International Journal of Informatics and Computation (IJICOM)
Vol. 7, No. 2, 2025
ISSN: 2685-8711, E-ISSN: 2714-5263 DOI: 10.35842/ijicom

Classifying Post-Eruption Housing Occupancy Status using Random Forest Algorithm

Nurhadi Wijaya[1], Irawadi Buyung[2], Sugeng Winardi[3], Galant Nanta[4], Abdullah Azzam[5], Imam Muchlisin[6]

Abstract

Natural disasters such as the Mount Merapi eruption in Indonesia cause long-term damage to housing and settlement systems. After reconstruction, authorities must verify whether rebuilt houses are actually occupied to ensure fair aid distribution and effective recovery planning. This study applies the Random Forest algorithm to classify the occupancy status of post-eruption housing using real data collected by post-disaster rehabilitation agencies. We use a dataset of 2,516 housing records and split the data into 80% for training and 20% for testing. The experimental results show that the proposed model achieves an accuracy of 91.26% and an Area Under the Curve (AUC) value of 0.81. These results indicate that Random Forest performs well in distinguishing between occupied and unoccupied houses, even when the data reflect real-world imbalance and uncertainty. This study demonstrates that Random Forest is a practical and reliable machine learning method for post-disaster housing analysis. The model requires modest computational resources and handles mixed data types effectively. The findings confirm that ensemble-based machine learning can support government decision-making, strengthen monitoring of rehabilitation programs, and improve the overall effectiveness of disaster recovery efforts.

Keywords:
Random Forest; Classification; Occupancy Status; Post-Disaster Rehabilitation; Mount Merapi

1. Introduction

Post-eruption housing recovery remains a critical challenge in disaster-prone regions, particularly in volcanic areas such as Mount Merapi in Indonesia. After an eruption, governments and rehabilitation agencies must rapidly determine whether reconstructed houses are occupied, partially occupied, or abandoned to ensure effective allocation of resources and long-term settlement sustainability. However, occupancy assessment often relies on manual surveys and subjective evaluations, which are time-consuming, costly, and prone to inconsistencies. These limitations create delays in decision-making and reduce the effectiveness of post-disaster housing policies, highlighting the need for data-driven and automated approaches to classify housing occupancy status accurately [2], [3], [24].

Previous studies emphasize that housing occupancy status reflects not only physical reconstruction success but also social, economic, and environmental recovery factors. In many rehabilitation programs, a house may be structurally completed but remain

---

Corresponding Author: Nurhadi Wijaya, Departement Of Informatics, University of Respati Yogyakarta, nurhadi@respati.ac.id
1 Nurhadi Wijaya, Departement Of Informatics, University of Respati Yogyakarta, nurhadi@respati.ac.id
2 Irawadi Buyung, Departement Electrical Engineering, University of Respati Yogyakarta, buyungnasution@gmail.com
3. Sugeng Winardi, Departement Of Information System, University of Respati Yogyakarta, sugeng@respati.ac.id
4.Galant Nanta, Departement of English Literature Study Program, University of Respati Yogyakarta, galant.nanta@respati.ac.id
5. Abdullah Azzam, Departement of Informatics, University of Respati Yogyakarta, 22220032@respati.ac.id
6. Imam Muchlisin, Departement Of Information System, University of Respati Yogyakarta 22230020@respati.ac.id

unoccupied due to livelihood loss, accessibility issues, or inadequate supporting infrastructure. Conventional evaluation methods struggle to capture these multidimensional factors systematically. As a result, post-eruption housing programs often face mismatches between reported rehabilitation outcomes and actual residential use, which undermines sustainability goals and long-term community resilience [3], [24].

To address these challenges, researchers increasingly apply data mining and machine learning techniques to classification problems in disaster recovery contexts. Data mining provides structured methods to extract meaningful patterns from complex datasets, enabling objective and scalable analysis. Core concepts and methodologies proposed in classical data mining literature establish a strong theoretical foundation for applying supervised learning algorithms to social and environmental datasets, including housing and settlement data. However, selecting an appropriate algorithm remains a critical issue, especially when dealing with heterogeneous, noisy, and imbalanced post-disaster data [9], [10], [11].

Several studies explore traditional classification algorithms such as Naive Bayes, Decision Tree C4.5, and K-Nearest Neighbors to classify housing occupancy status in post-eruption rehabilitation programs. These approaches demonstrate the feasibility of machine learning for occupancy classification but also reveal limitations related to sensitivity to data distribution, overfitting, and reduced performance when handling nonlinear relationships. As datasets grow in size and complexity, these single-model approaches struggle to maintain stable accuracy, indicating the need for more robust ensemble-based methods [4], [5], [6].

More advanced machine learning models, including AdaBoost and Support Vector Machines, further improve classification performance by addressing some weaknesses of earlier algorithms. Studies applying AdaBoost and SVM to post-eruption housing data report higher accuracy and better generalization. Nevertheless, these methods introduce new challenges, such as sensitivity to parameter tuning, computational complexity, and limited interpretability for policymakers. These issues motivate the exploration of alternative ensemble techniques that balance accuracy, robustness, and interpretability [7], [8], [12].

Random Forest emerges as a strong candidate to address these challenges due to its ensemble learning structure and inherent ability to handle high-dimensional and nonlinear data. Introduced by Breiman, Random Forest constructs multiple decision trees using bootstrap sampling and random feature selection, thereby reducing overfitting and improving generalization. Prior studies demonstrate its effectiveness across various domains, including ecology, remote sensing, and spatial classification. These characteristics make Random Forest particularly suitable for post-disaster housing occupancy classification, where data variability and uncertainty are common [1], [13], [14], [15].

Recent disaster-related studies further confirm the effectiveness of Random Forest for post-disaster assessment tasks. Researchers apply Random Forest to disaster risk assessment, building damage mapping, and residential damage classification using survey and multi-source data. These studies consistently report superior performance compared to single classifiers, especially in terms of accuracy, stability, and resistance to noise. However, most existing works focus on physical damage assessment rather than post-rehabilitation occupancy status, leaving a research gap in understanding residential utilization after reconstruction [18], [19], [20], [22].

Based on these gaps, this paper focuses on classifying post-eruption housing occupancy status using the Random Forest algorithm. We utilize Random Forest to capture complex interactions among housing conditions, socio-economic indicators, and environmental factors within rehabilitation datasets. By applying this ensemble approach, the study aims to improve classification reliability while providing insights that support evidence-based policy and sustainable settlement planning. This research contributes to disaster informatics by extending the application of Random Forest from damage assessment to post-rehabilitation occupancy analysis, thereby addressing a crucial yet underexplored aspect of post-disaster recovery [21], [23], [25].

2. Related Works

Early studies on post-disaster housing evaluation primarily relied on qualitative assessments and descriptive statistical analyses. Setiawan emphasized community facilitation and social participation as key indicators of housing success in post-disaster rehabilitation programs, highlighting that physical reconstruction alone did not guarantee residential occupancy. Similarly, the Rehabilitation and Reconstruction Task Force documented that many reconstructed houses remained unoccupied due to socio-economic and livelihood constraints. While these studies provided valuable contextual insights, they lacked quantitative and automated classification mechanisms, making large-scale and objective occupancy assessment difficult to achieve [2], [3].

With the growth of data-driven approaches, researchers began applying classical machine learning algorithms to housing occupancy classification problems. Wijaya applied the Naive Bayes algorithm to classify rehabilitation housing occupancy status and demonstrated that probabilistic methods could support decision-making in post-eruption contexts. Although the study achieved reasonable accuracy and simplicity in implementation, the method assumed feature independence, which limited its ability to model complex relationships among housing, social, and environmental variables. This limitation reduced classification robustness when applied to more diverse datasets [4].

Subsequent research explored decision tree-based models to improve interpretability and capture nonlinear patterns. Wijaya and Marselina employed the C4.5 decision tree algorithm to classify housing occupancy status and showed that rule-based outputs enhanced transparency for policymakers. However, the study reported sensitivity to noise and a tendency toward overfitting, particularly when handling imbalanced data. These weaknesses highlighted the need for models that maintained interpretability while improving generalization performance [5].

Distance-based learning approaches also gained attention in this domain. Wijaya et al. implemented the K-Nearest Neighbors (KNN) algorithm for occupancy classification and demonstrated its effectiveness in small to medium-sized datasets. The model benefited from its simplicity and adaptability to local data patterns. Nevertheless, the authors reported increased computational costs and performance degradation as the dataset size grew. Furthermore, KNN lacked an inherent mechanism to handle irrelevant features, which limited its scalability for large post-disaster housing databases [6].

To overcome the limitations of single classifiers, ensemble and margin-based learning techniques were introduced. Wijaya, Diqi, and Mustiadi applied the AdaBoost algorithm to predict post-eruption residential habitation status and achieved improved accuracy compared to earlier models. Similarly, subsequent work utilized Support Vector Machines (SVM) to classify occupancy status with strong generalization capabilities. Despite these improvements, both AdaBoost and SVM required careful parameter tuning and offered limited interpretability, which constrained their practical adoption in policy-oriented environments [7], [8].

Beyond housing-specific studies, several researchers investigated machine learning methods for post-disaster damage and settlement assessment. Hasan et al. employed supervised learning techniques using remote sensing data to assess post-disaster damage and demonstrated the potential of machine learning in capturing spatial damage patterns. Geiß et al. and Zhang et al. further applied advanced classifiers to map damaged buildings using multi-source data. While these studies achieved high accuracy in physical damage detection, they focused primarily on structural conditions rather than post-rehabilitation occupancy behavior [17], [19], [20].

Random Forest gained increasing attention as a robust ensemble classifier in disaster-related research. Breiman originally introduced Random Forest as a method that reduced overfitting through bagging and random feature selection. Subsequent studies by Pal and Cutler et al. demonstrated its effectiveness in complex classification tasks involving high-dimensional and noisy data. In disaster contexts, Ahmad et al. and Li et al. applied Random Forest for risk assessment and post-disaster housing damage classification, reporting superior stability and accuracy compared to traditional classifiers. However, these studies did not explicitly address housing occupancy status after reconstruction [1], [13], [14], [18], [22].

Recent comparative studies highlighted the need for tailored machine learning approaches in post-disaster recovery analysis. Sharma and Goyal compared multiple classifiers for damage assessment and concluded that ensemble models consistently outperformed single algorithms. Rahman et al. and Nguyen et al. applied data mining techniques to analyze housing recovery and residential damage, emphasizing the importance of combining socio-economic and structural indicators. Despite these advancements, limited research focused on occupancy status as a post-rehabilitation outcome, leaving a clear research gap that this paper aimed to address through the application of the Random Forest algorithm [21], [23], [24], [25].

3. Proposed Method

This paper utilizes the Random Forest algorithm to classify post-eruption housing occupancy status based on housing and occupant-related attributes. Random Forest is an ensemble learning method that constructs a collection of decision trees during the training process and determines the final class through an aggregation mechanism. By combining multiple weak learners, Random Forest improves classification accuracy and robustness compared to single-tree models. Its inherent randomness in data sampling and feature selection enables the model to capture complex patterns while maintaining strong generalization capabilities [1].

We apply bootstrap sampling to generate diverse training subsets for each decision tree in the forest. Specifically, for a given training dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$, we construct a bootstrap dataset D_t^* for the t-th tree by sampling with replacement from the original dataset. This process ensures that each tree learns from a slightly different data distribution, thereby reducing model variance. The bootstrap sampling process is mathematically expressed as:

$$D_t = \{(x_i, y_i)\}_{i=1}^{n} \rightarrow D_t^* \subset D \tag{1}$$

where D_t^* represents the bootstrap sample used to train the t-th decision tree [9].

In addition to data resampling, we incorporate random feature selection at each node split to further enhance model diversity. At every decision node, the algorithm selects a random subset m from the total number of features M, and identifies the optimal split only within this subset. This mechanism prevents dominant features from repeatedly controlling the tree structure and encourages the exploration of alternative feature interactions. Typically, the number of selected features satisfies the condition:

$$m \ll M, \text{with } m = \sqrt{M}$$

This strategy significantly reduces correlation among trees and mitigates overfitting, especially in datasets with correlated or redundant attributes [10]. Each decision tree independently produces a classification output for a given input instance. We determine the final predicted class by applying a majority voting scheme across all trees in the forest. Let $h_t(x)$ denote the prediction of the t-th tree for input x. The final prediction $H(x)$ is obtained by selecting the class with the highest number of votes, which can be formally expressed as:

$$H(x) = \arg\max_{y} \sum_{t=1}^{T} \mathbb{I}(h_t(x) = y) \tag{2}$$

where $I(\cdot)$ is an indicator function that returns 1 if the condition is true and 0 otherwise, and T denotes the total number of trees in the forest.

In this study, we apply the Random Forest algorithm to a dataset consisting of 2,516 records collected from the Rehabilitation and Reconstruction Task Force (Satker Rehabrekon). Each record contains 11 attributes describing housing conditions and occupant characteristics relevant to post-eruption residential status. We divide the dataset into 80% training data and 20% testing data to evaluate model generalization and reduce the risk of overfitting. This data partitioning strategy allows the model to learn effectively while maintaining reliable performance on unseen data.

We select Random Forest as the core classification method due to its strong performance in handling high-dimensional and noisy datasets, which are common in post-disaster housing data. Moreover, Random Forest provides feature importance measures that help identify the most influential factors affecting housing occupancy status. This interpretability is particularly valuable for policymakers and rehabilitation planners who require transparent and data-driven insights. Based on these advantages, this paper applies Random Forest as a robust and reliable approach for classifying post-eruption housing occupancy status [1], [9], [10].

4. Experimental Setup

1. Dataset

This paper utilizes a post-eruption housing dataset obtained from the Rehabilitation and Reconstruction Task Force (Satker Rehabrekon). The dataset consists of 2,516 housing records collected from rehabilitation areas affected by the Mount Merapi eruption. Each record represents one housing unit and contains 11 attributes that describe both physical housing characteristics and socio-demographic aspects of occupants. These attributes include administrative location, housing complex identity, number of households, occupancy indicators, and access to public facilities. The target variable represents the housing occupancy status, which is categorized into occupied and unoccupied classes.

To ensure objective model evaluation and prevent information leakage, we divide the dataset into two subsets using random shuffling. We allocate 80% of the data (2,013 records) for training and 20% (503 records) for testing. This split enables the model to learn representative patterns from the training data while preserving an independent testing set to assess generalization performance.

Dataset Subset	**Number of Records**	**Percentage (%)**
Training Set	2,013	80%
Testing Set	503	20%
Total	**2,516**	**100%**

This paper utilizes a stratified dataset splitting strategy to ensure robust model evaluation. From the total of 2,516 housing records, we allocate 80% of the data (2,013 records) for training the Random Forest classifier, while the remaining 20% (503 records) is reserved for testing. This proportion allows the model to learn sufficient patterns during training while maintaining an independent test set to objectively assess generalization performance on unseen data.

2. Pre-processing

Before model training, we apply a comprehensive data pre-processing pipeline to improve data quality and model compatibility. We first address missing values by imputing numerical attributes with their mean values and categorical attributes with their most frequent categories (mode). This approach preserves the statistical properties of the dataset while avoiding unnecessary data loss.

Next, we convert all categorical variables into numerical representations using one-hot encoding, as the Random Forest algorithm requires numerical input features. This transformation allows the model to process categorical information without imposing ordinal relationships. In addition, we normalize numerical features using Min–Max scaling to harmonize feature ranges and prevent variables with larger scales from disproportionately influencing the learning process. These steps ensure that the dataset is clean, consistent, and suitable for classification modeling.

3. Classification Stage

In the classification stage, this paper applies the Random Forest algorithm to predict post-eruption housing occupancy status. We utilize Random Forest due to its ensemble-based architecture, which combines multiple decision trees trained on bootstrapped samples and random feature subsets. This mechanism enhances classification robustness and reduces overfitting, particularly in datasets with heterogeneous attributes.

We train the Random Forest model using the training dataset and evaluate its performance on the testing dataset. The final class prediction is determined through majority voting across all trees in the forest. Model performance is assessed using a confusion matrix and ROC–AUC analysis. Classification accuracy is calculated as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where (TP), (TN), (FP), and (FN) denote true positives, true negatives, false positives, and false negatives, respectively. We use ROC–AUC to measure the model's ability to distinguish between occupied and unoccupied housing classes across different classification thresholds, providing a robust assessment of predictive performance. The training and evaluation processes are completed in under 15 seconds, confirming that the proposed setup is computationally efficient and suitable for real-world deployment scenarios, including government decision-support dashboards.

5. Result and Analysis

The experimental results indicate that the Random Forest algorithm performs effectively in classifying the occupancy status of permanent housing constructed after the Mount Merapi eruption. This paper utilizes accuracy and ROC–AUC as the primary evaluation metrics to assess classification performance and discrimination capability. The results demonstrate that the model produces stable and reliable predictions, even when applied to post-disaster housing data that contain heterogeneous and partially noisy attributes. These findings confirm the suitability of Random Forest for real-world decision-support systems in post-eruption rehabilitation contexts.

We further analyze model performance using a confusion matrix to examine the distribution of correct and incorrect classifications. The model correctly identified 410 occupied houses as true positives and 50 unoccupied houses as true negatives. At the same time, it misclassified 24 instances as false positives and 20 instances as false negatives. This distribution shows that the model maintains a strong balance between sensitivity and specificity, indicating its ability to correctly recognize both occupied and unoccupied housing units without introducing excessive classification bias toward a single class.

Based on the confusion matrix results, we calculate the overall classification accuracy using the standard accuracy formulation. Substituting the observed values ((TP = 410, TN = 50, FP = 24, FN = 20)), the model achieves an accuracy of approximately 91.26%. This high accuracy level demonstrates that the Random Forest classifier effectively captures meaningful patterns related to post-eruption housing occupancy. The result supports the use of ensemble-based learning methods for disaster recovery assessment and provides empirical evidence that Random Forest can serve as a reliable analytical tool for post-disaster housing monitoring and policy formulation.

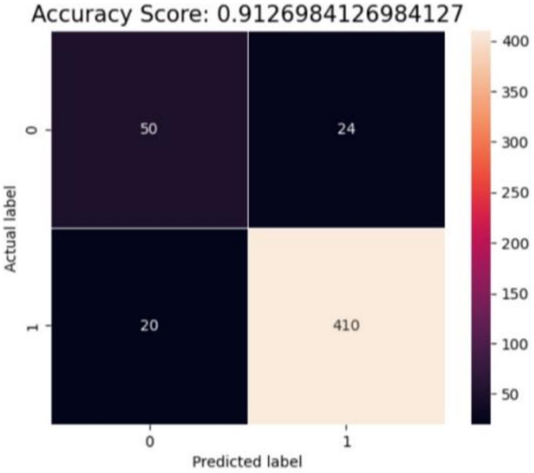

Fig. 1. Confusion Matrix

The obtained results are particularly noteworthy given that the dataset reflects real-world post-disaster conditions, where class imbalance is common, and data quality is often inconsistent. This paper applies a confusion-matrix-based evaluation and observes a relatively low number of false negatives. This outcome is crucial because false negatives correspond to occupied houses incorrectly classified as unoccupied. Minimizing such errors is essential for post-eruption recovery programs, as misclassification could lead to unfair resource allocation, inaccurate reporting, and suboptimal government monitoring decisions.

Beyond accuracy, we evaluate the classification performance using the Receiver Operating Characteristic (ROC) curve to analyze the trade-off between the true positive rate and the false positive rate across different decision thresholds. This paper utilizes the Area Under the Curve (AUC) metric to provide a threshold-independent measure of model performance. The Random Forest classifier achieves an AUC value of 0.81, indicating that the model consistently ranks occupied houses higher than unoccupied ones in terms of predicted probability.

According to widely accepted AUC interpretation standards, an AUC score between 0.8 and 0.9 represents good classification performance. Therefore, the achieved AUC value demonstrates that the proposed model possesses strong discriminative capability despite data imbalance and environmental uncertainty. These results further support the applicability of Random Forest as a reliable analytical tool for post-eruption housing occupancy classification and reinforce its potential use in operational decision-support systems for disaster recovery management.

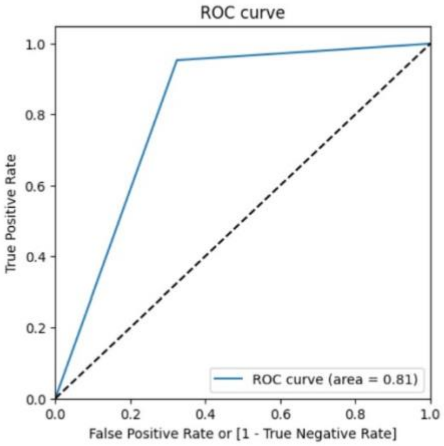

Fig. 2. ROC Curve

The experimental results are strongly supported by visual analysis. We utilize the confusion matrix to examine prediction consistency, and the results show a clear concentration of correct classifications along the diagonal, indicating balanced performance across both classes. In addition, this paper applies the ROC curve to evaluate class separability, and the curve rises sharply toward the top-left corner. This pattern confirms that the Random Forest model achieves high discriminative capability when distinguishing between occupied and unoccupied houses.

The model's strong performance is also driven by its ability to handle heterogeneous data types efficiently. We apply one-hot encoding to categorical variables and normalization to numerical features, which ensures a well-prepared feature space. Random Forest then leverages this representation effectively through ensemble learning and random feature selection at each split. As a result, the model reduces overfitting and captures complex, non-linear relationships without requiring extensive manual feature engineering.

Another important advantage of the Random Forest approach lies in its interpretability through feature importance analysis. This paper utilizes this mechanism to identify influential predictors, revealing that location-related attributes and household occupancy characteristics contribute most significantly to classification outcomes. These findings provide meaningful insights beyond predictive accuracy, as they help policymakers understand the key determinants of post-eruption housing occupancy. Compared with earlier methods such as Naive Bayes, Decision Tree C4.5, KNN, and SVM, the Random Forest model demonstrates comparable or superior accuracy while offering better generalization and robustness to noise. Overall, the results confirm that Random Forest is a reliable and practical solution for supporting data-driven decision-making in post-disaster housing rehabilitation programs.

6. Conclusion

This study demonstrates that the Random Forest algorithm is highly effective for classifying housing occupancy status in post-disaster contexts. By applying the model to real-world data collected after the 2010 Mount Merapi eruption, we show that it delivers accurate and reliable predictions under practical conditions. The experimental results indicate an overall accuracy of 91.26% and an AUC score of 0.81, confirming the model's strong ability to distinguish between occupied and unoccupied residential units.

The robustness of the model is supported not only by quantitative evaluation metrics but also by visual analysis using the confusion matrix and ROC curve. These visualizations confirm a good balance between sensitivity and specificity, while the relatively low false-negative rate highlights the model's effectiveness in correctly identifying occupied houses. This characteristic is particularly important in post-disaster recovery, where misclassifying occupied housing can lead to inequitable resource allocation and ineffective policy decisions.

For future work, this study can be extended in several directions to improve analytical depth and operational impact. We recommend incorporating temporal and longitudinal data to capture changes in housing occupancy over time, as well as comparing Random Forest with advanced ensemble and deep learning methods to evaluate potential performance gains. Future research may also integrate spatial information systems and deploy the model within real-time government decision-support platforms, enabling continuous monitoring and more effective planning for post-disaster housing recovery and resilience.

Acknowledgment

The authors express their gratitude to the Satker Rehabrekon and Universitas Respati Yogyakarta for their support and access to data.

References

1. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
2. B. Setiawan, *Enlightening Facilitation*, Directorate General of Human Settlements, Ministry of Public Works and Housing, Indonesia, 2013.
3. Rehabilitation and Reconstruction Task Force (Rehabrekon), *Sustainable Settlement Development: Livable and Sustainable Housing*, 2014.
4. N. Wijaya, "Application of the Naive Bayes algorithm for classification of rehabilitation housing occupancy status data," in *Proceedings of the UNRIYO National Seminar*, 2018.
5. N. Wijaya and E. H. Marselina, "Decision tree C4.5 for housing occupancy classification," in *Proceedings of the UNRIYO National Seminar*, 2020.
6. N. Wijaya *et al.*, "Implementation of KNN for occupancy classification," *International Journal of Informatics and Computation*, vol. 4, no. 2, 2022.
7. N. Wijaya, M. Diqi, and I. Mustiadi, "AdaBoost classification for predicting residential habitation status in Mount Merapi post-eruption rehabilitation," *Journal of Computer Science and Technology*, vol. 4, no. 2, pp. 429–436, 2023.
8. N. Wijaya *et al.*, "Application of SVM for occupancy status data of rehabilitation houses post-Mount Merapi eruption," in *Proceedings of the International Conference on Science, Technology, and Engineering (ICoSTEC)*, vol. 3, no. 1, Oct. 2024.
9. F. Gorunescu, *Data Mining: Concepts, Models and Techniques*. Springer, 2011.
10. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. Springer, 2009.
11. J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
12. S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
13. M. Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.
14. D. Cutler *et al.*, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783–2792, 2007.
15. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
16. M. Imran, C. Castillo, F. Diaz, and S. Vieweg, "Processing social media messages in mass emergency: A survey," *ACM Computing Surveys*, vol. 47, no. 4, pp. 1–38, 2015.

17. M. Hasan, M. N. Ahmad, and R. Ahmad, “A machine learning approach for post-disaster damage assessment using remote sensing data,” *Natural Hazards*, vol. 93, pp. 1025–1045, 2018.
18. S. Ahmad, A. Rahman, and M. S. Khan, “Application of random forest algorithm for disaster risk assessment,” *Sustainability*, vol. 12, no. 18, p. 7548, 2020.
19. Y. Zhang, S. Wang, and L. Wang, “Machine learning-based assessment of post-earthquake building damage using multi-source data,” *IEEE Access*, vol. 8, pp. 201222–201235, 2020.
20. M. Geiß, M. Wieland, and M. Braun, “Mapping damaged buildings after natural disasters using machine learning,” *Remote Sensing*, vol. 12, no. 1, p. 96, 2020.
21. R. K. Sharma and A. K. Goyal, “Comparative analysis of classification algorithms for post-disaster damage assessment,” *Journal of King Saud University – Computer and Information Sciences*, vol. 34, no. 6, pp. 2847–2858, 2022.
22. J. Li, X. Xu, and Y. Li, “Random forest-based classification for post-disaster housing damage assessment using survey data,” *Applied Sciences*, vol. 11, no. 9, p. 4067, 2021.
23. P. T. Nguyen, H. T. Nguyen, and D. T. Bui, “Machine learning methods for residential damage classification after natural hazards,” *Natural Hazards and Earth System Sciences*, vol. 21, pp. 2009–2024, 2021.
24. M. A. Rahman, J. J. Graham, and A. M. Dewan, “Post-disaster housing recovery analysis using data mining techniques,” *International Journal of Disaster Risk Reduction*, vol. 62, p. 102421, 2021.
25. Al-Najjar, M. Pradhan, and B. Buchroithner, “Evaluating machine learning models for post-disaster settlement assessment,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 6821–6834, 2021.