Convergence
A Journal for Young Researchers

ORIGINAL RESEARCH

Received: 10-15-25; Accepted: 10-31-25;
Managing Editor: Jonas Katona;
Action Editor: Jonas Katona

A Dynamic Risk Prediction Model for Rainfall-Triggered Landslides in Buncombe County, North Carolina

Wonjun Choi

Asheville School, Asheville, North Carolina, United States of America

Abstract

Rainfall-triggered landslides commonly occur across the Southern Appalachians, but effective, localized "nowcasts" that fuse dynamic precipitation with terrain characteristics remain limited. We develop a near-real-time landslide risk model for Buncombe County, North Carolina, integrating daily PRISM precipitation windows with static hydrological data. The novel dataset containing 302 mapped landslides (1981–2024) was compiled by integrating USGS Digital Elevation Model (DEM) and Soil Survey Geographic Database (SSURGO) soil data set. Across models, slope was the dominant feature, while short-duration rainfall (1–7 days and 3-day maxima) most strongly increased landslide probability. In particular, 30-day cumulative rainfall often exhibited a negative association with predicted risk, suggesting that prolonged rainfall without intense surges may reduce immediate triggering. The resulting model enables county-scale, data-driven susceptibility updates based on forecasted precipitation data, offering a practical foundation for near-nowcasting and a landslide alarm system during and after intense storms.

Keywords: rainfall-triggered landslides, nowcasts, Buncombe County, random forest, XGBoost, PRISM precipitation, landslide susceptibility modeling

1. Introduction

Landslides are devastating natural hazards that take place worldwide and cause wide-ranging damage to modern society, such as human casualties, economic losses, and infrastructure damage (Froude & Petley, 2018; Kirschbaum & Stanley, 2018; Petley, 2012). In the year 2024 alone, 766 instances of landslides took place around the world, killing 4,933 individuals (Petley, 2025). Landslides across the world incur an economic cost of 20 billion USD annually, taking 17% of the yearly mean expenses caused by all natural disasters worldwide between 1980 and 2013 (Klose et al. 2016). Social infrastructures, especially transportation and communication, are highly susceptible to landslides (Sim, Lee, Remenyte-Prescott, & Wong, 2022). Rural areas are likely to experience more impairments due to landslides, where these infrastructures are scarcer and

scattered (Klose et al. 2015).

In search of effective landslide prediction methods, researchers have proposed various theories on the sources of these disasters, ranging from geological activity such as volcanic eruptions and earthquakes to human-induced sources such as modification of slopes and obstruction of hydrological flows (Sidle & Ochiai, 2006; Jaboyedoff et al., 2016). Among those factors, precipitation is regarded as the leading cause of landslides worldwide; extreme rainfall triggered 70% of all landslide events across the world between 2004 and 2010 (Petley, 2012). Fittingly, the role of precipitation in landslides will only continue to grow with climate change; Gariano and Guzzetti (2016) predict that landslides triggered by rainfall will be increasingly catastrophic and recurrent as climate change increases the frequency of short, intense storm episodes. The predicted increase in landslide prevalence forebodes the urgency of an effective rainfall-induced landslide prediction model.

The Western North Carolina region has historically been dominated by rainfall-induced shallow landslides due to geographical and meteorological factors (Wooten et al., 2008). The Blue Ridge Mountains, as a part of the Southern Appalachian mountain range, are characterized by steep slope gradients (Khashchevskaya et al., 2025). Its bedrock layer mostly comprises metamorphic and igneous rocks, which, when decomposed, create saprolite, which constitutes the top layer of the horizons (Hatcher, 2010; Watterson & Jones, 2006). The saprolite-heavy soil absorbs water easily, which increases pore pressure, resulting in a higher probability of shallow landslides (Aydin, 2006). The region's geomorphic susceptibility has been called for focus again after Hurricane Helene's extensive landslide damage in Western North Carolina in 2024 (Lin et al., 2024). Allstadt et al. (2025) state that the hurricane caused unprecedented geographic disruptions in the region, particularly in Buncombe, Henderson, Rutherford, and Yancey counties. Freshwater flooding contributed to 95 of the 176 direct deaths that Helene caused, most of which were landslide-related; around 2,000 independent cases of landslides were reported to be related to Helene, with the majority located over western North Carolina (National Hurricane Center, 2025). Helene significantly damaged the communications infrastructure in western North Carolina, considering WNC's rural characteristics, which exacerbated the situation (Office of State Budget and Management, 2024). Helene's unexpected landslide damage calls for a real-time probabilistic forecasting model, localized for Western North Carolina.

At present, there is no near-nowcast model for Western North Carolina that captures dynamic changes in precipitation. In fact, nearly all of the current landslide hazard maps use minimal or no precipitation inputs. Specifically, the USGS Landslide Hazard Mapping Program's landslide susceptibility map identifies areas with elevated risk for landslips under extreme rainfall conditions (Fuemmeler et al., 2008); these, however, do not take into account real-time changes in precipitation since they were based on past correlations of landslides with geographical elements (Bauer et al., 2012). Academically, even though there have been numerous attempts at developing a physical/statistical model for the prediction of landslides (e.g., an ML-based model by Lin et al. (2024) that combines a variety of predictors such as the NC landslide database, soil surveys, digital elevation models, and water body locations), very few, if any, utilized dynamic rainfall as their primary predictor. To this end, an effective localized model of Western North Carolina for rainfall-induced landslides has proven necessary, as seen in the extensive damage Hurricane Helene caused in the region.

Thus, to address the absence of a localized prediction model that incorporates dynamic rainfall variability, the proposed research will address the question: How does incorporating static geographic data (elevation and soil depth) with the variability of rainfall data affect the predictability of shallow landslides in the Western North Carolina region? Ultimately,

this study hopes to create an efficient prediction model in near real-time (rainfall data are updated daily) for the area, developing an early warning system for residents of Western North Carolina, and specifically Buncombe County. Because the model relies on rainfall windows that end on the event day, it produces a near-nowcast rather than a true forecast with a defined lead time. Generating operational lead-time predictions would require integrating short-term rainfall forecasts, which is beyond the scope of the present study but represents a clear direction for future work.

Global models based on rainfall-triggered landslides have been developed and utilized for several years, with generally positive results. They are considered “nowcasts” because they incorporate both static geological characteristics and dynamic observations of rainfall to provide nearly real-time forecasts (Kirschbaum et al., 2012). For example, NASA's LHASA v2 model is the first operational global system to utilize dynamic satellite rainfall data along with static geological and environmental characteristics (Stanley et al., 2021). It is based on a Logistic Regression algorithm and examines static predictor variables as well as IMERG precipitation (1 km resolution) to generate probabilistic hazard labels. The new model is purported to be two-fold more accurate than the previous LHASA v1 system, which utilized a standard threshold algorithm (Kirschbaum & Stanley, 2018; NASA Earth Observatory, 2021).

Only recently has the volume of published landslide susceptibility studies grown substantially, from 31 in 2016 to 219 in 2024 (Ye et al., 2025). Supervised machine learning (ML) was a popular choice for its ability to capture nonlinear relationships between predictor types and landslide probability, resulting in a higher level of accuracy than conventional linear statistical models (Regorda et al., 2020). Mondini et al. (2023) developed a time-dependent landslide model based on Italy. Exploiting the strength of Recurrent Neural Networks (RNNs) in sequential data, they used continuous rainfall data for a window of 30 days, excluding geological or environmental factors. Similarly, Chan et al. (2018) built a logistic regression-based model aimed at predicting landslides caused by a typhoon’s heavy rain in the southern region of Taiwan. The model mainly used runoff flow depth, not pure precipitation data, taking into account soil saturation, which reached an accuracy of 80~85%. Kang et al. (2024) recently constructed a localized model for Yunnan Province, China, with its Random Forest model reaching an accuracy level of 0.906. All the above ML-based models center around rainfall data. Their high accuracy levels affirm the need for dynamic rainfall data in a localized landslide model.

Focusing on North America, Thomas et al. (2019) challenged whether satellite rainfall data can replace in situ hydrological data to evaluate the soil saturation threshold for a slope failure, reflecting a recent trend of increasing remote geospatial data-based models (Akosah et al., 2024). The findings from this localized model of California demonstrated that rainfall (specifically satellite-measured rainfall data) cannot be relied upon to predict landslides and encourage a hydrogeological gauge that calibrates rainfall data with on-site hydrologic data to develop a soil wetness index. Other studies that support the aforementioned findings include a comprehensive model (Lee et al., 2023) blending precipitation duration/intensity and normalized soil moisture capacity, which dropped false alarms (FA) from ~26 to 3. Both studies emphasize the importance of contextualizing raw rainfall data with local hydrologic features in evaluating the soil saturation level.

The contribution of this study is twofold. We build a localized, near-real-time landslide prediction model. For the modeling, an equal number of non-events were randomly chosen based on the landslide inventory of Buncombe County (North Carolina Department of Environmental Quality, 2024). To avoid pattern biasing in specific constellations of events, stratification and a spatial block cross-validation approach were utilized. Three different ML algorithms were tested overall, with three models per algorithm (F0, F1, and F2) based on feature type. The assessment of performance and feature analysis was interpreted using evaluation metrics, SHAP plots, and permutation importance.

To fit and evaluate the model, we designed a hydrologic/environmental dataset that collates daily precipitation data (PRISM Climate Group, 2024), a topography map (USGS DEM), and soil depth (USDA SSURGO). The dataset is open and available for further research at the specified repository.

The next section describes the datasets and methodology of the research. Section 3 illustrates the results of the findings. Section 4 delivers the conclusion of the research.

2. Data and Methodology

2.1. Dataset

This research aims to create a dynamic prediction model localized in Buncombe County, North Carolina, geographically located within the Blue Ridge Mountains (Figure 2a). The inventory of landslide events was extracted from the North Carolina Landslide Points dataset (via NC OneMap; North Carolina Department of Environmental Quality, 2024). The dataset includes event geometry (points) and traceability fields: *IsLandslide* (event flag), *Event_Date*, *Sort_Date*, X, Y, *County*, *GlobalID*, *OBJECTID*, *Data_Type*, and *Source_Period*. The inventory was both a temporal anchor (*Event_Date*) for rainfall-window calculations referred to by dates and a spatial anchor (X, Y) for topography and soil data extraction. Additionally, an equal number of non-event samples were created for control, randomly selecting the same time (January 1981–December 2021) and the same region of interest (Buncombe County) as those in the landslide dataset. Ultimately, an equal number of events and non-events were included in the final dataset used for modeling.

Gridded precipitation (4 km resolution) was extracted from PRISM daily products (PRISM Climate Group, 2024) throughout the entire analysis period (spanning from January 1981 to December 2024). Using the reference dates of the events and non-events—the event date (*Event_Date*) or the control (*Random_Date*)—historical sums over windows of 1-day (*R1d*), 3-days (*R3d*), 7-days (*R7d*), and 30-days (*R30d*) were computed. Other temporal windows included maximum rainfall sums calculated for both 3-day (*Max_Rainfall_3day*) and 30-day (*Max_Rainfall_30day*) intervals. All precipitation measurements were in units of millimeters, and all the calculation windows' end dates coincided with the reference date of the observation (Table 1 contains the definition and sources of all features).

Topography was taken from a DEM (Digital Elevation Model) encompassing Buncombe County, published by the U.S. Geological Survey (2022). The DEM already provided elevation (*Elevation_m*), while slope (*Slope_deg*) was computed from Horn's (1981) 3×3 finite difference gradients and converted into degrees (Figure 2; Equations 1, 2, and 3).

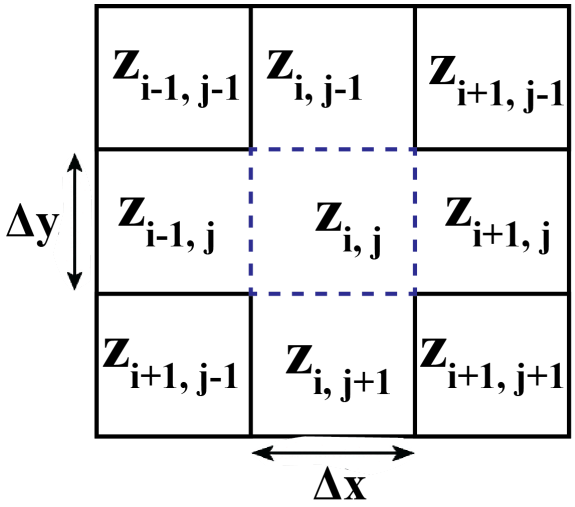


Figure 1: Visualization of Horn's (1981) 3×3 finite difference gradient calculator

Note: $z_{i,j}$ symbolizes the position of the elevation of a particular cell. Δx and Δy are the horizontal and vertical grid distances of the cells.

$$p = \frac{(z_{i+1,j-1}+2z_{i+1,j+1}-z_{i-1,j-1})-(z_{i+1,j+1}+2z_{i-1,j}+z_{i-1,j+1})}{8\Delta x} \quad (1)$$

$$q = \frac{(z_{i-1,j+1}+2z_{i,j+1}-z_{i+1,j+1})-(z_{i-1,j-1}+2z_{i-1,j}+z_{i-1,j-1})}{8\Delta y} \quad (2)$$

$$slope = tan^{-1}(\sqrt{p^2+q^2}) \quad (3)$$

The values p and q in Equations 1 to 3 represent the rate of elevation change in the east-west direction and the north-south direction, respectively. The DEM cell resolution for Buncombe County showed an equal width and height of 40 meters. To avoid unit discrepancy across different datasets, all datasets were under a consistent projected coordinate reference system (NAD83/North Carolina; EPSG:32119) and maintained a resolution of 40m.

Soil data originated from the USDA NRCS Soil Survey Geographic Database (SSURGO) map units (U.S. Department of Agriculture, Natural Resources Conservation Service, 2024). A point-polygon comparison was used to match the Map Unit Key (MUKEY) and Map Unit Symbol (MUSYM) to each sample point, linking depth information from the related tables. Soil depth was converted to a numeric variable (*Soil_Depth_cm*) by extracting data from raw strings (*Soil_Depth_cm_raw*) and converting the variable into a binary flag (*Soil_Depth_Deep200_Flag*), where entries larger than 200 were changed to 1, and any other entries were changed to 0. These soil properties were viewed as static covariates that reflect conditions of regolith associated with instability triggered by rainfall.

For map compilation and masking, administrative boundaries by county were employed to clip rasters and to define the sampling window for controls with a modest buffer to attenuate edge effects (U.S. Census Bureau, 2022). Land cover from NLCD was retained for descriptive mapping and voluntary sensitivity tests but was not included in the baseline prediction set (U.S. Geological Survey, 2021).

Gridded variables such as rainfall windows, elevation, and slope were sampled by bilinear interpolation at coordinates defined by points (X, Y) (Burrough & McDonnell, 1998). Soil data, specifically soil depth, came from the MUKEY and MUSYM area codes. Units of measurement have been standardized with precipitation in millimeters, elevation in meters, and slope measured in degrees. We verified the rainfall windows to check that the end dates coincide with the event reference date. Data instances with missing values were removed from the dataset.

The dataset created is a linked dataset, combining information from PRISM daily rainfall data, the USGS DEM, and the USDA SSURGO map specific to Buncombe County, and can be used for further analyses related to landslide risk. Linking information from the various datasets was not trivial, as it required gathering chronological information and summarizing it for the selected data points before linking it to the main dataset (Longley et al., 2015; Burrough & McDonnell, 1998). Other challenges included incompatible reference systems that had to be reconciled. The complete dataset includes the outcome variable (*IsLandslide*) and full fields for past rainfall, topography, and soil depth with comprehensive metadata for provenance and audit purposes, which comprises *Event_Date/Random_Date*, X, Y, *County*, *GlobalID*, *OBJECTID*, *Data_Type*, *Source_Period*, *MUKEY*, *MUSYM*, and *Soil_Depth_cm_raw*.

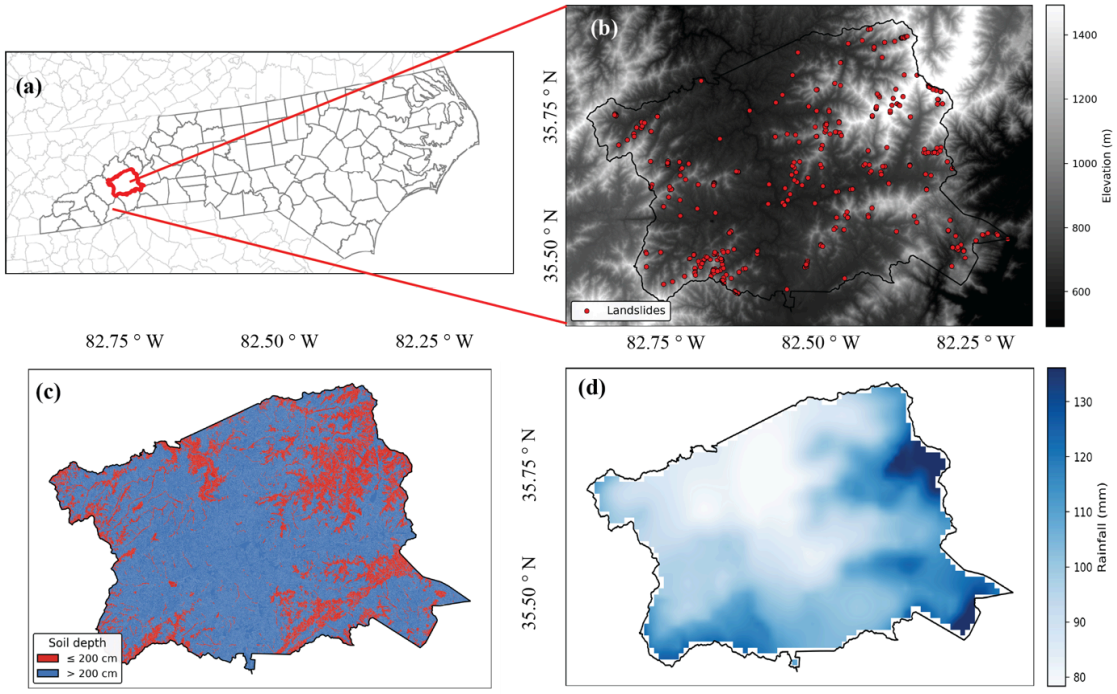


Figure 2: (a) Geographical location of the study area, (b) elevation, (c) soil depth, (d) 30-year average rainfall.

Table 1: Feature definitions.

Symbol (unit)	How it was calculated	Source
R1d (mm)	Sum of daily PRISM rainfall for 1 day	PRISM (800m)
R3d (mm)	Sum of daily PRISM rainfall for 3 days	PRISM (800m)
R7d (mm)	Sum of daily PRISM rainfall for 7 days	PRISM (800m)
R30d (mm)	Sum of daily PRISM rainfall for 30 days	PRISM (800m)
Max_Rainfall_3day (mm)	Rolling maximum of 3-day totals in past 30 days	PRISM (800m)
Max_Rainfall_30day (mm)	Rolling maximum of 30-day totals in past 90 days	PRISM (800m)
Elevation_m (m)	Extracted directly from DEM	USGS DEM (40m)
Slope_deg (°)	Derived from DEM using slope algorithm	USGS DEM (40m)
Soil_Depth_Deep200_Flag (—)	Flag for soils deeper than 200 cm	SSURGO Soil

Note: Rainfall timeframes are from the reference point of the corresponding event.

All data collection, processing, and model development steps were done in Python 3.9 in the PyCharm IDE. Libraries, including NumPy, Pandas, SciPy, Matplotlib, and Seaborn, were used for scientific computing; GeoPandas, Shapely, and Rasterio were used for data processing; and scikit-learn and XGBoost models combined with SHAP were chosen for model development and assessment. The models were run on a Mac Apple M2 processor.

2.1. Modeling

To assess which rainfall variables specifically influenced landslide probability, the temporal accumulation/maximum rainfall predictors were analyzed with all non-empty possible combinations. A primary concern was minimizing overfitting due to the large number of rainfall predictors (*R1d*, *R3d*, *R7d*, *R30d*, *Max_Rainfall_3day*, *Max_Rainfall_30day*). Thus, a training–test split was implemented via a stratified five-fold cross-validation approach using a constant seed and data point shuffling. In addition, the evaluation was repeated for every rainfall predictor subset. The model that showed the most accurate predictions on the test data was included in the final comparison across other types of models. The above processes were independently conducted for each machine learning (ML) algorithm model.

Three separate ML algorithms were employed for the actual study: Logistic Regression, Random Forest, and XGBoost (Breiman, 2001; Chen & Guestrin, 2016). Due to logistic regression's simple and linear design, the model was effective in serving as a benchmark to compare with nonlinear models. Receiver Operating Characteristic (ROC) curves plot the true positive rate against the false positive rate across probability thresholds. The Area Under the Curve (AUC) summarizes model discrimination. The diagonal dashed line represents the no-skill baseline, where the model performs no better than random chance. The performance of the model was quantified through ROC curves, confusion tables, and summary measures: accuracy, precision, recall, ROC–AUC, and the Brier score (Brier, 1950). For logistic regression, the coefficient magnitudes summarized how each variable influenced the model prediction, while standard errors, *p*-values, and confidence intervals

measured the degree of statistical significance of that prediction.

A Random Forest algorithm was used as the baseline nonlinear model. Interpretation of the model involved a combination of intrinsic feature importance as well as SHAP analysis. Intrinsic feature importance was determined as the mean decrease in impurity, indicating the effectiveness of each predictor in splitting the data. Shapley values for each variable and data point were illustrated via SHAP bar plots, SHAP beeswarm plots, and SHAP waterfall plots per sample (Lundberg & Lee, 2017). The bar plot visualizes the average contribution of each feature to the overall prediction, while the beeswarm plots show both the distribution as well as the direction of the predictor contributions. SHAP waterfall plots give local explanations regarding the most likely landslide samples. A full dataset, including SHAP values as well as both the imputed and original feature values, expected values, probabilities, and true labels, in addition to original metadata, was also provided so that both global and case-specific insights could be obtained. Model discrimination as well as calibration were also assessed as part of ROC curves and confusion matrices, as well as the same summary metrics that were used as part of Logistic Regression.

In answering the hypothesis, three separate models were created for each type of ML algorithm. Model F0 used just slope and elevation, aiming to assess the predictive potential based on the static predictor: terrain alone. Model F1 added rainfall accumulation intervals, alongside slope and elevation, as the basis to also evaluate the short- as well as longer-term rainfall triggers. Following on from F1, Model F2 also included soil depth so that the effect of the subsurface interacting with both rainfall and the terrain factors could be determined. Each algorithm was then trained across the three feature settings, ultimately leaving us with parallel models that allowed a direct comparison.

The F0, F1, and F2 configurations for each model (Logistic Regression, Random Forest, and XGBoost) were trained and tested on the same sample, and therefore any discrepancies in results were not driven by sampling but rather variable differences.

Comparison analysis conducted on XGBoost involved classification accuracy, showing ROC curves and confusion matrices, as well as the complete set of metrics, including accuracy, precision, recall, F1 score, and ROC–AUC, as well as the Brier score. Interpretable techniques, such as SHAP dependence plots or permutation importance, are only presented for the Random Forest model.

3. Results and Discussion

The compiled landslide dataset contained 9,092 events recorded within North Carolina between 1940 and 2024, of which 398 occurred within the study area of Buncombe County (Figure 3). Because the PRISM Weather daily rainfall data before 1981 was unavailable to the public, the inventory was limited to the time frame between 1981 and 2024, limiting the data points to 302 landslide events. Corresponding to these landslide points, an equal number of control points were created by sampling sites and dates randomly within the same county and time frame. As such, the final dataset contained 604 total entries, split evenly between events and controls, thus reducing potential biases for certain variables and increasing its robustness toward new data.

All the records had a collection of hydrological covariates. Rainy features comprised both cumulative windows (*R1d*, *R3d*, *R7d*, *R30d*) as well as highest intensities (*Max_Rainfall_3day*, *Max_Rainfall_30day*). Geological features included elevation as well as slope at 40 m resolution. Soil depth was included as an added predictor expressed as a binary flag (>200 cm vs. ≤200 cm), as the vast majority of the sites had extremely deep soils.

The final dataset merged the rainfall, terrain, and soil characteristics for each observation. Its well-balanced design (302 events + 302 controls) ensured the risk of class bias was kept small, and the range of covariate variability enabled the models to incorporate long-term conditioning factors (e.g., slope, soil depth) as well as short-term triggers (e.g., extreme rains).

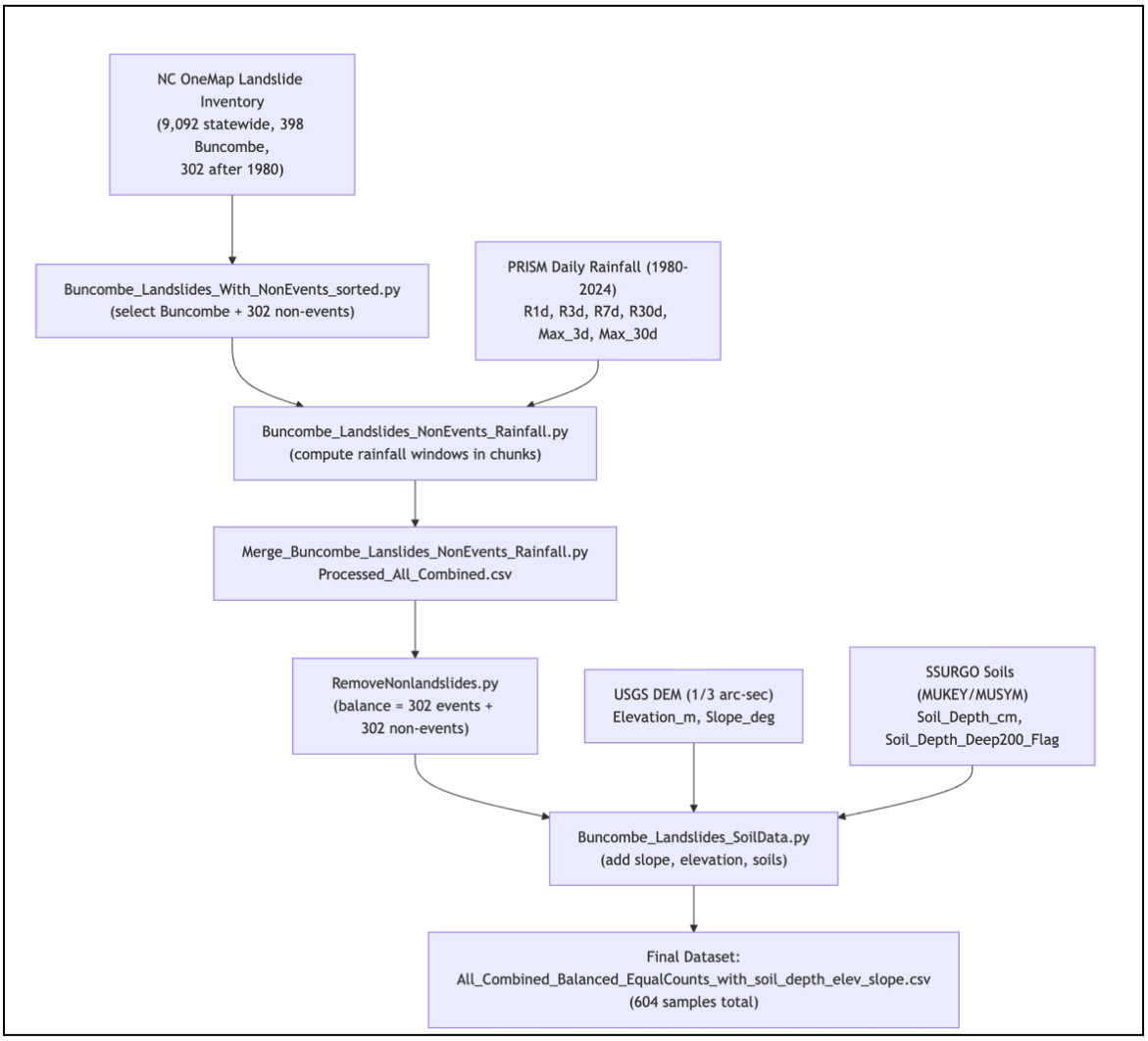


Figure 3: Flowchart of landslide events and non-events data processing

3.1. Feature selection

Feature selection indicated consistency across the nonlinear and linear models. In the Logistic Regression category, the optimum subset of rainfall data comprised Max_Rainfall_30day, R3d, R7d, and R30d, achieving an accuracy of 0.593 as well as an ROC–AUC of 0.614 (Table 2). Short-period rainfall accumulation, including three-day as well as seven-day buildup, was

frequently shown in the top-ranking combinations, showing how these predictors have a strong relationship with landslide risks. However, long-term predictors such as Max_Rainfall_30day and R30d also showed up occasionally, albeit not as frequently as short-term variables.

Table 2: Top-performing subsets of Rainfall Features for Logistic Regression

Logistic Regression (Top 5)	Accuracy	ROC–AUC
Max_Rainfall_30day, *R3d*, *R7d*, *R30d*	0.593	0.614
Max_Rainfall_30day, *Max_Rainfall_3day*, *R3d*, *R7d*	0.589	0.621
Max_Rainfall_30day, *R1d*, *R3d*, *R7d*, *R30d*	0.589	0.613
Max_Rainfall_30day, *R7d*, *R30d*	0.584	0.600
Max_Rainfall_30day, *Max_Rainfall_3day*, *R1d*, *R7d*, *R30d*	0.584	0.619

Table 3: Top-performing subsets of Rainfall Features for Random Forest

Logistic Regression (Top 5)	Accuracy	ROC–AUC
Max_Rainfall_30day, *Max_Rainfall_3day*, *R1d*, *R3d*	0.808	0.888
Max_Rainfall_30day, *R1d*, *R7d*, *R30d*	0.806	0.882
Max_Rainfall_30day, *R3d*, *R7d*, *R30d*	0.805	0.890
Max_Rainfall_30day, *Max_Rainfall_3day*, *R1d*, *R3d*, *R7d*	0.801	0.890
Max_Rainfall_30day, *Max_Rainfall_3day*, *R1d*, *R7d*, *R30d*	0.801	0.883

In the Random Forest output, the best-performing subset included *Max_Rainfall_30day*, *Max_Rainfall_3day*, *R1d*, and *R3d*, yielding exceptional accuracy up to 0.808 and ROC-AUC reaching as high as 0.888 (Table 3). Moreover, subsets including *R3d* and *R7d* also appeared among the very best, a point also supported by the results yielded by the Logistic Regression (Table 2). However, the Random Forest model also brought to light that short-term rainfall predictors like *R1d*, as well as *Max_Rainfall_3day*, significantly improved classification, bringing to the forefront nonlinear interactions. These findings indicate that although both nonlinear and linear prototypes invariably recognized the paramount significance of short-term rainfall, the Random Forest technique captured the additional nuances of the lengths of the rainfall affecting the probabilities of the landslide occurrences (Table 3).

Table 4: Logistic Regression model performance across feature sets (F0, F1, F2)

FeatureSet	Accuracy	Precision	Recall	F1 Score	ROC_AUC	Brier
F0	0.620	0.609	0.650	0.629	0.673	0.229
F1	0.702	0.688	0.733	0.710	0.771	0.193

F2	0.702	0.682	0.750	0.714	0.775	0.191

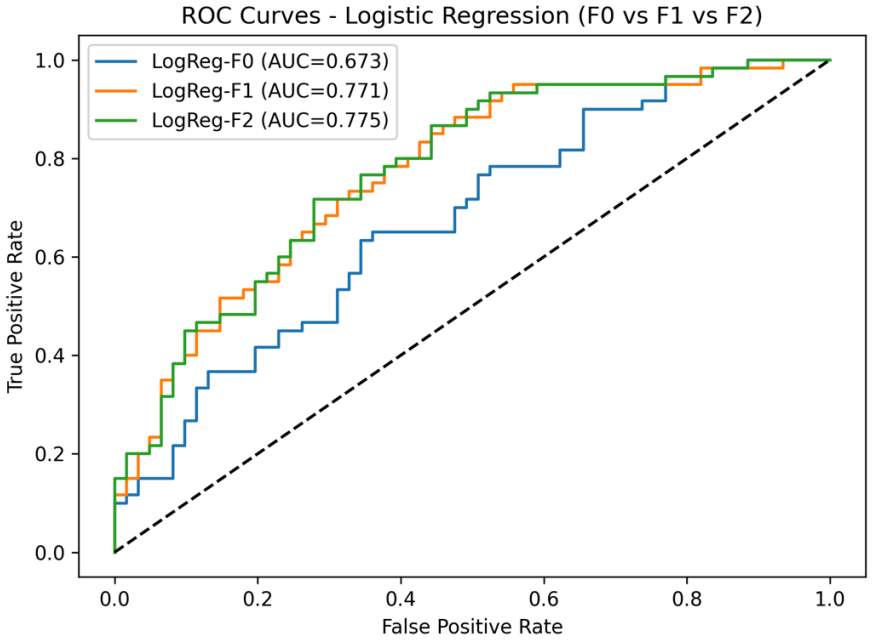
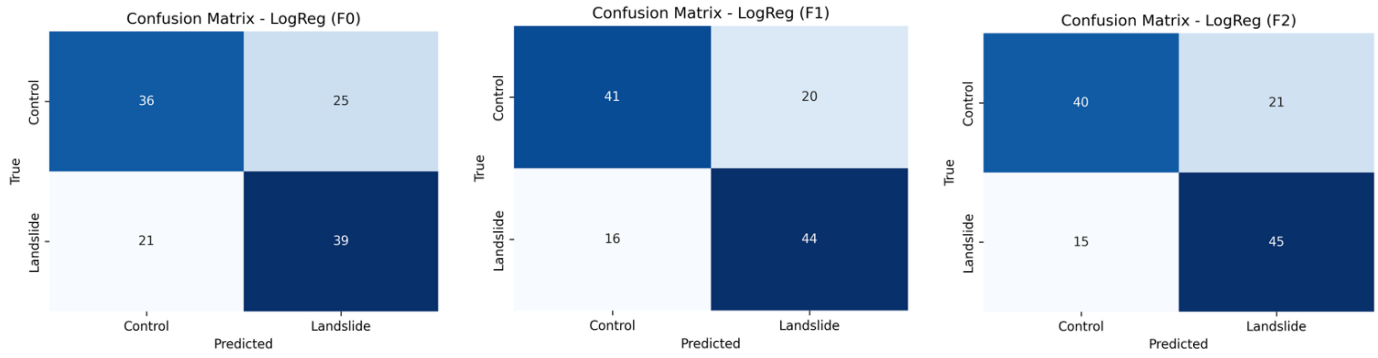


Figure 4: ROC Curves for the three Logistic Regression models: F0, F1, and F2.

Note: Colors may be difficult to distinguish for some readers; interpretation should rely on line style and the AUC values.



Figures 5-7: Confusion matrices for F0, F1, and F2 logistic regression models.

Table 5: Regression coefficients for each feature of the F0, F1, and F2 logistic regression models.

Feature	F0	F1	F2
Slope_deg	0.7150	1.3574	1.3813
Max_Rainfall_30day	-	0.5666	0.5871
R30d	-	-0.5968	-0.6327
R7d	-	0.2192	0.2136
Elevation_m	-0.0287	-0.1886	-0.1099
R3d		-0.0571	-0.0620
Soil_Depth_Deep200_Flag			0.1943

Note: Rainfall-related coefficients are intentionally blank for the F0 model, as F0 includes only static terrain predictors (slope and elevation) and therefore does not estimate coefficients for any rainfall variables.

Performance for Logistic Regression got progressively better as more predictors were added (Table 4). In the terrain-only version F0, accuracy was 0.62 with an AUC (Area Under Curve evaluates how well the model distinguishes between cases and controls) of 0.67. Inclusion of the rainfall predictors in F1 got the accuracy up to 0.70 and the AUC up to 0.77 (Figure 4). Inclusion of the full set of predictors in the complete version F2 got the best accuracy at 0.70, as well as the best AUC at 0.78. The confusion matrices further corroborated this ranking of models (Figures 5-7): F2 got 40 correct controls as well as 45 correct landslides, compared to 36 and 39 cases for the F0 model, and 41 and 44 cases for the F1 model, respectively.

No formal statistical test (e.g., McNemar's test or bootstrap confidence intervals) was conducted to evaluate whether these accuracy differences are statistically significant, and the improvements should therefore be interpreted cautiously.

Model interpretation relied on the regression coefficients (Table 5). The slope was the strongest predictor variable, followed closely by *Max_Rainfall_30day*, both having a positive effect on the risk of landslides, except for the F0 model. *R30d* represents the total cumulative rainfall in the 30 days preceding the event, whereas *Max_Rainfall_30day* represents the single wettest 30-day period within the past 90 days. These variables capture different hydrologic processes: *R30d* reflects gradual wetness buildup, while *Max_Rainfall_30day* captures past extreme episodes. Their differing definitions explain why one may be positive and the other negative in regression coefficients. Here, the cumulative 30-day rainfall (*R30d*) tended to have a negative coefficient, meaning long-term accumulation did not significantly increase the probability of slope failure after accounting for short-term accumulation or extreme events. These results strengthen the conclusion that the best explanation for landslides lies in the interaction between the region's steep slope and short-term, high-rate bursts of rain, as shown in the Brier Scores of Table 4.

Table 6: Random Forest model performance across feature sets (F0, F1, F2).

FeatureSet	Accuracy	Precision	Recall	F1 Score	ROC_AUC	Brier
F0	0.661	0.679	0.623	0.650	0.741	0.216
F1	0.810	0.797	0.836	0.816	0.888	0.139
F2	0.785	0.778	0.803	0.790	0.880	0.144

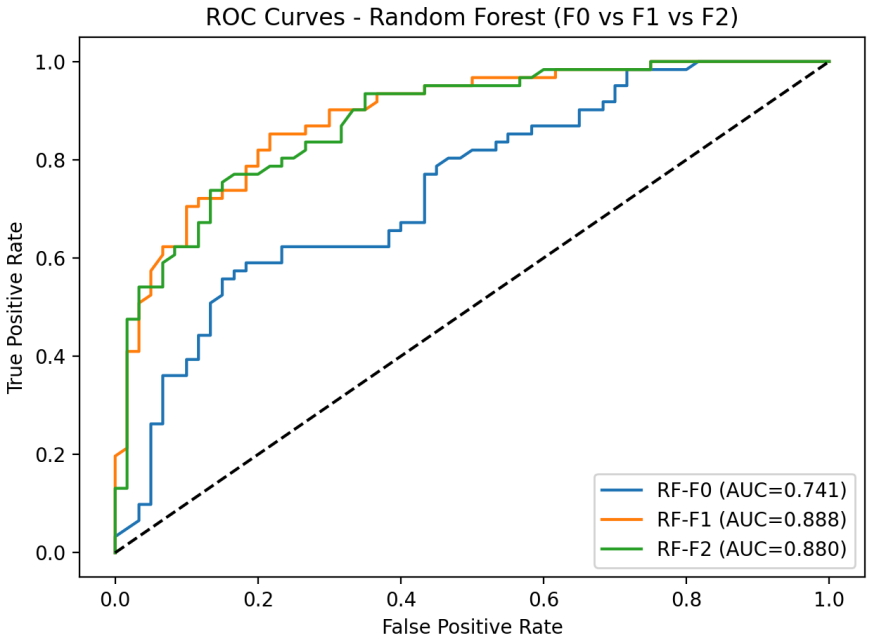
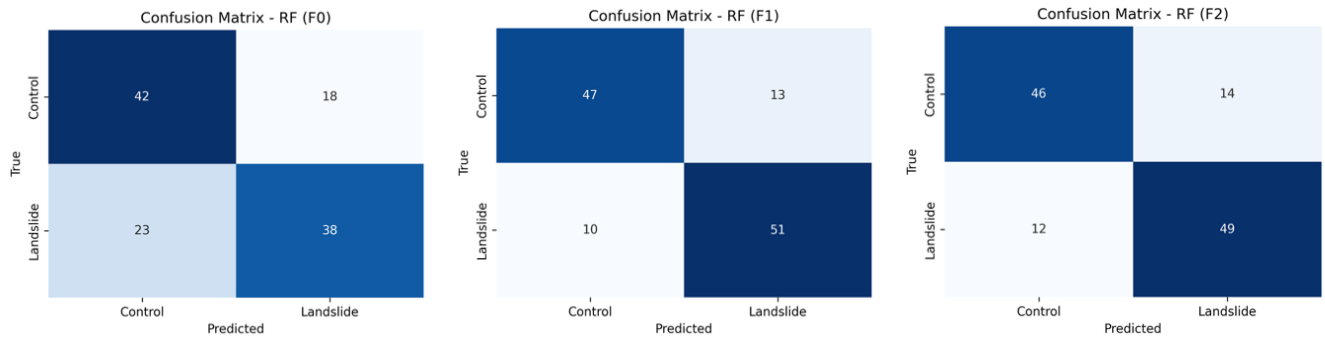


Figure 8: ROC Curves for the three Random Forest models F0, F1, and F2

Note: Colors may be difficult to distinguish for some readers; interpretation should rely on line style and the AUC values.



Figures 9-11: Confusion matrices of the three models F0, F1, and F2 of Random Forest

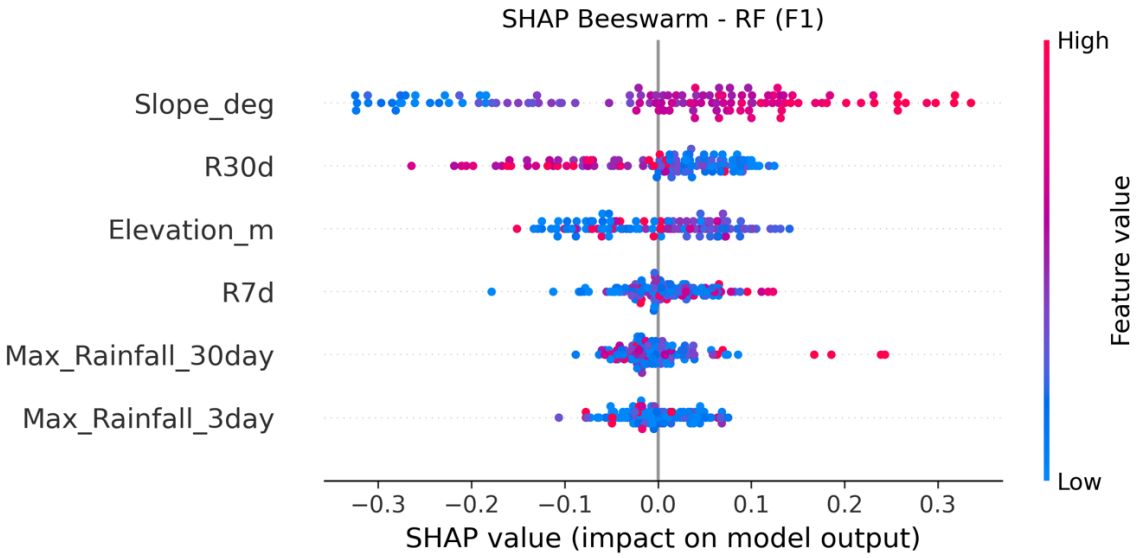


Figure 12: SHAP beeswarm chart for the F1 Random Forest model

The best-performing model out of all three algorithms was the Random Forest (RF) model. It was consistent across varying combinations of features (Table 5; Figure 8). Although the accuracy of the F0 model was 0.66 and the AUC 0.74—still notably higher than the accuracy of the Logistic Regression (LR) model—the F1 model setup, incorporating rainfall data, gave an accuracy of 0.81 and an AUC of 0.89, demonstrating the best performance of all. But the introduction of the soil feature in F2 detracted from the performance to an accuracy of 0.79 and an AUC of 0.88, departing from the trends of the LR. The slight performance decrease when soil depth is added (F1 to F2) may reflect the coarse spatial scale of SSURGO polygons relative to the 40-m DEM grid, introducing noise rather than meaningful stratification (Figure 9-11). Because most mapped soils in Buncombe County are uniformly deep (>200 cm), the limited variability may dilute stronger predictors such as slope and short-term rainfall.

Considering F1's superior performance, feature analysis was only conducted on the F1 model, focusing on SHAP figures (Figure 12). The figure confirmed that slope is the primary predictor, exerting the most influence on the model. Subsequently, it was followed by *R30d*, elevation, and short-term rainfall features such as *R7d*, *Max_Rainfall_30day*, and *Max_Rainfall_3day*. According to the SHAP beeswarm plot (Figure 12), slope indicated a positive relationship with landslide risk, while elevation showed a bidirectional influence. This pattern occurs because elevation functions as a proxy for diverse terrain settings in Buncombe County: higher elevations may correspond to steep, dissected hillslopes that increase susceptibility (positive SHAP values), whereas other high-elevation areas sit on stable ridge tops with gentler gradients (negative SHAP values). Likewise, some lower-elevation areas occur in broad, low-slope valleys that are less prone to failures. Because elevation captures multiple landscape conditions rather than a single physical mechanism, its SHAP values naturally spread across both positive and negative ranges. Short-term bursts of rain, along with extreme maxima, exerted powerful effects in increasing the probability of landslides; conversely, cumulative rains over a 30-day period revealed a negative relationship, reducing the likelihood of shallow failures.

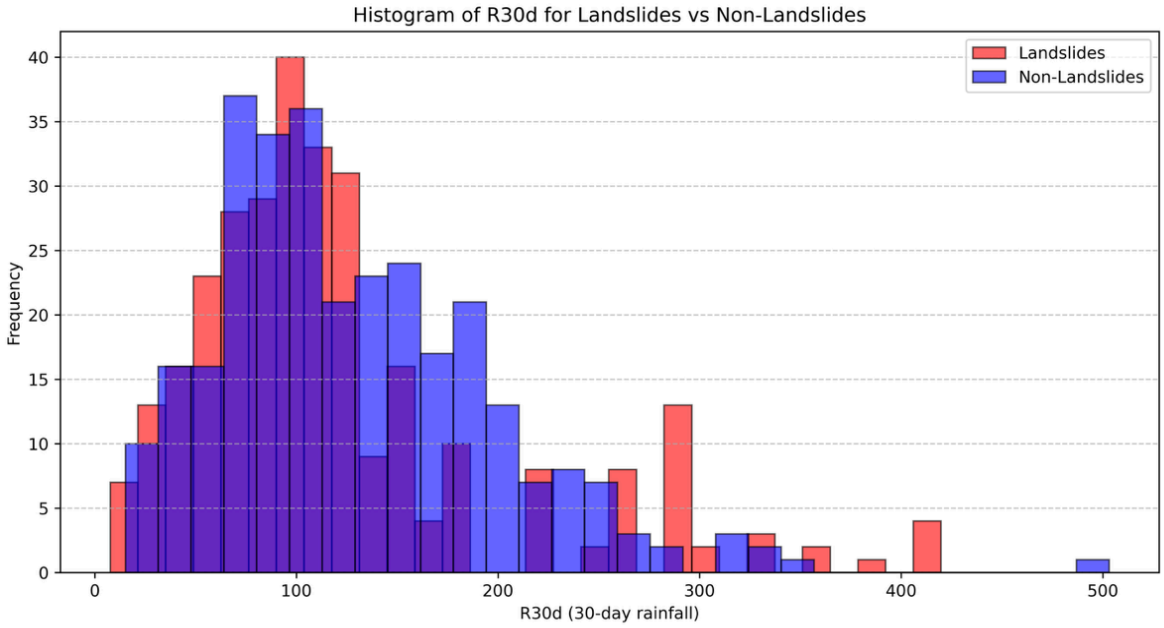


Figure 12: Distribution of 30-Day Cumulative Rainfall (R30d) for Landslide and Non-Landslide Events

Note: The darker bars represent landslide events, and the lighter bars represent non-event controls. All rainfall values are measured in millimeters (mm).

As shown in the R30d histogram (Figure 13), both distributions of rainfall values were comparable, confirming that rainfall data were similarly distributed across both categories, thus strengthening the model's credibility. These findings underscore the capability of Random Forest to capture both anticipated and unexpected dynamics regarding landslide susceptibility. Intriguing is the discovery of the negative correlation between long-term cumulative rains and the model's predicted

likelihood. This suggests that long-term rain accumulation, when not accompanied by brief extremes, may actually contribute to stabilizing the slope by facilitating gradual infiltration and percolation before pore pressurization persists long enough to initiate brief shallow failures. The results derived match past works citing landslide behavior on highly stepped landscapes as being mostly due to short-term high-magnitude events instead of being entirely the product of prolonged wet spells (e.g., Crozier, 2010; Tiranti & Rabuffetti, 2010; Bogaard & Greco, 2018). In addition, the decreased performance for F2 implies that soil depth had little bearing on enhancing predictive performance at the particular resolution utilized in this study. In conclusion, the results from applying the Random Forest model indicate that the landslide hazard in Buncombe is primarily caused by a combination of a highly stepped landscape on slope discontinuities and short periods of increased rainfall, rather than long periods of rainfall acting as an instantaneous trigger.

Table 7: XGBoost model performance across feature sets (F0, F1, F2).

FeatureSet	Accuracy	Precision	Recall	F1 Score	ROC_AUC	Brier
F0	0.711	0.719	0.683	0.701	0.814	0.179
F1	0.802	0.757	0.883	0.815	0.879	0.141
F2	0.802	0.765	0.867	0.813	0.885	0.138

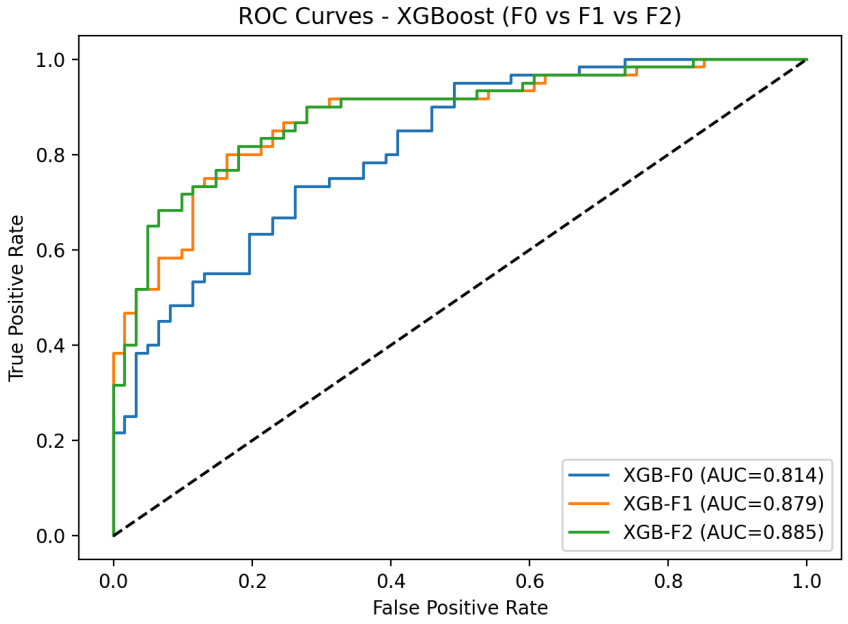
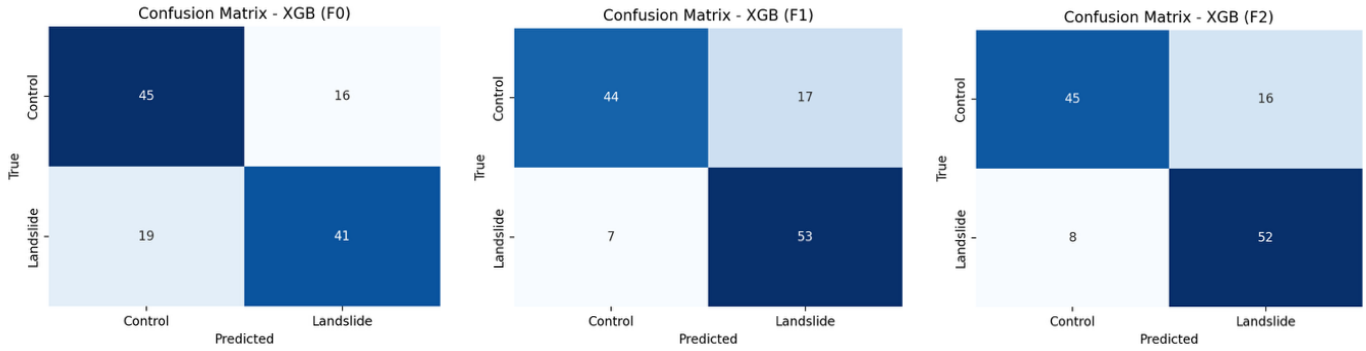


Figure 14: ROC curves of the three XGBoost models, F0, F1, and F2

Note: Colors may be difficult to distinguish for some readers; interpretation should rely on line style and the AUC values.



Figures 15–17: Confusion matrices of the three XGBoost models F0, F1, and F2

Surprisingly, XGBoost did not outperform the random forest model—instead, its F1 and F2 models showed similar prediction accuracy to the Random Forest Model (Table 6; Figure 14). The F0 XGBoost architecture excelled in model strength through an accuracy of 0.74 and an AUC of 0.81 in comparison to the other two F0 methods: Logistic Regression (LR) (accuracy 0.62, AUC 0.67) and Random Forest (RF) (accuracy 0.66, AUC 0.74). As expected, adding rainfall features improved the accuracy to 0.80 and AUC to 0.88, still better than the LR model, but similar to the performance of the RF model (accuracy: 0.81, AUC: 0.89). Likewise, these trends appear in the full model (F2) equivalently. The full model (F2), which incorporated soil depth, produced the highest overall performance with an accuracy of 0.82 and an AUC of 0.89. The confusion matrix for F2 (Figure 17) highlights its balance, correctly classifying 45 controls and 52 landslides.

Compared to Logistic Regression and Random Forest, XGBoost delivered higher accuracy and consistently stronger AUC across feature sets, confirming its robustness in handling complex nonlinear relationships. The results reinforce that steep slopes and short-term rainfall extremes are primary triggers, while long-term rainfall can act as a dampening factor rather than a direct trigger (Crozier, 2010; Bogaard & Greco, 2018). The improved performance of the full F2 model also suggests that soil depth, while secondary to slope, provides additional predictive power in boosted frameworks.

Accuracy improved across all models when rainfall was added to terrain predictors. Logistic Regression rose modestly from 0.62 (F0) to 0.70 (F2), Random Forest increased more sharply from 0.66 (F0) to 0.81 (F1) but dropped slightly with soil depth (0.79, F2), while XGBoost achieved the highest accuracies at every stage, from 0.74 (F0) to 0.82 (F2).

4. Conclusion

This study developed a localized near-real-time landslide prediction framework for Buncombe County, North Carolina, using machine learning models and environmental predictors. By compiling a balanced dataset of 302 landslides and 302 non-events, rainfall windows from PRISM, topography from USGS DEMs, and soil depth from SSURGO, we trained Logistic Regression, Random Forest, and XGBoost classifiers. The pipeline incorporated stratified and spatial-block cross-validation, allowing us to rigorously compare models while preserving spatial representativeness. SHAP analysis and intrinsic feature

importance made it possible to understand how predictors shaped landslide susceptibility in this mountainous region.

Across models, Random Forest consistently achieved the highest predictive performance (accuracy ≈ 0.82, AUC ≈ 0.89), followed by XGBoost and Logistic Regression. In all frameworks, slope emerged as the dominant predictor, confirming prior work showing terrain steepness as the primary control on landslide occurrence (Zheng et al., 2025; Xiao et al., 2024). Short-term rainfall accumulations (1–7 days, maxima over 3 days) were strongly associated with failures, consistent with global findings that intense short-duration rainfall is the principal trigger for shallow landslides (Crosta, 2004; Pennington et al., 2014; Barthélemy et al., 2024). A notable theoretical contribution was the identification of long-term cumulative rainfall (R30d) as a negative predictor in several models, suggesting that extended wet periods may sometimes stabilize slopes by allowing gradual drainage and infiltration rather than immediate pore-pressure buildup (Crosta, 2004; Fan et al., 2020). Soil depth contributed marginally, adding value only in boosted frameworks, which echoes recent findings that subsurface factors are often secondary to slope–rainfall interactions (Lee et al., 2023).

These results underscore both the practical and scientific value of localized machine-learning–based landslide modeling. The framework demonstrates that integrating rainfall windows with terrain predictors can generate near-real-time susceptibility maps, offering actionable information for hazard managers after storms like Hurricane Helene. At the same time, the findings reinforce a broader theoretical consensus: steep slopes combined with short-term rainfall extremes are the key drivers of shallow failures, while long-term rainfall plays a preparatory but not directly triggering role (Pennington et al., 2014; Gariano & Guzzetti, 2016). Limitations of this study include a simplified binary soil depth representation, a relatively small sample size, and a lack of landslide-type differentiation. A limitation of this study is the use of PRISM daily rainfall, which cannot capture short-duration, high-intensity bursts that often trigger shallow landslides. These sub-daily intensities (e.g., mm/hr) are important in operational thresholds, but they are smoothed in daily grids. Future work could incorporate higher-temporal-resolution rainfall from radar products such as MRMS or satellite–gauge datasets like IMERG to better capture intense triggering events. Because the current framework is based on observed rainfall up to the event date, it functions as a near-nowcast rather than a true lead-time forecast; integrating short-term precipitation forecasts would be necessary to produce actionable early-warning lead times. Ultimately, coupling such models with rainfall forecasts could enable truly real-time landslide "nowcasting," a direction increasingly emphasized in both local and global hazard research (Stanley et al., 2021; Khan et al., 2022).

5. References

Akosah, S., Gratchev, I., Kim, D.-H., & Ohn, S.-Y. (2024). Application of artificial intelligence and remote sensing for landslide detection and prediction: Systematic review. *Remote Sensing*, *16*(16), Article 2947. https://doi.org/10.3390/rs16162947

Allstadt, K. E., McBride, S. K., Godt, J. W., Slaughter, S. L., Baxstrom, K. W., Sobieszczyk, S., & Stull, A. (2025). *Preliminary field report of landslide hazards following Hurricane Helene* (Open-File Report 2025–1028). U.S. Geological Survey. https://doi.org/10.3133/ofr20251028

Aydin, A. (2006). Stability of saprolitic slopes: Nature and role of field scale heterogeneities. *Natural Hazards and Earth System Sciences*, *6*(1), 89–96. https://doi.org/10.5194/nhess-6-89-2006

Barthélemy, S., Bernardie, S., & Grandjean, G. (2025). Assessing rainfall threshold for shallow landslides triggering: A case

study in the Alpes Maritimes region, France. *Natural Hazards*, 121(4), 4023–4049. https://doi.org/10.1007/s11069-024-06941-2

Bauer, J., Fuemmeler, S., Wooten, R., Witt, A., Gillon, K., Douglas, T., Eberhardt, E., Froese, C., Turner, K., & Lerouell, S. (2012). Landslide hazard mapping in North Carolina–Overview and improvements to the program. In *Landslides and engineered slopes: Protecting society through improved understanding. 11th International Symposium on Landslides and 2nd North American Symposium on Landslides* (pp. 257–263). https://www.researchgate.net/publication/260077615

Bogaard, T., & Greco, R. (2018). Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: Proposing hydro-meteorological thresholds. *Natural Hazards and Earth System Sciences*, 18(1), 31–39. https://doi.org/10.5194/nhess-18-31-2018

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Burrough, P., & McDonnell, R. (1998). *Principles of geographic information systems*. https://www.researchgate.net/publication/37419765

Chan, H.-C., Chen, P.-A., & Lee, J.-T. (2018). Rainfall-induced landslide susceptibility using a rainfall–runoff model and logistic regression. *Water*, 10(10), Article 1354. https://doi.org/10.3390/w10101354

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Crozier, M. J. (2010). Deciphering the effect of climate change on landslide activity: A review. *Geomorphology*, 124(3–4), 260–267. https://doi.org/10.1016/j.geomorph.2010.04.009

Fan, L., Lehmann, P., Zheng, C., & Or, D. (2020). Rainfall intensity temporal patterns affect shallow landslide triggering and hazard evolution. *Geophysical Research Letters*, 47(1), Article e2019GL085994. https://doi.org/10.1029/2019GL085994

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, 18(8), 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Fuemmeler, S. J. (2008, April 10). *Landslide hazard mapping methodology and examples from North Carolina* [Conference presentation]. The Geological Society of America: Southeastern Section - 57th Annual Meeting, Charlotte, NC, United States. https://gsa.confex.com/gsa/2008SE/webprogram/Paper136714.html

Gariano, S. L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, 162, 227–252. https://doi.org/10.1016/j.earscirev.2016.08.011

Hatcher, R. D. (2010). The Appalachian orogen: A brief summary. In R. P. Tollo, M. J. Bartholomew, J. P. Hibbard, & P. M. Karabinos (Eds.), *From Rodinia to Pangea: The lithotectonic record of the Appalachian region* (pp. 1–19). Geological Society of America. https://doi.org/10.1130/2010.1206(01)

Horn, B. K. P. (1981). Hill shading and the reflectance map. *Proceedings of the IEEE*, 69(1), 14–47. https://doi.org/10.1109/PROC.1981.11918

Jaboyedoff, M., Michoud, C., Derron, M.-H., Voumard, J., Leibundgut, G., Sudmeier-Rieux, K., Nadim, F., & Leroi, E. (2016). Human-induced landslides: Toward the analysis of anthropogenic changes of the slope environment. In S. Aversa, L. Cascini, L. Picarelli, & C. Scavia (Eds.), *Landslides and engineered slopes. Experience, theory and practice* (pp. 217–232). CRC Press. https://doi.org/10.1201/b21520

Kang, J., Wan, B., Gao, Z., Zhou, S., Chen, H., & Shen, H. (2024). Research on machine learning forecasting and early warning model for rainfall-induced landslides in Yunnan province. *Scientific Reports*, 14(1), Article 14049. https://doi.org/10.1038/s41598-024-64679-0

Khan, S., Kirschbaum, D. B., Stanley, T. A., Amatya, P. M., & Emberson, R. A. (2022). Global Landslide Forecasting System for hazard assessment and situational awareness. *Frontiers in Earth Science*, 10, Article 878996. https://doi.org/10.3389/feart.2022.878996

Khashchevskaya, D., Owen, L. A., Wegmann, K., Scheip, C., & Figueiredo, P. M. (2025). The characteristics and timing of multiphase major landslides along the Blue Ridge Escarpment of the southern Appalachians revealed by combined cosmogenic nuclide dating and Schmidt hammer rebound measurements. *Geomorphology*, 485, Article 109857. https://doi.org/10.1016/j.geomorph.2025.109857

Kirschbaum, D. B., Adler, R., Hong, Y., Kumar, S., Peters-Lidard, C., & Lerner-Lam, A. (2012). Advances in landslide nowcasting: Evaluation of a global and regional modeling approach. *Environmental Earth Sciences*, 66(6), 1683–1696. https://doi.org/10.1007/s12665-011-0990-3

Kirschbaum, D., & Stanley, T. (2018). Satellite-based assessment of rainfall-triggered landslide hazard for situational awareness. *Earth's Future*, 6(3), 505–523. https://doi.org/10.1002/2017EF000715

Klose, M., Damm, B., & Terhorst, B. (2015). Landslide cost modeling for transportation infrastructures: A methodological approach. *Landslides*, 12(2), 321–334. https://doi.org/10.1007/s10346-014-0481-1

Klose, M., Maurischat, P., & Damm, B. (2016). Landslide impacts in Germany: A historical and socioeconomic perspective. *Landslides*, 13(1), 183–199. https://doi.org/10.1007/s10346-015-0643-9

Lee, S., Oh, S., Ray, R. L., Lee, Y., & Choi, M. (2023). Three-dimensional hydrological thresholds to predict shallow landslides. *Terrestrial, Atmospheric and Oceanic Sciences*, 34(1), Article 20. https://doi.org/10.1007/s44195-023-00052-4

Lin, S., Chen, S., Rasanen, R. A., Zhao, Q., Chavan, V., Tang, W., Shanmugam, N., Allan, C., Braxtan, N., & Diemer, J. (2024). Landslide prediction validation in Western North Carolina after Hurricane Helene. *Geotechnics*, 4(4), 1259–1281. https://doi.org/10.3390/geotechnics4040064

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic information science and systems* (4th ed.). Wiley.

Lundberg, S., & Lee, S.-I. (2017). *A unified approach to interpreting model predictions* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1705.07874

Mondini, A. C., Guzzetti, F., & Melillo, M. (2023). Deep learning forecast of rainfall-induced shallow landslides. *Nature Communications*, 14(1), Article 2466. https://doi.org/10.1038/s41467-023-38135-y

NASA Earth Observatory. (2021, June 10). *Machine learning model doubles accuracy of global landslide 'nowcasts'*. NASA. https://www.nasa.gov/missions/gpm/machine-learning-model-doubles-accuracy-of-global-landslide-nowcasts/

National Hurricane Center. (2025). *Tropical cyclone report: Hurricane Helene* (AL092024). National Oceanic and Atmospheric Administration. https://www.nhc.noaa.gov/data/tcr/AL092024_Helene.pdf

North Carolina Department of Environmental Quality. (2024). *North Carolina landslide inventory points* [Dataset]. NC OneMap. https://www.nconemap.gov/datasets/01965a193482438cb70332e5e524e38b_0/about

Office of State Budget and Management. (2024). *Hurricane Helene damage and needs assessment*. State of North Carolina. https://www.osbm.nc.gov/hurricane-helene-dna/open

Pennington, C., Dijkstra, T., Lark, M., Dashwood, C., Harrison, A., & Freeborough, K. (2014). Antecedent precipitation as a potential proxy for landslide incidence in South West United Kingdom. In K. Sassa, P. Canuti, & Y. Yin (Eds.), *Landslide science for a safer geoenvironment* (pp. 253–259). Springer. https://doi.org/10.1007/978-3-319-04999-1_34

Petley, D. (2012). Global patterns of loss of life from landslides. *Geology*, 40(10), 927–930. https://doi.org/10.1130/G33217.1

Petley, D. (2025, January 6). Global fatal landslides in 2024. *Eos*. https://eos.org/thelandslideblog/fatal-landslides-in-2024

PRISM Climate Group. (2024). *PRISM daily precipitation data (1981–present)* [Dataset]. Oregon State University. https://prism.oregonstate.edu/

Regorda, A., Lardeaux, J.-M., Roda, M., Marotta, A. M., & Spalla, M. I. (2020). How many subductions in the Variscan orogeny? Insights from numerical models. *Geoscience Frontiers*, 11(3), 1025–1052. https://doi.org/10.1016/j.gsf.2019.10.005

Sidle, R. C., & Ochiai, H. (2006). *Landslides: Processes, prediction, and land use* (Vol. 18). American Geophysical Union. https://doi.org/10.1029/WM018

Sim, K. B., Lee, M. L., Remenyte-Prescott, R., & Wong, S. Y. (2022). An overview of causes of landslides and their impact on transport networks. In *Advances in modelling to improve network resilience* (pp. 119–156). Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/c81e8bc9-1469-11ed-8fa0-01aa75ed71a1

Stanley, T. A., Kirschbaum, D. B., Benz, G., Emberson, R. A., Amatya, P. M., Medwedeff, W., & Clark, M. K. (2021). Data-driven landslide nowcasting at the global scale. *Frontiers in Earth Science*, 9, Article 640043. https://doi.org/10.3389/feart.2021.640043

Thomas, M. A., Collins, B. D., & Mirus, B. B. (2019). Assessing the feasibility of satellite-based thresholds for hydrologically

driven landsliding. *Water Resources Research*, 55(11), 9006–9023. https://doi.org/10.1029/2019WR025577

Tiranti, D., & Rabuffetti, D. (2010). Estimation of rainfall thresholds triggering shallow landslides for an operational warning system implementation. *Landslides*, 7(4), 471–481. https://doi.org/10.1007/s10346-010-0198-8

U.S. Census Bureau. (2022). *Cartographic boundary shapefiles – Counties* [Dataset]. https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html

U.S. Department of Agriculture, Natural Resources Conservation Service. (2024). *Soil Survey Geographic (SSURGO) database* [Dataset]. https://websoilsurvey.nrcs.usda.gov/app/

U.S. Geological Survey. (2021). *National Land Cover Database (NLCD) 2021* [Dataset]. https://www.usgs.gov/centers/eros/science/national-land-cover-database

U.S. Geological Survey. (2022). *3D Elevation Program (3DEP), 1/3 arc-second DEM seamless products* [Dataset]. https://www.sciencebase.gov/catalog/item/627f3798d34e3bef0c9a3198

Watterson, N. A., & Jones, J. A. (2006). Flood and debris flow interactions with roads promote the invasion of exotic plants along steep mountain streams, Western Oregon. *Geomorphology*, 78(1–2), 107–123. https://doi.org/10.1016/j.geomorph.2006.01.019

Wooten, R. M., Gillon, K. A., Witt, A. C., Latham, R. S., Douglas, T. J., Bauer, J. B., Fuemmeler, S. J., & Lee, L. G. (2008). Geologic, geomorphic, and meteorological aspects of debris flows triggered by Hurricanes Frances and Ivan during September 2004 in the Southern Appalachian Mountains of Macon County, North Carolina (southeastern USA). *Landslides*, 5(1), 31–44. https://doi.org/10.1007/s10346-007-0109-9

Xiao, X., Zou, Y., Huang, J., Luo, X., Yang, L., Li, M., Yang, P., Ji, X., & Li, Y. (2024). An interpretable model for landslide susceptibility assessment based on Optuna hyperparameter optimization and random forest. *Geomatics, Natural Hazards and Risk*, 15(1), Article 2347421. https://doi.org/10.1080/19475705.2024.2347421

Ye, C., Wu, H., Oguchi, T., Tang, Y., Pei, X., & Wu, Y. (2025). Physically based and data-driven models for landslide susceptibility assessment: Principles, applications, and challenges. *Remote Sensing*, 17(13), Article 2280. https://doi.org/10.3390/rs17132280

Zheng, W., Fan, W., Cao, Y., Nan, Y., & Jing, P. (2025). Landslide hazard assessment under record-breaking extreme rainfall: Integration of SBAS-InSAR and machine learning models. *Remote Sensing*, 17(13), Article 2265. https://doi.org/10.3390/rs17132265

Acknowledgements & Mentor Contribution Statement

Dr. Janine Haugh provided valuable guidance during the formative stages of this research project. Her mentorship was instrumental in helping refine the initial research direction, shape the core questions, and clarify the conceptual framework connecting hydrologic thresholds, soil characteristics, and geomorphology. Through early discussions, she offered thoughtful feedback on how to situate the study within broader landslide-risk research and helped the author identify a focused and meaningful gap in existing modeling approaches.

Dr. Haugh also advised on interpreting long-term rainfall patterns and incorporating terrain-based predictors, ensuring that the project maintained coherence between its scientific objectives and methodological decisions. Her insights strengthened the logical flow of the manuscript, particularly in the development of the introduction and the explanation of the environmental context.

Dr. Georgios Aivaliotis contributed additional guidance related to data acquisition and modeling design. His expertise in environmental prediction and machine-learning applications helped refine the selection and processing of key datasets, including PRISM rainfall windows, SSURGO soil depth, and the North Carolina landslide inventory. His feedback supported clearer alignment between the statistical methods and the goals of real-time hazard assessment.

The mentors' roles were limited to conceptual and structural guidance; all data processing, analysis, coding, interpretation, and writing were conducted independently by the author. Throughout the process, their support encouraged methodological precision, deeper critical thinking, and a clearer connection between the research objectives and final findings, providing an invaluable foundation for the study's development.

Author Biography

Wonjun Choi is a student researcher at Asheville School in North Carolina, where he focuses on data-driven modeling of rainfall-triggered landslides and the development of real-time hazard prediction systems for mountainous regions in Western North Carolina. His most recent work, "A Dynamic Risk Prediction Model for Rainfall-Triggered Landslides in Buncombe County, North Carolina," integrates PRISM rainfall data, USGS Digital Elevation Models, and SSURGO soil depth information using machine-learning frameworks such as Logistic Regression, Random Forest, and XGBoost. The study explores how dynamic rainfall windows and antecedent soil-moisture conditions shape landslide susceptibility, offering a near-nowcasting approach for operational hazard monitoring.

Beyond this project, Wonjun has engaged in a broad range of quantitative and technical pursuits, including robotics programming, geospatial modeling, and environmental data analysis. His previous work includes coding real-time rainfall ingestion pipelines, building geospatial visualizations, and assisting peers in advanced mathematics and computer science as a tutor and leader.

Wonjun's academic interests extend across environmental engineering, machine learning, and earth-systems science. He hopes to continue developing applied data-science tools that improve community resilience, especially in regions facing increased hydrological extremes due to climate change.

AI Use Statement

AI tools (ChatGPT) were used solely for grammar and formatting edits; all analysis and interpretation are original to the author.