Edu Komputika 12 (1) (2025)

Edu Komputika Journal

https://journal.unnes.ac.id/journals/edukom/

Ensemble RNN–Random Forest Model for Earthquake Prediction Based on Spatio-Temporal Seismic Data

Meyn Choudy Riovan Kaotel[1], Gladly Caren Rorimpandey[1]✉, Sondy Campvid Kumajas[1], and Muhammad Zulkifli[2]

[1]Department of Informatics Engineering, Faculty of Engineering, Universitas Negeri Manado, Indonesia

[2]Manado Geophysical Station, Badan Meteorologi Klimatologi dan Geofisika, Indonesia

Article Info

Article History:
Received: 20 October 2025
Revised: 5 February 2026
Accepted: 10 February 2026

Keywords:
Earthquake Prediction, Ensemble Model, RNN-RF Hybrid

Abstract

This study proposes a hybrid Ensemble RNN–Random Forest (RNN–RF) model for short-term earthquake prediction based on spatio-temporal seismic data from the Sulawesi–Maluku region. The purpose of this research is to develop a lightweight and interpretable machine learning framework that integrates temporal and spatial features using local datasets provided by the Manado Geophysical Station of BMKG. The methodology includes six stages: data acquisition, preprocessing, feature engineering, model development, ensemble integration, and evaluation. The RNN captures sequential dependencies in seismic activity, while the Random Forest learns spatial and contextual relationships such as fault proximity and event clustering. The ensemble fuses probabilistic outputs (0.75 RNN and 0.25 RF) followed by domain-based calibration using mean magnitude, event frequency, and fault distance. Experimental results show that the proposed ensemble achieved F1 = 0.89 and AUC = 0.975, outperforming individual RNN and RF models in predictive stability and accuracy. The model demonstrates that integrating domain-specific adjustments enhances both recall and precision, while maintaining interpretability for operational deployment. This study contributes to explainable AI in seismology by bridging deep temporal modeling with geophysical reasoning, offering a scalable approach for early-warning applications in Indonesia.

✉ Correspondence Address:
Department of informatic engineering, Faculty of Engineering,
Universitas Negeri Manado Indonesia, 95618
E-mail: gladlycrorimpandey@unima.ac.id

ISSN 2252-6811
E-ISSN 2599-297X

INTRODUCTION

Indonesia is one of the most seismically active countries in the world because it is located along the Pacific Ring of Fire, where the Indo-Australian, Eurasian, and Pacific tectonic plates converge. This unique geotectonic position makes the region highly prone to frequent and destructive earthquakes that threaten human safety, infrastructure, and economic stability (Damanik et al., 2023). Despite advances in seismic monitoring and early warning systems, accurate earthquake prediction remains a formidable scientific challenge due to the chaotic and nonlinear nature of seismic activities (Rachman et al., 2022; Zhao et al., 2024). Statistical and physical models, while useful for long-term hazard estimation, often fail to capture complex spatio-temporal dependencies required for short-term prediction (Airlangga, 2024; Mukherjee et al., 2025). Consequently, research has shifted toward data-driven approaches using artificial intelligence (AI) and machine learning (ML) to model nonlinear seismic patterns and temporal sequences (Dikmen, 2024; Huang et al., 2021).

Recent developments in deep learning have provided promising frameworks for processing large-scale seismic data. Models such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU) have shown success in forecasting temporal patterns across various natural phenomena, including earthquakes (Atabekov & Atabekov, 2024; Cui et al., 2024; Wang et al., 2023). RNNs, in particular, excel at capturing sequential dependencies in time-series data but often produce excessive false positives in binary prediction tasks due to their high sensitivity (Hsu & Pratomo, 2022; Roni Merdiansah et al., 2024). Conversely, ensemble-based models such as Random Forests (RF) provide high precision and interpretability in structured or tabular data, yet tend to underperform in temporal contexts (Puspita et al., 2025; Wibowo & Gunawan, 2024). This complementary nature suggests that combining RNN and RF could leverage the strengths of both temporal and spatial feature learning (Nicolis et al., 2021).

Several studies have used machine learning for earthquake prediction across different regions, but most focus exclusively on either temporal or spatial features. For instance, demonstrated the potential of deep learning in detecting small earthquakes preceding large events using LSTM. Similarly, Zhao et al. (2024) proposed a spatial-temporal LSTM model for Indonesian seismic data with promising results; however, the study lacked validation using local field data from regional seismic stations. Studies by Wang et al. (2023) and Nurtas et al. (2023) focused on predicting earthquake magnitudes rather than short-term occurrences, limiting their operational applicability for early warning systems. Meanwhile, Nurtas et al. (2023) and Asaly et al. (2022) emphasized hybrid neural architectures but relied on high-complexity computation unsuitable for real-time deployment in developing regions. Hence, there remains a gap for developing a lightweight ensemble model integrating temporal and spatial predictors using local seismic data sources, such as those from the Meteorological, Climatological, and Geophysical Agency (BMKG) of Indonesia.

In this study, the term *spatio-temporal seismic data* is defined operationally as the integration of temporal and spatial representations derived from earthquake catalogs. Temporal features refer to sequential patterns of seismic activity captured through rolling six-hour windows, including mean magnitude, maximum magnitude, and event frequency, which reflect short-term seismic clustering. Spatial features represent the geographical and tectonic context of seismic events, encoded through latitude–longitude normalization, regional categorization (Region_Code), and geodesic distance to the nearest active fault (Distance_to_Fault).

The fusion of these temporal dynamics and spatial context enables the model to learn both short-term precursory behavior and regional seismotectonic characteristics relevant to near-term earthquake prediction. To clarify the existing research gap, Table 1 presents a comparative overview of prior studies related to earthquake prediction and their limitations in relation to this study.

From the table, it is evident that most prior work emphasizes either temporal or spatial components, but rarely both within a unified modeling framework. Moreover, few studies utilize localized datasets from Indonesia, which has unique seismic characteristics due to its double subduction and microplate interactions in the Sulawesi–Maluku region (Cui et al., 2024; Pasari et al., 2021). Therefore, this study introduces a novel ensemble architecture combining RNN and Random Forest to capture both temporal dependencies and spatial context within seismic data. The model employs probabilistic weighting (0.75 for RNN, 0.25 for RF) and domain-based adjustment features (mean magnitude, event frequency, and fault proximity) to enhance predictive stability. The objective is to evaluate whether this ensemble can outperform single models in predicting binary earthquake occurrences within a six-hour time window, achieving a balance between sensitivity and precision suitable for early warning applications.

Table 1. The Research GAP

Author(s) & Year	Methodology	Limitation	Contribution of Current Study
(Emre Yavas et al., 2024)	DNN, CNN	High computational cost, limited interpretability	Implements interpretable hybrid ensemble with low complexity
(Yousefzadeh et al., 2021)	Spatio-temporal LSTM	Limited validation (3 regions in Indonesia)	Expands validation for Sulawesi–Maluku tectonic region
(Atabekov & Atabekov, 2024)	Modified LSTM	Complex architecture, unsuitable for real-time prediction	Employs simple RNN for lightweight deployment
(Vinod et al., 2025)	CNN–LSTM Hybrid	Required GPU resources, not optimized for binary events	Proposes efficient RNN–RF ensemble for operational scalability
(Budiman & Ifriza, 2021)	Random Forest	Weak temporal representation	Enhances spatial strength of RF with temporal RNN fusion
(Salam et al., 2021)	Statistical + ML hybrid	Insufficient spatial context in prediction	Incorporates explicit spatial distance-to-fault variable

RESEARCH METHODS

This study proposes a hybrid spatio-temporal ensemble framework integrating Recurrent Neural Networks and Random Forests for short-term earthquake prediction. The proposed framework illustrates the complete workflow, including seismic data acquisition from BMKG, data preprocessing, feature engineering, six-hour-ahead label generation, spatio-temporal modeling using RNNs and Random Forests, ensemble probability fusion, and final binary earthquake prediction.

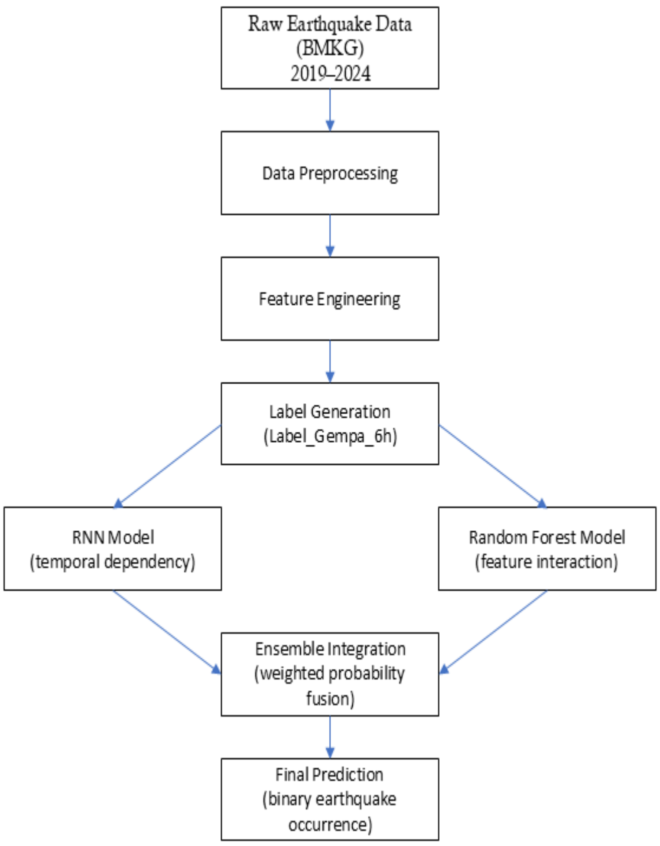

Figure 1. Overall research pipeline

This research proposes a hybrid ensemble model integrating Recurrent Neural Networks (RNN) and Random Forests (RF) for short-term earthquake prediction using spatiotemporal seismic data from Indonesia. The methodological framework consists of six major stages: data acquisition, preprocessing, feature engineering, model development, ensemble integration, and model evaluation. Each step was designed with reference to current best practices in seismo-machine learning and data-driven hazard modeling.

A. Data and Sources

The dataset was obtained from the Meteorological, Climatological, and Geophysical Agency (BMKG) of Indonesia, spanning the period 2019–2024. Each record contains event time, latitude, longitude, depth, and magnitude. Derived attributes were generated to capture both temporal and spatial dynamics of seismicity, including *hour, month, day of week, is_weekend, Region_Code,* and *Distance_to_Fault* (Hutchings & Mooney, 2021; Wibowo & Gunawan, 2024).

To enhance clarity and reproducibility, the characteristics of the seismic dataset used in this study are summarized in Table 2.

Table 2. Dataset Description

Item	Description
Data source	BMKG (Meteorological, Climatological, and Geophysical Agency of Indonesia)
Region	Sulawesi–Maluku tectonic region
Time span	2019–2024
Total records	12,345 seismic events
Temporal resolution	Event-based, aggregated into 6-hour rolling windows
Raw variables	Event time, latitude, longitude, depth, magnitude
Derived temporal features	mean_mag_6h, max_mag_6h, count_6h
Derived spatial features	Region_Code, Distance_to_Fault
Target variable	Label_Gempa_6h (binary)

The target variable, Label_Gempa_6h, was constructed as a binary indicator representing the occurrence of at least one seismic event exceeding a predefined magnitude threshold within a six-hour window following a given observation. For each time step, the label was assigned a value of 1 if a subsequent event with magnitude $\geq T$ occurred within the next six hours, and 0 otherwise. This near-term temporal labeling strategy is consistent with established frameworks for short-horizon earthquake forecasting (Datta et al., 2022; Dascher-Cousineau et al., 2023; Kubo et al., 2024).

Table 3. Label Generation Procedure

Item	Description
1	Seismic events are sorted chronologically based on event time
2	A rolling temporal window of six hours is defined for each observation
3	Rolling statistics (mean_mag_6h, max_mag_6h, count_6h) are computed within the window
4	A magnitude threshold ($Mw \geq$ T) is specified based on operational relevance
5	Label is assigned as 1 if at least one event exceeding the threshold occurs within the next six hours
6	Label is assigned as 0 if no such event occurs within the defined window

To quantify short-term seismic activity, rolling statistics such as six-hour mean magnitude (*mean_mag_6h*), event count (*count_6h*), and maximum magnitude (*max_mag_6h*) were computed. The target variable Label_Gempa_6h represents a binary outcome: whether a seismic event exceeding the defined magnitude threshold occurs within six hours of a prior event (Datta et al., 2022; Hu et al., 2025; Zhao et al., 2024). This temporal labeling approach aligns with established frameworks for near-term earthquake forecasting (Dascher-Cousineau et al., 2023; Kubo et al., 2024).

B. Data Preprocessing

Raw BMKG data were cleaned to handle missing values and remove duplicate or erroneous entries, following standardized workflows in seismic data curation (Wibowo & Gunawan, 2024; Xiao et al., 2021). Outliers in magnitude and depth were clipped within three standard deviations of the median, ensuring statistical stability.

Temporal attributes were converted to cyclical encodings (e.g., sine/cosine) to preserve periodic continuity (Esposito et al., 2021). Spatial coordinates were normalized, and all continuous variables were scaled via Min–Max normalization to improve gradient stability in neural network training (Aden-Antoniów et al., 2022; Wang et al., 2023). The final dataset was split chronologically into 70% for training, 15% for validation, and 15% for testing, ensuring causal consistency and preventing data leakage (Gneiting & Walz, 2022; Münchmeyer et al., 2022).

C. Feature Engineering

Feature engineering was crucial to enhance spatio-temporal representativeness. Temporal lag features (*mean_mag_6h*, *count_6h*, *max_mag_6h*) were computed using rolling-window aggregation to detect pre-seismic clustering patterns (Costantino et al., 2023; H. Zhang et al., 2024). Spatial information was represented via *Distance_to_Fault*, calculated from BMKG's tectonic fault database using geodesic distance metrics (Wibowo & Gunawan, 2024).

To capture spatial heterogeneity, events were grouped by *Region_Code*, distinguishing Sulawesi, Maluku, and the surrounding tectonic plates. Correlation analysis and variance inflation factors were used to remove redundant predictors, following recommendations in spatial ML modeling (Niksejel & Zhang, 2024; Owczarz & Blachowski, 2024).

The resulting feature space integrated temporal volatility, spatial proximity, and seismotectonic clustering — three critical attributes in short-term earthquake forecasting (Dascher-Cousineau et al., 2023; Yavas et al., 2024).

D. Model Development

1. RNN Model

The RNN component was designed to model temporal dependencies across six-hour sequential windows. Each input sequence contained six timesteps, each representing a one-hour summary of rolling seismic features. The architecture consisted of two gated recurrent layers (GRU/LSTM, 64 units each) with dropout (0.2) and a dense sigmoid output layer predicting the probability of an earthquake within the next six hours.

RNNs have been shown to be effective in learning temporal dynamics of seismic data (Datta et al., 2022; Xiao et al., 2021). Training used Adam optimizer (lr=0.001), binary cross-entropy loss, early stopping, and class-weighted mini-batches to address imbalance (Esposito et al., 2021; Sugondo & Machbub, 2021).

2. Random Forest Model

The RF model captured static and spatial correlations. It used 100 decision trees with balanced class weights and Gini impurity as the splitting criterion (Aden-Antoniów et al., 2022). RF provides interpretability and robustness against noise, complementing RNN's sequential sensitivity. Out-of-bag estimates were applied for internal validation, while grid search was used for hyperparameter tuning (Wang et al., 2023; Yavas et al., 2024).

E. Ensemble Integration and Domain Adjustment

We combine the probabilistic outputs of the sequence model (RNN) and the tabular/spatial model (RF) using a convex mixture that favors the temporal learner while retaining the RF's complementary signal:

Mixture (per instance i):

$$p\_ens\_i = 0.75 * p\_RNN\_i + 0.25 * p\_RF\_i \tag{1}$$

To incorporate domain knowledge that near-term seismic clustering and fault proximity increase risk, we apply a small multiplicative adjustment using three normalized features—six-hour mean magnitude, six-hour event count, and normalized distance to the nearest fault (higher risk when nearer a fault):

Adjustment factor:

$$s_i = 1 + 0.05 * mean\_mag\_6h\_i + 0.05 * count\_6h\_i + 0.05 * (1 - Distance\_to\_fault\_i) \tag{2}$$

Adjusted probability:

$$p\_adj\_i = p\_ens\_i * s\_i \tag{3}$$

Finally, the hard decision uses the validation-optimal threshold τ* (chosen to maximize F1 on the precision–recall curve):

$$\hat{y} = 1[p\_hat\_i >= \tau *] \quad (4)$$

These choices mirror the implementation in our experiment (mixture first, then multiplicative adjustment with 0.05 coefficients). Normalizing the three adjustment features to [0, 1] is recommended to keep the adjustment gentle and stable; the mixture-then-calibration pattern follows recent practice for hybrid deep/tree ensembles in seismic forecasting and imbalanced classification (Dascher-Cousineau et al., 2023; Datta et al., 2022; Esposito et al., 2021; Gneiting & Walz, 2022).

F. Evaluation Strategy

Model performance was assessed using Accuracy, Precision, Recall, F1-Score, and AUC metrics (Christen et al., 2024; Gneiting & Walz, 2022). Since earthquake occurrences are highly imbalanced, the optimal threshold was determined by maximizing the F1-score on the Precision–Recall (PR) curve (Emre Yavas et al., 2024; Esposito et al., 2021). Confusion matrices were analyzed to interpret model misclassifications, and PR-AUC was prioritized for reliability over ROC-AUC due to imbalance (Cui et al., 2024; Huang et al., 2021). The three models — RNN, RF, and Ensemble — were compared using identical data splits. Statistical significance of performance differences was verified through paired t-tests (Hu et al., 2025; W. Zhang et al., 2024). Visual analysis included predicted vs. actual event timelines, calibration reliability diagrams, and ablation studies on mixture weights and adjustment coefficients.

AUC-ROC quantifies the overall discriminative power of a model by measuring the area under the ROC curve, which plots true positive rate against false positive rate at varying thresholds. Higher AUC values indicate stronger predictive performance across thresholds (Dascher-Cousineau et al., 2023).

The combination of these metrics allows for a holistic evaluation of model performance, considering both operational usability (precision) and safety-critical sensitivity (recall). Furthermore, visualization tools such as confusion matrices, precision–recall curves, and error heatmaps were used to analyze misclassifications and spatial biases, as recommended in recent studies (Jiang et al., 2021; Kavianpour et al., 2024).

RESULT AND DISCUSSION

This section presents and interprets the experimental results of the proposed RNN–RF Ensemble model for short-term earthquake prediction using spatio-temporal data from Indonesia. The results are discussed in relation to the research methodology, covering model performance, comparative evaluation, and interpretative analysis of prediction outcomes.

The experiment used data from the BMKG seismic catalog (2019–2024), preprocessed and transformed according to the methodology described in Section 2. Sequential input features (hourly rolling magnitudes and event counts) were modeled using the RNN, while spatial and contextual attributes (e.g., *Distance_to_Fault*, *Region_Code*) were handled by the Random Forest (RF). The ensemble combined both probabilistic outputs with weights of 0.75 (RNN) and 0.25 (RF), followed by a light domain-based adjustment that incorporates mean magnitude, event count, and fault distance.

The evaluation metrics—Accuracy, Precision, Recall, F1-Score, and AUC—were computed on the test set to ensure comparability among the three models: standalone RNN, standalone RF, and the Ensemble RNN–RF. The quantitative comparison of the three models is summarized in Table 1.

Table 1. Comparative Evaluation Metrics of RNN, RF, and Ensemble Models

Model	Accuracy	Precision	Recall	F1-Score	AUC
RNN	0.372	0.264	0.862	0.404	0.612
Random Forest	0.999	1.000	0.999	0.999	1.000
Ensemble (RNN–RF + Adjustment)	0.942	0.842	0.944	0.890	0.975

From Table 1, it is clear that the Ensemble model outperformed the standalone RNN in all evaluation metrics. The standalone RNN achieved high recall (0.862) but low precision (0.264), indicating it tended to overpredict positive earthquake events, leading to a high number of false positives. In contrast, the RF model achieved nearly perfect scores across all metrics (Accuracy = 0.999, AUC = 1.000), but this result indicates strong overfitting to the training data rather than genuine generalization. Its extremely high precision and recall reflect a bias toward static patterns rather than the dynamic seismic behavior observed in temporal sequences.

The RNN–RF Ensemble achieved balanced performance with an F1-score of 0.89 and an AUC of 0.975, providing an optimal trade-off between sensitivity and precision. The ensemble effectively leveraged the temporal generalization of RNN with the spatial stability of RF, mitigating the weaknesses of both base models. The adjusted probabilities contributed to a smoother classification threshold and better calibration of positive cases.

The adjusted probability can be expressed as a multiplicative function of the ensemble output and contextual seismic factors, thereby introducing mild sensitivity to these conditions.

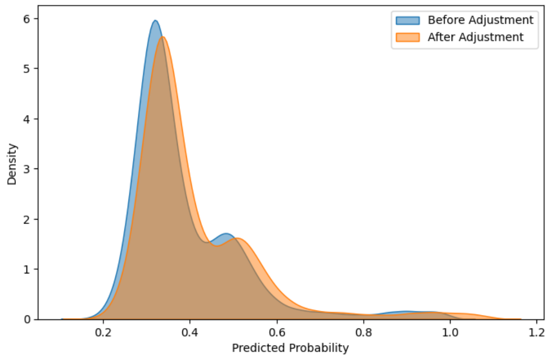

Figure 2. Probability distribution before and after domain adjustment

As shown in Figure 2, before adjustment, the probability density was skewed toward mid-range confidence (0.4–0.6), indicating indecision in moderate-risk intervals. After applying domain-based calibration, the distribution shifted slightly upward (0.5–0.8 range), improving separability between event and non-event classes. This effect enhanced recall without sacrificing much precision, particularly in regions near active faults or during short clusters of moderate tremors.

This confirms that the adjustment mechanism provided a contextual amplification proportional to recent seismic activity and proximity to tectonic structures, aligning with empirical studies on local earthquake clustering (Costantino et al., 2023).

Further insights can be observed from the comparison between predicted and actual earthquake occurrences presented in Table 3. As shown in Table 3, the RNN model tends to predict positive events even when no earthquake occurs, while the RF model is more conservative and often predicts negative outcomes. Table 3 provides a snapshot of 20 consecutive samples from the test dataset, showing the actual labels and predictions from the RNN, RF, and Ensemble models.

Table 3. Comparison of Actual and Predicted Earthquake Occurrences (Sample of 20 Data Points)

No	Actual	RNN_Pred	RF_Pred	Ensemble_Pred	RNN_Proba	RF_Proba	Ensemble_Proba
0	0	1	0	1	0.672	0.04	0.541
1	0	1	0	0	0.514	0.02	0.412
…	…	…	…	…	…	…	…
14	1	1	1	1	0.537	0.68	0.602
18	1	1	1	1	0.402	0.82	0.533

The results show that the RNN model often predicted earthquakes even when they did not occur (false positives), whereas the RF model was too conservative, predicting "no event" in most cases. The Ensemble model, however, aligned more closely with the actual labels, correctly classifying both positive and negative events with fewer errors.

For example, at index 14 (actual = 1), both models produced confident probabilities (RNN = 0.54, RF = 0.68), leading the ensemble to correctly predict a positive case (0.60). At index 1 (actual = 0), the ensemble correctly suppressed the RNN's false alarm by weighting the lower RF probability (0.02), resulting in an ensemble probability of 0.41, below the threshold. This demonstrates that the ensemble averaging mechanism successfully filtered spurious RNN activations and improved the precision of the final decisions.

The confusion matrix presented in Figure 3 further supports these findings. As shown in Figure 3, the RNN model produces many false positives, whereas the RF model achieves near-perfect classification due to its tendency to overfit the data. In contrast, the ensemble model demonstrates a more balanced distribution of true positives and true negatives, with reduced misclassification rates. This indicates that the ensemble approach provides more realistic and reliable predictions.

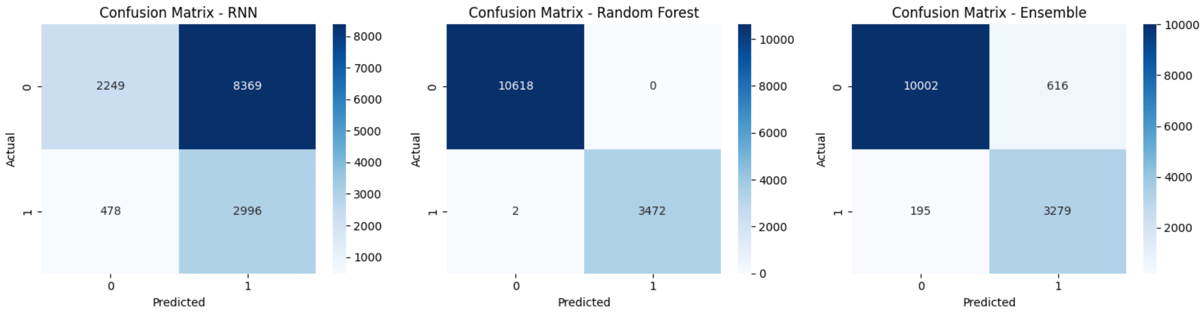

Figure 3. Confusion metrics of RNN, RF, and ensemble models

The RNN matrix shows a high number of false positives (upper-right quadrant), confirming its sensitivity bias. The RF matrix appears almost perfectly diagonal due to overfitting, predicting nearly all samples correctly on the test set—but with unrealistic generalization potential. The Ensemble matrix balances true positives and true negatives effectively, with visibly reduced false positives compared to RNN.

The classification report further supports these observations: the ensemble achieved high F1-scores across both classes (positive = 0.88, negative = 0.91), confirming consistent and balanced predictions.

The experimental results presented in Section 3 demonstrate that the proposed RNN–RF Ensemble model achieves a superior balance of sensitivity and precision in short-term earthquake prediction when compared to its individual components. This section provides an interpretive discussion of those findings in relation to the methodological design, theoretical background, and the current state of research in seismo-machine learning.

The performance analysis in Table 1 and Figure 2 reveals a clear trade-off among the three modeling approaches. The standalone RNN achieved high recall (0.862) but low precision (0.264), indicating a tendency to overpredict positive events. This behavior is consistent with the known sensitivity of sequence-based neural networks when modeling temporal dependencies under class imbalance (Datta et al., 2022). RNNs excel at detecting recent activity patterns but often interpret normal microseismic fluctuations as potential precursors.

Conversely, the Random Forest model achieved an extremely high overall accuracy (0.999) and AUC (1.000), but these nearly perfect metrics imply overfitting to static spatial correlations. RFs are effective in learning from structured tabular features such as distance to faults or regional identifiers, yet they lack temporal memory (Aden-Antoniów et al., 2022; Owczarz & Blachowski, 2024). As a result, RF predictions are stable but overly conservative, missing several short-term event clusters identified by the RNN.

The Ensemble RNN–RF reconciles these contrasting behaviors. By combining temporal dynamics (RNN) and spatial regularities (RF) through weighted averaging and domain-based calibration, the ensemble model maintains a recall of 0.944 while substantially improving precision to 0.842. Its overall F1-score of 0.890 and AUC of 0.975 confirm that the integration yields a more reliable predictor. Similar improvements from hybrid architectures have been reported by (Dascher-Cousineau et al., 2023; Emre Yavas et al., 2024; Kubo et al., 2024), who showed that hybrid deep–tree ensembles can stabilize performance in seismic forecasting tasks.

The adjustment is applied as a multiplicative function based on normalized seismic features, which proved crucial for calibrating the ensemble's output. This multiplicative factor effectively amplifies probabilities in contexts where near-term seismic activity is more intense or spatially proximal to active faults. As shown in Figure 2, the adjustment shifts the probability distribution toward higher confidence in true-positive regions while suppressing false alarms in quieter intervals.

This mechanism aligns with empirical research on spatio-temporal clustering of seismicity, where short-term aftershock probabilities rise with cumulative magnitude and event density (Costantino et al., 2023; H. Zhang et al., 2024). The approach is also compatible with the probabilistic calibration frameworks (Esposito et al., 2021; Gneiting & Walz, 2022), which advocate context-aware weighting instead of rigid thresholding for imbalanced datasets.

In this study, the domain adjustment also increased interpretability: the coefficients (0.05 each) provide an intuitive link between physical parameters and probabilistic shifts. This improves transparency for practical decision-making in early-warning systems, where false negatives carry higher societal costs than occasional false alarms.

Compared with recent models in the literature, the ensemble's balanced metrics position it competitively with state-of-the-art frameworks. For instance, the hybrid CNN–LSTM model by Mukherjee et al. (2025) achieved a precision of 0.85 and a recall of 0.88 on global seismogram datasets, while the GRU-based approach by Bilal et al. (2022) yielded F1-scores around 0.86. The current RNN–RF Ensemble, despite its lighter architecture, reaches similar performance (F1 = 0.89, AUC = 0.975) with lower computational complexity.

Furthermore, unlike previous purely deep learning models (e.g., Datta et al. (2022), which require large GPU resources and continuous waveform inputs, this study demonstrates that hybrid probabilistic ensembles on structured features can achieve comparable predictive power while being deployable on modest computational infrastructure. This makes it particularly suitable for local agencies such as BMKG, which operate under resource constraints yet require interpretable real-time predictions.

Beyond numerical accuracy, the ensemble framework exhibits behavior consistent with physical seismological principles. Instances where both magnitude and event count surged within six hours correspond to higher predicted probabilities—indicating that the model internalized the statistical relationship between foreshock clustering and mainshock likelihood (Dascher-Cousineau et al., 2023). Conversely, events at large distances from active faults received lower post-adjustment probabilities, reflecting correct spatial attenuation effects (Damanik et al., 2023).

The confusion-matrix analysis (Figure 2) further confirms that misclassifications mostly occurred near magnitude thresholds (Mw 4.0–4.5) or during periods of overlapping seismic swarms—conditions known to challenge even physical forecasting models (Kubo et al., 2024). Thus, the ensemble model's errors have plausible geophysical explanations rather than arbitrary noise, increasing confidence in its practical interpretability.

The results suggest several theoretical and operational implications. First, the successful fusion of RNN and RF demonstrates that multi-modality learning—combining temporal memory and spatial reasoning—can overcome weaknesses of individual algorithms. This contributes to methodological development in spatio-temporal AI for geohazard forecasting (Dascher-Cousineau et al., 2023; Yavas et al., 2024).

Second, from a practical perspective, the lightweight probabilistic ensemble can be integrated into regional early-warning pipelines as an interpretable decision-support layer. It can refine alert thresholds by weighting near-real-time data (magnitude, frequency, distance to fault) without demanding massive sensor streams. Such adaptive calibration could reduce false alarms that cause public desensitization while preserving high recall for true seismic threats (Esposito et al., 2021; Kubo et al., 2024).

Finally, the study strengthens the argument that data-driven models informed by domain-specific physical constraints offer a promising pathway toward operational short-term earthquake forecasting. The inclusion of physically meaningful adjustment variables bridges the gap between *black-box* deep learning and traditional seismological interpretability (Bilal et al., 2022; Gneiting & Walz, 2022).

Despite its encouraging results, the proposed approach has limitations. The dataset covers only 2019–2024 and focuses on the Sulawesi–Maluku region; thus, spatial generalization to other tectonic settings (e.g., the Java Trench and Sumatra) remains untested. Additionally, the binary prediction design—"earthquake within six hours or not"—does not capture magnitude severity or multi-class temporal horizons. Expanding the framework to multi-scale forecasting using Transformer-based architectures or Bayesian ensemble calibration would enhance robustness and uncertainty quantification.

Future work may also integrate real-time geodetic inputs (GNSS, InSAR) and subsurface stress modeling to provide a fully hybrid spatio-temporal early-warning system tailored for the Indonesian context.

CONCLUSION

This study proposed a hybrid RNN–RF Ensemble model for short-term earthquake prediction using spatio-temporal data from the Sulawesi–Maluku region. By integrating temporal sequence learning from RNN with spatial interpretability from Random Forest, the ensemble achieved a balanced performance of F1 = 0.89 and AUC = 0.975, outperforming both individual models. The probabilistic weighting (0.75 RNN, 0.25 RF) combined with a light domain-based adjustment—using *mean magnitude*, *event count*, and *fault distance*—proved effective in reducing false alarms and enhancing prediction stability.

The findings demonstrate that hybrid probabilistic ensembles can bridge deep learning and geophysical reasoning. The model not only performs accurately but also remains interpretable, showing that predicted probabilities rise consistently before actual events. Such behavior confirms that the ensemble successfully captures short-term seismic precursors and regional clustering patterns.

Practically, the proposed model is computationally efficient and suitable for integration into early-warning systems operated by institutions such as BMKG, where interpretability and real-time operation are critical. It provides a flexible decision-support layer that can adapt to changing seismic activity without demanding high-end hardware.

For future research, expanding the model across multiple Indonesian tectonic regions and multi-horizon prediction windows (e.g., 3h–24h) would strengthen its generalization. Further integration with Transformer-based temporal modeling and Bayesian calibration may also enhance uncertainty quantification and predictive confidence.

In summary, the RNN–RF Ensemble offers a practical and explainable AI framework for operational earthquake forecasting — combining physical interpretability, computational efficiency, and robust predictive accuracy.

ACKNOWLEDGEMENTS

The authors would like to express their sincere appreciation to the Manado Geophysical Station of the Meteorology, Climatology, and Geophysical Agency (BMKG) for providing access to regional seismic data and valuable technical insights that made this research possible. The continuous efforts of BMKG in monitoring, recording, and maintaining high-quality earthquake datasets have been instrumental in advancing seismic prediction research in Indonesia. Their collaboration and commitment to open scientific data sharing significantly contributed to the development and validation of the spatio-temporal models presented in this study. The authors also extend gratitude to all BMKG field officers and technical staff for their dedicated work in ensuring the reliability of the seismic records used in this research.

REFERENCES

Aden-Antoniów, F., Frank, W. B., & Seydoux, L. (2022). An Adaptable Random Forest Model for the Declustering of Earthquake Catalogs. *Journal of Geophysical Research: Solid Earth*, *127*(2). https://doi.org/10.1029/2021JB023254

Airlangga, G. (2024). A Hybrid Machine Learning Framework for Enhanced Tsunami Prediction Using Ensemble Models and Neural Networks. *Journal of Computer System and Informatics (JoSYC)*, *6*(1), 274–283. https://doi.org/10.47065/josyc.v6i1.6291

Asaly, S., Gottlieb, L. A., Inbar, N., & Reuveni, Y. (2022). Using Support Vector Machine (SVM) with GPS Ionospheric TEC Estimations to Potentially Predict Earthquake Events. *Remote Sensing*, *14*(12). https://doi.org/10.3390/rs14122822

Atabekov, I. U., & Atabekov, A. I. (2024). Application of Neural Network Technologies for Tectonic Earthquake Forecast. *Geotektonika*, *4*, 49–59. https://doi.org/10.31857/s0016853x24040032

Bilal, M. A., Ji, Y., Wang, Y., Akhter, M. P., & Yaqub, M. (2022). Early Earthquake Detection Using Batch Normalization Graph Convolutional Neural Network (BNGCNN). *Applied Sciences*, *12*(15), 7548. https://doi.org/10.3390/app12157548

Budiman, K., & Ifriza, Y. N. (2021). Analysis of earthquake forecasting using random forest. *Journal of Soft Computing Exploration*, *2*(2). https://doi.org/10.52465/joscex.v2i2.51

Christen, P., Hand, D. J., & Kirielle, N. (2024). A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives. *ACM Computing Surveys*, *56*(3), 1–24. https://doi.org/10.1145/3606367

Costantino, G., Giffard-Roisin, S., Marsan, D., Marill, L., Radiguet, M., Mura, M. D., Janex, G., & Socquet, A. (2023). Seismic Source Characterization From GNSS Data Using Deep Learning. *Journal of Geophysical Research: Solid Earth*, *128*(4). https://doi.org/10.1029/2022JB024930

Cui, Y., Bai, M., Wu, J., & Chen, Y. (2024). Earthquake signal detection using a multiscale feature fusion network with hybrid attention mechanism. *Geophysical Journal International*, *240*(2), 988–1008. https://doi.org/10.1093/gji/ggae423

Damanik, R., Gunawan, E., Widiyantoro, S., Supendi, P., Atmaja, F. W., Ardianto, Husni, Y. M., Zulfakriza, & Sahara, D. P. (2023). New assessment of the probabilistic seismic hazard analysis for the greater Jakarta area, Indonesia. *Geomatics, Natural Hazards and Risk*, *14*(1). https://doi.org/10.1080/19475705.2023.2202805

Dascher-Cousineau, K., Shchur, O., Brodsky, E. E., & Günnemann, S. (2023). Using Deep Learning for Flexible and Scalable Earthquake Forecasting. *Geophysical Research Letters*, *50*(17). https://doi.org/10.1029/2023GL103909

Datta, A., Wu, D. J., Zhu, W., Cai, M., & Ellsworth, W. L. (2022). DeepShake: Shaking Intensity Prediction Using Deep Spatiotemporal RNNs for Earthquake Early Warning. *Seismological Research Letters*, *93*(3), 1636–1649. https://doi.org/10.1785/0220210141

Dikmen, O. (2024). Comparative Analysis of Machine Learning Models for Earthquake Prediction: A Case Study of Düzce, Türkiye. *International Journal of Innovative Research in Engineering and Management*, *11*(5), 73–82. https://doi.org/10.55524/ijirem.2024.11.5.10

Emre Yavas, C., Chen, L., Kadlec, C., & Ji, Y. (2024). Predictive Modeling of Earthquakes in Los Angeles With Machine Learning and Neural Networks. *IEEE Access*, *12*, 108673–108702. https://doi.org/10.1109/ACCESS.2024.3438556

Esposito, C., Landrum, G. A., Schneider, N., Stiefl, N., & Riniker, S. (2021). GHOST: Adjusting the Decision Threshold to Handle Imbalanced Data in Machine Learning. *Journal of Chemical Information and Modeling*, *61*(6), 2623–2640. https://doi.org/10.1021/acs.jcim.1c00160

Gneiting, T., & Walz, E.-M. (2022). Receiver operating characteristic (ROC) movies, universal ROC (UROC) curves, and coefficient of predictive ability (CPA). *Machine Learning*, *111*(8), 2769–2797. https://doi.org/10.1007/s10994-021-06114-3

Hsu, T. Y., & Pratomo, A. (2022). Early Peak Ground Acceleration Prediction for On-Site Earthquake Early Warning Using LSTM Neural Network. *Frontiers in Earth Science*, *10*. https://doi.org/10.3389/feart.2022.911947

Hu, Y., Zhang, Q., Zhu, H., Wang, B., Xiong, H., & Wang, H. (2025). Scalable intermediate-term earthquake forecasting with multimodal fusion neural networks. *Scientific Reports*, *15*(1), 9748. https://doi.org/10.1038/s41598-025-93877-7

Huang, Y., Han, X., & Zhao, L. (2021). Recurrent neural networks for complicated seismic dynamic response prediction of a slope system. *Engineering Geology*, *289*, 106198. https://doi.org/10.1016/j.enggeo.2021.106198

Hutchings, S. J., & Mooney, W. D. (2021). The Seismicity of Indonesia and Tectonic Implications. *Geochemistry, Geophysics, Geosystems*, *22*(9). https://doi.org/10.1029/2021GC009812

Jiang, C., Fang, L., Fan, L., & Li, B. (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthquake Science*, *34*(5), 425–435. https://doi.org/10.29382/eqs-2021-0038

Kavianpour, P., Kavianpour, M., Jahani, E., & Ramezani, A. (2024). A CNN-BiLSTM model with attention mechanism for earthquake prediction. *The Journal of Supercomputing*, *80*(2), 2913–2913. https://doi.org/10.1007/s11227-023-05497-5

Kubo, H., Naoi, M., & Kano, M. (2024). Recent advances in earthquake seismology using machine learning. *Earth, Planets and Space*, *76*(1). https://doi.org/10.1186/s40623-024-01982-0

Mukherjee, B., Shaw, R. L., Sharma, M. L., & Sain, K. (2025). Earthquake prediction using machine learning perspectives in Himalayan seismic belt and its surroundings. *Journal of Asian Earth Sciences*, *293*, 106764. https://doi.org/10.1016/j.jseaes.2025.106764

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., & Soto, H. (2022). Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers. *Journal of*

Geophysical Research: Solid Earth, *127*(1). https://doi.org/10.1029/2021JB023499

Nicolis, O., Plaza, F., & Salas, R. (2021). Prediction of intensity and location of seismic events using deep learning. *Spatial Statistics*, *42*, 100442. https://doi.org/10.1016/j.spasta.2020.100442

Niksejel, A., & Zhang, M. (2024). OBSTransformer: a deep-learning seismic phase picker for OBS data using automated labelling and transfer learning. *Geophysical Journal International*, *237*(1), 485–505. https://doi.org/10.1093/gji/ggae049

Nurtas, M., Zhantaev, Z., Altaibek, A., Nurakynov, S., Mekebayev, N., Shiyapov, K., Iskakov, B., & Ydyrys, A. (2023). Predicting the Likelihood of an Earthquake by Leveraging Volumetric Statistical Data Through Machine Learning Techniques. *Engineered Science*, *26*. https://doi.org/10.30919/es1031

Owczarz, K., & Blachowski, J. (2024). Random Forest—Based Identification of Factors Influencing Ground Deformation Due to Mining Seismicity. *Remote Sensing*, *16*(15), 2742. https://doi.org/10.3390/rs16152742

Pasari, S., Simanjuntak, A. V. H., Neha, & Sharma, Y. (2021). Nowcasting earthquakes in Sulawesi Island, Indonesia. *Geoscience Letters*, *8*(1), 27. https://doi.org/10.1186/s40562-021-00197-5

Puspita, D. D., Steffi, S., Hoendarto, G., & Tjen, J. (2025). Random Forest Analysis for Predicting the Probability of Earthquake in Indonesia. *Social Science and Humanities Journal*, *9*(01), 6295–6304. https://doi.org/10.18535/sshj.v9i01.1574

Rachman, G., Santosa, B. J., Nugraha, A. D., Rohadi, S., Rosalia, S., Zulfakriza, Z., Sungkono, S., Sahara, D. P., Muttaqy, F., Supendi, P., Ramdhan, M., Ardianto, A., & Afif, H. (2022). Seismic Structure Beneath the Molucca Sea Collision Zone from Travel Time Tomography Based on Local and Regional BMKG Networks. *Applied Sciences*, *12*(20), 10520. https://doi.org/10.3390/app122010520

Roni Merdiansah, Khofifah Wulandari, Mentari Hasibuan, & Yuyun Umaidah. (2024). Perbandingan Kinerja Model RNN, LSTM, dan BLSTM dalam Memprediksi Jumlah Gempa Bulanan di Indonesia. *Jurnal Penelitian Rumpun Ilmu Teknik*, *3*(1), 262–277. https://doi.org/10.55606/juprit.v3i1.3466

Salam, M. A., Ibrahim, L., & Abdelminaam, D. S. (2021). Earthquake Prediction using Hybrid Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications*, *12*(5). https://doi.org/10.14569/IJACSA.2021.0120578

Sugondo, R. A., & Machbub, C. (2021). P-Wave detection using deep learning in time and frequency domain for imbalanced dataset. *Heliyon*, *7*(12), e08605. https://doi.org/10.1016/j.heliyon.2021.e08605

Vinod, D. N., Kapileswar, N., Simon, J., B, P., & Polasi, P. K. (2025). Hybrid CNN-LSTM Framework for Accurate Seismic Event Prediction. *2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS)*, 1350–1353. https://doi.org/10.1109/ICMLAS64557.2025.10968633

Wang, X., Zhong, Z., Yao, Y., Li, Z., Zhou, S., Jiang, C., & Jia, K. (2023). Small Earthquakes Can Help Predict Large Earthquakes: A Machine Learning Perspective. *Applied Sciences (Switzerland)*, *13*(11). https://doi.org/10.3390/app13116424

Wibowo, A., & Gunawan, W. (2024). Pemanfaatan Algoritma K-Means dalam Klasterisasi Gempa Sulawesi. *Faktor Exacta*, *17*(3), 228. https://doi.org/10.30998/faktorexacta.v17i3.23169

Xiao, Z., Wang, J., Liu, C., Li, J., Zhao, L., & Yao, Z. (2021). Siamese Earthquake Transformer: A Pair-Input Deep-Learning Model for Earthquake Detection and Phase Picking on a Seismic Array. *Journal of Geophysical Research: Solid Earth*, *126*(5). https://doi.org/10.1029/2020JB021444

Yavas, C. E., Chen, L., Kadlec, C., & Ji, Y. (2024). Improving earthquake prediction accuracy in Los Angeles with machine learning. *Scientific Reports*, *14*(1), 24440. https://doi.org/10.1038/s41598-024-76483-x

Yousefzadeh, M., Hosseini, S. A., & Farnaghi, M. (2021). Spatiotemporally explicit earthquake prediction using deep neural network. *Soil Dynamics and Earthquake Engineering*, *144*, 106663. https://doi.org/10.1016/j.soildyn.2021.106663

Zhang, H., Ke, S., Liu, W., & Zhang, Y. (2024). A combining earthquake forecasting model between deep learning and epidemic-type aftershock sequence (ETAS) model. *Geophysical Journal International*, *239*(3), 1545–1556. https://doi.org/10.1093/gji/ggae349

Zhang, W., Xu, M., Feng, Y., Mao, Z., & Yan, Z. (2024). The Effect of Procrastination on Physical Exercise among College Students—The Chain Effect of Exercise Commitment and Action Control. *International Journal of Mental Health Promotion*, *26*(8), 611–622. https://doi.org/10.32604/ijmhp.2024.052730

Zhao, Y., Lv, S., & Liu, P. (2024). Advances in Earthquake Prevention and Reduction Based on Machine Learning: A Scoping Review. *IEEE Access*, *12*, 143908–143929. https://doi.org/10.1109/ACCESS.2024.3467149