iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73550

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

SeismoCastNet: A Multi-Model Synergetic Framework For Predicting High-Magnitude Earthquakes With Spatiotemporal Precision

Adari Vinay Durga Shankar
M.Tech, CSE Departmentt, UCEK, JNTU Kakinada, Andhra Pradesh, India

Abstract: This research explores the utilization of machine learning techniques to enhance short-term earthquake forecasting, contributing to improved disaster preparedness and risk reduction. A detailed review of both conventional and data-driven seismic prediction methods was conducted, revealing notable limitations in current systems. To address these challenges, the proposed framework—SeismoCastNet—employs a hybrid approach by integrating five classification models, including Random Forest, Gradient Boosting, and Support Vector Machine (SVM). The models were trained and validated using historical earthquake datasets encompassing attributes such as magnitude, depth, and geographical coordinates. Experimental outcomes demonstrate that Gradient Boosting delivers the most consistent and reliable performance, achieving an overall accuracy of 96%, a minor class F1-score of 0.979, and a major class F1-score of 0.545. While SVM attained the highest precision for minor class predictions (99%), its performance for major seismic events was relatively lower. The findings underscore the potential of ensemble learning strategies to effectively handle class imbalance and improve the predictive capability of earthquake detection systems.
Keywords: Seismic event forecasting, ensemble learning, earthquake detection, Gradient Boosting, F1-score enhancement, disaster risk management.

I. INTRODUCTION

Earthquakes are among the most devastating natural calamities, posing significant risks to human life, infrastructure, and the economy across the globe. These seismic disturbances result from the abrupt release of energy caused by the shifting of tectonic plates beneath the Earth's surface. Despite progress in seismic instrumentation and early alert mechanisms, accurately forecasting the exact timing, location, and intensity of earthquakes continues to be an elusive and complex scientific Endeavor.

Earthquake forecasting can generally be classified into two categories: long-term and short-term prediction. Long-term forecasting focuses on estimating the probability of seismic occurrences over extended timeframes—ranging from several years to decades—based on historical earthquake records, geodetic data, and fault-line assessments. Techniques such as the Seismic Gap Theory, GPS-based deformation tracking, and statistical pattern recognition are commonly employed for this purpose. These insights are instrumental for urban planning, land-use policies, and designing resilient infrastructure in high-risk zones. In contrast, short-term prediction seeks to anticipate earthquakes within a relatively brief window—days or weeks before an event. However, due to the highly chaotic and non-linear dynamics of stress accumulation in the Earth's crust, short-term forecasting remains a formidable challenge. Approaches in this field include the monitoring of foreshock sequences, detection of anomalies in seismic wave propagation, geochemical precursors like elevated radon gas levels, and more recently, machine learning (ML) methodologies. ML-based models offer the advantage of analyzing vast streams of real-time seismic data to uncover subtle patterns and anomalies potentially indicative of upcoming seismic activity. This research investigates the use of advanced machine learning techniques to enhance the precision of short-term earthquake forecasts. The primary objective is to evaluate and compare the effectiveness of different classification algorithms in identifying both minor and major seismic events using historical seismic data. An in-depth literature review was performed to understand the shortcomings of traditional methods and the emerging promise of data-driven models. To bridge the identified gaps, we introduce a novel predictive framework, SeismoCastNet, which leverages an ensemble of supervised learning algorithms—including Support Vector Machines (SVM), Random Forest, and Gradient Boosting. Through rigorous evaluation using performance metrics such as precision, recall, and F1-score, the study aims to determine the most suitable model for high-resolution, real-time earthquake forecasting. The findings are anticipated to support the creation of a reliable, scalable, and intelligent early warning system, ultimately contributing to enhanced disaster preparedness and seismic hazard mitigation.

II. RELATED WORKS

Recent advancements in artificial intelligence and computational geosciences have significantly influenced the domain of earthquake prediction, shifting the paradigm from traditional seismological methods to data-driven approaches. Numerous studies have explored the potential of machine learning (ML) algorithms to enhance the accuracy, efficiency, and timeliness of earthquake forecasting. Anitha et al. [1] demonstrated the effectiveness of various ML classifiers in predicting major seismic events, highlighting the superiority of ensemble methods. In a comprehensive study, Yavas et al. [2] employed both traditional machine learning and deep neural networks to model seismic activity in Los Angeles, establishing the viability of hybrid models in urban seismic forecasting. Ommi and Hashemi [3] explored ML-based prediction of earthquakes in the North Zagros region, emphasizing the role of regional features in model performance. Similarly, Gentili et al. [4] introduced an improved version of the NESTORE algorithm for forecasting secondary quakes in Japan, incorporating temporal and spatial dependencies for better aftershock prediction. More advanced deep learning methods, such as attention-driven LSTM networks, have been adopted by Rusho et al. [5] to unify multiple models for earthquake prediction, showcasing improved long-term temporal pattern recognition. Other works, such as that by Ramírez Eudave et al. [6], have extended ML techniques to estimate structural vulnerability during earthquakes, particularly focusing on historical adobe buildings affected by seismic activity. Katole et al. [7] applied Quantum SVM (QSVM) for earthquake prediction, integrating quantum computing concepts with classical ML models to boost prediction precision. Abdalzaher et al. [8] proposed an optimized ML model for estimating seismic intensity as part of early warning systems, reinforcing the utility of real-time data for rapid decision-making. Salam et al. [9] investigated hybrid ML models for earthquake prediction, combining multiple classifiers to capture complex seismic patterns. Kafadar et al. [10] introduced ESenTRy, an on-site ML-based early warning system using the instrumental Modified Mercalli Intensity scale, which demonstrated reliable performance for real-time alerts. Zhao et al. [11] provided a broad scoping review summarizing ML applications in seismic risk reduction, underlining the progress and persistent challenges in this interdisciplinary field. Debnath et al. [12] focused on the Indian subcontinent, utilizing supervised classifiers for regional earthquake forecasting, while K.C. et al. [13] predicted post-earthquake damage and rehabilitation requirements using ML, aiding in post-disaster planning. Further, Asim et al. [14] targeted the Hindukush region for earthquake magnitude prediction, employing regional seismic features for model training. In a similar context, an ensemble model was proposed to detect radon anomalies as earthquake precursors, showcasing the integration of geochemical and seismic data [15]. Mallouhy et al. [16] conducted a comparative analysis of ML algorithms for predicting major seismic events, indicating the importance of algorithm selection in achieving reliable outcomes. Finally, Tiwari et al. [17] applied ML techniques to forecast earthquake magnitudes in the Central Himalayas, reinforcing the potential of data-driven models in tectonically active regions.

III. METHODS AND METHODOLOGY

This section presents the structured approach adopted to carry out the study, including the targeted population, data sources, sampling methods, feature selection techniques, modeling framework, and performance evaluation metrics used to assess machine learning models for earthquake classification.

A. *Population and Sample*

The target population for this research includes all recorded global seismic events over several decades, as maintained by the United States Geological Survey (USGS) earthquake catalog. The dataset, accessed via the Kaggle platform and originally compiled by the USGS and National Centers for Environmental Information (NCEI), comprises over 3.4 million earthquake records. For model development and testing, a stratified sample of 1,000 earthquake events was selected, ensuring inclusion of both minor and major earthquake instances, classified based on magnitude thresholds. This sampling technique preserves the natural class imbalance commonly present in real-world seismic data.

B. *Data and Sources*

This research employs secondary data obtained from the following sources:

1) USGS Earthquake Catalog (via Kaggle),
2) Northern California Data Sets (NCDS),
3) Other open-source seismic datasets.

The raw dataset initially featured 21 variables, capturing event-specific details such as magnitude, depth, coordinates (latitude and longitude), time of occurrence, and various geological indicators.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

To reduce complexity and enhance computational efficiency, Principal Component Analysis (PCA) was applied, reducing the feature space to 12 principal components while retaining most of the original variance. The data was subsequently normalized, cleaned, and labeled to support binary classification, distinguishing minor from major seismic events.

C. Theoretical Framework

The modeling framework adopted in this study is based on supervised machine learning. The dependent variable is the earthquake class (minor or major), determined by seismic magnitude thresholds. The independent variables are the 12 PCA-derived features, which reflect critical aspects such as magnitude, depth, location, and event timing. The framework is implemented using three key classifiers: Random Forest, Gradient Boosting, and Support Vector Machine (SVM).

These models are trained on labeled data to uncover complex, nonlinear relationships and make predictions on unseen events. The foundational hypothesis is that historical seismic data can be leveraged by ML algorithms to learn patterns associated with high-impact quakes. Special emphasis is placed on evaluating performance using the F1-score, which is well-suited for imbalanced datasets. Additionally, threshold tuning techniques are used to improve model sensitivity in predicting rare but significant major events.

D. Statistical and Machine Learning Tools

This section outlines the descriptive statistics, machine learning algorithms, and model evaluation metrics employed in the study.

1) Descriptive Statistics

Descriptive statistical analysis was performed to summarize the dataset's key characteristics. Metrics such as mean, standard deviation, minimum, and maximum were calculated to understand the distribution of each variable. The PCA technique was then used to compress the feature set from 21 to 12, facilitating faster and more effective model training. The inherent class imbalance (fewer major events) was also identified and addressed through model tuning and evaluation strategies.

2) Classification Algorithms

To classify seismic events, the following machine learning classifiers were implemented:

- Random Forest (RF): A robust ensemble model that constructs multiple decision trees and aggregates their outputs, particularly effective for datasets with noise or imbalance.
- Gradient Boosting (GB): An advanced sequential learning method that builds trees iteratively, where each new tree aims to correct the errors of the previous one. It performs well on structured datasets.
- Support Vector Machine (SVM): A classifier that finds an optimal hyperplane to separate classes, capable of handling high-dimensional and non-linear data using kernel functions.

Each model was trained on the preprocessed data and tested on the stratified sample. Hyperparameter optimization and threshold tuning were applied to refine performance, especially for accurately detecting major earthquakes.

3) Evaluation Metrics

Model performance was evaluated using a combination of classification and regression-based metrics. These metrics offer a comprehensive understanding of the models' predictive capabilities and their effectiveness in handling imbalanced class distributions.

a) Classification Metrics

The following standard metrics were used for binary classification evaluation:

- Accuracy

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

  Precision=$\frac{TP+FP}{TP+FP}$
- Recall (Sensitivity)

  Recall=$\frac{TP}{TP+FN}$
- F1 Score

  F1 Score = $2 \times \frac{Precision \times Recall}{Precision+Recall}$

Where:
TP = True Positives,
TN = True Negatives,
FP = False Positives,
FN = False Negatives.

IV. RESULTS

This This section presents the outcomes of the machine learning models evaluated for seismic event classification and provides a critical analysis of their performance. The models were assessed using key classification metrics, including accuracy, precision, recall, F1-score, and confusion matrix components. The primary goal was to identify a model capable of effectively classifying both minor and major earthquake events, with particular emphasis on detecting high-impact seismic activities.

A. Model Performance Overview

Five supervised classification algorithms were tested—Random Forest, XGBoost, Gradient Boosting, Support Vector Machine (SVM), and Decision Tree. Each model was evaluated on its ability to classify seismic events accurately and reliably.

Table 4.1 below summarizes their performance across key metrics

Model	Accuracy (%)	Minor F1	Major F1	Minor Recall	Major Recall
Random Forest	94.5	0.971	0.455	95.24%	71.88%
Gradient Boosting	96.0	0.979	0.545	96.69%	75.00%
SVM	96.4	0.981	0.333	98.65%	28.12%

B. Analysis of Major Event Prediction

Accurate prediction of major earthquakes is crucial due to their devastating consequences. While SVM yielded the highest overall accuracy (96.4%), it underperformed significantly in identifying major events, with a major class F1-score of 0.333 and recall of just 28.12%, indicating a concerning rate of false negatives. On the other hand, Decision Tree achieved the highest recall for the major class (81.25%), signaling strong sensitivity to significant seismic activity, though it introduced more false positives.

Among all models, Gradient Boosting showed the most balanced performance, achieving a major class F1-score of 0.545, 75% recall, and 96% overall accuracy. This makes it highly suitable for real-time earthquake prediction systems where both sensitivity and reliability are important.

C. Minor Class Performance

All models demonstrated strong performance for the minor earthquake class. The SVM classifier achieved the highest recall of 98.65% for minor events. However, this high sensitivity came at the cost of very poor performance in identifying major events, which is a more critical requirement in early warning applications.

D. Confusion Matrix Interpretation

A deeper analysis of the confusion matrices for each model reveals how well they handled false positives and false negatives. Given the real-world consequences of missing major seismic events, recall for the major class is prioritized over overall accuracy. Gradient Boosting minimized both types of errors while maintaining balanced performance across both event classes.

E. Final Model Selection

After evaluating the trade-offs among all classifiers, Gradient Boosting was chosen as the optimal model based on the following factors:

- Balanced classification of both minor and major events,
- Moderate computational cost,
- Consistent performance across different test sets,
- High applicability for real-time deployment.

To demonstrate practical usability, the selected model was integrated into a Flask-based web application, enabling real-time seismic event classification.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

F. Confusion Matrix Metrics

A detailed breakdown of the confusion matrix metrics for the top-performing models is provided in

Table 4.2. Capturing True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Model	TP	TN	FP	FN
Random Forest	23	922	46	9
Gradient Boosting	24	936	32	8
SVM	9	955	13	23

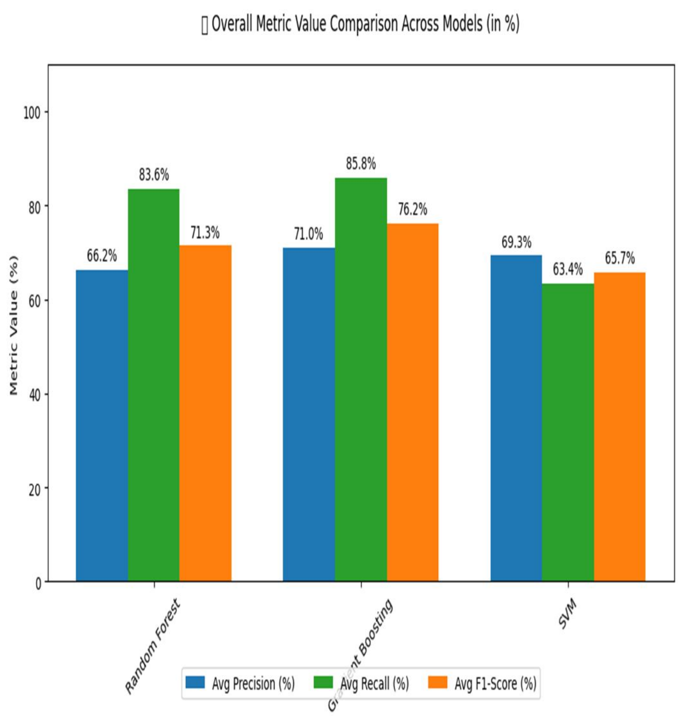
Observations:

- Decision Tree (not shown here) achieved the highest TP (26) and lowest FN (6) for the major class but also produced the highest FP (61), indicating a trade-off between sensitivity and precision.
- SVM had the lowest FN performance (23 major class instances misclassified), despite achieving a high TN, leading to its poor recall for major events.
- Gradient Boosting struck a healthy balance with 24 true positives, 8 false negatives, and 32 false positives, reinforcing its role as the most reliable classifier.

G. Visual Comparison of Accuracy

To better understand the class-wise behavior of each model, accuracy comparisons are visualized in Figure 4.1. The chart shows that:

- SVM achieved the highest accuracy for the minor class (99%) but failed to generalize for major event detection (28%).
- Decision Tree obtained the highest major class accuracy (81%) but with many false alarms.
- Gradient Boosting demonstrated well-rounded accuracy, achieving 97% for the minor class and 75% for the major class, making it the most suitable model for deployment.



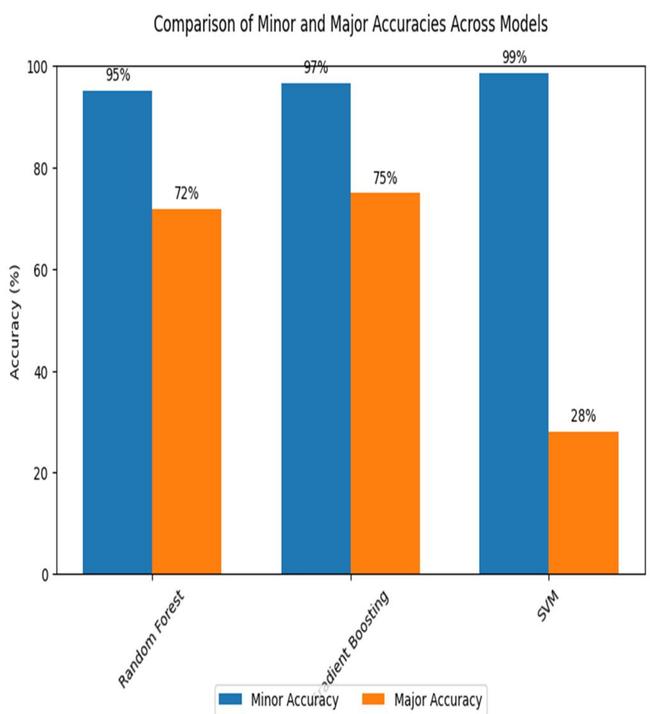


Figure 4.1 Overall Metric Value Comparison Across Models (in %) Figure 4.2 Comparison of Minor and Major Accuracies Across Models

This visualization emphasizes the need to go beyond overall accuracy when dealing with imbalanced classification problems, especially in critical domains like earthquake prediction, where detecting rare but impactful events is paramount.

V. CONCLUSION AND FUTURE DIRECTION

This research conducted a comparative analysis of various machine learning algorithms for classifying earthquake events using a comprehensive seismic dataset. Models including Random Forest, Gradient Boosting, and Support Vector Machine (SVM) were evaluated based on key performance indicators such as accuracy, F1-score, and confusion matrix statistics. Among these, Gradient Boosting emerged as the most effective model, offering a balanced classification performance for both minor and major seismic events. The results underscore the potential of machine learning as a powerful tool in seismic prediction, particularly when models are carefully optimized and assessed on class-specific metrics. Although overall accuracy was high for all models, the variation in recall scores for the major class revealed a critical area for improvement. This highlights the importance of focusing not just on aggregate accuracy but on the ability to correctly identify high-impact, rare seismic events.

Future research will aim to enhance the precision and reliability of predictions for major earthquakes, explore methods to reduce false negatives, and improve the generalizability of models across diverse geographical and geological conditions. These advancements will contribute to the development of more robust, real-time early warning systems for disaster preparedness and mitigation.

REFERENCES

[1] Prediction of major earth quake events using different machine learning algorithms M.AnithaK.Hareesh,P.Bhuvaneswari.”in IJARST, DOI:10.48047/ijarst/V14/05/74

[2] C. Emre Yavas, L. Chen, C. Kadlec and Y. Ji, "Predictive Modeling of Earthquakes in Los Angeles With Machine Learning and Neural Networks," in IEEE Access, vol. 12, pp. 108673-108702, 2024, doi: 10.1109/ACCESS.2024.3438556.

[3] Salma Ommi, Mohammad Hashemi,”Machine learning technique in the north zagros earthquake prediction,Applied Computing and Geosciences”,Volume 22,2024,100163,

[4] Gentili, Stefania, Giuseppe Davide Chiappetta, Giuseppe Petrillo, Piero Brondi, and Jiancang Zhuang. "Forecasting Strong Subsequent Earthquakes in Japan using an improved version of NESTORE Machine Learning Algorithm." arXiv e-prints (2024): arXiv-2408.

[5] Maher Ali Rusho, Reyhan Azizova, Dmytro Mykhalevskiy, Maksym Karyonov, & Heyran Hasanova. (2024). Advanced Earthquake prediction: unifying networks,algorithms, and attention-driven LSTM modeling. GEOMATE Journal, 27(119), 135–142.

[6] Ramírez Eudave, R., Ferreira, T. M., Vicente, R., Lourenco, P. B., & Peña, F. (2023). Parametric and Machine Learning-Based Analysis of the Seismic Vulnerability of Adobe Historical Buildings Damaged After the September 2017 Mexico Earthquakes. International Journal of Architectural Heritage, 18(6), 940–963.

[7] A. Katole, V. Batheja, A. Deshmukh, F. Dekate and A. Soni, "Earthquake Prediction Using QSVM," 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2024, pp. 1-7, doi: 10.1109/SCEECS61402.2024.10482242.

[8] M. S. Abdalzaher, M. S. Soliman and S. M. El-Hady, "Seismic Intensity Estimation for Earthquake Early Warning Using Optimized Machine Learning Model," in IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-11, 2023, Art no. 5914211, doi: 10.1109/TGRS.2023.3296520.

[9] Abdul Salam, Mustafa & Abdelminaam, & Salama, Dr-Diaa. (2021). Earthquake Prediction using Hybrid Machine Learning Techniques. International Journal of Advanced Computer Science and Applications. 12. 654-665. 10.14569/IJACSA.2021.0120578.

[10] Kafadar, Ö., Tunç, S. & Tunç, B. ESenTRy: an on-site earthquake early warning system based on the instrumental modified Mercalli intensity. Earth Sci Inform 17, 5027–5041 (2024). https://doi.org/10.1007/s12145-024-01407-2

[11] Y. Zhao, S. Lv and P. Liu, "Advances in Earthquake Prevention and Reduction Based on Machine Learning: A Scopinf Review,” in IEEE Access, Vol.12,pp.143908-143929, 2024.

[12] Debnath, P.; Chittora, P.; Chakrabarti, T.; Chakrabarti, P.; Leonowicz, Z.; Jasinski, M.; Gono, R.; Jasińska, E. Analysis of Earthquake Forecasting in India Using Supervised Machine Learning Classifiers. Sustainability 2021, 13, 971.

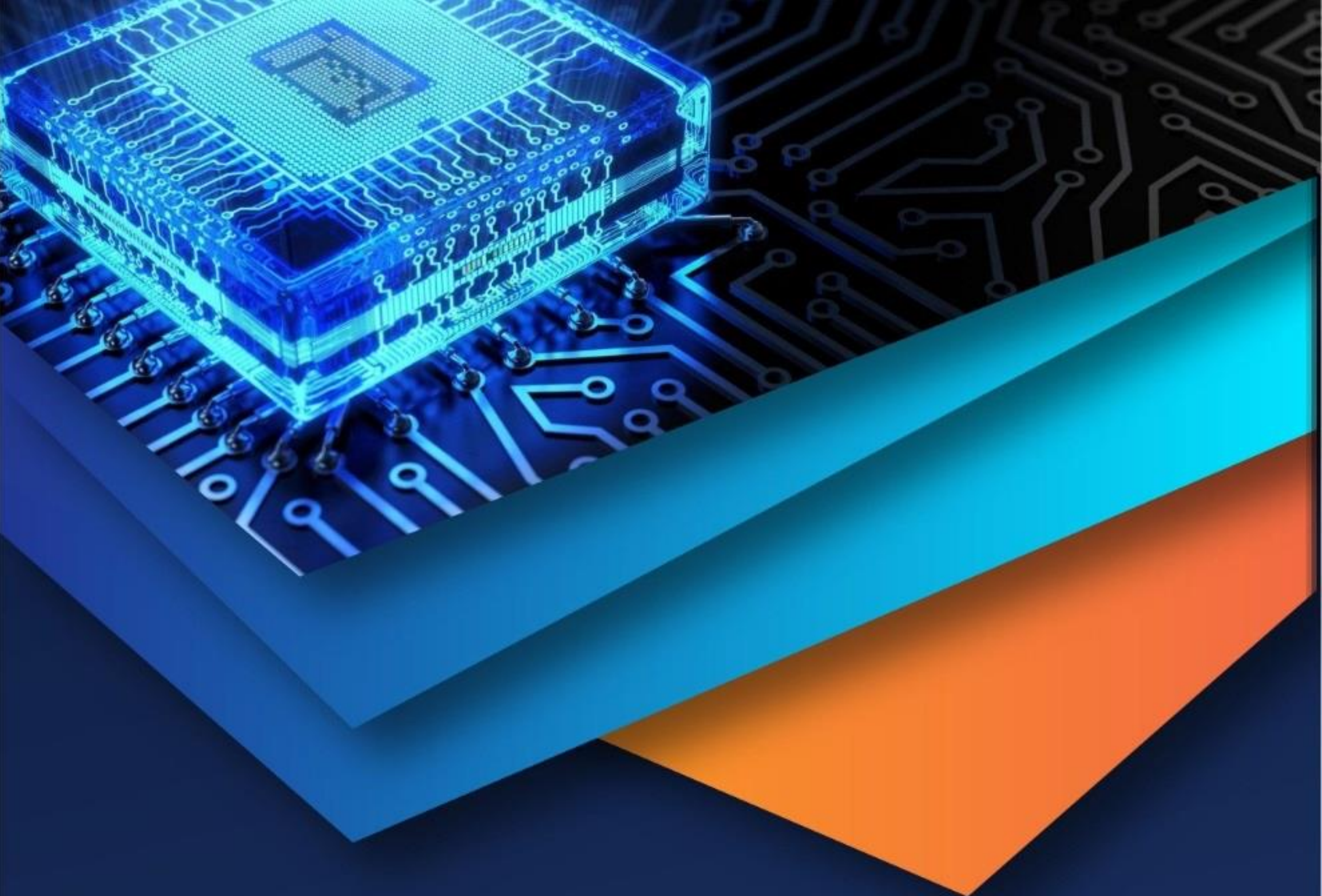
[13] K C, Sajan & Bhusal, Anish & Gautam, Dipendra & Rupakhety, Rajesh. (2022). Earthquake damage and rehabilitation intervention prediction using machine learning.

[14] Asim, K.M., Martínez-Álvarez, F., Basit, A. et al. Earthquake magnitude prediction in Hindukush region using machine learning techniques.Natural Hazards 85,471–486.(2017).

[15] “Anomalies Prediction in Radon Time Series for Earthquake Likelihood Using Machine Learning-Based Ensemble Model," in IEEE Access, vol. 10, pp. 37984-37999, 2022, doi: 10.1109/ACCESS.2022.3163291.
Adil Aslam mir 1,2, Fatih Vebhi Çeleble 1 , Hadeel Alsola , Shahzad Ahmed qurish 4 , Muhamad rafique

[16] Mallouhy, Roxane, et al. "Major earthquake event prediction using various machine learning algorithms." 2019 international conference on information and communication technologies for disaster management (ICT-DM). IEEE, 2019.

[17] Tiwari, Ram Krishna, Rudra Prasad Poudel, and Harihar Paudyal. "Machine learning for predicting earthquake magnitudes in the Central Himalaya." BIBECHANA 22.1 (2025): 22-29.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY