

Hybrid Deep Learning Framework for Landslide Risk Prediction Using Meteorological and Geospatial Parameters

Mohsin Hassan ^{1,✉}, Ayushi Kotwal ², Jatinder Manhas ², Tasleem Arif ¹

1. Department of Information Technology, Baba Ghulam Shah Badshah University, Rajouri, IND

2. Department of Computer Science and IT, University of Jammu, Jammu, IND

Abstract

Landslide prediction plays an important role in disaster risk reduction and protecting infrastructure and human lives when performed accurately and in a timely manner. Despite good performance, machine learning techniques are hindered by severe class imbalance in landslide datasets and the opacity of complex predictive models. To address these challenges, this study proposes an integrated framework for landslide risk prediction through systematic comparative analysis. The training data were first balanced using the Synthetic Minority Over-sampling Technique with meteorological and geospatial parameters. We then used Logistic Regression as a baseline, Random Forest as a machine learning benchmark, and two deep learning architectures, namely 1D-CNN and LSTM. A performance-weighted hybrid ensemble was also considered to combine the predictive capabilities of the individual models. Under the evaluated model configurations, Random Forest and the optimized 1D-CNN emerged as the strongest models, achieving test-set macro F1-scores of 0.91 and 0.90, respectively, with comparable cross-validation performance. The optimized 1D-CNN was further analyzed using SHapley Additive exPlanations to provide class-specific explanations of model predictions and identify the features driving landslide risk assessment. The outcomes illustrate an interpretable and reproducible workflow that combines class balancing, comparative model assessment, and explainable artificial intelligence for landslide risk prediction. The framework, however, needs to be validated using independent, real-world landslide inventories before it can be considered for operational early-warning applications.

Categories: AI applications, Explainable AI, Machine Learning (ML)

Keywords: synthetic minority oversampling technique (smote), one-dimensional convolutional neural network (1d-cnn), long short-term memory (lstm), shapley additive explanations (shap), landslide early warning system

Introduction

Globally, landslides are a permanent and devastating form of geological hazard, causing substantial economic losses, infrastructure damage, and loss of human life. The occurrence of landslides is governed by a complex, non-linear interaction among topographic, lithological, and soil-mechanical factors and is frequently triggered by hydro-meteorological events such as intense rainfall and rapid snowmelt [1,2]. Changing climate patterns have further intensified the urgency of addressing this hazard. Increasing frequencies of extreme weather events, coupled with anthropogenic environmental degradation and rapid urbanization, have significantly increased landslide susceptibility in vulnerable regions [1]. Consequently, the development of reliable predictive frameworks has become a critical objective for disaster risk reduction, enabling timely evacuation, informed land-use planning, and targeted mitigation strategies.

Landslide Susceptibility Mapping (LSM) serves as the principal analytical tool for identifying areas prone to future slope failures based on historical inventories and conditioning factors [2]. Traditionally, LSM has relied on heuristic approaches and statistical models such as logistic regression, which assume relatively simple relationships between predictors and outcomes [3,4]. However, as geospatial datasets have increased in complexity, volume, and resolution, the limitations of these conventional approaches in capturing highly nonlinear interactions have become increasingly evident [5]. This

limitation has encouraged the widespread adoption of machine learning and deep learning techniques within geomorphological research [5,6]. These approaches possess a superior capability to discover complex patterns within large, high-dimensional datasets while automatically learning informative feature representations [6].

Despite these advances, several methodological challenges remain. The first challenge is severe class imbalance, a common characteristic of landslide datasets where stable observations substantially outnumber failure events. Furthermore, catastrophic landslides constitute only a very small fraction of the dataset. Models trained on such imbalanced distributions often become biased toward the majority class, resulting in poor detection of high-risk events [7]. Consequently, a model may achieve high overall accuracy while failing to identify the most critical landslide occurrences.

The second challenge is the limited interpretability of advanced machine learning and deep learning models. Although these models achieve high predictive performance, their “black-box” nature restricts their practical adoption in disaster management and public safety applications. Decision-makers require transparent explanations to justify mitigation measures, evacuation decisions, and resource allocation. Therefore, model interpretability has become an essential requirement for operational deployment [8].

Although numerous machine learning approaches have been proposed for landslide susceptibility assessment, several research gaps remain. Comparative studies often suffer from inconsistent hyperparameter optimization procedures, leading to potentially biased performance evaluations [2]. Additionally, while the Synthetic Minority Over-sampling Technique (SMOTE) has been widely adopted to address class imbalance, systematic studies investigating its impact across different landslide risk categories remain limited [7]. In ensemble learning, traditional soft-voting approaches rely on simple averaging, which may reduce overall effectiveness when constituent models exhibit unequal predictive capabilities [9]. Furthermore, most applications of Explainable Artificial Intelligence (XAI) in geohazard analysis focus primarily on global feature importance rankings, offering limited insights into class-specific prediction mechanisms [10].

Recent studies have demonstrated the effectiveness of deep learning architectures, particularly convolutional neural networks, for landslide susceptibility mapping and geospatial pattern extraction [11,12]. LSTM networks have also been applied successfully to temporal landslide prediction tasks [13]. Furthermore, ordinal classification techniques provide a natural framework for exploiting the ordered structure of landslide risk categories and may improve discrimination between adjacent classes [14]. Alternative machine learning approaches, such as the Group Method of Data Handling (GMDH), have also demonstrated promising results in susceptibility assessment [15]. Advances in one-dimensional convolutional neural network architectures further support their applicability to structured environmental datasets [16].

The framework systematically compares traditional statistical methods, ensemble machine learning models, and deep learning architectures using consistent preprocessing, model configuration, and five-fold stratified cross-validation. It further investigates the impact of SMOTE through detailed ablation studies, develops a performance-weighted hybrid ensemble model to overcome the limitations of conventional soft voting, and employs SHapley Additive exPlanations (SHAP) to provide transparent, class-specific interpretations of model predictions [8]. To address these challenges, this study proposes an integrated and reproducible framework for landslide risk prediction. The contribution of this study lies not in introducing a new individual learning algorithm, but in systematically integrating complementary techniques within a common evaluation framework. Specifically, the study: (i) compares statistical, ensemble machine learning, and deep learning models under identical preprocessing and stratified validation procedures, with systematic hyperparameter optimization of the 1D-CNN; (ii) evaluates model performance under severe class imbalance using both overall and class-specific metrics; (iii) investigates performance-weighted ensemble learning when constituent models exhibit unequal predictive capabilities; and (iv) applies class-specific SHAP analysis not only for model interpretation but also for identifying potentially implausible patterns introduced during synthetic oversampling. The dataset, implementation code, and supplementary materials are openly available to facilitate reproducibility and further investigation [17].

Materials And Methods

Dataset description

The present study utilizes a synthetic landslide prediction secondary dataset generated via multivariate statistical distributions calibrated to the meteorological and geospatial parameters of landslide-prone regions [17]. It comprises 5,000 instances with five key features: temperature (°C), humidity, precipitation (mm), soil moisture, and elevation (m). The target variable possesses four ordinal classes: low, moderate, high, and very high risk. As detailed in Table 1, the dataset exhibits a severe class imbalance profile that mirrors empirical geohazard recording anomalies.

Risk Class	Count	Percentage (%)	Imbalance Ratio
Low	4,591	91.82	1.00 (majority)
Moderate	334	6.68	13.74:1
High	63	1.26	72.87:1
Very High	12	0.24	382.58:1
Total	5,000	100.00	—

TABLE 1: Class Distribution of Landslide Risk Dataset

Preprocessing and balancing

A multi-step preprocessing pipeline was implemented to prepare the data for modeling, with particular attention to preventing data leakage and maintaining evaluation integrity.

Feature Scaling and Target Encoding

To ensure that all features contribute equally to the model training process and to facilitate the convergence of gradient-based optimization algorithms, the numerical features were standardized using the StandardScaler implementation available in Scikit-learn [18]. This transformation scales the features to have a mean of '0' and a standard deviation of '1':

z_scaled = (z - μ) / σ (1)

where:

z is the original feature value for a given instance, μ is the mean of that feature computed exclusively from the training set, σ is the standard deviation of that feature computed exclusively from the training set, and z_scaled is the standardized feature value.

The categorical target variable was converted into a numerical format using LabelEncoder: Low = 0, Moderate =1, High = 2, Very High = 3. For the deep learning models, which utilize a categorical cross-entropy loss function, these integer labels were further transformed into a one-hot encoded format using the 'to_categorical' utility from Keras.

Stratified Data Splitting

The dataset was partitioned into a training set, i.e., 80% of the data, or 4,000 instances, and a testing set, i.e., 20% of the data, or 1,000 instances, using stratified random sampling. A stratified splitting strategy was employed to ensure that the proportional representation of each risk class in the original dataset was maintained in both the training and testing subsets. This is crucial for imbalanced datasets, as it guarantees that the models are evaluated on a test set that reflects the real-world class distribution.

For cross-validation, we employed 5-fold stratified cross-validation, where each fold maintains class proportions and every instance appears in the validation set exactly once across all folds.

Imbalance Handling with SMOTE

To address the class imbalance, the SMOTE was applied exclusively to the training data after stratified splitting to prevent data leakage. SMOTE works by identifying minority class instances and generating synthetic samples between neighboring minority observations [19].

$$\mathbf{x}_{synthetic} = \mathbf{x}_i + \delta \left(\mathbf{x}_{neighbour(i)} - \mathbf{x}_i \right) \quad (2)$$

where:

$\mathbf{x}_i$ is a randomly selected minority class instance (feature vector), $\mathbf{x}_{neighbour(i)}$ is one of the k -nearest neighbours of $\mathbf{x}_i$ chosen randomly from the same minority class, k is the number of nearest neighbours considered (set to $k=5$ in our implementation), δ is a random interpolation weight uniformly drawn from the interval $[0, 1]$, which determines the position of the new synthetic point along the line segment between $\mathbf{x}_i$ and $\mathbf{x}_{neighbour(i)}$, and $\mathbf{x}_{synthetic}$ is the newly generated synthetic feature vector. The effect of SMOTE is illustrated in Figure 1.

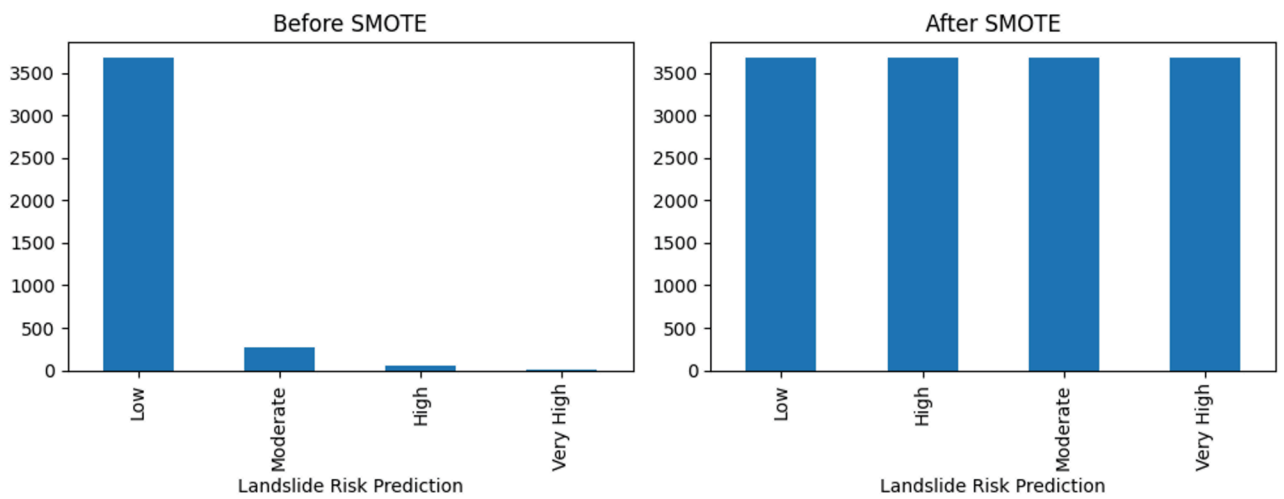


FIGURE 1: Dataset Before and After SMOTE

Results And Discussion

Experimental setup

All experiments were conducted on a PC equipped with 16 GB RAM, 512 GB SSD storage, and an NVIDIA GeForce RTX 2050 GPU. The software environment comprised Python 3.12, TensorFlow, Scikit-learn, and SHAP. Model configurations and optimization procedures are summarized in Table 5. The 1D-CNN hyperparameters were systematically optimized using Hyperband, while Logistic Regression, Random Forest, and LSTM were evaluated using the specified

configurations. Five-fold stratified cross-validation was used for model evaluation. Early stopping with a patience of 10 epochs was applied to the deep learning models. Evaluation metrics were computed on the held-out test set comprising 1,000 instances. The source code, Jupyter notebook, and supplementary materials are available on Zenodo [20].

Comparative model performance

Model	Accuracy (%)	Macro F1	Cohen's κ	5-Fold CV Macro F1, Mean ± SD	95% CI*
Logistic Regression	93.2	0.63	0.44	0.61 ± 0.05	0.55–0.67
Random Forest	99.2	0.91	0.87	0.91 ± 0.03	0.87–0.95
1D-CNN (Optimized)	98.7	0.9	0.85	0.89 ± 0.02	0.87–0.91
LSTM	96.3	0.59	0.46	0.57 ± 0.05	0.51–0.63
Naive Ensemble	97.1	0.6	0.48	0.58 ± 0.04	0.53–0.63
Weighted Ensemble	98.6	0.89	0.83	0.88 ± 0.03	0.84–0.92

TABLE 2: Overall Model Performance on the Held-Out Test Set and Cross-Validation

*95% confidence intervals estimated from the five-fold cross-validation mean and standard deviation using the t-distribution (df = 4). LSTM, Long Short-Term Memory

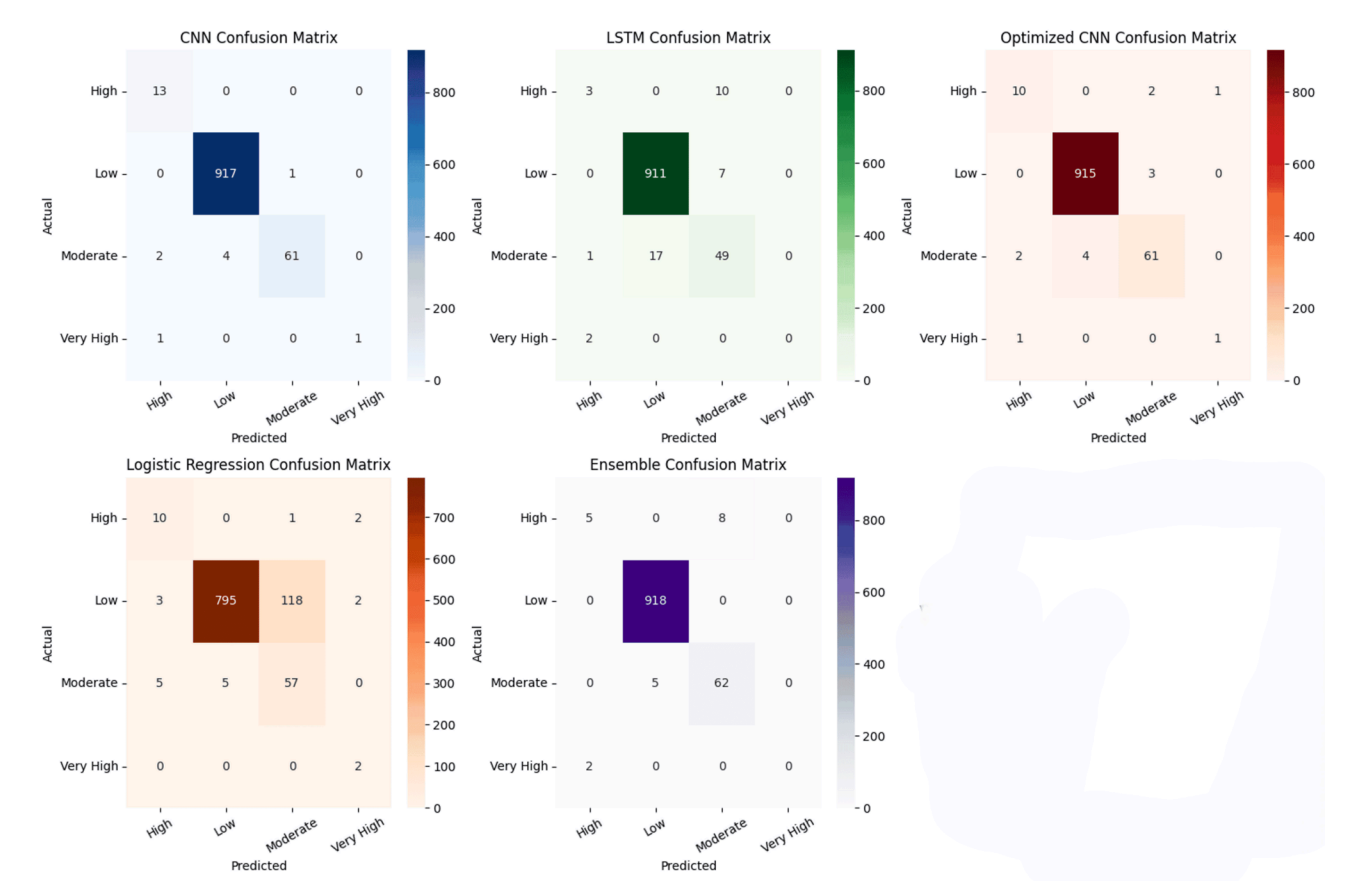


FIGURE 2: Confusion Matrices

The overall performance of all models on the test set is summarized in Table 2, while confusion matrices for selected models are presented in Figure 2. Random Forest (macro F1 = 0.91) and optimized 1D-CNN (macro F1 = 0.90) are the top-performing models. The weighted ensemble (macro F1 = 0.89) performs comparably to 1D-CNN but does not exceed Random Forest, indicating that the inclusion of weaker-performing models (LSTM and Logistic Regression) may limit ensemble performance despite reduced weighting.

A key observation is the performance of LSTM. Despite achieving 96.3% overall accuracy, LSTM detected only 4 of 13 High-risk instances (30.8% recall) and none of the two Very High-risk instances. This poor performance on the minority risk classes illustrates why overall accuracy alone is insufficient for evaluating models under severe class imbalance. The training and validation performance of the evaluated models is shown in Figure 3.

Per-class performance

Model	Class	Precision	Recall	F1-Score
Random Forest	Low	0.99	1	0.99
	Moderate	0.92	0.94	0.93
	High	0.85	0.85	0.85
	Very High	1	0.5	0.67
1D-CNN (Optimized)	Low	0.99	1	0.99
	Moderate	0.89	0.91	0.9
	High	0.85	0.85	0.85
	Very High	0.5	0.5	0.5
LSTM	Low	0.98	0.99	0.99
	Moderate	0.78	0.78	0.78
	High	0.29	0.31	0.3
	Very High	0	0	0

TABLE 3: Per-Class Metrics for Top Models (Test Set)

LSTM, Long Short-Term Memory

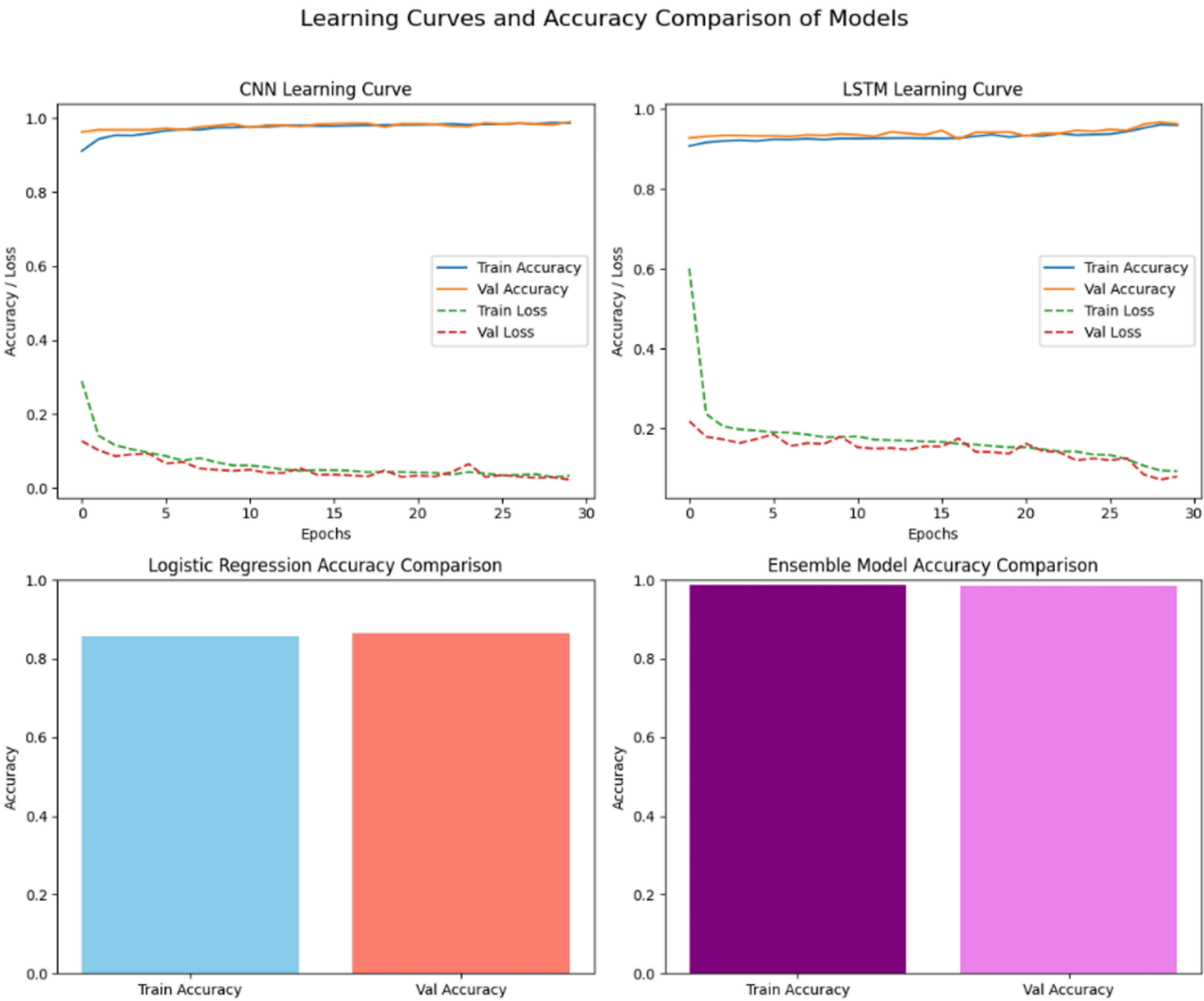


FIGURE 3: Training and Validation Performance of the Evaluated Models

Table 3 details per-class metrics for the top models (RF and 1D-CNN) and the LSTM for comparison. All models achieve near-perfect F1 (>0.98) for the low-risk class. RF and 1D-CNN maintain acceptable high-risk recall (85%). However, all models struggle with the very high-risk class (2 test instances) due to extreme rarity. In future operational evaluations, decision thresholds should prioritize recall for high- and very high-risk classes while considering the associated increase in false-alarm rates. Given that the Very High-risk class contains only two observations in the held-out test set, its precision, recall, and F1-score are highly sensitive to individual predictions and should therefore be interpreted as descriptive rather than statistically stable performance estimates.

SHAP explainability analysis

Global SHAP feature importance across 1,000 test instances ranked the predictors as follows: Elevation (mean $|\text{SHAP}| = 0.42$) $>$ Soil Moisture (0.38) $>$ Precipitation (0.35) $>$ Temperature (0.19) $>$ Humidity (0.16). This ordering is physically plausible, as elevation is associated with terrain characteristics relevant to slope instability, while soil moisture and precipitation influence subsurface water conditions. Figure 4 presents the global SHAP feature importance of the optimized 1D-CNN model based on mean absolute SHAP values.

Class-specific SHAP analysis further indicated that feature contributions varied across risk classes. Low-risk predictions were associated with lower elevation, soil moisture, and precipitation values, whereas Very High-risk predictions showed stronger contributions from elevated terrain, high soil moisture, and intense precipitation.

A counter-intuitive pattern was observed for humidity, with high humidity producing negative SHAP contributions for some Very High-risk predictions. Examination of the correlation structure indicated that SMOTE weakened the humidity-precipitation correlation in synthetic Very High-risk instances ($r = 0.62$ before oversampling and $r = 0.23$ after oversampling), suggesting that interpolation may have introduced less physically plausible feature combinations. This finding highlights SHAP's potential value as a model-diagnostic tool in addition to its role in model interpretation.

As a feature-level sensitivity analysis, feature ablation was performed and was consistent with the global SHAP ranking: removing elevation produced the largest single-feature reduction in performance ($\Delta F1 = -0.14$), while subsequent feature removal broadly followed the SHAP importance order of elevation > soil moisture > precipitation.

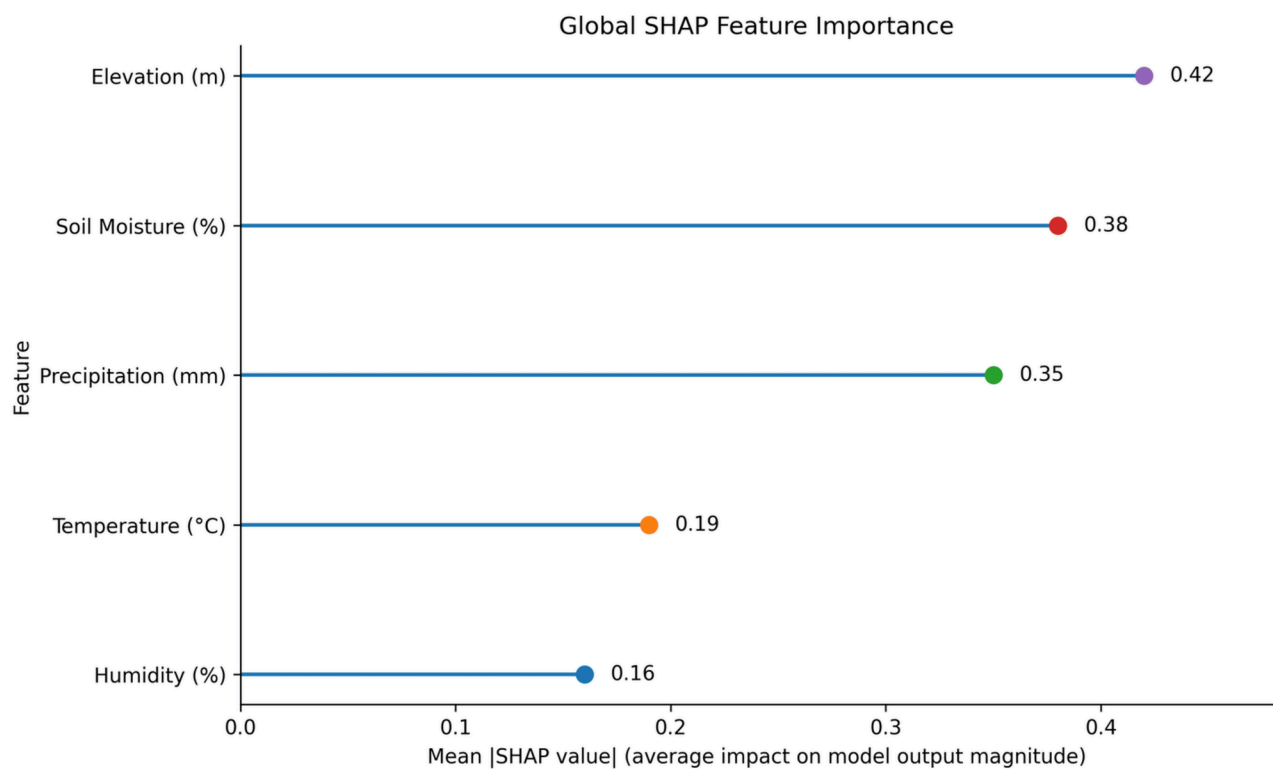


FIGURE 4: Global SHAP Feature Importance for the Optimized 1D-CNN Model

Model configuration and optimization impact

Model	Baseline F1	Final F1	$\Delta F1$	Final Configuration
Logistic Regression	0.57	0.63	+0.06	C = 0.1, solver = saga
Random Forest	0.88	0.91	+0.03	n_estimators = 200, max_depth = 30
1D-CNN	0.86	0.90	+0.04	filters = 64/128, dropout = 0.2/0.3
LSTM	0.51	0.59	+0.08	units = 64/32, dropout = 0.2/0.3

TABLE 4: Performance Impact of Model Configuration and Optimization

LSTM, Long Short-Term Memory

The performance differences associated with the selected model configurations and 1D-CNN hyperparameter optimization are summarized in Table 4. The 1D-CNN improved from a macro F1-score of 0.86 to 0.90 following Hyperband optimization, while the corresponding configurations and search space are reported in Table 5.

Model	Hyperparameter / Configuration	Values Evaluated / Used
Logistic Regression	Configuration	C = 0.1; solver = saga
Random Forest	Configuration	n_estimators = 200; max_depth = 30
1D-CNN	Filters	{32, 64, 96, 128}
1D-CNN	Dropout rate	{0.1, 0.2, 0.3, 0.4, 0.5}
1D-CNN	Learning rate	{0.01, 0.001, 0.0001}
LSTM	Architecture	Units = 64/32; dropout = 0.2/0.3

TABLE 5: Model Configuration and Hyperparameter Search Space

LSTM, Long Short-Term Memory

Model configurations and optimization procedures are summarized in Table 5. The 1D-CNN hyperparameters were systematically optimized using Hyperband, whereas Logistic Regression, Random Forest, and LSTM were evaluated using the specified configurations. The 1D-CNN hyperparameter search space and the configurations used for the remaining

models are reported in Table 5.

Recent research has applied a range of machine learning and deep learning approaches to landslide prediction and susceptibility assessment, with different study areas, input factors, and reported performance levels. A summary of these recent studies alongside the results of the present study is presented in Table 6. The present study achieved accuracies of 99.20%, 98.70%, and 96.30% using Random Forest, optimized 1D-CNN, and LSTM, respectively, reporting higher accuracy values than those reported by the recent studies summarized in Table 6, which range from 75.86% to 95.76%. Specifically, the proposed Random Forest and optimized 1D-CNN exceed the CNN-LSTM of Teng et al. [21] (95.60%), the best-performing GraphSAGE-CatBoost-RSA of Bahrami et al. [22] (93.20%), the image-wise CNN of Oh et al. [23] (75.86%), the Meta Classifier of Halder et al. [24] (95.60%), the CNN of Feng et al. [25] (95.76%), and the optimized UCNN1D of Mejia-Manrique et al. [26] (84.80%) in terms of accuracy. It should be noted, however, that the studies listed in Table 6 report accuracy as their primary metric and do not report macro F1-scores; a direct metric-aligned comparison is therefore not feasible, and the present study's principal evaluation criterion of macro F1-score cannot be benchmarked against these external results. Furthermore, direct performance comparisons across studies must be interpreted with caution, as the referenced works differ substantially in study area, dataset size, spatial resolution, number and type of conditioning factors, and validation protocol.

Dataset / Study Area	Model	Accuracy	Main Influencing Factors	Reference
Kerman Province, Iran	CNN-LSTM	95.60%	Slope, elevation, vegetation indices, rainfall, cumulative precipitation, reservoir-level fluctuations	Teng et al., 2026 [21]
Karaj-Chalus Road, Iran	GraphSAGE-CatBoost-RSA	93.20%	Lithology, distance to faults, roads and rivers, LULC, rainfall, slope, aspect, TWI, SPI	Bahrami et al., 2026 [22]
Karaj-Chalus Road, Iran	GoogleNet-CNN-HHO	88.70%	Lithology, distance to faults, roads and rivers, LULC, rainfall, slope, aspect, TWI, SPI	Bahrami et al., 2026 [22]
Karaj-Chalus Road, Iran	Autoencoder-XGBoost	84.10%	Lithology, distance to faults, roads and rivers, LULC, rainfall, slope, aspect, TWI, SPI	Bahrami et al., 2026 [22]
Jinbu-myeon, South Korea	CNN (image-wise)	75.86%	DEM derivatives, forest cover, soil properties, geological variables (20 factors)	Oh et al., 2026 [23]
Sub-Himalayan West Bengal, India	Meta Classifier	95.60%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]
Sub-Himalayan West Bengal, India	Extra Trees	94.60%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]
Sub-Himalayan West Bengal, India	Gradient Boosting	94.30%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]
Sub-Himalayan West Bengal, India	XGBoost	94.30%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]

Sub-Himalayan West Bengal, India	Random Forest	94.30%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]
Sub-Himalayan West Bengal, India	SVM	94.10%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]
Sub-Himalayan West Bengal, India	Logistic Regression	86.00%	Rainfall, geological conditions, elevation, slope, drainage, LULC, anthropogenic factors	Halder et al., 2025 [24]
Bakhtegan Watershed, Iran	CNN	95.76%	Elevation, slope, aspect, curvature, lithology, drainage, distance to river, weathering, vegetation, LULC, distance to faults, earthquakes, rainfall, temperature, distance to cities and roads	Feng et al., 2025 [25]
Puerto Rico / Hurricane Maria	Optimized UCNN1D	84.80%	Slope, rainfall accumulation, soil moisture, soil saturation, distance to roads, streams and faults, geology, LULC	Mejia-Manrique et al., 2025 [26]
Puerto Rico / Hurricane Maria	Random Forest	84.20%	Slope, rainfall accumulation, soil moisture, soil saturation, distance to roads, streams and faults, geology, LULC	Mejia-Manrique et al., 2025 [26]
Landslide Prediction Dataset	Random Forest	99.20%	Temperature, humidity, precipitation, soil moisture, elevation	Present study
Landslide Prediction Dataset	Optimized 1D-CNN	98.70%	Temperature, humidity, precipitation, soil moisture, elevation	Present study
Landslide Prediction Dataset	LSTM	96.30%	Temperature, humidity, precipitation, soil moisture, elevation	Present study

TABLE 6: Comparison with Recent Landslide Prediction and Susceptibility Assessment Studies (2025–2026)

LSTM, Long Short-Term Memory

Operational deployment considerations

In a potential operational setting, the trained framework could receive meteorological and geospatial observations from monitoring stations, remote-sensing products, or IoT-based sensors and generate class-specific landslide risk estimates for decision support. Once trained, RF and 1D-CNN models have relatively modest inference requirements, making near-real-time prediction computationally feasible. However, practical deployment would require reliable sensor networks, continuous data-quality control, location-specific calibration, appropriate warning thresholds, and validation using local landslide inventories. Therefore, the present framework should be considered a methodological prototype rather than a deployment-ready landslide early warning system.

Limitations

The principal limitation of this study is its reliance on a synthetic secondary dataset. Although the dataset enables controlled evaluation under severe class imbalance and provides meteorological and geospatial variables relevant to landslide susceptibility, it cannot fully reproduce the spatial heterogeneity, measurement uncertainty, geological complexity, temporal dependence, and site-specific conditions encountered in real-world landslide inventories. In particular, the Very High-risk class contains only 12 observations in the original dataset and two observations in the held-out test set. Consequently, performance estimates for this class should be interpreted as preliminary and descriptive rather than as statistically stable estimates of operational performance. Validation using independent real-world landslide inventories across different geographic regions is therefore required before conclusions regarding generalizability or operational deployment can be established. Future work should also include broader sensitivity analyses across alternative preprocessing and resampling configurations, comparison with additional state-of-the-art machine learning and deep learning approaches, and external validation using geographically diverse real-world datasets.

Conclusions

This study developed and evaluated an integrated framework for landslide risk prediction that combines data balancing, comparative assessment of machine learning and deep learning models, hyperparameter optimization, stratified cross-validation, and explainable artificial intelligence. The results demonstrate that careful treatment of class imbalance is important for improving minority-class detection; however, the effectiveness of alternative SMOTE sampling ratios requires further systematic evaluation. Under the evaluated configurations, Random Forest and 1D-CNN delivered the strongest predictive performance, clearly outperforming Logistic Regression and LSTM in terms of macro F1-score. Although LSTM attained high overall accuracy, its weak recall on critical high-risk events exposed the limitations of relying on accuracy alone in safety-critical contexts and underscored the importance of class-sensitive metrics such as macro F1-score and recall for geohazard prediction. The explainability analysis further explained model behavior, with SHAP interpretation identifying elevation, soil moisture, and precipitation as dominant contributors to landslide risk, consistent with established geomorphological understanding. This analysis also revealed artifacts introduced during synthetic data generation, demonstrating that explainable AI serves not only as a transparency mechanism but also as a diagnostic tool for validating machine learning workflows. Overall, the proposed framework offers an interpretable and reproducible approach that links data balancing, optimized modeling, performance evaluation, and explainability to support more trustworthy decision-support systems for disaster risk reduction and landslide early warning. Accurately predicting the rarest, very high-risk events remains a challenge, yet the framework provides a foundation for future validation using independent, real-world landslide inventories and geographically diverse datasets.

Additional Information

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with [the ICMJE uniform disclosure form](#), all authors declare the following:

- **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work.
- **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work.
- **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The dataset used in this study is publicly available on Kaggle (Reference 17). The source code, Jupyter notebook, and supplementary materials are available on Zenodo (Reference 20): <https://doi.org/10.5281/zenodo.21307099>.

References

1. Gariano SL, Guzzetti F: [Landslides in a changing climate](#). Earth-Science Reviews. 2016, 162:227-252. [10.1016/j.earscirev.2016.08.011](#)
2. Reichenbach P, Rossi M, Malamud BD, Mihir M, Guzzetti F: [A review of statistically-based landslide susceptibility models](#). Earth-Science Reviews. 2018, 180:60-91. [10.1016/j.earscirev.2018.03.001](#)
3. Pradhan B, Lee S: [Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling](#). Environmental Modelling & Software. 2010, 25:747-759. [10.1016/j.envsoft.2009.10.016](#)
4. Kanungo DP, Arora MK, Sarkar S, Gupta RP: [A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas](#). Engineering Geology. 2006, 85:347-366. [10.1016/j.enggeo.2006.03.004](#)
5. Bui DT, Tsangaratos P, Nguyen VT, Liem NV, Trinh PT: [Comparing the prediction performance of a Deep Learning Neural Network model with conventional machine learning models in landslide susceptibility assessment](#). CATENA. 2020, 188:104426. [10.1016/j.catena.2019.104426](#)
6. Zhang L, Zhang L, Du B: [Deep learning for remote sensing data: A technical tutorial on the state of the art](#). IEEE Geoscience and Remote Sensing Magazine. 2016, 4:22-40. [10.1109/mgrs.2016.2540798](#)
7. Vanto, Mahmud IS, Sugiyanto G: [The effect of spatial scales and imbalanced data treatment on landslide susceptibility mapping using Random Forest](#). AIP Conference Proceedings. 2023, 2482:050006. [10.1063/5.0111326](#)
8. Adadi A, Berrada M: [Peeking inside the black-box: A survey on explainable artificial intelligence \(XAI\)](#). IEEE Access. 2018, 6:52138-52160. [10.1109/access.2018.2870052](#)
9. Chen W, Zhang S, Li R, Shahabi H: [Performance evaluation of the GIS-based data mining techniques of best-first decision tree, random forest, and naïve Bayes tree for landslide susceptibility modeling](#). Science of the Total Environment. 2018, 644:1006-1018. [10.1016/j.scitotenv.2018.06.389](#)
10. Panahi M, Rahmati O, Rezaie F, Lee S, Mohammadi F, Conoscenti C: [Application of the group method of data handling \(GMDH\) approach for landslide susceptibility zonation using readily available spatial covariates](#). CATENA. 2022, 208:105779.

[10.1016/j.catena.2021.105779](https://doi.org/10.1016/j.catena.2021.105779)

11. Frank E, Hall M: [A simple approach to ordinal classification](#). Machine Learning: ECML 2001. Lecture Notes in Computer Science. De Raedt L, Flach P (ed): Springer, Berlin, Heidelberg; 2001. 2167:145-156. [10.1007/3-540-44795-4_13](https://doi.org/10.1007/3-540-44795-4_13)
12. Kiranyaz S, Avci O, Abdeljaber O, Ince T, Gabbouj M, Inman DJ: [1D convolutional neural networks and applications: A survey](#). Mechanical Systems and Signal Processing. 2021, 151:107398. [10.1016/j.ymssp.2020.107398](https://doi.org/10.1016/j.ymssp.2020.107398)
13. Segoni S, Pappafico G, Luti T, Catani F: [Landslide susceptibility assessment in complex geological settings: sensitivity to geological information and insights on its parameterization](#). Landslides. 2020, 17:2443-2453. [10.1007/s10346-019-01340-2](https://doi.org/10.1007/s10346-019-01340-2)
14. Wang Y, Fang Z, Hong H: [Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China](#). Science of The Total Environment. 2019, 666:975-993. [10.1016/j.scitotenv.2019.02.263](https://doi.org/10.1016/j.scitotenv.2019.02.263)
15. Lin Z, Sun X, Ji Y: [Landslide displacement prediction model using time series analysis method and modified LSTM model](#). Electronics. 2022, 11:1519. [10.3390/electronics11101519](https://doi.org/10.3390/electronics11101519)
16. Youssef AM, Pradhan B, Dikshit A, Al-Katheri MM, Matar SS, Mahdi AM: [Landslide susceptibility mapping using CNN-1D and 2D deep learning algorithms: comparison of their performance at Asir Region, KSA](#). Bulletin of Engineering Geology and the Environment. 2022, 81:165. [10.1007/s10064-022-02657-4](https://doi.org/10.1007/s10064-022-02657-4)
17. Nandha SR: [Landslide Prediction dataset](#). Kaggle. 2023,
18. Pedregosa F, Varoquaux G, Gramfort A, et al.: [Scikit-learn: Machine Learning in Python](#). Journal of Machine Learning Research. 2011, 12:2825-2830.
19. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP: [SMOTE: Synthetic Minority Over-sampling Technique](#). Journal of Artificial Intelligence Research. 2002, 16:321-357. [10.1613/jair.953](https://doi.org/10.1613/jair.953)
20. Hassan M: [Landslide Risk Prediction - Code and Supplementary Materials \(Jupyter Notebook\)](#). Zenodo. 2025, [10.5281/zenodo.21307099](https://doi.org/10.5281/zenodo.21307099)
21. Teng F, Ekraminia SS, Zarei A, Li Y: [Remote sensing-based landslide prediction and risk assessment using a hybrid CNN-LSTM deep learning model](#). Scientific Reports. 2026, 16:10687. [10.1038/s41598-026-43927-5](https://doi.org/10.1038/s41598-026-43927-5)
22. Bahrami Y, Maghsoudi A, Pour AB, Shirali M: [Advanced graph-based and deep learning frameworks for landslide susceptibility mapping in mountain transportation corridors](#). Scientific Reports. 2026, 1-29. [10.1038/s41598-026-60285-4](https://doi.org/10.1038/s41598-026-60285-4)
23. Oh J, Lee JH, Park HJ, Yoon D: [Principled XAI analysis of the deep learning-based landslide susceptibility prediction model](#). Scientific Reports. 2026, 16:27121. [10.1038/s41598-026-52786-z](https://doi.org/10.1038/s41598-026-52786-z)
24. Halder K, Srivastava AK, Ghosh A, et al.: [Improving landslide susceptibility prediction through ensemble recursive feature elimination and meta-learning framework](#). Scientific Reports. 2025, 15:5170. [10.1038/s41598-025-87587-3](https://doi.org/10.1038/s41598-025-87587-3)
25. Feng L, Zhang M, Mao Y, Liu H, Yang C, Dong Y, Nanekaran YA: [Convolutional neural network-based deep learning for landslide susceptibility mapping in the Bakhtegan watershed](#). Scientific Reports. 2025, 15:13250. [10.1038/s41598-025-96748-3](https://doi.org/10.1038/s41598-025-96748-3)
26. Mejia-Manrique SA, Ramos-Scharrón CE, Hughes KS, Gonzalez-Cruz JE, Khanbilvardi RM: [Dynamic landslide susceptibility for extreme rainfall events using an optimized convolutional neural network approach](#). Natural Hazards. 2025, 121:15383-15411. [10.1007/s11069-025-07396-9](https://doi.org/10.1007/s11069-025-07396-9)