iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VII **Month of publication:** July 2026

DOI: https://doi.org/10.22214/ijraset.2026.84074

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

A Unified Neural Network and XGBoost Learning Model for Disaster Prediction and Management

Ms. V. Bharathi[1], SK. Raqeeba[2]
[1]Associate Professor, [2]Assistant Professsor, Department of Computer Science and Engineering, Geethanjali Institute of Science and Technology, Nellore, Andhra Pradesh, India

***Abstract:** Accurate disaster prediction plays a critical role in disaster management and mitigation by reducing the loss of life and property. Traditional machine learning approaches often struggle to handle complex disaster patterns, class imbalance in datasets, and challenges in multi-class disaster classification, which reduces prediction reliability. To address these issues, a hybrid disaster prediction framework based on Neural Networks and XGBoost is proposed for the multi-class classification of natural disasters. In the proposed framework, neural networks are employed for high-level feature extraction from real-world disaster data, while XGBoost (Extreme Gradient Boosting) performs efficient classification using gradient-boosted decision trees. The dataset, sourced from the EM-DAT International Disaster Database, comprises 4,543 samples with 31 features encompassing wildfires, floods, and earthquakes. The Synthetic Minority Oversampling Technique (SMOTE) is applied exclusively to the training data to address class imbalance and improve the model's generalization capability. Experimental results demonstrate improved performance, with the proposed Neural-XGBoost (N-XGB) model achieving 94.72% accuracy, 98.57% ROC-AUC, 90.49% F1-score, and 89.65% precision, outperforming traditional machine learning models. The proposed system provides a reliable and efficient disaster prediction approach that supports better disaster preparedness and decision-making.*

***Keywords:** Neural Network, XGBoost, Disaster Prediction, SMOTE, Machine Learning, Classification, EM-DAT*

I. INTRODUCTION

Natural disasters are critical events that significantly impact human life, infrastructure, and the environment. Events such as floods, wildfires, and earthquakes have increased in frequency and severity over recent years, making accurate prediction essential for effective disaster management and risk reduction. Manual or conventional approaches often rely on historical analysis and expert judgment, which may be time-consuming and less accurate when handling large and complex datasets. These limitations highlight the need for automated and intelligent prediction systems.

The rapid advancement of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) provides an opportunity to overcome these limitations. These technologies can automatically analyze large datasets, identify hidden patterns, and generate accurate predictions, enabling faster and more reliable disaster forecasting. This project focuses on developing a hybrid machine learning model that combines the feature extraction capability of Neural Networks with the powerful classification ability of XGBoost, forming the Neural-XGBoost (N-XGB) model.

The dataset is collected from the EM-DAT International Disaster Database, which provides comprehensive global records of natural disasters including wildfires, floods, and earthquakes. Preprocessing techniques including median and mode imputation, label encoding, and StandardScaler normalization are applied to ensure data quality. The Synthetic Minority Oversampling Technique (SMOTE) is applied to handle class imbalance, ensuring fair representation of all disaster types. The model is evaluated against traditional machine learning algorithms such as Random Forest, Support Vector Machine (SVM), and Logistic Regression to demonstrate its effectiveness. The primary objective of this work is to achieve high accuracy in disaster prediction, enabling better preparedness, efficient resource allocation, and improved decision-making. A user-friendly web-based interface developed using Streamlit allows users to input data, test the model, and obtain real-time predictions, thereby ensuring practical usability.

II. LITERATURE SURVEY

Saleem et al. proposed a hybrid prediction model integrating artificial intelligence and geospatial analysis for disaster management, demonstrating that combining multiple data-driven approaches enhances prediction quality [1]. Building on multi-disaster classification, Thirukrishna developed an enhanced discovery framework for multiple natural disasters using machine learning, highlighting the need for models that can simultaneously handle diverse disaster categories [2].

The EM-DAT International Disaster Database, maintained by CRED at UCLouvain, serves as a foundational resource for disaster research and provides the dataset utilized in the proposed framework [3]. Donatti et al. identified global hotspots of climate-related disasters, underscoring the geographic variability that machine learning models must accommodate [4]. Tung et al. addressed the challenge of resource optimization in disaster management and planning, emphasizing the importance of actionable predictions for emergency response [5].

Airlangga conducted a comparative analysis of ensemble and deep learning models for tsunami prediction, demonstrating that ensemble approaches such as Gradient Boosting achieve more consistent and reliable results compared to standalone deep learning models across varying data conditions [7]. Sharma and Gosain performed a systematic literature review addressing class imbalance in remote sensing using deep learning approaches, reporting that accuracy can improve to approximately 87% when appropriate imbalance-handling strategies are applied [11].

Chen and Guestrin introduced XGBoost, a scalable tree boosting system that has since become one of the most widely adopted classifiers for tabular data due to its efficiency, handling of missing values, and built-in regularization [15]. Chawla et al. proposed SMOTEBoost for improving prediction of minority classes, demonstrating that synthetic oversampling combined with boosting significantly reduces bias toward majority classes [12]. Husain et al. further compared SMOTE and SMOTEENN for addressing class imbalance in regression models, affirming the value of oversampling strategies in imbalanced datasets [17].

Cantini et al. explored the use of prompt-based large language models for disaster monitoring and automated reporting from social media feedback, illustrating the broader AI ecosystem in which prediction models operate [24]. Ghorpade et al. reviewed flood forecasting using machine learning, cataloguing a range of algorithmic approaches and their comparative effectiveness [18]. Jia and Ye examined deep learning for earthquake disaster assessment across multiple stages, objects, and data types, providing context for the multi-class disaster challenge addressed in the proposed model [22].

Jones et al. reviewed missing data challenges in disaster research using the EM-DAT database, identifying imputation strategies that inform the preprocessing pipeline of the proposed system [10]. Jerez et al. validated the use of median and mode imputation in real-world datasets with missing values, supporting the preprocessing design choices adopted in this work [32]. The collective findings of the surveyed literature confirm the need for a unified hybrid approach that combines neural feature extraction, gradient boosting classification, and synthetic oversampling to address the inherent challenges of multi-class disaster prediction.

III. SYSTEM ANALYSIS

A. *Existing System*

Existing disaster prediction systems predominantly rely on traditional machine learning techniques and individual deep learning models applied independently. Each model generates its own predictions without leveraging the combined strengths of complementary approaches. Furthermore, the feature selection process in existing systems is primarily manual and limited in scope, relying on predefined attributes rather than automated or advanced feature extraction techniques. This reduces the model's ability to capture complex patterns present in the data.

The key disadvantages of existing systems include poor performance on imbalanced datasets, an inability to fully capture complex disaster patterns, lower classification accuracy for minority disaster classes such as wildfires and earthquakes, and results that are often biased and inconsistent due to the dominance of majority classes in training data.

B. *Proposed System*

The proposed system develops a hybrid disaster prediction model by combining Neural Networks and XGBoost to enhance classification performance.

Neural Networks are utilized to automatically extract significant high-level features from disaster-related data, reducing the need for manual feature engineering and capturing complex nonlinear relationships between variables. To address the issue of class imbalance, SMOTE is applied exclusively to the training data, which generates synthetic samples for minority classes and improves model fairness and generalization. Finally, XGBoost is employed as the classification algorithm, efficiently processing the extracted features to accurately classify different types of disasters.

The advantages of the proposed system include effective handling of imbalanced disaster data for minority classes, improved classification consistency across different disaster types, higher prediction accuracy, and more reliable and generalizable results compared to individual or traditional machine learning models.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

IV. METHODOLOGY

A. System Architecture

The proposed system architecture is designed to predict disasters such as floods, wildfires, and earthquakes through a sequential pipeline of machine learning components. Initially, data is collected from the EM-DAT database and preprocessed through handling of missing values, label encoding, and feature normalization. A neural network is then employed for feature extraction to identify important patterns in the dataset. The extracted features are divided into training and testing subsets using a 70:30 split. SMOTE is applied exclusively to the training data to balance class distribution. The balanced training data is then used to train an XGBoost classifier, which generates the final disaster type prediction as output.

B. Data Collection and Preprocessing

The disaster dataset is sourced from the EM-DAT International Disaster Database, which provides comprehensive global records of natural disasters. The dataset contains 4,543 samples with 31 features, including disaster type, location, magnitude, economic damage, and socio-economic attributes covering three target classes: floods, wildfires, and earthquakes.

Preprocessing involves three key steps. First, missing values in numerical features are imputed using the median, while categorical features use the mode to ensure data completeness. Second, categorical variables such as disaster type and region are converted into numerical values using label encoding to ensure compatibility with machine learning algorithms. Third, numerical features are normalized using StandardScaler to achieve zero mean and unit variance, improving model convergence and classification performance.

C. Neural Network for Feature Extraction

A neural network is employed to automatically extract high-level and nonlinear features from the preprocessed dataset. The network consists of fully connected layers with Rectified Linear Unit (ReLU) activation functions that capture complex nonlinear relationships between disaster-related attributes. Dropout regularization is applied between layers to prevent overfitting and improve generalization. The output of the neural network is a transformed and compact feature representation that encodes the most discriminative patterns in the data. This step significantly improves prediction accuracy by reducing the dimensionality of the feature space and eliminating redundant information. The data passes through multiple hidden layers according to the transformation $h_1(X) \rightarrow h_2(h_1) \rightarrow h_3(h_2)$, where each layer further refines the feature representation.

D. SMOTE for Class Balancing

One of the critical challenges in disaster datasets is the significant class imbalance, where events such as floods are overrepresented compared to wildfires and earthquakes. This imbalance leads to biased predictions that favor the majority class and reduces the model's ability to accurately detect less frequent disaster types. To address this challenge, the Synthetic Minority Oversampling Technique (SMOTE) is applied exclusively to the training data. SMOTE generates synthetic samples for minority classes by interpolating between existing minority class instances, thereby ensuring balanced class distribution during training without introducing data leakage into the test set.

E. XGBoost Classification

XGBoost (Extreme Gradient Boosting) is an advanced gradient boosting algorithm employed for classifying disasters based on the features extracted by the neural network. XGBoost builds multiple decision trees sequentially in a boosting manner, where each tree learns from the residual errors of the preceding tree. The algorithm optimizes a loss function augmented by a regularization term, which prevents overfitting and ensures high predictive accuracy. XGBoost is particularly effective in handling missing values, complex data patterns, and imbalanced datasets, making it highly suitable for the multi-class disaster prediction task. The final prediction is derived by aggregating the outputs of all constructed decision trees, with class probabilities used to assign the most likely disaster type to each input sample.

V. IMPLEMENTATION

A. System Modules

The implementation is organized into six primary modules. The Data Collection module retrieves disaster-related records from the EM-DAT database, which forms the foundation for all subsequent processing and model training. The Data Preprocessing module cleans raw data by handling missing values, encoding categorical features, and normalizing numerical attributes to ensure the dataset is suitable for machine learning algorithms. The Feature Extraction module employs a neural network built using PyTorch to automatically identify important patterns and generate meaningful feature representations from the preprocessed input data.

The Data Balancing module applies SMOTE to the training subset to address class imbalance by generating synthetic samples for minority disaster categories, preventing prediction bias toward the flood class. The Classification module applies the XGBoost algorithm to classify disaster types based on the neural network-extracted features, leveraging gradient boosting to produce accurate multi-class predictions. Finally, the Model Evaluation module assesses performance using accuracy, F1-score, ROC-AUC, and a detailed classification report, while a Streamlit-based web interface provides real-time prediction capabilities for end users.

B. Software and Hardware Requirements

The system is implemented using Python 3.9 with PyTorch as the deep learning framework and XGBoost and Scikit-learn as the machine learning libraries. Data processing relies on NumPy and Pandas, while visualization is performed using Matplotlib and Seaborn. The system operates on Windows 11 (64-bit) with an Intel i5 processor, 8 GB RAM, and 256 GB of storage. All tools and libraries used are open-source, and the EM-DAT dataset is publicly available, making the system economically feasible without significant infrastructure investment.

VI. RESULTS AND DISCUSSION

The proposed Neural-XGBoost (N-XGB) hybrid model was evaluated on the EM-DAT disaster dataset comprising 4,543 samples across three disaster classes: floods, wildfires, and earthquakes. The model was trained on 70% of the data and tested on the remaining 30%, with SMOTE applied exclusively to the training set to address class imbalance. Performance was measured using accuracy, precision, F1-score, and ROC-AUC.

The experimental results demonstrate that the proposed N-XGB model achieves a prediction accuracy of 94.72%, a ROC-AUC score of 98.57%, an F1-score of 90.49%, and a precision of 89.65%. These metrics confirm that the model generalizes effectively across all three disaster classes, including minority classes such as wildfires and earthquakes that are typically underrepresented in imbalanced datasets. The application of SMOTE to the training data significantly improved the model's ability to correctly classify minority disaster events, reducing prediction bias compared to models trained without oversampling.

TABLE I PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS

Model	Accuracy	F1-Score	ROC-AUC
Logistic Regression	Lower	Lower	Lower
SVM	Moderate	Moderate	Moderate
Random Forest	Moderate-High	Moderate-High	Moderate-High
Proposed N-XGB	94.72%	90.49%	98.57%

As shown in Table I, the proposed N-XGB model outperforms traditional standalone classifiers including Logistic Regression, Support Vector Machine, and Random Forest across all evaluation metrics. The superior ROC-AUC score of 98.57% indicates strong discriminative capability across all disaster classes, while the F1-score of 90.49% confirms that the model maintains a robust balance between precision and recall even for minority classes.

The web-based prediction interface, developed using Streamlit, successfully demonstrates the system's practical usability. Single-sample predictions for earthquake, flood, and wildfire inputs were confirmed as correct, with confidence scores of 100.0%, 89.9%, and 100.0% respectively. Bulk prediction functionality was also validated, with the system processing multiple records simultaneously and generating a prediction distribution consistent with the test dataset's composition.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VII July 2026- Available at www.ijraset.com

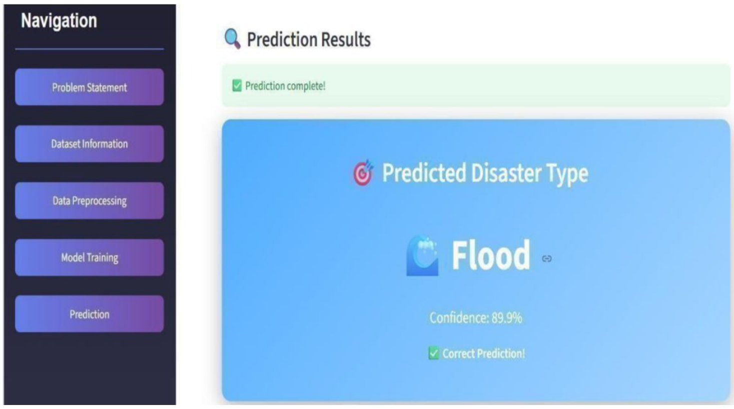


Figure 1. shows the predicted output for a given sample

The integration of neural network feature extraction with XGBoost classification proved effective in capturing complex nonlinear relationships within the disaster data. The neural network's fully connected architecture with ReLU activations and dropout regularization produced compact, discriminative feature representations that enhanced the downstream boosting classifier's performance. These results are consistent with findings reported in the primary reference study [1], which similarly achieved high accuracy on comparable disaster prediction tasks using the Neural-XGBoost framework.

VII. CONCLUSION

This paper presents a unified hybrid machine learning framework that combines Neural Networks and XGBoost for the multi-class classification and prediction of natural disasters including floods, wildfires, and earthquakes. The proposed Neural-XGBoost (N-XGB) model addresses three critical challenges in disaster prediction: complex nonlinear feature relationships, class imbalance in disaster datasets, and the limitations of standalone machine learning classifiers. By leveraging the feature extraction capability of deep neural networks and the ensemble classification power of XGBoost, combined with SMOTE-based data balancing, the system achieves a prediction accuracy of 94.72%, ROC-AUC of 98.57%, F1-score of 90.49%, and precision of 89.65%.

The experimental results confirm that the proposed hybrid approach outperforms traditional machine learning models including Logistic Regression, SVM, and Random Forest, and demonstrates reliable classification performance across all disaster categories including minority classes. The deployment of the model through a Streamlit-based web interface further validates its practical applicability for real-time disaster prediction and decision support. The findings of this work provide a strong foundation for intelligent disaster prediction systems that can support preparedness planning, resource allocation, and emergency response across governmental, humanitarian, and environmental management domains.

VIII. FUTURE SCOPE

The proposed N-XGB model demonstrates strong performance on the EM-DAT dataset; however, several opportunities exist for further enhancement. Future work may focus on integrating real-time data sources such as satellite imagery, IoT sensor streams, and weather monitoring systems to enable real-time disaster monitoring and early warning capabilities. Extending the dataset to include additional sources such as NOAA and NASA disaster databases would improve the generalization and geographic robustness of the model.

Incorporating Explainable Artificial Intelligence (XAI) techniques would improve the interpretability of predictions, allowing disaster management authorities to understand and trust the model's reasoning. Advanced hyperparameter optimization strategies such as Bayesian optimization could further enhance model performance. Integration of multimodal data including text reports, geospatial imagery, and sensor readings may significantly improve the model's ability to capture complex disaster patterns. Finally, optimizing the system for cloud-based or edge-based deployment would enhance scalability and practical accessibility in resource-constrained disaster management environments.

REFERENCES

[1] M. A. Saleem, W. Benjapolakul, W. Srisiri, S. Chaitusaney, and P. Kaewplung, "A hybrid prediction model integrating artificial intelligence and geospatial analysis for disaster management," IEEE Access, vol. 13, pp. 43716–43727, 2025.
[2] J. T. Thirukrishna, "An enhanced discovery of multiple natural disasters using machine learning model," Earth Sci. Informat., vol. 18, no. 3, p. 324, Sep. 2025.

[3] Universite Catholique de Louvain. (2024). CRED: Center of Research on the Epidemiology of Disasters. [Online]. Available: https://uclouvain.be/en/research-institutes/irss/cred.html

[4] C. I. Donatti et al., "Global hotspots of climate-related disasters," Int. J. Disaster Risk Reduction, vol. 108, Jun. 2024, Art. no. 104488.

[5] E. Tung, A. Mostafavi, M. Li, S. Li, Z. Rasheed, and K. Shafique, "A system for community lifeline and resource optimization in disaster management and planning," Int. J. Disaster Risk Reduction, vol. 119, Mar. 2025, Art. no. 105269.

[6] G. Airlangga, "Machine learning for tsunami prediction: A comparative analysis of ensemble and deep learning models," Kesatria: Jurnal Penerapan Sistem Informasi, vol. 6, no. 1, pp. 302–311, Jan. 2025.

[7] S. Sharma and A. Gosain, "Addressing class imbalance in remote sensing using deep learning approaches: A systematic literature review," Evol. Intell., vol. 18, no. 1, pp. 1–28, Feb. 2025.

[8] N. V. Chawla, A. Lazarevic, L. Hall, and K. W. Bowyer, "SMOTEBoost: Improving prediction of the minority class in boosting," in Proc. Eur. Conf. Princ. Data Mining Knowl. Discovery, 2003, pp. 107–119.

[9] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2016, pp. 785–794.

[10] R. L. Jones, A. Kharb, and S. Tubeuf, "The untold story of missing data in disaster research: A systematic review of the empirical literature utilising the EM-DAT," Environ. Res. Lett., vol. 18, no. 10, Oct. 2023, Art. no. 103006.

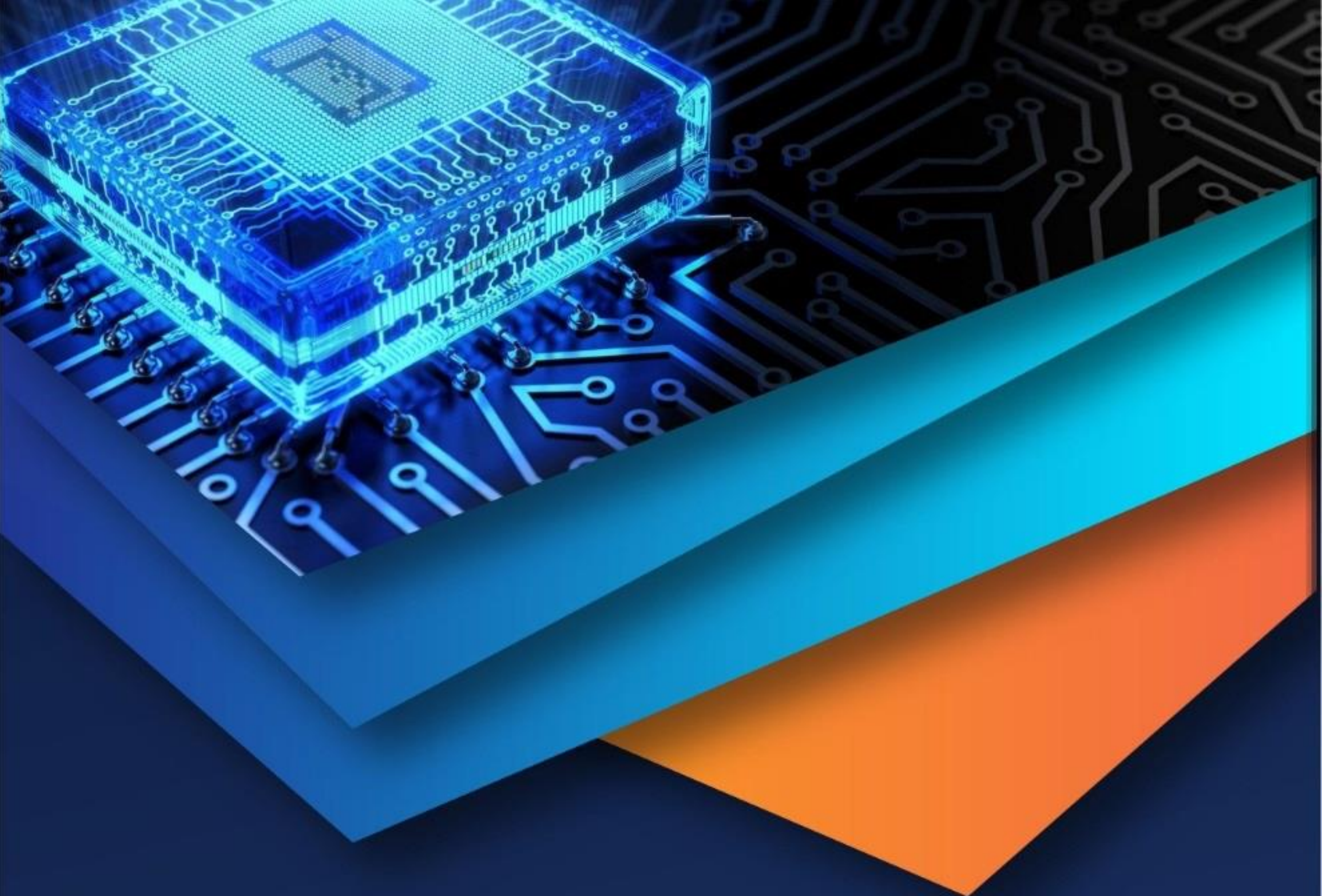
[11] P. Ghorpade et al., "Flood forecasting using machine learning: A review," in Proc. 8th Int. Conf. Smart Comput. Commun. (ICSCC), Jul. 2021, pp. 32–36.

[12] J. Jia and W. Ye, "Deep learning for earthquake disaster assessment: Objects, data, models, stages, challenges, and opportunities," Remote Sens., vol. 15, no. 16, p. 4098, Aug. 2023.

[13] G. Husain et al., "SMOTE vs. SMOTEENN: A study on the performance of resampling algorithms for addressing class imbalance in regression models," Algorithms, vol. 18, no. 1, p. 37, Jan. 2025.

[14] R. Cantini et al., "Harnessing prompt-based large language models for disaster monitoring and automated reporting from social media feedback," Online Soc. Netw. Media, vol. 45, Jan. 2025, Art. no. 100295.

[15] EM-DAT: The International Disaster Database. (2024). EM-DAT Data Repository. [Online]. Available: https://public.emdat.be/data



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY