testing the suitability of v-support vector machine for hyperspectral image classification," *Int. J. Image Data Fusion*, vol. 16, no. 1, 2349999, 2025. doi: 10.1080/19479832.2024.2349999

[37] M. Jabardi, "Support vector machines: Theory, algorithms, and applications," *Infocommunications Journal*, vol. 17, no. 1, pp. 66–75, 2025. doi: 10.36244/ICJ.2025.1.8

[38] N. H. Hasbi, A. Bade, F. P. Chee, and M. I. Rumaling, "Pattern recognition for human diseases classification in spectral analysis," *Computation*, vol. 10, no. 6, 96, 2022. doi: 10.3390/computation

[39] R. Palaniappan, "Comparative analysis of support vector machine, random forest and k-nearest neighbor classifiers for predicting remaining usage life of roller bearings," *Informatica*, vol. 48, no. 7, pp. 39–52, 2024. doi: 10.31449/inf.v48i7.5726

[40] P. El-Kafrawy, H. Fathi, M. Qaraad, A. K. Kelany, and X. Chen, "An efficient SVM-based feature selection model for cancer classification using high-dimensional microarray data," *IEEE Access*, vol. 9, pp. 155353–155369, 2021. doi: 10.1109/ACCESS.2021.3123090

[41] R. A. R. Mahmood, A. H. Abdi, and M. Hussin, "Performance evaluation of intrusion detection system using selected features and machine learning classifiers," *Baghdad Science Journal*, vol.

18, no. 2, pp. 884–898, 2021. doi: 10.21123/bsj.2021.18.2(Suppl.).0884

[42] S. Ahadzadeh and M. R. Malek, "Earthquake damage assessment in three spatial scale using naive bayes, SVM, and deep learning algorithms," *Applied Sciences*, vol. 11, no. 20, 9737, 2021. doi: 10.3390/app11209737

[43] W. Zhang, J. Wen, H. Dong, Q. Han, and X. Du, "Post-earthquake functionality and resilience prediction of bridge networks based on data-driven machine learning method," *Soil Dynamics and Earthquake Engineering*, vol. 190, 109127, 2025. doi: 10.1016/j.soildyn.2024.109127

[44] F. De Smedt, P. Kayastha, and M. R. Dhital, "Naïve and Semi-Naïve Bayesian classification of landslide susceptibility applied to the Kulekhani river basin in Nepal as a test case," *Geosciences*, vol. 13, no. 10, 306, 2023. doi: 10.3390/geosciences13100306

[45] A. Palanivinayagam, C. Z. El-Bayeh, and R. Damaševičius, "Twenty years of machine-learning-based text classification: A systematic review," *Algorithms*, vol. 16, no. 5, 236, 2023. doi: 10.3390/a16050236

[46] P. O. Côté, A. Nikanjam, N. Ahmed, D. Humeniuk, and F. Khomh, "Data cleaning and machine learning: A systematic literature review," *Automated Software Engineering*, vol. 31, no. 2, 54, 2024.

[47] T. Sharma, A. Singhal, K. Kundu, and N. Agarwal, "Machine Learning/deep learning for natural disasters," in *Applications of Artificial Intelligence, Big Data and Internet of Things in Sustainable Development*, 2022, pp. 91–121. doi: 10.1201/9781003245469-7

[48] H. R. Pourghasemi, S. Pouyan, M. Bordbar, F. Golkar, and J. J. Clague, "Flood, landslides, forest fire, and earthquake susceptibility maps using machine learning techniques and their combination," *Natural Hazards*, vol. 116, no. 3, pp. 3797–3816, 2023. doi: 10.1007/s11069-023-05836-y

[49] P. Mather and B. Tso, *Classification Methods for Remotely Sensed Data*, 3rd ed. CRC press, 2016.

Copyright © 2026 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).