

Monte Carlo Aggregating Method for Radon Precursor-Based Earthquake Parameter Prediction in Java, Indonesia



Wahyu Sukestyastama Putra^{1,2}, Sunarno^{3*}, I Wayan Mustika¹

¹ Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

² Faculty of Computer Science, Universitas Amikom Yogyakarta, Yogyakarta 55283, Indonesia

³ Department of Nuclear Engineering and Engineering Physics, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia

Corresponding Author Email: sunarno@ugm.ac.id

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijss.150217>

ABSTRACT

Received: 30 November 2024

Revised: 27 January 2025

Accepted: 13 February 2025

Available online: 28 February 2025

Keywords:

earthquake prediction, radon, Monte Carlo Aggregating (MCA)

This study introduces a novel method for predicting earthquake parameters using radon as a precursor to address uncertainties and limitations in the dataset. The dataset comprises radon observation data from Yogyakarta, Indonesia, and earthquake records collected from a radon monitoring site and the USGS earthquake database from December 11, 2022, to August 8, 2023. The proposed method was trained on 80% of the dataset, which was utilized to generate a probability distribution for the Monte Carlo process to handle the constraints of limited precursor data. The results from the Monte Carlo simulations were then used to develop a model for predicting earthquake parameters. Experimental results demonstrate that the proposed method performs well within a monitoring station's radius of 300 and 400 km. At 300 km, the method outperforms in predicting magnitude, distance, and time, with RMSE values of 0.48, 60.60 km, and 57.85 hours, respectively. At 400 km, it achieves excellent performance with RMSE values of 0.61, 76.29 km, and 46.69 hours. This study shows that the proposed method outperforms benchmark methods in predicting earthquake parameters using radon gas as an earthquake precursor.

1. INTRODUCTION

Earthquake early prediction technology is an exciting innovation to develop, especially in Indonesia. Indonesia is at the meeting point of the world's active tectonic plates, causing high seismic activity. Earthquakes with large magnitudes have the potential to create damage that impacts the lives of affected communities. If not anticipated, infrastructure damage, chaos, and even loss of human life are examples of the impacts of natural disasters. *Early prediction technology* is built to provide earlier information before an earthquake occurs. With this technology, mitigation can be achieved to reduce the impact of the disaster.

Before an earthquake occurs, several natural phenomena can be used as precursors. Potential earthquake precursors that can be used are the Earth's magnetic field anomalies, radon gas emissions, and soil temperature. Radon is a naturally occurring radioactive gas found in Earth's crust. Radon is a potential seismic precursor because it is released from the Earth's cavity during a seismic event. The radon gas anomaly is internationally recognized as one of the seven seismic precursors [1]. One significant advantage is its ability to serve as a short-term pre-seismic indicator due to its rapid emission changes preceding earthquakes [2].

In comparison to other seismic precursors like electromagnetic anomalies and soil temperature, radon emissions provide a distinct geochemical signal directly linked to subsurface stress changes [3]. This geochemical nature

allows radon monitoring to detect stress accumulations that may not produce immediate electromagnetic responses or soil temperature changes. Moreover, when used in conjunction with other precursors, radon data can enhance the robustness of earthquake forecasting models by providing complementary information [3]. The monitoring of radon gas concentrations for earthquake precursors has been established as a valuable approach given the abundance of this radioactive gas in groundwater with a short half-life, making it a suitable indicator for seismic activity [4]. However, the reliability of radon as a sole predictor is challenged by its sensitivity to environmental factors such as precipitation and atmospheric pressure changes, which can obscure seismic-related anomalies [5].

Radon anomalies observed before earthquakes have been a focal point of research, with findings suggesting that changes in radon release rates could be key precursory phenomena for earthquakes [6]. Previous studies have shown that changes in radon concentrations in groundwater and soil can be an early sign of an earthquake, because sudden changes in radon concentrations often occur before seismic events [7]. This is reinforced by many recent studies that showed changes in radon concentrations before major earthquakes [8]. Radon gas emissions can be a short-term pre-seismic precursor [9]. Radon concentration anomalies indicate impending earthquakes. This is reinforced by research that highlights the correlation between radon gas concentrations, meteorological data, hydrology, and seismic activity [10].

Radon gas concentration fluctuations can also be used as medium and long-term indicators to reveal abnormal underground air changes before an earthquake occurs [2]. Radon-based earthquake precursors in Indonesia have also been reported [11, 12]. The phenomenon of anomalous radon gas concentrations before the 5.6 M earthquake on June 8, 2023, south of Java Island, Indonesia, is shown in Figure 1.

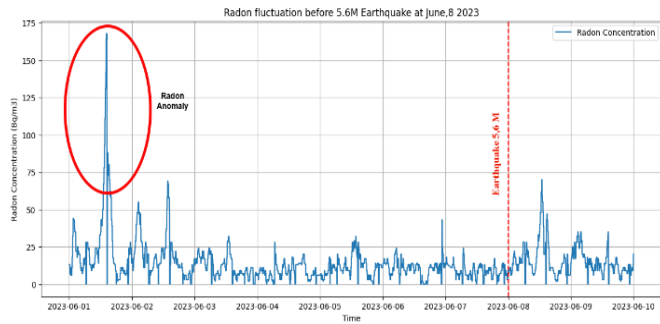


Figure 1. Radon concentration anomaly phenomenon

Earthquake parameter prediction using artificial intelligence (AI) technology is being increasingly developed along with advances in the field of computing. In particular, the application of machine learning as a branch of AI has great potential for the development of prediction algorithms that can provide more accurate and faster results. However, one of the main challenges in applying machine learning to earthquake prediction is the requirement for a large amount of high-quality data. These data are required for the model training process and for validating the accuracy of the algorithm. An earthquake is a natural event that occurs suddenly and in limited quantities; therefore, collecting sufficient data for machine learning is a challenge. In addition, specifically for data on radon gas concentration fluctuations, one of the precursors of earthquakes, this data collection takes a long time because significant changes in radon concentration occur only before an earthquake.

Earthquake event simulation efforts have been conducted previously [13-15]. Simulations were conducted using the Monte Carlo method. This work modified the machine learning method using the Monte Carlo approach to predict earthquakes based on radon gas precursors. This study is expected to increase the accuracy of predictions, so that earthquake early warning technology can be created to benefit the community. This study modified the limited training data to develop a more accurate prediction model. This modification was performed by increasing the training data using the Monte Carlo method to produce variations close to real conditions. The training data generated from this method is then used to build a more robust prediction model. The prediction model was then subjected to an aggregation process to produce more reliable prediction results.

The main contribution of this study is to solve the issue of limited datasets to improve the prediction accuracy of earthquake parameters based on radon gas precursors. This paper is divided into four sections. The first Section, the Introduction, discusses the background and objectives of the problem. The second Section, Related Work, provides a review of previous studies that have been conducted in earthquake prediction. Section Proposed Method describes the steps for implementing the Monte Carlo method and aggregation techniques. The fourth Section, Results and Discussion, presents the simulation results using radon-based

earthquake precursor data and the analysis of these results. Finally, the Conclusion Section summarizes the main findings of this study.

2. RELATED WORKS

Several researchers have reported radon gas fluctuations before earthquakes. Muto et al. [16] conducted observations and concluded that there was a decrease in radon gas concentration before the 2018 North Osaka earthquake. Attanasio and Maravalle [17] also revealed a relationship between radon emissions and earthquakes in Italy. However, the use of this precursor to predict the magnitude of earthquakes is still a challenge. D'Incecco et al. [2] have also designed a real-time monitoring system for radon gas to study earthquake prediction. Efforts to utilize radon precursors for prediction were carried out by Feng et al. [18]. The EMD-LSTM method was developed for the early detection of earthquakes using groundwater radon monitoring. The results showed that the developed method could predict earthquakes early. These studies have revealed the potential of radon as an important precursor in earthquake prediction.

Radon concentration fluctuations can be classified into three types: single (sudden spike), multiple, and persistent [19]. Three types of radon concentration fluctuations can be used as earthquake precursors: single fluctuations, multiple fluctuations, and persistent fluctuations. Single fluctuations refer to situations in which data samples experience significant fluctuations in a relatively short period. These fluctuations were characterized by one prominent peak. In contrast, multiple fluctuations include sudden changes in the value that occur over a certain time interval. These fluctuations were characterized by several prominent peaks. Persistent fluctuations describe situations where sample data show wide fluctuations over a long period of time and consistently exceed abnormal thresholds.

Machine learning methods have made significant contributions to earthquake prediction based on radon precursors. Mir et al. [20] analyzed the performance of several machine learning methods, such as boosted tree, bagged cart, linear model, support vector machine, and k-nearest neighbor, to predict anomalies in radon time-series data related to seismic activity. The results show that the boosted tree and support vector machine with radial kernel are better models for predicting anomalies in soil radon gas concentration during seismic activities. Zhu et al. [21] also used machine learning algorithms to detect anomalies in the hydrochemical data of hot springs to predict earthquakes based on radon concentration data. Wang et al. [22] explored the use of machine learning methods for seismic hazard evaluation. The exploration results showed that small earthquakes can help predict larger earthquakes. Jarah et al. [23] applied the Random Forest method to identify factors that precede earthquakes. Furthermore, Gitis and Derendyaev [24] and Deb et al. [25] discussed the use of machine learning methods for seismic hazard forecasting. These studies have shown that machine learning methods are popular for the early detection of radon-based earthquakes.

The integration of machine learning with radon monitoring for earthquake prediction has yielded promising results. Tehseen et al. [26] mapped various expert systems for earthquake prediction and analyzed research evolution. Asim et al. [27] developed a Support Vector Regression and Hybrid

Neural Network (SVR-HNN) model for earthquake prediction. The results show that the SVR-HNN model enhanced earthquake prediction capabilities in the Hindukush, Chile, and Southern California regions. The main challenge in machine learning methods is the requirement of large amounts of data to train the model. The development of accurate earthquake prediction models is obstructed by the limited availability of radon-precursor datasets. Machine learning models require larger datasets to identify significant patterns associated with seismic events.

Unfortunately, radon precursor datasets used for earthquake prediction are often limited. Researchers have used Monte Carlo to predict potential earthquake and tsunami hazards. Goda and Song [14] used Monte Carlo analysis simulation to assess the characteristics of earthquake sources in tsunami risk in the 2011 Tohoku tsunami, especially in the Rikuzentakata area, Japan. This study emphasized the importance of the uncertainty approach in tsunami risk models to make predictions more realistic and accurate. Muhammad et al. [15] used Monte Carlo simulation (MCS) techniques together with the Autoregressive Integrated Moving Average (ARIMA) method to find a significant relationship between radon gas anomalies and micro-seismic activity in the North Anatolian Fault Zone (NAFZ) of Turkey. The study showed a strong correlation between radon gas (Rn-222) concentration anomalies in the soil and micro-seismic activity around the fault zone. Meanwhile, Crowley and Bommer [13] used the Monte Carlo method to create multi-earthquake scenarios that generate ground motion based on seismicity models. The Monte Carlo method in the study provides advantages in modeling earthquake losses in more detail. The research shows that the Monte Carlo method can be used to overcome data limitations by simulating various scenarios using probability distributions.

One of the challenges in earthquake prediction is the need for more representative data [28]. Predicting earthquakes using precursors faces challenges due to data limitations and uncertainties. In this study, the Monte Carlo method is proposed as an alternative solution to address issues related to uncertainty and dataset limitations. The application of the Monte Carlo method to enhance data quantity for earthquake prediction using machine learning has yet to be explored in previous studies.

3. PROPOSED METHOD

3.1 Monte Carlo Aggregating

Fluctuations in radon gas concentrations at the radon monitoring station prior to the earthquake exhibited a sudden and significant increase. The gain at the monitoring station was defined as the ratio of the peak fluctuation value to the average radon gas concentration measured over the seven days preceding the fluctuation. This station gain value was used to predict the earthquake parameters. A visualization of the station gain is shown in Figure 2. Figure 2 shows that a single fluctuation occurred, which has been identified as a precursor to earthquakes [20]. The gain of the monitoring station is expressed by Eq. (1).

$$G_s = \frac{R_{\max}}{\frac{1}{n} \sum_{i=0}^n R_{t-i}} \quad (1)$$

where,

- G_s = Gain of the monitoring station,
- R_{\max} = Maximum Radon value (Bq/m³),
- R_{t-i} = Daily average Radon value (Bq/m³).

In this study, an algorithm was designed to predict earthquake parameters using the gain of the Radon Monitoring Station. Owing to the random nature of earthquakes, with a distribution that is challenging to determine, the algorithm was required to account for this randomness and uncertainty. One approach employed was the use of a Monte Carlo simulation. A simulation was conducted to obtain all possible output values of the prediction. The illustration of the Monte Carlo approach is presented in Figure 3.

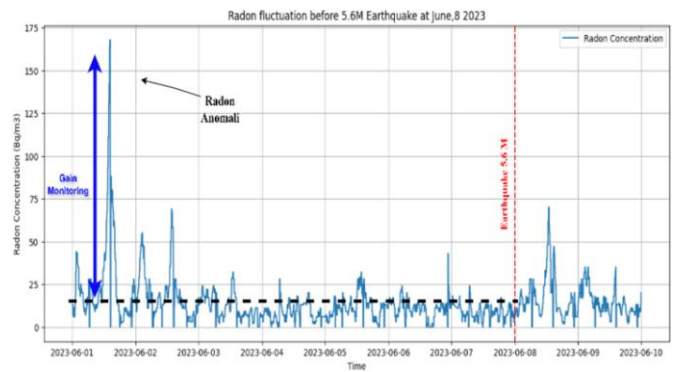


Figure 2. Gain monitoring of radon concentration anomaly

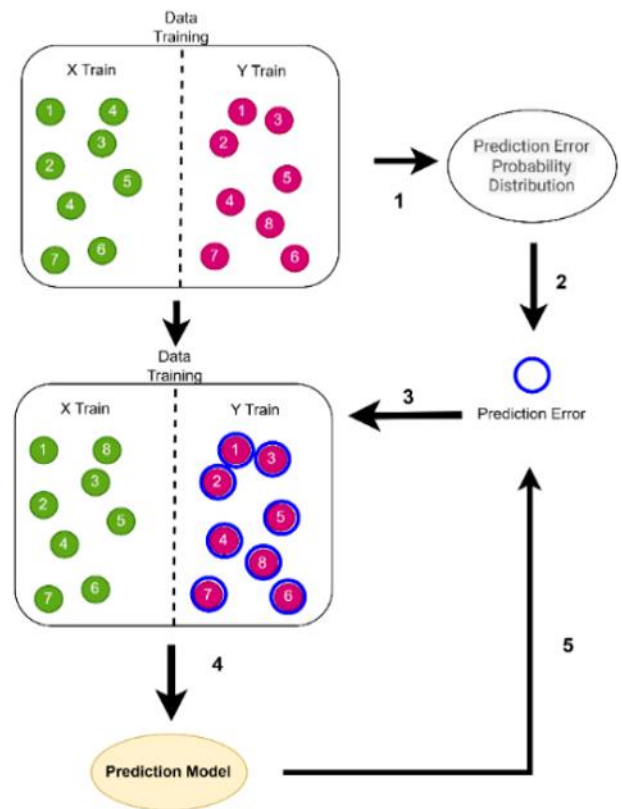


Figure 3. Monte Carlo approach in data training

The Monte Carlo method applied in this study utilized probability distributions derived from historical data of radon concentration anomalies and seismic activity. To ensure that the generated synthetic data closely aligned with observed patterns, an empirical distribution was selected, providing a more accurate representation of the variability present in real-world data. Random samples were subsequently generated based on the empirical distribution to simulate variations in prediction errors of earthquake parameters. These samples were specifically designed to replicate the potential uncertainties inherent in earthquake prediction. The proposed method was trained on historical data, ensuring that the simulations accurately captured the relationships present in the observed data. This approach was implemented to ensure the robustness and relevance of the Monte Carlo simulations for earthquake prediction, grounding the methodology in empirical data and realistic patterns to enhance its reliability.

The first step in the proposed method involved creating a probability distribution of prediction errors. At this stage, the prediction error (e) was obtained at this stage by measuring the difference between the output training data (Y train) and the predicted value using a prediction model with the input training data (X train). This probability distribution is then formed into a probability mass function ($f(e)$) expressed in Eq. (2).

$$f(e_i) = \begin{cases} P_1 & \text{for } e_1 \\ P_2 & \text{for } e_2 \\ \dots & \dots \dots \\ P_m & \text{for } e_m \end{cases} \quad (2)$$

P is the probability of a prediction error occurring, and i is an index with values ranging from 1 to m . The probability mass function was then used to create random prediction errors. The probability mass function is then used to form the cumulative probability function ($F(e)$), which is written as:

$$F(e_i) = \begin{cases} P_1 & \text{for } e_1 \\ P_1 + P_2 & \text{for } e_2 \\ \dots & \dots \dots \\ 1 & \text{for } e_m \end{cases} \quad (3)$$

The random number U (Uniform(0,1)) determines the prediction error value. If the random number value is $P_1 \leq U \leq P_1 + P_2$, then the prediction error value is e_2 . The resulting prediction error was then added to the output training data (Y train). The prediction model is then created using the Y train, which has added errors. The prediction model can then predict the Y -test with the input X -test.

Given that an earthquake is a random event with significant uncertainty in its prediction, the Monte Carlo approach was implemented to enhance prediction accuracy. Prediction errors, which were generated randomly through n iterations, produced n prediction models, allowing the prediction output to be approximated as the average of the predicted outputs from each model. By simulating all possible outcomes based on the probability distribution, the predictive capability was expected to improve. Based on this approach, the Monte Carlo Aggregating (MCA) method for earthquake prediction was developed. The Monte Carlo aggregation method is drawn in Figure 4.

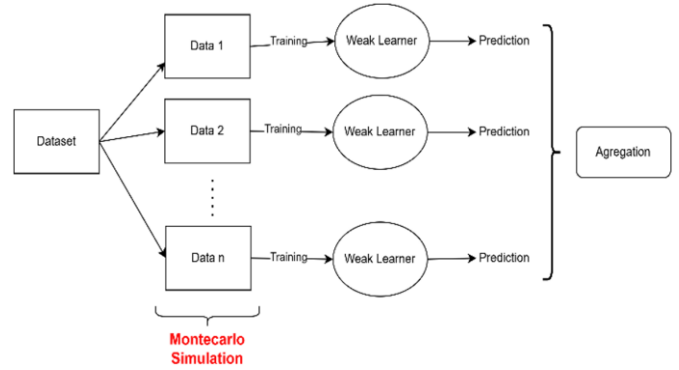


Figure 4. Monte Carlo Aggregating method

3.2 Data collection

The monitoring station was located in the Special Region of the Yogyakarta Province (lat. -7.7531627 S, long. 110.4215244 E), Indonesia. This area is located in the central part of Java. Java Island is adjacent to the subduction zone at the bottom of the Indian Ocean along the southern coast of Java. Therefore, the island is prone to earthquakes and tsunamis. The sensor used is the ion chamber RD200 sensor, and radon data can be accessed at <http://dataalamdiy.com/dataview/>, whereas earthquake data can be accessed from <https://earthquake.usgs.gov/earthquakes/map/>. The data used were earthquake and radon gas data from December 11, 2022, to August 8, 2023. The data were collected based on the radius of the earthquake event at the radon monitoring station. The earthquake profile used in this study is summarized in Table 1.

Table 1. Earthquake profile

		Radius (km)		
		200	300	400
Magnitude (M)	Min	4	4	4
	Max	5.9	7	7
	mean	4.72	4.62	4.59
Distance of earthquake event from monitoring station (km)	Min	92.87	92.87	92.87
	Max	197.21	285.85	393.19
	mean	155.12	218.88	256.47
Earthquake occurrence time after Anomaly (hr)	Min	3	2	3
	Max	164	168	168
	mean	72.33	87.06	88.282
Number of Earthquake Events		12	32	46

The earthquake data used in this study are presented in Table 1. This table provides a summary of seismic events based on three radial distances from the monitoring station: 200, 300, and 400 km. The table shows that the minimum earthquake magnitude recorded was 4, with the maximum magnitude being 5.9 at a radius of 200 km and 7 at radii of 300 km and 400 km. The distance of the closest earthquake event to the monitoring station was 92.87 km, while the maximum distance was 197.21 km at a radius of 200 km, 285.85 at a radius of 300 km and 393.19 km at a radius of 400 km. The time of occurrence of earthquakes after the anomaly with minimum time was recorded between 2 h and a maximum time of 168 h. Meanwhile, the number of earthquake events recorded was 12 at 200 km, 32 at 300 km, and 46 at 400 km.

3.3 Evaluation and validation

The prediction evaluation was performed using the RMSE metric [29]. RMSE is the square root of the mean of the squares of the prediction errors, which provides a measure of the error in native units. RMSE is expressed by:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{N}} \quad (4)$$

\hat{Y}_i is the predicted value, Y_i is the recorded earthquake parameter value and N denotes the number of data points to be evaluated. Model validation was performed using a 5-fold cross-validation technique. The average RMSE is used to evaluate the performance of the proposed method.

3.4 Baseline method

One of the challenges in earthquake prediction using radon gas as a precursor is the lack of a representative radon precursor dataset and appropriate baseline methods that could

help researchers compare their approaches and results [28]. To compare our study, we selected a baseline regression method from previous research based on performance results and data characteristics, enabling our approach to be applied to other studies. Random Forest has been previously used for earthquake prediction and produced good results [30]. In earlier studies, XGBoost [31] and linear regression [32] were also developed to predict earthquake events.

4. SIMULATION RESULT AND DISCUSSION

The proposed method has been implemented on a test dataset. The proposed method is then compared with the baseline method. This study compares several prediction methods' root mean square error (RMSE) values for three different parameters. The prediction parameters used are Magnitude (Mag), Distance (Dis), and Time for three different prediction radii: 200 km, 300 km, and 400 km. The methods compared are Linear Regression (LR), Random Forest (RF), XG-Boost (XGB), and MCA. The performance of the proposed methods is summarized in Table 2.

Table 2. Earthquake prediction performance

		RMSE Prediction			
		Linear Regression	Random Forest	XGBoost	Monte Carlo Aggregating
200 km Radius	Mag (M)	0.61	0.49	0.66	0.59
	Dis (km)	28.86	13.25	26.85	28.16
	Time (Hour)	45.62	57.29	63.90	41.98
300 km Radius	Mag (M)	0.48	0.55	0.57	0.48
	Dis (km)	62.88	68.39	81.29	60.60
	Time (Hour)	60.06	65.20	72.60	57.85
400 km Radius	Mag (M)	0.63	0.67	0.78	0.61
	Dis (km)	77.84	100.13	107.97	76.29
	Time (Hour)	47.18	57.60	63.75	46.69

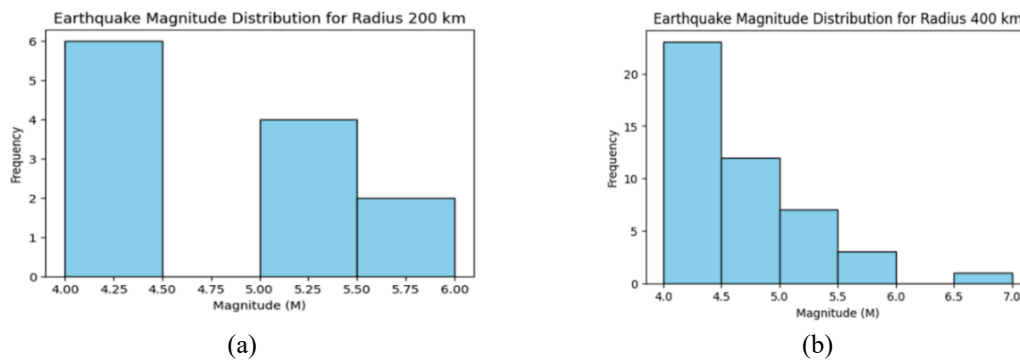


Figure 5. Frequency distribution of earthquake magnitudes for 200 km radius (a) and 400 km radius (b)

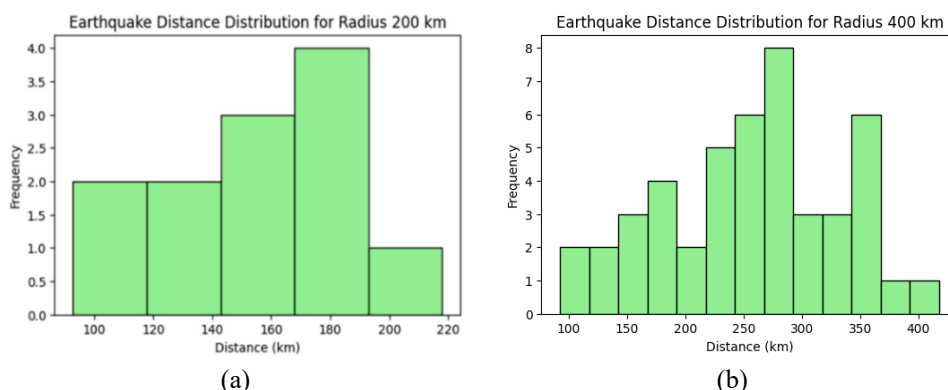


Figure 6. Frequency distribution of earthquake distance for 200 km radius (a) and 400 km radius (b)

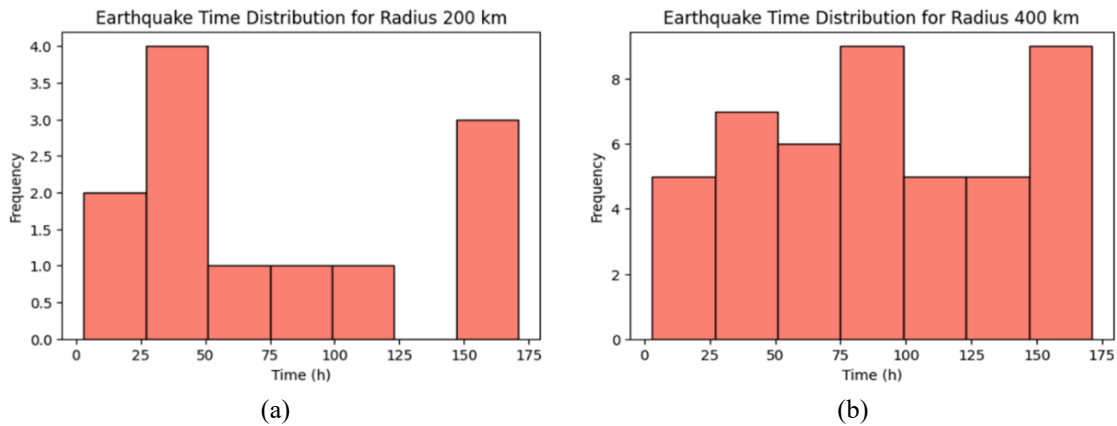


Figure 7. Frequency distribution of earthquake time occurrence for 200 km radius (a) and 400 km radius (b)

Table 2 presents the average RMSE value from the 5-fold cross-validation results for each algorithm in predicting the magnitude, distance, and time of an earthquake. At a prediction radius of 200 km, Monte Carlo Aggregating shows the lowest RMSE for distance prediction (41.98 km), outperforming Linear Regression (45.62 km), Random Forest (57.29 km), and XGBoost (63.90 km). From the magnitude prediction performance, the Random Forest method has the best performance with the lowest RMSE (0.49), while Monte Carlo Aggregating has an RMSE of 0.59. For event-time prediction, Monte Carlo Aggregating and linear regression show the best performance with an RMSE of 0.48. This proposed method is better than Random Forest (0.55) and XGBoost (0.57) in predicting the time of an earthquake.

At a prediction radius of 300 km, the Proposed Method showed the lowest RMSE for earthquake occurrence time prediction (57.85 h), outperforming other prediction methods, namely Linear Regression (60.06 h), Random Forest (65.20 h), and XGBoost (72.60 h). This method also performs well for predicting earthquake occurrence distance with an RMSE of 60.60 km. The Proposed Method can predict better than XGBoost (81.29 km), Random Forest (68.39 km), and linear regression (62.88 km). For magnitude prediction, the Proposed Method has an RMSE equal to Linear Regression (0.48) but still outperforms Random Forest (0.55) and XGBoost (0.57).

For prediction at a 400 km radius, the Proposed Method outperforms all the comparative methods by achieving the lowest RMSE for prediction of time (46.69 h), magnitude (0.61), and distance (76.29 km). In earthquake magnitude prediction, the Proposed Method outperforms XGBoost (0.78), linear regression (0.63), and Random Forest (0.67) in terms of RMSE. In distance prediction, the Proposed Method outperformed XGBoost (107.97), linear regression (77.84), and Random Forest (100.13).

In the 200 km prediction radius, the Random Forest method provides the best predictions for both magnitude and distance. Meanwhile, the proposed method achieved the most accurate predictions for the occurrence time. On the other hand, for the 400 km radius, the proposed method demonstrates the best performance in predicting magnitude, distance, and occurrence time. Uncertainty is a key challenge in making predictions. To analyze the factors contributing to the uncertainty, the frequency distributions of the magnitude, predicted distance, and occurrence time are presented in Figures 5-7.

Figure 5 shows the distribution of earthquake magnitudes for two different radii: 200 km and 400 km. From this

perspective, the difference between these distributions reveals important insights into the predictability and variability of the earthquake magnitudes within each radius. For the 200 km radius, the magnitudes range in the ranges of 4.0-4.5, 5.0-5.5 and 5.5-6.0.

Figure 5(a) shows that most earthquakes occurred with magnitudes between 4.0-4.5. However, the 400 km radius shows a wider range of magnitudes, ranging from 4.0 to 7.0. However, the distribution shows that most earthquakes occur between 4.0 and 5.0. The bin occupancy of the histogram shows the distribution of the magnitudes that occur, so it can be seen that in Figure 5(b) has a wider range of earthquake magnitudes than the distribution of earthquake magnitudes in Figure 5(a). The uncertainty increases because the probability of the earthquake magnitude increases. Monte Carlo Aggregating in magnitude prediction at a radius of 400 km is better to several comparison methods indicating that the Monte Carlo Aggregating method can overcome the problem of uncertainty than the comparison methods.

An analysis of the prediction distance distribution was also conducted to determine the effect of uncertainty on the prediction distance. Figure 6 shows the distribution of earthquake distances for radii of 400 km and 200 km from the monitoring station. The histogram uses a bin width of 25 km. In the distribution of the radius distance of 400 km in Figure 6(b), the number of filled bins was greater than the number of bins in Figure 6(a). This condition indicates that the variability in Figure 6(b) is higher than that in Figure 6(a) with respect to the earthquake distance. Variability was also observed in the height of the bins in Figure 6(b). This condition increases the uncertainty in predicting the distance between earthquake events. In the distance prediction test, the Monte Carlo Aggregating method exhibited a better prediction performance than the comparison method. From this test, it can be observed that the Monte Carlo Aggregating method overcomes the problem of uncertainty in distance prediction.

Figure 7 shows the distribution of earthquake occurrence times for radii of 200 km and 400 km. The distribution provides insight into the predictability of earthquake occurrence time for each radius. For the distribution, the histogram used a bin width of 24 h. Figure 7(b) shows that the bins are more occupied than those in Figure 7(a). Figure 7(b) also shows a relatively flat frequency of occurrence compared to that in Figure 7(a). Thus, the uncertainty in the prediction of the occurrence time for a radius of 400 km was higher than that for a radius of 200 km. The Monte Carlo Aggregating method shows that this method can overcome the uncertainty

in earthquake time prediction.

The main problems in earthquake prediction are data limitations and uncertainties. To ensure that the Monte Carlo Aggregating method can solve the problem of data limitation, a comparison is made between the Monte Carlo Aggregating and Bootstrap Aggregating methods. The Bootstrap method is commonly used for anticipating uncertainty. The comparison is made using a benchmark function written as:

$$y = 2x + 1 + \text{noise} \quad (5)$$

The data used as input (x) were random numbers from a random seed. The *noise* was generated using random numbers from random seeds. The obtained input and output data were divided into 80% for training and 20% for testing. The quantity of data used varied to determine the performance of the method on varying amounts of data. The performance of the Monte Carlo aggregation method is shown in Figure 8.

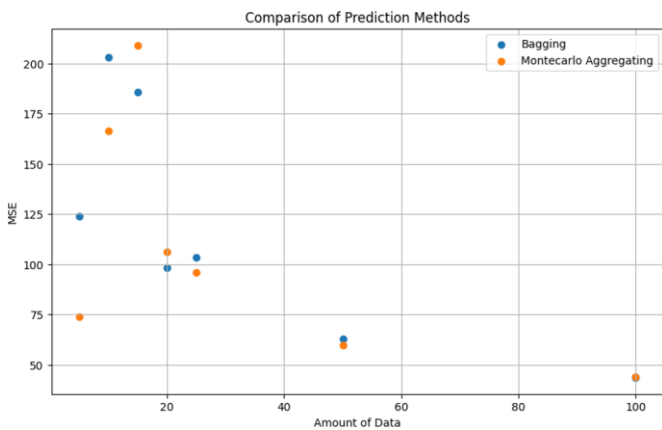


Figure 8. Comparison of Monte Carlo Aggregating and Bootstrap Aggregating (Bagging)

Figure 8 illustrates a comparison of the performance of two prediction methods, Bagging and Monte Carlo Aggregating, based on their Mean Squared Error (MSE) across varying data volumes. Both methods display a general trend of decreasing MSE as the amount of data increases, indicating enhanced prediction accuracy with larger datasets. For very small datasets (5 and 10 data points), Monte Carlo Aggregating outperforms Bagging by achieving a significantly lower Mean Squared Error (MSE). This advantage results from its ability to utilize probabilistic distributions to handle uncertainty, whereas Bagging tends to over fit due to its Bootstrap sampling, resulting in less diverse models. For moderate datasets (15 and 20 data points), Bagging outperforms Monte Carlo Aggregating with a lower MSE, attributed to its ensemble modeling capability, which reduces prediction variability. However, Monte Carlo Aggregating may struggle to represent distributions accurately. For larger datasets (50 and 100 data points), both methods perform well with low MSE, but Monte Carlo Aggregating consistently achieves slightly lower MSE than Bagging, despite the marginal difference.

These results indicate that Monte Carlo Aggregating performs well with very small datasets due to its strength in managing uncertainty, while Bagging is more effective for moderate datasets. For larger datasets, both methods demonstrate reliability, although Monte Carlo Aggregating retains a slight advantage in prediction accuracy. The strong

performance of the Monte Carlo Aggregating method suggests that it could serve as a viable alternative for developing earthquake early warning systems in scenarios with sparse or incomplete data.

Based on the experimental results, it was found that the proposed method could predict earthquake parameters better than several other methods. The data limitations were successfully overcome using the proposed method. It has been proven that the Monte Carlo Aggregating method can predict data better than the Bootstrap Aggregating method on limited data. The success of the proposed method in predicting the uncertainty input shows that it can overcome the problem of uncertainty in making predictions.

Based on these findings, this study is expected to significantly contribute to the development of early earthquake detection technology based on radon gas precursors. Further research should focus on developing more sophisticated learning methods for the proposed algorithm to improve the accuracy and reliability of the predictions. The current method cannot adapt to and learn when more data are available. The actual probability distribution may not be accurately represented when the data are limited. This leads to the potential use of an incorrect distribution, thereby reducing prediction accuracy. Adaptive learning methods that effectively utilize an increase in the amount of precursor data are crucial. With an increase in data over time, the algorithm is expected to learn more complex patterns and gradually improve predictions. Thus, in the future, early earthquake detection technology based on radon gas precursors could become a reliable tool for disaster mitigation.

5. CONCLUSION

From the research conducted, it is proven that the proposed Monte Carlo Aggregating algorithm for dataset limitation in earthquake parameter prediction performs better than several other methods. This algorithm can provide more accurate predictions, which is often the main obstacle to earthquake prediction. The findings demonstrate the performance of Monte Carlo Aggregating, particularly in managing uncertainty and achieving high accuracy with limited data. This approach offers significant potential for enhancing earthquake early warning systems, especially in scenarios with sparse or incomplete data monitoring. Moreover, this method accelerates the development of earthquake early warning systems without requiring prolonged radon observation periods, particularly in the Java Island region. By providing a reliable method for modeling and predicting earthquake parameters, this study contributes to mitigating the impacts of earthquake disasters and improving preparedness through timely and accurate warnings.

At a prediction radius of 200 km, the proposed method shows the best performance in time predictions, with RMSE values of 41.98 hours, respectively, compared to other methods. Meanwhile, the Random Forest has the lowest RMSE value (0.49) in the magnitude parameter and 13.5 km in distance prediction. For a prediction radius of 300 km, the proposed method also shows better performance in magnitude, distance, and time predictions with RMSE values of 0.48, 60.60 km and 57.85 hours, respectively. At a prediction radius of 400 km, the proposed method remains better in magnitude, distance, and time prediction, with RMSE values of 0.61, 76.29 km and 46.69 hours, respectively.

However, additional research is necessary to enhance the performance of this algorithm further. The current method cannot adapt to and learn when more data are available. The actual probability distribution may not be accurately represented when the data are limited. This leads to the potential use of an incorrect distribution, thereby reducing prediction accuracy. Future research should focus on developing an adaptive process that enables the algorithm to improve its prediction accuracy when more data are collected. With this adaptive component, the algorithm is expected to be more responsive to changes in the data patterns, resulting in greater accuracy over time. This advancement will elevate Monte Carlo Aggregating-based earthquake prediction technology, significantly mitigating natural disaster risk.

ACKNOWLEDGMENT

This research is supported by Universitas Gadjah Mada through Program Rekognisi Tugas Akhir 2024 (Grant No.: 5286/UNI.PI/PT.01.03/2024).

REFERENCES

- [1] Xie, L.F., Zou, S.L., Li, X.Y., Hong, C.S., et al. (2018). Effect of ultrasonic treatment on radon exhalation from porous media: An experimental case study. *Sustainability*, 10(9): 3005. <https://doi.org/10.3390/su10093005>
- [2] D’Incecco, S., Petraki, E., Priniotakis, G., Papoutsidakis, M., Yannakopoulos, P., Nikolopoulos, D. (2021). CO₂ and radon emissions as precursors of seismic activity. *Earth Systems and Environment*, 5(3): 655-666. <https://doi.org/10.1007/s41748-021-00229-2>
- [3] Nikolopoulos, D., Cantzos, D., Alam, A., Dimopoulos, S., Petraki, E. (2024). Electromagnetic and radon earthquake precursors. *Geosciences*, 14(10): 271. <https://doi.org/10.3390/geosciences14100271>
- [4] Masruoğlu, G., Altun, C., Şentürk, M.Z., İçhedef, M., Taşköprü, C. (2023). Variation of soil gas ²²²Rn/²²⁰Rn concentration ratios along the Pınarbaşı segment of İzmir fault. *Journal of Radioanalytical and Nuclear Chemistry*, 332(11): 4739-4743. <https://doi.org/10.1007/s10967-023-08910-8>
- [5] Chowdhury, S., Guha Bose, A., Das, A., Deb, A. (2024). A study of some research work on soil radon concentration and ionospheric total electron content as earthquake precursors. *Journal of Radioanalytical and Nuclear Chemistry*, 333(4): 1633-1659. <https://doi.org/10.1007/s10967-024-09409-6>
- [6] Lee, J.K. (2022). Basic study on the observation of earthquake precursor manifestation using radon variability in groundwater. *Crisisonomy*, 18(6): 39-51. <https://doi.org/10.14251/crisisonomy.2022.18.6.39>
- [7] Zhou, Z., Tian, L., Zhao, J., Wang, H., Liu, J. (2020). Stress-related pre-seismic water radon concentration variations in the Panjin observation well, China (1994-2020). *Frontiers in Earth Science*, 8: 596283. <https://doi.org/10.3389/feart.2020.596283>
- [8] Katsanou, K., Stratikopoulos, K., Zagana, E., Lambrakis, N. (2010). Radon changes along main faults in the broader Aigion region, NW Peloponnese. *Bulletin of the Geological Society of Greece*, 43(4): 1726-1736. <https://doi.org/10.12681/bgsg.11358>
- [9] Mehmood, T., Awais, M. (2021). Tukey control chart for radon monitoring in relation to the seismic activity. *Mathematical Problems in Engineering*, 2021(1): 9999500. <https://doi.org/10.1155/2021/9999500>
- [10] Aich, A. (2022). Preliminary studies on soil radon activity at geothermal hotspot of Bakreswar-Tantloi. *IOP SciNotes*, 3(2): 025201. <https://doi.org/10.1088/2633-1357/ac78ac>
- [11] Martha, A.A., Prayogo, A.S., Nugraha, J., Pakpahan, S., Riama, N.F. (2021). Network of radon gas concentration monitoring of research and development centre–BMKG for earthquake precursor research in Indonesia. *IOP Conference Series: Earth and Environmental Science*, 873(1): 012006. <https://doi.org/10.1088/1755-1315/873/1/012006>
- [12] Pratama, T.O., Sunarno, S., Hawibowo, S., Waruwu, M.M., Wijaya, R. (2021). Deterministic system for earthquake early warning system based on radon gas concentration anomaly at Yogyakarta Region-Indonesia. *AIP Conference Proceedings*, 2320(1): 040003. <https://doi.org/10.1063/5.0037683>
- [13] Crowley, H., Bommer, J.J. (2006). Modelling seismic hazard in earthquake loss models with spatially distributed exposure. *Bulletin of Earthquake Engineering*, 4: 249-273. <https://doi.org/10.1007/s10518-006-9009-y>
- [14] Goda, K., Song, J. (2016). Uncertainty modeling and visualization for tsunami hazard and risk mapping: A case study for the 2011 Tohoku earthquake. *Stochastic Environmental Research and Risk Assessment*, 30: 2271-2285. <https://doi.org/10.1007/s00477-015-1146-x>
- [15] Mohammed, D.H.K., Külahcı, F., Muhammed, A. (2021). Determination of possible responses of Radon-222, magnetic effects, and total electron content to earthquakes on the North Anatolian Fault Zone, Türkiye: An ARIMA and Monte Carlo Simulation. *Natural Hazards*, 108(3): 2493-2512. <https://doi.org/10.1007/s11069-021-04785-8>
- [16] Muto, J., Yasuoka, Y., Miura, N., Iwata, D., et al. (2021). Preseismic atmospheric radon anomaly associated with 2018 Northern Osaka earthquake. *Scientific Reports*, 11(1): 7451. <https://doi.org/10.1038/s41598-021-86777-z>
- [17] Attanasio, A., Maravalle, M. (2016). Some considerations between radon and earthquakes in the crater of L’Aquila. *Natural Hazards*, 81: 1971-1979. <https://doi.org/10.1007/s11069-016-2169-4>
- [18] Feng, X., Zhong, J., Yan, R., Zhou, Z., Tian, L., Zhao, J., Yuan, Z. (2022). Groundwater radon precursor anomalies identification by EMD-LSTM model. *Water*, 14(1): 69. <https://doi.org/10.3390/w14010069>
- [19] Qiao, Z., Wang, G., Fu, H., Hu, X. (2022). Identification of groundwater radon precursory anomalies by critical slowing down theory: A case study in Yunnan Region, Southwest China. *Water*, 14(4): 541. <https://doi.org/10.3390/w14040541>
- [20] Mir, A.A., Çelebi, F.V., Alsolai, H., Qureshi, S.A., et al. (2022). Anomalies prediction in radon time series for earthquake likelihood using machine learning-based ensemble model. *IEEE Access*, 10: 37984-37999. <https://doi.org/10.1109/access.2022.3163291>
- [21] Zhu, R., Yang, F., Zhou, X., Tian, J., et al. (2024). Anomaly detection using machine learning in

- hydrochemical data from hot springs: Implications for earthquake prediction. *Water Resources Research*, 60(6): e2023WR034748. <https://doi.org/10.1029/2023wr034748>
- [22] 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*, 13(11): 6424. <https://doi.org/10.3390/app13116424>
- [23] Jarah, N.B., Alasadi, A.H.H., Hashim, K.M. (2023). Earthquake prediction technique: A comparative study. *IAES International Journal of Artificial Intelligence*, 12(3): 1026-1032. <https://doi.org/10.11591/ijai.v12.i3.pp1026-1032>
- [24] Gitis, V.G., Derendyaev, A.B. (2019). Machine learning methods for seismic hazards forecast. *Geosciences*, 9(7): 308. <https://doi.org/10.3390/geosciences9070308>
- [25] Deb, A., Gazi, M., Barman, C. (2016). Anomalous soil radon fluctuations—signal of earthquakes in Nepal and eastern India regions. *Journal of Earth System Science*, 125: 1657-1665. <https://doi.org/10.1007/s12040-016-0757-z>
- [26] Tehseen, R., Farooq, M.S., Abid, A. (2020). Earthquake prediction using expert systems: A systematic mapping study. *Sustainability*, 12(6): 2420. <https://doi.org/10.3390/su12062420>
- [27] Asim, K.M., Idris, A., Iqbal, T., Martínez-Álvarez, F. (2018). Earthquake prediction model using support vector regressor and hybrid neural networks. *PloS ONE*, 13(7): e0199004. <https://doi.org/10.1371/journal.pone.0199004>
- [28] Al Banna, M.H., Taher, K.A., Kaiser, M.S., Mahmud, M., Rahman, M.S., Hosen, A.S., Cho, G.H. (2020). Application of artificial intelligence in predicting earthquakes: State-of-the-art and future challenges. *IEEE Access*, 8: 192880-192923. <https://doi.org/10.1109/ACCESS.2020.3029859>
- [29] Berhich, A., Belouadha, F.Z., Kabbaj, M.I. (2023). An attention-based LSTM network for large earthquake prediction. *Soil Dynamics and Earthquake Engineering*, 165: 107663. <https://doi.org/10.1016/j.soildyn.2022.107663>
- [30] Agarwal, N., Arora, I., Saini, H., Sharma, U. (2023). A novel approach for earthquake prediction using random forest and neural networks. *EAI Endorsed Transactions on Energy Web*. <https://doi.org/10.4108/ew.4329>
- [31] Jena, R., Pradhan, B., Al-Amri, A., Lee, C.W., Park, H.J. (2020). Earthquake probability assessment for the Indian subcontinent using deep learning. *Sensors*, 20(16): 4369. <https://doi.org/10.3390/s20164369>
- [32] Khan, T., Rabbani, M., Siddiquee, S.M.T., Majumder, A. (2019). An innovative data mining approach for determine earthquake probability based on linear regression algorithm. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, Coimbatore, India, pp. 1-4. <https://doi.org/10.1109/ICECCT.2019.8869286>

NOMENCLATURE

G_s	Gain of the monitoring station
R_{\max}	Maximum Radon value (Bq/m ³)
R_{t-i}	Daily average Radon value (Bq/m ³)
e	Prediction error
$f(e)$	Probability mass function
P	Probability
i	index
$F(e)$	Cumulative probability function
U	Random number Uniform
n	Number of iteration
N	Number of data
\hat{Y}_i	Predicted value
Y_i	Recorded earthquake parameter value