Evaluating Machine Learning Models for Identifying At-Risk Highway Slope Assets

Rakesh Salunke, Ph.D., M.ASCE[1]; and Sadik Khan, Ph.D., P.E., M.ASCE[2]

[1]Research Scientist, Dept. of Civil and Environmental Engineering, Jackson State Univ., Jackson, MS (corresponding author). Email: rakesh.salunke@jsums.edu
[2]Associate Professor, Dept. of Civil and Environmental Engineering, Jackson State Univ., Jackson, MS. Email: J00797693@jsums.edu

ABSTRACT

Geographic information system (GIS) based landslide susceptibility mapping is a proven methodology for understanding and forecasting infrastructure impacts during significant weather events. While researchers worldwide have increasingly applied GIS and machine learning methods to study landslide susceptibility on hillside slopes affected by geomorphological and hydrological factors, there is a noticeable lack of focus on highway slope (HWS) failures in the literature. This research addresses this gap by comprehensively evaluating HWS failure susceptibility in central Mississippi counties. The study focused on developing an inventory of HWS susceptible to failure, susceptibility mapping, and model validation using probabilistic and statistical methods. Several supervised machine learning (ML) classification models, including artificial neural networks, were compared with random forest and logistic regression to solve the classification problem of HWS failure susceptibility mapping. Various data sources were utilized to develop causative factors, including Digital Terrain Models (DTM) created from Remote Sensing methods such as satellites, drone sensors, and terrestrial LiDAR. The failed slopes investigated in this study were from four counties in central Mississippi. The resolution used was 3 ft × 3 ft per pixel, representing an area of 9 ft^2 per pixel. A ratio of 1:2 was maintained between failed and non-failed areas within the study area for developing the failure susceptibility prediction models. The causative factors considered in this study encompassed geotechnical and geomorphological attributes, such as slope, aspect, curvature, elevation, normalized vegetation difference index (NDVI), soil composition, and terrain from DTM. Hydrological factors were also incorporated, including precipitation, distance from the stream, groundwater depth, and Topographic Wetness Index (TWI). These causative factors were utilized as independent features to train the classification ML models for predicting vulnerable HWS. Based on the random forest model's classification results of failed vs. non-failed assets on the unseen data set, the influence of the features was calculated. Among the top four influencing factors, ground elevation was the highest contributing factor, followed by distance from streams, NDVI, and precipitation. The results of this study can significantly contribute to transportation agencies by offering valuable insights to target preventative maintenance efforts and mitigate catastrophic failures caused by significant rainfall and weather events on road networks and highway slopes. The findings advocate for the integration of an AI/ML-based approach within asset management programs, enabling transportation agencies to rapidly detect at-risk infrastructure. This ML-based automated detection is especially beneficial when identifying vulnerable sites before a forecasted extreme event, providing value to infrastructure resiliency efforts.

INTRODUCTION

Highway embankment slopes are one of the significant components of transportation geo-infrastructure assets. Embankment failure is a common problem due to various geotechnical, climatological, and environmental contributing parameters. Highway embankments are crucial for the economy because they ensure the safe and efficient movement of goods, services, and people. They improve connectivity between cities, ports, and industrial areas, reducing transportation costs, travel time, and vehicle wear, which boosts trade and economic productivity. Additionally, well-designed embankments enhance the resilience of transport networks against geohazards, landslides, and flooding, minimizing disruptions and maintenance costs and ensuring reliable access to markets and resources essential for economic growth.

Combining geographical information system (GIS) and machine learning (ML) methods has proven effective in mapping landslide and other hazard susceptibility. Studies have applied various methods for landslide susceptibility modeling, such as the frequency ratio (FR) method (Aditian et al., 2018; Ding et al., 2016; Pathak & Devkota, 2022), the weight of evidence (WOE) method (Anderson et al., 2022; Ding et al., 2016), and the analytical hierarchical process (AHP) (Abdi et al., 2021; Shahabi & Hashim, 2015). WOE and AHP assign scores to influence factors based on expert judgment or random assignment.

ML methods, particularly random forest models, have gained popularity for landslide susceptibility modeling, outperforming traditional statistical methods such as AHP and WOE, which require hypothesized weights for influencing factors. The random forest model generally outperforms other machine learning models as they leverage the performance of multiple decision trees while capturing complex non-linear relationships in the data.

Studies by W. Chen et al. (2017), Orhan et al. (2022), and Sevgen et al. (2019) have demonstrated the effectiveness of random forest models in susceptibility modeling. Implementing this proven highway embankment failure susceptibility method could aid the transportation asset management process, including inventory development of at-risk infrastructure.

Despite the proven technique of developing failure susceptibility models using ML and statistical methods, they have not been widely applied to highway embankment failure susceptibility mapping for integration into Geotechnical Asset Management (GAM). When a critical asset inventory is absent, the proposed GIS-based susceptibility modeling technique can develop this inventory, followed by implementing the smart GAM framework. This method represents the initial step in identifying critical highway embankment assets.

This research addressed the gap by evaluating Mississippi's highway embankment and slope failure susceptibility. The study developed an inventory of highway slopes susceptible to washouts and slide failures, created susceptibility maps, and validated models using probabilistic and statistical methods. Various ML classification models were evaluated, including random forest, naïve Bayes classifier, support vector machine, and logistic regression.

The causative factors were derived from multiple data sources, including digital elevation models (DEMs) from remote sensing methods such as satellites, drone sensors, and terrestrial LiDAR. These factors included geotechnical and geomorphological attributes (slope inclination, aspect, curvature, normalized vegetation difference index (NDVI), soil type) and hydrological factors (precipitation, distance from streams, topographic wetness index (TWI), groundwater depth). These factors have been consistently used in landslide and geohazard susceptibility modeling by researchers (Abdi et al., 2021; Chen & Chen, 2021; Kalantar et al., 2019; Mandal et al., 2021; Shahabi & Hashim, 2015; Zhou & Wang, 2019).

These factors were used as independent features for training ML models to predict vulnerable highway slopes and embankments. The results provide valuable insights for transportation agencies, aiding in targeting preventative maintenance efforts and mitigating catastrophic failures caused by significant weather events.

METHODS

The study was carried out in the following stages: (i) development of a failed embankment inventory database; (ii) determining the causative factors; (iii) building and comparing machine learning classification models for predicting failure, including random forest, logistic regression, Naïve Bayes and SVM; (iv) generation of a failure susceptibility map of unstable slopes and embankments using the better performing model.

Digital elevation models (DEM) from satellite imagery and LiDAR survey data, were used to develop the causative factor rasters. DEMs from LiDAR survey data were obtained from the Mississippi Automated Resource Information System (MARIS) from aerial LiDAR survey missions conducted during 2005-2017. The DEM was in the horizontal projection system: NAD 1983 (2011) UTM 15N and 16N, Vertical Datum: NAVD88 GEOID12b, with a resolution or pixel spacing of approximately 0.7 meters (or 3 ft.). Using the DEM data, several causative factors were processed into raster format. Machine learning models were then evaluated for classifying failed slopes, and the best-performing model was selected.

The software used for the susceptibility mapping exercises were ArcGIS Pro, GDAL 324, and Pandas library in Python and JupyterLab 3.6.3, a web-based integrated development environment in the Anaconda Navigator platform.

Highway Embankment Inventory

Approximately 64,000 pixels within a DEM of 3ft. x 3ft. cell sizes of failed embankment slopes in Mississippi were used for training the ML models. Around double the number of pixels representing non-failed areas were randomly selected within the study area for developing the failure susceptibility prediction models. The failed and non-failed highway embankment locations are shown in Figure 1, although the non-failed slopes are not visible on the map due to the scale.

MARIS-derived DEM for the study area. National highway corridors, state highways, and roads were selected and vectorized in Google Earth, and a buffer polygon of up to 1400 feet was generated on both sides of the roads.

Topographical causative factors[1]

Topographical, geological causative factor rasters comprising slope, aspect, Topographic wetness Index(TWI), curvature, plan and profile curvatures, flow accumulation, and distance from a stream were developed from the LiDAR-based DEM. Except Normalized Difference Vegetation Index (NDVI), all other topographical and geological causative factors were developed using the DEM with a 3'x3' cell size resolution obtained from LiDAR survey.

[1]The causative factors rasters are not presented in this paper due to the 10 page constraint. However, they can be provided upon request.

NDVI. NDVI aids in assessing vegetation health and density. Healthy vegetation means the roots are strong and contribute to the soil's shear strength and integrity. NDVI can be an influencing factor for embankment and slope failures. NDVI rasters were developed from Landsat 9 satellite imagery made available by USGS Earth Explorer. The multiple spectral bands of Landsat 9, including the red and near-infrared bands, were extracted in ArcGIS Pro, and NDVI was calculated using Eq.1

$$\boldsymbol{NDVI} = \frac{NIR-Red}{NIR+Red} \tag{1}$$

Where, NIR = Near Infrared reflectance, and Red = Red reflectance value

NDVI values typically range from -1 to +1; water bodies are close to -1; barren areas, rocks, or built-up environments are close to 0; and values of 0 to 1 are assigned to areas based on increasingly dense and healthy vegetation. The NDVI raster developed for the study area is presented in Figure B.8.

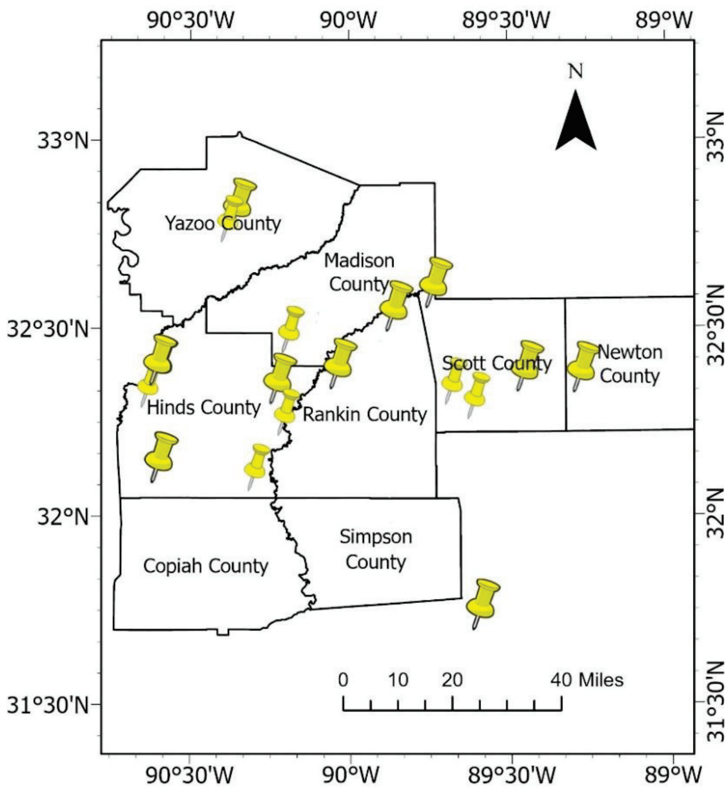

Figure 1. Training data with failed and non-failed highway embankment slopes

Hydrological causative factors

Topographic Wetness Index (TWI). The topographic wetness index is a dimensionless index calculated from digital elevation models (DEMs) that considers each cell's slope and contributing area within a study area. Areas with higher TWI might be more prone to saturation and increased pore water pressure, potentially contributing to slope instability or embankment failure. The TWI was calculated using Eq. 2.

$$\text{TWI}= \ln\left(\frac{a}{\tan(\beta)}\right) \tag{2}$$

Where, a is the specific contributing area (upslope area draining to a point per unit contour length), β is the slope in radians. The TWI raster is presented in Figure B.9.

Flow Accumulation. Flow accumulation represents the number of cells contributing flow to a specific cell in a raster grid, indicating potential runoff. The flow accumulation raster is shown in Figure B.10.

Precipitation. Precipitation, a critical trigger for landslides and slope failures, was analyzed using rainfall data from 2011 to 2020, obtained from NASA's power data access website for eight Mississippi counties. Average rainfall values were rasterized using empirical Bayesian kriging interpolation. The resulting precipitation raster is shown in Figure B.11.

Distance from Stream. Distance from stream measures how far a location or feature is from a stream or river. Proximity to streams can affect slopes due to soil saturation and groundwater level changes. Streams were identified, and distances from highway embankment slopes were calculated in ArcGIS and rasterized. The distance from stream raster is shown in Figure B.12.

Water Table Depth and Soil Type. Groundwater impacts pore water pressure, effective shear strength, and soil density, making groundwater table depth a key factor in embankment failure. The USA Soils dataset from ESRI's Living Atlas Library provided the groundwater table depth and soil type data, which were rasterized and are shown in Figures B.13 and B.14, respectively.

Machine Learning Models for Susceptibility Mapping

Training and validation datasets are crucial in susceptibility mapping. Data sampling strategies influence the accuracy of models. Therefore, in this study, failed embankment slopes were sampled as grid cells, a proven method in susceptibility modeling (Bui et al., 2011). Failed embankment slope polygons were imported in kmz format, converted to rasters in Esri ArcGIS Pro, and sampled to a 3'x3' pixel grid. From the 26 failed slope sites, 29,469 grid cells within eight Mississippi counties were used. Non-failed embankment slopes were randomly selected at a 1:2 ratio to failed slopes. For the purpose of binary classification, failed slopes were labeled 1, and non-failed slopes were labeled 0. This dataset included 14 features and one target column, resulting in 32,217 rows and 15 columns. The dataset was split into 70% training and 30% testing using the train-test split method in the sklearn library.

Data preparation was done in ArcGIS Pro and JupyterLab in Anaconda Navigator using Python programming language and Geospatial Data Abstraction Library (GDAL). A dedicated Python environment within Anaconda Navigator, RasterIO, and GDAL libraries was used to extract training and target values from the rasters as needed. Machine learning methods, random forest, Naïve Bayes, support vector machine classifier, and logistic regression were tested for their accuracy in classifying highway embankment/slope failures.

Random Forest (RF). Random Forest (RF) is an ensemble learning method that creates multiple decision trees to generate classification outputs. Known for robustness, high accuracy, and resistance to overfitting, it combines individual tree predictions for a stable classification. The predicted class probability *P(yi)* for a sample *i* in a binary classification scenario is given by Eq. 3.

$$P(y_i) = \frac{1}{N}\sum_{j=1}^{N} P_j(y_i) \tag{3}$$

where *N* is the number of trees in the forest, and $P_j(y_i)$ is the predicted probability from the *j*-th tree for class $y_{i..}$

Naïve Bayes (NB). Naïve Bayes (NB) is a probabilistic classifier based on Bayes' theorem, assuming independence between features. It is efficient and effective for text classification and spam filtering. The posterior probability The posterior probability, for a binary classification problem, $P(C_k|X)$ for class C_k given features X is computed using Eq.4.

$$P(C_k|X) = P(X)P(X|C_k) * P(C_k) \quad (4)$$

Where $P(C_k|X)$ is the likelihood, $P(C_k)$ is the prior probability of class C_k, and $P(X)$is the evidence.

Support Vector Machine Classifier (SVC). Support vector machine classifier (SVC) is suitable for regression and classification, identifying the hyperplane that best separates classes while maximizing the margin. Effective in high-dimensional spaces, it handles non-linear relationships through kernel functions. For binary classification, a linear SVM finds the hyperplane represented by w*x+b=, where w is the weight vector, x is the input vector, and b is the bias. The decision function is given by Eq 5.

$$f(x)=w*x+b \quad (5)$$

The predicted class is determined by the sign of f(x), with f(x)>0 corresponding to one class and f(x)< corresponding to the other class.

Logistic regression (LR). LR estimates the probability of an instance belonging to a class suitable for binary classification. It models the relationship between the dependent binary variable and independent variables using the logistic function. For binary logistic regression, the logistic function is shown in Eq 6.

$$P\left(Y = \frac{1}{X}\right) = \frac{1}{1+e^{-(\beta_0+\beta_1 X_1+..\beta_n X_n)}} \quad (6)$$

Where $\beta_0, \beta_1, ..\beta_n$ are the coefficients.

ML model performance evaluation.

Confusion matrices were created for each ML model to evaluate their performance on data with known true values. These matrices provide a detailed breakdown of the model's predictions, highlighting instances of correct and incorrect classifications, which is especially useful for binary and multiclass classification problems. A confusion matrix consists of four components:

- True Positive (TP): Correctly predicts a failed area as failed.
- False Positive (FP): Incorrectly predicts a non-failed area as failed.
- True Negative (TN): Correctly predicts non-failed areas as non-failed.
- False Negative (FN): Incorrectly predicts failed areas as non-failed.

Various performance metrics can be derived from these components listed in Eq. 7 through Eq. 12.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

$$\textit{Precision (Positive Predictive Value)} = \frac{TP}{TP+FP} \tag{8}$$

$$\textit{Recall (Sensitivity or True Positive Rate (TPR))} = \frac{TP}{TP+FN} \tag{9}$$

$$\textit{Specificity (True Negative Rate (TNR))}: \frac{TN}{TN+FP} \tag{10}$$

$$\textit{F1 Score} = \frac{Precision*Recall}{Precision+Recall} \tag{11}$$

$$\textit{False Positive Rate (FPR) = 1-Specificity = 1-TNR} \tag{12}$$

ANALYSIS & RESULTS

Comparison of Machine Learning Classification Models performance

The models implemented for failure susceptibility classification were evaluated by comparing their performance metrics. The unseen or 30% test data was used to obtain classification results and compare them with the ground truth data. Confusion matrices were developed from the results obtained from the four machine learning models, and then the dependent performance metrics, including recall, precision, the F-1 Score, and accuracy score, were calculated. Finally, the receiver operating characteristic (ROC) curves and the area under the curve (AUC) score were developed to evaluate the models' performance. The confusion matrixes of all four ML models are presented in Figure 2.

Random forest performed the best, with the highest number (9,655) of true positive and true negative classification results (8,833), followed by NB, SVC, and LR. Comparisons of performance metrics of the different models are presented in Figure 3.

Receiver Operating Characteristic (ROC) Curve

The ROC curve is a graphical representation that illustrates the performance of a binary classification model at various classification thresholds. ROC curve assesses the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity). The ROC curve is obtained by plotting two important parameters with false positive rates (1-specificity) on the x-axis and true positive rates (sensitivity value) on the y-axis. A comparison of ROC curves for all ML models tested in this study is presented in Figure 4. A model that performs no better than random chance would have an ROC curve along the diagonal line (from the bottom-left corner to the top-right corner). A model with a higher area under curve (AUC), closer to 1, generally indicates better classification for the set threshold. It is clear from the figure that the random forest is the best-performing model with the highest AUC (AUC=0.99~1.0).

Based on the evaluation of the four ML models, random forest performed the best and hence was chosen to predict highway embankment failure susceptibility.

Random Forest Results

The classification was performed using the trained RF model. Due to the high resolution of the raster data (3'x3'), the entire 8-county study area generated an immense amount of data (450 million rows or pixels). Processing to classify each pixel became extremely expensive computationally. As a result, the susceptibility prediction was only conducted for a portion of the national highway corridor along I55 and I20 in the Jackson, MS metro area. The susceptibility map is presented in Figure 5.

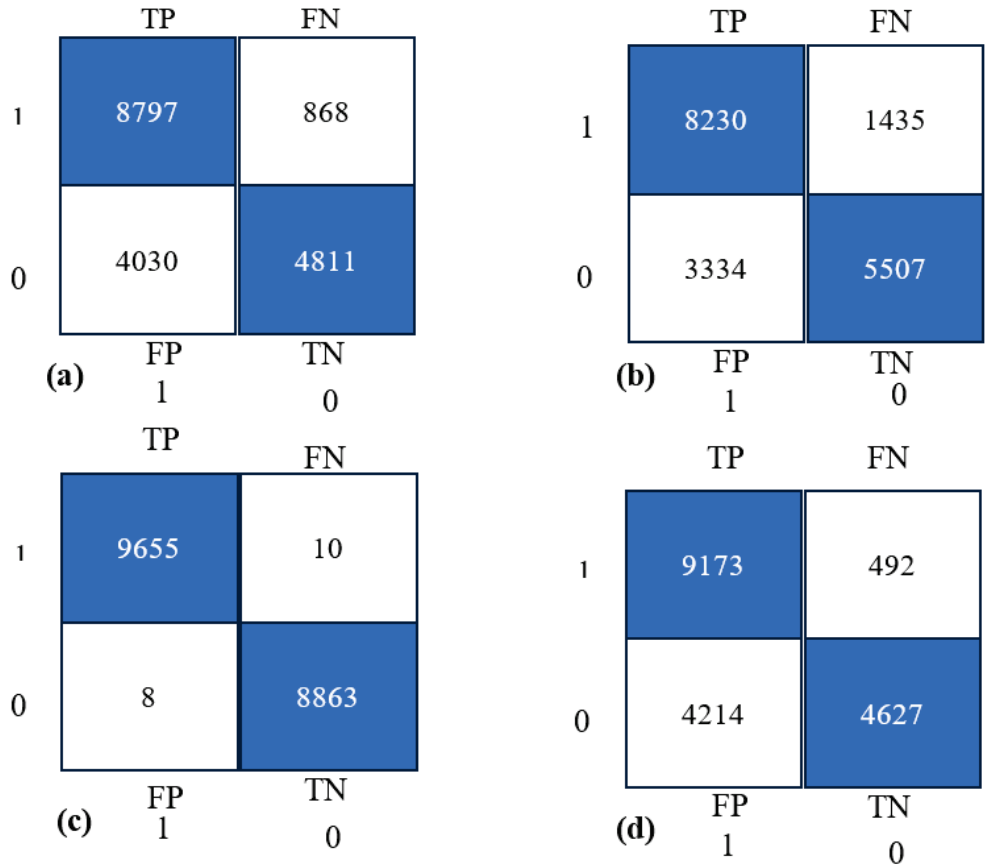

Figure 2. Confusion Matrices from Results of ML Classifications on Test Dataset (a) SVC (b) NB (c) RF (d) LR.

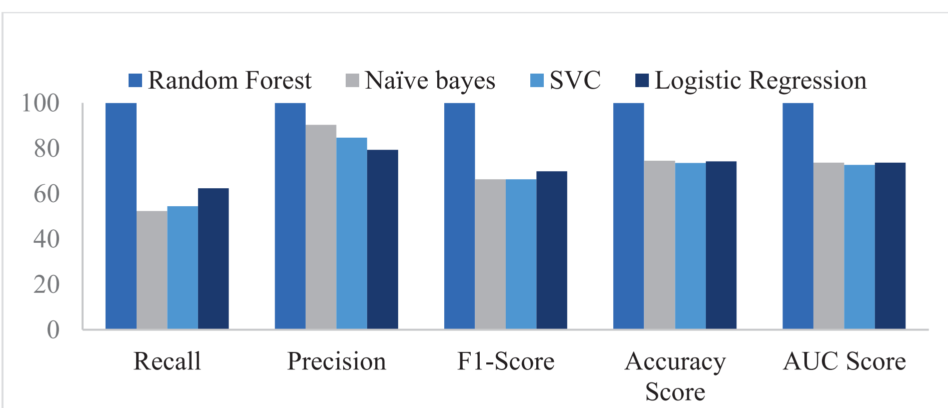

Figure 3. Performance Metrics of ML Models

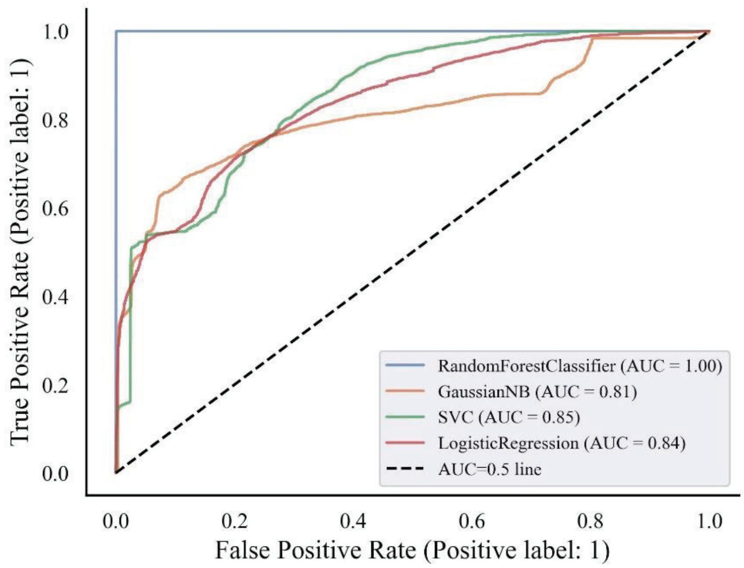

Figure 4. ROC Curve and AUC for the ML Models

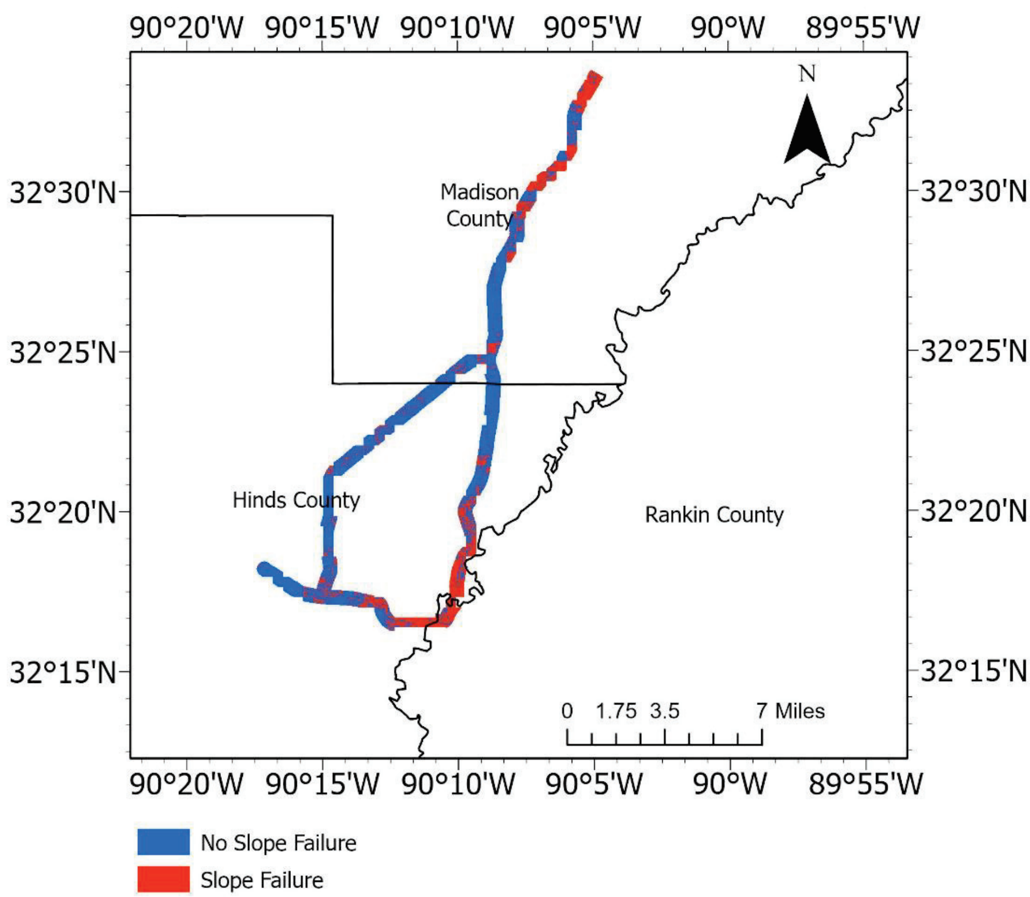

Figure 5. Embankment Failure Susceptibility Mapping for National Highway Corridor in Jackson MS Metro Area, using the Random Forest classifier

Contributing Factors

Based on the random forest model's classification results of failed vs. non-failed assets on the unseen dataset, the influence of the features was calculated. The importance of features in random forest analysis is determined using two methods. One is Gini Importance/Mean Decrease impurity, which is based on impurity reduction in nodes across all tree splits of a given feature. The other method is based on mean decrease accuracy, which measures feature importance by evaluating the drop in model accuracy on out-of-bag (OOB) data when the feature values are randomly permuted. If accuracy decreases significantly, the feature is considered important; if the decrease is minimal, the feature is less important. The resulting importance of features is presented in Figure 6. Among the top four influencing actors, elevation was the highest contributing factor, followed by distance from streams, NDVI, and precipitation.

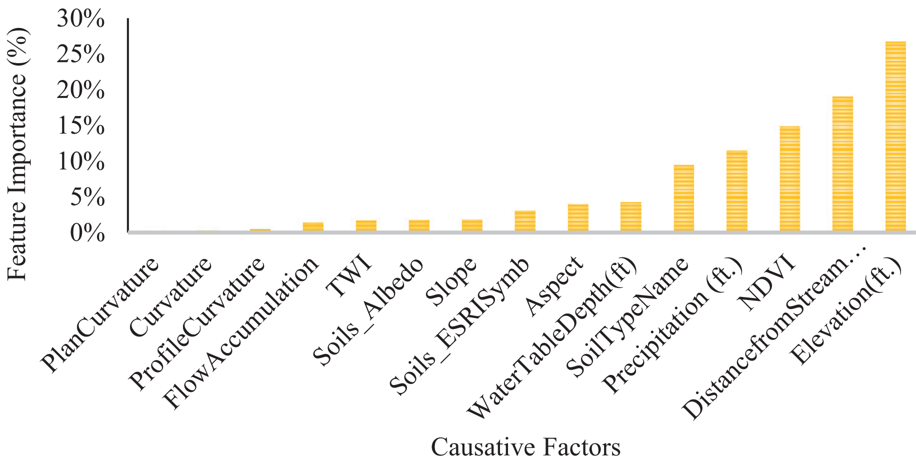

Figure 6. Importance of Features Contributing to Failure

DISCUSSION AND CONCLUSIONS

The random forest model was identified as the most effective classifier in this study, achieving an impressive area under the curve (AUC) of 0.99. This metric highlighted the model's robust performance, superior precision, strong recall, and excellent F-1 score metrics, confirming its status as the top-performing model in the analysis. Random Forest generally performed better than many other classification models because it uses an ensemble of diverse decision trees, reducing overfitting and capturing complex, non-linear relationships in the data. Its built-in randomness, feature selection, and ability to handle noisy and high-dimensional data make it robust and adaptable. When predicting highway embankment failures, the random forest model identified the four most influential factors: elevation, distance from a stream, NDVI, and precipitation. These findings provided insights into the critical variables contributing to the vulnerability of highway embankments and offered actionable strategies for effective risk mitigation.

The failure susceptibility modeling methodology demonstrated its efficacy in evaluating unstable slopes and embankments, supporting Geotechnical Asset Management (GAM). This methodology, validated by the high performance of the random forest model, promises to be a valuable tool for identifying potentially vulnerable highway slopes where interventions can be planned to enhance infrastructure resilience.

Additionally, the research introduced a progressive approach within the GAM framework by integrating remote sensing technologies. This methodology extended beyond traditional asset identification methods that relied on manual inspections or historical failure data. Instead, the presented technique leveraged geospatial analysis coupled with machine learning and statistical classification models. This innovation facilitated the automation of critical asset identification, streamlining the GAM process and enabling a more proactive and data-driven approach to geotechnical asset management. The study's results promote the integration of an AI/ML-based approach within the asset management program for the benefit of transportation agencies in pursuing rapid detection of at-risk infrastructure. Especially when used to detect vulnerable sites before a forecasted extreme event, this ML-based automated rapid detection of at-risk sites provides tremendous value to infrastructure resiliency efforts.

ACKNOWLEDGMENTS

The research presented in this study was partially supported by the National Science Foundation (NSF) projects CMMI #2152896, CMMI# 2046054, and State Study 316, funded by the Mississippi Department of Transportation (MDOT). The authors sincerely appreciate the support of the grants for this study. This paper's contents, statements, and views are strictly those of the authors and do not necessarily reflect the views of their affiliates or the funding agencies.

REFERENCES

Abdi, A., A. Bouamrane, T. Karech, N. Dahri, and A. Kaouachi. 2021. "Landslide Susceptibility Mapping Using GIS-Based Fuzzy Logic and the Analytical Hierarchical Processes Approach: A Case Study in Constantine (North-East Algeria)." *Geotechnical and Geological Engineering* 39 (8): 5675–91.

Aditian, A., T. Kubota, and Y. Shinohara. 2018. "Comparison of GIS-Based Landslide Susceptibility Models Using Frequency Ratio, Logistic Regression, and Artificial Neural Network in a Tertiary Region of Ambon, Indonesia." *Geomorphology* 318 (October):101–11.

Chen, W., X. Xie, J. Wang, B. Pradhan, H. Hong, D. T. Bui, Z. Duan, and J. Ma. 2017. "A Comparative Study of Logistic Model Tree, Random Forest, and Classification and Regression Tree Models for Spatial Prediction of Landslide Susceptibility." *CATENA* 151 (April):147–60.

Chen, X., and W. Chen. 2021. "GIS-Based Landslide Susceptibility Assessment Using Optimized Hybrid Machine Learning Methods." *CATENA* 196 (January):104833.

Ding, Q., W. Chen, and H. Hong. 2016. "Application of Frequency Ratio, Weights of Evidence and Evidential Belief Function Models in Landslide Susceptibility Mapping." *Geocarto International*, March, 1–21.

Kalantar, B., N. Ueda, H. A. H. Al-Najjar, M. B. A. Gibril, U. S. Lay, and A. Motevalli. 2019. "An evaluation of landslide susceptibility mapping using remote sensing data and machine learning algorithms in Iran." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-2/W5 (May):503–11.

Mandal, K., S. Saha, and S. Mandal. 2021. "Applying Deep Learning and Benchmark Machine Learning Algorithms for Landslide Susceptibility Modelling in Rorachu River Basin of Sikkim Himalaya, India." *Geoscience Frontiers* 12 (5): 101203.