Predicting Landslide Using Machine Learning Techniques

Mehul Patel[1*], Mittal Chavda[2], Rajesh Patel[3], Ankur Goswami[4], Jayesh Mevada[5]

Assistant Professor, Department of CE/IT, Sankalchand Patel University, Visnagar, India[1, 2, 4, 5]
Associate Professor, Department of CE/IT, Sankalchand Patel University, Visnagar, India[3]

mspatelit_spce@spu.ac.in[1*]; mittal2010chavada@gmail.com[2]; drrppatelce_spce@spu.ac.in[3]; ajgoswamice_spce@spu.ac.in[4]; jmmevadace_spce@spcevng.ac.in[5]

Abstract: In mountainous areas prone to landslides, it's crucial to map out where these hazardous events are likely to occur to mitigate risks effectively. This study focuses employing an integrated approach to assess landslide susceptibility using Random Forest (RF), Stacking, Vote, AdaBoostM1, and Bagging. 13 factors influencing landslide occurrence are identified for modeling purposes. To evaluate and compare the models' performance, multiple statistical methods are employed. The analysis highlights the effectiveness of employing machine learning models, Random Forest (RF), Stacking, Bagging, and Vote methods. The results demonstrate the efficiency of the models in accurately predicting landslide susceptibility. The study suggests that similar hybrid models can be effectively utilized in other sensitive regions with comparable geo-environmental conditions for landslide susceptibility studies. By integrating various techniques and leveraging ensemble algorithms, these models offer improved accuracy and reliability in assessing landslide hazards. This comprehensive approach provides valuable insights for disaster management and risk reduction efforts in landslide- prone areas worldwide.

Keywords: Landslide, Machine Learning, Remote Sensing

I. INTRODUCTION

Landslides are natural disasters that can cause a lot of damage to homes, roads, and other buildings, and they can even hurt or kill people. To try and stop this from happening, we're using really smart technology, like computers that learn things, along with pictures from satellites, to figure out where landslides might happen. But the way we were doing it before was kind of complicated and took a long time, which isn't good when we need to act fast in an emergency. Addressing the challenge of landslide detection involves bringing together different areas of expertise, like geology, engineering, remote sensing, meteorology, and data science. The goal is to effectively identify, assess, and reduce the risks associated with landslides. This problem is complex and requires a collaborative effort to tackle various aspects of landslide management. The tasks range from creating systems that can detect landslides in real-time and provide early warnings to conducting thorough assessments of the risks involved. Additionally, there's a focus on developing long-term strategies to mitigate these risks, and this is where machine learning algorithms come into play. These algorithms help us make sense of large and diverse datasets, such as satellite images and geological information.

A landslide happens when a lot of rocks or soil suddenly move down a hill, causing a lot of damage to the environment and sometimes harming people. This is a big problem in many parts of the world and can cause huge losses. Scientists use images taken from far away to help detect, map, predict, and understand landslides better. By studying these images and using special computer programs, like RF [1, 15], they can predict landslides more accurately. In mountainous areas prone to landslides, it's crucial to map where these dangerous events are likely to occur to prevent and mitigate risks. In India, landslides often occur in the Himalayas, Western Ghats, and Nilgiris Mountains [11]. If we can predict when a landslide might happen, we can try to prevent it and keep people safe. This study focuses on using these advanced techniques to predict landslides in western India more effectively.

This paper is organized as follows: Segment 2 presents the Literature Survey. Segment 3 depicts the methodology implemented in this study. Section 4 shows the strategy of implementation. Section 5 portrays the results of the landslide prediction model. Section 6 depicts the conclusion on the effect of the significant landslide prediction, and Segment 7 portrays the future work with other ML models to further develop the landslide prediction for further study.

II. LITURATURE SURVEY

The study [1] focuses on assessing the efficacy of three machine learning algorithms—random forest (RF), decision tree (DT), and support vector machine (SVM)—for landslide susceptibility mapping along the Kamyaran-Sarvabad road in Iran's Kurdistan province, a region prone to landslides. Fourteen factors, such as slope, elevation, land use, and geological features, were incorporated into the models based on field survey data identifying 64 landslide locations. Seventy percent of these locations were randomly selected for training the MLAs, while the remainder was reserved for validation. By analyzing these factors, the study aims to predict areas at higher risk of landslides, aiding in infrastructure planning and disaster management efforts. The research underscores the importance of leveraging advanced technologies like machine learning to mitigate the impact of natural hazards on critical infrastructure and human lives, particularly in geographically vulnerable regions like mountainous areas. Ultimately, the findings contribute to enhancing the understanding and preparedness for landslide events in the study area and beyond.

The absence of comprehensive global data inventories poses a significant obstacle to effectively modeling and responding to the hazards posed by landslides, which often result in significant loss of life and property damage. Addressing this challenge involves exploring innovative solutions [2], such as leveraging citizen science for active participation. Additionally, the emergence of social media as a non-traditional data source has prompted researchers to utilize it in disaster response and management efforts. Building on this trend, author propose harnessing social media data for automatic extraction of landslide-related information using artificial intelligence techniques. An approach involves developing a cutting-edge computer vision model capable of real-time detection of landslides in social media image streams. This process begins with the creation of a large dataset of landslide images labeled by experts, emphasizing a data-centric perspective. Subsequently, extensive model training experiments are conducted to optimize performance and accuracy, ultimately enhancing our ability to detect and respond to landslide events more effectively.

The absence of comprehensive global data inventories poses a significant obstacle to effectively modeling and responding to the hazards posed by landslides, which often result in significant loss of life and property damage. Addressing this challenge [3] involves exploring innovative solutions, such as leveraging citizen science for active participation. Additionally, the emergence of social media as a non-traditional data source has prompted researchers to utilize it in disaster response and management efforts. Building on this trend, Author propose harnessing social media data for automatic extraction of landslide-related information using artificial intelligence techniques. An approach involves developing a cutting-edge computer vision model capable of real-time detection of landslides in social media image streams. This process begins with the creation of a large dataset of landslide images labeled by experts, emphasizing a data-centric perspective. Subsequently, extensive model training experiments are conducted to optimize performance and accuracy, ultimately enhancing our ability to detect and respond to landslide events more effectively.

In research study [4], Author combined two different kinds of computer programs, called CenterNet and ResNet50, and they made some changes to them to make them easier to use and faster. They also added some extra things to help them see landslides better. With this new way, the computer can look at pictures really quickly and figure out where landslides might happen, which helps keep people and buildings safe.

In research study [5], Author shows map where landslides are likely to occur to help reduce the impact of these disasters. This kind of map using a mix of GIS (Geographic Information System) and machine learning algorithms. Author applied testing two different machine learning models, called RF and XGBoost, in two areas known for landslides: Cameron Highland and Penang Island in Malaysia. Author looked at 233 landslides in Cameron Highland and 443 in Penang Island. They gathered information about things like the slope of the land, how close it is to streams and roads, and other factors that might affect landslides. Author used GIS software to create maps with all this information. By combining the data and using machine learning, Author create a map that can predict where landslides might happen in these areas, which can help people be better prepared and keep them safe.

In research study [7], Author uses a combination of traditional machine learning methods (SVM, LR, RF) with CNN (Convolutional Neural Network). The models are trained and tested using spatial data from 3251 historical landslide sites, split into a 70:30 ratio. Initially, 16 factors influencing landslides are considered. Three hybrid models—CNN-SVM, CNN- LR, and CNN-RF—are constructed using this data. Final landslide susceptibility maps (LSMs) are generated using these hybrid models, alongside maps produced using individual ML methods for comparison. Statistical methods are then employed to validate and compare the models' performance. The results demonstrate the effectiveness of combining traditional machine learning techniques with CNN, suggesting that future hybrid approaches could be valuable for landslide susceptibility studies.

In simpler terms, the aim is to use technology and a combination of different fields to not only spot potential landslides quickly but also to come up with effective ways to reduce their impact. By using advanced tools and knowledge from different areas, this approach strives to create a comprehensive framework for detecting and managing landslides, making vulnerable regions safer and more resilient.

III. METHODOLOGY

WEKA is user-friendly software that helps people analyze data and make predictions. It was made by researchers in New Zealand and has many tools for things like organizing data, finding patterns, and making models. You can use it even if you're not great at programming because it has a simple interface. It works with different types of data and lets you see your data in graphs and charts to understand it better. People use WEKA a lot in both schools and businesses for things like figuring out trends in data or making guesses about what might happen in the future. It's free and lots of people work on it, so it's always getting better. Machine learning algorithms such as Support Vector Machines (SVM), Logistic Regression (LR), and Random Forest (RF), Geographic Information System (GIS) for analyzing remotely sensed spatial data, Thematic maps derived from Digital Elevation Models (DEM) for extracting topographical factors and These tools and technologies are utilized for the assessment of landslide susceptibility in the study area.

Looking ahead in the field of predicting landslides using machine learning, the focus should be on making these prediction models better in terms of accuracy, reliability, and how well they can be used in different situations. Right now, we've made progress in using machine learning to predict landslides, but there's still work to be done to make these models more trustworthy and useful in various conditions.

Improving accuracy means making sure the models can understand and predict landslides more precisely. This might involve using more advanced techniques and data to train the models. Robustness is about making sure these models can handle different situations and changes in the environment. We want them to be able to adapt to different landscapes and weather conditions.

Enhancing applicability means making these models useful in many different places and situations. We want the models to work well in different types of terrain and with different kinds of data. Additionally, it would be great if these models could work with less data, making them helpful even in places where data is not easily available.

TABLE I:
FEATURE/ATTRIBUTE

Sr. No.	Feature/ Attribute
01	Landslide
02	Aspect
03	Curvature
04	Earthquake
05	Elevation
06	Flow
07	Lithology
08	NDVI
09	NDWI
10	Plan
11	Precipitation
12	Profile
13	Slope

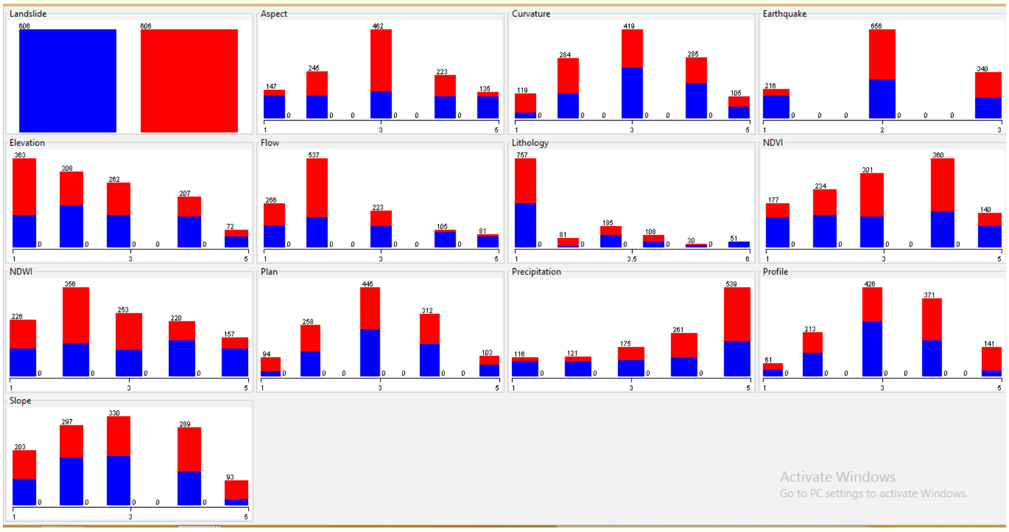

Fig. 1: Dataset Analysis

IV. IMPLEMENTATION

In WEKA, implementing data classification involves several steps:

A. Data Preparation:

Loading dataset into WEKA required formatted correctly and contains the needed attributes for classification and preprocess the data to handle missing values, normalize attributes, or address other data quality issues.

B. Selecting a Classifier:

Choose a classification algorithm from the many options in WEKA. Common classifiers include decision trees, support vector machines, k-nearest neighbors, naive bayes, random forest, stacking, vote, AdaBoostM1, and bagging.

C. Training the Classifier:

Split dataset into training and testing sets. Use the training set to train the chosen classifier on the labeled data, letting it learn patterns and relationships between the input attributes and the class labels.

D. Testing and Evaluation:

Once the classifier is trained, use the testing set to evaluate its performance. Apply the trained classifier to the testing data to predict class labels for each instance. Compare the predicted labels with the actual labels to assess the classifier's accuracy.

E. Performance Metrics:

Evaluate the classifier's performance using various metrics like accuracy, precision, recall, F1- score, and area under the ROC curve (AUC). These metrics give insights into different aspects of the classifier's performance, such as its ability to correctly classify instances from different classes and its robustness to imbalanced data.

F. Cross-Validation:

Perform a cross-validation to get more reliable estimates of the classifier's performance. This involves splitting the data into multiple subsets, training the classifier on one subset, and testing it on the remaining subsets, rotating through all possible combinations.

Achieving high accuracy in classification using WEKA involves selecting an appropriate classifier, optimizing its parameters, and ensuring that the data is well-preprocessed. Additionally, feature selection techniques can be applied to identify the most relevant attributes, which can improve the classifier's performance. By carefully tuning the classifier and evaluating its performance using rigorous testing procedures, you can maximize its accuracy in classifying unseen data instances.

V. RESULT

Our machine learning models was trained on a dataset consisting of 1212 instances with 13 features. After training and evaluation, overall random forest performed well as the following results were obtained:

A. *Accuracy*: The accuracy of the model on the test set was found to be 946. This indicates that the model correctly predicted 78.0528% of the instances in the test set.
B. *Precision*: The precision of the model was 0.781. This means that when the model predicted a positive outcome, it was correct 78.1% of the time.
C. *Recall*: The recall score of the model was 0.781. This shows that the model captured 78.1% of all positive instances in the dataset.
D. *F1-score*: The F1-score, which balances precision and recall, was 0.781. This metric provides a measure of the overall performance of the model, taking into account both false positives and false negatives.
E. *Area under the ROC Curve (AUC)*: The AUC score for the model was 0.8516. This indicates the model's ability to distinguish between positive and negative instances, with a higher score indicating better performance.

TABLE II
PERFORMANCE MATRIX

Model Name	Accuracy	Precision	Recall	F1-Score	AUC
Random Forest	78.05	0.781	0.781	0.780	0.851
Stacking	72.68	0.727	0.727	0.727	0.780
Vote	77.80	0.779	0.778	0.778	0.850
AdaBoostM1	78.05	0.781	0.781	0.781	0.829
Bagging	78.03	0.783	0.783	0.783	0.849

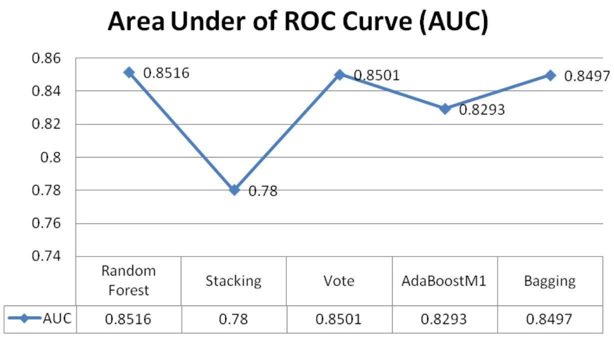

Fig. 2 : Area under of ROC Curve (AUC)

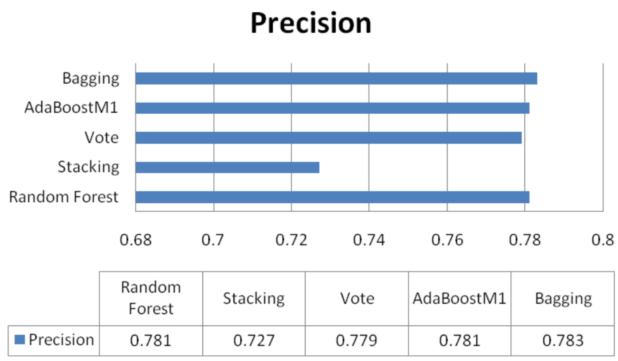

Fig. 3 : Precision

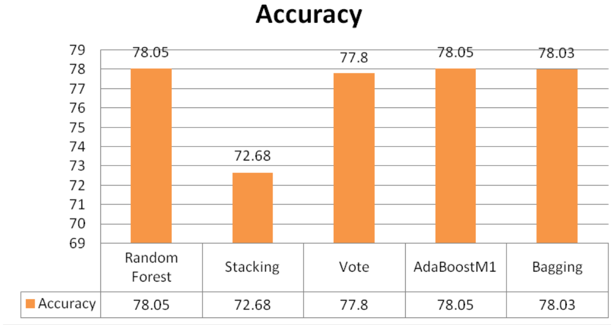

Fig. 4 : Accuracy

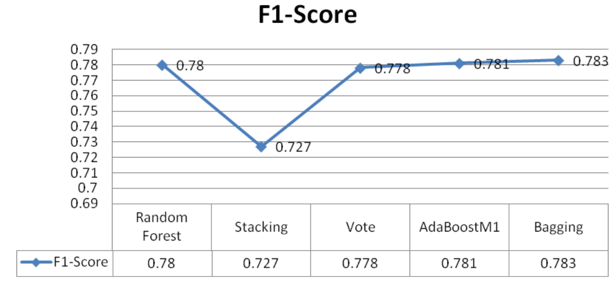

Fig. 5: F1-Score

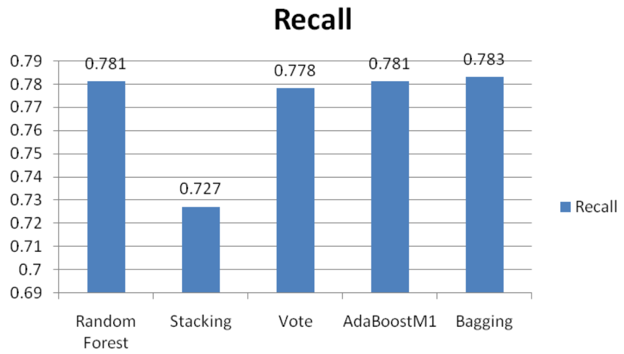

Fig. 6 : Recall

The results of our machine learning experiment suggest that the model performs reasonably well on the given dataset. The high accuracy, F1-score and Area Under of ROC Curve (AUC) indicate that the model is effective at making overall predictions, while the precision and recall scores provide insights into its ability to correctly identify positive instances.

However, further analysis is warranted to understand potential areas for improvement. This could involve examining misclassified instances, exploring feature importance, and experimenting with different algorithms or hyper parameters to optimize performance. Overall, while the model shows promise, ongoing refinement and evaluation are necessary to ensure its reliability and effectiveness in real-world applications.

VI. CONCLUSION

In summary, this study employs dataset classification to predict landslides, yielding promising results. By leveraging machine learning algorithms, we've developed alternate models for identifying potential landslide occurrences. Through feature selection techniques, we've enhanced the model's interpretability and efficiency. Future work involves refining the model with additional datasets and advanced algorithms, as well as validating it across diverse regions. Our efforts represent a significant step in utilizing data-driven approaches for natural disaster management, offering valuable insights to mitigate landslide risks and enhance community safety in vulnerable areas.

VII. FUTURE WORK

Future work entails enhancing model accuracy and feature selection methods. We plan to refine the model's accuracy through various techniques such as adjusting parameters, trying different algorithms, and optimizing training methods. Additionally, we will employ feature selection methods like correlation analysis and recursive feature elimination to identify the most important features, making the model more efficient and understandable. Cross-validation will be used to thoroughly test the model's performance with feature selection, ensuring its effectiveness across different datasets. Furthermore, validation and testing on independent datasets or real-world examples will validate the improved model's performance. Continuous monitoring and refinement will be essential to adapt the model to changing data and requirements, ensuring ongoing improvement in accuracy and usability. Overall, our focus on improving model accuracy through feature selection aims to create a more reliable and effective predictive model for practical applications.

REFERENCES

[1] Shahabi, Himan, et al. "*Landslide susceptibility mapping in a mountainous area using machine learning algorithms.*" Remote Sensing 15.12 (2023): 3112.
[2] Ge, Xuming, et al. "*Lightweight landslide detection network for emergency scenarios.*" Remote Sensing 15.4 (2023): 1085.
[3] Ofli, Ferda, et al. "*Landslide detection in real-time social media image streams.*" Neural Computing and Applications 35.24 (2023): 17809-17819.
[4] Ghorbanzadeh, Omid, et al. "*Landslide detection using deep learning and object-based image analysis.*" Landslides 19.4 (2022): 929-939.
[5] D. Martin and S. S. Chai, "*A study on development of landslide susceptibility map in malaysia landslide prone areas by using geographic information system (GIS) and machine learning,*" in 2022 IEEE 18th International Colloquium on Signal Processing & Applications (CSPA). IEEE, 2022, pp. 386–391.
[6] P. Varangaonkar and S. Rode, "*Research on efficient landslide prediction approaches using machine learning techniques,*" in 2022 5th International Conference on Advances in Science and Technology (ICAST). IEEE, 2022, pp. 64–68
[7] Aslam, Bilal, Adeel Zafar, and Umer Khalil. "*Development of integrated deep learning and machine learning algorithm for the assessment of landslide hazard potential.*" Soft Computing 25.21 (2021): 13493-13512.
[8] G. Danneels, E. Pirard & H. Havenith, "*Automatic landslide detection from remote sensing images using supervised classification methods,*" 2007 IEEE International Geoscience & Remote Sensing Symposium, Barcelona, 2007, pp. 3014-3017
[9] Mendez-Astudillo and M. Mendez-Astudillo, "*A machine learning approach to monitoring the UHI from GNSS data,*" IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–11, 2021.
[10] I. L. Mallika, D. V. Ratnam, S. Raman, and G. Sivavaraprasad, "*A new ionospheric model for single frequency GNSS user applications using klobuchar model driven by auto regressive moving average (SAKARMA) method over indian region,*" IEEE Access, vol. 8, pp. 54 535–54 553, 2020.
[11] A. Siemuri, H. Kuusniemi, M. S. Elmusrati, P. V ̈alisuo, and A. Shamsuzzoha, "*Machine learning utilization in GNSS—use cases, challenges and future applications,*" in 2021 International Conference on *Localization and GNSS (ICL-GNSS). IEEE, 2021, pp. 1–6*
[12] *T. Gravalon, L. Seoane, G. Ramillien, J. Darrozes, and L. Roblou, "Determination of weather-induced short-term sea level variations by*

GNSS reflectometry," Remote Sensing of Environment, vol. 279, p. 113090, 2022

[13] O. K. Isik, I. Petrunin, G. Inalhan, A. Tsourdos, R. V. Moreno, and R. Grech, "*A machine learning based gnss performance prediction for urban air mobility using environment recognition,*" in 2021 IEEE/AIAA 40th Digital Avionics Systems Conference (DASC). IEEE, 2021, pp. 1–5.

[14] Alcantara-Ayala, I.:*"Hazard assessment of rainfall-induced landsliding in Mexico"*, Geomorphology, 61, 19–40, 2004.

[15] Ayalew, L. and Yamagishi, H.: "*The application of GIS-based logistic regression for landslide susceptibility mapping in the KakudaYahiko Mountains*", Central Japan, Geomorphology, 65, 15–31, 2005.

[16] Y. Nakagawa, T. Miyauchi, T. Higashino, and M. Okada, "*Application of random forest to classify weather observation into rainfall using GNSS receiver*," in 2021 IEEE VTS 17th Asia Pacific Wireless Communications Symposium (APWCS). IEEE, 2021, pp. 1–5.

[17] Q. Yan, W. Huang, and C. Moloney, "*Neural networks based sea ice detection and concentration retrieval from GNSS-R delay-doppler maps*," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 10, no. 8, pp. 3789–3798, 2017.

[18] G. Danneels, E. Pirard & H. Havenith, "*Automatic landslide detection from remote sensing images using supervised classification methods*," 2007 IEEE International Geoscience & Remote Sensing Symposium, Barcelona, 2007, pp. 3014-3017.

[19] Blaschke T, "*Object based image analysis for remote sensing*". ISPRS J Photogramm Remote Sens 65:2–16, 2010