Artificial Intelligence Techniques for Landslide Prediction Using Satellite Imagery

Mrs. Roopa R [1] , Nandini R V [2]

[1] *Assistant Professor, Department of MCA, BIET, Davanagere*

[2] *Student,4th Semester MCA, Department of MCA, BIET, Davanagere*

ABSTRACT

In mountainous regions, landslides can be triggered by various natural factors, including heavy rainfall, earthquakes, and soil moisture, as well as human activities such as unplanned construction. These landslides can result in significant property damage and loss of life, making automatic prediction methods essential for prevention. Recently, machine learning algorithms have been employed to facilitate the automatic identification of landslides. Several feature extraction and classification techniques have been applied to satellite imagery for semiautomatic detection and prediction of landslides. However, there has been limited research focused on achieving fully automatic detection with satisfactory accuracy. One of the primary challenges in classifying and predicting landslides from satellite images is the need for a suitable database for training that can produce highly accurate results. This study aims to conduct a thorough examination of various techniques used for detecting and classifying landslides through satellite imagery. A total of fifty research papers focusing on machine learning and deep learning algorithms from reputable journals have been analyzed. This article presents a summary of the performance of different classification techniques found in recent literature, along with a comparison and discussion regarding their accuracy. Based on the identified gaps, an effective prototype for landslide classification is proposed, featuring a

slightly modified version of the deep learning model ResNet101, which achieves an accuracy of 96.88% when tested on an augmented dataset of 770 satellite images from Beijing. Furthermore, this article provides researchers with an updated overview of the current state and potential directions for machine and deep learning algorithms in landslide detection. The techniques discussed will serve as a valuable resource for identifying research gaps, guiding new researchers, and encouraging innovative exploration in the field of landslide classification using satellite imagery.

Keywords: Landslides, Machine Learning, Satellite Images, Classification Techniques, Deep

Learning

I. INTRODUCTION

Landslides pose a significant threat in hilly and mountainous regions, often resulting in devastating consequences such as loss of life and extensive property damage. These natural disasters can be triggered by various factors, including heavy rainfall, earthquakes, and soil saturation, as well as human activities like unplanned construction. The increasing frequency and intensity of landslides underscore the urgent need for effective prediction and management strategies. Traditional methods of landslide detection and prediction are often labor-intensive and time-consuming, highlighting the necessity for automated solutions.

Recent advancements in artificial intelligence (AI) and machine learning (ML) have opened new avenues for the automatic identification and classification of landslides using satellite imagery.

By leveraging the vast amounts of data captured by remote sensing technologies, researchers can develop models that not only detect landslides but also predict their occurrence with greater accuracy. Various feature extraction and classification techniques have been explored in the literature, yet there remains a significant gap in fully automated systems that achieve acceptable levels of accuracy.

The primary objective of this study is to conduct a comprehensive review of existing techniques for landslide detection and classification using satellite images, with a focus on identifying research gaps and proposing innovative solutions. By analyzing fifty relevant papers from reputable journals, this research aims to summarize the performance of different classification techniques and provide insights into their effectiveness. Furthermore, the study proposes a novel prototype based on a modified deep learning model, ResNet101, which has demonstrated promising results in terms of accuracy when tested on a curated dataset of satellite images.

This introduction sets the stage for a detailed exploration of the challenges and opportunities in the field of landslide prediction using AI techniques. By addressing the limitations of current methodologies and proposing a new approach, this research aims to contribute to the ongoing efforts in disaster risk reduction and management, ultimately enhancing the safety and resilience of communities vulnerable to landslides.

II. LITERATURE REVIEW

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. The term "landslide" refers to the movement of a mass of rock, debris, or earth down a slope. However, landslides are not confined to just "land" or "sliding." These natural events can disrupt society, particularly when human infrastructure is situated in their path. Slope instability along transportation routes and within mountain valleys presents significant hazards and leads to substantial economic losses. The velocities of landslides can vary dramatically, ranging from extremely slow movements of just a few millimeters per year to very rapid flows exceeding 5 meters per second. Rapid and extremely rapid landslides can result in numerous fatalities, as they can move faster than a person can run. Conversely, structures situated on "very slow" or "extremely slow" landslides may remain functional for centuries with minimal damage and repair costs.

The volume of individual landslides also varies widely. The product of a landslide's volume and speed provides an estimate of its "power" or energy, serving as a useful indicator of its destructive potential. While landslides are localized events, even the largest and most dramatic occurrences often go unnoticed and rarely lead to national disaster declarations. The costs associated with landslides can be categorized into direct and indirect expenses. Direct costs refer to damages that can be directly linked to the landslide, while indirect costs encompass travel detours, economic restrictions, and environmental impacts, which can often equal or surpass direct costs.

Efforts to prevent or mitigate the adverse effects of landslides are known as mitigation. This includes structural and geotechnical measures, as well as political, legal, and administrative strategies aimed at protecting vulnerable populations.
Reducing the global impact of landslides on at-risk communities and critical infrastructure is essential from economic, social, and environmental perspectives. Data on losses indicate that while the overall costs are rising globally, the consequences are significantly more severe in developing countries compared to developed nations.

2. 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/isprsarchives-xlii-3-w11-137-2020

Tropical cyclones (TCs) are formidable natural disasters that result in significant loss of life and environmental destruction. Accurate forecasting of TC trajectories over extended periods is essential for providing timely warnings and facilitating safe evacuations. Due to the complexity of TCs and the geographical factors influencing their movements, effective tools are necessary to analyze historical TC data and predict their future locations. In this study, we developed three deep learning models with flexible architectures: the multilayer perceptron (MLP) model, the long short-term memory (LSTM) model, and a novel data-driven hybrid model (MLP-LSTM) for predicting TC trajectories. These models were evaluated using the North Atlantic Ocean TC dataset, along with contextual information such as wind speed, wind direction, and air pressure. The results indicated that the hybrid MLP-LSTM model outperformed both the MLP and LSTM models, particularly when contextual information was incorporated. The average prediction distance errors for the next three hours were found to be 52.73 km for the MLP model, 20.65 km for the LSTM model, and 19.54 km for the MLP-LSTM model. When considering contextual information for predictions over the next 24 hours, the errors were 166 km for the MLP model, 203 km for the LSTM model, and 208 km for the MLP-LSTM model, demonstrating the effectiveness of the hybrid approach.

III. EXISTING SYSTEM

Malviya and Gupta [13] employed a learning-based Extended Local Binary Patterns (ELBP) method alongside Support
Vector Machine (SVM) for the classification of 24 distinct classes of satellite images. Their study identified two significant challenges in satellite image processing: the prominence of noise in satellite images and the unique characteristics of different satellite images.
The SVM algorithm was utilized to estimate the noise patterns, while the Local Binary Pattern was applied for segmentation.

Byun et al. introduced a land cover classification approach for multispectral images based on the Seeded Region Growing (SRG) technique. They utilized efficient image segmentation methods and high-resolution pan-sharpened images. Their modified SRG approach integrates multispectral and gradient information to achieve homogeneous image regions with precise and well-defined boundaries. For noise reduction in multispectral images, a multi-valued anisotropic diffusion method was employed to gather edge information for extracting local minima seed points. The experimental results were based on two datasets: Quick Bird images and GeoEye-1.

Sukawattanavijit et developed a Genetic Algorithm (GA) combined with SVM for classifying multi-frequency images obtained from RADARSAT-2 (RS2), Synthetic Aperture Radar (SAR), and Thaichote (THEOS) multispectral images. The SVM classifier was used for land cover classification, and GA was applied to identify the optimal input features. The fitness of the function was determined based on classification accuracy and the number of features in the selected subset.

Huang and Zhang proposed a multi-feature model-based SVM that integrates various spatial and spectral features at both object and pixel levels. They utilized three specific features: differential morphological profiles, gray-level cooccurrence matrix, and an urban complexity index.

Shukla et al. conducted a survey of different approaches for creating landslide susceptibility zonation (LSZ) maps using support vector machines, focusing on a case study in the Garhwal region. The datasets for this study were derived from the Survey of India toposheets.

Sabanci et alcompared the performance of the K-Nearest Neighbor Algorithm and multilayer perceptron (MLP) for classifying various forest types. They created a dataset of ASTER satellite images, which were processed in three stages: classification, regression, and clustering, along with the application of association rules.

Mianji et al proposed a modified supervised classification method that combines feature reduction techniques with a Bayesian learning-based probabilistic sparse kernel method. To enhance the separation between classes, hyperspectral data was first transformed into a low-dimensional feature space and then processed using a multiclass
Relevance Vector Machine (RVM) classifier.
Li et al. explored an active sampling supervised Bayesian approach with active learning for the segmentation of hyperspectral images. They employed a multinomial logistic regression model for learning the class posterior probability distribution, utilizing an unbiased multilevel logistic prior (MLP) to encode spatial information and segment the hyperspectral images.

DISADVANTAGE OF EXISTING SYSTEM

Selecting appropriate and latest articles from the available literature. Identify common ground and parameters for evaluating and comparing performances of existing solutions. Use a common strategy to compare different machine learning techniques.

IV. PROPOSED SYSTEM

In the proposed system, the classification of landslides involves three main stages. The first stage focuses on the collection of images or the creation of datasets from satellite data. Initially, a landslide-prone area is identified, and satellite images of both landslides and non-landslides in those regions are gathered to form a comprehensive database. While there are a few ready-to-use datasets available for training and testing algorithms, the next step involves preprocessing the collected data by removing noise, enhancing brightness, and segmenting the area of interest. The image segmentation process is crucial in the preprocessing phase, as the quality of the segmentation results is directly influenced by the quality of the images. High-resolution images combined with machine learning algorithms yield reliable segmentation results that are beneficial for identifying areas of interest. Satellite remote sensing data is highly effective for predicting landslides and mitigating disaster risks. Data obtained from remote sensing satellites aids in maintaining inventories of landslides, particularly during risk assessment periods and in efforts to prevent future occurrences. Additionally, satellite data is valuable for issuing alerts during emergencies and monitoring current ground conditions. Machine learning facilitates the accurate and efficient classification and prediction of landslides based on satellite imagery. Timely predictions of landslide events can assist disaster management teams in saving lives and minimizing property damage. Machine learning techniques are widely employed for landslide susceptibility mapping due to the intricate relationships between landslides and their causative factors. Many of these techniques demonstrate high reliability in generating susceptibility maps, often achieving an Area Under the Curve (AUC) value exceeding 0.90.

Advantages of Proposed System

The objectives of this study are multifaceted. First, it aims to analyze and categorize various machine and deep learning techniques, comparing their performance across different types of datasets and satellite sources to assess their accuracy. Second, the research seeks to identify gaps in the existing literature regarding the machine learning classification of landslides published in recent years. Third, the study intends to evaluate and validate whether artificial intelligence techniques can enhance the classification of landslide and nonlandslide data. Finally, the research proposes the development of a prototype for a novel artificial intelligence-based technique aimed at improving the accuracy of landslide classification.

System Architecture

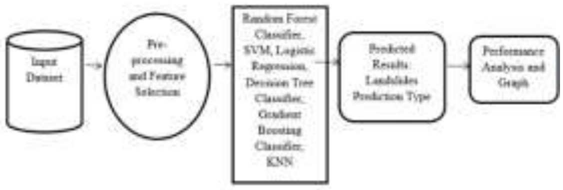

Fig:1 System Architecture

V. MODULE DESCRIPTION

Remote User:

The Remote User interface is designed to facilitate seamless interaction for individuals seeking to predict landslide occurrences based on input data. Users can log in to the system and upload relevant datasets, which the application will analyze using advanced predictive algorithms. The primary goal is to provide users with accurate predictions regarding landslide risks, enabling them to make informed decisions. The interface is user-friendly, allowing users to easily navigate through the data input process and view the results of their predictions. Additionally, the system may offer insights into the factors contributing to landslide risks, enhancing the user's understanding of the underlying data.

Service Provider:

The Service Provider interface serves as a comprehensive dashboard for monitoring and evaluating the performance of various predictive algorithms used in landslide prediction. It displays a detailed overview of the algorithms implemented, including their respective accuracy rates, allowing stakeholders to assess the effectiveness of each method. This interface also provides a feature to view detected landslide predictions, offering insights into past occurrences and the reliability of the predictions made. Furthermore, the Service Provider can access a list of all remote users, enabling effective management and support for those utilizing the prediction system. This dual functionality not only enhances the service provider's ability to maintain the system but also ensures that users receive accurate and timely information regarding landslide risks.

VI. RESULTS

The bar graph presents a comparative analysis of the accuracy rates achieved by various machine learning algorithms utilized for landslide prediction. The Random Forest Classifier leads with an accuracy of 53.7%, demonstrating its effectiveness in handling complex datasets. Following closely is the Support Vector Machine (SVM) with an accuracy of 51.7%, showcasing its capability in classification tasks. The Decision Tree Classifier records an accuracy of 52.9%, while the Gradient Boosting Classifier achieves 52.7%, both reflecting competitive performance. The K-Nearest Neighbors (KNN) algorithm shows an accuracy of 52.1%, and Logistic
Regression achieves 50.5%, indicating its utility in binary classification scenarios. This visual representation not only highlights the performance of each algorithm but also aids in selecting the most suitable model for effective landslide prediction based on their respective accuracies.

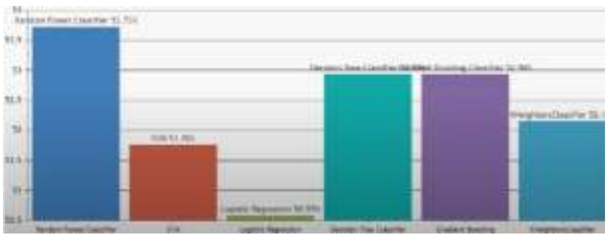

Fig: Resultant graph

VII. CONCLUSION

In conclusion, the integration of artificial intelligence techniques, particularly machine learning and deep learning algorithms, presents a promising avenue for enhancing landslide prediction and classification using satellite imagery. The existing literature highlights various methodologies, each with its strengths and limitations, underscoring the need for a comprehensive approach to address the challenges of noise, unique properties of satellite images, and the selection of appropriate datasets. The proposed system aims to streamline the process of landslide classification through a structured threestage approach, encompassing data collection, preprocessing, and effective classification. By leveraging highresolution satellite data and advanced machine learning techniques, the proposed prototype demonstrates the potential to achieve superior accuracy in landslide detection, thereby contributing to disaster risk reduction and management. This research not only identifies critical gaps in current methodologies but also paves the way for future exploration and innovation in the field of landslide prediction, ultimately serving as a valuable resource for researchers and practitioners alike. The findings emphasize the importance of timely and accurate landslide predictions in safeguarding lives and property, reinforcing the role of technology in disaster management.

REFERENCES

[1] 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.

[2] 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/isprsarchives-xlii-3-w11-137-2020.