IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

Integrated Machine Learning Framework for Flood and Landslide Hazard Assessment

Nameet Vyavahare, Darshan Chitale, Rajkumar Vishwakarma, Devang Nimbalkar
Department of AI&ML
ISBM College of Engineering, Nande, Pune, Maharashtra
vyavaharenameet@gmail.com, rk.1947.2003@gmail.com, nimbalkardevang@gmail.com

Abstract: *Floods and landslides are major natural hazards in India, requiring accurate and timely risk assessment. This paper presents BhoomiSense, an integrated machine learning framework that predicts flood probability, landslide probability, and overall hazard severity through a unified web-based platform. The proposed system employs ensemble models based on XGBoost, LightGBM, and Random Forest, combined with hyperparameter optimization, probability calibration, and domain-specific feature engineering. SHAP analysis is incorporated to improve model interpretability. Experimental results demonstrate the effectiveness of the framework. The flood prediction model achieved an accuracy of 79.33% and a ROC-AUC of 79.57%, while the landslide prediction model achieved an accuracy of 71.00% and a ROC-AUC of 72.41%. The hazard severity model obtained an R^2 score of 77.49% with a mean absolute error of 7.867. The trained models are deployed using a FastAPI backend on Railway Cloud and integrated with a React-based frontend through REST APIs for efficient real-time prediction. The proposed framework demonstrates the potential of ensemble machine learning for accurate and interpretable multi-hazard assessment.*

Keywords: Machine Learning, Flood Prediction, Landslide Susceptibility, Remote Sensing, GIS, Ensemble Learning, DEM, Hydrological Hazards

I. INTRODUCTION

Natural disasters such as floods and landslides continue to pose serious challenges to human life, infrastructure, agriculture, and economic development. In India, the frequency and intensity of these hazards have increased due to changing weather patterns, rapid urbanization, and environmental degradation. Their occurrence is influenced by multiple interconnected factors, including rainfall, topography, soil conditions, vegetation, drainage characteristics, and hydrological conditions. Since these variables interact in complex ways, conventional assessment methods often struggle to provide timely and location-specific hazard predictions.

Recent advances in machine learning have enabled the development of predictive models capable of identifying complex relationships within large environmental datasets. By learning from historical patterns, these models can support disaster preparedness through more accurate hazard assessment. However, many existing approaches focus on a single disaster type, limiting their ability to provide a comprehensive understanding of risk in regions where multiple hazards may occur simultaneously. Furthermore, the lack of model interpretability and practical deployment often reduces their effectiveness in real-world applications.

To address these challenges, this paper presents **BhoomiSense**, an integrated machine learning framework designed to estimate flood probability, landslide probability, and overall hazard severity through a unified web-based platform. The proposed framework combines separate ensemble learning models for flood prediction, landslide susceptibility assessment, and hazard severity estimation. Model performance is enhanced through systematic feature engineering, hyperparameter optimization, probability calibration, and SHAP-based explainability, enabling both reliable predictions and improved interpretation of the factors influencing model decisions.

The developed system follows a modular architecture in which environmental parameters are processed through a FastAPI backend before being evaluated by the trained machine learning models. The prediction results are delivered to a React-based frontend using REST APIs, providing users with an efficient and accessible interface for hazard assessment. The trained models are deployed on Railway Cloud and loaded at runtime, allowing predictions to be generated without retraining.

The major contributions of this work are summarized as follows:

- An integrated framework for predicting flood probability, landslide probability, and overall hazard severity within a single system.
- The use of ensemble machine learning models combined with hyperparameter optimization and probability calibration to improve predictive performance.
- The incorporation of domain-specific feature engineering to capture meaningful environmental relationships affecting disaster occurrence.
- The application of SHAP-based explainability to improve transparency and interpretation of model predictions.
- The deployment of the complete framework as a scalable web-based application capable of providing real-time hazard assessment.

The remainder of this paper is organized as follows. Section II reviews related research on machine learning-based disaster prediction. Section III describes the proposed methodology and system architecture. Section IV presents the experimental evaluation and results, followed by a discussion of the findings. Finally, Section V concludes the paper and outlines potential directions for future work.

II. LITERATURE REVIEW

The application of machine learning in natural hazard prediction has gained considerable attention due to its ability to model complex relationships among environmental and geographical variables. Researchers have explored various computational approaches to improve the prediction of floods and landslides, with a primary focus on enhancing predictive accuracy and supporting disaster risk management.

Several studies have employed machine learning algorithms such as Random Forest, XGBoost, and LightGBM for flood prediction using factors including rainfall, elevation, soil moisture, drainage characteristics, and river-related information. Ensemble learning techniques have demonstrated improved performance by combining the strengths of multiple classifiers, resulting in more robust and reliable predictions than individual models. In addition, feature engineering has been recognized as an important step for extracting meaningful information from environmental data, enabling predictive models to better capture the interactions among hydrological and meteorological variables.

Similarly, landslide susceptibility assessment has increasingly relied on machine learning methods to analyze terrain, geological, and climatic factors. Environmental variables such as slope, rainfall, vegetation cover, elevation, and soil conditions have been widely utilized to identify regions vulnerable to landslide occurrence. Ensemble-based approaches have shown promising results by improving model stability and reducing prediction errors across diverse geographical conditions. Recent studies have also emphasized the importance of model interpretability, encouraging the use of explainable artificial intelligence techniques to better understand the influence of individual features on prediction outcomes.

Despite these advancements, most existing research addresses flood prediction and landslide prediction as separate problems. As a result, users often need to rely on multiple independent systems to obtain a comprehensive understanding of disaster risk. Furthermore, many studies primarily emphasize predictive accuracy while giving comparatively less attention to probability calibration, explainability, and practical deployment in accessible web-based applications.

To address these limitations, the proposed **BhoomiSense** framework integrates flood prediction, landslide prediction, and hazard severity estimation within a single machine learning platform. The framework combines ensemble learning,

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 9, June 2026 Impact Factor: 8.2

domain-specific feature engineering, probability calibration, and SHAP-based explainability to improve both predictive performance and model transparency. In addition, the trained models are deployed through a FastAPI-based backend and made accessible via a web application, enabling efficient real-time hazard assessment. By integrating multiple hazard prediction tasks into a unified framework, the proposed system extends beyond conventional single-hazard approaches and provides a more comprehensive solution for environmental risk assessment.

III. METHODOLOGY

The proposed BhoomiSense framework adopts a systematic machine learning pipeline to predict flood probability, landslide probability, and overall hazard severity for locations across India. The methodology integrates environmental data processing, feature engineering, ensemble learning, model optimization, explainability, and cloud-based deployment into a unified prediction framework. Instead of relying on a single predictive model, the proposed system employs three independent machine learning models, each designed to perform a specific task while collectively providing a comprehensive hazard assessment. The overall workflow begins with the acquisition of environmental and geographical parameters corresponding to a user-selected location. These parameters undergo preprocessing and feature engineering before being supplied to the trained machine learning models. The generated predictions are then returned through a FastAPI backend and visualized using a React-based web interface. The complete methodology adopted in this work is described in the following subsections.

A. Dataset

The effectiveness of any machine learning model largely depends on the quality and relevance of the data used during training. Since flood occurrence and landslide susceptibility are influenced by multiple environmental processes, the proposed framework integrates diverse environmental, geographical, and hydrological datasets to represent the conditions responsible for disaster occurrence. Rather than relying on a single source of information, the framework combines multiple categories of environmental parameters to improve the robustness and reliability of prediction.

The dataset incorporates rainfall measurements, soil moisture information, elevation data, terrain characteristics, river-related parameters, vegetation information, weather variables, and drainage characteristics. These variables collectively capture the interaction between meteorological conditions and geographical features that influence flood generation and slope instability. Rainfall values recorded over different temporal windows provide information regarding both short-term precipitation events and accumulated rainfall conditions. Similarly, terrain attributes such as elevation and slope describe the physical characteristics of the landscape, while river distance, river discharge, and drainage density provide important hydrological information associated with water accumulation and flood propagation.

Vegetation-related attributes and land cover information contribute to understanding the influence of surface conditions on runoff generation, infiltration, and soil stability. Additional environmental variables, including humidity, temperature, mining proximity, and soil moisture, further enhance the representation of local environmental conditions that may contribute to hazard occurrence. By integrating these heterogeneous datasets into a unified feature space, the framework captures multiple aspects of the environmental system responsible for flood and landslide events.

The collected environmental variables serve as the foundation for subsequent preprocessing and feature engineering stages. Instead of directly utilizing the raw attributes, the framework derives several additional domain-specific features that better represent complex environmental interactions, thereby improving the predictive capability of the machine learning models.

Environmental datasets often contain missing values, inconsistencies, and variations arising from differences in data collection procedures and measurement conditions. Such irregularities can adversely affect model performance if they are not addressed appropriately before training. Therefore, a comprehensive preprocessing pipeline is implemented to ensure that the input data is reliable, consistent, and suitable for machine learning.

The preprocessing process begins with data cleaning, where incomplete or inconsistent records are identified and corrected wherever possible. Missing numerical values are handled using median imputation, which replaces

DOI: 10.48175/568

unavailable observations with the median value of the corresponding feature. Median imputation is selected because it is less sensitive to extreme values and preserves the overall distribution of environmental variables more effectively than mean-based replacement.

Following missing value treatment, the environmental variables are organized into a structured feature matrix suitable for machine learning. The preprocessing pipeline also ensures that all required input variables are consistently represented before model training. After preprocessing, the complete dataset is divided into training and testing subsets using an 80:20 train-test split. This separation enables independent evaluation of the trained models while ensuring that prediction performance is assessed on previously unseen data. The resulting processed dataset forms the basis for the feature engineering stage and subsequent model development.

B. Feature Engineering

Feature engineering constitutes one of the most important components of the proposed framework. While the original environmental variables provide valuable information regarding weather and terrain conditions, many hazard-generating processes arise from interactions among multiple factors rather than from individual variables. Consequently, the framework derives several engineered features that explicitly represent these complex relationships.

The engineered feature set includes Terrain Wetness, Terrain Instability, Slope-Rain Stress, Vegetation Stability, Antecedent Precipitation Index, Rainfall Accumulation, Rainfall Intensity, Flood Pressure Index, and Drainage Deficiency. These features are designed to better characterize environmental conditions associated with flood formation and landslide initiation.

For example, rainfall accumulation and antecedent precipitation describe the cumulative influence of previous rainfall events, whereas rainfall intensity represents the severity of recent precipitation. Terrain Wetness combines terrain characteristics with moisture conditions to indicate areas more susceptible to water retention. Similarly, Terrain Instability and Slope-Rain Stress represent the interaction between topographical features and rainfall conditions, improving the ability of the landslide prediction model to identify unstable regions. Flood Pressure Index and Drainage Deficiency provide additional information regarding surface runoff and drainage capacity, enabling the flood prediction model to capture localized flood susceptibility more effectively.

These engineered variables supplement the original environmental attributes and enable the machine learning models to learn meaningful nonlinear relationships that may not be directly observable from the raw input variables. The resulting feature space significantly enhances the overall predictive capability of the proposed framework.

C. Machine Learning Model Development

The BhoomiSense framework consists of three independent machine learning models developed to perform different prediction tasks. Separating the prediction process into specialized models enables each algorithm to focus on a specific hazard while collectively providing a comprehensive assessment of disaster risk.

The first model predicts flood probability using a soft voting ensemble classifier composed of XGBoost, LightGBM, and Random Forest classifiers. Each individual classifier independently estimates the likelihood of flood occurrence based on the engineered environmental features. The final prediction is obtained by combining the probability outputs of all constituent classifiers through soft voting, thereby improving robustness and reducing the limitations associated with any individual algorithm.

The second model estimates landslide probability using the same ensemble architecture. Although the underlying algorithms remain identical, the model learns a different relationship between environmental variables and landslide occurrence. By combining multiple tree-based classifiers, the ensemble effectively captures nonlinear interactions among terrain characteristics, rainfall conditions, vegetation, and soil-related variables that contribute to slope instability.

The third model estimates overall hazard severity using a Voting Regressor consisting of XGBoost Regressor, LightGBM Regressor, and Random Forest Regressor. Unlike the classification models, the regression model predicts a continuous severity value representing the combined impact of environmental conditions on disaster intensity. This additional prediction enables users to understand not only the likelihood of individual hazards but also the overall level of environmental risk associated with a selected location.

D. Model Training and Optimization

To improve predictive performance, each machine learning model undergoes systematic training and optimization. Hyperparameter tuning is performed using RandomizedSearchCV, which efficiently explores different combinations of model parameters while reducing computational complexity compared to exhaustive search techniques. The optimized parameter combinations are subsequently used to train the final ensemble models.

The flood and landslide classification models are evaluated using Stratified K-Fold Cross Validation. This validation strategy preserves the class distribution within each fold, producing a more reliable estimate of model performance and reducing the possibility of biased evaluation.

Following model training, probability calibration is performed using CalibratedClassifierCV with Isotonic Regression. Calibration improves the reliability of predicted probabilities by aligning the confidence scores generated by the classifiers with the actual likelihood of hazard occurrence. Since disaster management applications often rely on probability estimates for decision-making, calibrated predictions provide more meaningful information than uncalibrated classifier outputs.

The trained models are evaluated using multiple performance metrics. The classification models are assessed using Accuracy, Precision, Recall, F1-Score, ROC-AUC, and Confusion Matrix, while the regression model is evaluated using Mean Absolute Error (MAE) and the Coefficient of Determination (R^2 Score). The use of multiple evaluation metrics provides a comprehensive assessment of predictive performance across different aspects of classification and regression.

E. Model Explainability

Although ensemble machine learning models generally provide high predictive performance, their internal decision-making process is often difficult to interpret. To improve transparency, the proposed framework incorporates SHAP (SHapley Additive exPlanations) for model explainability.

SHAP quantifies the contribution of individual features toward each prediction by measuring their influence on the model output. Feature importance analysis is performed for both the flood prediction model and the landslide prediction model, allowing the framework to identify the environmental variables that most strongly affect prediction outcomes. This interpretability not only improves user confidence in the generated predictions but also provides valuable insight into the environmental factors responsible for hazard occurrence.

F. Deployment Workflow

After model development and validation, the trained machine learning models are serialized using Joblib and deployed within a FastAPI backend hosted on the Railway Cloud Platform. During application startup, the serialized models are loaded into memory, enabling efficient prediction without retraining.

When a user submits a location through the web interface, the backend retrieves the corresponding environmental parameters and performs the required preprocessing and feature engineering operations. The processed feature vector is then sequentially supplied to the flood prediction model, landslide prediction model, and hazard severity regression model. The prediction results are packaged as REST API responses and transmitted to the React-based frontend, where they are presented through an interactive user interface.

This deployment architecture separates the frontend from the prediction engine, improving modularity, scalability, and maintainability. The use of FastAPI enables efficient communication between the machine learning backend and the client application, while Railway Cloud provides a reliable deployment environment capable of serving real-time prediction requests without requiring model retraining or manual intervention.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

IV. RESEARCH GAPS

The literature demonstrates that machine learning has significantly improved the accuracy of flood prediction and landslide susceptibility assessment by leveraging environmental, geographical, and hydrological data. Ensemble learning techniques and feature engineering have further enhanced the predictive capabilities of these models. Despite these advancements, several limitations remain that restrict the practical applicability of existing approaches.

A major limitation observed in existing studies is the focus on single-hazard prediction. Most research addresses either flood prediction or landslide susceptibility independently, despite the fact that both hazards are often triggered by common environmental factors such as intense rainfall, terrain characteristics, and soil moisture. This separation prevents a comprehensive assessment of disaster risk, particularly in regions where multiple hazards may occur simultaneously.

Another important limitation is the limited integration of prediction and deployment. Many studies primarily emphasize model development and evaluation without providing an operational framework capable of delivering predictions through an accessible web-based platform. As a result, the transition from research models to practical decision-support systems remains limited.

Furthermore, while several studies report high predictive performance, comparatively fewer incorporate probability calibration to improve the reliability of prediction confidence or explainability techniques to interpret model decisions. The absence of transparent prediction mechanisms can reduce user confidence and make it difficult to understand the environmental factors influencing hazard predictions.

Although environmental variables are widely used in existing models, relatively few studies emphasize extensive domain-specific feature engineering to capture complex interactions among rainfall, terrain, drainage, vegetation, and hydrological characteristics. The predictive potential of engineered environmental features therefore remains insufficiently explored in integrated hazard assessment systems.

To address these research gaps, this work proposes BhoomiSense, an integrated machine learning framework that combines flood prediction, landslide prediction, and hazard severity estimation within a unified web-based platform. The proposed framework incorporates ensemble learning, systematic feature engineering, hyperparameter optimization, probability calibration, SHAP-based model explainability, and cloud deployment to provide accurate, interpretable, and real-time multi-hazard assessment. By integrating these components into a single architecture, BhoomiSense aims to bridge the gap between machine learning research and practical disaster risk assessment.

V. RESULTS AND DISCUSSIONS

The proposed BhoomiSense framework was evaluated to assess its capability in predicting flood probability, landslide probability, and overall hazard severity using environmental, geographical, and hydrological parameters. Three independent machine learning models were developed and optimized for their respective prediction tasks. The performance of each model was analyzed using appropriate evaluation metrics to determine its effectiveness and generalization capability. In addition to predictive accuracy, the framework was evaluated in terms of model reliability, interpretability, and practical deployment.

A. Flood Prediction Model Performance

The flood prediction model was implemented using a soft voting ensemble comprising XGBoost, LightGBM, and Random Forest classifiers. Hyperparameter optimization was performed using RandomizedSearchCV, followed by Stratified K-Fold Cross Validation to ensure reliable model evaluation. The probability estimates generated by the ensemble were further refined using CalibratedClassifierCV with isotonic regression, resulting in more reliable confidence scores.

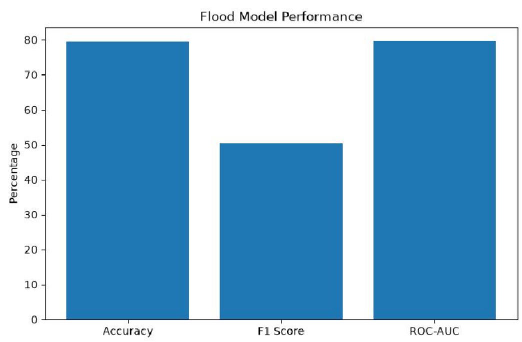
The experimental results indicate that the proposed flood prediction model achieved an accuracy of 79.33%, an F1-score of 50.33%, and a ROC-AUC score of 79.57%. The confusion matrix obtained during evaluation is presented in Table I.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

Table I. Confusion Matrix of Flood Prediction Model

Actual / Predicted	Negative	Positive
Negative	2754	194
Positive	633	419

The confusion matrix demonstrates that the model correctly classified a large proportion of non-flood instances while maintaining satisfactory detection of flood events. The achieved ROC-AUC score indicates that the ensemble effectively distinguishes between flood and non-flood conditions across different classification thresholds. The combination of multiple classifiers through soft voting contributed to improved robustness by reducing the dependence on a single learning algorithm.



B. Landslide Prediction Model Performance

The landslide prediction model employed the same ensemble architecture consisting of XGBoost, LightGBM, and Random Forest classifiers. Similar optimization and validation procedures were adopted to ensure consistency across both classification models.

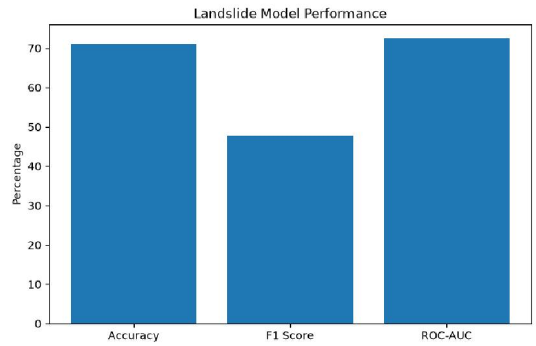
The proposed landslide prediction model achieved an accuracy of 71.00%, an F1-score of 47.79%, and a ROC-AUC score of 72.41%. The corresponding confusion matrix is shown in Table II.

Table II. Confusion Matrix of Landslide Prediction Model

Actual / Predicted	Negative	Positive
Negative	2309	325
Positive	835	531

The results demonstrate that the ensemble model successfully captures the nonlinear relationship between environmental variables and landslide susceptibility. Although landslide prediction remains inherently challenging because of the complex interaction of terrain, rainfall, vegetation, and soil conditions, the ensemble learning approach provides reliable classification performance. The obtained ROC-AUC value further indicates good discrimination capability between landslide-prone and stable locations.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Volume 6, Issue 9, June 2026
ISSN: 2581-9429
Impact Factor: 8.2



C. Hazard Severity Prediction Performance

In addition to predicting individual hazards, the proposed framework estimates the overall severity of environmental risk using a Voting Regressor composed of XGBoost Regressor, LightGBM Regressor, and Random Forest Regressor.

The regression model achieved a Mean Absolute Error (MAE) of 7.867 and an R^2 score of 77.49%. These results indicate that the model explains a substantial proportion of the variation in hazard severity while maintaining a relatively low prediction error. The regression-based severity estimation complements the classification models by providing a quantitative measure of overall disaster intensity instead of limiting the output to binary predictions.

D. Impact of Feature Engineering

Feature engineering played a significant role in improving the predictive capability of the proposed framework. Rather than relying solely on the original environmental attributes, the introduction of engineered variables enabled the models to better represent the complex interactions responsible for flood generation and landslide occurrence.

Features such as Terrain Wetness, Terrain Instability, Slope-Rain Stress, Antecedent Precipitation Index, Rainfall Accumulation, Rainfall Intensity, Flood Pressure Index, and Drainage Deficiency provided additional contextual information that enhanced the learning capability of the ensemble models. These engineered features allowed the models to capture environmental relationships that are difficult to identify using raw variables alone, thereby contributing to improved prediction performance.

E. Model Explainability

To improve transparency, SHAP (SHapley Additive exPlanations) was incorporated into both classification models. SHAP analysis quantified the contribution of each environmental feature toward individual predictions, enabling interpretation of the decision-making process of the ensemble models.

The explainability analysis demonstrated that environmental variables associated with rainfall, terrain characteristics, soil moisture, elevation, and drainage conditions contributed significantly to hazard prediction. By providing feature importance information, SHAP improves confidence in the generated predictions and enables users to better understand the factors influencing flood and landslide risk.

F. Deployment and Practical Implementation

The trained machine learning models were serialized using Joblib and deployed within a FastAPI backend hosted on the Railway Cloud Platform. During runtime, the backend loads the trained models into memory, allowing prediction requests to be processed without model retraining.

The React-based frontend communicates with the backend through REST APIs, enabling users to submit a location and receive predictions for flood probability, landslide probability, and hazard severity in real time. This deployment architecture demonstrates the practical applicability of the proposed framework by integrating machine learning models with a scalable web-based system capable of providing efficient hazard assessment.

G. Discussion

The experimental results demonstrate that integrating ensemble learning, domain-specific feature engineering, probability calibration, and explainability produces a robust framework for multi-hazard prediction. The use of soft voting ensembles improves prediction stability by combining the strengths of multiple machine learning algorithms, while probability calibration enhances the reliability of the generated confidence scores. Furthermore, SHAP-based explainability increases the transparency of model predictions, making the framework more suitable for practical decision-support applications.

Unlike conventional systems that focus on a single hazard, BhoomiSense simultaneously predicts flood probability, landslide probability, and overall hazard severity within a unified platform. The integration of these three prediction tasks, combined with a cloud-based deployment architecture, demonstrates the feasibility of developing an intelligent and scalable decision-support system for environmental hazard assessment. The obtained experimental results indicate that the proposed framework effectively combines predictive accuracy, interpretability, and real-time deployment, making it a promising solution for multi-hazard risk assessment.

VI. FUTURE SCOPE

The proposed **BhoomiSense** framework establishes a strong foundation for intelligent multi-hazard prediction. However, several enhancements can be incorporated to further improve its accuracy, scalability, and practical applicability. The potential future directions are discussed below.

A. Real-Time Environmental Data Integration

The current framework can be extended by integrating continuously updated environmental and weather data to enable real-time hazard monitoring. Incorporating dynamic environmental information would allow the system to respond to rapidly changing conditions and generate more timely predictions. Such an enhancement would improve the framework's capability to support proactive disaster preparedness and emergency response.

B. Expansion of Environmental Datasets

The predictive capability of the proposed models can be further improved by incorporating a wider range of environmental datasets with greater geographical coverage. Expanding the diversity of training data would enable the models to better capture regional variations in climate, terrain, and hydrological conditions across India. This would enhance the robustness of the framework and improve its ability to generalize to previously unseen locations.

C. Advanced Multi-Hazard Prediction

Although the current framework focuses on flood prediction, landslide prediction, and hazard severity estimation, future versions can be extended to include additional natural hazards such as droughts, cyclones, forest fires, and earthquakes. Developing a unified multi-hazard prediction platform would provide a more comprehensive assessment of environmental risks and support integrated disaster management strategies.

D. Enhanced Explainability and Visualization

Future work may focus on improving the interpretability of prediction results by incorporating more advanced visualization techniques alongside SHAP-based explanations. Interactive feature importance analysis and intuitive graphical representations would enable users to better understand the environmental factors influencing predictions, thereby increasing transparency and supporting more informed decision-making.

E. Scalable Cloud-Based Deployment

The deployment architecture can be further optimized to support large-scale real-time prediction requests from multiple users simultaneously. Enhancing cloud infrastructure and backend scalability would improve system performance, reduce response time, and ensure reliable operation during periods of high demand, making the framework more suitable for large-scale deployment.

F. Decision Support and Early Warning Systems

The proposed framework can be integrated with disaster management and early warning systems to assist government agencies and emergency response organizations. By providing timely hazard predictions and severity assessments, the system could support resource allocation, evacuation planning, and disaster mitigation efforts, ultimately contributing to improved public safety and more effective disaster risk management.

VII. CONCLUSION

This paper presented **BhoomiSense**, an integrated machine learning framework developed for predicting flood probability, landslide probability, and overall hazard severity through a unified web-based platform. The proposed framework combines environmental, geographical, and hydrological parameters with ensemble machine learning techniques to provide a comprehensive approach for multi-hazard assessment. By integrating three independent prediction models within a single architecture, the system enables simultaneous evaluation of multiple disaster risks while maintaining an efficient and scalable prediction workflow.

To improve predictive performance, the framework incorporates systematic data preprocessing, domain-specific feature engineering, hyperparameter optimization, probability calibration, and ensemble learning. Furthermore, SHAP-based explainability enhances the transparency of model predictions by identifying the contribution of individual environmental features, making the prediction process more interpretable and reliable.

Experimental evaluation demonstrates the effectiveness of the proposed approach. The flood prediction model achieved an accuracy of **79.33%** with a **ROC-AUC of 79.57%**, while the landslide prediction model achieved an accuracy of **71.00%** with a **ROC-AUC of 72.41%**. In addition, the hazard severity model obtained an **R^2 score of 77.49%** and a **Mean Absolute Error (MAE) of 7.867**, indicating satisfactory regression performance. These results demonstrate that the proposed ensemble-based framework is capable of effectively modeling the complex relationships among environmental factors associated with flood and landslide occurrence.

The trained models were successfully deployed through a **FastAPI** backend hosted on the **Railway Cloud Platform** and integrated with a **React-based** frontend using REST APIs, enabling efficient real-time hazard prediction without requiring model retraining during inference. This deployment demonstrates the practical applicability of the proposed framework as an accessible decision-support tool.

Overall, **BhoomiSense** demonstrates that the integration of ensemble machine learning, domain-specific feature engineering, model explainability, and cloud-based deployment can provide an effective solution for multi-hazard risk assessment. The proposed framework offers a scalable foundation for intelligent disaster prediction systems and has the potential to support future developments in environmental monitoring, disaster preparedness, and risk-informed decision-making.

REFERENCES

[1] S. Kratzert, G. Klotz, K. Brenner, K. Schulz, M. Herrnegger, A. Sampson, S. Hochreiter, and G. Klambauer, “Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks,” *Hydrology and Earth System Sciences*, vol. 23, no. 12, pp. 5089–5110, 2019.

[2] D. T. Bui, H. Shahabi, A. Shirzadi, B. B. Khosravi, and B. Pradhan, “Landslide susceptibility mapping using machine learning algorithms: A comparative study,” *Geomatics, Natural Hazards and Risk*, vol. 10, no. 1, pp. 87–112, 2019.

[3] J. Yaseen, M. Deo, and M. Kisi, “Flood susceptibility mapping using machine learning techniques and environmental variables: A comprehensive review,” *Journal of Hydrology*, vol. 610, Art. no. 127845, 2022.

[4] H. Moayedi, B. Kalantar, B. Pradhan, and A. Foong, “Flood susceptibility assessment using ensemble machine learning models,” *Science of the Total Environment*, vol. 701, Art. no. 134548, 2020.

[5] A. Aditian, T. Kubota, and Y. Shinohara, “Comparison of Random Forest and Support Vector Machine for landslide susceptibility mapping,” *Geosciences*, vol. 8, no. 11, Art. no. 403, 2018.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 9, June 2026
Impact Factor: 8.2

[6] B. Pradhan and S. Lee, "Regional landslide susceptibility analysis using back-propagation neural networks, support vector machines, and decision tree models," *Computers & Geosciences*, vol. 36, no. 6, pp. 747–759, 2010.
[7] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, pp. 785–794, 2016.
[8] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 3146–3154, 2017.
[9] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, pp. 4765–4774, 2017.
[10] M. Dou, M. E. Krampe, and A. B. Smith, "UNet-based flood segmentation: A review and benchmark comparison," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 182, pp. 216–230, 2021.
[11] S. Hong, S. Kim, and J. Park, "A convolutional neural network-based flood extent mapping using Sentinel-1 SAR imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 8, pp. 5379–5390, 2020.
[12] P. Sahana, M. Mohammed, and S. Patel, "Ensemble learning techniques for landslide susceptibility mapping," *Natural Hazards*, vol. 110, pp. 1857–1881, 2022.