RESEARCH ARTICLE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
www.rspsciencehub.com
Vol. 08, Issue 06 June

Check for updates

http://dx.doi.org/10.47392/IRJASH.2026.027

A Multi-Modal Parallel Deep Learning Framework Integrating Attention U-Net and Bidirectional LSTM for Spatio-Temporal Landslide Prediction

Farhan V A[1], Drupath Prasad[2], Alia Teresa T M[3]
[1] PG – Computer Science, Sacred Heart College, Kochi, Kerala
[2] PG – Computer Science, Sacred Heart College, Kochi, Kerala
[3] Assistant Professor, Computer Science, Sacred Heart College, Kochi, Kerala
Emails: farhanva91@gmail.com[1], drupathsprasad@gmail.com[2], aliateresa@shcollege.ac.in[3]

Article history
Received: 26 May 2026
Accepted: 05 June 2026
Published: 14 July 2026

Keywords:
Landslide Prediction, Multi-Modal Deep Learning, Attention U-Net, Bidirectional LSTM, Spatio-Temporal Modeling, Western Ghats. .

Abstract
Landslides are among the most significant natural hazards in mountainous regions, particularly under intense rainfall conditions. Accurate prediction requires the integration of both static terrain characteristics and dynamic hydrometeorological factors. However, many existing approaches rely primarily on spatial variables or aggregated rainfall indicators, limiting their ability to represent spatio-temporal relationships associated with slope failure. This study proposes a dual-stream multi-modal deep learning framework for spatio-temporal land-slide prediction in the Kerala Western Ghats, India. The framework combines an Attention U-Net branch for spatial feature extraction and a Bidirectional Long Short-Term Memory (BiLSTM) branch for temporal sequence modeling. The spatial input consists of a seven-channel tensor derived from Sentinel-2 spectral bands, the Normalized Difference Vegetation Index (NDVI), elevation, and slope layers, while the temporal input comprises a 30-day antecedent rainfall sequence obtained from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) dataset. Feature representations extracted from the spatial and temporal branches are integrated through a late-fusion mechanism and subsequently used for binary classification. Model performance was evaluated using an independent geographic holdout dataset. The proposed framework achieved an accuracy of 94.79%, precision of 89.74%, recall of 97.22%, F1-score of 93.33%, and an AUC-ROC value of 0.9757. Comparative experiments with Random Forest, ResNet-18, and unidirectional LSTM models demonstrated improved predictive performance of the proposed architecture. The results indicate that the integration of spatial and temporal information can enhance landslide prediction in data-driven frameworks for monsoon-dominated mountainous environments.

1. Introduction

Globally, landslides are one of the most dangerous types of geomorphological hazards that cause extensive infrastructure damage, significant economic loss, and many deaths in mountainous areas. These slope failures are a result of complex interactions between static pre-existing factors and dynamic environmental factors. One of the most prominent examples of a high-risk area for landslides in the world is tropical or subtropical regions (e.g., the Western Ghats mountain range in Kerala, India). This area is characterized by steep, weathered denudational slopes and fractured lithology. Under systemic conditions, these characteristics result in highly unstable slopes. The primary causal mechanism of these failures is heavy seasonal monsoon rains. Severe rainfall will eventually cause rainwater to accumulate within the porous surface soil layer and gradually infiltrate deeper into the soil. As the soil becomes saturated due to the accumulation of water, the excess water cannot drain quickly enough into the underlying bedrock; therefore, excess porewater pressure will create a rise in pressure along the soil-bedrock contact at the slope. A rapid increase in porewater pressure at the soil-bedrock interface will effectively negate the internal shear strength and cohesive friction of the slope. This creates rapid and catastrophic mass-movement of materials down the slope.

1.1. Problem Statement

Even though there is a strong need for precise systems that can predict landslides, developing reliable Landslide Susceptibility Mapping (LSM) frameworks is still a tough job because of certain limitations of data modelling. Conventional hazard mapping techniques usually assess data using a limited number of perspectives or criteria. For example, either: (1) static geo-morphological data (e.g., slope angle, elevation, land cover) OR; (2) dynamic meteorological data (i.e., rainfall depth). Conventional machine-learning techniques (e.g., Random Forest & Support Vector Machine) or standard deep-learning (DL) networks that take these static and dynamic data types treat them as “flat” or “simple,” thus producing “unstructured” data or simple vectors from the different types of data. The process of “flattening” a multispectral spatial image patch eliminates important “contextual” relationships between the local structural or geometric connections produced by that patch of terrain. Similarly, when “sequential rain events” are treated as a single “rain event,” the accumulated amount of water stored in the soil over time is not considered. These two examples illustrate why single-modality frameworks have difficulty balancing these two types of data, which can result in high false positive rates in relatively flat, high rainfall coastal zones and failing to capture all landslides in extreme high-risk locations high in the mountains.

1.2. Motivation

The pressing need to create an integrated approach to deep learning that can analyse both space (through multi-spectral geo-morphological images) and time (by analysing long-term real-time changes in rainfall data) simultaneously in order to simulate how landslides happen is the driving force behind this research. Landslides do not happen in isolation from one another. A heavy rainfall event does not cause a landslide to occur on flat terrain, while a steep mountain slope might remain stable during a dry season. To correctly represent how spatially vulnerable areas and temporally triggering events interact, an automated network must analyse multi-spectral geo-morphological images and real-time precipitation data concurrently. Because the majority of the background noise present in large mountain landscapes (i.e., stable ridges and open plains) is significant, an attention-driven filtering mechanism will allow this network to focus its computational resources on only those active hazard zones so that the network provides sufficient level of predictiveness to allow for reliable operation in real-world applications.

1.3. Research Objectives

To resolve these open methodological and data fusion challenges, this study establishes the following core research objectives:

High Resolution Spatio-Temporal Pipeline:- To develop a multi-modal geospatial data pipeline to efficiently associate 7-channel spatial imagery patches (7 × 128 ×128) from Sentinel-2 and SRTM DEM data with continuous 30-day temporal antecedent rainfall sequences (30 × 1) using CHIRPS data.

- Attention-Driven Spatial Encoding:- To develop and implement a parallel deep learning framework with an advanced Attention U-Net spatial encoder integrated with multi-scale Additive Attention Gates (AGs) to effectively suppress background terrain noise and extract localized geo-morphological vulnerabilities.
- Bidirectional Hydrological Sequence Tracking:- To build a Bidirectional Long Short-Term Memory (BiLSTM) temporal network to learn the forward and backward

dependencies between antecedent rainfall sequences and to capture the critical pore-water pressure variations.

- Late Feature Fusion Paradigm:- Late stage feature concatenation and fusion layer to merge these heterogeneous spatial and temporal feature streams into a common feature tensor for accurate, binary landslide susceptibility classification (PLSM).
- Empirical Performance Benchmarking:- To rigorously evaluate the framework for prediction accuracy, precision and sensitivity against single modality baseline models and traditional machine learning classifiers on an independent, field-validated holdout dataset from the Kerala Western Ghats.

1.4. Contributions

The primary scientific contributions of this research are summarized as follows:

- Novel Parallel Multi-Modal Architecture:- We propose a novel dual-stream deep learning architecture that simultaneously extracts spatial terrain features and temporal rainfall sequences, thus not losing structural information as in traditional flattening approaches.
- Multi-Scale Feature Selection via AGs:- The network is trained to automatically identify high-risk slope features in the spatial encoder pipeline by directly embedding specialized Additive Attention Gates (AG 1 and AG 2) that filter out non-vulnerable topographic background noise.
- Bidirectional Hydrological Modeling:- We successfully model the cumulative impact of antecedent rainfall on soil saturation levels and slope destabilization thresholds using a BiLSTM network over a sliding 30- day temporal lookback window.
- High-Sensitivity Operational Design:- The proposed framework achieves a critical Recall rate of 97.22% and an overall Accuracy of 94.79% on real field data, significantly reducing dangerous false negatives to provide a highly dependable tool for real-world disaster risk management and early warning networks.

1.5. Paper Organization

For the sake of structural clarity, the remainder of this manuscript is organized into separate, sequential sections. Section 2 provides a comprehensive review of the related work in the area of landslide susceptibility mapping and multi-modal deep learning architectures. Section 3 comprises the geographical profile, geological features and the field inventory database of the area of study. Section 4 is devoted to the data acquisition and pre-processing pipelines for the spatial imagery layers and temporal CHIRPS rainfall tracks. In Section 5, the mathematical formulation and structural description of the proposed Attention U-Net and BiLSTM hybrid network are presented. Section 6 describes the experimental setup, the training environment and the hyperparameter tuning protocols. Section 7 describes the quantitative evaluation metrics, comparison to baseline analysis, ablation studies and extensive discussion of the visualization results. Finally, Section 8 summarizes the main conclusions of this study and suggests directions of future work.

2. Literature Review

Landslide Susceptibility Mapping (LSM) has undergone considerable methodological evolution, transitioning from rule-based empirical approaches toward sophisticated data-driven frameworks capable of processing high-dimensional geospatial information. The inherent complexity of landslide processes, driven by the interplay of environmental, geological, and hydrometeorological variables, has historically directed researchers toward two complementary analytical perspectives. The first emphasizes static land surface characterization — examining spectral reflectance, terrain morphology, and spatial vulnerability indicators derived from satellite and topographic data. The second focuses on temporal dynamics, modeling the probabilistic conditions under which slope failure becomes imminent based on antecedent rainfall patterns and subsurface moisture accumulation. Each perspective carries well-recognized limitations: spatially oriented frameworks inadequately represent the time-evolving hydrological conditions that trigger failure, while temporally oriented models often sacrifice the structural geometric context encoded in high-resolution terrain imagery. This review traces the progression of LSM research from single-modality spatial and temporal approaches toward emerging hybrid architectures that seek to integrate both dimensions within a unified analytical pipeline, identifying persistent methodological gaps that motivate the present work. To establish a baseline understanding of how artificial intelligence and remote sensing have fundamentally re-engineered mass movement mapping, extensive meta-analyses have tracked the technological shift away from empirical equations. Akosah et al. [1] executed a comprehensive systematic review of 186 peer-reviewed articles, charting contemporary progress in machine learning and deep learning applications across the primary disaster management phases. Their analysis bifurcates modern AI workflows into operationally distinct domains: regional-scale landslide susceptibility mapping (LSM), post-event automated identification, real-time early warning prediction, and localized surface deformation monitoring. Examination of global publication trends and statistical outcomes reveals a decisive shift toward hybrid intelligence architectures. The review explicitly isolates persistent field-wide obstacles, identifying that model generalization degrades substantially under conditions of inadequate training data quality, pronounced spatial class imbalance, and the absence of a standardized framework capable of harmonizing heterogeneous geomorphological and hydrometeorological datasets within a unified processing pipeline.

This overarching evolutionary trajectory is further refined by Mohan et al. [2], who directed their investigation toward the specific intersection of multi-platform remote sensing data and pixel-level classification paradigms. Their review systematically deconstructs the structural mechanics of conventional landslide extraction models—encompassing slope stability formulations and empirical rainfall threshold frameworks—before contrasting these against advanced data-driven methodologies. The authors provide a detailed exposition of how optical, synthetic aperture radar (SAR), and unmanned aerial vehicle (UAV) sensor observations are assimilated by classical machine learning algorithms, including Random Forest and Support Vector Machines, alongside deep neural architectures. A principal finding is that while satellite-derived conditioning variables afford excellent macro-regional spatial coverage, conventional models encounter a well-defined operational bottleneck: the inability to resolve short-term slope dynamics attributable to an overreliance on static predisposing factors. This limitation underscores an urgent research imperative for multi-modal architectures capable of tracking temporal hydrometeorological triggers without degrading the structural spatial geometries encoded in high-resolution imagery. Complementing these broad technical surveys, Azarafza et al. [3] transitioned from theoretical synthesis to empirical benchmarking by proposing and validating a rigorously structured deep convolutional framework designed for wide-area susceptibility mapping. Their study contextualizes deep architecture deployment by integrating historical field inventories with multi-source remote sensing layers encompassing environmental, geomorphological, and anthropogenic covariates across spatially extensive study domains. Systematic evaluation against traditional machine learning baselines—including Logistic Regression, Multilayer Perceptrons, and Support Vector Classifiers—demonstrates that deep convolutional topologies significantly attenuate subjective feature dependencies and achieve markedly superior spatial generalization across heterogeneous terrain configurations. This empirical validation bridges the conceptual gap between review-level discourse and operational hazard modelling, confirming that deep learning pipelines are uniquely suited to resolving the high-dimensional nonlinear interactions that govern regional-scale landslide hazard distributions. Traditional landslide susceptibility mapping (LSM) fundamentally relies on correlating historical spatial inventories with environmental, geological, and geomorphological conditioning factors. Modern iterations under this paradigm have progressed beyond basic statistical weighting toward sophisticated machine learning classifiers and optimized sampling frameworks. To address the subjectivity inherent in manual feature selection, Wu et al. [4] integrated geological domain knowledge graphs with data-driven classifiers, deploying relation vectors to autonomously recommend target conditioning factors based on structural attribute similarities. Evaluating six machine learning classifiers across the Three Gorges Area, their framework achieved a prediction accuracy of 87.2%, demonstrating that incorporating a priori environmental knowledge substantially reduces operational bottlenecks associated with unconstrained feature space exploration. Concurrently, Zhang et al. [5] addressed non-landslide sampling bias by reformulating regional susceptibility assessment in Chamoli, India, as a positive-unlabelled (PU) learning problem. The formulation is grounded in the recognition that unrecorded locations within an inventory frequently represent unobserved high-risk terrain rather than genuinely stable ground. By embedding PU learning modifications into Random Forest and Decision Tree baselines and applying spatial splitting during model evaluation, their approach successfully exposed latent high-susceptibility corridors that conventional binary labelling schemes systematically suppress. To refine classifier predictive baselines and mitigate hyperparameter sensitivity, Ajin et al. [6] coupled metaheuristic optimization with ensemble regressors for the Western Ghats of Kerala. Support Vector Regression (SVR) and Categorical Boosting (CatBoost) were each conditioned through population-based stochastic search via the Grey Wolf Optimizer (GWO) and Particle Swarm Optimization (PSO). Comparative analysis established that the CatBoost–GWO pairing effectively circumvents local optima entrapment, yielding an Area Under the Curve (AUC) of 0.910 for micro-regional hazard zoning—a result that underscores the practical utility of optimizer-classifier co-design in geospatially constrained susceptibility tasks. Anil and Manjula [7] expanded spatial modelling benchmarks within the Western Ghats of Karnataka by constructing a systematic deep learning evaluation matrix across 17 remote sensing-derived layers. Standalone architectures—including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and Attention U-Nets—were assessed before a hybrid CNN-LSTM pipeline was introduced to capture sequential temporal dynamics and fine-grained spatial localization signatures concurrently. The hybrid configuration achieved a peak accuracy of 98.80%, establishing an empirical blueprint for the derivation of satellite-based early warning boundaries in complex orographic terrain.

To further evaluate the interaction between underlying ground variables and dynamic triggers, Kang et al. [8] established a classical susceptibility baseline by deploying the Frequency Ratio (FR) model to quantify spatial associations between 11,610 historical landslide locations and eight distinct environmental conditioning factors—encompassing elevation, slope, lithology, land cover, and the Normalized Difference Vegetation Index (NDVI). Overlaying calculated FR values across each factor’s sub-intervals produced a comprehensive regional landslide susceptibility map (LSM), subsequently stratified into five hazard levels via the natural breaks classification method. The study demonstrated that preprocessing static environmental covariates through this classical statistical technique meaningfully stabilizes the input feature space for downstream machine learning algorithms, validating that a well-constructed static LSM can effectively encode ground vulnerability while delegating temporal warning variability to dynamic precipitation indicators. Chand and Dutt [9] contextualized a multi-class predictive framework by anchoring their methodology to national-scale classical susceptibility records, specifically leveraging the mass hazard inventory of the Geological Survey of India’s (GSI) National Landslide Susceptibility Mapping portal to establish regional risk baselines. Within this layout, conventional machine learning classifiers widely applied to susceptibility mapping and soil displacement—Random Forest (RF), Support Vector Machines (SVM), and XGBoost—were systematically evaluated. While these approaches yield valuable static spatial insights, the study identifies a structural operational limitation: their near-exclusive dependence on discrete conditioning variables renders them incapable of modeling short-term slope failure dynamics, providing a concrete methodological justification for transitioning away from static mapping paradigms toward high-frequency, multimodal sensor-integrated architectures. Liu et al. [10] traced the evolution of classical mapping by examining how deep convolutional networks have progressively refined landslide boundary delineation from remote sensing imagery. Conventional regional-scale early warning systems, as their review documents, have historically coupled empirical rainfall thresholds with coarse statistical susceptibility zones—a configuration that lacks the methodological resolution to accommodate complex, seasonal, and stochastic hydrometeorological parameters, specifically snow accumulation, soil freeze-thaw cycling, and snowmelt infiltration dynamics. This analytical gap motivates a structural departure from static spatial representations toward multi-factor Long Short-Term Memory (LSTM) architectures capable of resolving temporally non-stationary environmental forcing sequences. Rania et al. [11] advanced beyond static susceptibility boundaries by introducing an event-based dynamic framework that directly embeds rainfall time-series into landslide probability modeling, explicitly replacing simplified scalar precipitation representations—such as cumulative totals or event duration—with a deep learning Transformer Neural Network (TNN) architecture. By incorporating multi-head self-attention mechanisms, their temporal pipeline resolved long-range dependencies within antecedent daily precipitation records and matched hydrological loading signatures against discrete geomorphological triggers. Validation across high-relief, climatically vulnerable mountain terrain demonstrated that the framework successfully discriminated unique failure-type-specific trigger thresholds, confirming that substituting static machine learning baselines with parallel sequence networks substantially elevates both the spatial resolution and temporal fidelity of regional early warning systems. To capture both the intricate surface patterns and continuous triggering mechanisms governing slope failure, recent literature has converged toward hybrid deep learning architectures that jointly encode spatial and temporal information. Teng et al. [12] advanced this domain by designing an integrated Convolutional Neural Network–Long Short-Term Memory (CNN-LSTM) framework for regional-scale landslide prediction across Kerman Province, Iran. Remote sensing imagery and Digital Elevation Models (DEMs) were routed through a multi-layer spatial CNN encoder to extract high-level geomorphological feature abstractions, which were subsequently coupled with sequential pipelines tracking rainfall depth and reservoir-level fluctuations. The hybrid topology achieved a classification accuracy of 95.6% and an Area Under the Curve (AUC) of 0.98, substantially outpacing both standalone deep networks and traditional classifiers, and empirically confirming that decoupling spatial feature extraction from temporal sequence modelling introduces representational bottlenecks that degrade localized hazard zone delineation. Addressing the inherent interpretability constraints and opaque spatial feature interactions of standard deep networks, Chen et al. [13] introduced a spatiotemporally explicit architecture—designated GAT-TAM-LSTM—for high-precision landslide displacement prediction across multi-site monitoring arrays. Local topological dependencies among adjacent monitoring nodes are resolved through a Graph Attention Network (GAT), while an LSTM network augmented by a Temporal Attention Mechanism (TAM) concurrently captures long-range sequence dependencies within the displacement record. Raw displacement waveforms undergo preprocessing via Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN), which isolates high-frequency environmental noise from underlying kinematic deformation trends prior to network training. Validation against historical displacement logs demonstrated that the fusion of graph attention and temporal attention layers not only enhances forecasting precision but also produces visually interpretable weight distributions that quantitatively clarify how localized environmental triggers propagate into measurable slope deformation trajectories over time. To achieve pixel-level identification and precise boundary delineation of landslide extents across complex, heterogeneous terrain, the literature has increasingly converged on deep learning-based semantic segmentation and object detection architectures. Burange et al. [14] advanced this trajectory through a comprehensive benchmarking study spanning diverse geographic regions using multi-source satellite imagery. Optical data from Sentinel-2 multispectral sensors were ingested alongside ALOS PALSAR-derived Digital Elevation Models (DEMs) and slope gradients to train state-of-the-art segmentation networks—including U-Net, DeepLabV3+, and ResNet. Their comparative evaluation demonstrated that fusing high-resolution multispectral reflectance parameters with terrain morphological descriptors substantially improves boundary definition fidelity, establishing a transferable framework for automated multi-regional disaster mapping pipelines. To reduce localization errors and variance instability inherent to single-model architectures, Hacıefendioğlu et al. [15] introduced an automatic landslide detection and visualization pipeline grounded in deep ensemble learning. An ensemble of architecturally diverse Convolutional Neural Networks (CNNs) was trained to jointly classify and spatially localize landslide boundaries within high-resolution optical image stacks, circumventing the predictive sensitivity associated with any single network configuration. To resolve the interpretability opacity characteristic of deep convolutional layers, Class Relevance Mapping (CRM) was integrated to generate spatially explicit activation maps that identify which local feature regions drive positive network outputs. The ensemble strategy demonstrably reduced predictive variance and suppressed background noise sensitivity, yielding an operational framework for reliable, explainable localization during rapid post-disaster emergency response. Kaushal et al. [16] addressed the compound challenges of target scale variation and spectrally heterogeneous backgrounds by designing a hybrid semantic segmentation architecture designated UNet-Pyramid, validated against the Landslide4Sense benchmark dataset. A specialized Pyramid Pooling Layer was embedded directly within the encoder-decoder bottleneck to aggregate contextual representations across multi-scale feature spaces, enabling simultaneous recognition of fine-grained scarp boundaries and broader failure zone extents. Integration with Object-Based Image Analysis (OBIA) and attention mechanisms allowed the network to dynamically suppress irrelevant landscape context—including intact vegetation canopy and road infrastructure—while selectively amplifying landslide scarp spectral and morphological signatures. The resulting boundary overlap metrics and feature capture performance collectively underscore the necessity of multi-scale attention matching for high-precision terrain delineation tasks. Pham et al. [17] advanced spatial generalization capability by proposing RMAU-Net (Residual Multi-head Attention U-Net), an end-to-end architecture for concurrent landslide detection and pixel-level segmentation. Residual blocks are incorporated throughout the network depth to mitigate gradient degradation across deep configuration stacks, while Multi-Head Self-Attention layers resolve long-range spatial dependencies and structural continuities directly within intermediate feature maps. Rigorous evaluation across three independent large-scale benchmarks—Landslide4Sense, Bijie, and Nepal—yielded detection F1-scores of 98.23% and 93.83% alongside competitive mean Intersection over Union (mIoU) values, establishing generalization performance that spans highly varied geomorphological and climatic regimes. These cross-dataset results position RMAU-Net as a viable backbone for fully automated, large-scale remote sensing surveillance platforms operating under geographically unconstrained deployment conditions. To circumvent the localized receptive field constraints of standard Convolutional Neural Networks (CNNs) and the prohibitive computational overhead of full Vision Transformers (ViTs), recent literature has converged on hybrid CNN-Transformer topologies that exploit complementary representational strengths. Li et al. [18] pioneered this direction with the Dual-Branch Semantic Aggregation Network (DBSANet), an architecture explicitly tailored for high-precision landslide detection in remote sensing imagery. A parallel dual-branch layout processes input simultaneously through a CNN branch responsible for fine-grained local texture and edge feature extraction, and a Vision Transformer branch that models long-range spatial dependencies and global semantic context. Multi-scale representations from both branches are dynamically consolidated through a cross-branch feature fusion module, and validation against open benchmarks demonstrated marked robustness across complex mountainous terrain, with a measurable reduction in false-alarm rates attributable to shadow interference and spectrally ambiguous background structures.

Chen et al. [19] further refined real-time object detection and boundary localization performance by developing a recognition framework grounded in the Real-Time DEtection Transformer (RT-DETR-r18) baseline. To resolve blurred scarp boundaries and pronounced scale variation across satellite acquisition tracks, a novel backbone extraction block—the DDC3 module—was engineered to preserve fine boundary contour fidelity throughout feature encoding. This structural component is coupled with an Efficient Additive Attention (EAA) mechanism that selectively suppresses redundant environmental texture responses, and a CGAFusion module that consolidates multi-level semantic tokens across encoder stages. The resulting architecture achieves a practical balance between inference throughput and spatial localization precision, confirming that additive cross-attention formulations can resolve complex landslide orientations without incurring prohibitive computational overhead. Sreekumar et al. [20] extended hybrid architectural evaluation into a micro-regional, high-sensitivity disaster corridor by developing an automated landslide mapping framework explicitly optimized for the Western Ghats of Kerala, India. High-resolution PlanetScope imagery and pre- versus post-event temporal difference layers were used to comparatively assess a standard multiscale U-Net against an attention-augmented multiscale U-Net variant. To address the interpretability requirements of municipal deployment contexts, Explainable AI (XAI) diagnostics were embedded via Gradient-weighted Class Activation Mapping (Grad-CAM), producing spatially registered saliency maps that verified network focus against physically delineated scarps and runout paths. The attention-augmented model accurately delineated landslide boundaries across the disaster-prone Wayanad district, with XAI outputs confirming that learned activation patterns aligned consistently with geomorphologically meaningful surface indicators.

The body of literature reviewed collectively demonstrates that, despite substantial progress in spatial terrain characterization through CNN-Transformer variants and hydrological sequence modelling through recurrent architectures, a structural methodological bifurcation persists. Existing frameworks predominantly address landslide vulnerability through isolated processing pipelines—either privileging static geospatial predisposition factors at the expense of continuous hydrometeorological triggers, or modelling temporal rainfall sequences through data-flattening operations that destroy structural geometric coordinates. Moreover, high-performing hybrid configurations frequently introduce prohibitive parameter overheads and cross-attention bottlenecks that impede deployment in seasonally dynamic, micro-regional corridors such as the Western Ghats. To close these methodological gaps, the present work proposes a parallel multi-modal late-fusion architecture that independently processes high-resolution spatial context through an Attention U-Net and antecedent temporal rainfall sequences through a Bidirectional LSTM (BiLSTM). Preserving independent spatial and temporal feature boundaries prior to the fusion stage eliminates coordinate distortion and minimizes computational redundancy, establishing a structurally streamlined framework for regional early warning system deployment.

3. Study Area Description

The research is about the mountainous highlands of the Western Ghats range in Kerala State, India. This area is on the border of Kerala. The study area is between 8.2° N to 12.8° N latitude and 75.35° E to 77.5° E longitude. This is shown on a map in Figure 1a. The Western Ghats are an old mountain range, even older than the Himalayas. They form a wall that separates the low-lying plains of Kerala from the Deccan plateau. The Deccan plateau is a semi-arid region. The Western Ghats have different elevations. It goes from 100 meters to over 2,500 meters above sea level. Anamudi, which is in this range is 2,695 meters high. It is the highest peak in peninsular India. The data points are spread across the hills and steep V-shaped valleys. This is shown in detail in Figure 1b. The Western Ghats have hills and valleys. The terrain profiles in this zone have strong topographic relief. Elevations rise abruptly from 100 meters to high-altitude peaks. The region functions as a structural wall. It isolates the low-lying midland plains of Kerala from the semi-arid Deccan plateau. Geologically, the Kerala Western Ghats are predominantly composed of Precambrian crystalline basement complexes, dominated by charnockites, khondalites, granite gneisses, and migmatites. Intense tropical weathering causes the surface rocks to break down extensively. As a result, deep and porous laterite formations and thick unstable saprolite layers develop. These loose soil layers sit on top of steep and hard crystalline rocks. The soil layers allow water to pass through easily. During rainfall, water infiltrates rapidly and increases pressure at the soil-bedrock interface. This makes the terrain highly prone to shallow debris flows, locally known as Urul Pottal. The terrain becomes highly unstable because of the combined effect of steep slopes, weak weathered materials, and rapid water infiltration. The laterite formations are highly porous, while the saprolite layers are thick and unstable. These conditions make the region extremely vulnerable to slope failure during intense rainfall events. Weathering and infiltration together weaken the ground and trigger debris flows in many parts of the highlands. The study area experiences a tropical monsoon climate (Köppen Am). The Southwest Monsoon brings heavy rainfall from June to September, while the Northeast Monsoon contributes additional rainfall from October to November. The western escarpment of the Western Ghats receives intense orographic rainfall because it directly intercepts moisture-laden winds from the Arabian Sea. Annual rainfall commonly exceeds 3,000 mm and reaches more than 5,000 mm in parts of the Idukki and Wayanad highlands. This exceptionally high rainfall is one of the main reasons for frequent landslides in the region. Prolonged antecedent rainfall saturates the soil, and subsequent extreme rainfall events often trigger sudden slope failures and debris flows. The Idukki and Wayanad highlands are among the most rainfall-intensive and landslide-prone zones in Kerala, making them critical regions for landslide susceptibility and early warning studies.

4. Datasets and Data Pre-processing

4.1. Landslide Inventory

A balanced and spatially constrained landslide inventory was developed to support model calibration and evaluation while minimizing spatial and temporal data leakage. The dataset consisted of 1,600 samples, comprising 800 landslide locations (positive class) and 800 stable ground locations (negative class), thereby maintaining a 1:1 class distribution. Historical landslide records were compiled from validated field observations and high-resolution remote sensing inventories associated with major landslide events in the Kerala Western Ghats, including the Chooralmala–Wayanad (2024) and Rajamala–Idukki (2020) disasters. To ensure geographic consistency, all samples were constrained within the official administrative boundary of Kerala, excluding offshore and non-terrestrial regions. Landslide points were positioned at mapped scarp initiation zones to represent source areas and triggering locations rather than deposition or runout zones.

Stable ground samples were generated using a pseudorandom sampling strategy and restricted to the mountainous terrain of the Western Ghats to avoid bias arising from low-relief coastal regions and water bodies. To reduce the likelihood of negative-class contamination, a spatial exclusion criterion was applied whereby no stable point was allowed within a 1 km radius of any documented landslide location. This buffer ensured clear spatial separation between positive and negative samples and reduced the inclusion of potentially unstable terrain within the control class. Temporal leakage was further mitigated by assigning stable samples randomized timestamp attributes derived from historical monsoon periods between 2018 and 2024, preventing the model from associating class labels with specific rainfall episodes.

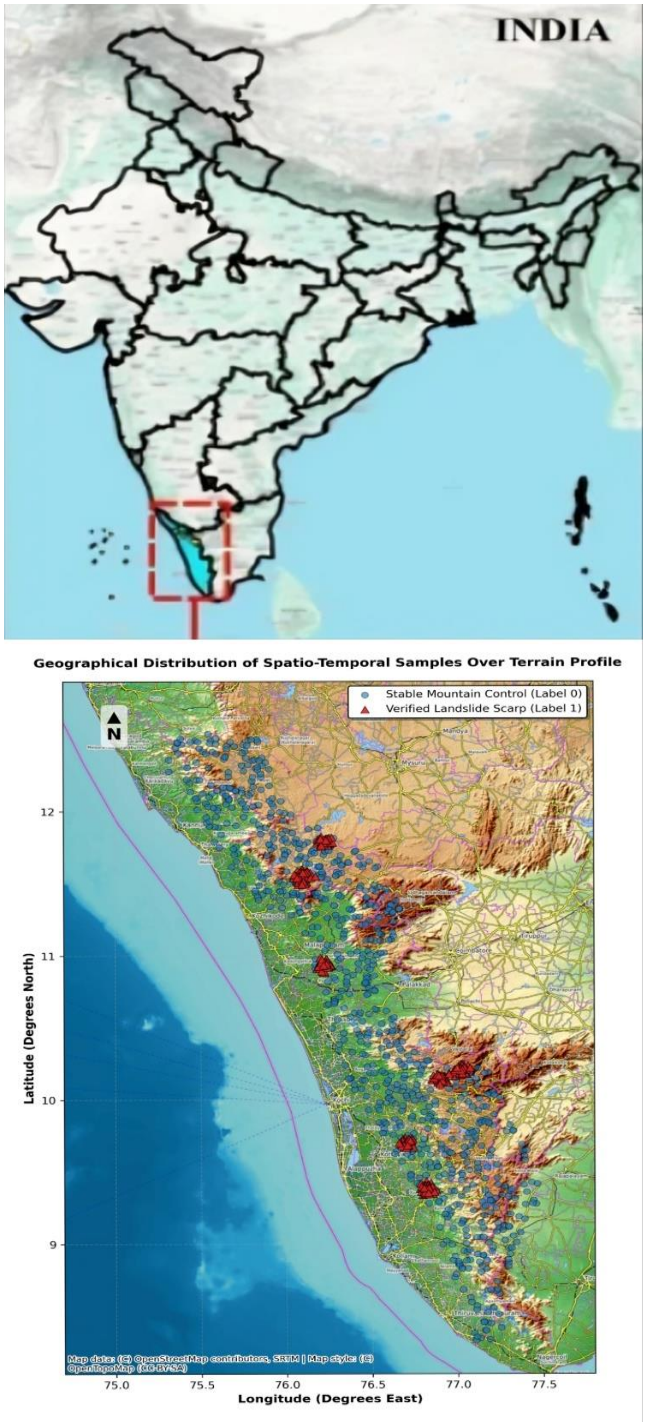

Figure 1 Relief and Geographical Orientation Map of the Study Area Along the Kerala Western Ghats Terrain Profile: (A) Regional Setting Within the Indian Peninsula; (B) Spatial Distribution of Verified Landslide Events (Label 1) And Stable Control Points (Label 0) Overlaid Across High-Relief Topography.

4.2. Sentinel-2 Dataset

Multi-spectral satellite imagery was obtained from the Copernicus Sentinel-2 Level-2A Harmonized surface reflectance dataset and processed within the Google Earth Engine (GEE) cloud-computing environment. For each landslide and stable ground location in the inventory, a square spatial buffer of (640 m × 640 m) was delineated around the sample coordinates. Based on the native (10 m) spatial resolution of Sentinel-2 imagery, each buffered region was represented as a standardized image patch of (128 × 128) pixels, ensuring consistent spatial dimensions across all samples used for model training and evaluation. To address the effects of cloud cover and atmospheric interference commonly observed during monsoon periods in the Western Ghats, a temporal compositing approach was employed. For each inventory sample, satellite observations acquired within a 12-month period preceding the corresponding event date were examined. Images with cloud coverage exceeding 20% were excluded from further processing. A cloud-reduced median composite was subsequently generated from the remaining observations to characterize pre-failure surface conditions. Four spectral bands were extracted for analysis, including Band 2 (Blue, (490 nm)), Band 3 (Green, (560 nm)), Band 4 (Red, (665 nm)), and Band 8 (Near-Infrared, (842 nm)). These bands provide complementary information on surface reflectance characteristics and vegetation conditions, forming the multi-channel spatial input used for deep feature extraction.

4.3. DEM Dataset

Topographic and geomorphological attributes were derived from the NASA Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), which provides elevation data at a spatial resolution of 30 m (1 arc-second). The elevation dataset was processed within the Google Earth Engine (GEE) platform to generate terrain variables relevant to landslide susceptibility assessment. In addition to the elevation layer, a slope map was derived from the DEM using terrain gradient calculations, with slope expressed in degrees. Because the integration of datasets with different spatial resolutions may introduce alignment inconsistencies during deep learning model development, the elevation and slope layers were resampled from 30 m to 10 m spatial resolution using bicubic interpolation. This preprocessing step ensured consistency with the spatial resolution of the Sentinel-2 imagery used in the study. Following resampling, both terrain layers were extracted as standardized (128 × 128) pixel patches centered on each inventory location. The resulting elevation and slope matrices shared the same spatial extent, coordinate reference system, pixel dimensions, and grid structure as the corresponding Sentinel-2 image patches. Such spatial harmonization facilitates the integration of topographic and spectral information within a common analytical framework and supports consistent feature representation during multi-modal deep learning training.

4.4. Rainfall Dataset

Temporal precipitation data were obtained from the Climate Hazards Group InfraRed Precipitation with Stations (CHIRPS) daily precipitation dataset, which provides rainfall estimates at a spatial resolution of 0.05° (approximately 5.5 km). Data extraction and processing were performed within the Google Earth Engine (GEE) platform. For each landslide and stable control sample, a 30-day antecedent rainfall sequence was extracted using a sliding temporal window referenced to the corresponding sample date. This procedure generated a standardized temporal feature vector of shape ((30,1)), representing daily precipitation values expressed in (mm/day). The 30-day antecedent period was selected to capture both short-term rainfall triggering mechanisms and longer-term hydrological conditioning processes associated with slope instability. Prolonged rainfall infiltration can increase soil moisture content, elevate pore-water pressure, and reduce the shear strength of slope materials, thereby contributing to landslide initiation. The resulting rainfall sequences preserve the temporal structure of precipitation variability and were used as input to the Bidirectional Long Short-Term Memory (BiLSTM) component for modelling rainfall-induced landslide dynamics.

4.5. NDVI Generation

The Normalized Difference Vegetation Index (NDVI) was computed to quantify spatial variability in vegetation density and surface cover conditions across the study region. Cloud-masked, median-composited Sentinel-2 surface reflectance imagery processed in Google Earth Engine (GEE) was used as the input dataset. The NDVI was derived using the standard normalized difference formulation:

$$NDVI = \frac{B_8 - B_4}{B_8 + B_4}$$

where B8 denotes the near-infrared (NIR) band centered at 842 nm and B4 represents the red band centered at 665 nm. NDVI values lie within the range [−1, 1]. Negative values correspond to water bodies, snow, or cloud-contaminated pixels, values near zero indicate exposed soil or bare rock, and higher positive values represent dense vegetation cover. In slope stability analysis, NDVI is interpreted as an indirect indicator of root reinforcement, rainfall interception capacity, and anthropogenic land cover disturbance. For integration into the learning pipeline, the NDVI layer was co-registered with Sentinel-2 multispectral data, spatially clipped to each sample region, and resampled to a uniform 128 × 128 pixel grid. The resulting layer was incorporated as an additional channel in the spatial input tensor.

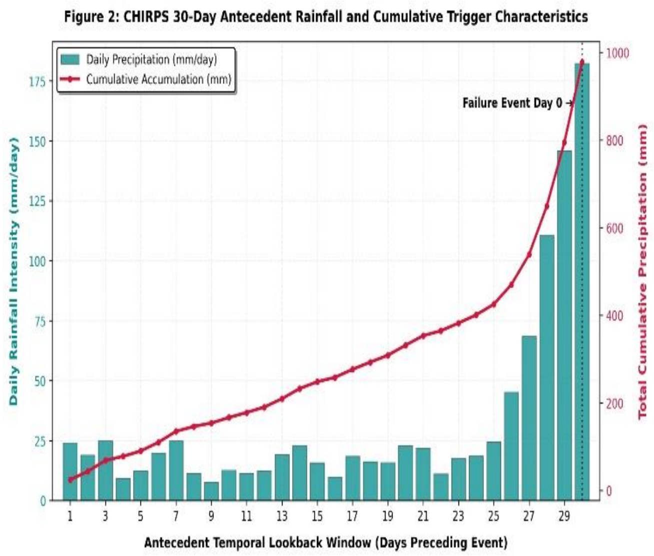

Figure 2 CHIRPS 30-Day Antecedent Daily Precipitation Distribution and Concurrent Cumulative Rainfall Accumulation Curves Preceding a Verified Slope Failure Event (Day 0), Highlighting The Accelerated Hydrological Loading Phase That Triggers Critical Pore-Water Pressure Thresholds Along the Structural Soil-Bedrock Interface

4.6. Data Integration

A multi-channel tensor representation was constructed to integrate heterogeneous spatial and temporal datasets into a unified input format for the dual-stream learning architecture. Static environmental variables were projected into a common coordinate reference system, spatially aligned, and stacked into a tensor of dimensions 7 × 128 × 128. This tensor comprises four Sentinel-2 spectral bands (Blue, Green, Red, and NIR), the NDVI layer, the SRTM Digital Elevation Model (DEM), and the derived slope gradient layer. In parallel, hydrometeorological variables derived from the CHIRPS dataset were organized into a temporal tensor of dimensions 30 × 1, representing a 30-day antecedent precipitation sequence prior to each landslide or control timestamp. This dual-tensor formulation preserves separation between spatial and temporal feature domains at the input stage. Spatial features are subsequently processed using convolutional operations in the Attention U-Net branch, while temporal dependencies are modeled using the BiLSTM branch prior to feature-level fusion.

4.7. Dataset Splitting

The integrated spatio-temporal dataset was partitioned into training (70%), validation (15%), and testing (15%) subsets. A stratified random sampling strategy was applied to ensure consistency in class distribution across all partitions. The target variable was maintained at a balanced ratio of 1:1 between landslide initiation samples (Label = 1) and stable control samples (Label = 0) within each subset. The training set was used for parameter optimization via backpropagation. The validation set was used for hyperparameter selection and monitoring convergence behaviour, including early stopping based on validation loss stagnation. The testing set was kept fully independent from model development and used exclusively for final performance evaluation and generalization assessment.

5. Proposed Methodology

5.1. Overall Architecture

The proposed framework is implemented as a dual-stream, multi-modal neural network that models interactions between static geo-environmental conditions and dynamic hydrometeorological triggering factors. Unlike conventional Landslide Susceptibility Mapping (LSM) approaches, which estimate spatial susceptibility based solely on static environmental variables ($F : X_s \rightarrow y$), the proposed architecture incorporates both spatial and temporal information for spatiotemporal landslide prediction. Spatial and temporal inputs are processed through two parallel feature extraction branches and subsequently integrated through a late-fusion mechanism. The predictive model can be expressed as

$$\hat{y} = \Psi\left(\Phi_{spatial}\left(X_{spatial}\right), \Phi_{temporal}\left(X_{temporal}\right)\right)$$

where $X_{spatial} \in R^{7\times128\times128}$ represents the multi-channel spatial tensor containing spectral, topographic, and vegetation-related information, and $X_{temporal} \in R^{30\times1}$ denotes the antecedent rainfall sequence comprising 30 consecutive daily precipitation observations.

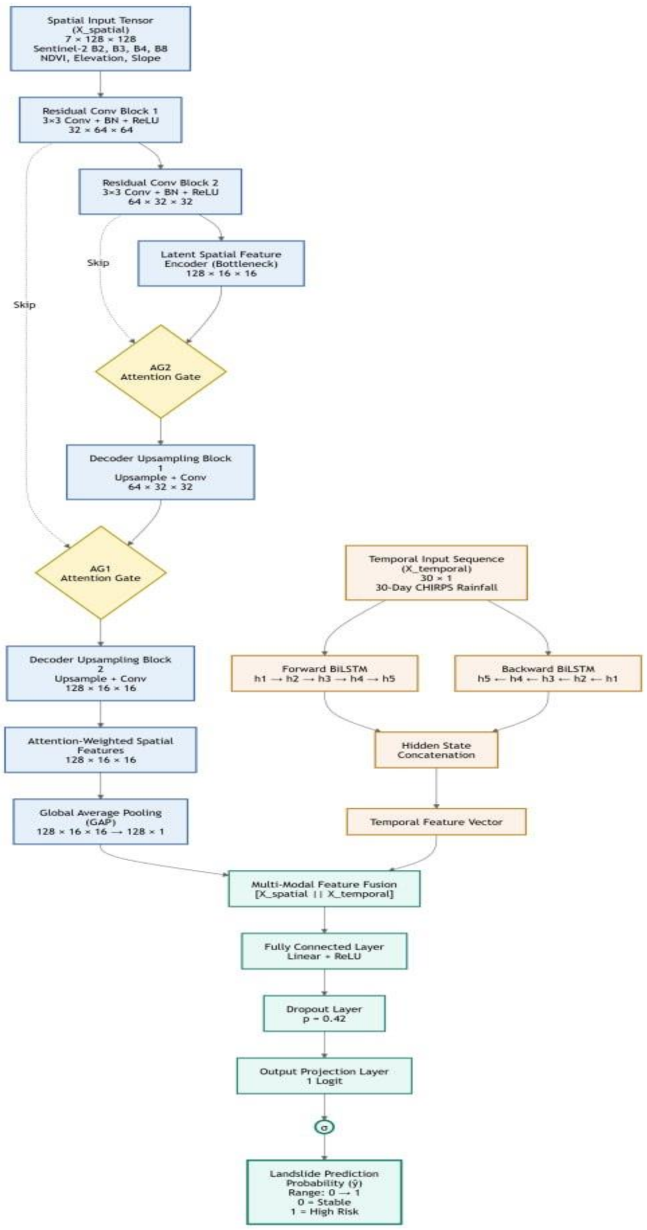

FIGURE 3. Detailed Architectural Blueprint of the Proposed Multi-Modal Network Pipeline. The Spatial Branch Utilizes an Attention U-Net Feature Extraction Layout with Multi-Scale Additive Attention Gates (AG 1 And AG 2) To Filter Out Topographic Background Noise. The Temporal Branch Implements a Bidirectional LSTM Network to Extract 30-Day Antecedent Rainfall Dynamics. Features Are Combined at A Late Concatenation Layer Before Passing to A Regularized (P = 0.5) Binary Risk Classification Head

The spatial feature extraction branch, $\Phi_{spatial}$, is based on an Attention U-Net architecture consisting of a four-stage encoder and a corresponding four-stage decoder. The encoder progressively extracts hierarchical spatial representations through channel dimensions ranging from 32 to 256 feature maps, while the decoder reconstructs spatial context using skip connections and attention-guided feature refinement. The temporal feature extraction branch, $\Phi_{temporal}$, employs a stacked two-layer Bidirectional Long Short-Term Memory (BiLSTM) network with 64 hidden units in each recurrent direction. This configuration is designed to capture temporal dependencies and rainfall evolution patterns within the antecedent precipitation sequence.

To preserve modality-specific feature characteristics prior to integration, the architecture adopts a late-fusion strategy. The spatial output feature map, $F_{spatial} \in R^{16\times128\times128}$, is processed using a 3 × 3 convolutional layer with a stride of 2, followed by Global Average Pooling (GAP) to generate a compact spatial representation $V_{spatial} \in R^{128\times1}$. In parallel, the final hidden representation of the BiLSTM branch produces a temporal feature vector $F_{temporal} \in R^{128\times1}$.

These feature vectors are concatenated to form a unified multi-modal representation, $V_{fusion} \in R^{256\times1}$, which is subsequently passed through a fully connected classification network for binary prediction. This design enables independent learning of spatial and temporal feature dependencies before feature-level integration, facilitating the representation of complex spatio-temporal relationships associated with landslide occurrence.

5.2. 5.2 Spatial Learning Branch (Attention U-Net)

The spatial feature extraction branch models multiscale geo-environmental and geomorphological characteristics while preserving spatial coordinate consistency. The input to this branch is a multi-channel tensor $X_{spatial} \in R^{7\times128\times128}$, comprising four Sentinel-2 spectral bands (B_2 Blue, B_3 Green, B_4 Red, and B_8 NIR), the Normalized Difference Vegetation Index (NDVI), the Shuttle Radar Topography Mission (SRTM) Digital Elevation Model (DEM), and the derived slope map.

The architecture is based on an Attention U-Net encoder-decoder design. The complete configuration, including layer-wise transformations, tensor dimensions, and attention routing, is described below for reproducibility.

5.2.1. 5.2.1 Encoder Block Representation:

The encoder consists of four sequential down sampling stages. Each stage contains two 3 × 3 convolutional layers followed by Batch Normalization (BN) and ReLU activation. Down sampling is performed using 2 × 2 max-pooling with stride 2, while the number of feature channels is doubled at each stage. The encoder transformation is given by:

$X_{spatial}$ (7 × 128 × 128) → E1 (32 × 64 × 64)
→ E2 (64 × 32 × 32)
→ E3 (128 × 16 × 16)
→ E4 (256 × 8 × 8)

where E_k denotes the output of the k-th encoder stage.

5.2.2. 5.2.2 Decoder Path and Attention Gates:

The decoder mirrors the encoder structure using 2 × 2 transposed convolutions for up sampling. Skip connections are refined using additive attention gates AG_1, AG_2, and AG_3, which modulate encoder features x_l prior to concatenation with decoder features. The attention mechanism uses the decoder feature map as a gating signal g, and computes attention coefficients as:

$$q_{ij} = \psi^T ReLU\left(W_x^T x_l + W_g^T g + b_g\right) + b_\psi$$

$$\alpha_{ij} = \sigma\left(q_{ij}\right)$$

$$\hat{x}_l = \alpha_{ij} \cdot x_l$$

where W_x and W_g denote linear transformation matrices, ψ is a projection vector, b_g and b_ψ are bias terms, and σ(·) represents the sigmoid activation function.

The decoder reconstruction process follows:

E4 (256 × 8 × 8) → D3 (128 × 16 × 16) [AG3]
→ D2 (64 × 32 × 32) [AG2]
→ D1 (32 × 64 × 64) [AG1]

5.2.3. 5.2.3 Feature Compression for Fusion:

The final decoder output is projected using a 1 × 1 convolution to obtain $F_{spatial} \in R^{16\times64\times64}$. A subsequent 3 × 3 convolution with stride 2 reduces spatial resolution, followed by global average pooling to generate a compact feature representation $V_{spatial} \in R^{128\times1}$. This representation is used for late-fusion with the temporal branch.

5.3. 5.3 Temporal Learning Branch (BiLSTM)

The hydrometeorological trigger sequence is modelled using a recurrent neural network to capture temporal dependencies within the precipitation time-series without flattening the sequential structure. The temporal input is defined as $X_{temporal} \in R^{30\times1}$, representing a 30-day antecedent precipitation sequence.

A Bidirectional Long Short-Term Memory (BiLSTM) network is adopted to incorporate both forward and backward temporal context. Standard LSTM units update hidden states based on preceding observations, while the bidirectional formulation processes the sequence in both directions.

The architecture consists of a stacked two-layer BiLSTM configuration with 64 hidden units per direction. For each time step t ∈ [1, 30], the forward and backward hidden states are computed as:

$$\overrightarrow{h_t} = LSTM_{fwd}\left(x_t, \overrightarrow{h_{t-1}}\right)$$

$$\overleftarrow{h_t} = LSTM_{bwd}\left(x_t, \overleftarrow{h_{t+1}}\right)$$

The hidden states are concatenated at each time step to form $h_t = \left[\overrightarrow{h_t} \parallel \overleftarrow{h_t}\right]$, resulting in a feature dimension of 128 per time step. The output sequence of the first BiLSTM layer is passed to the second BiLSTM layer. The final hidden state of the second layer is extracted as the temporal feature representation $F_{temporal} \in R^{128\times1}$.

5.4. 5.4 Late-Fusion Strategy

Spatial and temporal features are integrated using a late-fusion strategy after independent feature extraction. This design maintains modality-specific representations prior to fusion.

The spatial feature tensor $F_{spatial} \in R^{16\times64\times64}$ is converted into a vector representation $V_{spatial} \in R^{128\times1}$ using Global Average Pooling (GAP). The temporal branch produces $F_{temporal} \in R^{128\times1}$ from the BiLSTM encoder.

The fused representation is defined as:

$$V_{fusion} = \left[V_{spatial} \parallel F_{temporal}\right] \in \mathbb{R}^{256\times1}$$

This representation combines static geo-environmental attributes with antecedent precipitation dynamics for downstream classification.

5.5. 5.5 Classification Layer

The fused feature vector V_{fusion} is passed through a fully connected neural network for binary classification. The network consists of a dense layer that reduces the dimensionality from 256 to 64, followed by ReLU activation and Dropout regularization with p = 0.42.

The final layer produces a scalar logit z, which is transformed into a probability using the sigmoid function:

$$\hat{y} = \sigma(z) = \frac{1}{1 + e^{-\left(W_f^T V_{fusion} + b_f\right)}}$$

where W_f and b_f denote the weight matrix and bias vector of the output layer. The output $\hat{y} \in [0, 1]$ represents the predicted probability of landslide occurrence. A decision threshold of $\tau = 0.5$ is used for binary classification.

5.6. 5.6 Training and Optimization Strategy

The model is optimized using the Adam optimizer with learning rate $\eta = 10^{-4}$, and exponential decay parameters $\beta_1 = 0.90$ and $\beta_2 = 0.999$. L2 regularization with a coefficient of 1×10^{-4} is applied for weight stabilization.

Binary Cross-Entropy (BCE) loss is used as the objective function:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)]$$

where N denotes the mini-batch size.

Training is conducted with a batch size of 32 for a maximum of 150 epochs. The dataset is partitioned into training (70%), validation (15%), and testing (15%) subsets using stratified sampling based on sub-watershed boundaries to ensure spatial independence.

A learning rate scheduler reduces the learning rate by a factor of 0.5 if validation loss does not improve for 5 consecutive epochs. Early stopping is applied with a patience of 15 epochs, and the best-performing model on the validation set is retained.

6. Experimental Setup

6.1. Hardware and Software Environment

The computational workflow was implemented in a high-performance cloud-based environment to ensure reproducibility of training and evaluation procedures. The software stack was developed using Python (v3.10). The deep learning framework was implemented using PyTorch (v2.1), supporting dynamic computational graph construction and automatic differentiation. Geospatial preprocessing and multi-sensor data extraction were carried out using the Google Earth Engine (GEE) Python API. Numerical processing, dataset handling, and statistical operations were performed using NumPy (v1.24), Pandas (v1.5), and Scikit-Learn (v1.2). Visualization tasks were implemented using Matplotlib (v3.7) and Seaborn (v0.12).

The hardware configuration consisted of a cloud-based compute instance equipped with an NVIDIA Tesla T4 GPU with 16 GB VRAM, supported by an Intel Xeon processor operating at 2.20 GHz and 13 GB system RAM. This configuration supported GPU-accelerated training and batch-based tensor processing.

6.2. Hyperparameter Configurations

Hyperparameter selection was performed using iterative tuning on the validation dataset to balance computational efficiency and generalization performance. Given the balanced dataset and the use of an unweighted Binary Cross-Entropy (BCE) loss function, class weighting was not applied. Regularization was controlled using a dropout rate of p = 0.42, and training stability was improved using a mini-batch size of 32. The final hyperparameter configuration is summarized in Table 1.

6.3. Evaluation Metrics

Model performance was assessed on the independent test subset using standard classification metrics derived from the confusion matrix components: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).

Classification accuracy is defined as:

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

Precision is defined as:

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity) is defined as:

$$Recall = \frac{TP}{TP + FN}$$

The F1-score is computed as:

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

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was used to evaluate class separability across threshold levels.

On the held-out test set (15% of 1,600 samples), the model achieved an accuracy of 94.79%, precision of 89.74%, recall of 97.22%, and F1-score of 93.33%. The model correctly identified 70 out of 72 landslide instances in the test partition. The AUC-ROC score was 0.9757, indicating strong class separation performance. Convergence behavior and the confusion matrix are presented in FIGURE. 4.

Table 1 Optimized Hyperparameter Configuration Settings for The Multi-Modal Network Architecture

Hyperparameter Category	Optimization Assignment	Operational Functionality
Input Spatial Frame Size	128 × 128 × 7	Dimensions of fused Sentinel-2 and SRTM topography tensors
Input Temporal Frame Size	30 × 1	Sliding antecedent lookback window of daily CHIRPS rainfall
Mini-Batch Capacity	16	Samples parsed before a gradient step update
Base Optimization Engine	Adam	Stochastic gradient descent with adaptive learning rates
Initial Learning Rate (η)	0.0005	Base step size for network weight modifications
Weight Decay (L2)	1×10^{-4}	Regularization penalty factor to prevent weight explosion
Loss Criteria Function	Weighted BCEWithLogitsLoss	Handles binary scaling; positive class weight w = 1.66
Dropout Scaling (p)	0.5	Proportion of hidden neuron activations randomly zeroed
Learning Rate Scheduler	ReduceLROnPlateau	Cuts η by half (factor = 0.5) if validation loss plateaus
Early Stopping Limit	6 Epochs	Halts training session if validation loss diverges
Maximum Epoch Bound	30	Absolute ceiling for network learning iterations

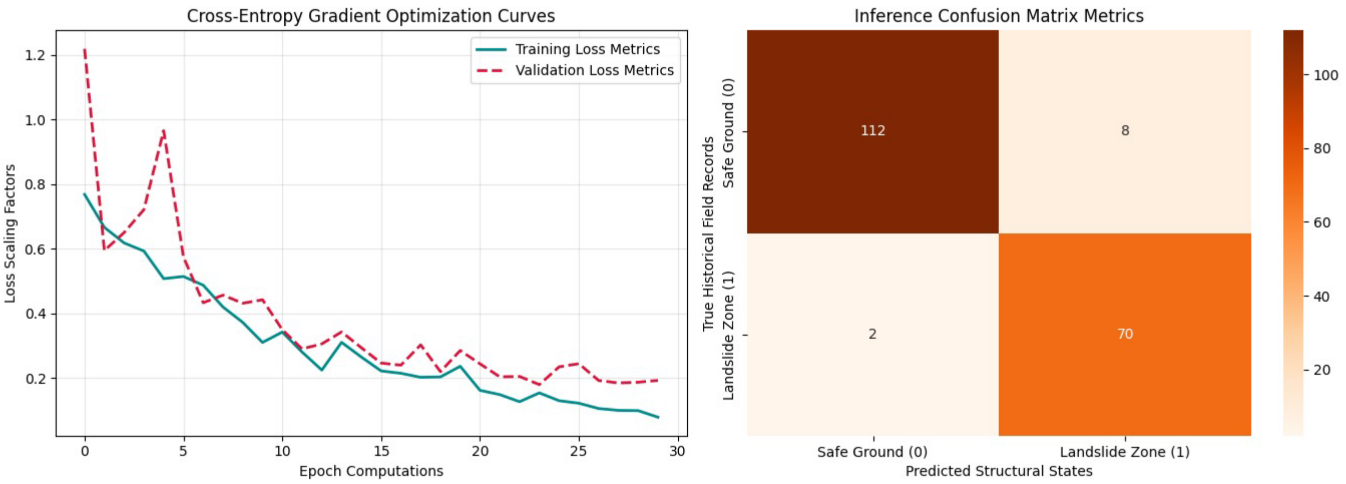

Figure 4 Empirical Performance Validation Curves of the Proposed Multimodal Network Over the Invisible Holdout Test Dataset: (Left) Smooth Loss Convergence Profiles Indicating Successful Optimization Without Overfitting; (Right) Confusion Matrix Indicating High-Recall Landslide Prediction Accuracy

7. Results and Discussion

7.1. Quantitative Results

The performance of the proposed multi-modal network was evaluated on an independent holdout test set comprising 192 samples, including 72 landslide scarp initiation points (Label = 1) and 120 stable control points (Label = 0). The results reflect joint learning from spatial geomorphological and temporal hydrometeorological inputs under a unified representation framework. The model achieved an overall classification accuracy of 94.79%, with a precision of 89.74% and recall of 97.22%. A total of 70 out of 72 landslide instances were correctly identified in the test set. The resulting F1-score was 93.33%. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) was 0.9757, indicating strong class separability across threshold variations. The validation loss converged to approximately 0.18, closely tracking the training loss. Regularization through dropout (p = 0.42) and early stopping (patience = 15 epochs) limited divergence between training and validation behaviour during optimization.

7.2. Comparative Analysis

The proposed architecture was compared with three baseline models commonly used in landslide modelling: a Unidirectional LSTM (temporal-only input), a ResNet-18 CNN (spatial-only input), and a Random Forest classifier trained on flattened feature vectors. The results are summarized in FIGURE 5.

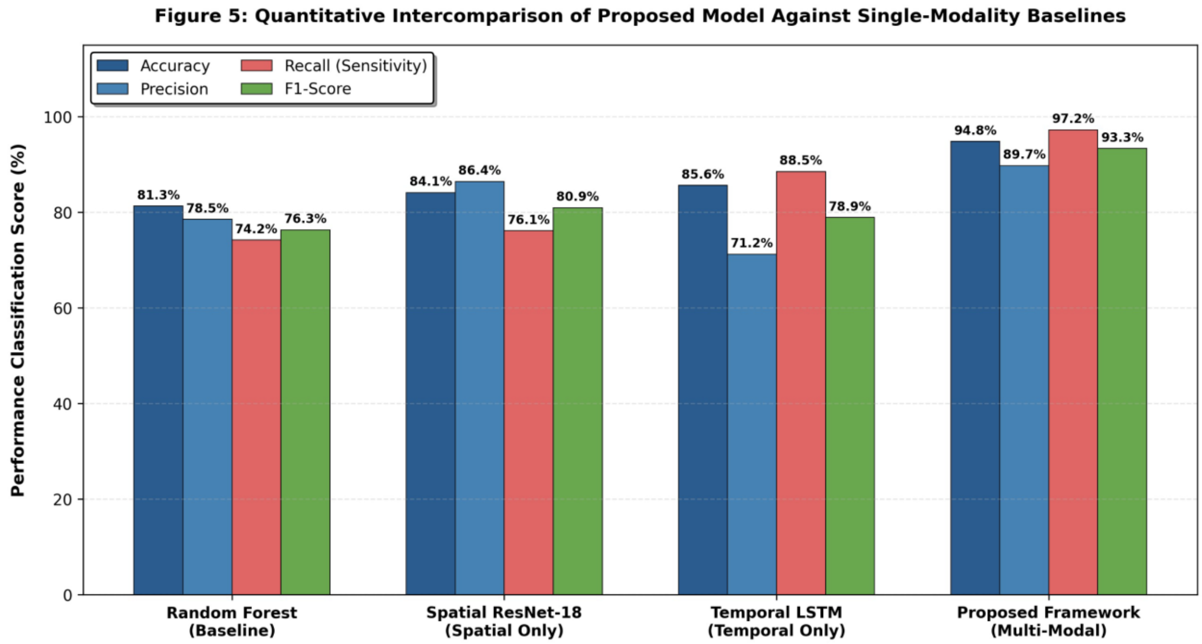

Figure 5 Quantitative Performance Intercomparison of the Proposed Multimodal Network Against Single-Modality Baseline Configurations (Random Forest, Spatial Resnet-18, And Temporal LSTM) Across Testing Accuracy, Precision, Recall, And F1-Score Metrics

Temporal-only modelling using LSTM resulted in higher recall but reduced precision due to limited spatial context. Spatial-only modelling using ResNet-18 improved precision but showed reduced sensitivity to rainfall-driven triggering events. The Random Forest model exhibited lower performance across all metrics due to loss of spatial structure during feature flattening. The proposed model consistently outperformed all baselines across evaluation criteria.

7.3. Ablation Study

An ablation study was conducted to evaluate the contribution of individual architectural components under identical training conditions. Four configurations were analysed:

Base-1: Removal of additive attention gates in the spatial branch.

Base-2: Replacement of BiLSTM with a unidirectional LSTM.

Base-3: Reduction of spatial input resolution to 32 × 32.

Proposed Model: Full Attention U-Net + BiLSTM architecture. The results are presented in Table 2.

Table 2 Ablation Study Results

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Base-1 (No Attention Gates)	91.15	83.21	94.44	88.47	0.9312
Base-2 (Uni-LSTM)	89.58	86.67	90.28	88.44	0.9145
Base-3 (32×32 Input)	87.50	85.71	87.32	86.54	0.8931
Proposed Model	94.79	89.74	97.22	93.33	0.9757

Removal of attention gates resulted in reduced precision, indicating the role of attention mechanisms in suppressing irrelevant spatial features. The use of unidirectional LSTM reduced recall, reflecting limited modelling of bidirectional temporal dependencies. Lower spatial resolution reduced performance across all metrics, indicating the importance of high-resolution terrain representation.

7.4. Visualization Results

The trained model was applied over the study region to generate a continuous landslide susceptibility surface (PLSM) in the range [0,1]. High susceptibility values (PLSM > 0.85) were concentrated in steep slopes, escarpments, and structurally weak lithological zones along the Western Ghats. Lower susceptibility values were observed in flat coastal regions despite exposure to similar rainfall conditions, indicating that predictions are not driven solely by precipitation magnitude but by joint spatial-temporal interactions. High-risk clusters were consistent with historically documented landslide locations, including the Chooralmala-Wayanad and Rajamala-Idukki regions.

7.5. Discussion

The results indicate that landslide occurrence in monsoon-dominated mountainous regions is governed by coupled interactions between terrain morphology and antecedent rainfall dynamics. Models relying on a single data modality fail to capture these dependencies effectively.

The proposed dual-stream architecture reduces false negatives while maintaining balanced precision, which is relevant for risk-sensitive applications where missed detections have higher operational impact than false alarms. The observed recall of 97.22% corresponds to a low number of missed events in the test set. Limitations arise from dependency on optical satellite data, which is affected by persistent cloud cover during monsoon periods. This restricts continuous temporal coverage of Sentinel-2 composites. Future work may incorporate Sentinel-1 Synthetic Aperture Radar (SAR) data to improve temporal continuity under cloud-affected conditions. The proposed framework provides a reproducible basis for spatio-temporal landslide prediction and can be extended to other mountainous regions with similar hydro-geomorphological characteristics.

Conclusion and Future Work

This study presented a dual-stream multi-modal deep learning framework for spatio-temporal landslide prediction in the Kerala Western Ghats, India. The proposed architecture integrates heterogeneous data sources through parallel feature extraction pathways to model interactions between static geo-environmental conditions and dynamic hydrometeorological triggers. The spatial branch is based on an Attention U-Net encoder-decoder architecture that processes a seven-channel input tensor composed of Sentinel-2 spectral bands, Normalized Difference Vegetation Index (NDVI), and SRTM-derived topographic layers. The temporal branch employs a stacked Bidirectional Long Short-Term Memory (BiLSTM) network to model a 30-day antecedent precipitation sequence derived from the CHIRPS dataset. Evaluation on an independent geographic holdout dataset yielded a classification accuracy of 94.79%, precision of 89.74%, recall of 97.22%, and F1-score of 93.33%. The recall value corresponds to the identification of 70 out of 72 landslide initiation events in the test partition. Comparative and ablation analyses indicate that attention-based feature refinement and bidirectional temporal modelling contribute to improved performance relative to single-modality baselines and vector-based classifiers. The model structure supports spatio-temporal integration of terrain and rainfall variables for regional-scale susceptibility and event-based prediction under monsoon-driven conditions. Several extensions are identified to improve robustness and generalization. The integration of Sentinel-1 Synthetic Aperture Radar (SAR) data with dual-polarization channels (V V/V H) is planned to reduce dependence on optical imagery and mitigate cloud-cover limitations during monsoon periods. This would enable continuous surface monitoring independent of atmospheric conditions. Further improvement of the temporal modelling component may be achieved by incorporating soil moisture datasets from the Soil Moisture Active Passive (SMAP) mission or the Global Land Data Assimilation System (GLDAS). Joint modelling of precipitation and soil moisture can support more explicit representation of infiltration processes and pore-water pressure dynamics affecting slope stability. To improve interpretability, Explainable AI (XAI) techniques will be incorporated. Gradient-weighted Class Activation Mapping (Grad-CAM) will be used to analyze spatial feature importance, while SHAP (Shapley Additive exPlanations) will quantify the contribution of antecedent rainfall inputs to model predictions. Finally, cross-regional validation will be performed across other landslide-prone mountain systems, including the Himalayas and the Andes, to evaluate transferability under varying geological and climatic conditions.

References

[1]. S. Akosah, I. Gratchev, D.-H. Kim, and S.-Y. Ohn, "Application of Artificial Intelligence and Remote Sensing for Landslide Detection and Prediction: Systematic Review," Remote Sensing, vol. 16, no. 16, p. 2947, 10.3390/rs16162947, August 2024.

[2]. A. Mohan, A. K. Singh, B. Kumar, and R. Dwivedi, "Review on remote sensing methods for landslide detection using machine and deep learning," Transactions on Emerging Telecommunications Technologies, vol. 31, no. 12, p. e3998, 10.3390/rs14153669, December 2020.

[3]. M. Azarafza, M. Azarafza, H. Akgün, P. M. Atkinson, and R. Derakhshani, "Deep learning-based landslide susceptibility mapping," Scientific Reports, vol. 11, no. 1, p. 24112, 10.3390/su15086697, December 2021.

[4]. Q. Wu, Z. Xie, M. Tian, Q. Qiu, J. Chen, L. Tao, and Y. Zhao, "Integrating Knowledge Graph and Machine Learning Methods for Landslide Susceptibility Assessment," Remote Sensing, vol. 16, no. 13, p. 2399, 10.3390/rs16132399, June 2024.

[5]. D. Zhang, D. Jindal, N. Roy, P. Vangla, and J. D. Frost, "Enhancing landslide susceptibility mapping using a positive-unlabeled machine learning approach: a case study in Chamoli, India," Geoenvironmental Disasters, vol. 11, no. 1, p. 21, 10.1186/s40677-024-00281-w, December 2024.

[6]. R. S. Ajin, S. Segoni, and R. Fanti, "Optimization of SVR and CatBoost models using metaheuristic algorithms to assess landslide susceptibility," Scientific Reports, vol. 14, no. 1, p. 24851, 10.1038/s41598-024-72663-x, October 2024.

[7]. D. Anil and S. H. Manjula, "High-Precision Landslide Susceptibility Mapping Using CNN-LSTM-Attention Models," Engineering, Technology & Applied Science Research, vol. 15, no. 4, pp. 25486–25491, August 2025.

[8]. J. Kang, B. Wan, Z. Gao, S. Zhou, H. Chen, and H. Shen, "Research on machine learning forecasting and early warning model for rainfall-induced landslides in Yunnan province," Scientific Reports, vol. 14, no. 1, p. 14049, 10.1038/s41598-024-64679-0, June 2024.

[9]. M. Chand and V. Dutt, "A Deep Learning Integrated Stacked Ensemble Framework for Futuristic Landslide Prediction using Multimodal Sensor Data," Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP), Art. no. 101, 10.1145/3774521.3774622, December 2024.

[10]. Y. Liu, L. Brezzi, Z. Liang, F. Gabrieli, Z. Zhou, and S. Cola, "Image analysis and LSTM methods for forecasting surficial displacements of a landslide triggered by

snowfall and rainfall," Landslides, vol. 22, no. 3, pp. 619–635, 10.1007/s10346-024-02328-3, March 2025.

[11]. R. Rania, A. Dahal, M. Berti, and L. Lombardo, "Transformer, more than meets the eye: A deep learning approach to integrate rainfall time-series in multi-type landslide probability modelling," Geoscience Frontiers, vol. 17, p. 102295 10.1016/j.gsf.2026.102295, February 2026.

[12]. F. Teng, S. S. Ekraminia, A. Zarei, and Y. Li, "Remote sensing-based landslide prediction and risk assessment using a hybrid CNN-LSTM deep learning model," Scientific Reports, vol. 16, no. 1, p. 10687, 10.1038/s41598-026-43927-5, March 2026.

[13]. F. Chen, X. Yuan, K. Zhou, K. Ma, Y. Zhang, J. Zhang, and L. Wu, "Interpretable deep learning prediction of landslide displacement based on graph attention networks and temporal attention mechanisms across selected sites," Discover Applied Sciences, vol. 8, no. 3, p. 120, , February 2026.

[14]. R. A. Burange, H. K. Shinde, and O. Mutyalwar, "Landslide Detection and Mapping Using Deep Learning Across Multi-Source Satellite Data and Geographic Regions," arXiv preprint arXiv:2507.01123, July 2025.

[15]. K. Hacıefendioğlu, N. Varol, V. Toğan, Ü. Bahadır, and M. E. Kartal, "Automatic landslide detection and visualization by using deep ensemble learning method," Neural Computing and Applications, vol. 36, no. 19, pp. 10761–10776, 10.1007/s00521-024-09638-6, March 2024.

[16]. A. Kaushal, A. K. Gupta, and V. K. Sehgal, "A semantic segmentation framework with UNet-pyramid for landslide prediction using remote sensing data," Scientific Reports, vol. 14, no. 1, p. 30071, December 2024.

[17]. L. Pham, C. Le, H. Tang, K. Truong, T. Nguyen, J. Lampert, A. Schindler, M. Boyer, and S. Phan, "RMAU-NET: A Residual-Multihead-Attention U-Net Architecture for Landslide Segmentation and Detection from Remote Sensing Images," arXiv preprint arXiv:2507.11143, July 2025.

[18]. Y. Li, W. Zhu, J. Wu, R. Zhang, X. Xu, and Y. Zhou, "DBSANet: A Dual-Branch Semantic Aggregation Network Integrating CNNs and Transformers for Landslide Detection in Remote Sensing Images," Remote Sensing, vol. 17, no. 5, p. 807, 10.3390/rs17050807, February 2025.

[19]. C. Chen, C. Yu, and S. Cai, "Advancing landslide recognition through multi-dimensional feature fusion and transformer architectures," The Visual Computer, vol. 41, no. 8, pp. 11311–11325, 10.1007/s00371-025-04108-z, August 2025.

[20]. A. Sreekumar, H. Thirugnanam, S. R. Meena, and M. V. Ramesh, "Enhancing landslide detection in Western Ghats of Kerala, India with deep learning and Explainable AI," Scientific Reports, vol. 15, no. 1, p. 45151, 10.1038/s41598-025-33065-9, May 2025.