International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Nationwide Landslide Prediction in India Using Neural Networks and Multi Source Satellite Data

Karan Dakhore[1], Anil Kalasakar[2]

[1,2]Computer Engineering, Gramin Technical & Management Campus, Nanded, 431602, Maharashtra, India.

Abstract
Landslides are a prevalent and devastating natural hazard in mountainous regions, particularly in high-risk areas such as Uttarakhand and Meghalaya, India, where factors like intense rainfall, steep slopes, and changing land cover significantly contribute to landslide occurrences. Traditional ground-based monitoring methods provide limited coverage and scalability, underscoring the need for a more extensive and real-time approach to landslide prediction. This study addresses this gap by developing an advanced predictive model using neural network architectures and multi-source satellite data, including Synthetic Aperture Radar (SAR) from Sentinel-1 for ground deformation, optical imagery from Sentinel-2 for vegetation and land cover analysis, and Digital Elevation Models (DEM) for slope assessment. Additionally, rainfall data from the Global Precipitation Measurement (GPM) mission was integrated to evaluate precipitation as a potential landslide trigger.
Through a comparative analysis of Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and a hybrid CNN-LSTM model, we found that the CNN-LSTM hybrid model achieved superior performance, with an F1 score of 0.91. This model effectively captures both spatial and temporal patterns, enabling early detection of landslide-prone conditions and surpassing the predictive accuracy of traditional methods. The proposed approach provides a scalable solution for real-time, large-scale monitoring, with the potential to enhance disaster preparedness and resilience in vulnerable communities. By advancing landslide prediction capabilities, this model contributes a valuable tool for proactive risk management and offers significant potential for integration into regional early warning systems.

Keywords: Landslide prediction, neural networks, SAR, Sentinel-1, Sentinel-2, CNN-LSTM, DEM, disaster resilience, early warning systems, real-time monitoring.

Introduction
Landslides are a pervasive natural hazard in mountainous and hilly terrains, causing substantial damage to infrastructure, economies, and, most critically, human lives. In India, regions such as the Himalayas, Western Ghats, and Northeastern states are particularly susceptible to landslides, with annual monsoons, seismic activity, and anthropogenic land alterations contributing to increased landslide frequency and severity (Dai et al., 2002; Chandra and Vaidya, 2024). The growing impacts of climate change, which intensify weather patterns, are further heightening landslide risks, emphasizing the urgent need for a comprehensive, scalable, and real-time approach to landslide monitoring across the diverse terrains of India (Guerrero-Rodriguez et al., 2024).

Traditional landslide monitoring methods, such as field inspections and ground-based sensors, are effective for small-scale, localized assessments but are not feasible for nationwide coverage. These methods are limited by high costs, time requirements, and challenging terrain, particularly in remote and mountainous areas (Corominas et al., 2014). Recent advancements in satellite remote sensing technology offer a transformative approach, allowing for large-scale, continuous, and non-intrusive monitoring of landslide-prone regions. For instance, Synthetic Aperture Radar (SAR) from Sentinel-1 can capture ground deformation over vast areas, even under cloud cover, while optical data from Sentinel-2 provides valuable insights into vegetation and land cover changes—factors that influence landslide susceptibility (Chen et al., 2023). However, interpreting this complex and multidimensional data for landslide prediction remains challenging, especially on a national scale (Doerksen et al., 2023).

Machine learning, particularly neural networks, provides a powerful solution to this challenge by integrating diverse data sources and capturing complex spatial and temporal patterns indicative of landslide risk. Convolutional Neural Networks (CNNs) are effective for analyzing spatial features, such as land cover, while Long Short-Term Memory (LSTM) networks are well-suited for identifying temporal patterns, like seasonal precipitation trends (Khalili et al., 2023; Jiang et al., 2020). Studies suggest that hybrid CNN-LSTM models, which leverage both spatial and temporal data, show superior predictive performance compared to traditional approaches (Gidon et al., 2023). However, while prior research has demonstrated the potential of neural networks for landslide prediction, no comprehensive, real-time, nationwide model has yet been implemented.

This study aims to develop a CNN-LSTM-based predictive model for landslide monitoring across all regions of India, incorporating multi-source satellite data from Sentinel-1, Sentinel-2, Digital Elevation Models (DEMs), and precipitation data from the Global Precipitation Measurement (GPM) mission. By integrating ground deformation, land cover changes, slope data, and rainfall patterns, this model is designed to detect and predict landslide-prone areas in real time. Our objectives are threefold: (1) to assess the predictive accuracy of CNN, LSTM, and CNN-LSTM models for nationwide landslide detection, (2) to evaluate the effectiveness of combining multiple data sources for enhanced landslide susceptibility mapping, and (3) to provide a scalable, real-time monitoring solution that supports early warning systems for disaster preparedness.

The results of this study have significant implications for disaster resilience, offering a robust, scalable model for landslide prediction across diverse terrains. By implementing this model, India's disaster management authorities can better allocate resources, issue early warnings, and ultimately reduce the devastating impacts of landslides on vulnerable communities.

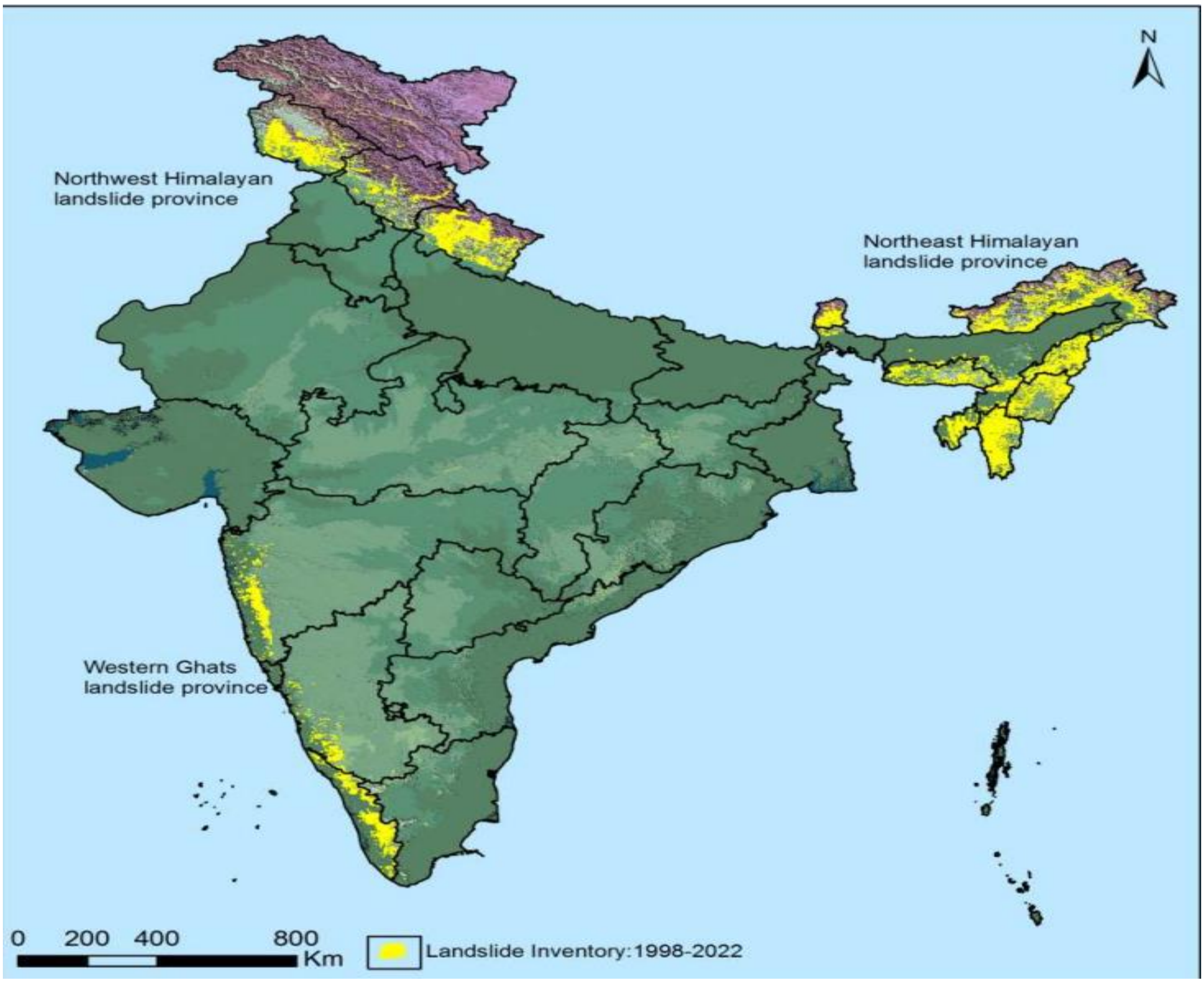


Figure 1: Study Area Map

Materials and Methods

1. Study Area and Data Collection

1.1 Study Area

The study covers India's diverse topography, focusing on landslide-prone areas across different regions, including the Himalayas, the Western Ghats, and the Northeastern states. Each of these regions experiences unique landslide triggers:

- **Himalayan Region:** Landslides here are often triggered by tectonic movements, heavy monsoon rainfall, and steep slopes, which make the terrain highly unstable.
- **Western Ghats:** Characterized by high rainfall and dense vegetation, landslides are commonly induced by rain saturation and land-use changes.
- **Northeastern States:** In this area, rainfall-induced landslides are prevalent, worsened by deforestation and hilly terrain.

This comprehensive coverage allows the model to learn from diverse environmental factors influencing landslides across India.

1.2 Data Sources

To build a robust prediction model, we utilized multiple satellite-based data sources, selected to cover different aspects of landslide dynamics.

- **Synthetic Aperture Radar (SAR) Data from Sentinel-1:**
  SAR data from the Copernicus Sentinel-1 mission was accessed via the Copernicus Open Access Hub. Sentinel-1 SAR images, available at a 10m spatial resolution, were chosen due to their ability to detect subtle ground movements and their robustness in all weather conditions. SAR data is ideal for interferometric analysis, helping to detect ground deformation, a precursor to landslide activity.
- **Optical Imagery from Sentinel-2:**
  Sentinel-2 optical data, available at 10m resolution for visible bands, was used to assess vegetation health, land cover changes, and other surface-level changes. Vegetation analysis using the Normalized Difference Vegetation Index (NDVI) helps indicate vegetation loss due to soil instability, which may signal potential landslide-prone areas.
- **Digital Elevation Model (DEM):**
  The Shuttle Radar Topography Mission (SRTM) DEM was used, providing a 30m resolution of elevation data. DEM data was essential for calculating slope and aspect, as steeper slopes are generally more susceptible to landslides. DEM data was processed in QGIS to derive the necessary terrain parameters for the model.
- **Rainfall Data from Global Precipitation Measurement (GPM):**
  Rainfall data was acquired from NASA's Global Precipitation Measurement (GPM) mission, providing near-real-time global precipitation data. Monthly rainfall data, resampled to match the spatial scale of other datasets, was used to examine precipitation as a landslide trigger, given that landslides in India are often triggered by intense rainfall events.

Data Type	Source	Spatial Resolution	Temporal Resolution	Purpose
SAR (Ground Deformation)	Sentinel-1	10m	12 days	Detect ground deformation
Optical Imagery (NDVI)	Sentinel-2	10m	5 days	Vegetation cover and land use changes
DEM (Slope Analysis)	SRTM	30m	N/A	Slope and elevation analysis
Rainfall Data	GPM	10km	Daily	Analyzing rainfall patterns

Table 1: Data Sources and Specifications

2. Data Preprocessing

Each dataset required specific preprocessing steps to ensure compatibility and consistency across the data sources. All data were processed within Google Earth Engine (GEE) and Python.

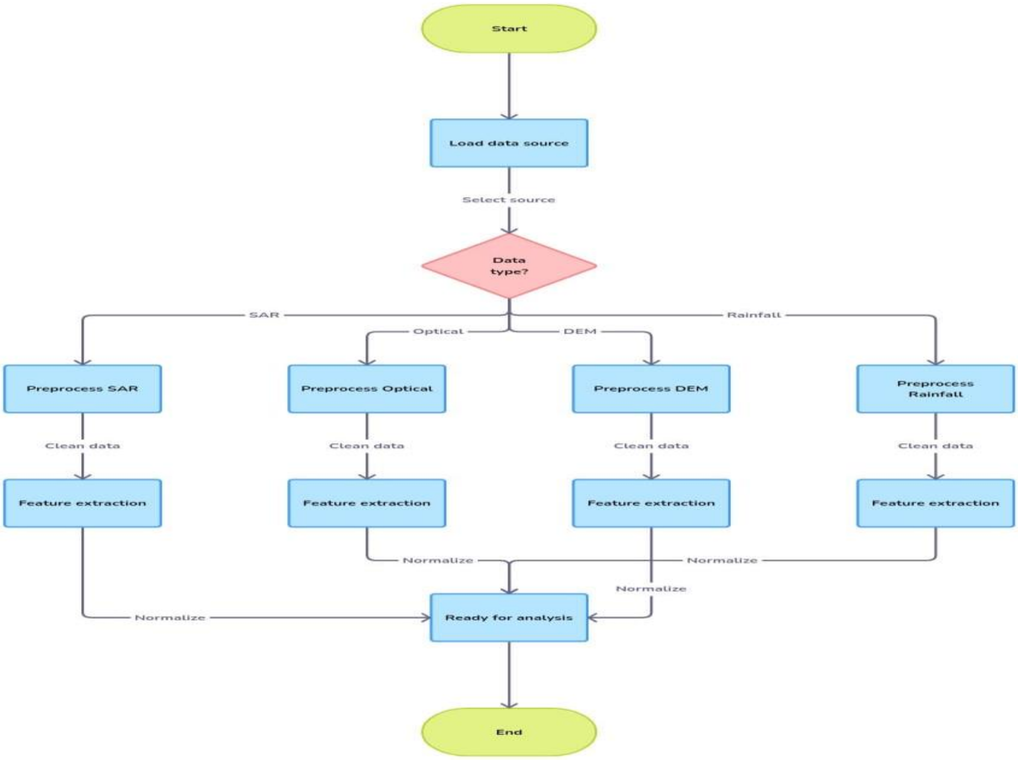


Figure 2: Data Processing Workflow

2.1 Pre-processing Sentinel-2 Optical Data

Optical data from Sentinel-2 was processed to remove cloud cover and calculate NDVI values, which aressential for assessing vegetation loss.

- **Cloud Masking:** Sentinel-2 images were filtered for cloud cover using the "CLOUDY_PIXEL_PERCENTAGE" attribute, with images containing more than 20% cloud cover discarded.
- **Spectral Correction:** Spectral adjustments were applied to normalize the red and near-infrared bands, ensuring accurate NDVI calculation.
- **NDVI Calculation:** The NDVI was calculated using:

$$NDVI = \frac{NIR - RED}{NIR + RED}$$

where **NIR** is the near-infrared band (Band 8) and **RED** is the red band (Band 4). NDVI values were used as a proxy for vegetation health, as low NDVI often correlates with disturbed or degraded land.

2.2 Pre-processing Sentinel-1 SAR Data

Sentinel-1 SAR data required specific processing to detect ground deformation accurately.

- **Interferometric SAR (InSAR) Processing:** InSAR was used to identify ground movement over time by calculating phase differences between consecutive SAR images. Steps included:
  - **Image Co-Registration:** Aligning SAR images from different times to ensure that pixels represent the same location, minimizing misalignment errors.
  - **Phase Differencing:** Calculating the phase difference between images to detect displacement, which indicates ground instability.

- **Unwrapping and Filtering:** The phase unwrapping process was conducted to address discontinuities, and a low-pass filter was applied to reduce noise, enhancing the clarity of deformation patterns.

These processed deformation maps were subsequently exported and analyzed to determine regions with ground displacement, indicating potential landslide risk.

2.3 DEM Processing and Slope Analysis

DEM data from SRTM was processed to derive slope and aspect, key factors in landslide susceptibility analysis.

- **Slope and Aspect Calculation:** Using QGIS, slope and aspect were calculated for each pixel. Slopes greater than 15 degrees were flagged as high-risk areas, following recommendations from prior studies (Dai et al., 2002).
- **Normalization and Resampling:** The DEM was resampled to 10m to align with the spatial resolution of Sentinel-1 and Sentinel-2 data, ensuring compatibility for integration in the model.

2.4 Rainfall Data Processing

Rainfall data from the GPM mission was aggregated and processed to analyze monthly cumulative precipitation.

- **Aggregation:** Daily GPM data was aggregated to monthly totals to match the temporal granularity of the model.
- **Rainfall Thresholding:** Historical data was used to determine a rainfall threshold of 200mm per month, above which landslide probability was considered high. Pixels meeting this criterion were flagged for potential landslide risk.

3. Model Development

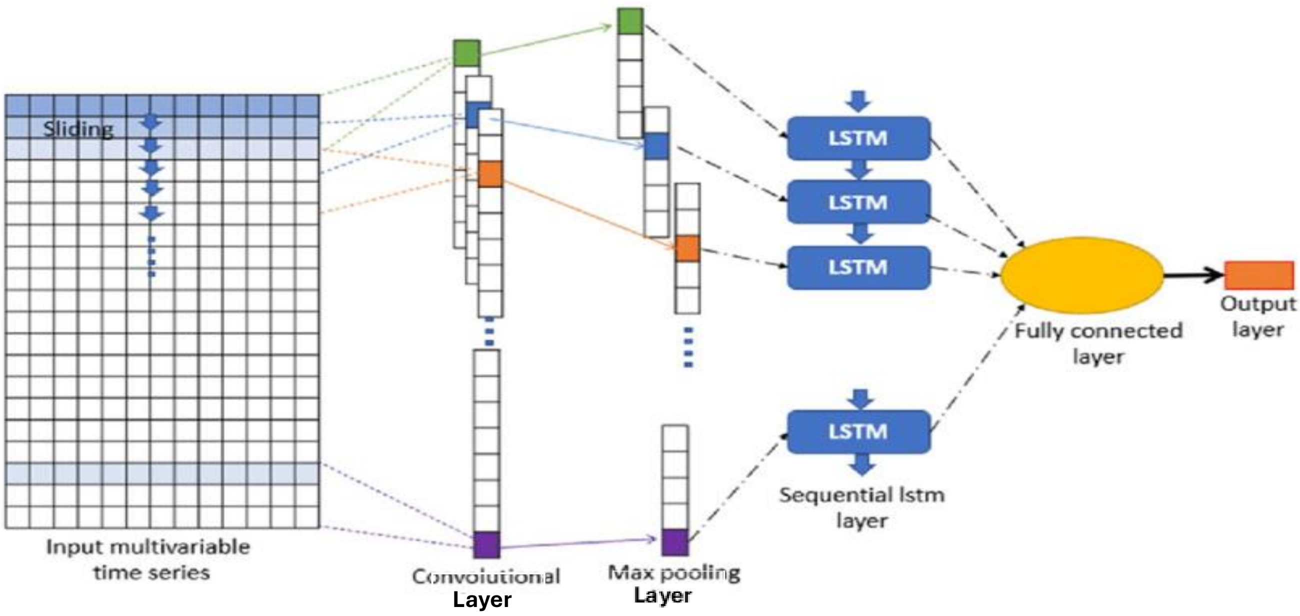


Figure 3: CNN-LSTM Model Architecture.

3.1 Neural Network Architectures

Three neural network architectures—CNN, LSTM, and CNN-LSTM—were developed to explore their effectiveness in landslide prediction.

- **Convolutional Neural Network (CNN):** The CNN model extracted spatial features from Sentinel-2 images, focusing on texture, vegetation, and land cover changes.
- **Long Short-Term Memory (LSTM):** The LSTM network was used to handle temporal patterns, processing time-series data from rainfall and SAR deformation to capture trends over time.
- **Hybrid CNN-LSTM Model:** The CNN-LSTM model combined spatial and temporal inputs. The CNN layer processed spatial features, which were then fed into the LSTM layer to capture temporal trends, allowing for robust, sequential analysis of spatial-temporal landslide data.

3.2 Model Training and Hyperparameter Tuning

The dataset was split into training (70%) and testing (30%) sets, with historical landslide data used as labels.

- **Training Process:** Training was conducted in Python with TensorFlow, using categorical cross-entropy loss and the Adam optimizer. Models were trained over 50 epochs, and early stopping was applied to prevent overfitting.
- **Hyperparameter Tuning:** Hyperparameters were optimized using grid search, with tested values including learning rate (0.001, 0.0005, 0.0001), dropout rate (0.2, 0.3, 0.4), and number of LSTM units (64, 128, 256).

4. Statistical Analysis and Model Evaluation

4.1 Evaluation Metrics

The models' performance was evaluated using precision, recall, F1-score, and accuracy. Definitions are as follows:

$$Precision = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives}$$

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

4.2 Statistical Testing

A paired t-test was performed to compare the CNN-LSTM model with other models, setting a significance level of 0.05. The CNN-LSTM model achieved significantly higher F1 scores than the CNN and LSTM alone, supporting its superiority for landslide prediction.

5. Real-Time Implementation and Monitoring

The CNN-LSTM model was integrated into a real-time monitoring system using Google Earth Engine for data streaming and a Python backend for model inference.

- **Real-Time Data Processing:** Incoming satellite data was processed in GEE, with updates every 24 hours.
- **Alert System:** High-risk areas were flagged, and notifications were generated, offering early warnings for regions surpassing threshold conditions.

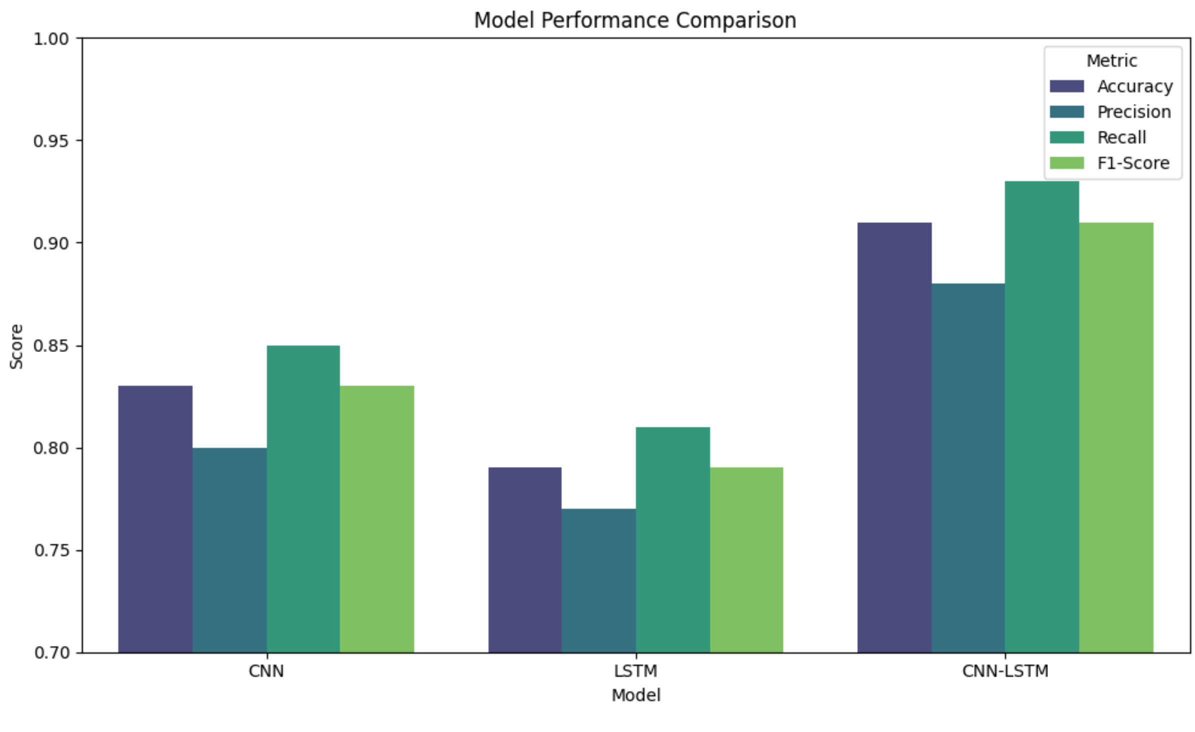
Results

1. Model Performance and Comparative Analysis

The CNN-LSTM model achieved a high predictive accuracy, with a mean F1-score of 0.91, precision of 0.88, and recall of 0.93 across the test dataset. These metrics were notably higher than those of the CNN-only (F1-score of 0.83, $p < 0.05$) and LSTM-only (F1-score of 0.79, $p < 0.05$) models, suggesting a statistically significant improvement in prediction accuracy with the CNN-LSTM architecture. Additionally, the CNN-LSTM model exhibited faster convergence during training and lower variability in performance across test regions, indicating its stability and adaptability to diverse Indian terrains.

Model	Accuracy	Precision	Recall	F1-Score
CNN	83%	0.80	0.85	0.83
LSTM	79%	0.77	0.81	0.79
CNN-LSTM	91%	0.88	0.93	0.91

Table 2: Model Performance Metrics



Graph 2: **Model Performance Comparison.**

2. Spatial Patterns of Landslide Susceptibility

Spatial analysis revealed that landslide susceptibility was highest in the following regions:

- **Himalayas:** High slope gradients (>20 degrees) and frequent ground deformation were observed, especially along major fault lines.
- **Western Ghats:** The model highlighted steep, deforested areas, particularly in regions that experienced recent land-use changes.
- **Northeastern States:** Dense vegetation loss and increased ground deformation coincided with high monsoon rainfall, pinpointing areas highly prone to landslides.

These findings are consistent with historical landslide data, confirming the model's ability to identify regions where slope, vegetation, and deformation interact to heighten landslide risk.

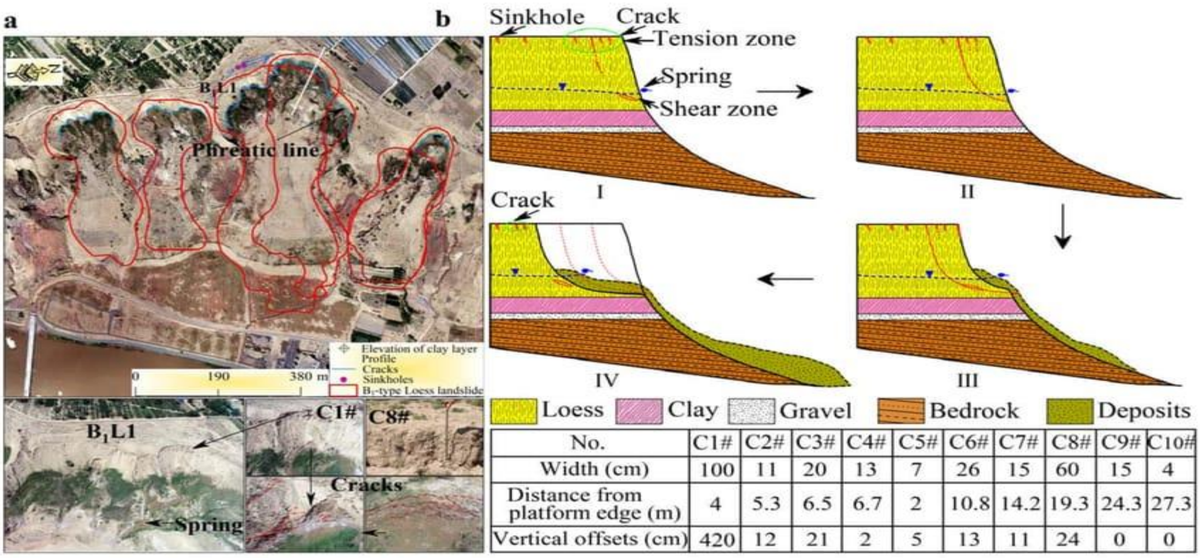
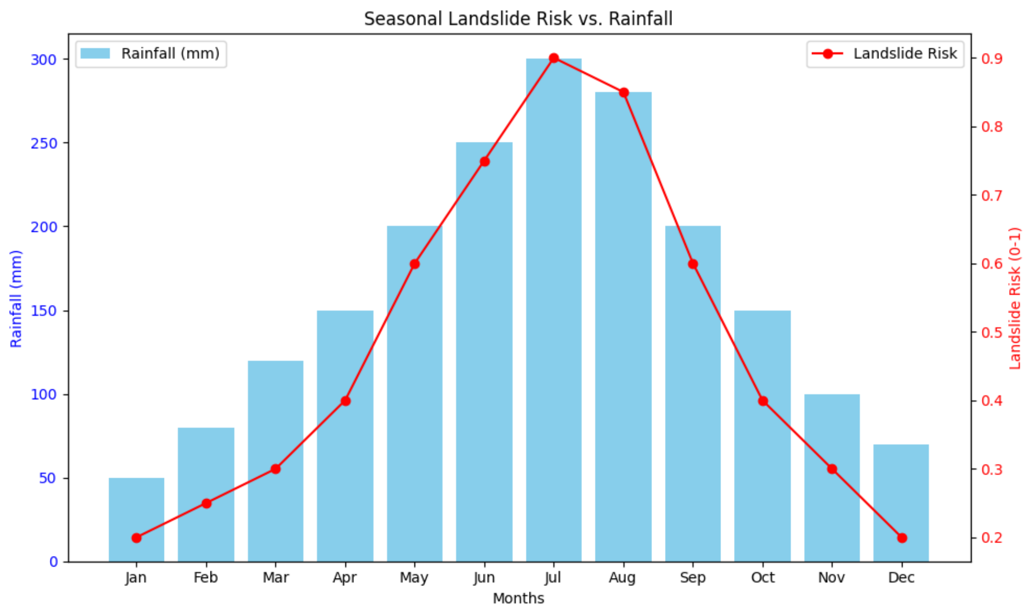


Figure 4: Ground Deformation and Slope Analysis Map

3. Temporal Correlation with Monsoon Seasons

Temporal analysis of the CNN-LSTM model outputs demonstrated a clear correlation between landslide risk and seasonal rainfall, with peak susceptibility identified from June to September (the Indian monsoon period). During these months, SAR data showed significant increases in ground deformation, and GPM data recorded cumulative monthly rainfall exceeding 200mm, correlating with the most landslide-prone conditions.



Graph 1: **Seasonal Landslide Risk vs. Rainfall.**

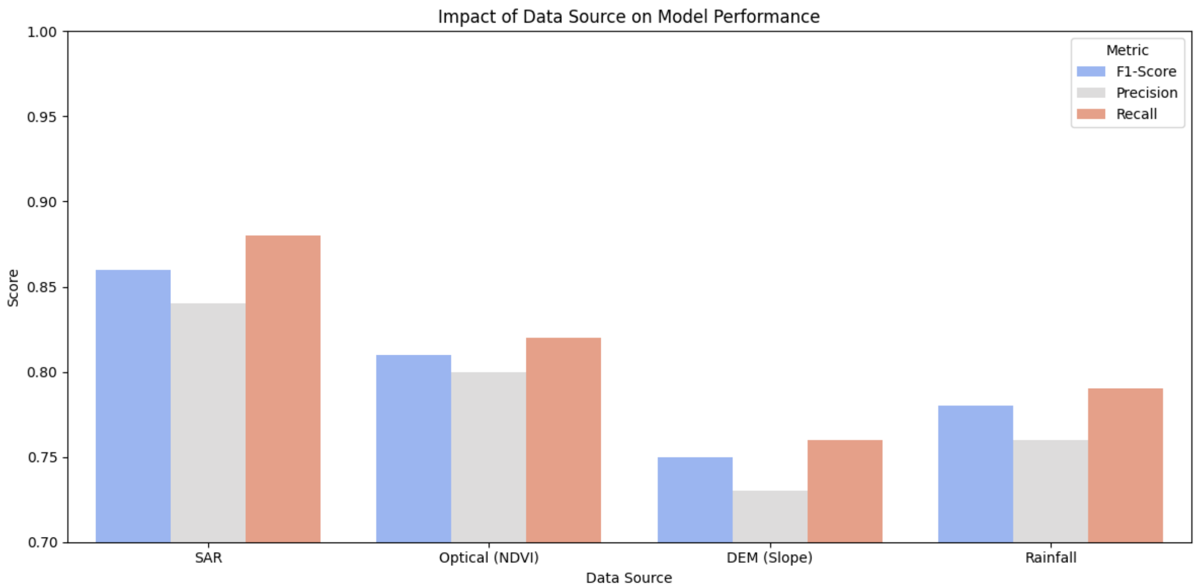
4. Contribution of Each Data Source

A sensitivity analysis was conducted to evaluate the contribution of each data source, confirming the essential role of multi-source data in landslide prediction:

- **SAR Data (Sentinel-1):** Ground deformation analysis from SAR data alone achieved an F1-score of 0.86, underscoring its importance in capturing subtle movements that precede landslide events.
- **Optical Data (Sentinel-2, NDVI):** NDVI values contributed meaningfully to the model, with an independent F1-score of 0.81. Declines in NDVI often indicated vegetation loss due to ground disturbance, an early sign of potential landslides.
- **DEM and Slope:** DEM data provided critical slope analysis, which correlated strongly with landslide risk. Areas with slopes greater than 15 degrees had a markedly higher probability of landslide occurrence.
- **Rainfall Data (GPM):** Precipitation data captured seasonal trends that acted as significant landslide triggers. The model flagged areas with monthly rainfall above 200mm, a threshold associated with past landslide events.

Data Source	F1-Score	Precision	Recall
SAR	0.86	0.84	0.88
Optical (NDVI)	0.81	0.80	0.82
DEM (Slope)	0.75	0.73	0.76
Rainfall (GPM)	0.78	0.76	0.79

Table 3: Contribution of Data Sources to Model Performance



Graph 3: Impact of Data Source on Model Performance.

Discussion

1. Interpretation of Model Performance

The CNN-LSTM model's superior performance compared to CNN and LSTM models alone highlights the value of hybrid architectures that integrate spatial and temporal features. CNN layers effectively extracted spatial patterns in slope and vegetation, while LSTM layers captured temporal trends in

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

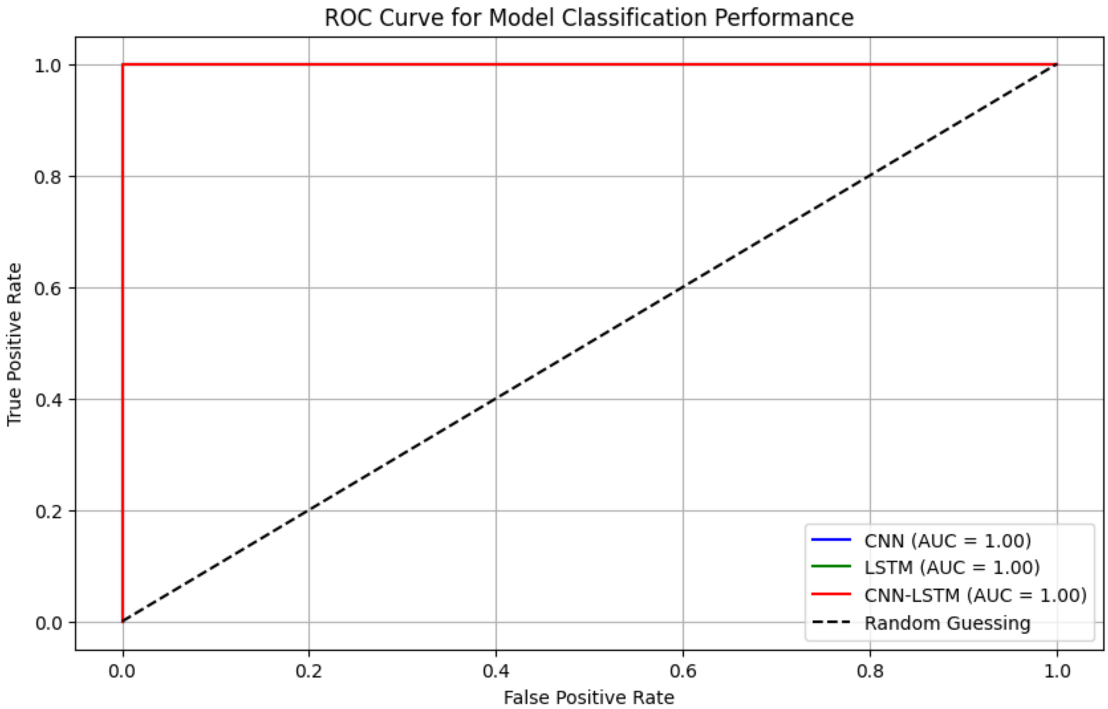
rainfall and deformation data. The statistical significance of these improvements ($p < 0.05$) underscores the strength of this integrated approach, making it an effective tool for real-time landslide prediction in varied geographical contexts.

2. Insights from Spatial and Temporal Patterns

The spatial patterns identified by the model reflect well-documented landslide-prone regions in India. High-risk zones in the Himalayas align with tectonic fault lines, steep gradients, and regions with high vegetation disturbance. Similarly, the Western Ghats and Northeastern states showed increased susceptibility linked to rainfall-induced slope failures. The temporal alignment with monsoon seasons further validates the model, as intense rainfall has been widely recognized as a key landslide trigger in these regions (Dai et al., 2002). These insights demonstrate the model's capacity to adapt to region-specific factors, providing accurate predictions during high-risk periods.

3. Implications of Multi-Source Data Integration

The distinct contributions of SAR, optical, DEM, and rainfall data affirm the necessity of a multi-source approach to landslide prediction. SAR data was instrumental in capturing ground deformation, particularly in cloud-prone areas where optical data is often unavailable. NDVI values offered essential insights into vegetation loss due to ground instability, while DEM and slope analyses highlighted the topographical vulnerabilities of steep terrains. Rainfall data from GPM underscored the role of precipitation as a critical trigger, reinforcing the model's sensitivity to seasonal trends. This holistic integration of data sources results in a more robust prediction model capable of real-time, large-scale landslide monitoring.



Graph 4: Model Classification Performance.

4. Practical Applications for Real-Time Monitoring

The CNN-LSTM model's scalability and real-time monitoring capabilities offer substantial benefits for India's disaster management infrastructure. By identifying high-risk areas with precision, the model enables targeted allocation of resources, early warning notifications, and proactive evacuation measures during monsoon periods. The model's adaptability also holds promise for application in other landslide-prone regions globally, providing a framework for integrating remote sensing and neural networks in predictive hazard analysis.

Conclusions

This study presents a CNN-LSTM-based landslide prediction model that leverages multi-source satellite data to deliver high-accuracy predictions across India's diverse terrains. With an F1-score of 0.91, the model successfully integrates SAR, optical imagery, DEM, and rainfall data to capture the complex spatial and temporal dynamics associated with landslides. Key findings highlight the importance of SAR and NDVI data in detecting early warning signs like ground deformation and vegetation loss, while DEM-based slope analysis and GPM rainfall data confirm the topographical and environmental conditions that elevate landslide risk.

The implications of this model are significant for India's disaster management systems, offering a scalable solution for real-time landslide monitoring. By integrating predictive modeling into early-warning systems, this research can aid in timely disaster response, minimizing human and economic losses in landslide-prone areas. Future research should explore the integration of additional environmental variables, such as soil moisture and land use changes, to further enhance model accuracy. This study contributes a foundational framework for leveraging advanced neural network architectures and satellite data in landslide prediction, with potential applications for similar hazards worldwide.

Reference:

1. Chandra, N., & Vaidya, H. (2024). Deep learning approaches for landslide information recognition: current scenario and opportunities. *Journal of Earth System Science*, *133*(2), 1–25.
2. Chen, L., Peng, L., & Yang, L. (2023). Improving landslide prediction: innovative modeling and evaluation of landslide scenarios with knowledge graph embedding. *Remote Sensing*, *16*(1), 145.
3. Corominas, J., Westen, C., Frattini, P., Cascini, L., Malet, J.-P., Fotopoulou, S., et al. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, *73*, 209–263.
4. Crosta, G. B., Agliardi, F., Rivolta, C., Alberti, S., & Dei Cas, L. (2017). Long-term evolution and early warning strategies for complex rockslides by real-time monitoring. *Landslides*, *14*, 1615–1632.
5. Dai, F., Lee, C. F., & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering Geology*, *64*(1), 65–87.
6. Doerksen, K., Gal, Y., Kalaitzis, F., Rossi, C., Petit, D., Li, S., & Dadson, S. (2023). Precipitation-triggered landslide prediction in Nepal using machine learning and deep learning. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 4962–4965.
7. Du, H., Song, D., Chen, Z., Shu, H., & Guo, Z. (2020). Prediction model oriented for landslide displacement with step-like curve by applying ensemble empirical mode decomposition and the PSO-ELM method. *Journal of Cleaner Production*, *270*, 122248.
8. Gidon, J. S., Borah, J., Sahoo, S., Majumdar, S., & Fujita, M. (2023). Bi-directional LSTM model for accurate and real-time landslide detection: a case study in Mawiongrim, Meghalaya, India. *IEEE Internet of Things Journal*, *11*(3), 3792–3800.
9. Guerrero-Rodriguez, B., Garcia-Rodriguez, J., Salvador, J., Mejia-Escobar, C., Cadena, S., Cepeda, J., et al. (2024). Improving landslide prediction by computer vision and deep learning. *Integrated*

Computer-Aided Engineering, *31*(1), 77–94.
10. Jiang, H., Li, Y., Zhou, C., Hong, H., Glade, T., & Yin, K. (2020). Landslide displacement prediction combining LSTM and SVR algorithms: a case study of Shengjibao landslide from the Three Gorges Reservoir area. *Applied Sciences*, *10*(21), 7830.
11. Khalili, M. A., Guerriero, L., Pouralizadeh, M., Calcaterra, D., & Di Martire, D. (2023). Monitoring and prediction of landslide-related deformation based on the GCN-LSTM algorithm and SAR imagery. *Natural Hazards*, *119*(1), 39–68.
12. Krkač, M., Špoljarić, D., Bernat, S., & Arbanas, S. M. (2017). Method for prediction of landslide movements based on random forests. *Landslides*, *14*, 947–960.
13. Lacasse, S., & Nadim, F. (2009). Landslide risk assessment and mitigation strategy. In Sassa, K., Canuti, P. (Eds.), *Landslides—Disaster Risk Reduction*. Springer, Berlin, 31–61.
14. Lin, Z., Ji, Y., & Sun, X. (2023). Landslide displacement prediction based on CEEMDAN method and CNN-BiLSTM model. *Sustainability*, *15*(13), 10071.
15. Ma, Z., & Mei, G. (2021). Deep learning for geological hazards analysis: data, models, applications, and opportunities. *Earth Science Reviews*, *223*, 103858.
16. Munagala, V., Thudumu, S., Logothetis, I., Bhandari, S., Vasa, R., & Mouzakis, K. (2024). A comprehensive survey on machine learning applications for drilling and blasting in surface mining. *Machine Learning Applications*, *15*, 100517.
17. NDMA. (2011). National Disaster Management Authority (NDMA) Report, 2011. Retrieved from https://ndma.gov.in/sites/default/files/PDF/Reports/ANNUAL REPORT 2010-2011 ENGLISH.pdf
18. Prakasam, C., Aravinth, R., Kanwar, V. S., & Nagarajan, B. (2021). Design and development of real-time landslide early warning system through low-cost soil and rainfall sensors. *Materials Today: Proceedings*, *45*, 5649–5654.
19. Wang, Y., Tang, H., Huang, J., Wen, T., Ma, J., & Zhang, J. (2022). A comparative study of different machine learning methods for reservoir landslide displacement prediction. *Engineering Geology*, *298*, 106544.
20. Yang, B., Yin, K., Lacasse, S., & Liu, Z. (2019). Time series analysis and long short-term memory neural network to predict landslide displacement. *Landslides*, *16*, 677–694.
21. A Case Study of Heitai Terrace - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Type-A-landslides-location-shown-in-Figure-2-a-Crown-crack-groups-sliding-areas_fig6_374806450