IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, April 2025
Impact Factor: 7.67

Maharashtra's Climate and Environmental Analysis on Monsoon, Cyclone, Floods, and Earthquakes using Random Forest and LSTM

Mr. Swaroop Santosh Sutar, Mr. Aditya Suresh Namdas
Mr. Rushi Kishor Munde, Mr. Suyog Sanjay Pawar, Prof. Anurag Kumar
Department of Computer Engineering
DY Patil University, Ambi, Pune, India

Abstract: *Natural and human-induced disasters cause significant socio-economic and environmental harm, making accurate forecasting and efficient risk management essential. Traditional disaster prediction models often rely on statistical techniques, which struggle to capture the complex and nonlinear patterns within large datasets. However, advancements in machine learning (ML) and deep learning (DL) have enhanced disaster analysis by utilizing both historical and real-time data.*

This research introduces a hybrid model combining Random Forest (RF) and Long Short-Term Memory (LSTM) networks to improve disaster prediction and management. The RF algorithm, an ensemble learning technique, is used for selecting key features and classifying regions based on disaster vulnerability. Meanwhile, LSTM, a type of recurrent neural network, is employed for time-series forecasting of disaster events. By integrating these models, the system effectively analyses large datasets, leading to improved predictive accuracy and better decision-making.

The model's performance was evaluated using historical data on monsoons, cyclones, floods, and earthquakes in Maharashtra, India. Results show that the RF-LSTM hybrid model outperforms traditional ML and statistical approaches in identifying high-risk areas and forecasting disaster trends. The findings suggest that this approach enhances early warning systems, optimizes resource allocation, and strengthens disaster mitigation strategies.

This study contributes to AI-driven disaster management by presenting a scalable and efficient framework for analysing and predicting disaster patterns. The insights gained can assist policymakers, emergency responders, and disaster management agencies in implementing proactive risk reduction measures. Future research will focus on deploying the model in real-time environments, integrating it with IoT-based disaster monitoring systems, and optimizing it for large-scale disaster datasets..

Keywords: Disaster Management, Machine Learning, Deep Learning, Random Forest, LSTM, Time-Series Prediction, Risk Assessment, Early Warning Systems unforeseen..

I. INTRODUCTION

Disasters, both natural and human-induced, present serious risks to human life, infrastructure, and economies worldwide. Effective disaster management depends on accurate prediction, real-time monitoring, and swift response strategies to minimize damage and enhance resilience. The advancement of machine learning (ML) and deep learning (DL) has revolutionized disaster forecasting, risk assessment, and decision-making by improving upon traditional methods that primarily rely on historical data and heuristic approaches.

Among various ML and DL techniques, Random Forest (RF) and Long Short-Term Memory (LSTM) networks have shown considerable promise in disaster prediction and analysis. Random Forest, an ensemble learning algorithm, is highly effective in classification and regression tasks as it aggregates predictions from multiple decision trees, increasing accuracy and reducing overfitting. Meanwhile, LSTM, is particularly well-suited for time-series forecasting, as it captures long-term dependencies within sequential data.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, April 2025
Impact Factor: 7.67

This study integrates RF and LSTM models to enhance disaster prediction and response, with a focus on extreme weather events such as cyclones, floods, earthquakes, and monsoons. The key objectives of this research are:

- Analysing historical disaster data to identify critical patterns and risk factors.
- Developing a hybrid model that integrates RF's classification capabilities with LSTM's time-series forecasting strengths.
- Evaluating the model's effectiveness in predicting disaster occurrences and intensities.
- Offering actionable insights to disaster management agencies to improve preparedness and response strategies.
- Analysing historical disaster data to identify critical patterns and risk factors.
- Developing a hybrid model that integrates RF's classification capabilities with LSTM's time-series forecasting strengths.
- Evaluating the model's effectiveness in predicting disaster occurrences and intensities.
- Offering actionable insights to disaster management agencies to improve preparedness and response strategies.

Disasters, whether natural or caused by human activities, have devastating effects, impacting millions of lives and causing substantial economic losses. With climate change and rapid urbanization contributing to an increase in both the frequency and intensity of disasters, effective disaster management has become a global priority. Traditional disaster prediction and response methods, which rely on historical records, meteorological observations, and expert assessments, often fail to provide timely and highly accurate forecasts. However, recent advancements in ML and DL offer new opportunities by utilizing large-scale data, recognizing complex patterns, and enhancing predictive precision.

Among ML techniques, Random Forest has gained popularity due to its ability to process high-dimensional datasets, classify risks, and predict disaster occurrences with high accuracy. RF operates as an ensemble model, building multiple decision trees and aggregating them outputs to increase reliability while minimizing overfitting. This makes it particularly useful for categorizing disaster-prone regions, assessing vulnerabilities, and identifying key risk factors from meteorological, geological, and environmental data.

On the other hand, deep learning models, particularly LSTM networks, have been widely adopted for disaster forecasting because of their capability to process sequential and time-series data. LSTM, a specialized RNN, is designed to address the vanishing gradient problem by preserving long-term dependencies in data. This makes it highly effective for predicting monsoons, floods, cyclones, and earthquakes, where past trends and real-time data inputs are crucial for accurate forecasting.

Need for a Hybrid Approach

While both RF and LSTM are highly effective in their respective domains, they have limitations when used independently. RF excels at feature selection and classification but struggles with capturing sequential dependencies. LSTM is powerful for time-series forecasting but can be computationally expensive and sensitive to noisy data. A hybrid RF-LSTM model can capitalize on the strengths of both techniques:

- RF extracts essential features and classifies disaster risks based on historical data.
- LSTM models time-series dependencies, enhancing real-time disaster prediction accuracy.

This research introduces a hybrid RF-LSTM model for disaster management, aimed at improving early warning systems, risk assessment, and mitigation efforts. The primary objectives of this study include:

1. Data Collection & Preprocessing – Gathering and refining large-scale disaster-related datasets, including meteorological, seismic, and hydrological records.
2. Feature Engineering & Selection – Employing RF to identify the most significant factors influencing disaster occurrences.
3. Model Development – Creating a hybrid predictive model that combines RF for classification and LSTM for forecasting.
4. Performance Evaluation – Measuring the model's accuracy and efficiency in comparison with traditional ML and statistical methods.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 9, April 2025 Impact Factor: 7.67

Application in Disaster Management – Delivering practical insights for government agencies, emergency responders, and policymakers to strengthen disaster preparedness and mitigation strategies.

II. MATERIALS AND METHODS

1. Data Collection and Preprocessing

1.1 Dataset Selection

This study utilizes historical disaster-related datasets obtained from multiple credible sources, including:

- Indian Meteorological Department (IMD): Data on cyclones, monsoons, and rainfall patterns.
- National Disaster Management Authority (NDMA): Records of floods, earthquakes, and extreme weather incidents.
- USGS Earthquake Catalogue: Seismic activity data relevant to Maharashtra.
- Remote Sensing and GIS Data: Satellite-based environmental and climatic information.
- Hydrological and Weather Stations: Temperature, humidity, wind speed, and atmospheric pressure readings.
- The dataset spans over two decades (2000–2024) to ensure comprehensive training and validation of the proposed model.

1.2 Data Preprocessing

To enhance model accuracy and reliability, the following preprocessing steps were applied:

- Data Cleaning: Addressing missing values, removing outliers, and eliminating redundant entries using interpolation and data imputation techniques.
- Feature Engineering: Extracting key attributes such as rainfall intensity, soil moisture, seismic activity levels, and cyclone wind speeds.
- Normalization & Standardization: Applying min-max scaling and z-score normalization to maintain consistency in feature magnitudes.
- Temporal Alignment: Synchronizing time-series data from multiple sources to optimize LSTM processing.

III. METHODOLOGY

3.1 Model Architecture

The hybrid RF-LSTM model integrates machine learning and deep learning techniques to enhance disaster analysis and forecasting:

Random Forest (RF):

- Used for feature selection and classification of disaster risk levels.
- Constructs multiple decision trees to determine disaster-prone areas.
- Outputs categorical risk assessments (e.g., low, moderate, high, extreme).

Long Short-Term Memory (LSTM):

- A specialized deep learning model for analysing time-series data.
- Captures long-term dependencies and sequential patterns in disaster trends.
- Forecasts future disaster occurrences and intensities

3.2 Model Implementation Steps

Feature Selection using Random Forest

- RF is trained on historical disaster data to identify the most influential predictive features.
- The most relevant features are selected for further processing in the LSTM model.

Disaster Time-Series Forecasting using LSTM

- A stacked LSTM architecture is implemented with multiple layers for improved feature extraction.
- The model is trained using a sliding window approach, leveraging past disaster events to predict future occurrences.

DOI: 10.48175/IJARSCT-25780

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, April 2025
Impact Factor: 7.67

- Hyperparameters such as neuron count, dropout rates, and learning rates are fine-tuned using Bayesian optimization.

Integration of RF and LSTM Outputs

- RF categorizes risk zones, while LSTM predicts temporal disaster trends.
- A final decision fusion layer integrates both outputs to enhance overall predictive accuracy.

3.3 Model Training and Validation

Data Splitting:

70% Training, 20% Validation, 10% Testing.

Evaluation Metrics:

- For Random Forest: Accuracy, Precision, Recall, F1-score, AUC-ROC.
- For LSTM: RMSE (Root Mean Squared Error), MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error).

Cross-Validation:

- K-Fold Cross-Validation (k=5) was applied to enhance model generalization.

Optimization Techniques:

- LSTM Training: Adam Optimizer.
- Hyperparameter Tuning: Grid Search andRandom Search.

3. Tools and Technologies Used

The study was implemented using the following tools and frameworks:

Programming Language: Python 3.x.

Libraries & Frameworks:

- Scikit-learn: For Random Forest classification.
- TensorFlow/Keras: For LSTM model development.
- Pandas & NumPy: For data preprocessing and manipulation.
- Matplotlib & Seaborn: For data visualization.

Computing Environment:

- NVIDIA GPU: For accelerated LSTM training.
- Google Colab/Jupyter Notebook: For experimentation and implementation**.**

4. Experimental Setup & Case Study

A case study was conducted in Maharashtra, India, focusing on high-risk disaster events, including monsoons, cyclones, floods, and earthquake-prone regions. The model was trained using real-world disaster datasets, and its predictions were validated against actual occurrences from recent years.

IV. RESULTS AND DISCUSSION

1. Model Performance Evaluation

The effectiveness of the RF-LSTM hybrid model was assessed using historical datasets related to monsoons, cyclones, floods,and earthquakes in Maharashtra, India. The model's performance was compared with conventional machine learning models and statistical approaches to determine its predictive accuracy and classification efficiency.

Performance Metrics

To measure the reliability of the model, various evaluation metrics were used:

Random Forest Classification Metrics (for disaster risk classification):

- Accuracy: Indicates the overall correctness of classification.
- Precision & Recall: Measures the ability to correctly identify disaster-prone regions.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, April 2025
Impact Factor: 7.67

- F1-Score: Balances precision and recall to optimize classification performance.
- AUC-ROC Curve: Assesses the classification efficiency at different threshold levels.

LSTM Time-Series Forecasting Metrics (for disaster trend prediction):

- Root Mean Squared Error (RMSE): Quantifies the error magnitude.
- Mean Absolute Percentage Error (MAPE): Evaluates relative prediction accuracy.
- Mean Absolute Error (MAE): Measures the average deviation of predictions from actualvalues.

1.2 Experimental Results

The RF-LSTM model was tested using real-world disasterdatasets, and the obtained results are summarized below.

Table 1: Random Forest Classification Performance

Disaster Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
Flood	92.4	89.6	91.2	90.4	95.3
Cyclone	94.1	91.8	93.2	92.5	96.7
Earthquake	90.5	88.2	89.5	88.8	93.1
Monsoon	95.3	94.0	94.8	94.4	97.5

The RF model achieved high classification accuracy across different disaster types, demonstrating its effectiveness in identifying high-risk zones based on historical disaster occurrences and environmental parameters.

Table 2: LSTM Time-Series Forecasting Performance

Disaster Type	RMSE	MAE	MAPE (%)
Flood	4.12	2.35	7.89
Cyclone	3.78	2.19	6.92
Earthquake	4.56	2.78	8.45
Monsoon	3.12	1.98	5.67

The low RMSE and MAPE values indicate that the LSTM model effectively captures time-series trends, making it well-suited for disaster forecasting.

1.3 Comparative Analysis

To further assess the effectiveness of the RF-LSTM model, its performance was compared against standalone models such as Random Forest, LSTM, Decision Trees, Support Vector Machines (SVM), and the ARIMA statistical model.

Table 3: Model Comparison for Disaster Forecasting

Model	Accuracy (%)	RMSE	MAPE (%)	Execution Time (sec)
Decision Tree	85.2	6.21	12.45	2.8
SVM	88.5	5.98	10.34	5.2
ARIMA	82.3	7.12	14.02	6.7
Random Forest	92.4	5.43	9.15	3.4
LSTM	94.5	4.11	6.92	12.3

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 9, April 2025
Impact Factor: 7.67

Model	Accuracy (%)	RMSE	MAPE (%)	Execution Time (sec)
RF-LSTM (Proposed)	96.2	3.78	5.67	9.8

The proposed RF-LSTM model demonstrated superior performance across multiple metrics, achieving the highest accuracy while maintaining a lower prediction error. This highlights the advantages of combining RF's feature selection with LSTM's time-series forecasting capabilities.

2. Discussion

2.1 Effectiveness of the RF-LSTM Hybrid Model:

- The integration of Random Forest for feature selection and LSTM for sequential prediction resulted in improved disaster risk classification and forecasting accuracy.
- High AUC-ROC scores confirm that the model effectively distinguishes between different risk levels.
- Low RMSE and MAPE values indicate that LSTM efficiently captures long-term dependencies in disaster trends.

2.2 Key Findings and Insights:

- Flood and monsoon predictions exhibited the highest accuracy, benefiting from extensive meteorological data availability.
- Earthquake forecasting showed relatively higher errors due to the inherent unpredictability of seismic events.
- Random Forest's classification effectively pinpointed high-risk zones, offering valuable insights for disaster preparedness.
- The model's computational efficiency makes it suitable for real-time disaster monitoring applications.

2.3 Limitations and Challenges:

- Data Quality Issues: Incomplete or inconsistent records posed challenges in model performance.
- Real-Time Application Constraints: The study relied on historical datasets; real-time deployment requires further optimization.
- Computational Complexity: Training deep learning models like LSTM demands significant GPU resources and time.

2.4 Future Scope

To enhance the model's real-world applicability, the following improvements are recommended:

- Integration with IoT-based monitoring systems to enable real-time disaster forecasting.
- Enhanced earthquake prediction using seismic wave analysis combined with deep learning architectures such as RF-LSTM.
- Development of a web-based or mobile application for real-time risk assessment and early warning dissemination.
- Exploration of attention mechanisms in deep learning to further improve long-term dependency modelling in LSTM networks.

V. CONCLUSION

The integration of LSTM and Random Forest models provides a comprehensive approach to analyzing Maharashtra's climate-related hazards. While LSTM effectively forecasts monsoon trends, Random Forest aids in classifying high-risk areas. These findings can contribute to enhancing early warning systems and disaster management strategies. Future research will focus on incorporating real-time data to further improve predictive accuracy.

REFERENCES

[1]. M. A. Patil, "Flood Prediction Using Artificial Intelligence: Western Maharashtra Case Studies," International Journal of Disaster Risk Science, vol. 15, no. 2, pp. 120-135, 2024.

[2]. S. Sharma and R. Kumar, "Towards Efficient Disaster Management: Role of Machine Learning Algorithms," International Journal of Intelligent Systems and Applications in Engineering, vol. 9, no. 3, pp. 145-152, 2023.

[3]. A. Gupta, P. Singh, and M. Verma, "Quantitative Analysis of Natural Calamities for Disaster Management through Machine Learning Techniques Using Satellite Data," Pacific and Asian Journal of Energy, vol. 31, no. 4, pp. 210-225, 2023.

[4]. L. Zhang and Y. Li, "Machine Learning Applications in Flood Forecasting and Predictions: A Review," AIMS Environmental Science, vol. 12, no. 1, pp. 50-70, 2025.

[5]. B. Singh and N. Patel, "Role of Machine Learning, Deep Learning, and WSN Technologies in Disaster Management," International Journal of Intelligent Systems and Applications in Engineering, vol. 9, no. 4, pp. 200-210, 2023.

[6]. C. Wang and H. Zhao, "Machine Learning in Earth Sciences: Applications and Challenges," Earth Science Reviews, vol. 220, pp. 103-118, 2024.