Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-156

Empirical Model of Apparent Residual Shear Strength for Enhanced Landslide Prediction Using Machine Learning

Batool Alomari, **L. Sebastian Bryson**, Houston Eatherly
Civil Engineering Department, University of Kentucky, Lexington, United states, Sebastian.bryson@uky.edu

ABSTRACT: In the Appalachian region of the United States, landslides predominantly occur as planar slides within shallow colluvial soils. Traditional limit equilibrium analyses often overestimate soil shear strength, leading to inaccurate slope stability predictions. This overestimation occurs because many landslides in mountainous terrains are reactivations of previous slides, making residual shear strength parameters more appropriate than peak or critical state parameters. This study hypothesizes that residual shear strength parameters can be empirically correlated with peak shear strength parameters, a relationship difficult to model through physics-based approaches alone due to the complexity of hillslope environments. Machine learning offers a promising alternative to derive this empirical relationship. The primary objective was to develop a machine learning-based empirical function relating residual shear strength parameters using observable field factors. This was achieved through dynamic landslide mapping in selected areas of Kentucky with documented slope failures. A key aspect involved back calculating friction angles at failure using a limit equilibrium equation that integrated dynamic changes in soil properties, such as volumetric water content (VWC) and soil suction. These back-calculated phi angles, referred to as apparent residual friction angles, were then compared with initial peak phi angle estimates. A random forest regression model with cross validation was used to establish a feature-driven relationship between apparent residual phi angles and factors such as slope angle, silt content percentage, bulk density, suction stress, and VWC at failure. Model results showed a strong predictive capability, with apparent residual friction angles estimated within approximately one degree using these variables. This approach significantly enhances landslide risk assessment by refining shear strength estimations in colluvial soils, enhancing landslide susceptibility model accuracy. This research highlights the potential of machine learning to overcome limitations in traditional geotechnical methods, especially in complex terrains where precise, field-calibrated data inputs are essential for reliable hazard assessment.

KEYWORDS: Residual shear strength, phi angle, machine learning, landslides.

1 INTRODUCTION

In the mountainous Appalachian region of the United States, landslides predominantly occur as planar slides in shallow colluvial soils. Colluvial soils are characterized by loose deposits accumulated at hillslope bases through weathering and gravitational processes. Analyses of these landslides using limit equilibrium methods have frequently overestimated soil shear strength by employing peak or critical state friction angles instead of residual friction angles.

Traditionally, the peak shear strength of hillslope soils is assumed to dictate stability; however, this approach is inappropriate for evaluating potential landslide sites within active landslide regions. Indeed, many contemporary landslides occur at sites of previous landslides (Mesri and Huvaj-Sarihan, 2012; Brooker and Peck, 1993; Early and Skempton, 1972). Mesri and Huvaj-Sarihan (2012) found that once slopes have undergone significant shear strain, their factor of safety can remain close to unity for extended periods, potentially centuries. Similarly, Mesri and Shahien (2003) analyzed forty-two reactivated landslides and concluded that the mobilized shear strength matched the residual shear strength values obtained from laboratory reversal direct shear or ring shear tests, regardless of time elapsed since initial sliding. Consequently, characterizing potential landslide sites using residual shear strength, rather than peak strength, is advisable. However, quantifying residual shear strength in situ is complicated by the intricate interactions among environmental, geotechnical, and temporal variables. Machine learning methodologies thus offer a valuable alternative for empirically estimating residual strength parameters based on observable field data.

Over the years, numerous landslide susceptibility mapping studies have employed a combination of limit equilibrium methods, Geographic Information Systems (GIS), and predictive as well as physical rainfall data. Chowdhury et al. (2024) conducted a GIS-based susceptibility assessment in Chattogram District, Bangladesh, employing machine learning methods with training variables such as slope, profile curvature, and geology. Guzzetti et al. (2019) discussed twenty-six geographic landslide early warning systems, noting their predominant reliance on real-time rainfall monitoring combined with spatial and temporal datasets. However, such approaches typically provide limited insights, primarily emphasizing rainfall thresholds and historical landslide occurrences without adequately integrating additional geotechnical or environmental factors.

Among machine learning algorithms adept at capturing complex relationships through nonlinear modeling, the Random Forest algorithm is widely utilized. Unlike alternatives such as artificial neural networks, Random Forest requires minimal hyperparameter tuning and is computationally efficient compared to other ensemble methods (Taalab, 2018). A Random Forest model generates predictions by constructing an ensemble of decision trees, each trained on randomly selected subsets of data, and aggregates their outcomes through unweighted averaging for regression analyses (Cutler et al., 2012).

This study aims to demonstrate the predictive capabilities of a Random Forest regression model for estimating residual friction angles, thereby addressing limitations inherent in traditional geotechnical analyses. By leveraging data collected directly from documented landslide sites, the research enhances the validity and practical applicability of the predictive model, ensuring outcomes that accurately reflect realistic soil and site conditions.

2 STUDY AREA

The study area encompasses Breathitt, Floyd, Johnson, and Pike counties in eastern Kentucky, United States, as shown in Figure 1. These counties were selected due to the high incidence of landslides in the region, driven by its mountainous terrain. In

areas such as Breathitt County, where the topography transitions into rolling hills with gentler slopes, landslide occurrences diminish rapidly. In contrast, counties farther east, i.e., Floyd, Johnson, and Pike, feature steeper slopes underlain by shallow colluvial soils atop bedrock primarily composed of sandstone and shale. This eastern landscape is also more susceptible to flooding and rapid watershed flow, further contributing to slope instability.

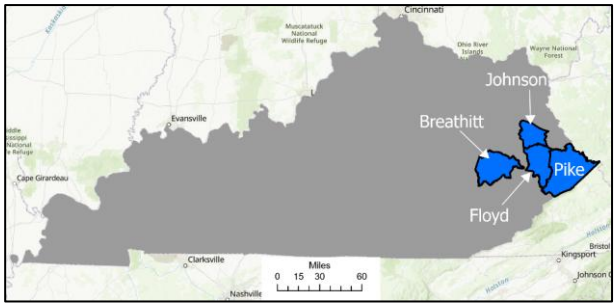

Figure 1. Selected counties in Eastern Kentucky used in this study.

3 METHODS AND PROCEDURES

3.1 *Landslide temporal and geographic data*

Landslide failure dates and geographic locations were obtained from the Kentucky Geological Survey Landslide Inventory (https://uknowledge.uky.edu/kgs_data/7/). The inventory includes data such as the date of failure, county, landslide length, type of slope failure, and a confidence score indicating the likelihood that the event was indeed a landslide. Maintained and regularly updated for public access, the inventory serves as a valuable resource for monitoring and learning from landslide events (Crawford, 2014).

For this study, only landslides within the selected area that had a documented or observed failure date after 2007 were included. This temporal limitation was necessary due to the lack of available precipitation and evapotranspiration data prior to 2007, as further explained in Section 3.6.

3.2 *Physical soil parameters*

Physical soil parameters, including percent sand, silt, clay, plasticity index, and bulk density, were obtained from the National Resources Conservation Center NRCS Web Soil Survey (WSS) (https://websoilsurvey.nrcs.usda.gov/app/) and imported into ArcGIS Pro (v3.0) as a polygon feature class. Figure 2 presents the resulting maps, with stretched symbology applied to improve clarity.

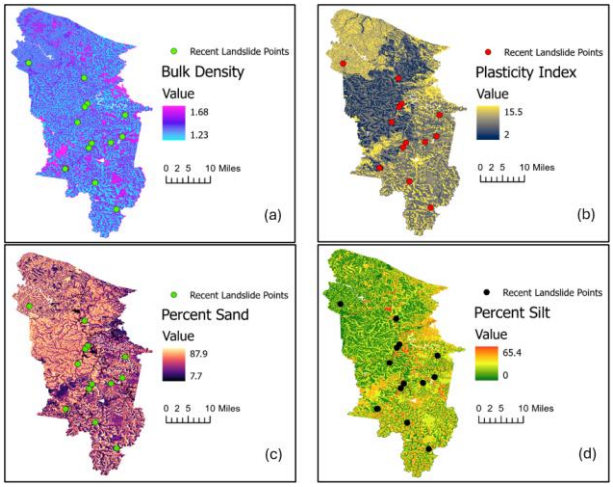

Figure 2. Soil property distribution maps within Floyd and Johnson counties overlaid with recent landslide locations (a) Bulk Density (g/cm³), (b) Plasticity Index (c) Percent Sand and (d) Percent Silt.

Attribute values (soil parameters) were manually added to the feature class table, and the 'Feature to Raster' tool was used to convert these fields into raster datasets at a specified resolution.

Figure 2a shows bulk density in Floyd and Johnson counties, used later to compute soil unit weight for factor of safety calculations. Figure 2b displays the plasticity index, indicating the soil's plastic behavior and shrink-swell potential. Figure 2c and 2d represent the percentage by weight of sand and silt, respectively, classified by Unified Soil Classification System (USCS) as particles ranging from 2–0.075 mm for sand and 0.075–0.002 mm for silt.

3.3 *Prediction of friction angles*

An empirical third order polynomial dependent on plasticity index was used to predict critical state friction angles. Francis and Bryson (2025) developed the third order polynomial relationship from the Terzaghi et al. (1996) study. It should be noted that this equation assumes that the soil being investigated is clayey, plotting above the A-line on an Atterberg chart (Francis and Bryson 2025). The third order polynomial is as in Equation (1)

$$\phi' = -0.00001PI^3 + 0.0033PI^2 - 0.3554PI + 36.194 \quad (1)$$

where ϕ'= friction angle in degrees and PI= Plasticity Index. Figure 3 shows the empirically predicted critical state phi angles across the entire study area by using the 'calculate field' tool in ArcGIS Pro. The areas with higher phi angles in Figure 3 align well with areas with a higher sand percentage in Figure 2c as sand is known to have higher phi angles than fine soils.

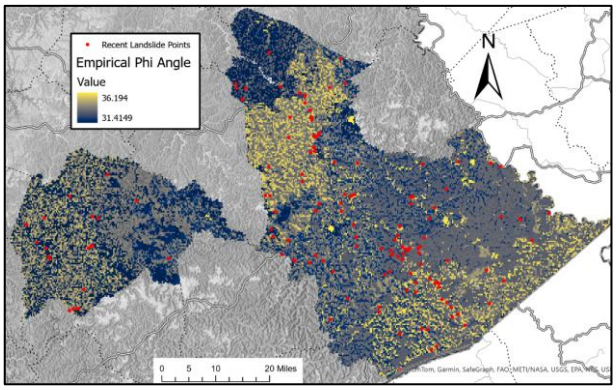

Figure 3. Color graded map showing the empirically derived critical state phi angles (degrees).

3.4 *Depth to bedrock*

Due to the sparse depth-to-bedrock data available from the Web Soil Survey (WSS), a raster layer generated from an internal preliminary dynamic landslide mapping study was utilized in ArcGIS Pro. This preliminary study incorporated data from over 3,000 boreholes across eastern Kentucky, sourced from the Kentucky Transportation Cabinet's (KYTC) online database (https://kgs.uky.edu/kgsmap/kytclinks.asp). The borehole locations, along with their respective depth-to-bedrock measurements, were imported into ArcGIS Pro as (x, y) coordinate points.

A depth-to-bedrock raster layer was then created using the kriging geoprocessing tool in ArcGIS Pro. Kriging is a widely used geostatistical interpolation method in soil science and geology that leverages spatial autocorrelation to produce a

prediction surface with associated uncertainty. It fits a mathematical function to a defined number of sample points and applies inverse distance weighting to estimate values at unsampled locations (Webster and Oliver, 2007).

In this study, depth to bedrock is also referred to as the height of the slip surface, as many shallow colluvial slope failures in eastern Kentucky occur along the contact plane between soil and bedrock. This parameter is critical for slope stability analysis using limit equilibrium methods, where the slope is typically divided into slices. A greater slip surface height results in heavier slices, thereby increasing the driving moment and affecting the factor of safety (Kumar et al., 2023).

3.5 *Calculating slope using digital elevation model data*

Digital Elevation Model (DEM) tiles were obtained from Kentucky From Above (https://kyfromabove.ky.gov/maps/) and imported to ArcGIS Pro. These 1.5 m resolution DEMs, using the state plane coordinate, were converted to a slope (in degrees) raster using the 'slope' tool.

An 18-m buffer was created around each landslide point, and 'extract by mask' tool was used to identify the maximum slope angle in degrees within each buffer. Maximum slope values were selected based on proximity to the slope failure and dominant slope color. As illustrated in Figure 4, purple raster cells in proximity would be used for the max slope value.

For larger DEM, 'spatial join' tool can be used along with the 'slope' tool to extract statistical data across several landslide at once. However, it may select the absolute highest value with no supportive biased for slopes that more accurately describe the landscape. If this technique was used for the landslide shown in Figure 4, it could yield a value of nearly 90 degrees.

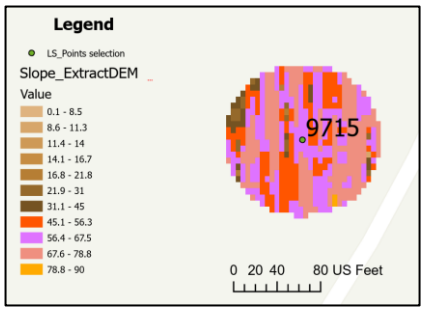

Figure 4. Extract by Mask tool output displaying slope in degrees within the 18-meter buffer zone around the landslide point.

Note that when using the limit equilibrium equation used in this study, any slope equal to or less than 30 degrees yields a factor of safety of 10.

3.6 *Precipitation and evapotranspiration temporal data*

Precipitation and evapotranspiration (ET) data was obtained by Irrigation Manager (http://weather.uky.edu/php/cal_et.php). Irrigation Manager allows users to select any Kentucky county with Mesonet monitoring station and specify a date range to get daily precipitation and evapotranspiration data dating back to 2007. Since Floyd County does not have a Mesonet monitoring station, data from the adjacent Johnson County was used.

The data used in the analysis was dependent on the failure date of the landslide sites. The KGS Landslides inventory includes both "observed" failure dates and actual failure dates; when available, actual failure dates were used to ensure higher temporal accuracy.

3.7 *Predicting volumetric water content on day of failure using Hydrus-1D*

Hydrus-1D is a public Windows-based modeling environment for analysis of water flow and solute transport in variably saturated porous media. The software package includes a one-dimensional finite element model for simulating the movement of water, heat, and multiple solutes in variably saturated media.

Precipitation and ET data covering the period from one month before to one month after the failure was used in the Hydrus models. Days were used as a time-based boundary to illustrate daily variations in soil volumetric water content (VWC).

The van Genuchten – Mualem model was used as soil hydraulic model (van Genuchten, 1980). Hydrus uses a neural network prediction based on the percent sand, percent silt, percent clay, and bulk density collected from WSS to determine van Genuchten model parameters including saturated VWC (θ_s) and residual VWC (θ_r), alpha (α), and fitting parameters (n) and (m). The m parameter is given as shown in Equation (2)

$$m = \frac{n-1}{n} \tag{2}$$

Atmospheric BC with surface layer and free drainage were used for upper and lower boundary conditions, respectively. Hydrus enables specify observation nodes at any depth in the soil profile to view the daily variations in VWC at that depth. In this study, a node was placed one meter below to obtain VWC values on the failure date that was used in the factor of safety limit equilibrium equation.

Although Hydrus effectively captures soil moisture trends, it consistently underestimates volumetric water content (VWC) by approximately 0.10 m^3/m^3 compared to physical monitoring stations, as noted in the preliminary dynamic landslide study. While this difference does not affect the random forest model's ability to learn or classify apparent residual friction angles, it does influence the final predicted values of residual friction angle.

4 ANALYSIS METHODS

4.1 *Back calculating apparent residual friction angles*

Factor of safety equation presented in the preliminary dynamic landslide study presents the factor of safety for subaerial infinite slopes in the saturated zone and is conceptualized as a function of the depth within the vadose zone (Lu and Godt 2008). The equation is given as the sum of three distinct contributions to slope strength; the frictional strength component, the cohesion strength component, and the strength component derived from matric suction (Dashbold et al. 2023). A generalized form of the equation is given in Equation (3)

$$FS = \frac{\tan\phi'}{\tan\beta} + \frac{2c'}{\gamma H_{ss} \sin 2\beta} + r_u(\tan\beta + \cot\beta)\tan\phi' \tag{3}$$

where ϕ'= friction angle (degrees), c'= cohesion (kPa) but it is assumed to be negligible in this study, γ= soil unit weight (kN/m^3), H_{ss}= height of slip surface (m), β= slope angle (degrees), and r_u= pore pressure ratio which is defined as shown in Equation (4)

$$r_u = \frac{\sigma^s}{\gamma H_{ss}} \tag{4}$$

where suction stress σ^s (kPa) is defined as the product of effective degree of saturation S_e and matric suction s (kPa) as shown in Equation (5)

$$\sigma^s = S_e\, s \tag{5}$$

Effective degree of saturation in terms of VWC is defined as shown in Equation (6)

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} \tag{6}$$

where θ = VWC, θ_r = residual VWC, and θ_s = saturated VWC.

The matric suction equation used in this study is derived from rearranging the van Genuchten model (van Genuchten 1980) and is determined by Equation (7)

$$s = \frac{1}{\alpha}\left[\left(\frac{1}{S_e}\right)^{\frac{1}{m}} - 1\right]^{\frac{1}{n}} \tag{7}$$

The factor of safety for each site was calculated in Excel by applying predicted phi angles and other variables to the limit equilibrium equation. Since a slope failure had occurred at each location on a known date, the factor of safety should not exceed 1. For any site where the calculated FS was greater than 1, the predicted phi angle was back calculated using Excel's Solver function. Solver was constrained to adjust only the phi angle, ensuring the resulting FS was less than or equal to 1.

Figure 5 presents a comparison between the phi angles predicted by Francis and Bryson (2025), and the back calculated apparent residual phi angles, all plotted against the soil plasticity index.

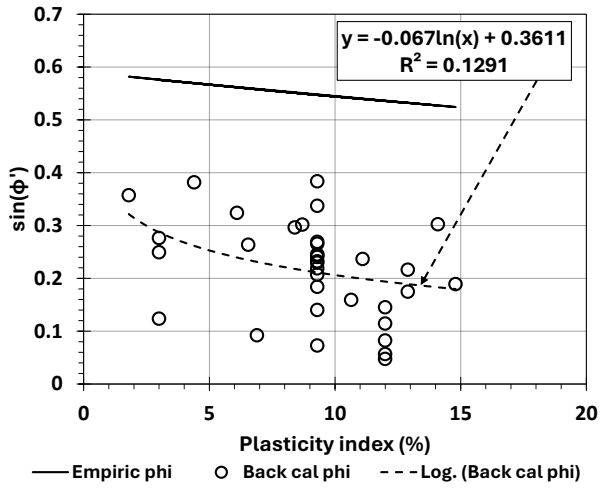

Figure 5. The comparison of empirically predicted phi angles to back calculated apparent residual phi angles.

The empirical equation proposed by Francis and Bryson (2025) consistently overestimates the back calculated apparent residual friction angles, as it was originally intended to predict critical state friction angles.

A logarithmic regression function was applied to evaluate the relationship between plasticity index (PI) and the sine of the back-calculated apparent friction angle (sin φ′). Equation (8) is the resulting regression with a coefficient of determination R^2 = 0.1291. Also, the predicted apparent friction angles show deviation from their actual values in the unity plot shown in Figure 6.

$$sin(\varphi') = -0.067\, ln(PI) + 0.3611 \tag{8}$$

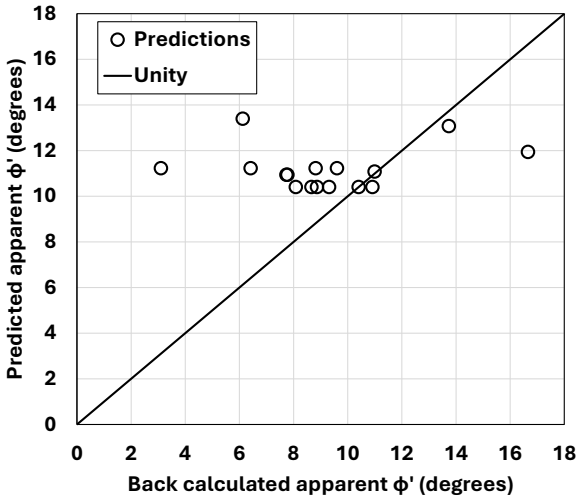

Figure 6. Predicted vs. back calculated apparent φ'.

This very low R^2 value (12.91%) indicates that the regression model explains only a small portion of the variability in sin(φ′). This weak correlation suggests that plasticity index alone is not a reliable predictor of the apparent friction angle.

Thus, univariate regression using only the plasticity index is insufficient to predict complex parameters like apparent friction angle due to the nonlinear and scattered nature of the data. More advanced methods like Random Forest or multivariate models are recommended to account for additional factors.

4.2 *Random Forest regression model*

In this study, a machine learning approach using a Random Forest regression model was employed to predict the target variable, apparent friction angle, based on a set of input features, including percent sand, percent silt, percent clay, bulk density (g/cm^3), plasticity index (%), depth to bedrock (m), VWC at failure day (cm^3/cm^3), suction stress (kPa) and slope angle (degrees). A random forest model consists of multiple decision trees trained on random subsets of data and combined by unweighted averaging for regression (Cutlerand et al. 2012). By combining results from multiple trees, Random Forest models offer higher accuracy when compared to other simpler models (Speiser et al. 2019).

The dataset was initially imported from an Excel file, then divided into predictor variables and the target output. A random 80/20 train-test split was performed using a fixed random seed to ensure the reproducibility of results.

Hyperparameters are configuration settings such as, number of estimators (ranging from 50 to 400), maximum tree depth, minimum samples per split, used to control the behavior of tree-based learning models and how the models will build the decision trees. To enhance model performance and to mitigate overfitting risks due to the relatively small dataset size, hyperparameter optimization was conducted using GridSearchCV with 5-fold cross-validation. Figure 7 shows the general cross-validation process from Scikit-learn.

The model selection criterion was based on the minimization of the negative mean squared error (MSE), see Equation (9) across the validation folds. After the optimal hyperparameters were determined through 5-fold cross-validation, the model was retrained on the entire training dataset

(80% of the total data) using the best configuration that is shown in Table 1.

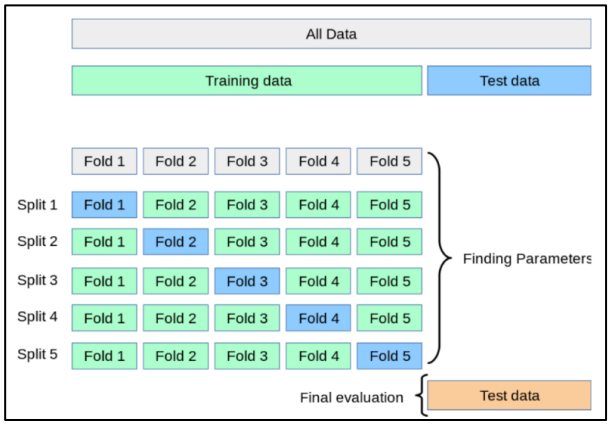

Figure 7. Cross-Validation Process, (Pedregosa et al., 2001).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (9)$$

The test set was withheld throughout the entire model selection and tuning process to ensure unbiased evaluation. It was only used once at the final stage to assess model generalization.

Table 1. Tuned hyperparameters.

Hyperparameter	Value
Number of estimators	200
Maximum tree depth	20
Minimum samples to split	3

Model performance was then assessed on the test dataset using standard regression metrics, including the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE), with the formulas presented in Equation (10), (11), (12), respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{\mathbf{y}}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{\mathbf{y}}_i)^2} \quad (10)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (12)$$

where the n = number of samples, y_i = the actual back calculated apparent friction angle (observed values), $\hat{y}_i$ =the predicted value by the random forest model, and $\bar{y}$ = the mean of the observed values.

4.3 *Results and observations*

The model currently shows a strong predictive accuracy and its generalization capability, as shown by the metrics, listed in Table 2.

Table 2. Standard regression metrics for the model.

R^2	MAE	RSME
0.854	1.468	1.856

To interpret the trained model, feature importances (shown in Table 3) were extracted to determine the relative contribution of each input variable to the prediction and offer insights into the most influential factors. Suction stress (66.26%) and the slope angel (26.25%) were identified as the most contributing factors affecting the apparent friction angle, while depth to bedrock (0.19%) and bulk density (1.16) were the least important.

In addition, Figure 8 shows the RMSE and MAE across training sizes using fivefold cross-validation learning curves. Errors are high with fewer than five data points due to unstable folds. As data points increase, both RMSE and MAE decrease, suggesting general model robustness with increased data.

Table 3. Random Forest model features importance.

Feature	Unit	Importance (%)
Soil suction	kPa	56.26
Slope angle	degrees	26.25
Percent silt	%	9.63
VWC at failure day	cm^3/cm^3	1.85
Percent clay	%	1.74
Plasticity index	%	1.70
Percent sand	%	1.21
Bulk density	g/cm^3	1.16
Depth to bedrock	ν	0.19

Finally, Figure 9 presents a scatter plot comparing actual and predicted values, which allows for visual inspection of model performance and potential prediction biases. The plot shows that the model tends to overpredict both the highest and lowest values of the apparent friction angle.

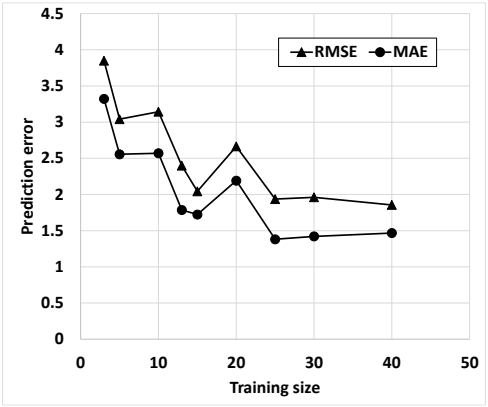

Figure 8. Learning curves of random forest regression model across multiple training sizes.

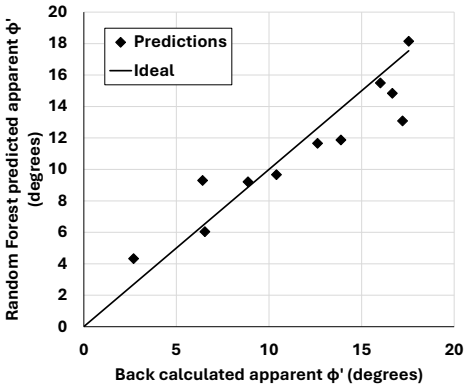

Figure 9. Random Forest predicted vs. back calculated apparent φ'.

5 DISCUSSION

The empirical equation developed by Francis and Bryson (2025) consistently overestimates the apparent residual friction angle, with deviations reaching up to 21 degrees above the back-calculated values. This discrepancy underscores the limitations of relying solely on the plasticity index (PI) in empirical models. Factors such as soil suction and slope angle, which are both excluded from the equation, can significantly influence the apparent friction angle, even among soils with similar PI values.

Results indicate that both Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) decrease rapidly as sample size increases. However, the error metrics exhibit temporary spikes before stabilizing and gradually declining again. This pattern suggests that while additional data generally enhances model performance, the presence of outliers can cause intermittent increases in error, potentially indicating overfitting. Overfitting occurs when a model memorizes the training data rather than generalizing from it (Pedregosa et al., 2011).

Slopes steeper than 45 degrees are typically too steep to retain substantial depths of colluvial soil (Dashbold et al., 2023). As a result, the random forest regression model was required to learn from apparent residual friction angles derived from factors of safety on unrealistically steep slopes. While this does not affect the model's internal learning process or precision, it can diminish its predictive accuracy.

One of the primary limitations affecting the model's accuracy is the quality of the field-collected data, particularly regarding depth to bedrock and volumetric water content. Large gaps between documented borehole locations led the kriging function in ArcGIS Pro to oversmooth the data, reducing spatial autocorrelation. Consequently, nearly all depth-to-bedrock values extracted for the respective sites were nearly identical. Verifying and enhancing the accuracy of field data would significantly improve the model's reliability and could support broader applications in slope assessment, landslide susceptibility, and hazard evaluation efforts.

6 CONCLUSIONS

This study demonstrates the effectiveness of implying machine learning, particularly random forest regression, to predict apparent residual friction angles for shallow landslides in the colluvial soils of the Kentucky Appalachian region. This study demonstrates the potential of using machine learning, particularly random forest regression, to predict apparent residual friction angles for shallow landslides in colluvial soils of the Kentucky Appalachian region.

The high coefficient of determination ($R^2 = 0.854$) and low error metrics (MAE = 1.468, RMSE = 1.856) indicate that the model provides accurate and reliable estimates of residual shear strength based on a diverse set of soil and environmental properties. These results suggest that data-driven models can significantly enhance our understanding of slope stability in complex terrains where traditional analytical approaches face real limitations.

The comparison between traditional limit equilibrium methods using peak strength parameters and those using residual parameters highlights a crucial issue in landslide risk assessment. Models based on peak friction angles significantly overestimated the stability of failed slopes, often predicting unrealistic factors of safety (>1) for landslides that had clearly failed. This highlights the importance of using residual strength parameters especially for slopes that have previously failed or are composed of colluvial materials prone to reactivation.

It is important to incorporate an apparent residual friction angle when moving forward in developing landslide susceptibility hazard mapping efforts. If there is any apparent healing in previously failed hillslope samples, it would likely come from a laboratory error like those listed in Mesri and Huvaj-Sarihan (2012) or reduction of overburden stress from complete soil redistribution and should not be relied on when performing slope stability calculations.

7 REFERNCES

Chowdhury, M.S., Rahman, M.N., Sheikh, M.S., Sayeid, M.A., Mahmud, K.H. and Hafsa, B., 2024. GIS-based landslide susceptibility mapping using logistic regression, random forest and decision and regression tree models in Chattogram District, Bangladesh. *Heliyon*, *10*(1).

Crawford, M.M., 2014. Kentucky Geological Survey landslide inventory: From design to application.

Cutler, A., Cutler, D.R. and Stevens, J.R., 2012. Random forests. Ensemble machine learning: Methods and applications, pp.157-175.

Dashbold, B., Bryson, L.S. and Crawford, M.M., 2023. Landslide hazard and susceptibility maps derived from satellite and remote sensing data using limit equilibrium analysis and machine learning model. *Natural Hazards*, *116*(1), pp.235-265.

Francis, D.M. and Bryson, L.S., 2025. Rainfall-induced landslide hazard analyses using spatiotemporal retrievals of soil moisture and geomorphologic data. *Environmental Earth Sciences*, *84*(8), pp.1-20.

Guzzetti, F., Gariano, S.L., Peruccacci, S., Brunetti, M.T., Marchesini, I., Rossi, M. and Melillo, M., 2020. Geographical landslide early warning systems. *Earth-Science Reviews*, *200*, p.102973.

Hu, S., Lu, Y., Liu, X., Huang, C., Wang, Z., Huang, L., Zhang, W. and Li, X., 2024. Stability prediction of circular sliding failure soil slopes based on a genetic algorithm optimization of random forest algorithm. *Electronic Research Archive*, *32*(11).

Kumar, S., Choudhary, S.S. and Burman, A., 2023. The effect of slope height and angle on the safety factor and modes of failure of 3D slopes analysis using limit equilibrium method. *Beni-Suef University Journal of Basic and Applied Sciences*, *12*(1), p.84.

Lu, N. and Godt, J., 2008. Infinite slope stability under steady unsaturated seepage conditions. *Water Resources Research*, *44*(11).

Mesri, G. and Huvaj-Sarihan, N., 2012. Residual shear strength measured by laboratory tests and mobilized in landslides. *Journal of Geotechnical and Geoenvironmental Engineering*, *138*(5), pp.585-593.

Mesri, G. and Shahien, M., 2003. Residual shear strength mobilized in first-time slope failures. *Journal of Geotechnical and Geoenvironmental Engineering*, *129*(1), pp.12-31.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. and Vanderplas, J., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, pp.2825-2830.

Speiser, J.L., Miller, M.E., Tooze, J. and Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, *134*, pp.93-101.

Taalab, K., Cheng, T. and Zhang, Y., 2018. Mapping landslide susceptibility and types using Random Forest. *Big Earth Data*, *2*(2), pp.159-178.

Terzaghi, K., Peck, R.B. and Mesri, G., 1996. *Soil mechanics in engineering practice*. John Wiley and Sons.

van Genuchten, M.T., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, *44*(5), pp.892-898.

Webster, R., and Oliver, M. A., 2007. Geostatistics for Environmental Scientists (2nd ed.). John Wiley and Sons.