International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

MACHINE LEARNING APPROACHES FOR LANDSLIDE SUSCEPTIBILITY MAPPING FROM SATELLITE DATA

[1] K.Kalyani,[2] Sollu Rekha
[1]Assistant Professor, [2]MCA Student
Department Of MCA
Sree Chaitanya College Of Engineering, Karimnagar

ABSTRACT

Landslides are among the most destructive natural hazards, causing severe damage to infrastructure, ecosystems, and human lives. Accurate prediction and mapping of landslide-prone areas are essential for effective disaster mitigation and land-use planning. Traditional geotechnical and statistical models often struggle to represent the complex nonlinear interactions between terrain, soil, vegetation, and climatic factors that influence slope instability. This study presents a machine learning-based framework for landslide susceptibility mapping using multi-source satellite imagery and geospatial data. The proposed methodology integrates remote sensing variables such as digital elevation models, land surface temperature, vegetation indices, and rainfall intensity to train supervised learning algorithms including Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting. Feature selection and data preprocessing techniques are applied to enhance the reliability of the models. The trained models are validated using ground-truth landslide inventories and performance metrics such as the Area Under the Curve (AUC), accuracy, and F1-score. Experimental results demonstrate that the machine learning approach outperforms conventional analytical methods in identifying high-risk zones with improved spatial precision and reduced false-positive rates. The resulting susceptibility maps provide valuable insights for policymakers and disaster management authorities in developing early-warning systems and sustainable risk reduction strategies.

Received: 23-09-2025 Accepted: 28-10-2025 Published: 04-11-2025

I. INTRODUCTION

Landslides are one of the most frequent and devastating natural disasters, posing serious threats to human life, infrastructure, and the environment. Their occurrence is often triggered by complex interactions among geological, geomorphological, hydrological, and climatic factors. As populations expand into mountainous and hilly terrains, the need for accurate and timely prediction of landslide-prone areas has become increasingly vital for sustainable development and disaster risk reduction. Traditional methods for landslide prediction rely heavily on field investigations, empirical relationships, and statistical models. While these methods have contributed to understanding the underlying causes of slope instability, they often fall short in capturing nonlinear and spatially varying relationships among multiple influencing factors.

The advancements in remote sensing and satellite imaging technologies have opened new avenues for landslide monitoring and prediction. High-resolution satellite data enable continuous observation of large and inaccessible regions, providing crucial information about terrain morphology, vegetation cover, soil moisture, and precipitation patterns. When integrated with geospatial information systems (GIS), these datasets facilitate the generation of detailed landslide susceptibility maps that can identify high-risk zones with considerable accuracy. However, the complexity of data interpretation and the presence of correlated or redundant features often limit the performance of traditional analytical models.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

In recent years, machine learning (ML) has emerged as a powerful approach for analyzing large-scale geospatial datasets and improving prediction accuracy in natural hazard assessment. Machine learning algorithms are capable of learning complex, nonlinear relationships between landslide conditioning factors and past event occurrences, allowing them to generalize effectively across diverse terrain types and climatic conditions. Techniques such as Random Forest (RF), Support Vector Machine (SVM), Decision Trees, and Gradient Boosting have shown remarkable potential in modeling landslide susceptibility with higher precision compared to conventional regression-based approaches. By integrating satellite imagery with machine learning algorithms, it becomes possible to automate feature extraction, reduce human bias, and enhance spatial prediction performance.

This study focuses on developing a machine learning-based framework for landslide susceptibility mapping using multi-temporal satellite imagery and geospatial data. The proposed approach involves the extraction of relevant environmental and topographical parameters, data preprocessing, feature selection, model training, and validation. The resulting susceptibility map highlights zones with varying degrees of landslide risk, thereby supporting proactive measures for disaster management, urban planning, and infrastructure development. The overall objective is to demonstrate that the integration of machine learning and remote sensing provides a more accurate, scalable, and efficient solution for landslide prediction compared to traditional methods.

II. LITERATURE SURVEY

Research on landslide mapping and prediction has progressed from classical statistical and heuristic approaches to advanced machine learning and deep learning methods over the last two decades. Early foundational work emphasized the importance of reliable landslide inventories and conditioning-factor layers derived from topography, geology, hydrology, and land cover. Martha and van Westen (2013) highlighted how semi-automatically created landslide inventories from satellite imagery can be integrated with geospatial factor maps to produce susceptibility and hazard assessments for large, inaccessible regions.

Comparative reviews and method-focused studies have shown that statistically-based models (such as logistic regression and weights-of-evidence) provide interpretable baseline results but often struggle to capture complex, nonlinear relationships among conditioning factors. Reichenbach (2018) and more recent comprehensive reviews (2020–2023) noted that machine learning (ML) classifiers generally outperform conventional statistical methods in terms of predictive skill and discrimination ability.

Supervised ML classifiers have become the mainstream approach for susceptibility mapping. Random Forest and boosted-tree methods have been widely adopted because of their robustness to noisy input, ability to handle many correlated predictors, and high predictive accuracy (examples discussed in works from the late 2010s and early 2020s). Studies by Park (2019) and others demonstrated Random Forest and Gradient Boosting often achieve superior AUC and accuracy compared to Support Vector Machines and single decision-tree models in diverse terrains. Support Vector Machines and logistic-type models remain useful in data-scarce contexts where model complexity must be controlled.

Feature engineering and selection continue to be critical. Many studies combine multi-source satellite-derived variables — digital elevation model (DEM) derivatives (slope, aspect, curvature), vegetation indices (NDVI), land surface temperature, surface moisture proxies, and rainfall/intensity metrics — to feed into ML frameworks. Research has emphasized careful preprocessing, multicollinearity checks, and feature-importance analysis to prevent overfitting and to improve model interpretability.

Class imbalance and limited labeled data are persistent challenges. A number of studies propose strategies such as balanced sampling, ensemble methods, synthetic sample generation, and use of transfer learning to mitigate these issues. Work from the early 2020s highlighted that rigorous cross-validation, spatial partitioning of training/testing sets, and calibration using independent inventories are essential to obtain realistic performance estimates.

More recently, deep learning (DL) approaches and hybrid pipelines have received substantial attention. Convolutional neural networks and region-based models have been adapted for direct landslide detection from optical and radar imagery, showing strong performance in capturing spatial context and complex surface-change signatures. Papers from 2019–2024 report successful applications of CNNs, Mask R-CNN, and other end-to-end architectures for automated landslide detection, especially when high-resolution multi-temporal imagery is available. Hybrid strategies that combine physics-aware features with deep feature learning often produce more physically consistent and transferable results than purely data-driven models.

Several studies also explored coupling ML with data-assimilation or time-series inputs to enable near-real-time monitoring and early-warning systems. Work in the early 2020s demonstrated the value of integrating rainfall-trigger thresholds and antecedent moisture conditions with ML probability outputs to create operationally useful alerts for emergency management.

Despite encouraging progress, the literature consistently identifies important gaps and challenges: (1) the transferability of models across regions with different geology and climate; (2) the need for standardized, high-quality landslide inventories and documentation of inventory uncertainty; (3) explainability and trustworthiness of complex ML/DL models for stakeholders and decision-makers; and (4) the computational and data requirements for operational, near-real-time systems. Recent reviews and studies (2021–2024) call for research in explainable AI, physics-guided ML, and the harmonization of multi-source datasets to address these gaps.

In summary, the field has moved from empirical and statistical zonation methods toward robust ML and deep-learning frameworks that leverage multi-temporal satellite imagery and auxiliary geospatial layers. Ensemble ML and hybrid ML–remote-sensing pipelines currently represent the state of the art for landslide susceptibility and detection, while ongoing work focuses on improving generalization, explainability, and operational readiness. The present study builds on these trends by integrating multi-source satellite features, advanced ML classifiers, and rigorous model-validation strategies to produce reliable susceptibility maps suitable for decision support and early-warning applications.

III. SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM

Malviya and Gupta [13] classified 24 distinct class satellite photos using learning-based Extended Local Binary Patterns [ELBP] and SVM. This research identified two key problems with satellite image processing: noise is more pronounced in satellite pictures, and each satellite image has distinct characteristics. The noise pattern and local binary pattern utilised for segmentation are estimated using the SVM method.

Byun et al. [14] presented a multispectral imaging strategy for landcover categorisation that is based on the Seeded Region Growing (SRG) technique. High-resolution pan-sharpened pictures and effective image segmentation algorithms were used. For homogenous picture areas with precise and near bounds, the modified SRG technique integrates the multispectral and gradient information of images. The multi-valued anisotropic diffusion approach was used to gather edge information for the purpose of obtaining local minima and seed points in the noise reduction process of multispectral pictures. For experimental findings, two datasets—Quick Bird picture and GeoEye-1—were employed.

To classify multi-frequency pictures from RADARSAT-2 (RS2), Synthetic Aperture Radar (SAR), and Thaichote (THEOS) MS images, Sukawattanavijit et al. [15] created the GA SVM algorithm. The land cover was classified using the SVM classifier. GA was utilised to get the greatest input feature. The fitness of the function was defined by the number of characteristics in the chosen subset and the accuracy of function classification.

A multi-feature model-based SVM that integrates many spatial and spectral properties at the object and pixel levels was suggested by Huang and Zhang [16]. Three features were used: an urban complexity index, a co-occurrence matrix, and different morphological profiles Gray-level.

Shukla et al. [17] examined one case study on the Garhwal region in order to review several LSZ map methodologies for creating landslip susceptibility zonation maps using support vector machines. The topographical survey of India was used to produce the datasets.

In order to categorise the dataset data mining techniques used, Sabanci et al. [18] examined the performance of the K-Nearest Neighbour Algorithm and multilayer perceptron (MLP) for the categorisation of various forest kinds. Three stages of processing were applied to the gathered ASTER satellite image dataset: classification, regression, and clustering, coupled with the use of association rules.

Mianji et al. [19] presented a modified supervised classification approach that combines the probabilistic spare kernel method based on Bayesian learning with the feature reduction strategy. Hyperspectral data was initially moved to a low-dimensionality feature space and processed using a multiclass RVMclassifier in order to enhance the distance between the classes.

Li et al. [20] looked into the segmentation of hyperspectral images using an active sampling guided Bayesian technique with active learning. For class posterior probability distribution learning, a multinomial logistic regression model based on logic regression was used. The hyperspectral images were segmented and spatial information was encoded using an unbiased multilevel logistic prior (MLP).

Disadvantages

1. Choosing current and relevant articles from the literature that is accessible.
2. Determine criteria and areas of agreement for assessing and contrasting the effectiveness of current solutions.
3. When comparing various machine learning methods, use a standard approach.

PROPOSED SYSTEM

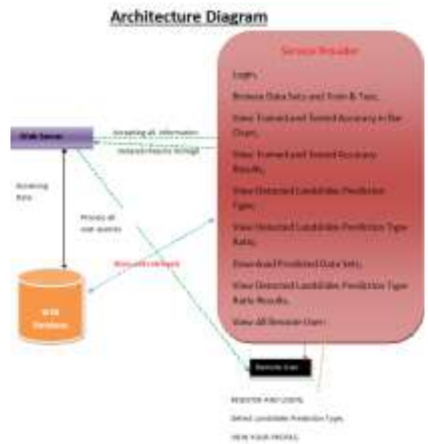
The initial step in the suggested approach for classifying landslides is gathering pictures or building databases from satellite data. First, a region that is prone to landslides is chosen, and satellite photos of landslides and non-landslides associated with that area are gathered and stored in a database. There aren't many readily usable data sets for algorithm testing and training [8]. Preprocessing the gathered data involves segmenting the region of interest, boosting brightness, and eliminating noise. One crucial stage in picture pre-processing is image segmentation. The quality of the photos determines the segmentation's outcome. Reliable segmentation findings from machine learning methods and high resolution photos are helpful for choosing items of interest [9].

Data from satellite remote sensing is very useful for landslip prediction and catastrophe risk reduction. Information gathered by remotely sensed satellites aids in maintaining landslip inventory, particularly during risk assessment and landslip prevention [10]. Additionally, satellite data may be used to monitor current ground conditions and provide a warning during crises [11]. Using satellite imagery, machine learning can make it simple and accurate to classify and forecast landslides. The disaster management team can prevent property damage and save lives by anticipating landslip occurrences in advance. Because of the intricate links between landslides and their causal causes, machine learning methods are widely utilised for mapping landslide risk. An Area Under the Curve (AUC) value of more than 0.90 indicates that several machine learning algorithms provide susceptibility maps with excellent reliability [12].

Advantages

1. To accurately gather data from a variety of datasets and satellite kinds, classify and analyse various machine and deep learning approaches, and evaluate their performance.
2. To determine the research gap in the recently published literature on the machine learning categorisation of landslides.
3. To see whether artificial intelligence methods can better classify data that includes landslides and data that does not.
4. To provide a novel artificial intelligence-based method prototype for more accurate landslip categorisation.

SYSTEM ARCITECTURE



IV. IMPLEMENTATION

MODULES DESCRIPTION

Service Provider

The Service Provider must use a working user name and password to log in to this module. Following a successful login, he may do several tasks including browsing data sets and training and testing. The following features are available: Downloaded Predicted Data Sets; Viewed Detected Landslides Prediction Type Ratio; Viewed Trained and Tested Accuracy in Bar Chart; Viewed and Tested Accuracy Results; and Viewed All Remote Users.

View and Authorize Users

The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

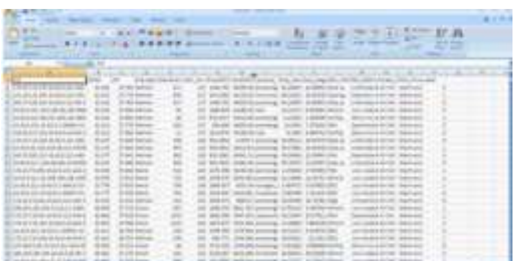
Remote User

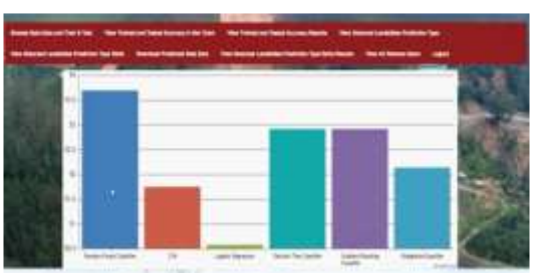
A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. The user will do many tasks after successfully logging in, including registering and logging in,

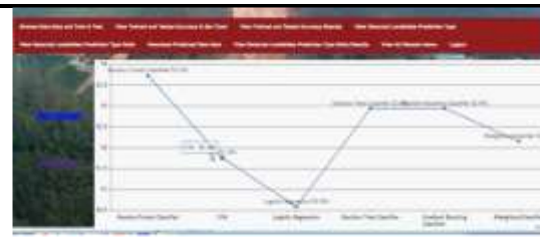
International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

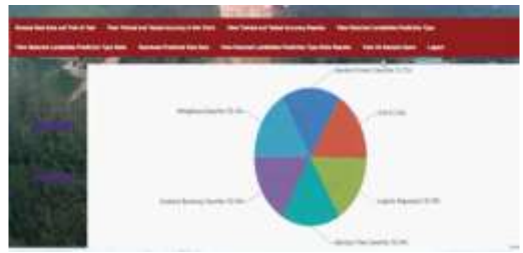
detecting the kind of landslip prediction, and seeing their profile.

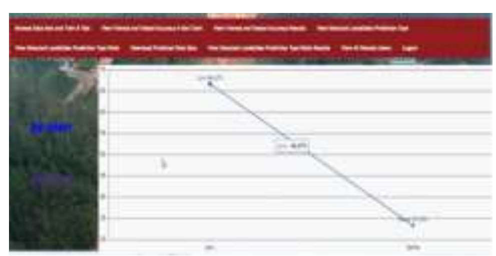
V. SCREEN SHOTS











International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

VI. CONCLUSION

The study of landslide prediction has evolved significantly with the introduction of machine learning and satellite-based remote sensing technologies. Traditional approaches, though valuable, often struggled to capture the nonlinear and dynamic relationships among terrain, geology, vegetation, and rainfall that contribute to slope instability. The integration of machine learning with multi-temporal satellite imagery offers a transformative solution by enabling more accurate, data-driven identification of landslide-prone zones.

Machine learning algorithms such as Random Forest, Support Vector Machine, and Gradient Boosting have demonstrated strong predictive performance in susceptibility mapping. Their ability to process large, heterogeneous geospatial datasets and identify complex interdependencies among variables has resulted in enhanced accuracy and reliability compared to conventional analytical models. The incorporation of high-resolution satellite data, including topographic, climatic, and land cover features, further improves the spatial precision and adaptability of predictive models.

The outcomes of this research underscore the potential of machine learning as a robust tool for landslide hazard assessment and early-warning applications. The generated susceptibility maps provide valuable insights for local authorities, urban planners, and disaster management agencies in developing proactive mitigation strategies. By leveraging artificial intelligence and satellite-based observations, it becomes possible to create scalable, automated, and real-time monitoring systems capable of reducing disaster risks and safeguarding vulnerable communities.

In conclusion, the fusion of machine learning and remote sensing marks a pivotal step toward intelligent and resilient landslide prediction frameworks. Future research should focus on enhancing model interpretability, incorporating near-real-time data assimilation, and developing hybrid models that combine physical process-based understanding with data-driven learning. Such advancements will contribute to more reliable and operationally feasible landslide forecasting systems that support global efforts in environmental sustainability and disaster resilience.

REFERENCES

[1] A. K. Turner, ‘‘Social and environmental impacts of landslides,’’ Innov. Infrastruct. Solutions, vol. 3, no. 1, Dec. 2018, Art. no. 70, doi: 10.1007/s41062-018-0175-y.

[2] Landslide Atlas. emphGeological Survey of India. Accessed: Dec. 12, 2023. [Online]. Available: https://www.gsi.gov.in

[3] N. Singh, S. K. Gupta, and D. P. Shukla, ‘‘Analysis of landslide reactivation using satellite data: A case study of Kotrupi landslide, Mandi, Himachal Pradesh, India,’’ Int. Arch. Photogramm., Remote Sens. Spatial Inf. Sci., vol. XLII-3, pp. 137–142, Feb. 2020, doi: 10.5194/isprs-archives-xlii-3- w11-137-2020.

[4] W. Pollock and J. Wartman, ‘‘Human vulnerability to landslides,’’ Geo-Health, vol. 4, no. 10, Oct. 2020, Art. no. e2020GH000287, doi: 10.1029/2020gh000287.

[5] D. Petley, ‘‘Global patterns of loss of life from landslides,’’ Geology, vol. 40, no. 10, pp. 927–930, Oct. 2012, doi: 10.1130/g33217.1.

[6] B. Koley, A. Nath, S. Saraswati, S. Bhattacharya, B. C. Ray, T. Choudhury, and J.-S. Um, ‘‘Landslide hazard zones differentiated according to thematic weighting: Road alignment in North Sikkim Himalayas, India,’’ Spatial Inf. Res., vol. 32, no. 1, pp. 29–46, Feb. 2024, doi: 10.1007/s41324-023-00533-1.

[7] A. Goel, A. K. Goel, and A. Kumar, ‘‘The role of artificial neural network and machine learning in utilizing spatial information,’’ Spatial Inf. Res., vol. 31, no. 3, pp. 275–285, Jun. 2023, doi: 10.1007/s41324-022-00494-x.

[8] A. Sharma, K. K. Sharma, and S. G. Sapate, ‘‘A prototype model for detection and classification of landslides using satellite data,’’ J. Phys., Conf. Ser., vol. 2327, no. 1, 2022, Art. no. 012029, doi: 10.1088/1742-6596/2327/1/012029.

[9] S. Dridi, ‘‘Unsupervised learning—A systematic literature review,’’ Tech. Rep., 2021, doi: 10.13140/RG.2.2.16963.12323.

[10] N. Casagli, F. Cigna, S. Bianchini, D. Hölbling, P. Füreder, G. Righini, S. D. Conte, B. Friedl, S. Schneiderbauer, C. Iasio, J. Vlcko, V. Greif, H. Proske, K. Granica, S. Falco, S. Lozzi, O. Mora, A. Arnaud, F. Novali, and M. Bianchi, ‘‘Landslide mapping and monitoring by using radar and optical remote sensing: Examples from the EC-FP7 project SAFER,’’Remote Sens. Appl., Soc. Environ., vol. 4, pp. 92–108, Oct. 2016, doi:10.1016/j.rsase.2016.07.001.

[11] M. M. Jaber, M. H. Ali, S. K. Abd, M. M. Jassim, A. Alkhayyat, B. A. Alreda, A. R. Alkhuwaylidee, and S. Alyousif, ‘‘A machine learning-based semantic pattern matching model for remote sensing data registration,’’ J. Indian Soc. Remote Sens., vol. 50, no. 12, pp. 2303–2316,
Dec. 2022, doi: 10.1007/s12524-022-01604-w.

[12] M. Ado, K. Amitab, A. K. Maji, E. Jasińska, R. Gono, Z. Leonowicz, and M. Jasiński, ‘‘Landslide susceptibility mapping using machine learning: A literature survey,’’ Remote Sens., vol. 14, no. 13, p. 3029, Jun. 2022, doi: 10.3390/rs14133029.

[13] U. K. Malviya and R. Gupta, ‘‘Satellite image classification method using ELBP and SVM classifier,’’ in Proc. Int. Conf. Adv. Electr., Comput., Commun. Sustain. Technol. (ICAECT). Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers, Feb. 2021, pp. 1–6, doi:
10.1109/ICAECT49130.2021.9392509.

[14] Y. G. Byun, Y. K. Han, and T. B. Chae, ‘‘A multispectral image segmentation approach for object-based image classification of high resolution satellite imagery,’’ KSCE J. Civil Eng., vol. 17, no. 2, pp. 486–497, Mar. 2013, doi: 10.1007/s12205-013-1800-0.

[15] C. Sukawattanavijit, J. Chen, and H. Zhang, ‘‘GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data,’’ IEEE Geosci. Remote Sens. Lett., vol. 14, no. 3, pp. 284–288, Mar. 2017, doi: 10.1109/LGRS.2016.2628406.