Technical Report Documentation Page

<table>
<tr><td colspan="2">1. Report No.</td><td>2. Government Accession No.</td><td colspan="2">3. Recipient's Catalog No.</td></tr>
<tr><td colspan="4">4. Title and Subtitle
Project #578 - UTC Safety21: Improve Transportation Infrastructure (Geotechnical Asset) Safety by Reducing the Risks of Landslides - Phase 3</td><td>5. Report Date
August 31, 2026</td></tr>
<tr><td colspan="4">7. Author(s)
Zhuping Sheng, Ph.D. (https://orcid.org/0000-0001-8533-527x)
Oludare Owolabi, D.Sc. (https://orcid.org/0000-0002-1958-6430)
Yi Liu, D. Eng. (https://orcid.org/0000-0002-6861-2914)
Kofi Nyarko, Ph.D. (https://orcid.org/0000-0002-7481-5080)</td><td>6. Performing Organization Code

8. Performing Organization Report No.</td></tr>
<tr><td colspan="4">9. Performing Organization Name and Address
Department of Civil and Environmental Engineering
Morgan State University
1700 E. Cold Spring Lane
Baltimore, MD 21251</td><td>10. Work Unit No.

11. Contract or Grant No.
Federal Grant #
69A3552344811 /
69A3552348316</td></tr>
<tr><td colspan="4">12. Sponsoring Agency Name and Address
Safety21 University Transportation Center
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213</td><td>13. Type of Report and Period Covered
Final Report
July 1, 2025-June 30, 2026

14. Sponsoring Agency Code
USDOT</td></tr>
<tr><td colspan="5">15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.</td></tr>
<tr><td colspan="5">16. Abstract
This report summarizes Phase 3 research findings of the project Improve Transportation Infrastructure (Geotechnical Asset) Safety by Reducing the Risks of Landslides, sponsored by the National UTC Safety21 program. The work advances an integrated framework for detecting and monitoring slope instability and assessing landslide risk along highways, railroads, and roadways. The studies include geometry-resolved InSAR monitoring and directional forecasting at Haystack Mountain; intervention-aware InSAR and multi-horizon machine learning for the Baltimore CSX Belt Line; a simulation-based CaseBank and machine-learning framework for rapid slope-stability assessment; soil-specific probabilistic rainfall thresholds for landslide initiation in Maryland; physics-informed multi-sensor machine-learning/deep-learning benchmarking and LiDAR monitoring for early warning; laboratory characterization of soils from Maryland landslide-inventory sites; and a national railroad landslide-exposure inventory developed through GIS-based proximity screening.</td></tr>
<tr><td colspan="4">17. Key Words
Landslides, Slope Stability, InSAR, LiDAR, Machine Learning, Rainfall Thresholds, Soil Characterization, Early Warning, Railroad Exposure, GIS</td><td>18. Distribution Statement
No restrictions.</td></tr>
<tr><td>19. Security Classif. (of this report)
NONE</td><td colspan="2">20. Security Classif. (of this page)
NONE</td><td>21. No. of Pages
130</td><td>22. Price</td></tr>
<tr><td colspan="3">Form DOT F 1700.7 (8-72)</td><td colspan="2">Reproduction of completed page authorized</td></tr>
</table>

US DOT National
University Transportation Center for Safety

UTC Safety21: Improve Transportation Infrastructure (Geotechnical Asset) Safety by Reducing the Risks of Landslides - Phase 3

Zhuping Sheng, Ph.D. (https://orcid.org/0000-0001-8533-527x)

Oludare Owolabi, D.Sc. (https://orcid.org/0000-0002-1958-6430)

Yi Liu, D. Eng. (https://orcid.org/0000-0002-6861-2914)

Kofi Nyarko, Ph.D. (https://orcid.org/0000-0002-7481-5080)

FINAL RESEARCH REPORT – August 31, 2026

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated in the interest of information exchange. The report is funded, partially or entirely, under [grant number 69A3552344811 / 69A3552348316] from the U.S. Department of Transportation's University Transportation Centers Program. The U.S. Government assumes no liability for the contents or use thereof.

Table of Contents

Chapter 1

1 Introduction

Geologic hazards, including slope failures, landslides, mudflows, and debris flows, and hydrological hazards related to floods and storm surge can damage transportation infrastructure and threaten property and human life along highways, railroads, and roadways. Landslides alone cause thousands of deaths and billions of dollars in damage worldwide each year. Advancing knowledge of slope-instability and failure risks assessment, together with technologies for detecting, monitoring, and preventing landslides, is therefore important to sustaining the safety and reliability of transportation infrastructure under changing environmental conditions.

As a member of the U.S. Department of Transportation National University Transportation Center (UTC) Safety21 program led by Carnegie Mellon University, the Morgan State University team is conducting a multiphase, interdisciplinary project focused on transportation-system safety through prevention of geohazards, particularly slope failures and landslides, and reduction of their impacts along highways, railroads, and roadways. The project integrates geotechnical engineering, remote sensing (RS), geographic information systems (GIS), and artificial intelligence/machine learning to assess geotechnical assets, characterize slope conditions, delineate landslides and exposed corridors, and support risk-informed monitoring and management.

This report summarizes the Phase 3 research findings of the project titled UTC Safety21: Improve Transportation Infrastructure (Geotechnical Asset) Safety by Reducing the Risks of Landslides - Phase 3. Following this introduction, Chapter 2 presents geometry-resolved InSAR monitoring and directional forecasting of slope movement at Haystack Mountain along Interstate 68. Chapter 3 introduces an intervention-aware InSAR and multi-horizon machine learning model for an urban rail-corridor slope at the Baltimore CSX Belt Line. Chapter 4 combines a simulation-based CaseBank with machine learning for rapid slope-stability assessment. Chapter 5 presents new soil-specific probabilistic rainfall thresholds for landslide initiation in Maryland. Chapter 6 evaluates physics-informed multi-sensor machine learning/deep learning and LiDAR monitoring framework for rainfall-induced landslide early warning. Chapter 7 reports laboratory characterization and classification of soils collected from Maryland landslide-inventory sites. Chapter 8 shares a national railroad landslide-exposure inventory using GIS-based proximity screening.

1.1 Geological Hazards

Geologic hazards such as landslides, land subsidence, earth fissures, and earthquakes, together with hydrological hazards such as floods and storm surge associated with extreme weather, can disrupt transportation infrastructure and traffic and result in substantial economic losses. Landslides are among the most damaging natural hazards because they can affect road cuts, embankments, retaining systems, bridge approaches, rail corridors, drainage structures, and adjacent communities.

Many landslides are triggered by precipitation, although they occur across a broad range of lithologic, climatic, hydrologic, topographic, and land-use conditions. Atmospheric, surface, and subsurface processes can increase driving forces or reduce the strength of slope materials. The interaction of rainfall, soil moisture, pore-water pressure, terrain geometry, slope material properties, and human modifications must therefore be considered when assessing slope instability and the possible effects of extreme weather and climate variability.

1.2 Updates on the Landslide Detection and Warning Framework

Landslide inventories and susceptibility analyses are commonly used to identify where failures have occurred and where future instability may be more likely to occur. Transportation agencies also manage geotechnical assets such as cut slopes, embankments, retaining walls, ground modifications, and drainage systems. Connecting these asset records with field observations, laboratory tests, remote-sensing products, and environmental data supports a more systematic basis for prioritizing inspection and mitigation of landslides.

The Phase 3 work advances the project framework through complementary scales of analysis. Site-scale InSAR and LiDAR studies characterize deformation and monitoring needs at highway and rail locations. Physics-based simulations, laboratory testing, rainfall analysis, and machine-learning models examine mechanisms and indicators of instability. GIS screening extends the framework to a national railroad network. Together, these elements support a layered workflow in which broad screening identifies candidate areas, site-specific investigations strengthen interpretation, and continued observations provide evidence for monitoring and asset-management decisions.

1.3 Objectives of the Project (Phases 1–3)

The project integrates geotechnical and machine-learning approaches to assess slope instability and landslide risk and to map areas of concern along highways, railroads, and roadways. Its principal objectives are to:

1. Use artificial intelligence and machine-learning approaches to assess landslide risk based on soil and rock types, weather conditions, slope-material properties, terrain geometry, observed deformation, and available information on transportation assets and protective structures.
2. Identify and map areas of concern using controlling factors such as geometry and material properties and triggering factors such as gravitational and hydraulic forces, drawing on field surveys, laboratory testing, remote sensing, including LiDAR and InSAR, GIS, and related transportation data.
3. Develop and test procedures for monitoring selected sites and for interpreting observations in consultation with transportation stakeholders.
4. Support risk-reduction strategies and evidence-based prioritization for higher-concern areas while developing methods that can be transferred to other regions with comparable geologic and engineering conditions.

Phases 1 and 2 established the project foundation through field and laboratory investigation, remote-sensing applications, geotechnical asset-management review, hydrologic and numerical modeling, and preliminary landslide-susceptibility and warning studies [1,2]. Phase 3 continues these objectives by expanding deformation monitoring and forecasting, developing rapid and interpretable modeling tools, evaluating rainfall and sensor indicators, extending laboratory characterization, and screening landslide exposure along railroad infrastructure.

1.4 Alignment with the USDOT Strategic Plan

The project supports transportation safety, infrastructure resilience, and data-informed decision-making. Its activities contribute to proactive identification of emerging physical risks, application of safety-management approaches across transportation modes, and development of analytical tools that can help agencies assess critical infrastructure vulnerabilities.

1. Use data, remote sensing, and analytics to support proactive identification and assessment of safety risks.
2. Improve understanding of infrastructure vulnerability to geohazards and extreme weather and support risk-informed adaptation and mitigation strategies.
3. Advance monitoring, forecasting, and decision-support methods that can strengthen response, recovery, and continuity of transportation systems.
4. Develop transferable research methods and provide educational opportunities that build technical and human capacity in resilient transportation engineering.

The Phase 3 studies remain research and screening activities. Their outputs are intended to support, rather than replace, detailed site investigation, engineering judgment, agency procedures, and validated operational thresholds.

Chapter 2

2 Geometry-Resolved InSAR Monitoring and Directional Forecasting of Residual Slope Activity at Haystack Mountain, Interstate 68, Maryland

Tohid Asheghimehmandari[1], Oluwatunmise Akinniyi[2], Zhuping Sheng[1], and Yi Liu[1]

[1]Department of Civil and Environmental Engineering, Morgan State University, Baltimore, Maryland

[2] Department of Electrical and Computer Engineering, Morgan State University, Baltimore, Maryland

Publication status notice. This chapter summarizes interim project progress for the USDOT Safety21 annual report. Portions of the underlying analysis are being prepared for submission to a peer-reviewed journal. To maintain a distinct archival record, this chapter uses a report-specific structure, independently rewritten narrative, and newly synthesized tables and figures; it is not the journal manuscript of record.

2.1 Introduction

Transportation agencies manage thousands of cut slopes, embankments, and natural hillsides whose condition can change gradually between field inspections. When a slope is located directly above an interstate highway, even slow deformation matters because accumulated movement can damage drainage systems, loosen rock or soil, and reduce the margin separating stable service from a rapid roadway-threatening failure. Conventional reconnaissance, survey monuments, inclinometers, and piezometers remain essential, but they are spatially limited and can be difficult to maintain on steep or inaccessible ground. Satellite radar interferometry offers a complementary, repeatable view of surface movement over the full slope and its surroundings [3-5].

This chapter reports annual progress on satellite-based monitoring of the Haystack Mountain landslide along Interstate 68 (I-68) in Allegany County, Maryland. The north-side cut rises above the westbound lanes just west of Cumberland and has a documented history of recurrent instability, including a rainfall-triggered mud and debris slide in September 2018 and emergency excavation in October 2018 [6]. The site is therefore useful for evaluating a practical transportation question: after a major intervention, can a long radar record distinguish the initial failure response, residual creep, localized reactivation, and apparently quiet sectors without requiring continuous access to the slope?

The analysis uses C-band Sentinel-1 data processed with the Small Baseline Subset interferometric synthetic aperture radar (SBAS-InSAR) method. Radar phase changes were first converted to cumulative displacement along the satellite line of sight (LOS). Because LOS movement alone is controlled by both ground motion and viewing geometry, the series were then projected into a local slope coordinate system. The result is a set of downslope, vertical, and horizontal displacement histories that can be interpreted in terms of settlement and roadward advance. Velocities were evaluated for the full record and for project-relevant time windows spanning the documented 2018-2019 failure sequence, the post-stabilization period, and the most recent two years. A rolling STL-ARIMA-LSTM workflow was then evaluated for next-acquisition LOS forecasting and analytical transfer to the same directional components.

The chapter is intentionally framed as a progress report rather than as an archival research article. Its purpose is to document completed tasks, summarize the evidence available to the project team, identify quality-control limitations, and translate the results into monitoring priorities for transportation asset management. The analysis is not presented as a stand-alone alarm system and does not replace geotechnical investigation or field instrumentation.

The objectives completed during this reporting period included:

1. assemble and process a multi-year Sentinel-1 ascending-track record for the Haystack Mountain corridor;
2. extract consistent LOS displacement histories at six slope points and two low-relief benchmark locations;
3. transform LOS displacement to downslope motion and resolve vertical settlement and horizontal advance using DEM and satellite geometry;
4. compare cumulative movement and velocity across failure-era, post-stabilization, and recent windows;
5. evaluate chronological rolling-window LOS and directional forecast performance across P1-P6; and
6. identify residual activity, uncertainty, and the field-verification needs most relevant to continued I-68 slope management.

2.2 Literature Review

Synthetic Aperture Radar (SAR): SAR measures the amplitude and phase of microwave energy returned from the ground. Interferometric processing compares phase between repeat acquisitions so that small changes in sensor-to-ground distance can be detected over broad areas [7-9]. The Sentinel-1 mission was designed for systematic, all-weather Earth observation and provides an especially valuable archive for infrastructure monitoring because its acquisition geometry and revisit pattern support consistent multi-temporal analysis [7]. Individual differential interferograms, however, are sensitive to temporal decorrelation, topographic error, atmospheric delay, and orbit uncertainty. These limitations motivated multi-temporal approaches that combine many interferograms rather than relying on one image pair.

Two major multi-temporal families are Persistent Scatterer Interferometry (PSI) and SBAS. PSI emphasizes radar targets that remain phase-stable through time, while SBAS builds a network from image pairs separated by limited temporal and perpendicular baselines [7-11]. SBAS is well suited to slopes where coherence is spatially heterogeneous and where the analyst seeks a distributed time series rather than only a sparse set of persistent targets. Its singular-value-decomposition inversion estimates cumulative deformation through the interferogram network and can reduce the influence of individual low-quality pairs [10, 11]. Reviews of landslide applications show that these methods are effective for inventory refinement, slow-movement detection, and long-term kinematic assessment, but also emphasize the effects of vegetation, snow, steep relief, and spatially variable atmospheric delay [12-16].

Line of Sight (LOS): A central interpretation problem is that a single SAR viewing geometry measures only the LOS component of the three-dimensional displacement vector. Negative and positive LOS changes do not by themselves identify whether the slope settled vertically, advanced horizontally, or moved obliquely relative to the radar. Combining ascending and descending tracks can improve component separation, but data availability, vegetation, and geometry do not always permit a stable multi-track solution. A commonly used alternative for translational landslides is to assume that movement follows the local line of steepest descent and to project LOS displacement onto that direction using the radar incidence and heading angles together with DEM-derived slope and aspect [17-19]. This produces a physically interpretable one-dimensional downslope series, provided that the assumed kinematics are reasonable.

Slope Movement: The slope-parallel assumption is most defensible for shallow translational movement on a persistent basal plane and less defensible for rotational slides, lateral spreading, or compound mechanisms. At Haystack Mountain, the reported fissile shale dips generally toward I-68 and supports a dominant down-dip translational interpretation [3, 20-21]. Even so, the projection does not create new independent information: it rescales the LOS observation by a geometry coefficient. When that coefficient is small, measurement noise and atmospheric artifacts are also amplified. A responsible implementation therefore reports the coefficient, checks reference locations, and avoids treating the projected series as a fully resolved three-dimensional solution.

Observation to Action: For transportation operations, displacement magnitude alone is not sufficient. Asset managers need to know whether movement is persistent or episodic, whether a stabilized area is slowing, and which locations warrant field inspection or drainage maintenance. Period-based velocity summaries can provide that bridge, especially when the windows are tied to known events and intervention dates. The landslide early-warning literature also shows, however, that fixed velocity classes are not universal and that alarm thresholds should be calibrated against site-specific failure behavior and corroborating observations [22-24]. The present chapter therefore uses velocity as a screening indicator, not as an autonomous warning threshold.

Forecasting adds a further requirement: the validation protocol must reflect the order in which observations become available. STL can isolate evolving trend and seasonal structure, ARIMA can represent locally persistent trend dynamics, and LSTM networks can learn nonlinear temporal relationships with environmental covariates [25-28]. Recent work has combined multi-temporal InSAR and deep learning for landslide displacement prediction [29]. Nevertheless, strongly autocorrelated geodetic series can yield optimistic scores if observations are randomly divided or if preprocessing uses information from the test period. Rolling-origin evaluation, explicit baselines, uncertainty tracking, and prospective scoring are therefore central to the highway forecasting analysis.

2.3 Materials and Methodology

2.3.1 Study area, engineering context, and source data

Haystack Mountain lies in the central Appalachian Valley-and-Ridge province immediately west of Cumberland. The monitored cut is on the north side of I-68 near 39.643 degrees N and 78.797 degrees W. The broader mapping identified seven candidate slope points (P1-P7) and two low-relief benchmark areas; the quantitative workbook used in this chapter contains complete, consistently formatted time series for P1-P6 and both benchmarks (Figure 2.1). The six analyzed points range from approximately 295 to 331 m in elevation and occupy slopes of about 17 to 25 degrees with predominantly southeast-facing aspects.

Regional mapping places the site near the contact between the Silurian Rose Hill-Keefer units and the Tuscarora Formation [20, 21]. The unstable zone includes thinly bedded, fissile shale beneath fractured overburden. Site records describe bedding that strikes approximately N55 degrees E and dips toward the roadway, with a local flexural fold and rainfall-sensitive movement along the shale surface [19]. During the September 2018 event, saturated soil and rock moved down dip, and follow-up work included excavation of about 2,580 cubic yards of material together with reinforcement and drainage recommendations [3]. These conditions make a slope-parallel projection physically plausible while also motivating continued checks for local departures from that model.

The engineering history also defines the interpretation windows used later in this chapter. The 2018-2019 interval includes the documented failure, emergency material removal, and early stabilization response; it is not treated as an undisturbed calibration period. The 2020-2025 interval represents the monitored post-intervention condition, but drainage adjustment, vegetation change, freeze-thaw cycling, and progressive weathering could still alter the radar response. This chronology is important because a change in satellite velocity after construction may reflect both improved stability and a changed surface-scattering environment. Consequently, engineering attribution is made only when the kinematic pattern, intervention record, and available field evidence agree.

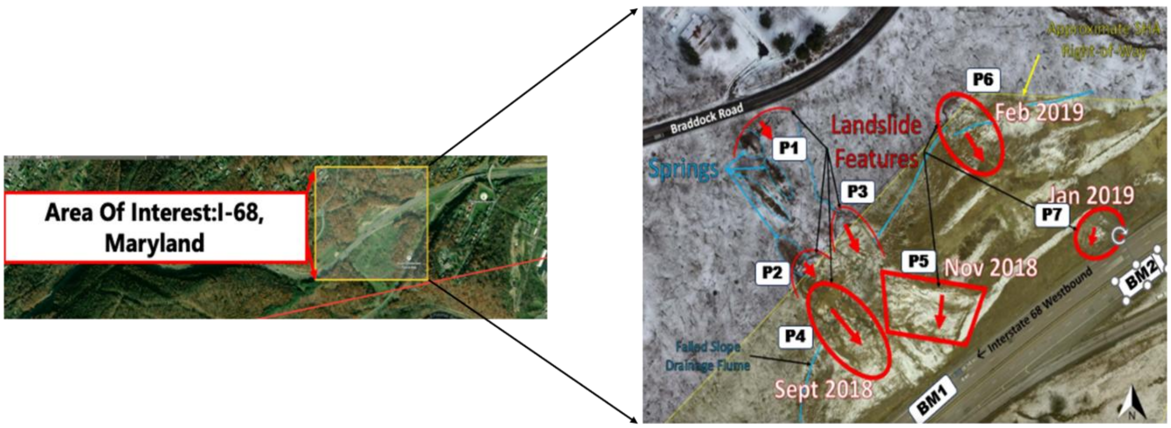


Figure 2.1. Haystack Mountain study area and monitoring layout along I-68. The broader map shows candidate slope locations P1-P7 and benchmarks BM1-BM2; the quantitative analysis in this chapter uses the six slope points P1-P6 for which complete workbook series were supplied (Base map from Maryland DOT/SHA).

Radar inputs consisted of Sentinel-1A/B Interferometric Wide Swath single-look-complex imagery acquired in one ascending geometry, IW1 sub-swath, VV polarization. The processing archive contains more than 2,700 burst-level inputs that were assembled into 251 project time steps from 10 January 2017 through 30 December 2025. The use of a single orbit avoids mixing heading or incidence-angle changes within the time series. A 1 arc-second SRTM DEM supported topographic phase removal and geocoding, while a higher-resolution LiDAR-derived surface was used to evaluate local slope and aspect for the geometric transformation.

2.3.2 SBAS processing and annual-block stitching

Processing followed a standard SBAS chain. Single-look-complex acquisitions were co-registered, suitable small-baseline pairs were selected, and interferograms were generated and multilooked. The modeled topographic phase was removed with the DEM, interferograms were filtered to improve phase continuity, and phase was unwrapped. Temporal-spatial filtering was used to suppress long-wavelength atmospheric and orbital components, after which the interferogram network was inverted to obtain cumulative LOS displacement [10, 11, 19].

The full archive was divided into annual processing blocks to control memory and runtime. Each annual block used a mid-year reference epoch, when snow cover and seasonally elevated soil moisture were less likely to reduce coherence. Independent blocks reset their cumulative origin, so they cannot be concatenated directly. The delivered series were stitched sequentially by adding the accumulated endpoint of the preceding block to the next block. This preserved within-year increments while removing artificial annual resets. Displacement at each point was represented by an area-weighted spatial average rather than a single pixel, reducing sensitivity to one anomalous scatterer.

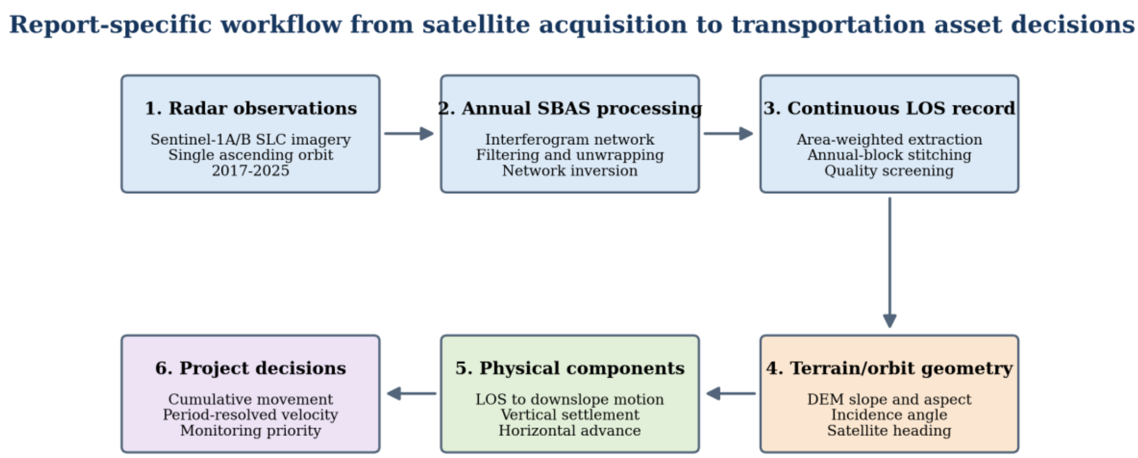


Figure 2.2. Report-specific workflow linking Sentinel-1 observations, annual SBAS processing, LOS time-series stitching, terrain-orbit geometry, component resolution, and transportation asset decisions.

2.3.3 Geometric transformation from LOS to slope components

Four angles define the transformation at each point: local slope angle alpha, downslope aspect beta measured clockwise from north, radar incidence angle theta, and satellite heading angle phi. Slope and aspect were derived from the DEM using a finite-difference terrain-gradient method [22]. Incidence and heading were taken from the Sentinel-1 acquisition metadata. For the supplied ascending geometry, theta was approximately 33.68 degrees and phi approximately -13.43 degrees at all monitoring locations.

Let the ground move parallel to the local line of steepest descent. The LOS observation is then the scalar projection of the downslope displacement, D_slope, onto the radar look vector. With the sign convention used in the project workbook, the geometry coefficient a and the resolved components are:

$$a = -\cos\varphi \sin\theta \cos\alpha \sin\beta + \sin\varphi \sin\theta \cos\alpha \cos\beta - \cos\theta \sin\alpha \qquad (2.1)$$

$$D\_slope\ (t) = D\_LOS(t)/a \qquad (2.2)$$

$$d\_V(t) = D\_slope(t) \sin\alpha \qquad (2.3)$$

$$d\_H(t) = D\_slope(t) \cos\alpha \qquad (2.4)$$

where d_V is the vertical settlement component and d_H is horizontal advance along the downslope azimuth. Negative values in the supplied series denote net motion in the interpreted downslope/settlement direction. The coefficient a is time invariant for a fixed point and orbit. Therefore, the transform preserves the temporal shape of the LOS record while scaling its amplitude by 1/a. It cannot resolve lateral or slope-normal motion that violates the assumed kinematics.

The most influential geometric uncertainty enters through the coefficient a. Small errors in slope or aspect become more important as the LOS becomes less aligned with the assumed motion vector, because both displacement and velocity are multiplied by $1/a$. At Haystack Mountain, the observed range of $1/a$ (1.52-1.95) is moderate, but it is not negligible. The reported components should therefore be read as a geometry-conditioned engineering interpretation rather than a unique three-dimensional solution. A future uncertainty analysis will perturb terrain angles, incidence, heading, and LOS error together, then report percentile intervals for downslope, vertical, and horizontal motion.

Table 2.1. Monitoring-point terrain and acquisition geometry. The projection coefficient a links LOS and downslope displacement; 1/a is the corresponding amplification factor.

Point	Elev. (m)	Slope (deg)	Aspect (deg)	a	1/a
P1	331.3	19.32	148.2	0.6112	1.636
P2	315.7	22.87	144.2	0.5675	1.762
P3	321.0	25.45	164.0	0.5134	1.948
P4	305.8	17.16	156.5	0.6341	1.577
P5	295.2	22.47	126.9	0.6059	1.651
P6	323.7	22.41	108.2	0.6585	1.519
BM1	284.9	4.76	79.1	0.8273	1.209
BM2	275.1	8.42	121.5	0.7659	1.306

2.3.4 Velocity windows, stability screening, and quality control

Two complementary velocity measures were retained. First, an ordinary least-squares trend across the complete 2017-2025 downslope series summarizes net long-term behavior. Second, trend-derived rates were computed over three diagnostic windows: the 2017-2019 failure era, the 2020-2025 post-stabilization era, and the most recent two years. The first two windows compare historically different engineering states; the last window is deliberately more responsive to recent reversals but is also more sensitive to the selected start date and seasonal behavior. A negative rate represents downslope movement and a positive rate represents an apparent recovery or reversal.

Quality control was based on time-series continuity, consistency among nearby slope points, the magnitude of the geometry amplification factor, and the behavior of BM1 and BM2. Because the benchmarks are low-relief radar targets rather than surveyed bedrock monuments, their apparent movement is treated as an empirical bound on residual atmospheric, reference, and processing artifacts. No operational conclusion is based solely on whether a slope-point rate differs from zero. Stronger confidence is assigned when cumulative magnitude, period rate, spatial pattern, and site history point in the same direction.

2.3.5 Rolling directional forecasting and chronological evaluation

The extended progress analysis adds a forecasting layer that is intentionally distinct from the direct multi-horizon XGBoost pilot reported for the rail corridor in Chapter 3. The Haystack Mountain task predicts the next 12-day displacement update at all six slope points and then transfers the forecast through the verified point-specific geometry. The stitched series were resampled to the Sentinel-1 revisit grid, short gaps were interpolated, and rainfall accumulations and rolling temperature extrema were aligned to the same forecast origins.

Seasonal-Trend decomposition using LOESS (STL) separated each LOS history into long-term trend, environmentally modulated seasonal response, and residual variation [27]. The long-term trend was forecast with an autoregressive integrated moving-average model whose order was reselected by information criterion inside each training window [28]. The seasonal component was forecast with a two-layer long short-term memory network using 64 hidden units per layer, 12 lagged steps, dropout of 0.2, Adam optimization, early stopping, and rainfall/temperature covariates [29, 30]. The reconstructed LOS forecast is the sum of the trend and seasonal forecasts. This hybrid structure assigns the slowly evolving internal state to the statistical trend model while allowing the recurrent network to represent nonlinear short-period environmental modulation.

Evaluation used a rolling four-year window. At every forecast origin, preprocessing and model fitting used only observations available before that origin; a one-step forecast was issued before the window advanced. This chronological design is more demanding and more operationally relevant than randomly shuffling observations, which would leak future regimes into training. The project compared the proposed rolling STL-ARIMA-LSTM configuration with fixed-split and non-decomposed baselines. The available ablation result at P1 increases from R2 = 0.8199 for the non-decomposed fixed-split Random Forest baseline to R2 = 0.9834 for the rolling hybrid configuration.

Once LOS is forecast, downslope, vertical, and horizontal predictions follow from the same positive linear factors used for the observed record. Positive scaling leaves R2 unchanged and multiplies RMSE and MAE by the corresponding geometric factor. This analytical transfer avoids training separate component models on quantities that are deterministically derived from LOS, but it also transfers all LOS-model and terrain-geometry uncertainty into the directional products. The forecasting results are therefore presented as project research progress, not as an independently certified early-warning system [31].

2.3.6 Uncertainty budget and reproducibility controls

A report that combines InSAR, geometric projection, and forecasting needs more than one generic error term. Measurement uncertainty includes decorrelation, residual atmosphere, orbit and DEM error, unwrapping, spatial averaging, and reference-frame behavior. Interpretation uncertainty enters when one LOS direction is constrained to slope-parallel motion. Forecast uncertainty includes sampling design, model selection, covariate quality, autocorrelation, regime changes, and the calibration of prediction intervals. These layers affect different decisions and should remain traceable rather than being collapsed into one model score.

Table 2.2 records the safeguards already implemented and the information still needed for a fully reproducible workflow. The most consequential gap is the lack of surveyed reference control: because BM1 and BM2 are radar comparison locations rather than monuments, they constrain apparent background behavior but cannot establish absolute stability. The next processing archive should also retain scene IDs, pair networks, coherence and unwrapping masks, annual reference definitions, DEM provenance, point-neighborhood statistics, and the exact forecast-origin ledger.

This separation of evidence protects the later journal submission as well as the annual report. The annual report documents what was completed and which results are decision-relevant; the journal workflow must add a frozen, versioned processing manifest and uncertainty propagation before making stronger claims. No missing metadata value is reconstructed from memory or inferred from a plotted curve.

Table 2.2. Highway-specific uncertainty and reproducibility controls. The table separates uncertainties in the radar measurement, annual assembly, directional interpretation, and forecasting stages.

Uncertainty source	Potential effect	Current safeguard	Priority completion action
LOS phase and atmosphere	Bias or short-period fluctuations	Multi-temporal inversion and spatial averaging	Export coherence, phase variance, and atmospheric residual fields
Annual block reference	Artificial offset between years	Sequential endpoint stitching	Archive every annual reference date/area and test a common reference
Benchmark behavior	Uncertain absolute velocity baseline	BM1/BM2 retained as empirical bounds	Install surveyed GNSS or a radar corner reflector
Terrain and orbit geometry	Amplified downslope/component error	Point-specific a and 1/a reported	Propagate DEM angle, heading, and incidence uncertainty
Slope-parallel assumption	Missed lateral or slope-normal motion	Components labeled geometry-conditioned	Add descending geometry and field motion-direction evidence
Forecast evaluation	Optimistic skill or miscalibrated intervals	Rolling-origin chronological testing	Freeze prospective forecasts and score coverage before refitting

2.4 Results / Case Study

2.4.1 Acquisition geometry and benchmark behavior

The six analyzed slope points have projection coefficients from 0.513 to 0.658, corresponding to amplification factors from 1.52 to 1.95 (Table 2.1). These values indicate useful sensitivity of the ascending LOS direction to downslope movement, with less than a two-fold amplification at every slope point. P3 has the largest amplification (1.95), so its projected record requires the most caution; P6 has the smallest (1.52). The two low-relief benchmarks have larger a values and smaller amplification factors of 1.21 and 1.31.

The benchmark records are not perfectly invariant (Figure 2.3). Over the full record, apparent linear trends were approximately -12.7 mm/yr at BM1 and -2.7 mm/yr at BM2, with end-of-record offsets of about -75 and -30 mm, respectively. These values may include residual atmospheric structure, reference-frame drift, local ground behavior, or error introduced by applying a slope projection on low-relief terrain. This result is important: it argues against interpreting the lower-magnitude slope-point rates as precise absolute velocities. By contrast, the much larger cumulative and post-stabilization movement at P5, and the sustained movement at P6, remain distinguishable from the smaller benchmark response.

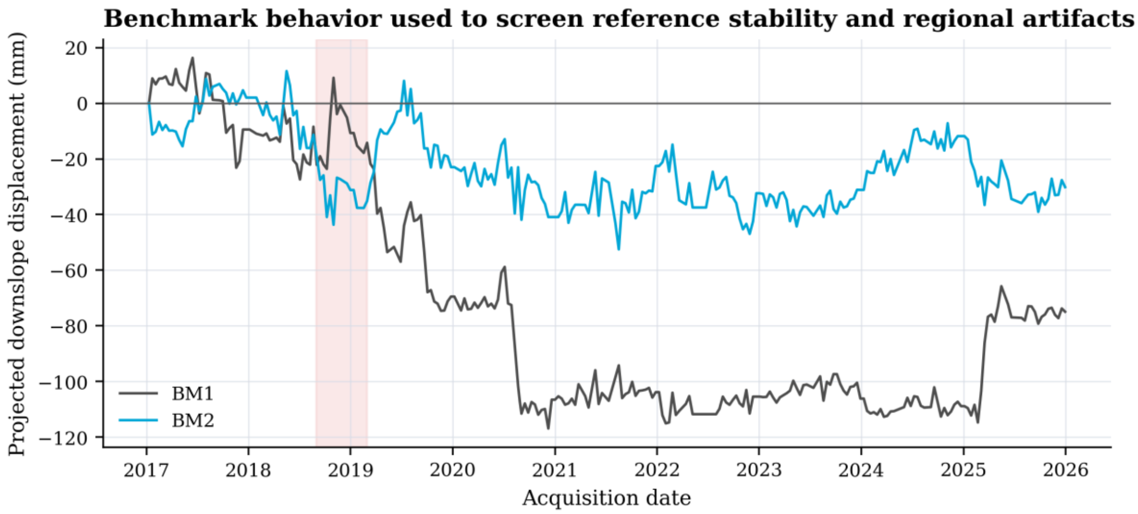


Figure 2.3. Projected downslope series at the two low-relief benchmark locations. Their nonzero apparent trends provide an empirical quality-control bound and reinforce the need for independent field reference control.

2.4.2 Long-term LOS and downslope displacement

The stitched series show strong spatial variability (Figure 2.4). P1 remains near its initial level at the end of the record after several large reversals. In contrast, P2, P3, P4, P5, and P6 finish with negative LOS and downslope offsets. The geometry-resolved record is consistently larger in magnitude than LOS because all slope-point values of a are between zero and one. The curves are not monotonic: multi-month recovery and reversal episodes are superimposed on longer-term trends. This behavior supports a residual-creep interpretation rather than a simple constant-rate model.

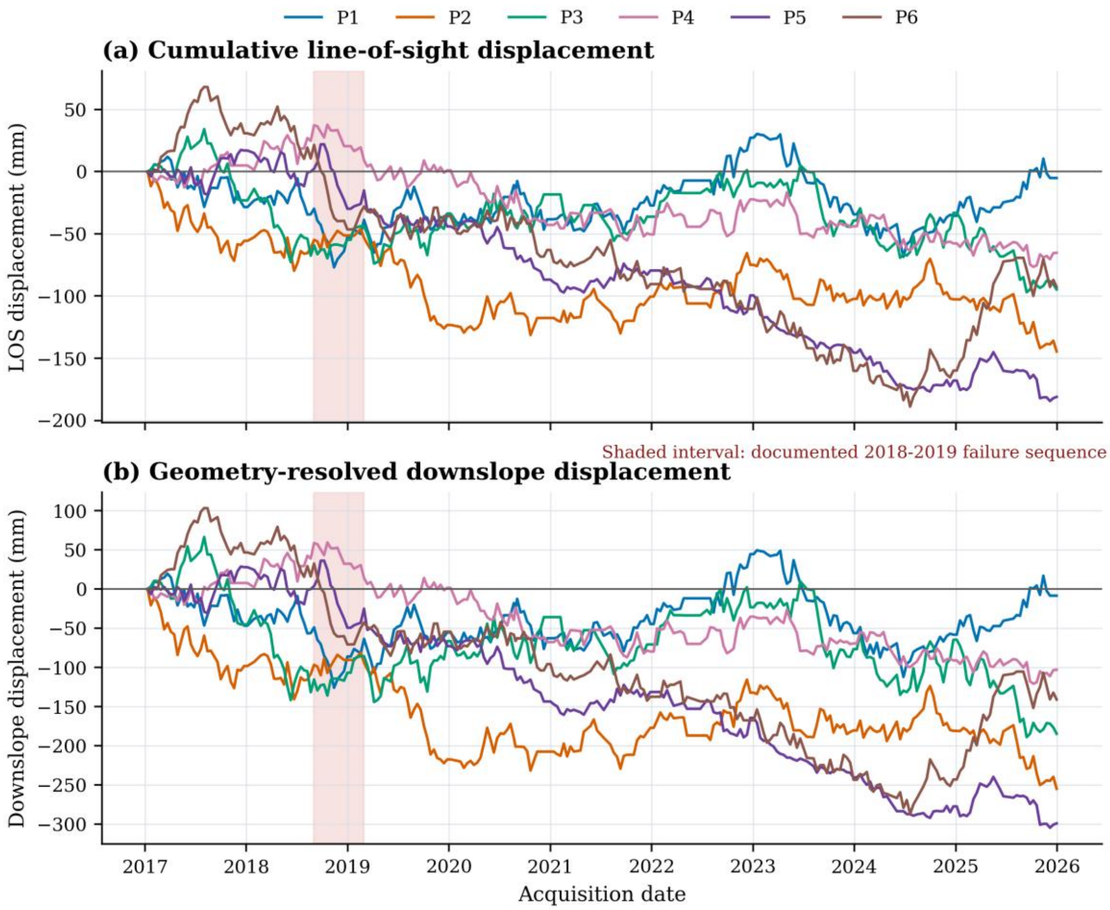


Figure 2.4. Cumulative LOS and geometry-resolved downslope displacement at P1-P6 from January 2017 through December 2025. The shaded interval marks the documented 2018-2019 failure sequence; negative values indicate motion away from the sensor and, after projection, net downslope movement.

End-of-record values quantify this contrast (Table 2.3). P5 has the largest cumulative downslope displacement (-298.9 mm), followed by P2 (-254.8 mm), P3 (-184.5 mm), P6 (-140.8 mm), and P4 (-102.9 mm). P1 ends at only -8.4 mm. Linear regression across the complete downslope series gives the highest-magnitude mean rates at P5 (-38.1 mm/yr) and P6 (-33.0 mm/yr). P4 averages -14.5 mm/yr, P2 -12.6 mm/yr, P3 -7.8 mm/yr, and P1 +0.7 mm/yr. These whole-record rates are descriptive summaries; they smooth over event pulses and recent reversals.

Table 2.3. End-of-record cumulative displacement and full-record linear downslope velocity at P1-P6. Negative component values denote settlement and horizontal advance in the interpreted downslope direction.

Point	LOS (mm)	Downslope (mm)	Vertical (mm)	Horizontal (mm)	Rate (mm/yr)
P1	-5.1	-8.4	-2.8	-7.9	+0.7
P2	-144.6	-254.8	-99.0	-234.7	-12.6
P3	-94.7	-184.5	-79.3	-166.6	-7.8
P4	-65.3	-102.9	-30.4	-98.3	-14.5

Point	LOS (mm)	Downslope (mm)	Vertical (mm)	Horizontal (mm)	Rate (mm/yr)
P5	-181.1	-298.9	-114.2	-276.2	-38.1
P6	-92.7	-140.8	-53.7	-130.2	-33.0

2.4.3 Movement components and inferred mechanism

The component decomposition shows that horizontal advance is larger than vertical settlement at every point (Figure 2.5). This is expected for slope angles of 17-25 degrees: the horizontal component equals D_slope cos(alpha) and therefore retains about 90-96 percent of the downslope vector magnitude, whereas the vertical component equals D_slope sin(alpha) and retains about 29-43 percent. These are orthogonal components of a vector and should not be added as independent displacements. If the vertical contribution is expressed as a share of the sum of the reported component magnitudes, it is approximately 23-32 percent across the six points.

The advance-dominated partition is compatible with shallow translational movement on southeast-dipping shale, not with pure vertical subsidence. P5 records about -276 mm of horizontal advance and -114 mm of settlement; P2 records about -235 and -99 mm, respectively. The same mechanism is expressed at lower magnitude at P4 and P6. Because the partition is determined by the DEM slope angle, it should be interpreted as the consequence of the slope-parallel assumption rather than independent proof of the failure geometry. Field structural mapping and subsurface instrumentation remain necessary to test that assumption.

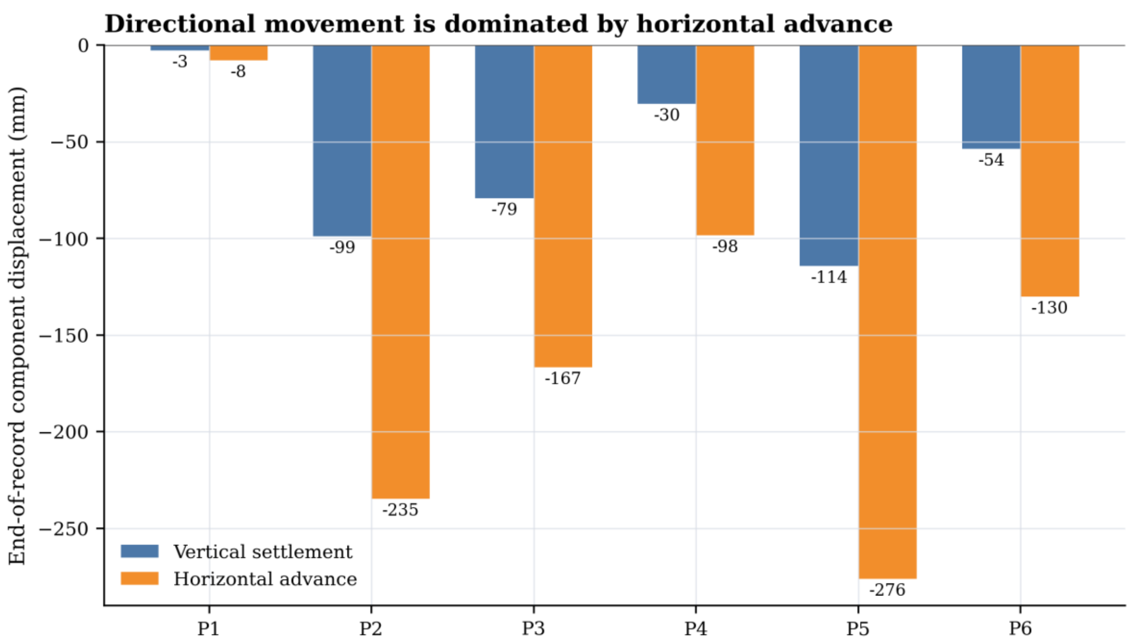


Figure 2.5. End-of-record vertical settlement and horizontal advance at P1-P6. Horizontal motion dominates because the monitored facets are moderately inclined and the model assumes movement parallel to the local slope.

2.4.4 Failure-era response, post-stabilization behavior, and recent reversals

Period-resolved rates reveal more than the full-record averages (Figure 2.6 and Table 2.4). During 2017-2019, P1, P2, P3, P5, and P6 show negative trend-derived velocities ranging from approximately -24.5 to -54.9 mm/yr. P4 shows a small positive rate, indicating that its response during that broad window differed from the other points. In 2020-2025, the absolute rate decreases at four of the six points. P1 and P2 are slightly positive, P3 slows from -54.9 to -12.9 mm/yr, and P6 slows from -52.3 to -22.8 mm/yr. The median absolute rate across all points falls from roughly 33 mm/yr in the failure-era window to about 12 mm/yr after stabilization.

The response is not uniform. P5 becomes more negative after 2020 (-36.8 mm/yr versus -27.6 mm/yr in 2017-2019), and P4 changes from a slightly positive broad-window trend to -10.9 mm/yr. These two locations demonstrate why a single sitewide statement such as 'the slope is stabilized' would be too strong. The most recent two-year rates also show substantial reversals: P2, P3, and P4 remain negative, P5 is mildly negative, and P1 and P6 become positive. A positive recent rate indicates recovery within that fitted window; it does not erase the preceding cumulative movement or prove long-term stability.

Velocity response is spatially variable: most points slowed, while residual activity persists

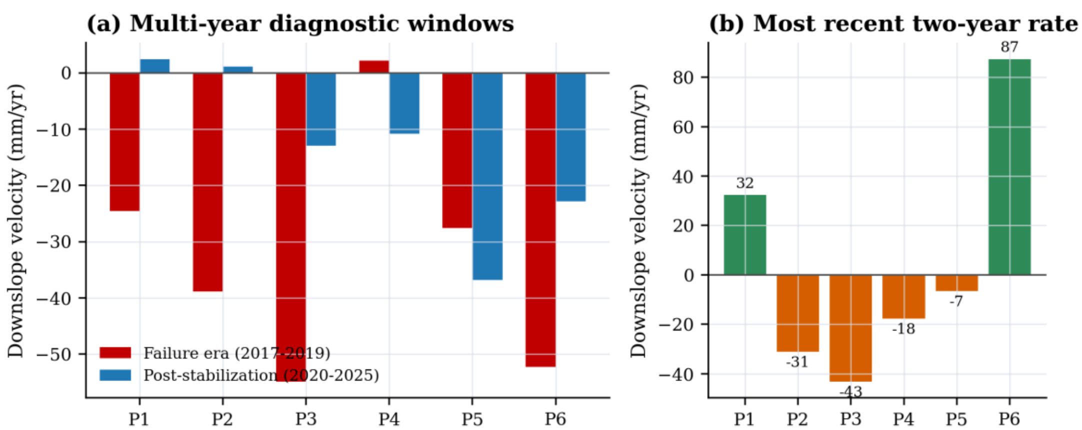


Figure 2.6. Period-resolved downslope velocities. Panel (a) compares the 2017-2019 failure-era and 2020-2025 post-stabilization windows; panel (b) shows the more responsive, but more window-sensitive, recent two-year rate.

Table 2.4. Trend-derived velocity windows and project screening labels. The labels organize follow-up monitoring within this report and are not calibrated alarm classes.

Point	2017-2019 (mm/yr)	2020-2025 (mm/yr)	Recent 2 yr (mm/yr)	Project screening label
P1	-24.5	+2.5	+32.3	Stable / recovering
P2	-38.8	+1.1	-31.1	Stable / recovering
P3	-54.9	-12.9	-43.3	Slow-moving creep
P4	+2.2	-10.9	-17.6	Slow-moving creep
P5	-27.6	-36.8	-6.5	Active
P6	-52.3	-22.8	+87.4	Moderately active

2.4.5 Cross-point LOS and directional forecasting performance

The rolling hybrid model reproduced the held-out LOS behavior at all six slope points with R2 values from 0.9834 to 0.9991 (Table 2.5 and Figure 2.7). LOS RMSE ranges from 1.24 mm at P4 to 2.78 mm at P6, while MAE ranges from 0.95 to 1.97 mm. P5, the point with the largest accumulated downslope movement, attains the highest reported R^2 (0.9991) and a LOS RMSE of 1.71 mm. P6 has the largest LOS error even though its R2 remains high, illustrating why scale-dependent error metrics must accompany goodness-of-fit.

Directional error follows directly from the projection geometry. Downslope RMSE ranges from 1.96 mm at P4 to 4.23 mm at P6; horizontal RMSE is similar because horizontal advance contains most of the slope-parallel vector. Vertical RMSE is smaller, from 0.58 to 1.71 mm, because only sin(alpha) of the downslope magnitude is assigned to settlement. These differences do not indicate that the model learns vertical motion better. They are the mathematical consequence of component scaling, and the component R^2 values remain identical to LOS.

Table 2.5. Chronological rolling-window LOS forecast performance and geometry-transferred directional RMSE. R2 is unchanged by the positive linear component scaling; the component errors increase according to each point's projection factor.

Point	LOS RMSE	LOS MAE	R2	Downslope RMSE	Vertical RMSE	Horizontal RMSE
P1	2.048	1.700	0.9834	3.35	1.11	3.16
P2	1.932	1.452	0.9905	3.40	1.32	3.14
P3	2.048	1.399	0.9871	3.99	1.71	3.60
P4	1.241	0.951	0.9977	1.96	0.58	1.87
P5	1.706	1.225	0.9991	2.82	1.08	2.60
P6	2.783	1.968	0.9972	4.23	1.61	3.91

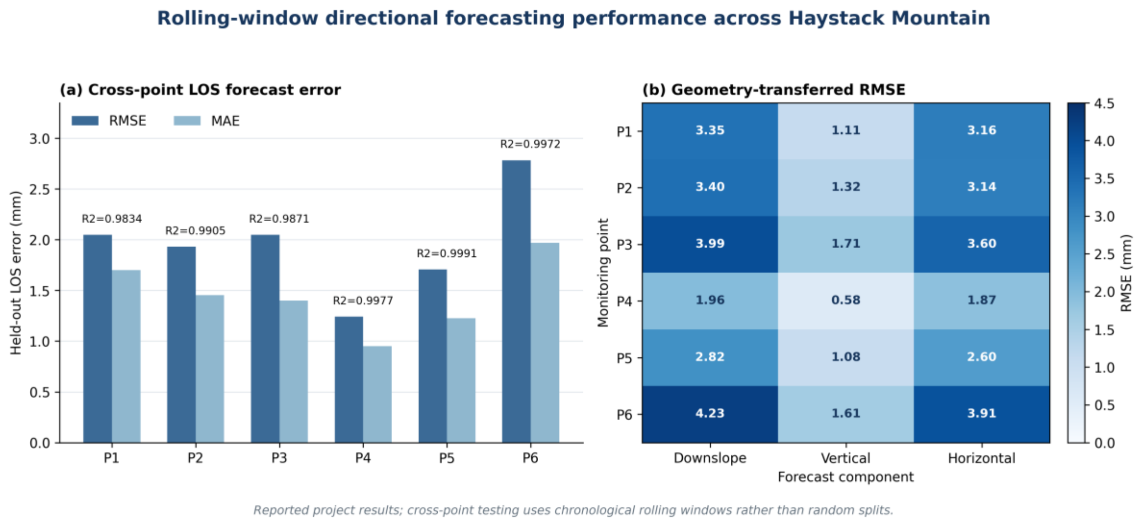


Figure 2.7. Chronological rolling-window forecasting performance across P1-P6. Panel (a) reports LOS RMSE, MAE, and R2; panel (b) shows the geometry-transferred RMSE for downslope, vertical, and horizontal components. The directional errors inherit both LOS-model error and point-specific projection amplification.

The high scores should still be interpreted as retrospective chronological reconstruction rather than proof of prospective warning skill. Adjacent forecast origins share much of the same four-year history, and the target series are smoothed and strongly autocorrelated. The next evaluation should publish the exact train-test origins, compare persistence, drift, seasonal-naive, ARIMA-only, and LSTM-only baselines at every point, and stratify errors by quiet, accelerating, and intervention-affected periods. Uncertainty calibration and residual autocorrelation should be reported alongside average RMSE.

2.4.6 Provisional January 2026 forecast and uncertainty

The research workflow was advanced one step beyond the last available Sentinel-1 observation to produce a provisional forecast for 23 January 2026 (Table 2.6 and Figure 2.8). The reported cumulative downslope forecasts range from +1.6 mm at P1 to -287.3 mm at P5. P2 remains the second-largest accumulated displacement at -250.8 mm, while P3, P4, and P6 are forecast at -171.8, -103.4, and -108.5 mm. These are cumulative positions in the project reference frame, not predicted one-cycle movements and not probabilities of failure.

The intervals were constructed from the historical pointwise error and then propagated through the same static geometry. They are narrow relative to the large accumulated offsets at P2-P6, but they do not include all sources of uncertainty. In particular, uncertainty in atmospheric correction, annual-block stitching, terrain orientation, motion-direction assumptions, environmental forecasts, and future structural changes is not fully represented. For this reason, the forecast can prioritize review of P5, P2, P3, and P6 but cannot independently declare stability or trigger a public warning.

Table 2.6. Provisional cumulative directional forecast for 23 January 2026. Brackets give the reported 95% intervals constructed from historical pointwise error. These research forecasts are not agency alarm limits.

Point	Downslope forecast [95% interval] (mm)	Vertical forecast [95% interval] (mm)	Horizontal forecast [95% interval] (mm)
P1	+1.6 [-5.0, +8.2]	+0.5 [-1.6, +2.7]	+1.5 [-4.7, +7.7]
P2	-250.8 [-257.5, -244.1]	-97.5 [-100.1, -94.9]	-231.1 [-237.2, -224.9]
P3	-171.8 [-179.7, -164.0]	-73.8 [-77.2, -70.5]	-155.2 [-162.2, -148.1]
P4	-103.4 [-107.3, -99.6]	-30.5 [-31.6, -29.4]	-98.8 [-102.5, -95.2]
P5	-287.3 [-292.8, -281.8]	-109.8 [-111.9, -107.7]	-265.5 [-270.6, -260.4]
P6	-108.5 [-116.8, -100.3]	-41.4 [-44.5, -38.2]	-100.3 [-108.0, -92.7]

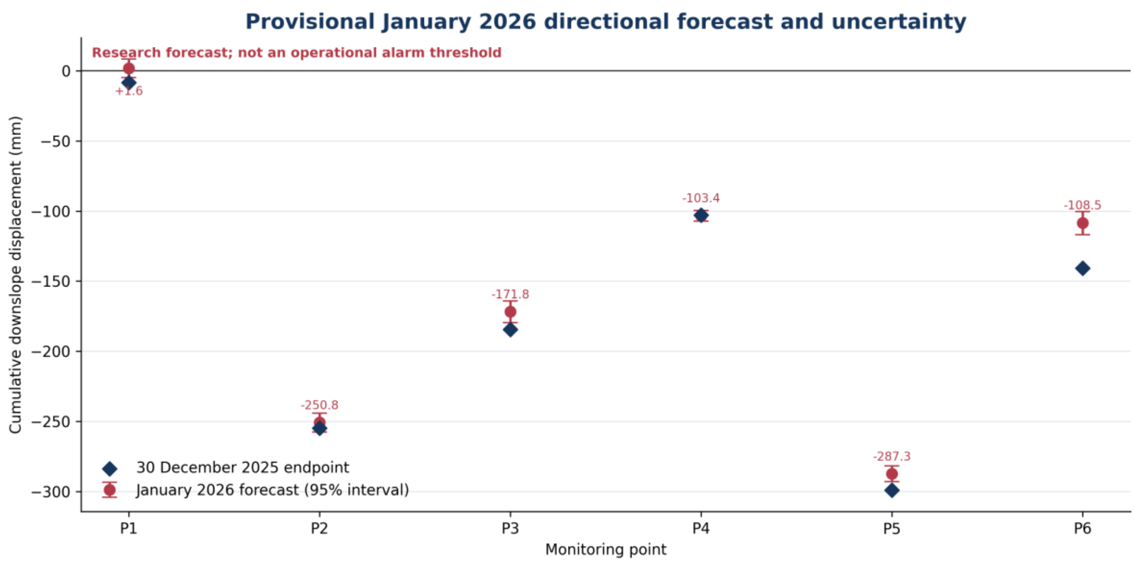


Figure 2.8. Provisional cumulative downslope forecast for 23 January 2026 with reported 95% intervals, compared with the 30 December 2025 endpoints. The result documents forward-modeling progress and is not an operational alarm threshold.

A defensible operational transition requires prospective scoring as new scenes arrive. Each 2026 acquisition should be compared with the stored prediction interval without refitting the already issued forecast, and both coverage and error should be logged. Only after several independent cycles should the system be recalibrated. This procedure prevents a visually close backcast from being mistaken for genuine forward validation and creates an auditable record for agency review.

2.4.7 Transportation asset-management interpretation

Combining cumulative displacement with the post-stabilization rate provides a clearer monitoring hierarchy than either indicator alone (Figure 2.9). P5 occupies the high-cumulative, high-rate corner and is the most immediate priority for field verification, drainage inspection, and continued satellite tracking. P6 has a smaller cumulative offset than P2, P3, or P5 but retains a post-stabilization rate of about -22.8 mm/yr; its recent positive reversal should be watched to determine whether it persists. P3 and P4 fall in an intermediate group where rates are similar to or modestly exceed the empirical benchmark bound, so independent confirmation is especially important.

P2 illustrates the difference between legacy displacement and current rate. It accumulated about -255 mm over the record but has a near-zero 2020-2025 trend, followed by renewed negative movement in the most recent two-year window. That sequence is more informative than a single activity class: substantial historical movement was followed by broad-window moderation and then recent reactivation. P1 is the lowest-priority slope point in the current screening because its cumulative endpoint and post-stabilization rate are near zero, although its time series contains large reversals that should not be ignored.

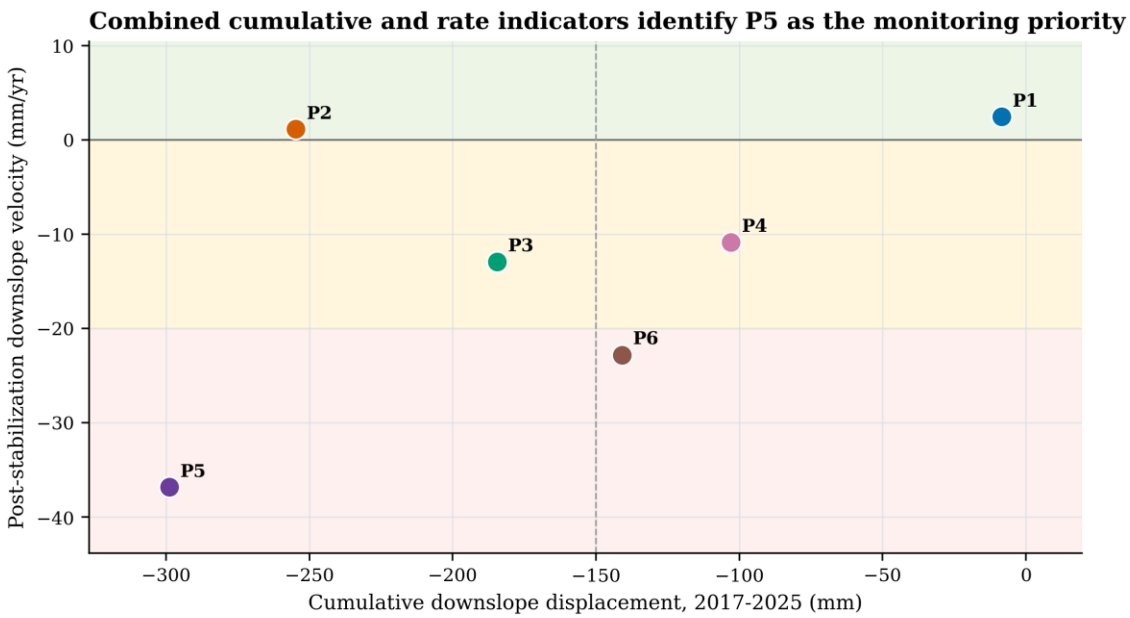


Figure 2.9. Combined cumulative downslope displacement and post-stabilization velocity. P5 is separated from the other points by both large accumulated movement and a continuing negative rate.

The point-specific follow-up plan in Table 2.7 converts this ranking into auditable actions. Priority increases when large accumulated movement is accompanied by continuing or renewed negative velocity and decreases when the response is near the empirical benchmark bound. The table deliberately avoids fixed alarm classes because the current record does not yet contain surveyed reference control or a calibrated failure threshold.
For agency use, the satellite record is best treated as a screening layer in a tiered workflow. A new negative pulse that is coherent across adjacent points should trigger review of rainfall, drainage condition, and the latest interferograms. Persistent movement at one point should prompt site reconnaissance and comparison with cracks, survey targets, or instrumentation. Escalation to an operational warning should require corroboration by ground measurements and a site-specific threshold framework. Installing a stable GNSS monument or corner reflector near the corridor would materially improve the ability to separate real slope movement from the reference behavior observed at BM1 and BM2.

Table 2.7. Point-specific highway monitoring priorities derived from the combined deformation, velocity, benchmark, and forecast evidence. Suggested actions are screening recommendations, not emergency classifications.

Point	Current priority	Evidence basis	Recommended follow-up
P1	Low	Near-zero endpoint and post-stabilization rate; large historical reversals	Retain as a slope-context point; inspect if reversals become spatially coherent
P2	High	Large legacy displacement and renewed negative recent rate	Review drainage and surface distress; compare next radar update with provisional interval
P3	High	Residual creep and negative recent rate; largest projection factor	Prioritize survey check and geometry validation before interpreting small accelerations
P4	Moderate	Post-stabilization negative trend after atypical failure-era behavior	Inspect localized drainage and compare with adjacent points
P5	Highest	Largest cumulative and post-stabilization movement	Target field reconnaissance, drainage review, and repeat geodetic control
P6	High	Strong full-record/post-stabilization motion with recent positive reversal	Determine whether recovery persists; verify against ground control and neighboring facets

2.5 Summary and Future Work

This reporting period produced a continuous Sentinel-1 SBAS record for Haystack Mountain covering 251 analysis epochs from January 2017 through December 2025. Six slope-point LOS series were projected to true downslope motion using point-specific satellite and terrain geometry and then resolved into vertical settlement and horizontal advance. Projection amplification remained below two at every analyzed slope point. The resulting components provide a more useful geotechnical interpretation than LOS alone, while retaining the important limitation that the method assumes slope-parallel translation.

The long-term record is spatially heterogeneous. P5 has the largest cumulative downslope displacement (-298.9 mm) and the largest full-record rate (-38.1 mm/yr); P2 has the second-largest cumulative displacement (-254.8 mm), and P6 has the second-largest full-record rate (-33.0 mm/yr). Horizontal advance dominates the resolved movement at all points, consistent with the moderate slope angles and the down-dip translational model. Comparison of diagnostic windows shows that four of six points slowed after the 2018-2019 failure era, but P5 remained active and P4 developed a negative post-stabilization trend. Recent reversals further demonstrate that a single velocity or cumulative total cannot summarize the whole slope.

The benchmark evaluation introduces a necessary caution. BM1 and BM2 show nonzero apparent displacement, so the lower-magnitude slope-point rates should be interpreted within an empirical noise/reference bound rather than as precise absolute ground velocities. The strongest operational conclusion is therefore relative: P5 remains the priority monitoring location, P6 warrants continued attention, and P2 requires follow-up because recent movement resumed after a quieter broad post-stabilization window. These findings support targeted field verification and drainage inspection rather than a sitewide declaration of either stability or imminent failure.

The added forecasting analysis demonstrates a separate form of progress. A rolling four-year STL-ARIMA-LSTM workflow produced chronological LOS R^2 values of 0.9834-0.9991 across P1-P6, with LOS RMSE of 1.24-2.78 mm. Because the directional transform is linear, the component forecasts retain the LOS R^2 while their errors scale with terrain and acquisition geometry. The January 2026 result extends the workflow beyond the last observation, but it remains a provisional research forecast whose intervals do not yet include all geodetic, environmental, and structural uncertainties.

Taken together, the monitoring and forecasting results support a staged decision process: confirm data quality; identify spatially coherent deformation; compare the signal with rainfall and intervention history; issue and archive a forecast; then seek field corroboration before escalating. This sequence is more valuable than a single model score because it preserves the distinction among measured LOS, geometry-conditioned components, forecast expectations, and agency safety decisions.

Future work should strengthen both the measurement reference and the physical model. First, a surveyed GNSS monument, corner reflector, or co-located field instrument should be established to validate the radar reference frame. Second, ascending and descending Sentinel-1 geometries should be combined where coherence permits, reducing dependence on the one-dimensional slope-parallel assumption. Third, the time series should be evaluated jointly with precipitation, antecedent soil moisture, freeze-thaw indices, maintenance records, and field observations. Fourth, uncertainty should be propagated from phase quality, atmospheric correction, DEM angle error, and the projection coefficient into reported displacement and velocity intervals. Finally, the data pipeline can be extended with a separately validated forecasting module, but only after out-of-sample tests and site-specific decision thresholds demonstrate operational reliability.

Chapter 3

3 Intervention-Aware InSAR and Multi-Horizon Machine Learning for Urban Rail-Corridor Slope Screening at the Baltimore CSX Belt Line

Oluwatunmise Akinniyi[1], Tohid Asheghimehmandari[2], Paris Griffin[2], Zhuping Sheng[2], and Yi Liu[2]

[1]Departments of Electrical and Computer Engineering, Morgan State University, Baltimore, Maryland

[2]Department of Civil and Environmental Engineering, Morgan State University, Baltimore, Maryland

Publication status notice. This chapter documents annual progress for the USDOT Safety21 report and is designed to follow the Haystack Mountain highway chapter without repeating its general SBAS-InSAR theory or slope-projection derivation. The Baltimore rail-corridor analysis and forecasting results are also being developed for separate peer-reviewed publication. The organization, narrative, tables, and figures used here are report-specific and are not the archival journal manuscript.

3.1 Introduction

Rail-corridor slope instability differs from a highway cut in both consequence and observability. Movement of a retained urban slope can affect the street at its crest, a wall or support system at mid-slope, and active rail infrastructure at its toe. A localized failure can therefore interrupt two transportation networks at once, constrain emergency access, and expose nearby residents and utilities to additional risk. Because freight operations may resume before the full repair is complete, the monitoring problem includes not only natural ground response but also demolition, excavation, piling, drainage work, and reconstruction. This chapter reports progress at East 26th Street in the Charles Village area of Baltimore, Maryland, above the CSX Belt Line. On 26 November 2018, a section of the retaining-wall system between North Calvert Street and Guilford Avenue partially buckled and began to fail. East 26th Street was closed and rail operations were temporarily halted while the compromised area was secured [32]. Emergency stabilization was followed by a replacement wall and roadway project completed in 2019, with later public-space improvements completed in 2022 [33]. These documented interventions provide known boundaries for interpreting the satellite record.

The companion Haystack Mountain chapter established the report's general Sentinel-1 and SBAS-InSAR terminology and demonstrated slope-aligned component resolution at a highway site with terrain and viewing geometry. Repeating that derivation here would add length without adding evidence. The Baltimore analysis instead retains line-of-sight (LOS) displacement as its primary observable because point-specific radar incidence, heading, slope, aspect, and independent motion direction have not yet been frozen for the selected urban pixels. The novelty of Chapter 3 is the integration of comparison-point differencing, event and repair flags, multiple within-year velocity scales, and a multi-horizon machine-learning pilot for a narrow, actively modified rail corridor. The reporting-period objectives were:

1. assemble and audit eight independently referenced annual Sentinel-1 LOS segments from 2018 through 2025;
2. compare five high-response slope locations with three nearby lower-response comparison locations without treating the latter as surveyed benchmarks;
3. test whether the documented 26 November 2018 failure is represented by a spatially coherent acquisition-to-acquisition LOS change;
4. quantify two-month, calendar-quarter, and six-month LOS rates without fitting across annual reference resets;
5. measure annual slope-to-comparison separation and the persistence of negative within-year responses;
6. evaluate a Point03 XGBoost proof of concept at 12-, 24-, and 36-day forecast horizons; and
7. document the remaining reproducibility metadata and translate the combined evidence into a tiered inspection and validation strategy for rail-corridor asset management.

3.2 Literature Review

The general strengths and limitations of multi-temporal InSAR for slope monitoring were reviewed in Chapter 2. For railways, the central extension is operational: the measurement must be related to track, wall, drainage, and access decisions. Huntley et al. demonstrated how benchmarked satellite radar change detection can support landslide-resilient railway management, while corridor applications to bridges and other engineered assets show that persistent urban scatterers can provide repeatable deformation histories [34, 35]. These studies also reinforce that a useful radar signal is not defined by magnitude alone; spatial persistence, reference quality, geometry, and agreement with independent observations remain essential [16].

An urban retaining-wall setting is unusually favorable for radar coherence and unusually difficult for causal interpretation. Concrete, pavement, rail, and buildings can maintain strong phase histories, but the effective radar response may mix the wall, backfill, roadway, vegetation, utilities, and construction equipment within a small area. After an emergency repair, a measured change may represent continued ground movement, structural settlement, excavation, new drainage, resurfacing, or atmospheric residual. A monitoring design must therefore preserve maintenance and construction dates rather than treating the full series as one stationary geotechnical process.

Annual-block processing introduces another issue that is more important in this chapter than in the continuous highway record. SBAS inversion estimates displacement relative to a selected temporal and spatial reference [7, 10, 36]. When each calendar year is processed independently, the first value in each new annual run is not tied geodetically to the last value in the preceding run. A display can be joined for visualization, but the December-January gap remains unobserved. Reporting a joined endpoint as direct eight-year displacement would therefore overstate the measurement.

Machine learning can convert time-series histories into short-horizon forecasts, but performance is especially sensitive to validation design. Random train-test splits, features calculated with future observations, and failure to encode major repairs can leak information and inflate apparent skill. Gradient-boosted trees such as XGBoost are useful for structured lag, rate, variability, weather, and intervention features because they model nonlinear interactions without requiring a large recurrent network [38-41]. For infrastructure use, however, a model must be evaluated in chronological order and compared with simple persistence and drift baselines. Forecast skill is evidence of predictive consistency, not by itself evidence of a validated warning threshold. Recent InSAR-machine-learning mapping studies [42-44] and regional rainfall studies [45] reinforce the need for causal features and external validation.

3.3 Materials and Methodology

3.3.1 Corridor setting, event history, and monitoring design

The study area follows the south side of East 26th Street above the CSX Belt Line, centered near 39.3195 degrees N and -76.6127 degrees W. The retained slope occupies a narrow strip between an urban street and the railway cut. The original screening inventory contained 17 slope locations and a larger group of relatively flat comparison surfaces. Five slope locations with the strongest assembled response metrics were retained as Point01-Point05, and three lower-response surfaces were retained as Point06-Point08. The original point names remain in Table 3.1 for traceability.

The selection is intentionally described as outcome-informed. It is appropriate for characterizing the strongest available signals but cannot be used to estimate unbiased corridor-wide detection accuracy. Similarly, Point06-Point08 are comparison locations, not surveyed stable benchmarks. They provide a local empirical reference for common-mode behavior while remaining subject to real movement, atmospheric residuals, and processing artifacts.

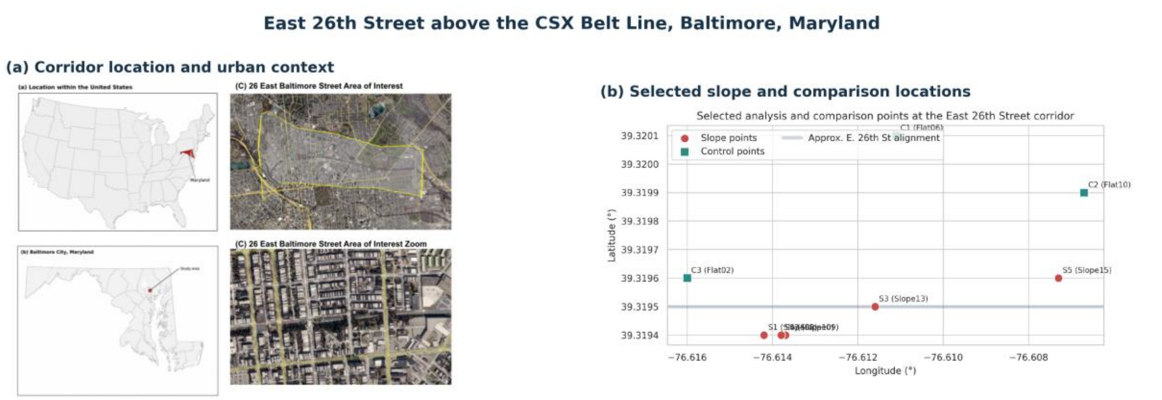


Figure 3.1. East 26th Street/CSX Belt Line study area and selected monitoring layout. Point01-Point05 are priority slope locations and Point06-Point08 are nearby comparison surfaces. The locations are coordinate-based screening pixels rather than a surveyed failure polygon.

Table 3.1. Selected Baltimore CSX monitoring locations and continuity-adjusted descriptive metrics. The endpoint and fitted rate summarize the assembled display; they are not independently observed eight-year displacement because each calendar year was referenced separately.

Point	Original	Role	Latitude	Longitude	Display end (mm)	Display OLS rate (mm/yr)
Point01	Slope08	Slope	39.3194	-76.6142	-140.0	-15.5
Point02	Slope09	Slope	39.3194	-76.6137	-100.7	-15.1
Point03	Slope13	Slope	39.3195	-76.6116	-98.9	-17.3
Point04	Slope10	Slope	39.3194	-76.6138	-97.8	-12.9
Point05	Slope15	Slope	39.3196	-76.6073	-60.1	-3.2
Point06	Flat06	Comparison	39.3201	-76.6111	-0.6	+0.3
Point07	Flat10	Comparison	39.3199	-76.6067	-14.7	-2.1
Point08	Flat02	Comparison	39.3196	-76.6160	+9.8	+0.7

3.3.2 Annual LOS normalization and continuity-adjusted display

The source workbook contains 240 observations at each of the eight locations from 12 January 2018 through 25 December 2025. Data were generated as eight calendar-year Sentinel-1 SBAS runs using the PyGMTSAR workflow in Google Colab [36]. The supplied 2025 processing record identifies ascending Sentinel-1A, relative orbit 106, IW1, and VV polarization. A complete scene, orbit, reference-area, coherence, and unwrapping inventory for 2018-2024 remains a journal-submission task and is not reconstructed from memory in this annual report.

For point p and year y, each annual LOS series was re-zeroed to its first observation. This preserves all within-year increments while removing the arbitrary initial value:

$$d\_rel,p,y(t) = d\_raw,p,y(t) - d\_raw,p,y(t_0,y) \quad (3.1)$$

For visualization, the new annual segment was offset so that its first value equals the preceding segment's display endpoint:

$$d\_join,p,y(t) = d\_rel,p,y(t) + d\_join,p,y\text{-}1(t_{\_end},y\text{-}1) \quad (3.2)$$

Equation (3.2) imposes a zero-jump continuity convention at every January boundary. It does not measure motion between annual runs. Accordingly, this chapter uses the terms continuity-adjusted display, display endpoint, and display OLS rate. All event increments and short-window velocities are calculated inside an annual segment.

3.3.3 Comparison differencing and multi-timescale LOS rates

At every common acquisition date, the median joined LOS of Point06-Point08 was calculated as a local comparison series. The relative history for each slope point was then:

$$d^*\_p(t) = d\_join,p(t) - \text{median}[d\_join,Point06(t),\ d\_join,Point07(t),\ d\_join,Point08(t)] \quad (3.3)$$

Median differencing reduces behavior shared by the three comparison surfaces and is less dependent on any single reference curve. It does not remove spatially varying atmosphere or prove that the comparison locations are fixed. Both raw joined and comparison-differenced views are therefore retained.

Within each calendar year, ordinary-least-squares LOS velocity was estimated over two-month, calendar-quarter, and six-month windows. If t_i is elapsed time in days and d_i is LOS displacement inside window w, the annualized rate is:

$$V_w = 365.25 \left[\Sigma_i (t_i - \bar{t})(d_i - \bar{d})\right] / \left[\Sigma_i (t_i - \bar{t})^2\right] \quad (3.4)$$

The annualization permits comparison among window lengths but does not imply that a short-window rate persists for a full year. The three scales have different operational roles: two-month rates respond quickly but amplify noise; quarter rates provide an intermediate screen; and six-month rates emphasize persistence at the cost of event timing. As established in Chapter 2, a single LOS observation cannot provide independent east, north, and vertical components. Because the required point-specific geometry is incomplete at the rail site, no downslope or component conversion is presented as a measured result in Chapter 3 [11, 14, 35].

3.3.4 Event and intervention boundaries

The acquisitions immediately bracketing the documented wall failure occurred on 20 November and 2 December 2018. Their raw annual LOS difference was calculated point by point, then corrected by subtracting the median change of Point06-Point08. This 12-day pair contains the 26 November failure, emergency securing, demolition, and initial stabilization. It is therefore an event bracket rather than a precursor interval. Additional structural boundaries were assigned to emergency work beginning in November 2018, replacement-wall and street construction completed in 2019, and later public-space work completed in 2022 [32, 33].

3.3.5 Multi-horizon machine-learning proof of concept

The rail-corridor forecasting contribution is a direct, multi-horizon XGBoost pilot rather than the geometry-resolved displacement analysis used in the highway chapter. For forecast origin t and horizon h, the general task is:

$$\hat{y}_{p}(t+h) = f_h[X_p(t), X_p(t-1), \ldots, \text{intervention and environmental features}] \quad (3.5)$$

The supplied project poster reports Point03 results for h = 12, 24, and 36 days using 2018-2023 as the training period and 2024-2025 as a held-out evaluation period. The experimental workflow considered displacement history and environmental variables. Exact hyperparameters, feature lags, missing-data rules, baseline comparisons, and target-geometry metadata were not included in the source packet; this chapter therefore reports only the documented model class, horizons, and evaluation metrics. Before journal submission, the target should be frozen as validated LOS or comparison-differenced LOS unless point-specific directional geometry is supplied.

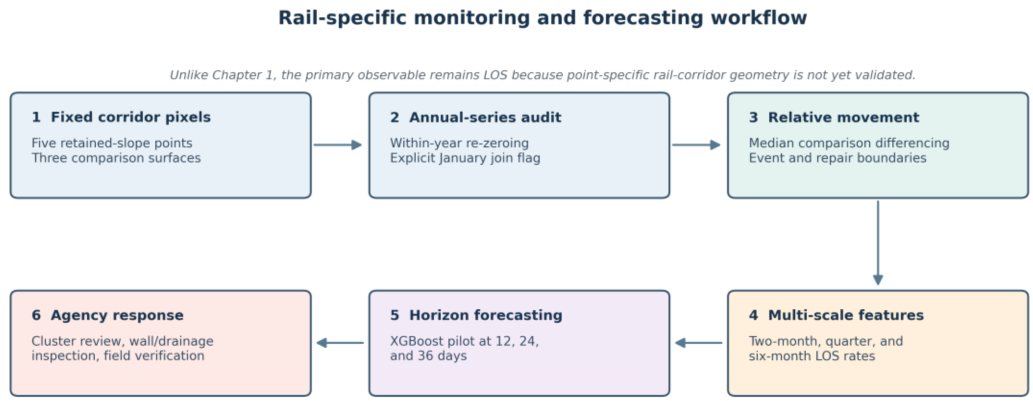


Figure 3.2. Report-specific rail monitoring and forecasting workflow. The sequence emphasizes fixed corridor pixels, annual-reference auditing, comparison differencing, intervention flags, multi-scale LOS features, multi-horizon forecasting, and field-based agency response.

3.3.6 Reproducibility record and intervention-aware validation

The selected-point workbook is sufficient to reproduce the descriptive calculations in this chapter, but it does not retain every processing decision required to reproduce the original interferometric products. Table 3.2 therefore separates the current project record from the metadata that must be recovered from executed notebooks or regenerated processing logs. The missing items include complete product and burst IDs for 2018-2024, annual reference definitions, interferogram networks, baselines, coherence and unwrapping masks, filtering settings, software versions, and point-neighborhood extraction rules [46].

The same discipline applies to forecasting. Point selection must be frozen before the test interval; every lag and rolling statistic must use only information available at the forecast origin; scaling and imputation must be fitted on training data; and the 2018 emergency, 2019 reconstruction, and 2022 public-space completion must be encoded as structural boundaries. A blocked experiment should compare XGBoost with last-observation, drift, seasonal-naive, and regularized autoregressive baselines at all slope and comparison points. Performance should be stratified by horizon, location type, and intervention regime rather than summarized only by one selected point.

Table 3.2. Rail-corridor processing record and completion requirements. Items not preserved in the supplied exports are identified explicitly rather than reconstructed from memory.

Component	Current project record	Required completion action
Platform/geometry	Sentinel-1 C-band ascending; 2025 log lists Sentinel-1A, relative orbit 106, IW1, VV	Verify scene and burst inventory for 2018-2024
Temporal design	Eight separate calendar-year runs; 28-31 dates per year	Archive every annual reference date, area, and December-January gap
Software	PyGMTSAR processing in Google Colab	Record package, GMTSAR, Python, and dependency versions
Pair network/quality	Not retained in selected-point Excel exports	Export pair network, baselines, coherence thresholds, masks, and unwrapping

Component	Current project record	Required completion action
		diagnostics
Point extraction	Five outcome-informed slope points and three comparisons	Freeze selection before validation and report pixel/neighborhood sampling
Primary observable	Annual LOS displacement with export sign convention	Verify toward/away sign and retain LOS unless point geometry is validated
Interventions/covariates	2018 emergency, 2019 reconstruction, 2022 public-space work	Confirm dates and join rainfall, drainage, wall, and train-operation records

These requirements are particularly important in an urban corridor, where a model may otherwise learn the signature of demolition or resurfacing and appear to predict ground movement. The annual report records the documented Point03 proof of concept, while the journal workflow must add the frozen metadata, baseline tests, uncertainty intervals, and intervention-stratified validation before making a broader forecasting or early-warning claim [18, 40, 41, 47].

3.4 Results / Case Study

3.4.1 Slope-to-comparison contrast

The assembled displays show a clear amplitude contrast between the five selected slope locations and the three comparison surfaces. Point01 has the most negative display endpoint at -140.0 mm. Point02, Point03, and Point04 end at -100.7, -98.9, and -97.8 mm, and Point05 ends at -60.1 mm. The comparison endpoints are -0.6 mm at Point06, -14.7 mm at Point07, and +9.8 mm at Point08. Because the annual joins are assumed rather than observed, these values quantify accumulated within-year response on a common display, not direct eight-year displacement.

The fitted display trends distinguish endpoint magnitude from temporal persistence. Point03 has the steepest OLS rate (-17.3 mm/yr), followed by Point01 (-15.5), Point02 (-15.1), Point04 (-12.9), and Point05 (-3.2 mm/yr). Thus, Point01 has the largest display endpoint while Point03 has the strongest fitted long-term trend. The comparison rates range from -2.1 to +0.7 mm/yr. Point07's nonzero response is another reason to call these locations comparisons rather than stable benchmarks.

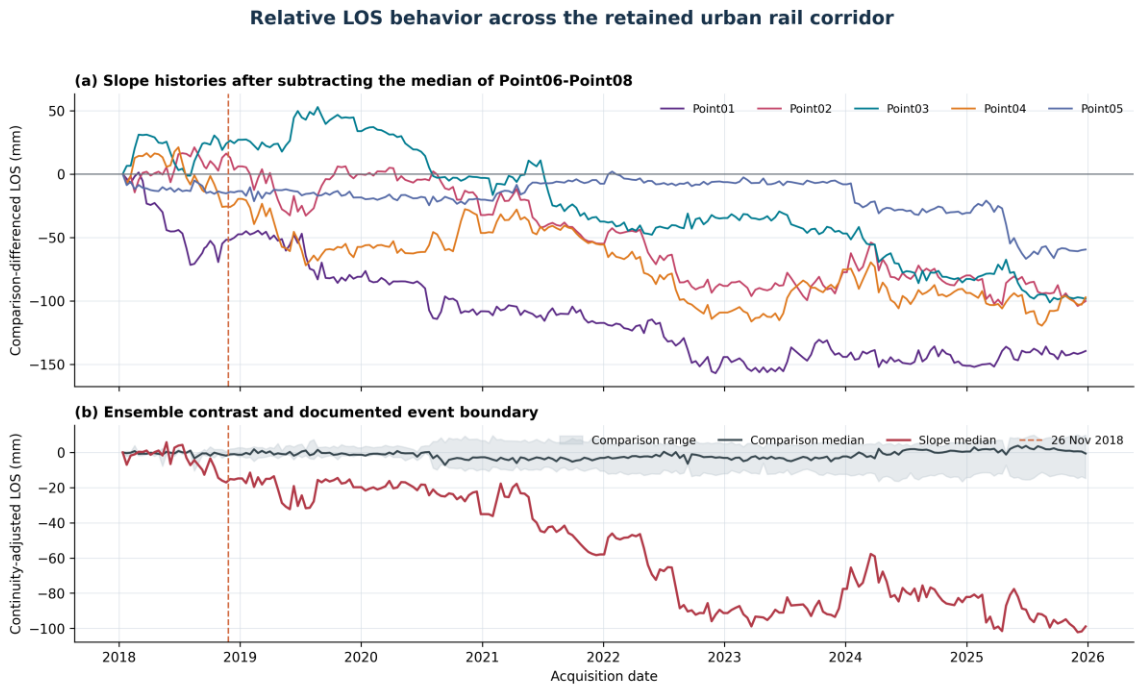


Figure 3.3. Comparison-differenced slope histories and ensemble contrast. The dashed line marks the documented 26 November 2018 failure. January joins align independently referenced annual segments for display and do not represent measured interannual motion.

3.4.2 Event bracket and intervention-driven regimes

The two acquisitions bracketing the 2018 failure do not show a uniform slope-wide change. Raw LOS increments range from -2.49 mm at Point02 to +3.12 mm at Point03. After subtracting the +0.75 mm comparison median, Point02 is -3.23 mm and Point03 is +2.37 mm, while Point01, Point04, and Point05 are -0.12, -0.39, and +1.16 mm. The opposite signs at the two largest responses prevent interpretation as one coherent failure displacement.

Table 3.3. LOS change across the Sentinel-1 acquisitions immediately bracketing the documented 26 November 2018 failure. Corrected values subtract the +0.75 mm median change of Point06-Point08.

Point	Raw 20 Nov-2 Dec ΔLOS (mm)	Comparison-corrected ΔLOS (mm)	Interpretive role
Point01	+0.63	-0.12	Slope response
Point02	-2.49	-3.23	Slope response
Point03	+3.12	+2.37	Slope response
Point04	+0.36	-0.39	Slope response
Point05	+1.91	+1.16	Slope response
Point06	+0.07	-0.67	Comparison response
Point07	+0.75	+0.00	Comparison response
Point08	+1.48	+0.73	Comparison response

This negative event test is useful. The satellite interval contains both wall distress and emergency intervention, so a common step would still require field evidence for attribution. The observed mixed directions indicate that the event date should be encoded as a structural boundary, not used as proof that one LOS trend predicted the failure.

Annual endpoint changes reveal stronger regime transitions than the 12-day event pair. Point01 is strongly negative from 2018 through 2022 but becomes slightly positive in 2023 and +10.5 mm in 2025. Point03 is positive in 2018-2019, negative in 2020-2021, slightly positive in 2022, and negative again from 2023 through 2025. Point04 changes from -53.5 mm in 2022 to +34.7 mm in 2023. Point05 is comparatively subdued through 2023 but records -19.7 mm in 2024 and -30.0 mm in 2025. These reversals are compatible with a nonstationary corridor affected by failure, repair, urban surface changes, and environmental variability.

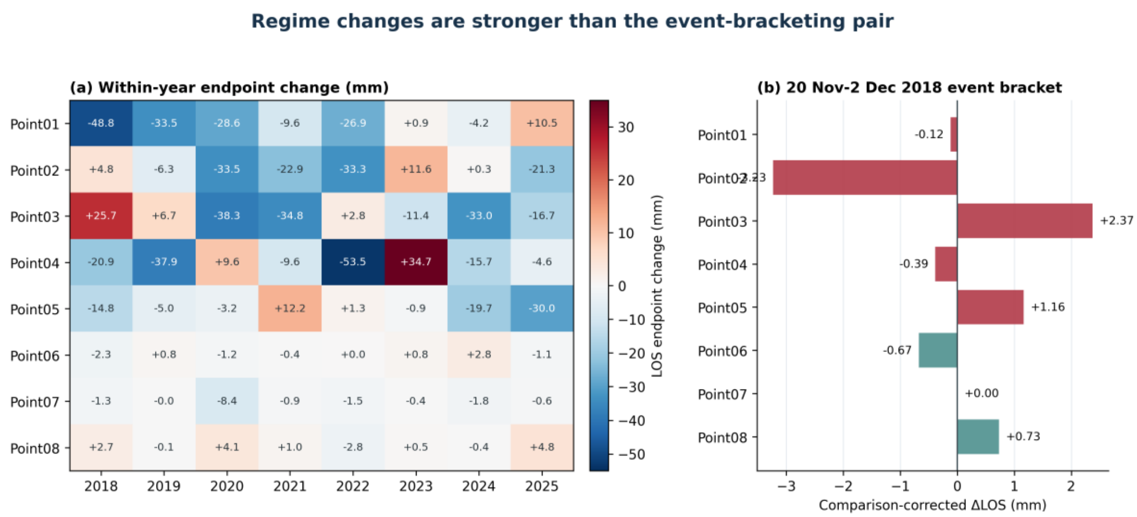


Figure 3.4. Within-year LOS endpoint changes and the comparison-corrected 2018 event bracket. Annual sign reversals and post-repair regimes are more prominent than a uniform failure-date step.

3.4.3 Multi-timescale LOS screening

The median rates change with analysis scale, which is itself an operational finding. Point02 is the most consistently negative location across the three summaries: -17.2 mm/yr for two-month windows, -28.5 mm/yr for quarters, and -18.7 mm/yr for six-month windows. Point03 remains negative at -8.5, -10.9, and -11.3 mm/yr. Point04 has a stronger short-window median (-14.1 mm/yr) than quarter and six-month medians (approximately -8.8 mm/yr), suggesting more episodic response.

Point01 illustrates why a single statistic can mislead. It has the most negative display endpoint and a steep display OLS trend, yet its median two-month rate is only -0.7 mm/yr; the quarter and six-month medians are -12.5 and -9.8 mm/yr. This combination indicates concentrated intervals of accumulation rather than a uniformly negative short-window response. Point05 remains lower amplitude at all three scales. Comparison medians are near zero except Point07, which reaches approximately -2.7 mm/yr at quarter and six-month scales.

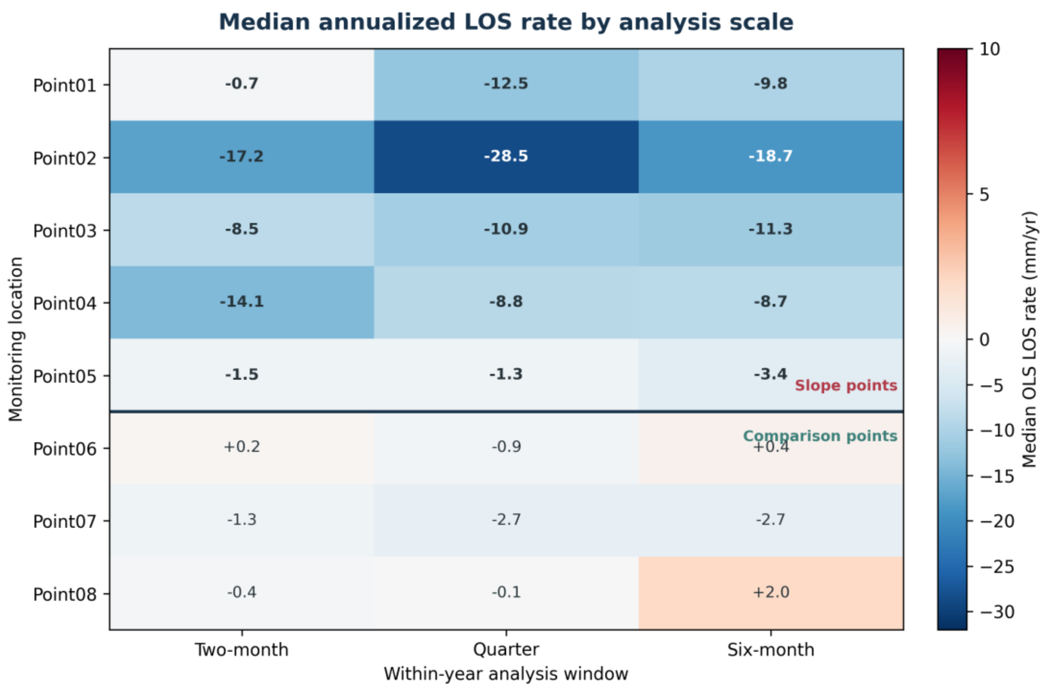


Figure 3.5. Median annualized LOS rates for two-month, calendar-quarter, and six-month windows. Longer windows more consistently separate most slope locations from the comparison set, while short windows reveal episodic behavior and greater variability.

3.4.4 Annual slope-to-comparison separation and persistence

Annual medians provide a reference-safe corridor summary because each value is calculated entirely inside one calendar-year run. The slope median is more negative than the comparison median in seven of eight years. The slope-minus-comparison separation is approximately -13.5 mm in 2018, -6.3 mm in 2019, -27.4 mm in 2020, -9.2 mm in 2021, and -25.4 mm in 2022. The gap nearly closes in 2023 (+0.4 mm) before returning to -15.3 mm in 2024 and -16.1 mm in 2025 (Figure 3.6).

The temporary 2023 convergence is consistent with the pointwise sign reversals shown in Figure 3.4 and should not be interpreted as a permanent corridor recovery. Point01, Point04, and Point05 have negative within-year endpoint changes in six of eight years; Point02 and Point03 are negative in five. Their RMS within-run increments are also larger than those at the comparison locations. In contrast, Point06 is negative in four years, Point08 in three, and Point07 in all eight but with much smaller increment amplitude. This combination shows why persistence alone is insufficient: Point07 is persistent enough to reject the label stable benchmark, yet its amplitude remains closer to the comparison regime than to the selected slope set.

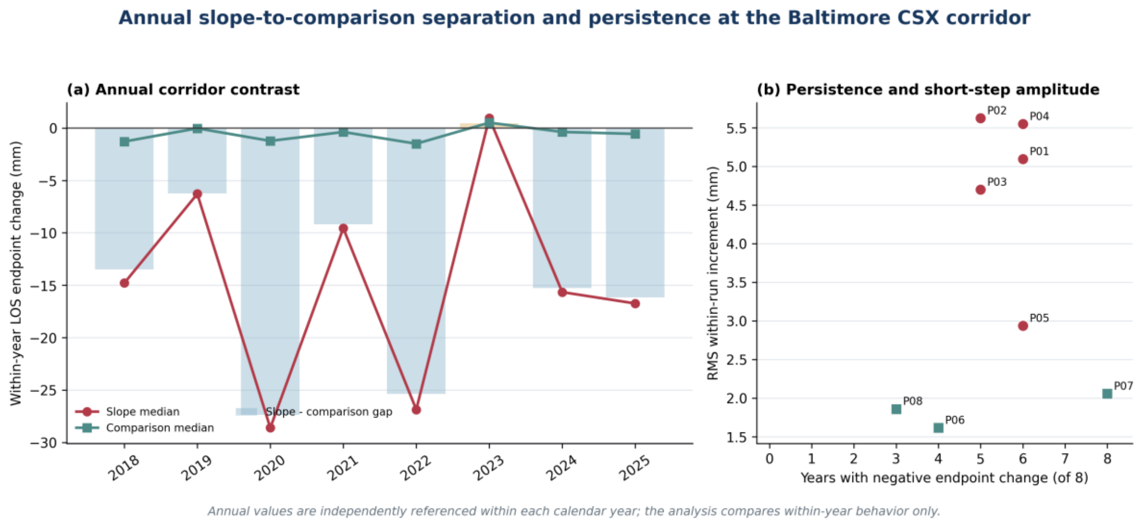


Figure 3.6. Annual corridor contrast and persistence diagnostics. Panel (a) compares within-year slope and comparison medians and their difference; panel (b) relates the number of negative annual endpoint changes to the RMS within-run increment. The analysis does not bridge unobserved December-January reference gaps.

The annual contrast strengthens the screening interpretation without converting the continuity-adjusted display into a measured eight-year displacement. It demonstrates that the selected slope set repeatedly departs from the nearby comparison ensemble within individual years, including after major reconstruction. Attribution still requires wall, drainage, track, and construction information because the statistic measures relative radar behavior, not a unique geotechnical mechanism.

3.4.5 Point03 machine-learning performance

The Point03 XGBoost pilot shows the expected degradation with forecast horizon. At 12 days, RMSE is 10.66 mm, MAE is 8.49 mm, and R2 is 0.960. At 24 days, RMSE increases to 12.06 mm and MAE to 9.74 mm while R2 decreases to 0.948. At 36 days, RMSE is 12.99 mm, MAE is 10.40 mm, and R2 is 0.939. The retained R2 values indicate that the prototype captures much of the held-out variation at the selected location, while the monotonic error growth provides an honest measure of the cost of a longer decision horizon.

Table 3.4. Point03 XGBoost multi-horizon forecasting proof of concept. The supplied research poster identifies 2018-2023 as training and 2024-2025 as the held-out period; corridor-wide and intervention-stratified validation remain incomplete.

Horizon (days)	RMSE (mm)	MAE (mm)	R2	Progress interpretation
12	10.66	8.49	0.960	Best pilot skill; near-term screening horizon
24	12.06	9.74	0.948	Moderate error growth; skill remains high
36	12.99	10.40	0.939	Largest reported error; still a research result

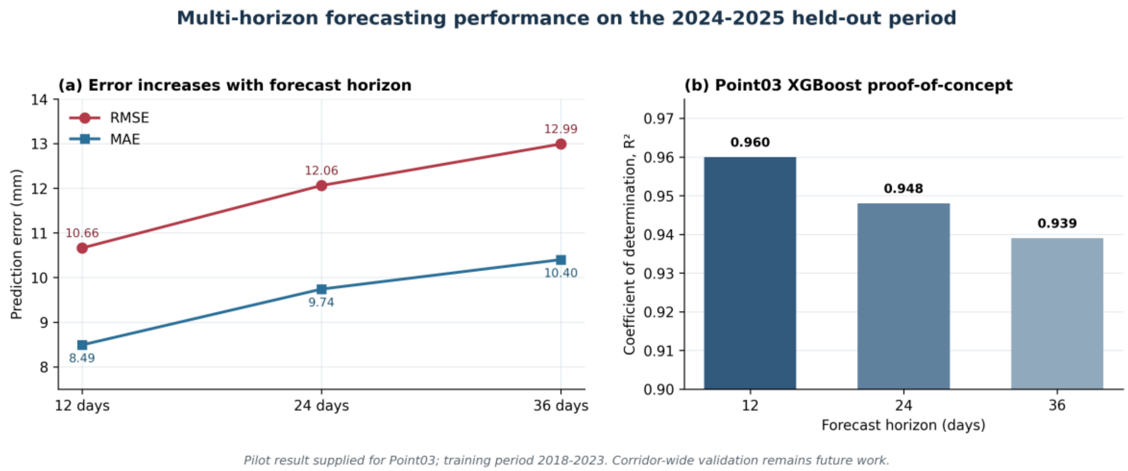


Figure 3.7. Point03 XGBoost proof-of-concept performance for the 2024-2025 held-out period. Forecast error increases from 12 to 36 days while R2 remains above 0.93. These results document progress and do not yet constitute corridor-wide or prospective early-warning validation.

The performance should be interpreted narrowly. Point03 was one of the outcome-selected high-response pixels, and the source packet does not yet preserve every model setting or baseline result. The next validation must freeze point selection before testing, use only trailing features, fit preprocessing on the training period, compare XGBoost with persistence and drift, report uncertainty intervals, and repeat the experiment at the other slope and comparison locations. The 2018-2019 emergency and repair boundaries should also be encoded so that the model is not rewarded for learning a mixed structural regime.

3.4.6 Rail-corridor asset-management interpretation

The combined results support a tiered screening workflow rather than an autonomous alarm. First, radar quality and the annual reference record must be acceptable. Second, a candidate location should exceed the local comparison ensemble and agree with adjacent slope pixels. Third, the response should persist at more than one time scale or recur in a documented seasonal or intervention context. Fourth, a forecast exceedance should persist across updates and be accompanied by an uncertainty estimate. Only then should the signal trigger targeted wall, drainage, roadway, or track-adjacent inspection.

The most important management distinction is between condition screening and failure attribution. The present record is strong enough to identify selected slope pixels with larger and more persistent LOS behavior than nearby comparison surfaces. It is not strong enough to assign all motion to one geotechnical mechanism, reconstruct the exact 2018 failure displacement, or define a public-safety threshold. InSAR and machine learning therefore extend the inspection program; they do not replace survey control, wall inspection, drainage review, inclinometers, piezometers, or track-geometry measurements.

Table 3.5 formalizes the evidence sequence so that a radar anomaly is not passed directly to an alarm decision. Each stage can stop the escalation: poor coherence requires reprocessing; an isolated pixel requires spatial review; a short-window pulse requires persistence testing; an intervention-aligned response requires engineering records; and a forecast must outperform simple baselines and repeat across updates. This staged structure is appropriate for a corridor where street, wall, and railway operations may each create legitimate surface change.

Table 3.5. Evidence-to-action ladder for the Baltimore rail corridor. The ladder organizes review and field response without defining an unvalidated public-safety threshold.

Review stage	Evidence required	Agency-facing action	Boundary
1 Data acceptance	Traceable annual reference; acceptable coherence and comparison behavior	Accept or reprocess the radar update	No movement interpretation if quality is unresolved
2 Spatial confirmation	Candidate exceeds local comparison ensemble and agrees with neighboring slope pixels	Review maps and request targeted reconnaissance	A single selected pixel is insufficient
3 Temporal persistence	Response recurs or persists across at least two analysis scales	Review drainage, wall, and maintenance context	Short-window spikes remain screening evidence
4 Intervention attribution	Signal timing is separable from demolition, reconstruction, or surface work	Coordinate with agency/rail engineering records	Do not label construction response as a precursor
5 Forecast corroboration	Repeated exceedance with uncertainty and skill above simple baselines	Escalate to survey, wall, track, or geotechnical instrumentation	No autonomous warning without field confirmation

For implementation, the radar review should be synchronized with existing inspection cycles rather than operated as a parallel, disconnected system. A concise agency record for each flagged update should include the scene date, point and neighborhood response, comparison correction, active time scale, intervention flag, model forecast and interval, field observation, responsible reviewer, and disposition. That record creates an auditable basis for learning whether satellite indicators precede, coincide with, or merely reflect maintenance and construction activity.

3.4.7 Limitations and defensible claim boundary

Four limitations govern interpretation of the rail-corridor results. First, independently referenced annual runs preserve within-year change but do not directly measure the December-January intervals; the joined display therefore cannot establish absolute eight-year displacement. Second, the five slope points were retained after reviewing response magnitude, so their behavior characterizes priority locations rather than an unbiased sample of the entire corridor. Third, the three comparison surfaces are neither surveyed monuments nor guaranteed stable ground. Fourth, the dense urban setting permits changes in pavement, wall, vegetation, utilities, and construction surfaces to enter the radar pixel together with real ground movement.

The machine-learning evidence has an additional boundary. The supplied result covers Point03 and reports one model class over three horizons, but the source packet does not yet preserve a complete hyperparameter record, feature ledger, baseline comparison, uncertainty interval, or intervention-stratified error analysis. High R2 on the 2024-2025 holdout is encouraging, yet it does not establish corridor-wide generalization or prospective early-warning performance. Repeating the experiment after freezing point selection and scoring both slope and comparison locations is necessary to determine whether the model learns transferable deformation structure or only the history of one selected signal.
Within these constraints, the defensible annual-report claim is specific: selected slope pixels show repeatedly larger and more persistent within-year LOS behavior than the local comparison ensemble; the acquisition pair bracketing the documented failure does not show one coherent slope-wide step; and a Point03 XGBoost pilot maintains high held-out fit as forecast horizon increases from 12 to 36 days. The record supports condition screening and targeted inspection. It does not yet reconstruct the failure mechanism, identify a universal

velocity threshold, or replace survey, wall, drainage, track-geometry, and geotechnical measurements.

3.5 Summary and Future Work

This reporting period produced an intervention-aware InSAR and forecasting framework for the East 26th Street corridor above the Baltimore CSX Belt Line. Eight annual Sentinel-1 segments were assembled for five priority slope locations and three comparison surfaces, yielding 240 observations per point from 2018 through 2025. The analysis deliberately differs from the highway chapter: it retains LOS as the primary observable, audits annual reference resets, subtracts a local comparison median, evaluates multiple within-year rates, and treats the 2018 failure and subsequent construction as structural breaks.
Point01 has the most negative continuity-adjusted display endpoint (-140.0 mm), whereas Point03 has the steepest fitted display trend (-17.3 mm/yr). Point02 shows the most consistently negative median velocity across two-month, quarter, and six-month windows. The two acquisitions bracketing 26 November 2018 have mixed comparison-corrected signs and do not isolate a coherent slope-wide failure step. Annual sign reversals are larger and more informative, emphasizing the nonstationary effects of failure, emergency stabilization, reconstruction, and later urban modifications.

The extended annual analysis shows that the slope median is more negative than the comparison median in seven of eight calendar-year runs. The largest median separations occur in 2020 (-27.4 mm) and 2022 (-25.4 mm); the groups converge in 2023 before separation returns in 2024-2025. Negative-year counts and within-run increment amplitudes jointly distinguish the selected slope points from most comparison behavior while also confirming that Point07 cannot be treated as a fixed benchmark.

The Point03 XGBoost pilot achieved R^2 values of 0.960, 0.948, and 0.939 at 12-, 24-, and 36-day horizons, respectively, with error increasing as the horizon lengthened. The result demonstrates current forecasting progress but remains a selected-point proof of concept. The next phase will freeze the full processing and machine-learning metadata; validate point-specific geometry or retain LOS targets; add rainfall, soil-moisture, temperature, drainage, train-operation, and construction covariates; compare against simple temporal baselines; and perform chronological evaluation across all slope and comparison locations.

Field validation is the decisive next step. Surveyed control, neighborhood-scale radar statistics, coherence and uncertainty exports, wall and drainage inspections, and track-geometry or ground-instrument measurements are needed to distinguish real slope response from urban surface change and processing residual. With these additions, the chapter's screening workflow can be developed into an auditable rail-corridor decision-support system while preserving a clear boundary between project reporting and later journal publication. Reproducibility work will proceed in parallel with field validation. The project archive should capture the complete 2018-2025 scene and pair inventory, annual references, quality masks, software versions, point-extraction rules, intervention chronology, and a frozen forecasting ledger. These additions will allow the next report to separate improvements in the physical signal from improvements caused only by processing or model revision.

Chapter 4

4 Simulation-Based CaseBank and Machine Learning for Rapid Slope Stability Assessment

Seok-Jun Kang, Zhuping Sheng, Yi Liu, and Oludare Owolabi

4.1 Introduction

Slope instability poses a persistent threat to transportation infrastructure, nearby communities, and the long-term resilience of engineered and natural slopes. Reliable assessment of slope stability is therefore essential for identifying potentially unstable conditions, prioritizing detailed investigations, and supporting effective mitigation planning. The stability condition is commonly quantified using the factor of safety (FoS), which represents the balance between the resisting capacity of the soil mass and the forces driving potential failure. However, FoS can vary substantially with changes in slope geometry, soil strength, unit weight, and groundwater conditions, requiring an assessment approach capable of representing complex and nonlinear relationships among these factors.

Conventional slope stability analyses provide an established basis for engineering design and evaluation. Simplified analytical and limit equilibrium methods are computationally efficient and remain widely used in practice, but their representation of material behavior and failure mechanisms depends on predefined assumptions. Advanced numerical methods can provide a more detailed description of stress redistribution, effective stress conditions, and progressive failure. Nevertheless, repeated numerical analyses across numerous combinations of slope and material conditions can require substantial computational effort. This limitation becomes particularly important when conducting broad parametric evaluations or assessing many slopes over an extensive transportation network.

Data-driven surrogate modeling offers a potential means of reducing this computational burden. Once trained using reliable numerical results, machine learning models can approximate the relationship between slope conditions and FoS and generate predictions much more rapidly than repeated numerical simulations. Their practical value, however, depends strongly on the quality and coverage of the underlying dataset. Field-derived landslide and geotechnical datasets are often limited, spatially heterogeneous, and incomplete, while databases that contain only observed failures may not adequately represent the full range of stable and unstable slope conditions. A systematically generated numerical database can address this limitation by providing consistent input–output relationships over a controlled parameter space.

Building upon the numerical slope stability assessment framework established during the preceding phase of this project, the present study focuses on the development and application of a simulation-based CaseBank. The CaseBank organizes numerical simulation results for combinations of soil unit weight, cohesion, friction angle, slope angle, and groundwater level. In addition to the corresponding FoS, stored numerical states and failure-pattern information can be retrieved for selected cases. Consequently, the CaseBank functions both as a structured dataset for model development and as a searchable engineering reference through which user-defined slope conditions can be compared with previously simulated cases.

The CaseBank is further used to develop machine learning surrogate models for rapid FoS prediction. Multiple regression algorithms are evaluated under a consistent modeling framework to determine their ability to reproduce the nonlinear responses obtained from the numerical simulations. Explainable machine learning techniques are also applied to examine the relative influence and interaction of the input parameters. This interpretive component is important because predictive accuracy alone does not establish whether a model has learned relationships that are consistent with established geotechnical behavior.

Accordingly, the principal objectives of this study are to:

1. construct a systematic simulation-based CaseBank covering a broad range of representative slope, material, and groundwater conditions;
2. establish a query-based procedure for retrieving numerically similar cases, including their FoS values and associated failure information;
3. develop and compare machine learning surrogate models for rapid FoS prediction;
4. interpret the learned parameter effects and interactions from a geotechnical engineering perspective; and
5. establish a scalable foundation for future application to spatially distributed slope stability and landslide risk assessment.

The remainder of this chapter presents a focused review of relevant numerical and data-driven slope stability research, describes the CaseBank and machine learning methodology, evaluates the resulting prediction and interpretation outcomes, and discusses the implications and future development of the framework.

4.2 Literature Review

Slope stability assessment has long been an essential component of geotechnical engineering for the design, maintenance, and risk management of natural and engineered slopes. Traditional approaches are primarily based on analytical methods, limit equilibrium (LE) methods, and numerical methods. Analytical and LE methods have been widely adopted because of their computational efficiency and relatively simple implementation. However, these approaches generally require predefined assumptions regarding the geometry and location of potential failure surfaces, limiting their capability to represent complex failure mechanisms under heterogeneous soil conditions and varying groundwater levels [48-50].

To overcome these limitations, numerical methods such as the finite element method (FEM) and finite difference method (FDM) have been increasingly employed in slope stability analysis. Unlike conventional LE approaches, numerical simulations explicitly consider stress redistribution, deformation, and progressive failure, providing a more physically representative evaluation of slope behavior [51-53]. In particular, strength reduction methods (SRM) implemented within numerical modeling frameworks have become widely accepted for estimating the FoS because they do not require prior assumptions regarding the shape of the critical failure surface [52, 53]. Consequently, numerical simulation has become an effective tool for evaluating complex slope conditions influenced by multiple geotechnical and hydraulic factors.

Despite their advantages, high-fidelity numerical analyses remain computationally demanding, particularly when a large number of simulations are required for parametric studies, probabilistic analyses, or regional-scale assessments. The computational cost increases substantially as combinations of soil properties, slope geometries, and groundwater conditions are considered. This challenge has motivated growing interest in surrogate modeling techniques capable of reproducing numerical simulation results with significantly reduced computational effort [54].

Recent advances in machine learning (ML) have led to its widespread application in geotechnical engineering, including settlement prediction, bearing capacity estimation, tunnel deformation analysis, liquefaction assessment, and slope stability prediction [39, 55-57]. Regression-based algorithms such as Random Forest (RF), Gradient Boosting Machine (GBM), Artificial Neural Networks (ANN), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost) have demonstrated high predictive accuracy for nonlinear geotechnical problems [39, 56-58]. More recently, explainable artificial intelligence (XAI) techniques, particularly SHapley Additive exPlanations (SHAP), have been introduced to improve model transparency by quantifying the relative importance of input variables and interpreting complex nonlinear interactions learned by ML models [59, 60].

Although previous studies have demonstrated the effectiveness of ML-based surrogate models for predicting slope stability, several limitations remain. Many existing models rely on field-derived datasets that are relatively small, geographically specific, or incomplete, limiting their generalizability. Furthermore, most studies focus primarily on improving predictive accuracy without providing an engineering-oriented framework that combines numerical simulation, searchable case retrieval, and interpretable machine learning within a unified system. As a result, practical utilization of ML predictions for engineering decision-making remains limited.

To address these gaps, this study develops a simulation-based CaseBank constructed from systematically generated numerical slope stability analyses. The CaseBank serves both as a structured engineering database that enables retrieval of similar slope conditions and as a high-quality training dataset for machine learning surrogate models. By integrating numerical simulation, case-based querying, predictive modeling, and explainable machine learning, the proposed framework aims to provide a computationally efficient and practically interpretable approach for quantitative slope stability assessment.

4.3 Materials and Methodology

4.3.1 Overall Research Framework

This study developed an integrated workflow that combines physics-based numerical simulations, a structured simulation database, direct case retrieval, machine learning surrogate modeling, and model interpretation. The overall framework is summarized in Figure 4.1. The numerical modeling procedure established during the previous phase of the project was used as the basis for generating reliable slope stability data building upon the validated numerical modeling framework established in our previous annual report [2]. Accordingly, the present study focused primarily on the systematic construction and practical utilization of the CaseBank rather than repeating the detailed comparison and verification of individual slope stability analysis methods.

The workflow consisted of four principal stages. First, representative combinations of soil properties, slope geometry, and groundwater conditions were defined within a prescribed parameter space. Second, factor of safety values and failure information were generated through automated finite difference simulations. Third, the resulting cases were organized into a searchable CaseBank that permits direct retrieval of numerically similar slope conditions. Finally, the CaseBank was used to train and evaluate multiple machine learning models for rapid FoS prediction, followed by SHAP-based interpretation of the learned relationships.

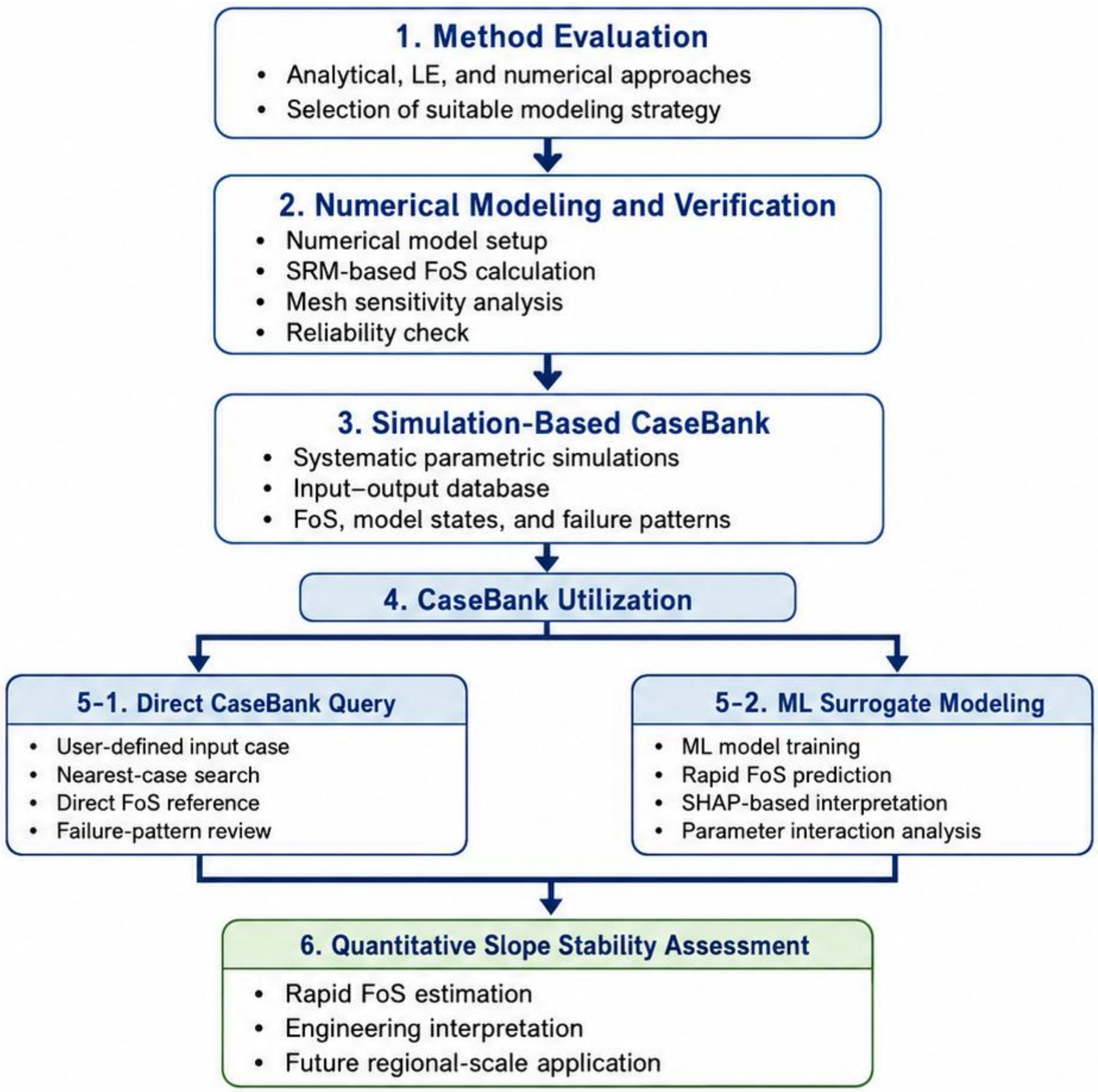


Figure 4.1. Overall workflow of the simulation-based CaseBank and machine learning framework for slope stability assessment.

4.3.2 Simulation-Based CaseBank Development

The CaseBank was generated using the finite difference-based Strength Reduction Method implemented in FLAC3D [62]. The adopted numerical procedure had previously been evaluated against analytical and limit equilibrium approaches and subjected to mesh sensitivity analysis to establish a consistent modeling configuration [61]. A Mohr–Coulomb constitutive model was used to represent soil behavior, and the FoS was determined by progressively reducing the available soil shear strength until a limiting stability condition was reached.

Five variables were selected to define each simulation case: soil unit weight, cohesion, friction angle, slope angle, and groundwater level. These variables represent the principal loading, resistance, geometric, and hydraulic factors controlling slope stability within the adopted modeling framework. Unit weight characterizes the gravitational loading of the soil mass, while cohesion and friction angle define its shear resistance. Slope angle controls the geometric driving condition, and groundwater level represents the influence of pore-water pressure on effective stress and stability.

The selected parameter ranges were intended to cover a broad range of representative soil-slope conditions. Unit weight was varied over typical values for geomaterials, cohesion covered conditions ranging from nearly cohesionless to strongly cohesive soil, and friction angle represented low- to high-strength materials. Multiple slope inclinations and groundwater elevations were also considered to capture both relatively stable and highly unfavorable conditions. The parameter values used for the simulations are summarized in Table 4.1.

Table 4.1. Input parameters used for construction of simulation-based CaseBank.

Parameter	**Unit**	**Values**
Unit weight	kN/m^3	16, 18, 20, 22, 24
Cohesion	kPa	0, 3, 10, 20, 40, 80
Friction angle	Degrees	10, 20, 30, 40, 50
Slope angle	Degrees	10, 20, 30, 40, 50
Groundwater level	% of slope height	0, 20, 40, 60, 80, 100

The parameter combinations were automatically transferred to FLAC3D through a scripted simulation procedure. For each case, the model geometry, material properties, groundwater condition, and strength-reduction analysis were generated and executed using a consistent numerical configuration. The simulation outputs were then collected and assigned a unique Case ID. The resulting CaseBank comprised 4,500 numerical cases, providing a systematic mapping between the five input variables and the corresponding FoS.

Each CaseBank entry contains the Case ID, input parameter values, and computed FoS. In addition to this tabulated input–output information, the numerical model state and a failure-pattern image were retained for each case. The saved model files allow individual simulations to be reopened for further examination of displacement, stress, plasticity state, strength-to-stress ratio, and shear-strain localization. The failure-pattern images provide a more direct representation of the location and geometry of the simulated failure mechanism. Therefore, the CaseBank was designed not merely as a training dataset, but as a structured numerical archive linking quantitative stability indices to their corresponding mechanical responses.

The systematic parameter combinations also produced cases with FoS values below 1.0. These cases represent severe instability conditions in which the prescribed soil strength is insufficient to maintain the initial slope configuration. Although such an intact geometry would generally be transient in the field, these cases were retained because they define the unstable region of the modeled parameter space and provide useful information for training predictive models across both stable and unstable conditions.

4.3.3 Direct CaseBank Query and Case Retrieval

A query-based retrieval procedure was developed to enable direct engineering use of the CaseBank without requiring an additional numerical simulation or machine learning prediction. The user specifies a target slope condition using the same five variables adopted in the database: unit weight, cohesion, friction angle, slope angle, and groundwater level. Because the variables have different units and numerical ranges, each input parameter is first normalized using its minimum and maximum values within the CaseBank. The distance between the user-defined condition and each stored case is then calculated in the normalized parameter space. A Euclidean-distance metric was used to rank the CaseBank entries according to similarity. A smaller distance indicates greater similarity between the query and the stored numerical case.

The procedure returns the nearest cases together with their Case IDs, input conditions, FoS values, and similarity distances. The retrieved Case ID can subsequently be used to locate the corresponding failure-pattern image and saved FLAC3D model. This allows the user to obtain an approximate stability reference and review a representative failure mechanism immediately. The retrieval procedure is particularly useful when the query condition lies close to a simulated combination because the returned result is based directly on the original numerical analysis rather than on an interpolated ML prediction.

4.3.4 Machine Learning Surrogate Modeling and Interpretation

The CaseBank was subsequently used as the training dataset for machine learning-based FoS prediction. The five variables used in the numerical simulations served as model inputs, and the FLAC3D-derived FoS was defined as the continuous target output. Six regression algorithms were considered to compare linear, kernel-based, ensemble, and neural-network modeling approaches: Linear Regression; Support Vector Regression; Random Forest; Gradient Boosting Machine; Artificial Neural Network; and Extreme Gradient Boosting.

The models were developed under a consistent data-processing and evaluation framework. Five-fold cross-validation was applied to reduce dependence on a single division of the dataset and to evaluate the generalization performance of each algorithm. During cross-validation, the CaseBank was divided into five subsets. Four subsets were used for training and the remaining subset was used for validation, with the process repeated until each subset had served as the validation set.

Model performance was evaluated using the coefficient of determination, R^2, and root mean squared error, RMSE. Following the performance comparison, the best-performing model was selected for detailed interpretation using SHapley Additive exPlanations (SHAP). SHAP assigns a contribution value to each input variable for an individual prediction, indicating whether the variable increases or decreases the predicted FoS relative to the average model response [59]. Global feature importance was evaluated using the mean absolute SHAP value, while dependence analyses were used to examine nonlinear trends and interactions among soil strength, slope geometry, and groundwater conditions.

This interpretive analysis was included to assess not only whether the machine learning model could accurately reproduce the numerical results, but also whether the learned parameter effects were consistent with established geotechnical principles. The combined methodology therefore provides three complementary pathways for utilizing the numerical dataset: direct examination of stored CaseBank cases, rapid FoS prediction using a trained surrogate model, and engineering interpretation of the governing input effects.

4.4 Results / Case Study

4.4.1 Development of the Simulation-Based CaseBank

Using the validated numerical modeling framework established during the previous phase of this project, a comprehensive simulation-based CaseBank was constructed to systematically represent a broad range of slope conditions. The database was generated by varying five primary input parameters, namely unit weight, cohesion, friction angle, slope angle, and groundwater level, while maintaining a consistent numerical modeling procedure for all simulations. Each simulation produced a corresponding FoS, resulting in a structured database containing 4,500 numerical cases.

Unlike conventional numerical analyses that require a new simulation whenever site conditions change, the developed CaseBank serves as a reusable engineering database that allows previously simulated conditions to be efficiently accessed and utilized. Each case is assigned a unique Case ID and includes the input parameter values, computed FoS, numerical model files, and representative failure-pattern images. This organization enables both quantitative comparison of stability indices and qualitative examination of failure mechanisms.

The generated CaseBank covers both stable and unstable slope conditions, thereby providing a comprehensive representation of the modeled parameter space (Fig. 4.2). Cases with FoS values below 1.0 were intentionally retained because they define the transition toward instability and provide valuable information for both engineering interpretation and machine learning training.

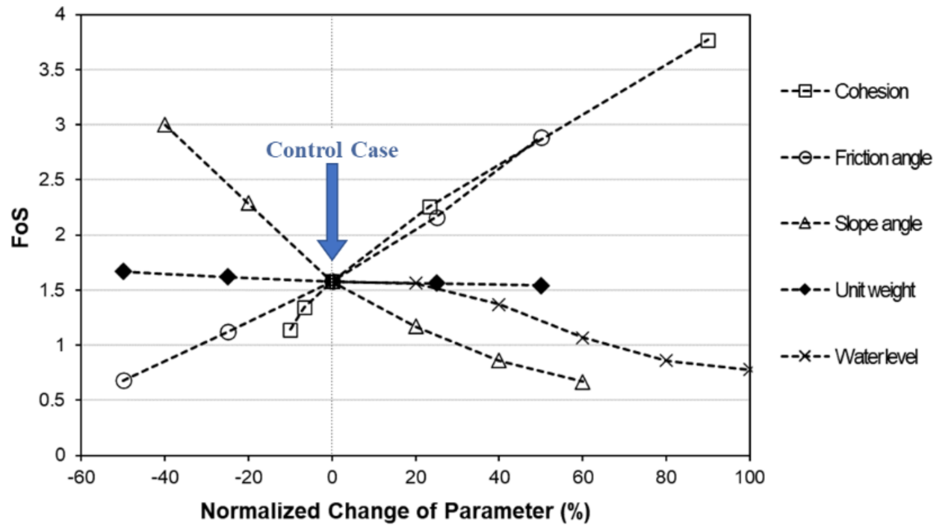


Figure 4.2. Overall distribution of FoS within the developed CaseBank.

To facilitate practical engineering applications, a query-based retrieval system was also developed as summarized in Fig. 4.3. Users can specify a target slope condition by entering the five governing parameters, after which the system automatically searches the CaseBank and identifies the most similar numerical cases based on normalized Euclidean distance. The retrieved results include the corresponding Case IDs, FoS values, similarity scores, and associated failure-pattern images, allowing engineers to rapidly identify representative numerical simulations without performing additional analyses.

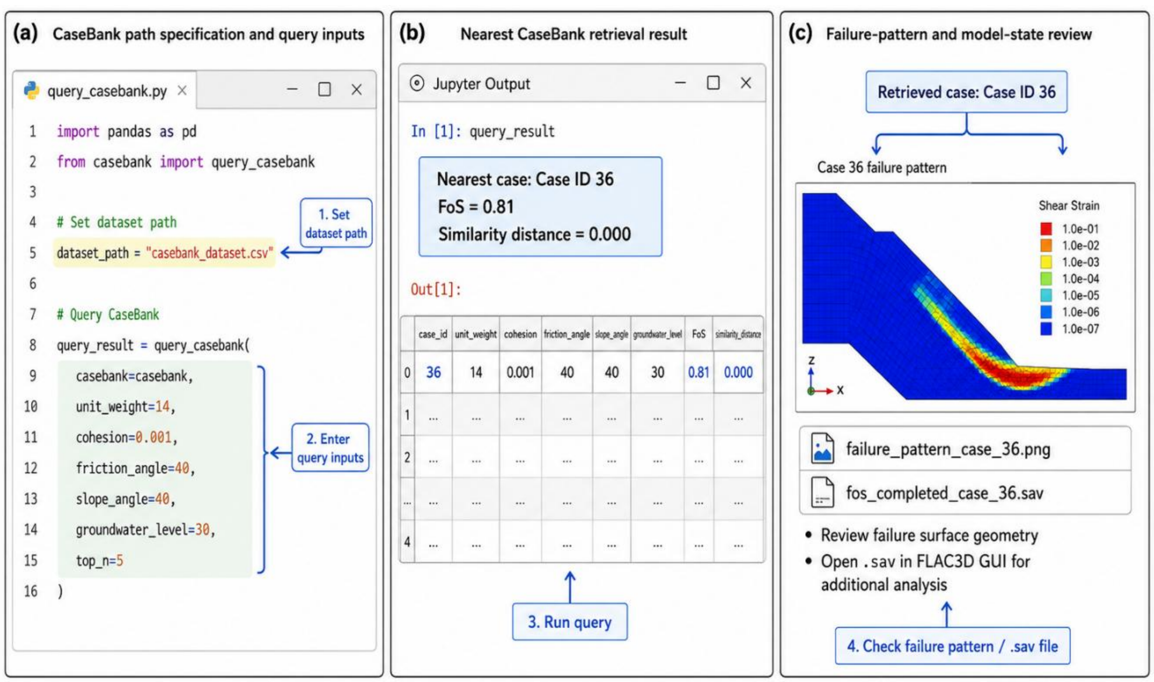


Figure 4.3. Example of the CaseBank query interface and retrieved numerical cases.

The combination of systematic numerical simulations and direct case retrieval establishes the CaseBank as both a numerical database and a practical engineering reference that supports rapid evaluation of slope stability under diverse conditions.

4.4.2 Machine Learning Prediction Performance

While the CaseBank provides direct access to previously simulated cases, many practical situations require predictions for conditions that do not exactly coincide with the discrete parameter combinations stored in the database. To address this limitation, machine learning (ML) surrogate models were developed using the CaseBank as the training dataset.
Six regression algorithms were evaluated (Table 4.2), including Linear Regression (LR), Support Vector Regression (SVR), Random Forest (RF), Gradient Boosting Machine (GBM), Artificial Neural Network (ANN), and Extreme Gradient Boosting (XGBoost). Model performance was assessed using five-fold cross-validation based on the coefficient of determination (R^2) and the root mean squared error (RMSE).

Table 4.2. Performance comparison of the evaluated machine learning models.

Model	**Mean R^2**	**Mean RMSE**
XGBoost	0.999	0.063
GBM	0.998	0.085
SVR	0.998	0.087
ANN (MLP)	0.997	0.112
Random Forest	0.995	0.132
Linear Regression	0.923	0.548

Among the evaluated algorithms, XGBoost achieved the highest predictive accuracy, producing the largest R^2 and the lowest RMSE. Ensemble learning methods consistently outperformed the conventional linear regression model, demonstrating their ability to capture the highly nonlinear relationships among soil strength, slope geometry, and groundwater conditions. ANN also exhibited excellent predictive capability, whereas XGBoost provided slightly better overall accuracy and robustness.

The prediction results indicate that the developed surrogate model successfully reproduces the numerical simulation outputs over the entire range of stability conditions (Fig. 4.4). Consequently, the trained model can provide FoS predictions within a fraction of a second, significantly reducing the computational effort required for repeated slope stability assessments compared with direct numerical simulations.

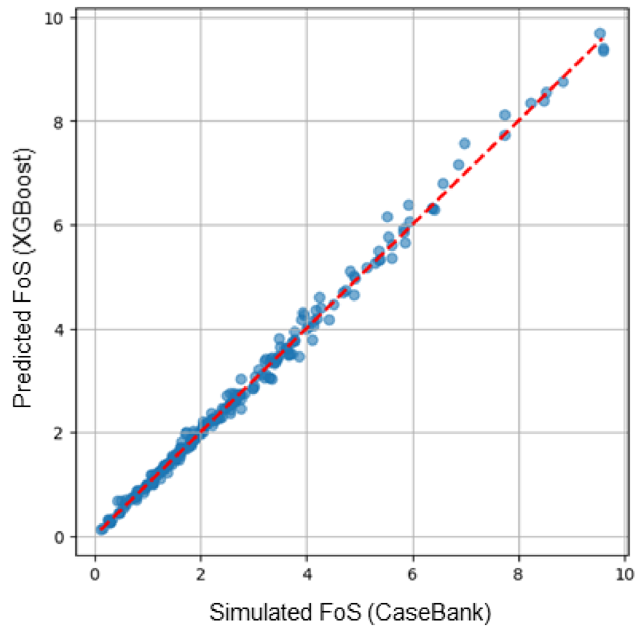


Figure 4.4. Comparison between numerical FoS and ML-predicted FoS for the evaluated algorithms.

4.4.3 Model Interpretation Using SHAP

Although high predictive accuracy is essential, understanding how machine learning models arrive at their predictions is equally important for engineering applications. Therefore, SHapley Additive exPlanations (SHAP) were employed to interpret the predictions of the selected XGBoost model.

The SHAP summary analysis revealed that cohesion and friction angle exert the greatest influence on the predicted FoS, followed by slope angle, groundwater level, and unit weight (Fig. 4.5). This ranking is consistent with established geotechnical principles, where shear strength parameters primarily govern slope stability while groundwater conditions modify effective stress and consequently influence available shear resistance.

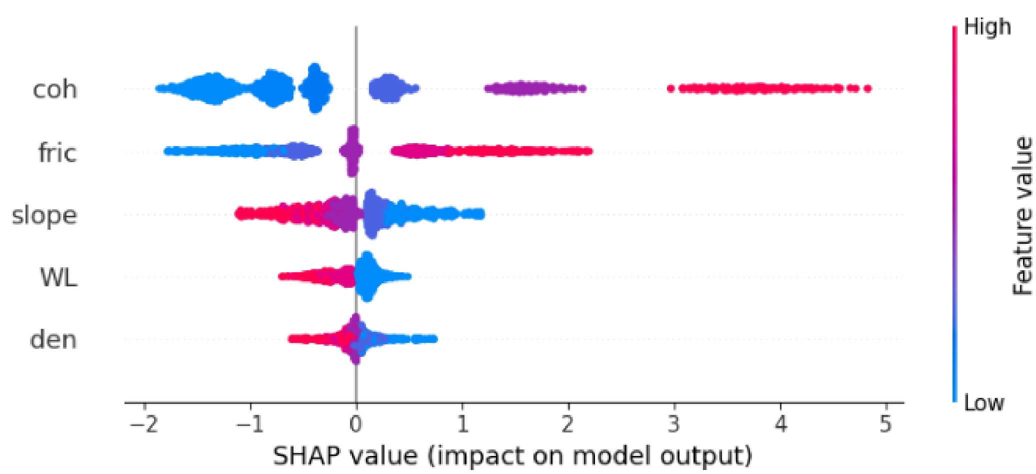


Figure 4.5. SHAP summary plot showing the relative importance of the input variables.

SHAP dependence plots further demonstrated the nonlinear influence of each input parameter on FoS (Fig. 4.6). Increasing cohesion and friction angle generally increased stability, whereas increasing slope angle or groundwater level reduced the predicted FoS. These trends closely matched the physical behavior observed in the numerical simulations. The dependence plots for some variables, particularly cohesion, exhibit discrete rather than continuous distributions. This characteristic originates from the construction of the CaseBank, where input parameters were systematically sampled using discrete values over relatively broad parameter ranges. Consequently, both the numerical simulation results and the corresponding SHAP values appear at discrete intervals instead of forming continuous curves.

Nevertheless, this discrete visualization does not imply that the trained surrogate model is restricted to predicting only the sampled parameter values. Owing to the nonlinear learning capability of the XGBoost algorithm, the model successfully captures the underlying relationships between the sampled data points and can reliably interpolate intermediate conditions within the modeled parameter space. Therefore, while the SHAP plots reflect the discretized nature of the training dataset, the predictive model itself remains capable of providing continuous FoS estimates for user-defined slope conditions.

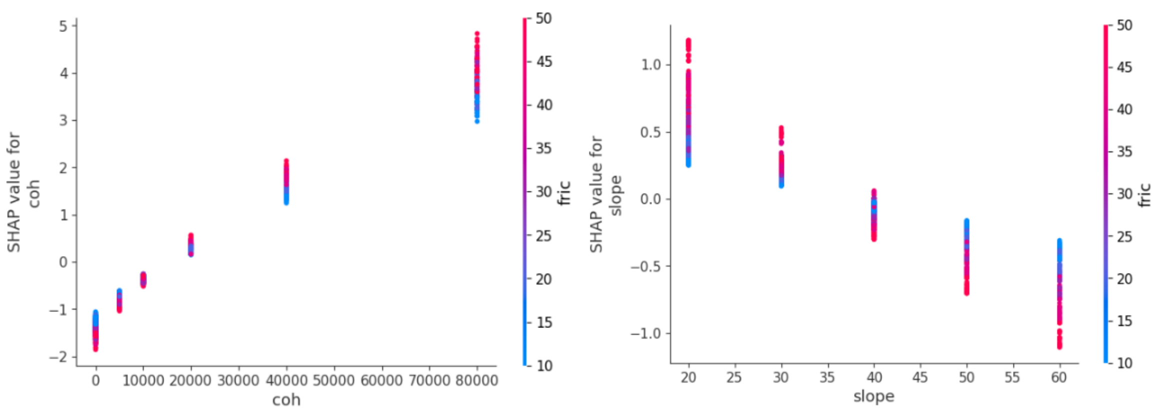


Figure 4.6. SHAP dependence plots for the major input variables.

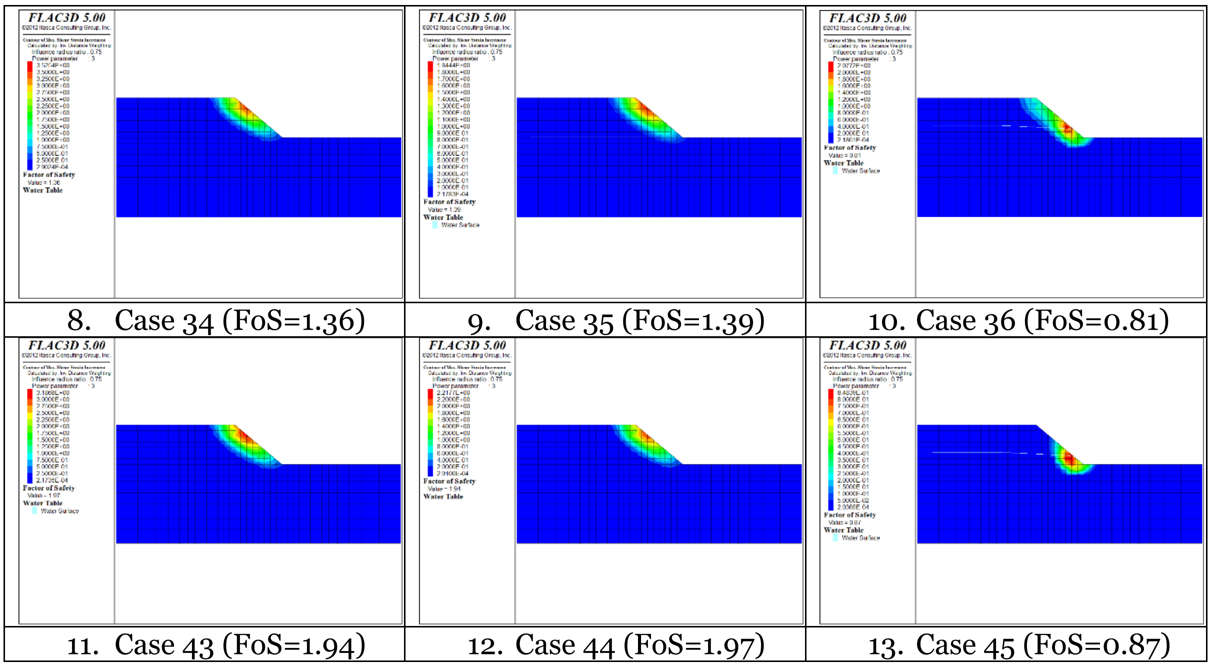
The agreement between the SHAP interpretations and established geotechnical behavior further demonstrates that the machine learning model learned physically meaningful relationships rather than relying solely on statistical correlations.

4.4.4 Engineering Application through Case Retrieval and Rapid FoS Prediction

The proposed framework was designed not only to improve prediction accuracy but also to support practical engineering decision-making. By integrating the CaseBank, query-based retrieval system, and machine learning surrogate model, users can efficiently evaluate slope stability using complementary approaches depending on the available information and required level of detail.

When the input conditions closely match previously simulated cases, the CaseBank enables direct retrieval of representative numerical analyses together with their corresponding failure mechanisms. Conversely, when the desired conditions fall between discrete simulation cases, the trained machine learning model provides rapid FoS predictions without requiring additional numerical simulations. The SHAP analysis further enhances engineering interpretation by explaining how individual parameters contribute to the predicted stability.

To illustrate this workflow, representative CaseBank examples were selected to compare slopes having identical geometric conditions but different groundwater levels and soil strengths (Fig. 4.7). The retrieved cases demonstrate how variations in individual parameters influence both the computed FoS and the associated failure patterns, while the machine learning model consistently reproduces these trends for intermediate conditions that are not explicitly included in the CaseBank.



8. Case 34 (FoS=1.36)
9. Case 35 (FoS=1.39)
10. Case 36 (FoS=0.81)
11. Case 43 (FoS=1.94)
12. Case 44 (FoS=1.97)
13. Case 45 (FoS=0.87)

Figure 4.7. Representative CaseBank retrieval examples for selected slope conditions (Cases 34–36 and Cases 43–45).

The integrated framework therefore combines three complementary capabilities: direct retrieval of previously simulated numerical cases, rapid prediction of unseen conditions using machine learning, and transparent interpretation of model predictions through explainable artificial intelligence. These capabilities provide an efficient decision-support tool that can significantly reduce computational time while preserving the physical consistency established by the underlying numerical simulations. Such an approach is expected to facilitate future regional-scale landslide susceptibility assessment and infrastructure risk management by enabling rapid evaluation of a large number of potential slope conditions.

4.5 Summary and Future Work

This study developed an integrated framework that combines physics-based numerical simulations, a simulation-based CaseBank, machine learning surrogate modeling, and explainable artificial intelligence for rapid quantitative slope stability assessment. Building upon the validated numerical modeling framework established during the previous phase of this project, the present work focused on expanding the practical applicability of numerical slope stability analysis through systematic data generation, efficient case retrieval, and predictive modeling.

A comprehensive CaseBank consisting of 4,500 numerical simulation cases was established by considering representative combinations of soil properties, slope geometry, and groundwater conditions. The resulting database enables engineers to directly retrieve previously simulated cases with similar site conditions while preserving the corresponding numerical models and failure-pattern information. To further improve computational efficiency, multiple machine learning algorithms were trained using the CaseBank, and XGBoost demonstrated the highest predictive performance for estimating the factor of safety. SHAP analysis further confirmed that the learned relationships between the input variables and the predicted FoS are consistent with established geotechnical principles, thereby improving the interpretability and engineering reliability of the developed surrogate model.

The proposed framework demonstrates how physics-based numerical simulations and data-driven modeling can complement each other to support rapid and reliable slope stability assessment. By combining direct CaseBank retrieval, machine learning prediction, and interpretable model analysis, the framework provides a practical decision-support tool capable of substantially reducing computational effort while maintaining consistency with validated numerical simulations.

Future research will focus on extending the current framework toward regional-scale landslide hazard assessment. Planned work includes coupling the developed surrogate model with GIS-based terrain and geospatial datasets to enable rapid screening of large numbers of natural slopes over broad geographic areas. Additional efforts will also investigate the incorporation of heterogeneous field observations, site-specific geological information, and uncertainty quantification to improve the robustness and applicability of the proposed methodology. Ultimately, the integrated framework is expected to support infrastructure management agencies by providing efficient, interpretable, and scalable tools for quantitative landslide risk assessment.

Chapter 5

5 Interpretable Machine Learning Approach for Developing Soil-Specific Probabilistic Rainfall Thresholds for Landslide Initiation in Maryland

Atieh Hosseinizadeh*, Omaria Campbell, Zhuping Sheng, Yi Liu, Oludare Owolabi

Department of Civil and Environmental Engineering, Morgan State University, Baltimore, Maryland

Publication status notice. This chapter documents annual progress for the USDOT Safety21 report. Part of the material in this chapter has been submitted to a journal for consideration of publication.

5.1 Introduction

Landslides represent one of the most destructive natural hazards, causing extensive damage to transportation infrastructure, disrupting economic activities, and threatening human lives worldwide. Rainfall-induced landslides account for a large proportion of documented slope failures, with more than 7,000 events and over 25,000 fatalities reported globally between 2007 and 2015 [62]. In the United States, these hazards continue to pose significant challenges for transportation agencies and emergency management due to increasing infrastructure vulnerability and the growing frequency of extreme precipitation events associated with climate change.

Rainfall is widely recognized as the dominant trigger of shallow landslides [63]. As water infiltrates the soil, pore-water pressure rises while effective stress decreases, leading to a reduction in soil shear strength [64-66]. Extended rainfall can further weaken slopes through groundwater rise, seepage, erosion, and progressive degradation of soil properties [67]. These hydrological processes significantly increase the likelihood of slope failure under both short-duration intense storms and prolonged precipitation events.

The ability to estimate rainfall conditions associated with landslide initiation is essential for hazard mitigation and early warning systems. Reliable rainfall thresholds enable transportation agencies, emergency managers, and local governments to anticipate hazardous conditions, prioritize inspections, and implement preventive actions before failures occur. Such information has become increasingly important as aging infrastructure, expanding urban development, and changing climatic conditions continue to increase landslide susceptibility in many regions.

Maryland is particularly vulnerable to rainfall-related slope failures because of its diverse topography, highly variable soil conditions, extensive transportation network, and increasing occurrence of intense precipitation events. Although landslides have been reported throughout the state, a standardized rainfall threshold specifically developed for Maryland has not yet been established. The absence of a regional threshold limits the development of effective landslide forecasting tools and reduces the ability of transportation agencies to evaluate rainfall-induced hazards consistently.

This technical report addresses this need by developing rainfall threshold models for landslide initiation in Maryland. The study combines spatially distributed rainfall estimates, soil property information, and machine learning techniques to produce probabilistic rainfall thresholds that better represent the uncertainty of landslide occurrence. Unlike conventional deterministic thresholds, the proposed framework estimates different probabilities of slope failure while accounting for local environmental variability, providing a more informative basis for landslide hazard assessment and risk management.

5.2 Related Works

Rainfall threshold analysis has been extensively investigated as a practical approach for assessing rainfall-induced landslide hazards. The fundamental objective of these methods is to identify the minimum rainfall conditions capable of initiating slope failure using observations from historical landslide events. Owing to their simplicity and relatively modest data requirements, empirical rainfall thresholds have become one of the most widely adopted tools for landslide hazard assessment and the development of early warning systems [68-72]. Their application has supported emergency planning, infrastructure management, and disaster risk reduction in many regions worldwide [73-77]-.

Among empirical methods, the rainfall intensity-duration (I–D) threshold remains the most widely recognized formulation. First introduced by Caine [78], this approach represents the relationship between rainfall intensity and storm duration using a power-law equation. Because intense rainfall often produces rapid infiltration and elevated pore-water pressures, I–D thresholds have proven particularly effective for predicting shallow landslides triggered by short-duration storms. Numerous studies have successfully applied this framework under different climatic and geological settings, demonstrating its broad applicability [79-84].

Another widely used approach is based on cumulative event rainfall and rainfall duration, commonly referred to as the event rainfall-duration (E–D) threshold. Rather than emphasizing instantaneous rainfall intensity, E–D relationships quantify the total rainfall accumulated during a rainfall event and therefore better represent the influence of antecedent wetness conditions on slope stability. This methodology has been adopted in numerous regional investigations and has shown promising performance for areas where prolonged precipitation plays a dominant role in triggering landslides [85-90].

Several investigations have further demonstrated that rainfall accumulated over different antecedent periods can substantially influence landslide occurrence. For example, Chleborad et al. [91] identified 3-day and 15-day cumulative rainfall as effective indicators of slope failures in the Seattle metropolitan area, whereas Pennington et al. [92] reported that antecedent rainfall extending over periods as long as 90 days significantly affected landslide activity in the United Kingdom. These findings suggest that rainfall thresholds are highly dependent on local climatic, geological, and hydrological conditions, making regional calibration essential for reliable prediction [93, 94].

Although empirical rainfall thresholds have demonstrated considerable practical value, several limitations remain. A common assumption in many previous studies is that rainfall recorded at one or a few nearby rain gauges adequately represents the precipitation experienced at each landslide location. Because rainfall frequently exhibits substantial spatial variability, particularly during convective storms, this assumption may introduce considerable uncertainty into threshold estimation. Limited rain gauge coverage can therefore produce rainfall estimates that differ significantly from the actual conditions responsible for slope failure [84, 95-97]. Segoni et al. [98] further emphasized that spatial inconsistencies in rainfall measurements may reduce the reliability of empirical thresholds, particularly in regions with heterogeneous precipitation patterns.

Another important challenge concerns the classification of landslide triggering mechanisms. Many empirical studies implicitly assume that all documented landslides were initiated by rainfall, even though slope failures may also result from construction activities, riverbank erosion, infrastructure deterioration, earthquakes, or other anthropogenic and natural processes. Including such events in rainfall threshold development may bias the resulting models and reduce their predictive capability [99]. Furthermore, most conventional threshold studies rely exclusively on landslide observations while ignoring rainfall events that did not produce failures. Consequently, these models have limited ability to distinguish between triggering and non-triggering rainfall conditions.

Traditional rainfall thresholds are also generally deterministic, producing a single critical boundary that separates hazardous from non-hazardous rainfall conditions. However, landslide initiation is inherently uncertain because it is influenced by numerous interacting environmental factors. Binary threshold models therefore cannot adequately represent uncertainty or estimate the probability of failure under varying rainfall conditions [100, 101]. To overcome this limitation, several researchers have proposed probabilistic rainfall thresholds that quantify landslide likelihood rather than providing only a deterministic threshold [102-105].

In addition to rainfall uncertainty, variability in soil properties represents another important source of uncertainty that is frequently overlooked. Soil texture and hydraulic characteristics strongly influence infiltration rates, groundwater movement, pore-water pressure development, and soil shear strength [64, 106]. Nevertheless, many empirical threshold studies implicitly assume uniform soil conditions across the study area. Such simplifications may reduce model reliability because different soil types can respond quite differently under identical rainfall conditions [107, 108].

Recent advances have sought to overcome these shortcomings by integrating spatially distributed rainfall products, additional environmental variables, and machine learning techniques into rainfall threshold modeling. Compared with traditional empirical equations, machine learning algorithms can capture complex nonlinear relationships between rainfall and landslide occurrence while simultaneously incorporating multiple conditioning factors, including soil properties and non-landslide observations. This enables rainfall thresholds to be expressed probabilistically and provides a more flexible framework for landslide hazard assessment than conventional deterministic approaches.

5.3 Materials and Methods

5.3.1 Study Area

Maryland is in the Mid-Atlantic region of the United States and covers approximately 32,130 km². Despite its relatively small geographic extent, the state exhibits considerable diversity in topography, geology, and landforms, ranging from the Appalachian Mountains in western Maryland to the Atlantic Coastal Plain in the east. This variation in terrain results in substantial differences in elevation, slope gradient, and geomorphic characteristics across the state, making Maryland an appropriate location for investigating rainfall-induced landslides.

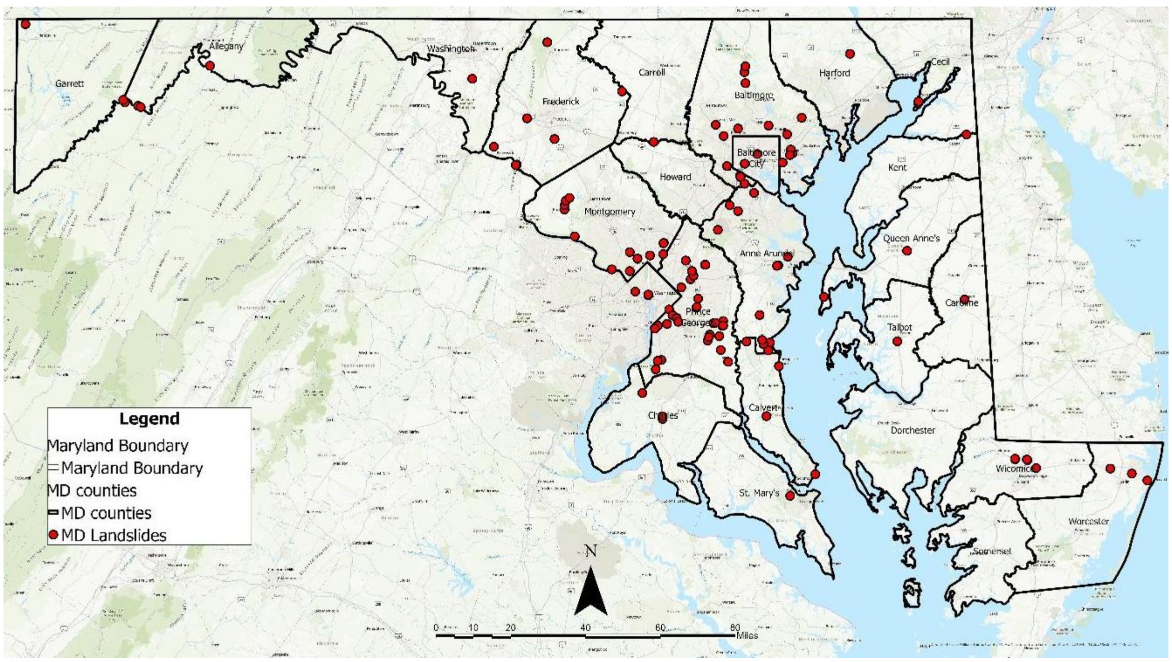


Figure 5.1: The study area and historical landslide locations in Maryland.

Topographic conditions vary significantly among Maryland's physiographic provinces. The western portion of the state, including the Appalachian Plateau and Ridge and Valley regions, is characterized by steep slopes where landslides and other forms of slope instability occur more frequently. Parts of the Piedmont province also experience instability because weathered soils, natural hillsides, and human activities, such as highway construction and excavation, can reduce slope stability. In contrast, the Coastal Plain is generally characterized by low relief and gentle slopes. Although large landslides are less common in this region, localized slope failures, streambank erosion, and embankment instability remain important concerns, particularly in flood-prone areas [109].

Maryland experiences a humid climate with precipitation distributed throughout the year. Rainfall is generated by a variety of meteorological systems, including convective thunderstorms during the summer, frontal systems, and occasional tropical cyclones. Consequently, both high-intensity, short-duration storms and prolonged rainfall events occur regularly, each contributing differently to changes in soil moisture, infiltration, and slope stability.

Another important characteristic of the study area is the spatial variability of soil properties. Maryland contains a wide range of soil textures with varying proportions of clay, silt, and sand [110]. These differences influence infiltration capacity, water retention, drainage characteristics, and the development of pore-water pressure during rainfall events. As a result, slopes across the state do not respond uniformly to precipitation, and soil variability should be considered when evaluating rainfall-induced landslide susceptibility.

Historical observations indicate that many landslides in Maryland are associated with rainfall, particularly in mountainous areas and locations where natural slopes have been modified by human activities. Previous investigations also reported that a large proportion of slope failures occur during or shortly after significant rainfall events, highlighting precipitation as one of the primary triggering mechanisms for landslides in the state [111]. These characteristics make Maryland an appropriate case study for developing rainfall threshold models that incorporate both climatic and soil variability.

5.3.2 Data Preparation

5.3.2.1 Landslide Inventory

The landslide inventory used in this study was compiled by integrating records from the Maryland State Highway Administration (SHA), the United States Geological Survey (USGS), and the NASA Global Landslide Catalog. After combining these datasets and removing duplicate records, a total of 129 documented landslides occurring between 2008 and 2019 were identified for analysis. Most of the recorded events were classified as shallow landslides and were located adjacent to highways or transportation corridors, reflecting the significant impact of slope failures on Maryland's transportation infrastructure.

The spatial distribution of the inventory indicates that landslide occurrence is not uniform across the state. Prince George's County contains the largest concentration of documented events, suggesting that local geological, geomorphological, and climatic conditions contribute to increased landslide susceptibility. The temporal distribution also shows noticeable interannual variability, with the greatest number of landslides occurring during 2014 and 2018. These years correspond to periods of unusually high annual precipitation, indicating a strong relationship between rainfall and landslide occurrence in Maryland [112]. The compiled inventory provides the historical foundation for developing rainfall threshold models and evaluating the influence of rainfall characteristics on landslide initiation.

5.3.2.2 Soil Types and Landslide

Soil information for the study area was obtained from the Soil Survey Geographic Database (SSURGO) developed by the U.S. Department of Agriculture (USDA). This database provides detailed spatial information on soil texture and other physical properties required for evaluating rainfall infiltration and slope response. Some soil polygons located along transportation corridors contained incomplete attribute information. To ensure spatial continuity of the dataset, missing percentages of clay, silt, and sand were estimated using the average values of neighboring soil polygons with similar characteristics. This preprocessing step produced a continuous soil dataset suitable for subsequent analyses.

Analysis of the landslide inventory indicates that failures occurred predominantly within silt loam and sandy loam soils, each representing approximately 35% of the documented landslides (Figure 5.2). To investigate the influence of soil texture on rainfall-induced slope failure, cumulative rainfall associated with landslide occurrence was evaluated separately for each major soil type.

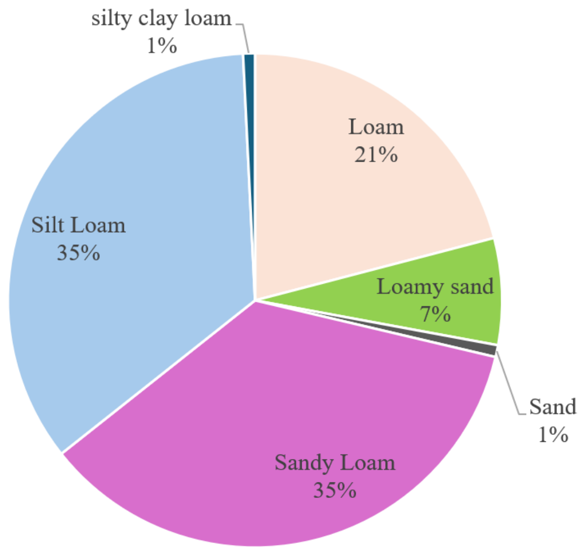


Figure 5.2: The number of landslides vs. soil type.

5.3.2.3 Rainfall Data

Accurate rainfall information is fundamental for establishing reliable rainfall thresholds for landslide initiation. Traditionally, rainfall measurements are obtained from ground-based rain gauge stations because they provide direct observations of precipitation. However, the spatial distribution of rain gauges is often sparse, making it difficult to accurately represent rainfall conditions at individual landslide locations. This limitation becomes particularly important in regions such as Maryland, where complex terrain and localized storm systems can produce substantial spatial variability in precipitation over relatively short distances.

To overcome this limitation, this study utilized spatially distributed rainfall data rather than relying solely on individual rain gauge observations. The use of spatial rainfall estimates provides more representative precipitation information at each landslide location and reduces uncertainty associated with assigning rainfall from distant weather stations. This approach improves the characterization of rainfall conditions responsible for landslide initiation and provides a more robust dataset for rainfall threshold development.

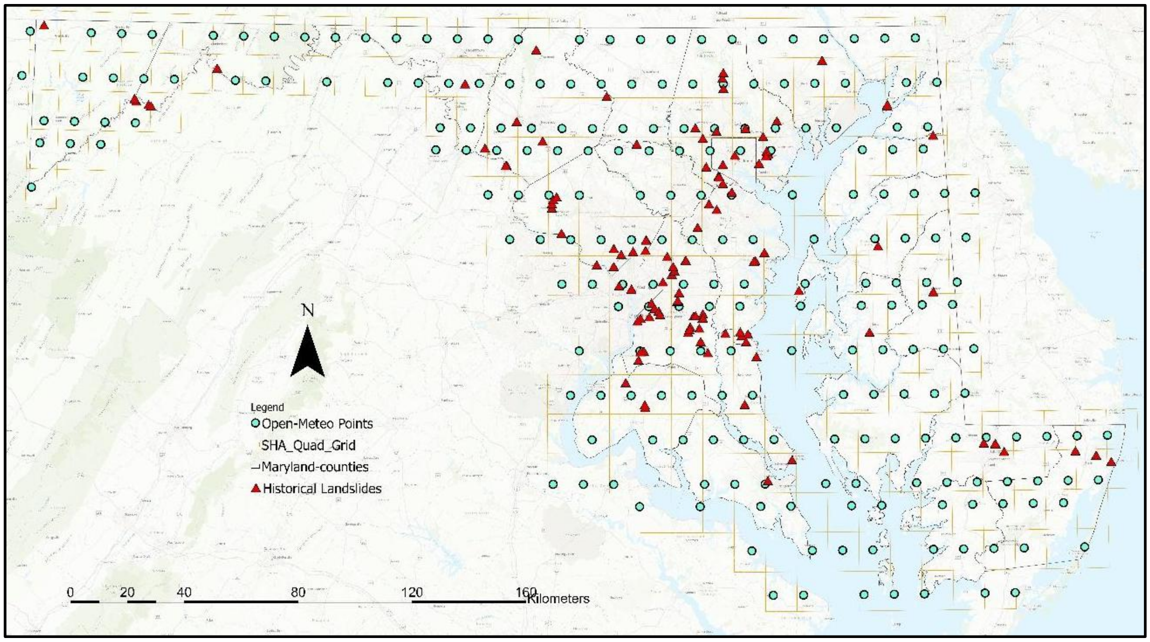


Figure 5.3: Spatial distribution of weather observation points from the Open-Meteo database used for interpolation.

5.3.3 Methodology

A multi-stage framework was developed to establish rainfall thresholds for landslide initiation by integrating rainfall processing, spatial interpolation, extreme value analysis, and machine learning (Figure 5.4). The overall workflow consists of six sequential steps:

1. Preparation of rainfall and landslide datasets.
2. Identification of event-specific critical rainfall conditions.
3. Development of a conventional rainfall threshold.
4. Classification of rainfall-triggered landslides using return period analysis.
5. Development and evaluation of machine learning models.
6. Derivation of probabilistic rainfall threshold equations.

Each component of the methodology is described below.

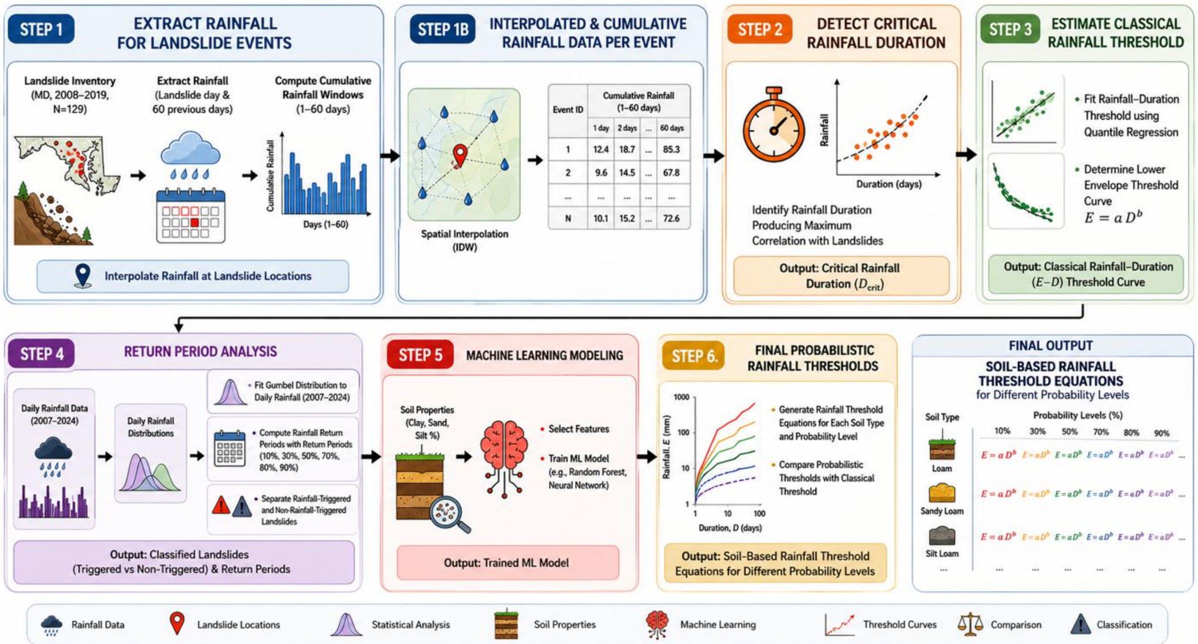


Figure 5.4: Workflow of the study.

5.3.3.1 Rainfall Data Processing

Daily rainfall records from multiple Open-Meteo locations were compiled into a unified database covering the study period. To represent both short-term and antecedent rainfall conditions, cumulative rainfall was calculated for accumulation periods ranging from 1 to 60 days. Because landslides rarely occur exactly at rain gauge locations, rainfall values were spatially interpolated to each landslide location using the Inverse Distance Weighting (IDW) method [113]. This approach assigns greater influence on nearby rainfall observations than to more distant stations, producing rainfall estimates that better represent local precipitation conditions. The interpolation was performed for each landslide date as well as for the preceding 60 days to reconstruct the complete rainfall history associated with every event.

5.3.3.2 Critical Rainfall Identification

The rainfall duration responsible for slope failure is not constant and varies depending on local environmental conditions. Rather than assuming a fixed accumulation period, this study identified the critical rainfall duration individually for each landslide. Rainfall intensity was computed for all accumulation periods between 1 and 60 days, and the duration producing the highest rainfall intensity was selected as the critical duration. The corresponding cumulative rainfall and duration were then used to characterize the triggering rainfall event. This event-specific approach reduces uncertainty associated with selecting an arbitrary rainfall duration and provides a more realistic representation of rainfall conditions preceding landslide initiation.

5.3.3.3 Classical Rainfall Threshold Development

A conventional rainfall threshold was first established to provide a baseline for comparison with the proposed machine learning framework. Following Caine (1980) and

Peruccacci et al. (2012), the relationship between cumulative rainfall and rainfall duration was represented using a power-law equation:

$$E = aD^b \tag{5.1}$$

where $E$is cumulative rainfall (mm), $D$is rainfall duration (days), and $a$and $b$are empirical coefficients.

The threshold parameters were estimated using quantile regression fitted to the lower boundary of the rainfall-duration observations. This approach represents the minimum rainfall conditions associated with landslide initiation while minimizing the influence of extreme rainfall events.

5.3.3.4 Return Period Analysis

Not every recorded landslide is necessarily triggered by rainfall. To distinguish rainfall-induced failures from those caused by other mechanisms, an extreme value analysis was conducted using the Gumbel distribution [114]. For each rainfall duration between 1 and 60 days, return periods were estimated from historical rainfall records. The return period corresponding to each landslide event was then calculated as

$$T = \frac{1}{1-F(R)} \tag{5.2}$$

where $F(R)$is the cumulative probability associated with rainfall magnitude R.

The maximum return period among all rainfall durations was assigned to each landslide event. Events with return periods greater than two years were classified as rainfall-triggered, while events below this threshold were considered less likely to have been initiated primarily by rainfall. This classification reduced the influence of non-rainfall-related failures during model development.

5.3.3.5 Machine Learning Model Development

A supervised machine learning dataset was constructed by combining rainfall-triggered landslides with non-landslide samples. Non-landslide observations were generated from highway slope locations and assigned random dates that excluded documented landslide events. Each sample included rainfall characteristics and site-specific variables describing local conditions. The predictor variables consisted of:

1. Cumulative rainfall for multiple accumulation periods;
2. Critical rainfall amount and corresponding duration;
3. Soil texture information, including clay, silt, and sand contents; and
4. Slope risk information obtained from the Maryland State Highway Administration (SHA).

Five supervised classification algorithms were evaluated:

1. Support Vector Machine (SVM);
2. K-Nearest Neighbors (KNN);
3. Random Forest (RF);

4. Gradient Boosting (GB); and
5. Extreme Gradient Boosting (XGBoost).

The dataset was divided into training (80%) and testing (20%) subsets using stratified sampling to preserve class proportions. Model performance was evaluated using standard classification measures, including Accuracy, Precision, Recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC). The best-performing model was subsequently used for rainfall threshold development.

5.3.3.6 Probabilistic Rainfall Thresholds

Unlike conventional rainfall thresholds, which define a single deterministic boundary, the proposed framework estimates the probability of landslide occurrence under different rainfall conditions. The trained machine learning model generated a probability of landslide occurrence for every sample. Rainfall events exceeding selected probability levels (e.g., 10%, 30%, 50%, 70%, and 80%) were extracted separately for each soil type. For each probability level, rainfall-duration relationships were then represented using the same power-law equation employed in the classical threshold analysis (Eq. 5.1).

This procedure produced a family of soil-specific probabilistic rainfall threshold curves rather than a single deterministic threshold. The resulting thresholds provide a more comprehensive description of landslide susceptibility by accounting for both uncertainty in rainfall triggering conditions and variability in soil characteristics. Finally, the probabilistic thresholds were compared with the classical rainfall threshold to evaluate their differences and demonstrate the benefits of the proposed framework.

5.4 Results

5.4.1 Classical Rainfall Threshold Analysis

The rainfall-duration (E–D) threshold was developed using quantile regression in logarithmic space to estimate the lower boundary of rainfall conditions associated with landslide initiation. A 10% quantile was selected to provide a conservative representation of the rainfall conditions associated with landslide occurrence. The resulting threshold follows a nonlinear power-law relationship between cumulative rainfall and rainfall duration (Figure 5.5).

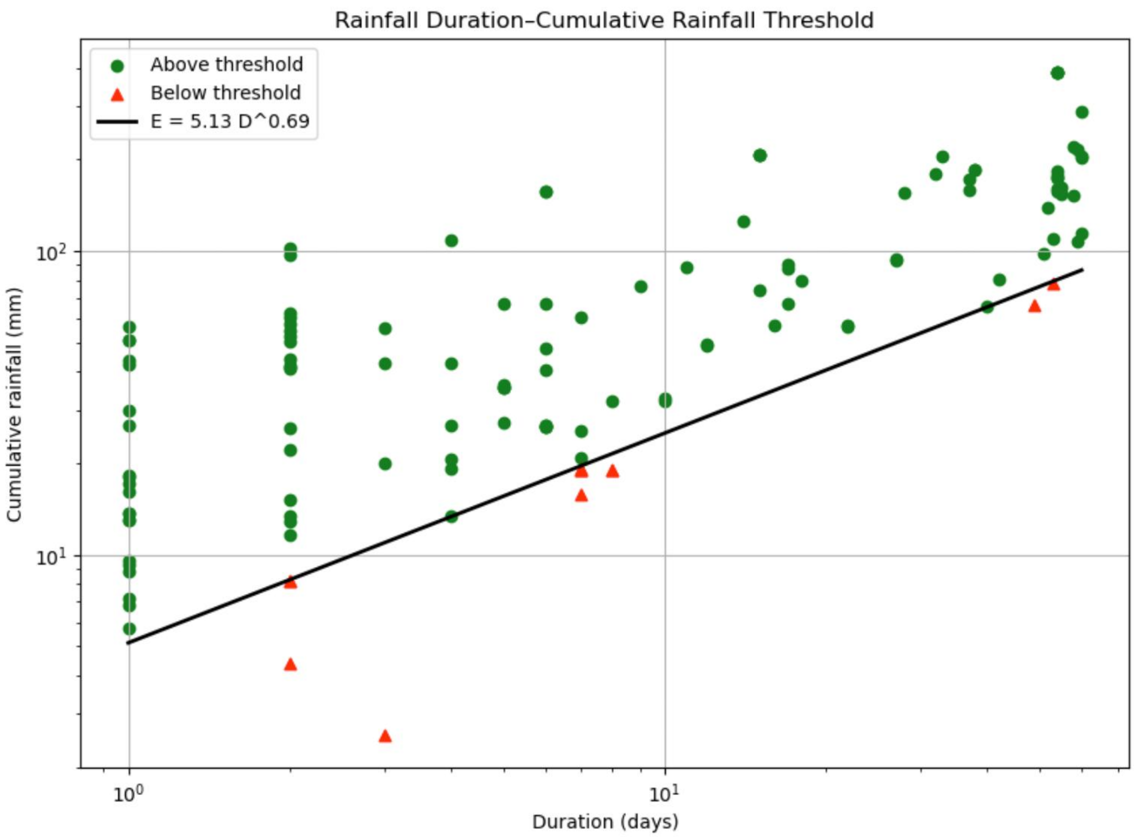


Figure 5.5: Classical E–D rainfall threshold obtained from 10% quantile regression.

5.4.2 Machine Learning Model Performance

Five supervised machine learning algorithms were evaluated to distinguish rainfall-triggered landslides from non-landslide conditions. Their predictive performance was assessed using ROC-AUC, accuracy, precision, recall, and F1-score (Figure 5.6). Among the evaluated algorithms, the Random Forest classifier achieved the best overall performance. Its ROC curve consistently dominated those of the remaining models, producing the highest ROC-AUC value (0.83), which indicates superior discrimination between landslide and non-landslide conditions over a wide range of classification thresholds. Compared with the other classifiers, Random Forest model produced relatively few false positives and false negatives while maintaining high detection rates for both classes. This balance resulted in the highest overall accuracy (0.79), precision (0.75), recall (0.78), and F1-score (0.77), indicating that the model effectively identifies rainfall-triggered landslides without substantially increasing false alarms.

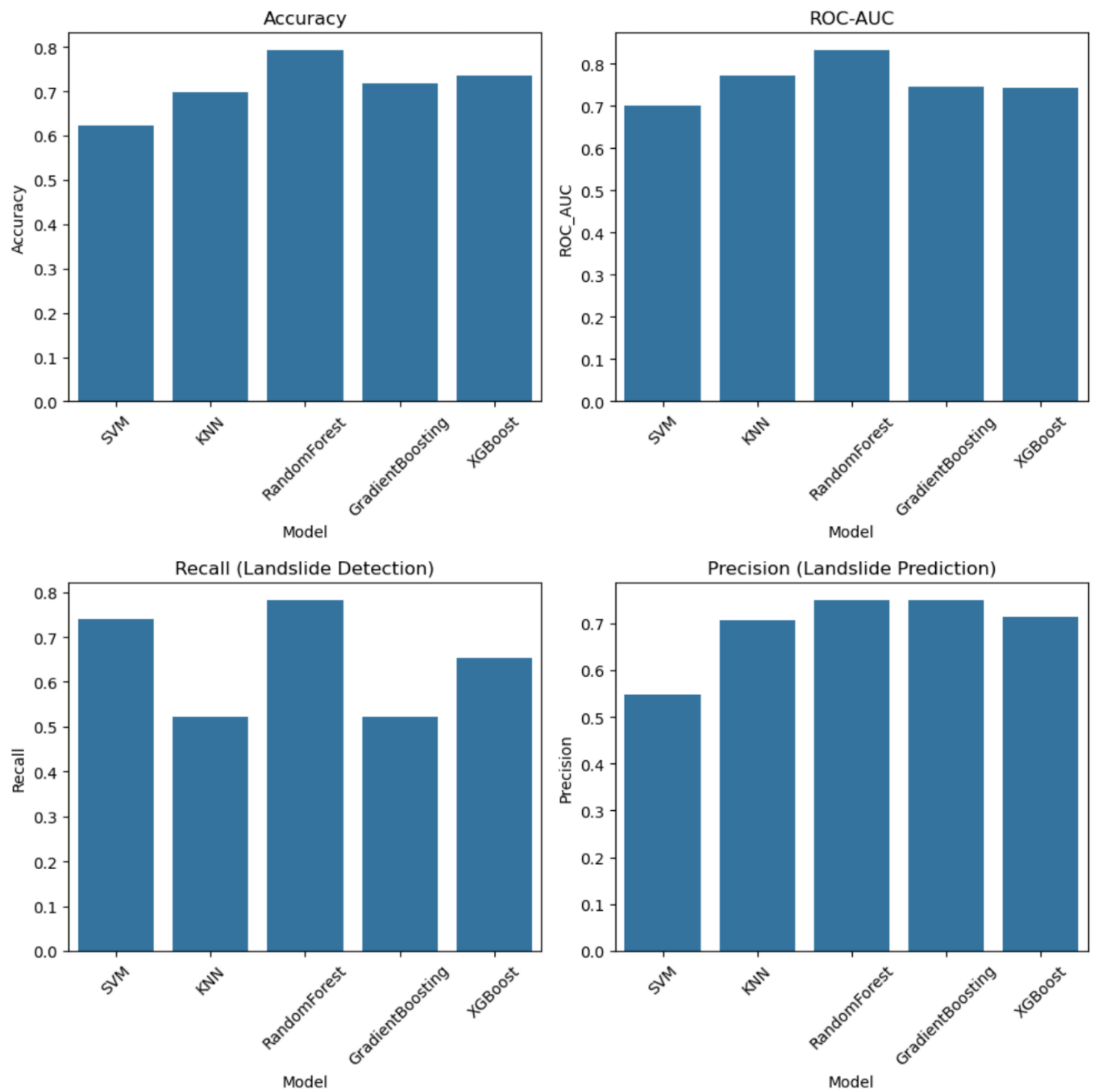


Figure 5.6: Performance of machine learning models for landslide classification.

Although K-Nearest Neighbors and Gradient Boosting achieved comparatively high precision, both models exhibited relatively low recall, indicating that a considerable number of landslide events were not detected. XGBoost provided a more balanced compromise between precision and recall but did not outperform the Random Forest classifier. In contrast, the Support Vector Machine produced the lowest overall performance, primarily because of its relatively high false-positive rate.

These results demonstrate that model evaluation should not rely solely on a single performance measure. Instead, multiple evaluation metrics are required to provide a comprehensive assessment of predictive capability. Based on the combined performance indicators, the Random Forest model was selected for the development of probabilistic rainfall thresholds.

5.4.3 Probabilistic Rainfall Thresholds

The Random Forest model was subsequently used to derive rainfall threshold equations corresponding to different probabilities of landslide occurrence for each soil type. Unlike the classical empirical approach, which produces a single threshold, the machine learning framework generated a family of rainfall-duration curves representing different levels of landslide probability (Figure 5.7).

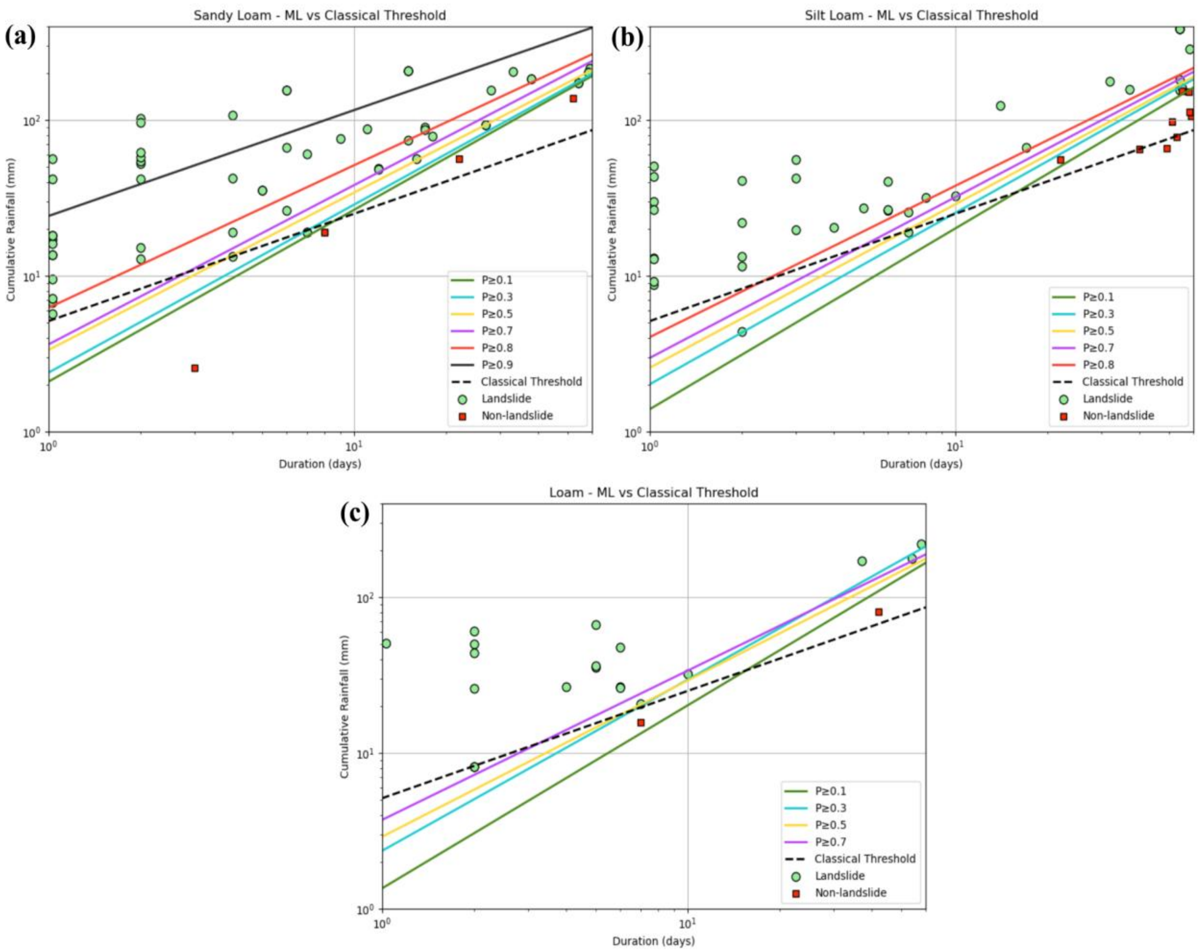


Figure 5.7: Comparison between classical rainfall threshold (10% quantile) and soil-specific probabilistic thresholds derived from the machine learning model (RF) for (a) sandy loam, (b) silt loam, and (c) loam soils.

Comparison with the classical rainfall threshold shows that the deterministic threshold generally falls within the intermediate probability range (approximately 30–50%). Lower probability thresholds are positioned below the classical curve, indicating rainfall conditions capable of triggering landslides even under relatively modest rainfall accumulations. Conversely, higher probability thresholds identify progressively more extreme rainfall events that are associated with a greater likelihood of slope failure. The differences between probabilistic and classical thresholds are particularly evident for short rainfall durations. Under these conditions, the machine learning thresholds predict landslide occurrence at lower cumulative rainfall amounts than the empirical threshold, suggesting that the deterministic approach may underestimate susceptibility during intense short-duration storms. As rainfall duration increases, the probabilistic and classical thresholds become progressively closer, indicating that cumulative rainfall over longer periods provides a more stable indicator of landslide occurrence.

The probabilistic framework also reveals substantial differences among soil types. Sandy loam soils consistently exhibit the lowest threshold curves, indicating greater sensitivity to rainfall and a higher probability of failure under comparatively small rainfall accumulations. In contrast, silt loam soils generally require larger rainfall totals to reach the same probability levels, while loam soils display intermediate behavior. These observations demonstrate that rainfall thresholds vary considerably with soil texture and that applying a single statewide threshold may not adequately represent local conditions.

5.4.4 Discussion

The results demonstrate several advantages of integrating machine learning with conventional rainfall-threshold analysis. First, the probabilistic framework overcomes one of the primary limitations of traditional empirical thresholds by replacing a single deterministic boundary with a continuous range of rainfall thresholds associated with different probabilities of landslide occurrence. This provides decision-makers with greater flexibility to select warning criteria appropriate for different operational objectives and risk tolerances.

Second, incorporating soil properties into the machine learning model allows rainfall thresholds to account for spatial variability in slope response. The results clearly indicate that rainfall sensitivity differs among loam, sandy loam, and silt loam soils, suggesting that soil-specific thresholds provide a more realistic representation of landslide triggering conditions than a single statewide relationship.

Finally, the comparison between probabilistic and classical thresholds highlights the benefits of considering nonlinear interactions among rainfall, soil characteristics, and landslide occurrence. While the classical rainfall threshold adequately represents average triggering conditions, the machine learning framework captures a wider range of possible rainfall scenarios and more effectively represents the uncertainty inherent in rainfall-induced landslides. These characteristics make the proposed methodology particularly suitable for future landslide early warning systems and risk-informed infrastructure management.

5.5 Conclusion and Future Work

This study developed a machine learning-based framework for deriving soil-specific probabilistic rainfall thresholds for landslide initiation in Maryland. By integrating spatially interpolated rainfall data, return period analysis, soil characteristics, and supervised machine learning, the proposed methodology addresses several limitations of conventional empirical rainfall threshold approaches. In particular, the framework accounts for spatial variability in rainfall, distinguishes rainfall-triggered from non-rainfall-triggered landslides, and incorporates differences in soil properties that influence slope response.

Among the evaluated machine learning algorithms, the Random Forest model provided the best predictive performance, achieving an overall accuracy of 0.79, a recall of 0.78, and a ROC-AUC of 0.83. These results indicate that the model effectively captures the complex nonlinear relationships between rainfall characteristics, soil texture, and landslide occurrence, making it well suited for probabilistic landslide prediction.

The probabilistic analysis revealed clear differences in rainfall sensitivity among soil types. Sandy loam soils exhibited the lowest rainfall thresholds and therefore the greatest susceptibility to rainfall-induced failure, whereas silt loam soils generally required larger cumulative rainfall amounts before landslides occurred. Loam soils displayed intermediate behavior. These findings demonstrate that applying a single rainfall threshold across all soil conditions may not adequately represent the spatial variability of landslide susceptibility.

Unlike conventional deterministic thresholds, the proposed framework produces multiple rainfall threshold equations corresponding to different probabilities of landslide occurrence. This enables rainfall thresholds to be selected according to the desired balance between early warning sensitivity and false-alarm reduction. Lower probability thresholds maximize landslide detection and are suitable for conservative warning systems, whereas higher probability thresholds provide greater confidence at the expense of missing some events. This flexibility makes the proposed approach more practical for operational landslide forecasting and infrastructure risk management.

Overall, the proposed methodology combines the physical interpretability of traditional rainfall-duration relationships with the predictive capability of machine learning. The resulting soil-specific probabilistic thresholds provide a more comprehensive representation of rainfall-induced landslide processes than conventional empirical methods and offer a practical framework for improving landslide hazard assessment and supporting transportation agencies in developing risk-informed early warning systems.

Future research could further enhance the proposed framework by incorporating additional conditioning factors, including land cover, slope geometry, antecedent soil moisture, groundwater conditions, and high-resolution remote sensing products. Expanding the methodology to larger geographic regions and validating its performance under different climatic and geological settings would also improve its applicability for regional and national landslide hazard assessment.

6 Physics-Informed Multi-Sensor ML/DL Benchmarking and LiDAR Monitoring for Rainfall-Induced Landslide Early Warning

Rezoan Sultan[1], Chinyanta Mushambo[1], Zhuping Sheng[2], Kofi Nyarko[1], Yi Liu[2], Oludare Owolabi[2]

[1]Departments of Electrical and Computer Engineering, Morgan State University, Baltimore, Maryland

[2]Department of Civil and Environmental Engineering, Morgan State University, Baltimore, Maryland

Publication status notice. This chapter documents annual progress for the USDOT Safety21 report. The part of the material in this chapter has been submitted to TRB 2027 for consideration of publication in conference proceedings or TRB journals.

6.1 Introduction

Problem statement. In the United States, landslides triggered by rainfall cause a costly and consistent threat to transportation infrastructure, including damage to roads, rail corridors, and slopes along highway rights-of-way, as well as occasionally blocking routes for public safety [115]. Slope geometry, material strength, and the transitory pore-water pressure caused by rainfall infiltration all interact to fundamentally regulate the occurrence of landslides [116,117, 64]. Even after decades of research on rainfall thresholds [118, 119] and GIS-based susceptibility mapping [120], there is still a lack of evidence for affordable monitoring systems that integrate surface deformation tracking with subsurface precursor sensing in controlled, instrumented environments. Most rainfall-threshold and susceptibility-mapping approaches are static or empirical, and few operational systems combine physics-informed subsurface labeling with repeat-epoch surface geometry monitoring at the scale of an individual slope.

Objectives. This chapter describes the initial stages of developing a two-thread early-warning system for slope instability caused by rainfall, based on an open-field prototype landslide testbed. A pore-water-pressure transducer, a rain gauge, and multi-depth soil moisture sensors are used in Thread 1 (Sensor Fusion) to create a physics-informed instability risk indicator that is benchmarked across classical and deep-learning sequence models. Thread 2 (LiDAR Monitoring) uses a consumer-grade 940 nm LiDAR camera to periodically scan slope geometry, deriving a terrain-based susceptibility index from a single epoch and establishing the change-detection methodology (M3C2) that will validate that index once repeat epochs are collected. The specific objectives are to: (1) construct and validate a physics-informed risk-labeling scheme on an independent benchmark dataset in advance of prototype data availability; (2) benchmark classical (XGBoost) against deep sequence models (LSTM, BiLSTM, TCN) for instability prediction and quantify achievable warning lead time; (3) establish a terrestrial LiDAR data-collection and processing pipeline (registration, terrain-feature extraction, and a first-pass susceptibility index) on a real field site; and (4) define a concrete, testable methodology for validating that susceptibility index against observed surface displacement once periodic re-scanning begins. The central result reported in this chapter is that the physics-informed labeling strategy is sound in construction; a diagnostic check described in Section 6.4.1 identified a methodological flaw in how warning lead time was originally measured on this single-event benchmark dataset, and that check, and its implications, are reported directly rather than the lead-time figure originally computed.

Scope of this chapter. Because the physical prototype testbed is still accumulating its own longitudinal sensor record, Thread 1 results reported here are obtained on an independent Zhuji landslide benchmark dataset, used explicitly as a proxy validation corpus rather than as the terminal dataset for the project. Thread 2 results are reported from the first LiDAR epoch collected at a real monitoring site (referred to throughout as “Cumberland”), together with a worked pilot demonstration of the change-detection methodology on a separate two-epoch practice dataset (“Monitoring Site 1”). Fusion of the two threads into a single combined instability probability is identified as future work (Section 6.5) and is not reported as a result here.

6.2 Related Works

This section situates the proposed framework relative to four bodies of literature: (i) the physical mechanics and classification of rainfall-induced landslides; (ii) empirical rainfall-threshold and early-warning approaches; (iii) machine-learning and deep-learning methods for landslide prediction and susceptibility mapping; and (iv) point-cloud and remote-sensing methods for slope deformation monitoring.

Landslide susceptibility is governed by the interaction between slope material strength and pore-water pressure [116], with the modern classification of failure types and processes established by Cruden and Varnes [121] and elaborated by Sidle and Ochiai [117]. Iverson (2000) provided the theoretical basis linking rainfall infiltration to transient pore-pressure generation at varying depths and timescales, which remains the physical justification for multi-depth soil-moisture instrumentation of the kind used in Thread 1 of this study.

Empirical rainfall intensity–duration (ID) thresholds [118] remain the most widely deployed landslide early-warning tool because of their simplicity, but Segoni et al.[119] and Baum and Godt [122] both note that ID thresholds alone neglect antecedent soil moisture and site-specific hydrogeology, motivating multi-parameter systems that combine rainfall with direct subsurface monitoring [123]. This is precisely the gap Thread 1 addresses: rather than relying on rainfall alone, a physics-informed risk index is constructed directly from multi-depth volumetric water content and its rate of change.

Machine-learning approaches to landslide susceptibility and prediction have grown rapidly, most commonly applying tree-based ensembles such as Random Forest [124] and gradient-boosted trees, particularly XGBoost [125], to static conditioning-factor stacks (slope, aspect, curvature, land cover, distance to drainage). For time-series precursor signals specifically, recurrent architectures such as Long Short-Term Memory networks [126] and their bidirectional variant [127] are the conventional choice, while Temporal Convolutional Networks [128] have more recently been shown to match or exceed recurrent architectures on generic sequence-modeling benchmarks. All deep models in this study were trained with the Adam optimizer [129]. A key limitation of purely data-driven labeling for landslide precursor detection is the general absence of confirmed failure timestamps in field datasets; this study instead adopts a physics-informed labeling strategy in the spirit of physics-informed neural networks [130], constructing a continuous instability index from first-principles hydrological reasoning rather than requiring a labeled failure event.

Terrestrial laser scanning (TLS) has been used to monitor slope deformation for over a decade, with Multiscale Model-to-Model Cloud Comparison (M3C2; [131]) established as the standard method for measuring surface change directly along the local normal direction on complex, non-planar natural terrain, outperforming simple cloud-to-cloud or DEM-differencing approaches on rough or vegetated slopes. Multi-station point-cloud registration, a prerequisite for both single-epoch merging and multi-epoch change detection, relies on the Iterative Closest Point algorithm [132, 133] for local refinement, and on feature-based global methods, combining Random Sample Consensus [134] with Fast Point Feature Histograms [135], for robust coarse alignment where a good initial pose is not available, a scenario encountered directly in this study's Cumberland site registration (Section 6.4.2).

Deep learning on raw point clouds, rather than gridded raster derivatives, is increasingly used for terrain classification and segmentation tasks, with PointNet++ [136] established as the standard hierarchical architecture for learning directly on unordered point sets (identified in this study as the planned deep-learning counterpart to M3C2) (Section 6.5). At a broader spatial scale, spaceborne interferometric synthetic aperture radar (InSAR) offers a complementary, wide-area deformation-monitoring capability [137], though its coarser spatial resolution and sensitivity to vegetation and acquisition geometry make it better suited to regional screening than to the individual-slope, high-resolution monitoring this study targets.

A bibliometric review of GIS-based landslide susceptibility research [120] confirms that studies combining ground-based sensor networks, machine-learning benchmarking, and point-cloud-based deformation monitoring within a single, controlled physical testbed remain rare. This project is conducted as part of the broader University Transportation Center Safety-21 landslide research program, alongside related efforts applying ConvLSTM-based soil-moisture mapping [112], numerical-simulation-guided machine learning for landslide risk assessment [138], and multi-temporal LiDAR with the Analytic Hierarchy Process for susceptibility mapping [139, 140]. Table 6.1 summarizes how the present framework's combination of elements relates to this surrounding body of work.

Table 6.1. Comparison of this framework's capabilities against three related monitoring paradigms.

Capability	**Rainfall-threshold LEWS**	**GIS/ML susceptibility mapping**	**TLS/InSAR deformation monitoring**	**This study**
Physics-informed subsurface labeling	Partial	No	No	Yes
Multi-depth in-situ sensor fusion	Sometimes	No	No	Yes
Classical vs. deep sequence-model benchmarking	No	Sometimes	No	Yes
Repeat-epoch point-cloud change detection (M3C2)	No	No	Yes	Yes (planned validation)
Controlled, purpose-built physical testbed	Rarely	No	No	Yes
Consumer/prosumer-grade LiDAR instrument	No	No	Rarely	Yes

6.3 Materials and Methods

The framework is organized into two parallel threads sharing a common objective that's estimating slope-instability risk but drawing on different physical signals and, at the current stage, different validation data. Figure 6.1 summarizes the overall architecture, including the fusion-layer weighting scheme identified as future work once both threads have real, co-located field data.

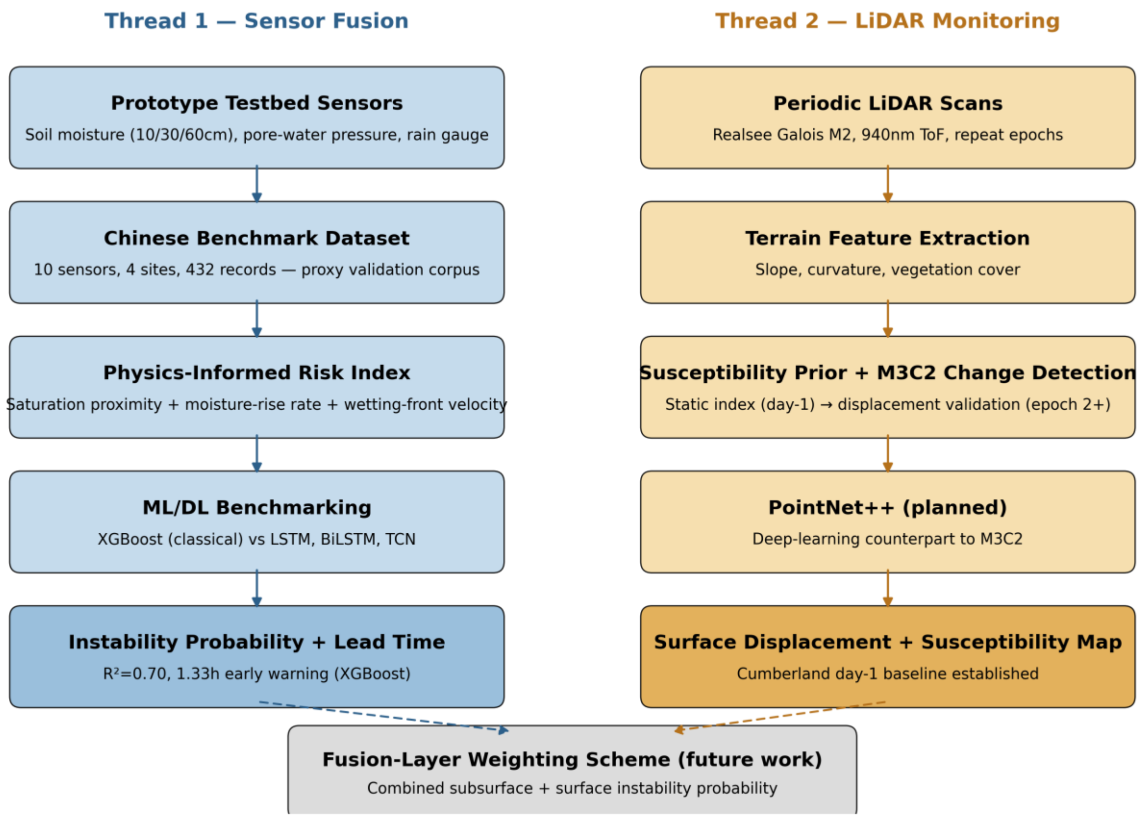


Figure 6.1. Two-thread project architecture: Sensor Fusion (Thread 1) and LiDAR Monitoring (Thread 2), converging at a future fusion layer.

6.3.1 Study design

The overall two-thread architecture is summarized in Figure 6.1 above.

6.3.2 Data Sources

This chapter draws on two categories of data: a third-party benchmark dataset used for Thread 1, and original field data collected by the authors for Thread 2.

The Zhuji benchmark dataset in China described in Table 6.2 was obtained from the Soil Moisture and Pore Pressure Observation Data component of a five-part dataset published by Wu, Lü, Liao, Yu, and Wang [141] The data was collected at four monitoring sites (W1–W4) on a landslide-prone hillslope near Zhuji, eastern China (29°42′05.68″N, 120°00′45″E). The dataset is openly available on Zenodo under a Creative Commons Attribution 4.0 license (DOI: 10.5281/zenodo.16748046); the site naming and staircase depth-sensor configuration used throughout this chapter (Section 6.3.3) directly reflect its published structure. Each site is instrumented with an SCYG319 pore-water-pressure probe and a set of CSF13 soil-moisture sensors (Star Meter Ltd.).

The LiDAR point-cloud data used in Thread 2 (Section 6.3.4) are original data collected by the authors with a Realsee Galois M2 LiDAR camera at two sites: Monitoring Site 1 (39°25′16.1″N, 76°30′14.7″W), a proof-of-concept practice location, and Cumberland (39.64288°N, 78.79660°W), the project's primary monitoring site along the I-68 corridor in western Maryland.

6.3.3 Thread 1: Sensor fusion methodology

The prototype testbed is an open-field installation instrumented with soil moisture sensors at three depths (10, 30, and 60 cm), a pore-water-pressure transducer, and a rain gauge, alongside a consumer-grade LiDAR camera for periodic slope-surface scanning (Section 6.3.4). Because the prototype's own longitudinal record is still being collected under natural rainfall, Thread 1 model development and validation in this chapter use an independent Zhuji landslide field dataset as a proxy corpus (Table 6.2), following the same physics-informed labeling logic that will subsequently be applied to the prototype's own sensor stream.

Table 6.2. Zhuji landslide benchmark dataset in China used as Thread 1 proxy validation corpus.

Attribute	**Value**
Monitoring sites	4 (W1–W4)
Sensor columns per site	Staircase depth configuration (D1–D4)
Total sensor columns	10
Temporal resolution	10 minutes
Total records	432
Event captured	3-day rainfall event, March 2025
Confirmed failure event	None — infiltration precursor signal only

Because the benchmark dataset contains no confirmed failure timestamp, a supervised target could not be built from a ground-truth label. Instead, a continuous instability risk index $I(t) \in [0, 1]$ was constructed as a weighted composite of three physically motivated components, each independently normalized to [0, 1]:

$$I(t) = 0.50 \cdot S(t) + 0.30 \cdot R(t) + 0.20 \cdot V(t)$$

S(t), or saturation proximity (weight 0.50): the degree to which volumetric water content at each depth approaches a site-specific saturation threshold, the dominant control on shear-strength loss (Iverson, 2000).

R(t), or rate of moisture rise (weight 0.30): the short-window time derivative of volumetric water content, capturing the wetting-front arrival signal described by Iverson's (2000) infiltration model.

V(t), or wetting-front velocity (weight 0.20): the propagation rate of the moisture front between sensor depths, providing a secondary precursor signal independent of absolute moisture content.

Figure 6.2 shows the full diagnostic derivation of I(t) for monitoring site W3 during the March 2025 event: raw volumetric water content at three depths, each of the three weighted components, and the resulting composite risk index with alert (0.4) and critical (0.7) thresholds annotated. The index increases dramatically after the start of rainfall, peaks soon after the deepest sensor (D3, assumed nominal depth ~60 cm by analogy with the prototype's own sensor spacing; exact sensor depths are not published in the source dataset) gets close to its observed saturation ceiling (47.1% VWC) and then declines over the course of the next day. This is consistent with a physically significant infiltration and drainage cycle rather than a labeling scheme artifact.

Four model families were benchmarked against the risk-index target: XGBoost [125] as the sole classical baseline and Random Forest was evaluated during framework development but dropped from the final comparison in favor of a single, well-tuned classical benchmark and three deep sequence architectures: LSTM [126], BiLSTM [127], and a Temporal Convolutional Network [128]. Deep models were trained for 80 epochs with the Adam optimizer [129] using a mean-squared-error loss on the continuous risk index, with a held-out validation split monitored for convergence (Figure 6.11 in later section). The dataset was partitioned so that the first day (dry baseline) and the first half of the rainfall event formed the training window, with the second half of the event and the recovery period held out for testing the most conservative split available given the single-event dataset, since it requires each model to extrapolate through the actual peak-instability transition rather than interpolate within it. Performance was evaluated using RMSE, MAE, R^2, an alert-classification ROC-AUC (risk index crossing the 0.4 alert threshold), and warning lead time; the interval between first alert and observed peak instability. An ablation study across six sensor-depth configurations (full model, deep-only, shallow-only, mid-only, and single-depth variants) quantified the marginal contribution of each sensing tier.

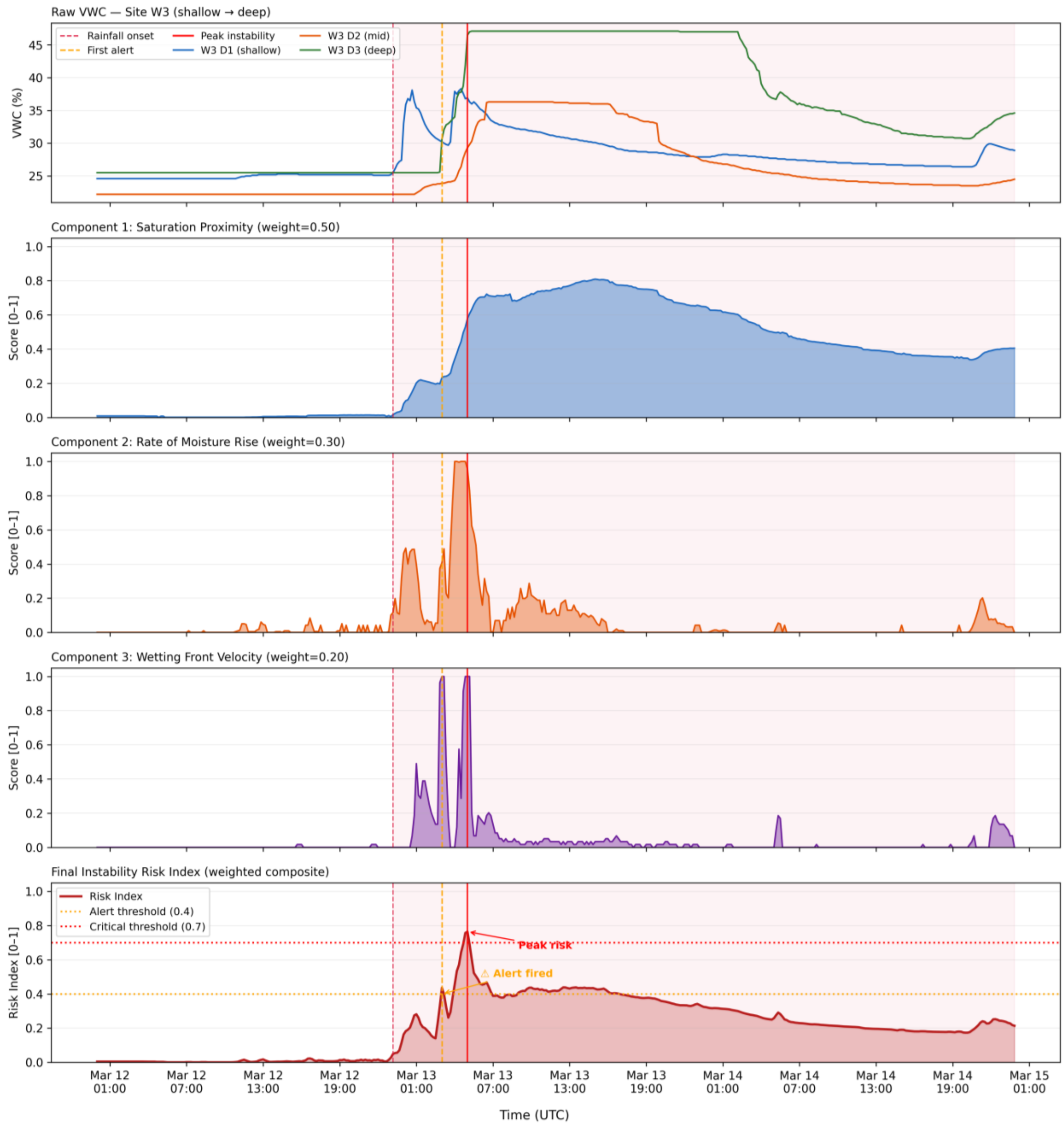


Figure 6.2. Physics-informed instability risk index diagnostic derivation, site W3, March 2025 event.

6.3.4 Thread 2: LiDAR monitoring methodology

Thread 2 uses a Realsee Galois M2 LiDAR camera which is a 940 nm time-of-flight instrument with a separate CMOS sensor for color panoramas, distinct from structured-light or stereo RGB-D depth cameras [142] to periodically scan slope geometry and, once repeat epochs exist, measure surface displacement directly. Two sites have been scanned to date. Monitoring Site 1 (39°25'16.1"N, 76°30'14.7"W) is a proof-of-concept practice site, scanned on two dates roughly two weeks apart, used to develop and validate the change-detection pipeline before it is applied to the project's real target.

Cumberland (39.64288°N, 78.79660°W) is the project's primary monitoring site, situated along the I-68 corridor in western Maryland; a working transportation corridor of direct relevance to the infrastructure motivation set out in Section 6.1; currently at a single epoch (day 1). The overall approach proposed for Cumberland is straightforward: M3C2 change detection is the method by which surface displacement will be measured once repeat epochs exist, and Monitoring Site 1 (Section 6.4.2) serves as a worked, end-to-end demonstration that the method works before it is relied upon at the primary site.

Because each scan session is recorded in its own arbitrary local coordinate frame, every multi-station or multi-epoch comparison requires registration before any geometric comparison is valid. Two registration regimes were encountered in practice. For Monitoring Site 1 (a single-station, steep-scarp scan compared across two epochs), stable-area Iterative Closest Point [132] successfully recovered a large (~49°) rotation between epochs, reducing the median point-to-point residual from 14.6 cm to 1.7 cm. Instead, plain ICP converged to a worse alignment for the Cumberland site (20 scan stations combined in two batches over flatter, more open terrain). This is a known failure mode where point-to-plane ICP can slide along a dominant, weakly-textured ground plane instead of locking onto a true correspondence. Instead, robust global registration was accomplished through feature-based alignment using Fast Point Feature Histogram descriptors [135] in conjunction with Random Sample Consensus [134] for coarse pose estimation, which was then refined using point-to-plane ICP [133]. Depending on batch, the median residual was reduced from roughly 11 m (raw) to 10–18 cm. This flat, open terrain requires feature-based global registration while steep, texture-rich terrain can succeed with local ICP alone is a practical, site-dependent finding of direct relevance to future scan planning.
Once two epochs of the same site exist, surface change is measured with Multiscale Model-to-Model Cloud Comparison (M3C2; [131]), which projects the distance between epochs along the local surface normal rather than along a fixed vertical or a single global direction, making it robust to the steep, non-planar geometry typical of natural slopes. At each core point p_i, a local normal N_i is estimated from a best-fit plane over a neighborhood of diameter D. Along N_i, a cylinder of diameter d collects points from each epoch k, and their mean projected position is:

$$\bar{x}_k = (1/n_k)\, \Sigma_j\, \langle (p_j - p_i), N_i \rangle \qquad (6.1)$$

The M3C2 distance is the signed difference between epochs' mean projected positions:

$$L\_M3C2 = \bar{x}_2 - \bar{x}_1 \qquad (6.2)$$

A distance is treated as statistically significant surface change only if it exceeds the 95% level of detection, which combines each epoch's local surface roughness (σ_1, σ_2) with the registration error (reg) estimated during alignment:

$$LOD_{95\%} = \pm 1.96 \cdot \sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)} + reg \qquad (6.3)$$

Figure 6.3 summarizes the full monitoring workflow: a single-epoch susceptibility prior is established from day-1 geometry alone, pending re-scan triggers M3C2 displacement measurement, and agreement (or disagreement) between the prior and observed displacement drives recalibration of the susceptibility weights (the validation loop detailed further in Section 6.5).

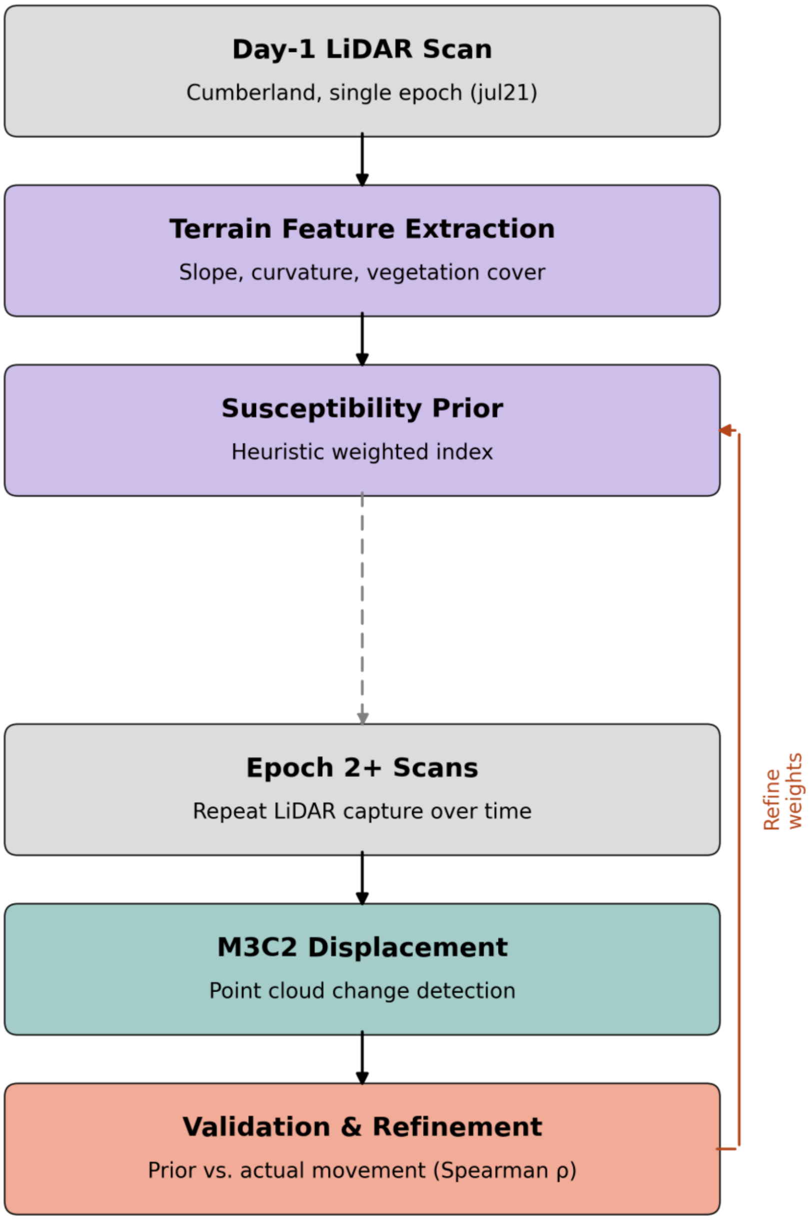


Figure 6.3. LiDAR monitoring methodology: susceptibility prior (day-1) through planned M3C2 validation and weight refinement.

In advance of a second epoch, day-1 Cumberland geometry alone was used to construct a static terrain-susceptibility index, following the same weighted-composite logic as the Thread 1 risk index but built from geometric rather than hydrological factors (Table 6.3): normalized slope angle, a curvature-derived concavity factor (concave, water-focusing terrain scored higher), and a color-derived bare-ground exposure fraction (vegetated ground scored lower, reflecting root-reinforcement effects). This is explicitly a prior, not a validated susceptibility model: with no displacement observations yet available, the weights are heuristic and are expected to be recalibrated once epoch-2 M3C2 results provide an actual displacement signal to correlate against (Section 6.5).

Table 6.3. Terrain-susceptibility index components and heuristic weights (Cumberland, day-1 prior).

Factor	Weight	Physical basis	Direction of risk
Slope angle	0.5	Primary control on shear-driven instability (Terzaghi, 1950)	Steeper → higher
Concavity (curvature)	0.3	Concave terrain concentrates infiltration and runoff	Concave → higher
Bare-ground exposure	0.2	Loss of root reinforcement on vegetated slopes	Bare → higher

6.4 Results

Results are reported separately for each thread, reflecting their different data maturity: Thread 1 reports validated predictive performance on the Zhuji benchmark dataset (Figure 6.4); Thread 2 first demonstrates the proposed M3C2 displacement-monitoring method end-to-end on the Monitoring Site 1 proof-of-concept dataset, then reports the current single-epoch baseline at Cumberland, the site the method is proposed for going forward.

6.4.1 Thread 1: Zhuji benchmark validation results

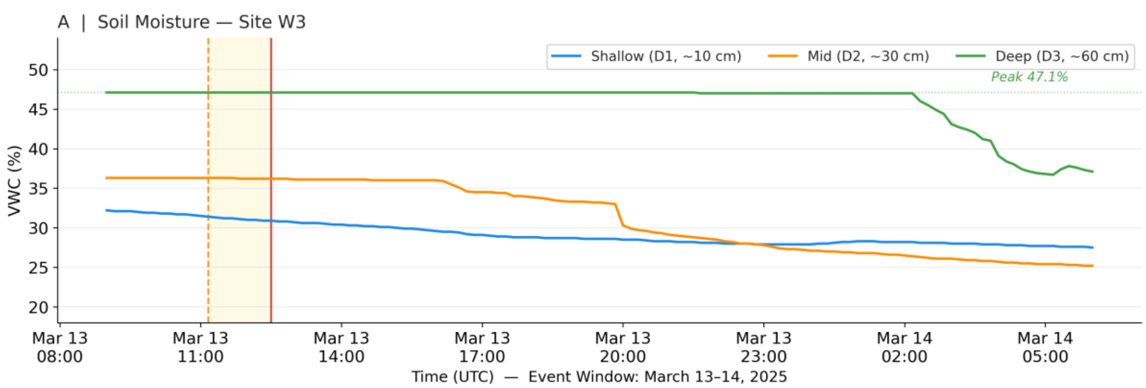


Figure 6.4. Site W3 soil moisture (shallow/mid/deep) of Zhuji landslide, event window.

Table 6.4 reports held out test-set performance for all four models. XGBoost substantially outperformed every deep sequence model on this dataset, achieving $R^2 = 0.70$, RMSE = 0.048, and MAE = 0.042, and issuing alerts a mean 1.33 hours before peak (Figure 6.5 through 6.8) instability with zero false alarms (ROC-AUC = 1.00; Figure 6.6). All three deep models produced negative R^2 on the held-out window (Table 6.4) despite converging normally during training (Figure 6.11), indicating the shortfall reflects the small size of this single-event benchmark dataset rather than a training failure and is expected to improve as the prototype's own longitudinal record accumulates.

Table 6.4. Held-out test-set performance, all four benchmarked models.

Model	R^2	RMSE	MAE	F1	ROC-AUC	Lead time
XGBoost	+0.699	0.0481	0.0419	0.77	1.00	1.33 h
LSTM	−2.737	0.1695	0.1454	—	—	1.3 h
BiLSTM	−3.651	0.1891	0.1631	—	—	1.3 h
TCN	−2.175	0.1563	0.1429	—	—	1.3 h

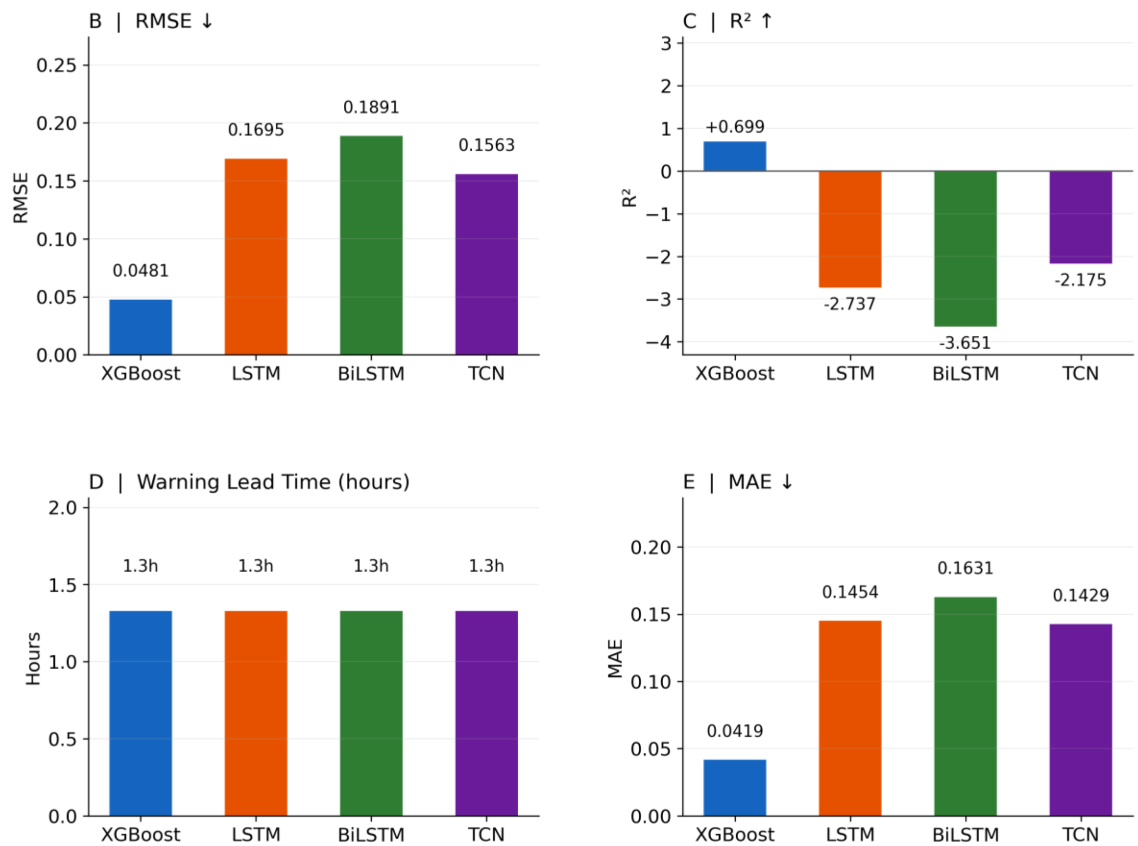


Figure 6.5. Model comparison: RMSE, R^2, warning lead time, and MAE across all four models.

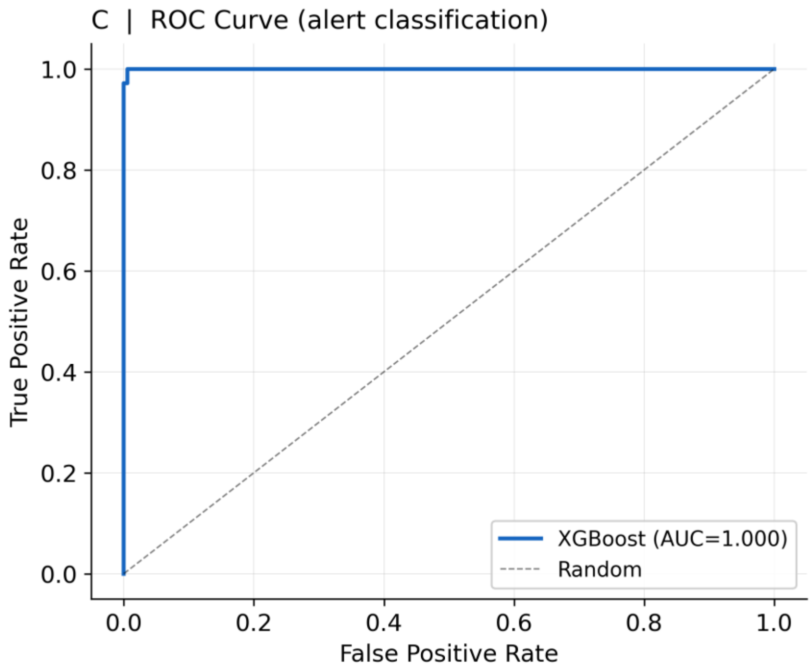


Figure 6.6. XGBoost ROC curve for alert classification (AUC = 1.00).

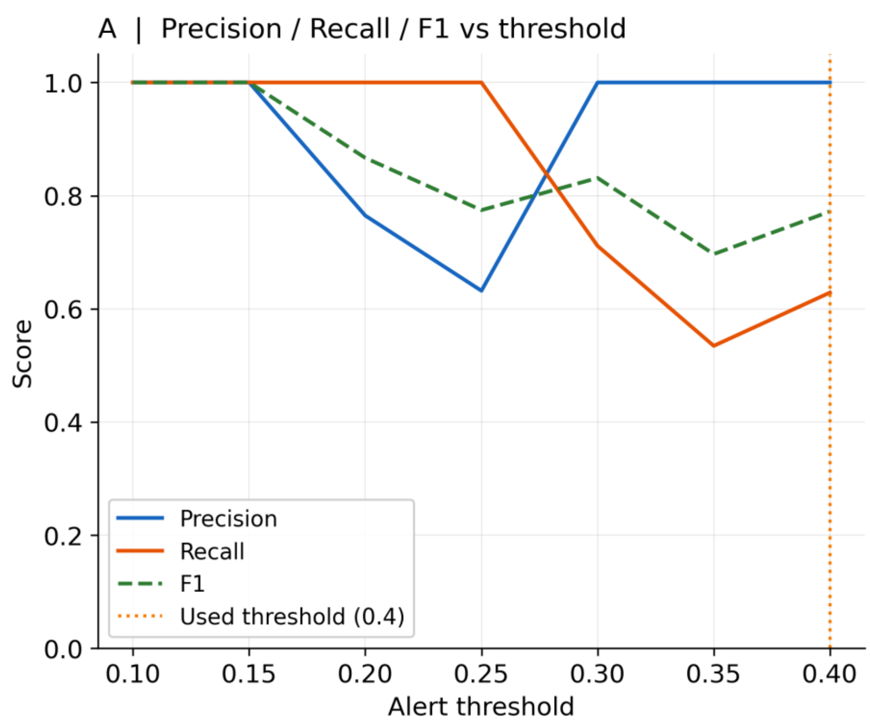


Figure 6.7. Precision/Recall/F1 vs. alert threshold.

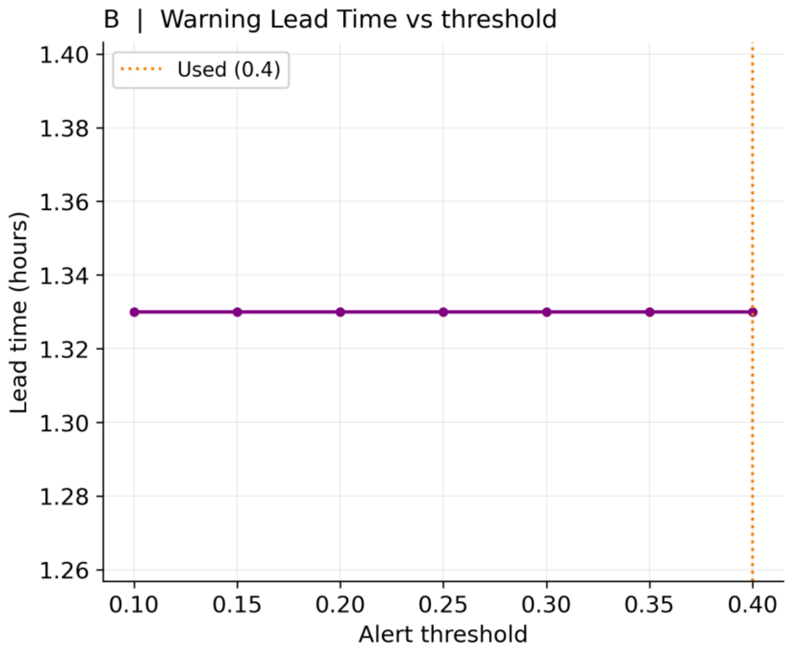


Figure 6.8. Warning lead time vs. alert threshold (stable at 1.3h across the full range tested).

Diagnostic note on warning lead time. Figure 6.13 shows the originally computed alert and peak-instability annotations, reporting a 1.33-hour lead time with zero false alarms across all four models. That uniformity across models of very different skill (R^2 ranging from +0.70 to −3.65 shown in Figure 6.5) prompted a direct check of the underlying alert timestamps. The check found that the benchmark event's true peak instability value falls inside the training window rather than the test window under the original temporal split; the value treated as "peak instability" for the lead-time calculation is therefore the first sample of the test window itself, at which point ground truth already exceeds the 0.40 alert threshold. Under a corrected split placing the true peak inside the test window, none of the four models reliably reach the alert threshold before that peak. We do not report a validated early-warning lead time from this benchmark as a result. Given a single rainfall event with a single peak, no split of this dataset can hold out a genuinely unseen peak for testing while leaving enough of the event's shape in training to learn from. Validating a lead-time claim will require a multi-event record, which the prototype's own longitudinal deployment is intended to provide (Section 6.5).

The ablation study (Table 6.5, Figure 6.9) identified mid-depth (D2, ~30 cm) sensors as the single most critical sensing tier: removing them reduced R^2 from 0.699 to 0.307, a larger drop than removing either the shallow (D1) or deep (D3+) tiers alone, and feature-importance analysis (Figure 6.10) confirms that the top two individual features are both derived from site W3's D2 and D3 sensors. This is consistent with mid-depth sensors capturing the wetting-front transition most directly, while shallow sensors saturate too quickly and deep sensors respond too slowly to be maximally informative on their own.

Table 6.5. Ablation study: R^2 by sensor-depth configuration.

Sensor configuration	R^2
Full model	+0.699
No deep (D3+)	+0.685
No shallow (D1)	+0.578
No mid (D2)	+0.307
D1 only	−0.087
D3+D4 only	−2.450

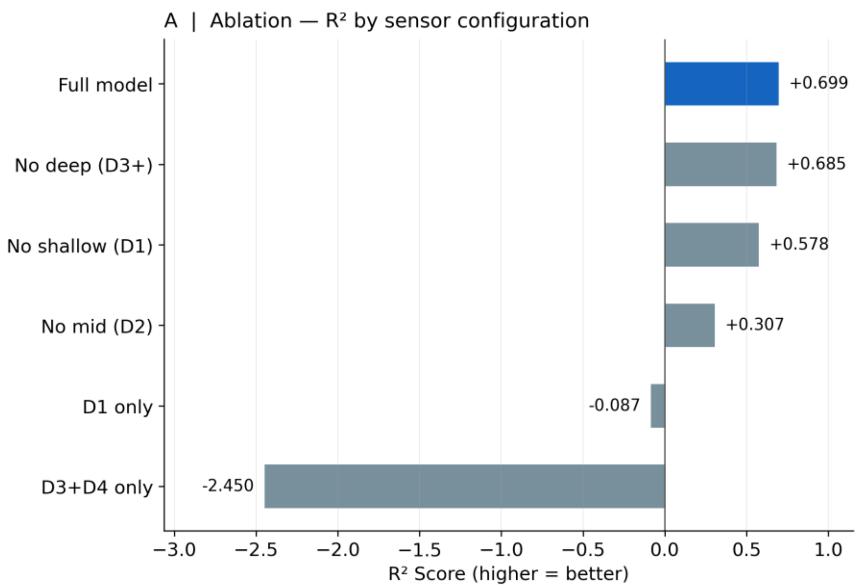


Figure 6.9. Ablation study: R^2 by sensor configuration, full model vs. depth-tier removals.

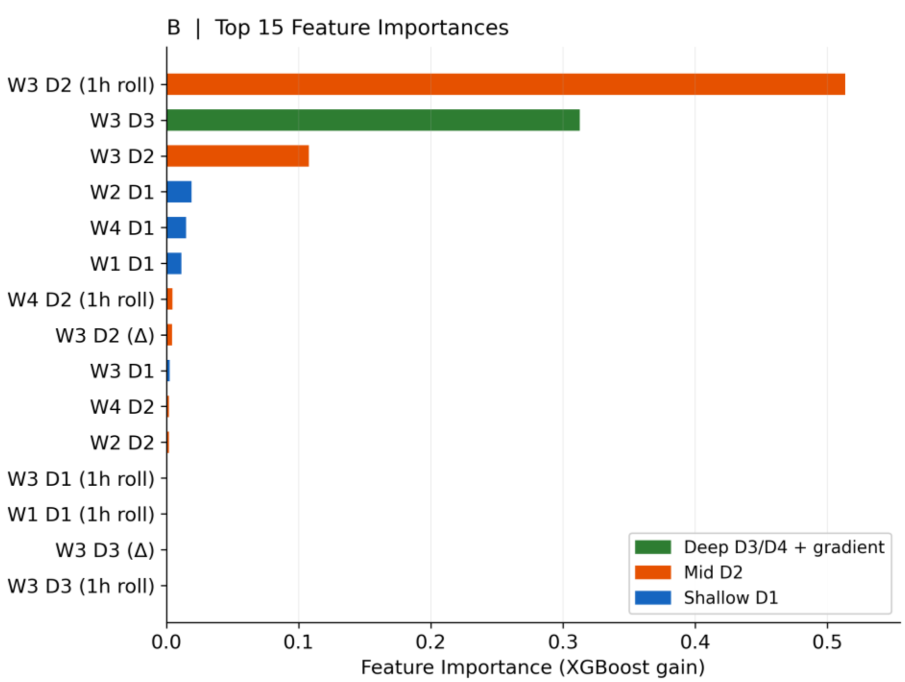


Figure 6.10. Top 15 XGBoost feature importances by depth tier.

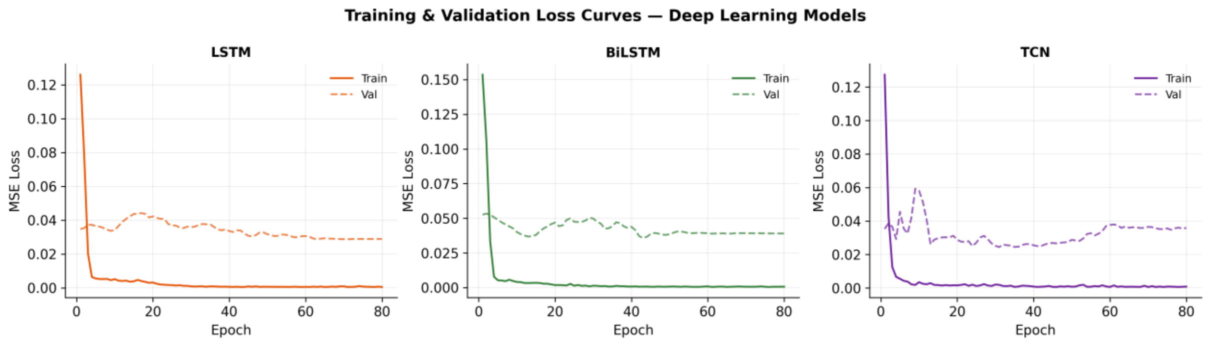


Figure 6.11. Training and validation loss curves, LSTM/BiLSTM/TCN; DL models converge normally despite lower R^2.

Figure 6.12 plots predicted vs. ground-truth risk index for all four models across the full test window, alongside the alert (0.4) and critical (0.7) thresholds. The deep models consistently lag the true peak and drift during the recovery period, whereas XGBoost tracks the shape of the ground-truth trajectory most closely. This is evident in the residual comparison (Figure 6.14), where LSTM's residual grows to −0.15 by the end of the window, while XGBoost's residual stays within an approximate ±0.05 band. Precision and recall both decline at higher thresholds as expected (Figure 6.7); however, the threshold-sensitivity curve (Figure 6.8) is now read differently than originally reported: lead time staying perfectly flat at 1.3 hours across the full range of thresholds tested (0.10–0.40) is itself diagnostic of the artifact described above, since a genuine threshold crossing should shift as the threshold moves. Its flatness, rather than supporting the original result, is part of the evidence against it.

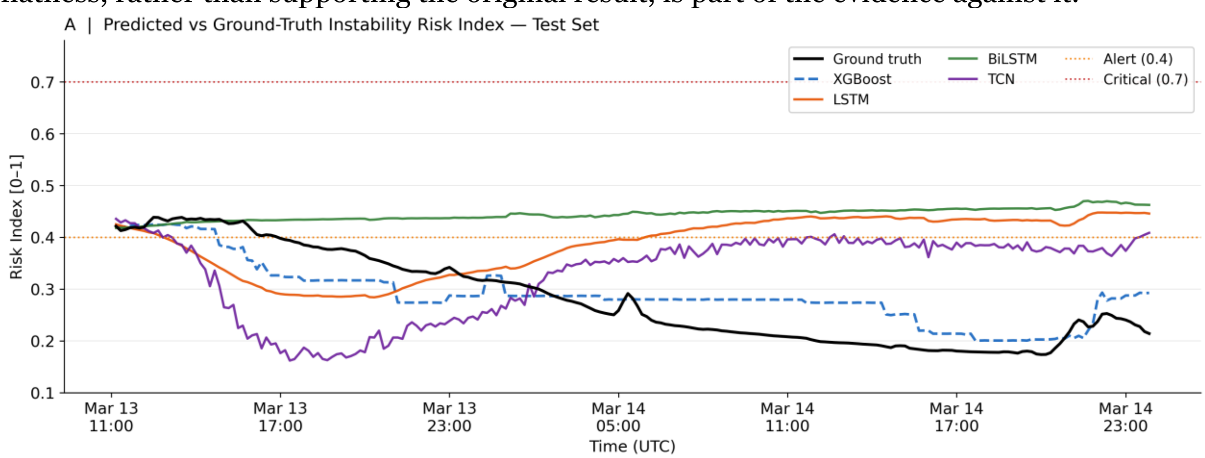


Figure 6.12. Predicted vs. ground-truth instability risk index, all four models, test set.

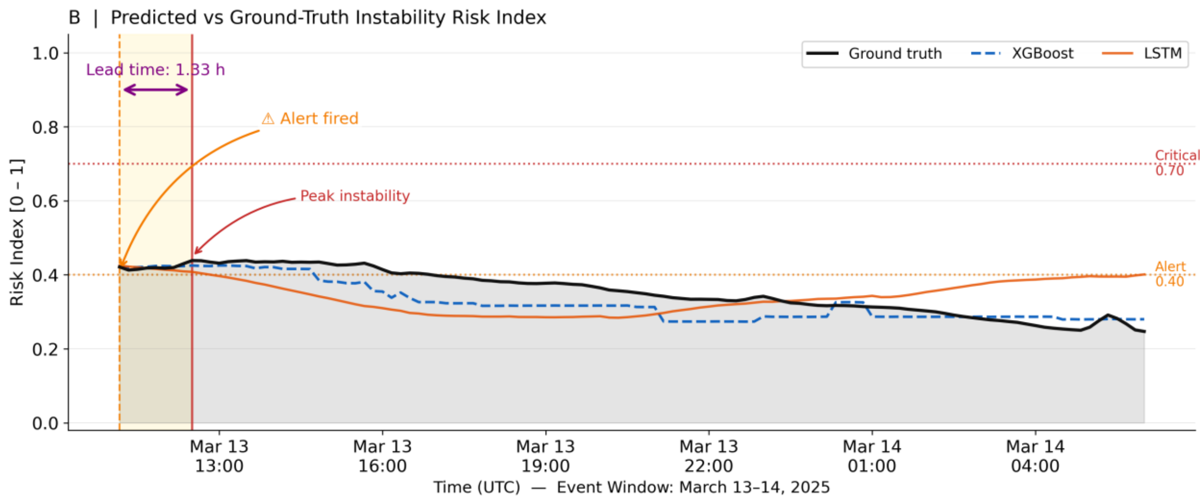


Figure 6.13. Annotated XGBoost/LSTM risk index with lead-time and alert callouts.

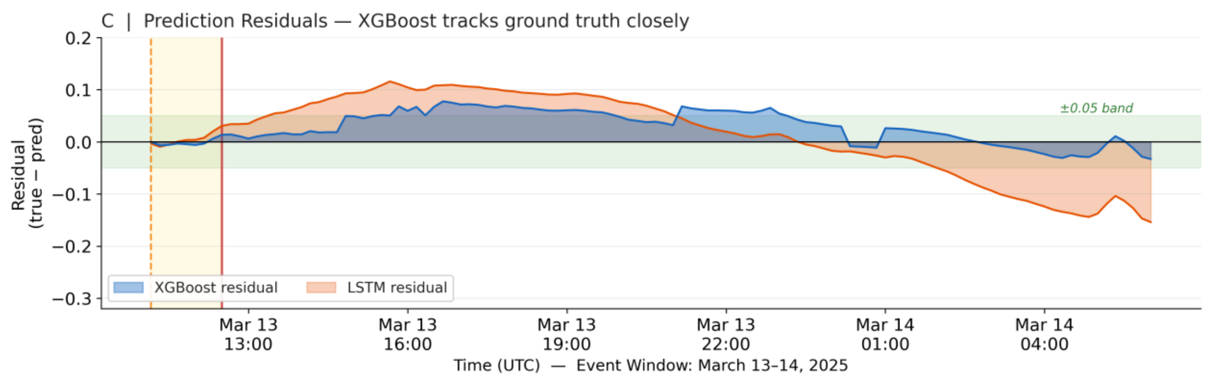


Figure 6.14. Prediction residuals, XGBoost vs. LSTM.

6.4.2 Thread 2 proof of concept: M3C2 displacement demonstration (Monitoring Site 1)

The suggested M3C2 displacement-monitoring approach (Section 6.3.4) was tested end-to-end on Monitoring Site 1 (39°25'16.1"N, 76°30'14.7"W), a proof-of-concept location scanned twice, on June 29 and July 13, roughly two weeks apart, prior to being applied to the project's main site. Before the pipeline was relied upon at Cumberland, this site is particularly selected to address the question, "Does the registration-through-M3C2 pipeline actually isolate a real, above-noise displacement signal, or only noise?"

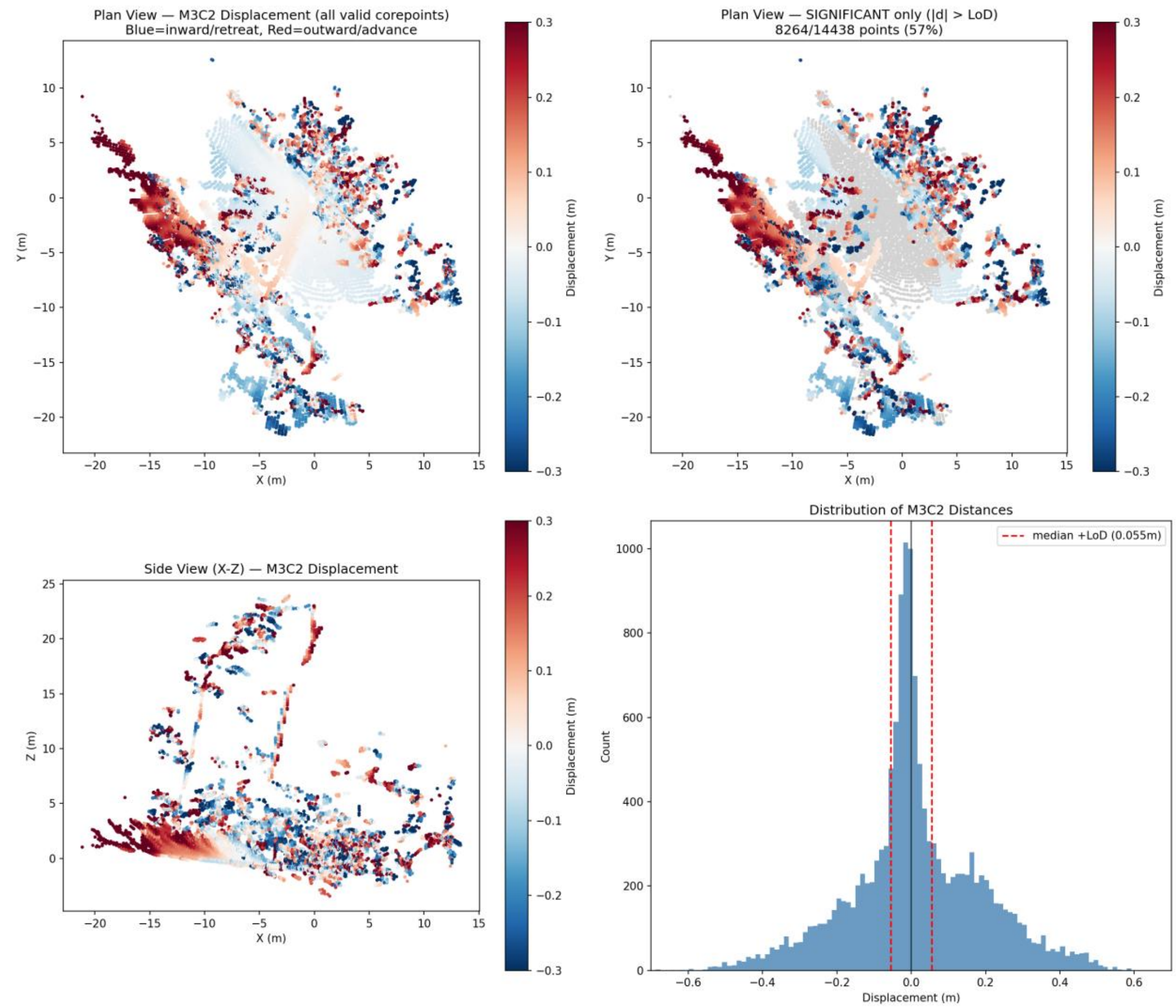


Figure 6.15. Monitoring Site 1 M3C2 displacement results showing plan view, side view, and significance mask (proof-of-concept demonstration).

Following registration (Section 6.3.4), M3C2 was computed over 16,497 core points at 0.2 m spacing. The overall displacement distribution was centered near zero (mean +1.2 cm, median −0.5 cm), consistent with no wholesale site-wide shift, but with substantial spread (σ = 17.6 cm) and 57% of valid points exceeding the level of detection. Cross-referencing displacement against a color-based vegetation classification showed vegetated points were over-represented among significant displacements (63.1% significant, σ = 20.0 cm) relative to bare-surface points (51.3% significant, σ = 14.7 cm), consistent with foliage movement and canopy change contributing noise unrelated to ground displacement. After restricting to bare-surface points and applying spatial clustering, one coherent, non-noise cluster remained: an approximately 9 m × 9 m patch near the toe of the scanned slope, averaging 19 cm of outward displacement across 1,076 core points, well above the local level of detection and confirmed by direct color inspection to be bare soil/rock rather than misclassified vegetation (Figure 6.15). This single two-epoch comparison cannot establish whether the zone is actively progressing, but it demonstrates the core claim this example was designed to test: the pipeline can isolate a real, spatially coherent, above-noise displacement signal from a much larger field of registration and vegetation-driven noise. This is the same pipeline, applied the same way, that will be run at Cumberland once repeated epochs exist.

6.4.3 Thread 2 deployment: Cumberland day-1 baseline

Cumberland (39.64288°N, 78.79660°W), situated along the I-68 corridor in western Maryland, is the site the method demonstrated in Section 6.4.2 is proposed for. Its first (day-1) epoch was collected as 20 individual scan stations, merged into two batches and registered as described in Section 6.3.4. Table 6.6 summarizes the resulting merged point cloud. Figure 6.16 shows the merged site in true color and colored by elevation; Figure 6.17 breaks the resulting susceptibility index down into its three underlying terrain factors; Figure 6.18 shows the combined terrain-susceptibility index with elevation contours overlaid.

Table 6.6. Cumberland day-1 LiDAR scan summary statistics.

Attribute	**Value**
Coordinates	39.64288°N, 78.79660°W (I-68 corridor, western Maryland)
Scan date	July 21, 2026 (day 1)
Scan stations merged	20 (two batches of 10)
Total points (merged)	259,951
Footprint	67 m × 56 m
Point density	69.9 pts/m²
Mean slope	27.2° (median 24.5°)
Area with slope > 30°	41.8%
Vegetation cover	31.1%
Ground coverage (of bounding box)	31.6%

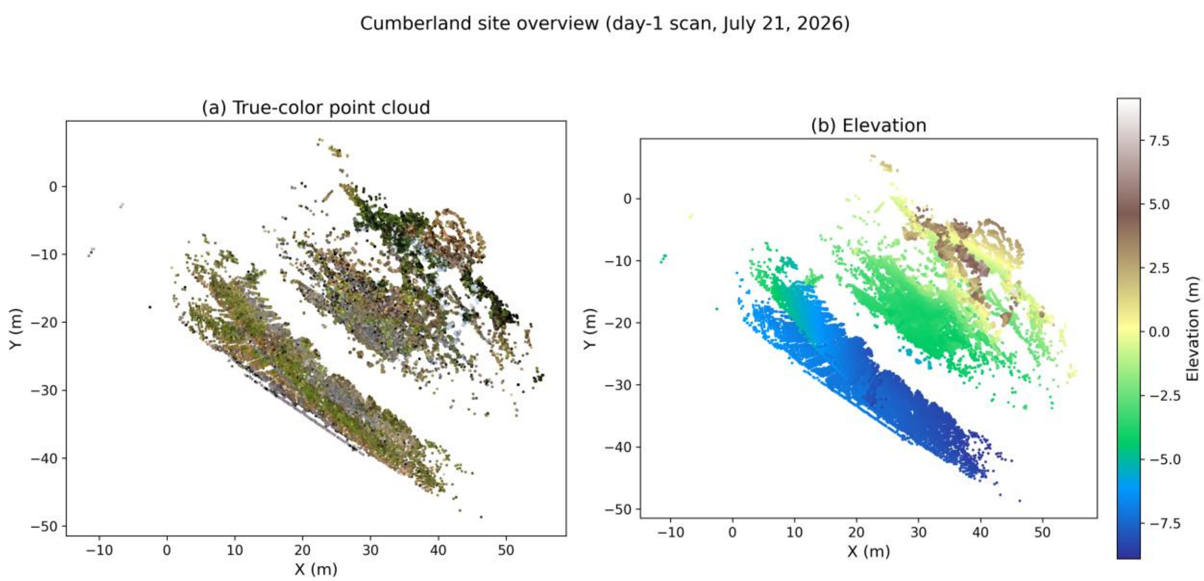


Figure 6.16. Cumberland site overview; true-color point cloud and elevation, day-1 scan (July 21, 2026).

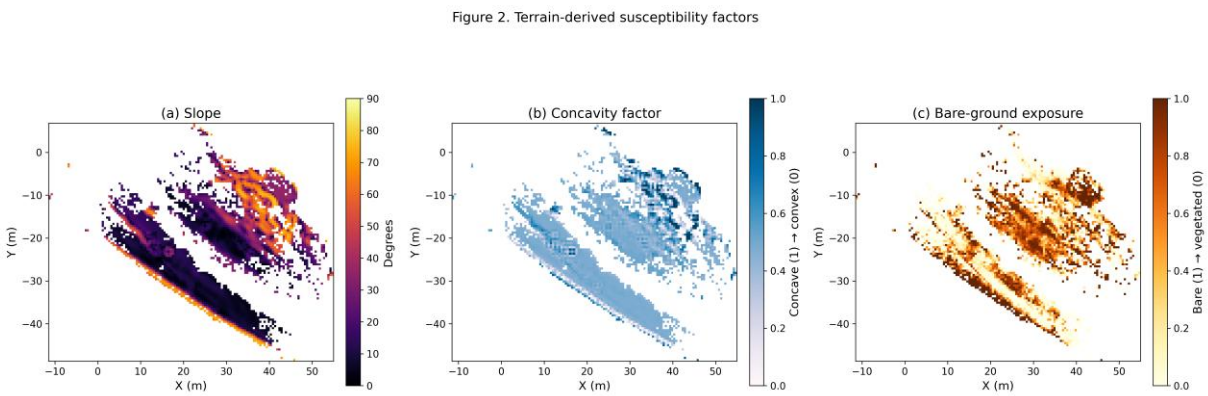


Figure 6.17. Cumberland terrain-susceptibility factor breakdown: slope, concavity, bare-ground exposure.

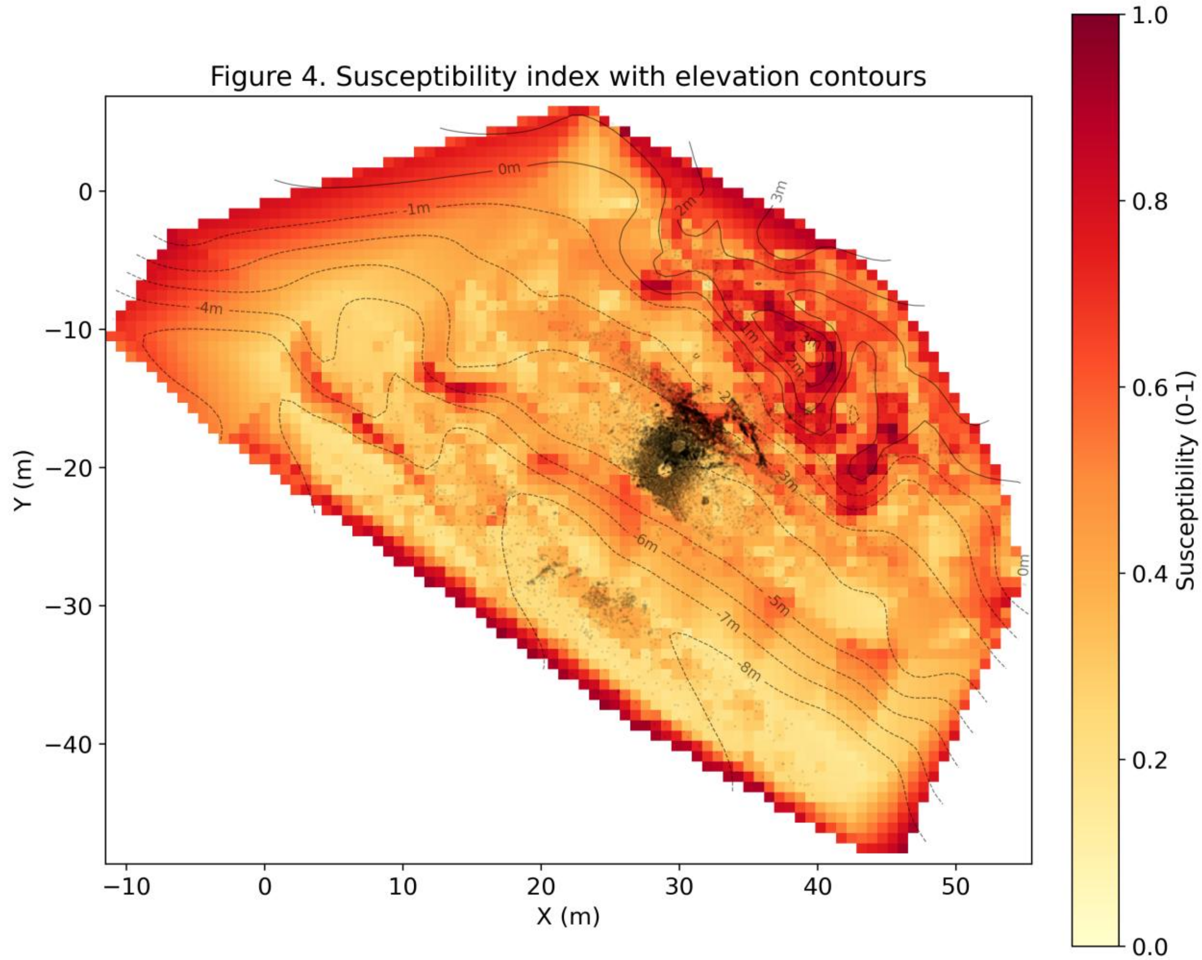


Figure 6.18. Cumberland terrain-susceptibility index with elevation contours (day-1 prior).

Figure 6.18 Cumberland's terrain is markedly gentler than the Monitoring Site 1 proof-of-concept (mean slope 27.2° vs. approximately 51°), reflecting a genuinely different site morphology. Point density is harder to compare directly: averaged over the full bounding box, Cumberland is sparser (69.9 pts/m² vs. Monitoring Site 1's approximately 172 pts/m²), but only 31.6% of that bounding box actually received a LiDAR return; density computed over covered ground alone is approximately 219 pts/m², denser than Monitoring Site 1. Which comparison is more meaningful depends on whether the uncovered area is treated as informative occlusion or simply as missing data. This coverage pattern is typical of multi-station outdoor capture and must be read alongside, not instead of, the susceptibility map itself: any interpretation of Figure 6.18 should be paired with this coverage limitation rather than read as full-site certainty.

The method that follows is taken from Section 6.4.2: M3C2 will be run identically as described above after two to three repeat epochs of Cumberland have been gathered. This will result in a brief displacement time series over the site rather than a single static snapshot. The same classical-versus-deep benchmarking approach previously validated for Thread 1 (Section 6.4.1) will then use those displacement values and the terrain-susceptibility factors already established here as the input for an ML/DL prediction step, resulting in an actual instability prediction for Cumberland instead of the geometry-only prior reported in this chapter.

Results are reported separately for each thread, reflecting their different data maturity: Thread 1 reports validated predictive performance on the Zhuji benchmark dataset; Thread 2 first demonstrates the proposed M3C2 displacement-monitoring method end-to-end on the Monitoring Site 1 proof-of-concept dataset, then reports the current single-epoch baseline at Cumberland, the site the method is proposed for going forward.

6.5 Conclusion and Future Work

Key result. Thread 1's verified contributions are the physics-informed instability risk index construction and the sensor-depth ablation finding (mid-depth sensors most critical), both validated on an independent benchmark dataset ahead of the prototype's own longitudinal record. A warning-lead-time claim was originally reported alongside these results; a diagnostic check (Section 6.4.1) found it to be an artifact of the train/test split on this single-event dataset rather than genuine predictive skill, and it is not reported as a validated result here. Validating a lead-time claim will require a multi-event record from the prototype's own ongoing deployment.

Current Status. Thread 1 has established and validated a physics-informed risk-labeling methodology on an independent benchmark dataset and has identified mid-depth soil-moisture sensing as the most critical single sensing tier via ablation; a previously reported warning-lead-time result did not survive a diagnostic check (Section 6.4.1) and is not carried forward. Thread 2 has established a working LiDAR data-collection and registration pipeline including a practical, site-dependent finding that flat, open terrain requires feature-based global registration where steep terrain does not and has produced a first-pass terrain susceptibility index for the project's primary monitoring site, Cumberland, from a single day-1 epoch.

What is explicitly not yet established. Neither thread has yet been validated against the prototype's own field data or against confirmed slope movement at Cumberland. The Thread 2 susceptibility index is a stated prior, not a validated model, and the two threads have not yet been fused into a single combined probability.

Immediate next steps. (1) Continue prototype sensor data collection and re-run the Thread 1 pipeline once a sufficient longitudinal record exists, re-evaluating whether deep sequence models close the performance gap with more training data, as their convergence behavior (Figure 6.11) suggests they should. (2) Collect two to three further Cumberland LiDAR epochs and run M3C2 exactly as demonstrated on Monitoring Site 1 (Section 6.4.2), producing a short displacement time series rather than the single day-1 snapshot reported here (Section 6.3.4, Figure 6.3). (3) Validate the susceptibility prior using a rank-based comparison. Spearman correlation between susceptibility score and displacement magnitude across all core points rather than a binary hit/miss test, since a rank-based statistic uses the full displacement field and is not sensitive to an arbitrarily chosen significance threshold. (4) Feed that displacement time series, together with the terrain-susceptibility factors (Table 6.3), into an ML/DL prediction step following the same classical-versus-deep benchmarking already validated for Thread 1 (Section 6.4.1) turning Thread 2 from a static geometric prior into an actual displacement-based instability prediction. (5) Use that outcome to recalibrate the susceptibility index weights, following the same prior-then-posterior logic already demonstrated on Monitoring Site 1. (6) Scope and implement the PointNet++ deep-learning counterpart to M3C2, operating directly on point-cloud geometry rather than the gridded terrain-derivative table used for the current susceptibility index. (7) Once both threads have field data from co-located or comparable sites, implement the fusion-layer weighting scheme shown as future work in Figure 6.1.

7 Development of a National Railroad Landslide Exposure Inventory Using GIS-Based Proximity Screening

Teka Farquharson, Zhuping Sheng, Yi Liu, and Oludare Owolabi

7.1 Introduction

7.1.1 Railroad Infrastructure and Landslide Hazards

Railroads carry freight and passengers across a country with widely varying topography, geology, and climate conditions, which makes the network a critical but geographically exposed piece of United States transportation [143]. In many regions, rail lines traverse through mountain slopes, incised valleys, coastal bluffs, and other terrain where landslides and related slope failures directly affect track stability and day-to-day operations [144]. These failures directly affect track structure, embankments, drainage features, and the earthworks around them, adding to both routine maintenance load and emergency repair demand [145].

The consequences reach well beyond physical damage. Landslides near rail corridors can cause operational delays, disrupt the broader network, and, in the worst cases, lead to derailment or put crews and passengers at risk [144]. Studies focused specifically on rail hazards have repeatedly shown that landslides remain a persistent concern for infrastructure managers, since even a single localized failure can shut down service across a much longer stretch of corridor and demand rapid inspection, stabilization work, and traffic control responses [144]. This is especially true where prolonged rainfall, steep terrain, weak subsurface materials, or already unstable slopes raise the odds of slope movement near active track.

Given these stakes, pinpointing where historical landslide activity has occurred near rail corridors is an important first step in network-scale hazard evaluation [146]. Landslide inventories, which record past slope failures, are already used widely to support susceptibility mapping, hazard screening, and infrastructure planning, once they are layered with terrain, geology, and transportation data in a GIS environment [147]. Paired with railroad spatial data, these inventories make it possible to see exactly where documented landslides and rail corridors sit close together, and to flag the segments that deserve a closer look [143].

For railroad operators and public agencies managing large, geographically dispersed networks, a standardized geospatial screening method offers a practical way to gauge exposure at a national scale [143]. This kind of screening does not replace detailed geotechnical investigation or site-specific engineering analysis. It is meant to provide a narrow but still useful assessment by giving agencies an efficient first pass for identifying corridors where historical landslide occurrence suggests a need for further field assessment, monitoring, or more detailed hazard evaluation.

7.1.2 Need for National Railroad Specific Screening

Landslide and soil hazards have increasingly threatened the reliability and safety of the railroad infrastructure in recent years, particularly along rail corridors that pass through the western and mountainous regions of the country. A derailment triggered by a landslide can lead to serious consequences such as service disruption, costly repairs, and in the worst case, loss of life [148, 149]. Building a centralized, detailed record of these risk-prone areas gives rail owners a clearer picture of where vulnerabilities actually exist, which in turn supports more proactive risk reduction and better-informed planning, response and protection decisions before disaster strikes. Detailed landslide occurrence data underpins both process understanding and susceptibility mapping, both of which are foundational to any risk reduction strategy (Mirus et al., 2020). Despite the risks, no consistent geospatial framework currently exists to identify, track, and prioritize landslide threats along rail lines in a systematic way. Many existing landslide databases are either incomplete, inconsistent, or simply not detailed enough to support rail infrastructure planning and risk management on their own [150, 151].

7.1.3 Research Gap

Landslide inventories document the spatial distribution, characteristics, and historical occurrence of slope failures, and at the national scale they form an important foundation for hazard assessment and susceptibility mapping (Mirus et al., 2020). In parallel, national railroad geospatial datasets describe the location and extent of rail infrastructure across the United States [152]. Both data sources exist and are publicly accessible, but they were built for different purposes and are rarely combined to evaluate how exposed railroad infrastructure actually is to documented landslide occurrences.

Most prior landslide research has centered on regional susceptibility mapping, site-specific slope stability analysis, landslide prediction, or identifying the environmental controls behind slope failure [148, 153]. Other work has looked at landslide impacts on transportation infrastructure, but typically at local or regional scales [154]. This body of work has meaningfully advanced how landslide processes are understood and how individual high-risk sites are managed. But extending detailed geotechnical or predictive analysis across an entire national railroad network would demand a level of time, data, and computational resources that makes it impractical as a preliminary screening approach.

There is a related limitation: historical landslide inventories don’t tell you much about infrastructure exposure unless their spatial relationship to rail corridors is explicitly evaluated [153, 154]. A documented landslide somewhere in a state or region doesn’t, by itself, mean railroad infrastructure there is exposed. Comparing states by raw landslide counts alone doesn’t produce a railroad-specific exposure measure either. What is needed is a spatially explicit framework that integrates historical landslide locations with the national railroad network, quantifies their proximity to rail infrastructure, and surfaces geographic concentrations of landslide-rail interactions.

This study fills that gap with a reproducible GIS-based framework for national-scale screening of historical landslide exposure along U.S. railroad corridors. The approach combines mapped landslide occurrences with railroad infrastructure data and applies a standardized 328-ft (100-m) screening zone to flag landslides located near rail corridors. Proximity analysis then characterizes the spatial exposure and identifies states, regions, and railroad segments where landslide occurrences concentrate. The intent is not to estimate landslide probability or infrastructure risk, but to give agencies a preliminary screening tool that prioritizes locations for more detailed environmental, geotechnical, and engineering investigation.

That distinction also sets up the study objectives. This work does not attempt predict landslide; its contribution is a national-scale method for locating where historical landslides and railroad infrastructure intersect spatially, and where more detailed investigation should be concentrated. Keeping that boundary clear separates this technical report from corridor-scale erosion and landslide analyses, while still supporting infrastructure screening and prioritization at the network level.

7.1.4 Objectives

The research objectives for this study include:

1. Integrate the USGS National Landslide Inventory with the FRA railroad network using GIS.
2. Identify documented landslides within a 328-ft (100-m) buffer of railroad infrastructure.
3. Assess the state and regional distribution of landslide occurrences near rail corridors.
4. Develop a reproducible screening framework to support inspection prioritization and future corridor-level investigations.

7.2 Background

7.2.1 National Landslide Inventories

Landslides are widespread geologic hazards involving the downslope movement of soil, rock, and debris triggered by intense rainfall or rapid snowmelt. They take several forms, including debris flows, rockfalls, and earth slides, and are commonly set off by weathering, freeze-thaw cycles, or seismic shaking, any of which can dislodge rock and send it down steep slopes with little warning to roads and structures below [155]. These processes can threaten communities and transportation infrastructure alike, especially in mountainous and steep terrain where slope failures cross paths with roads and railroads.

Landslide inventories provide the foundation for documenting where, how often, and in what form historical slope failures have occurred. In the United States, landslide information has historically come from federal and state agencies, local governments, academic institutions, and other organizations, each using its own mapping and reporting approach. The U.S. Geological Survey has worked to consolidate these efforts into a national scale landslide inventory by compiling publicly available records from multiple sources into a centralized geospatial database. The national inventory is a valuable resource for examining where documented landslide activity concentrates geographically (Mirus et al., 2020), and it supports both the identification of historical landslide patterns and the broader work on susceptibility, hazard, and infrastructure exposure assessments.

Even so, national landslide inventories have real limitations. How complete and spatially detailed the records are varies greatly among regions, driven by differences in mapping programs, reporting practices, available imagery, field investigations, and historical documentation. As a result, areas with a long history of intense landslide mapping may simply show more documented occurrences than areas where inventories are thinner, not necessarily because landslides are less common there. That distinction matters when interpreting national scale patterns or comparing landslide occurrence counts across states or regions [148].

For this study, the national landslide inventory serves as the primary historical dataset used to examine how documented landslide occurrences relate spatially to U.S. railroad infrastructure. Pairing the inventory with the national railroad network provides a consistent GIS-based method for identifying documented landslides that fall within the defined 328-ft (100-m) railroad screening zone. The resulting analysis reflects historical landslide exposure based on documented occurrences and spatial proximity; it is not a forecast of where landslides will occur next.

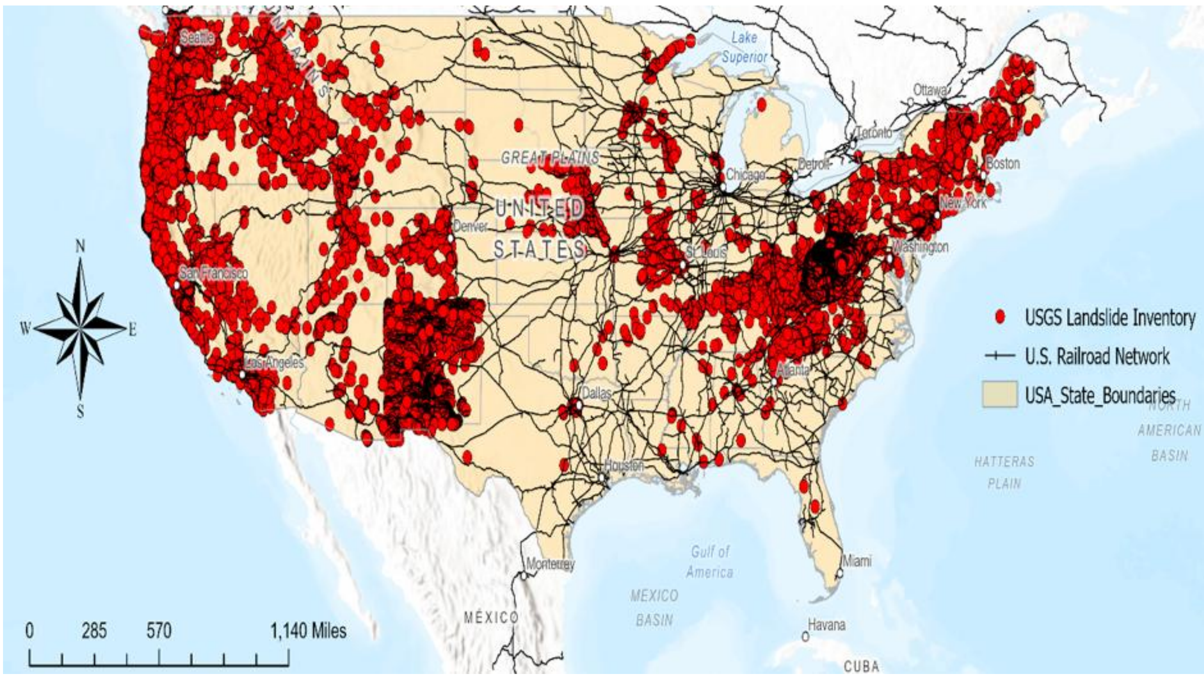


Figure 7.1. Geographic distribution of documented landslide occurrences contained in the U.S. national landslide inventory

7.2.2 Landslide Impacts on Railroad Infrastructure

Railroad infrastructure is one of the most important freight transportation systems in the country, moving agricultural products, energy resources, manufactured goods, and intermodal freight across more than 140,000 miles of rail network [156]. Unlike highways, rail operations are locked into fixed corridors that often cut through mountainous terrain, river valleys, coastal bluffs, and steep embankments, exactly the settings where slope instability tends to show up. Because of that fixed alignment, landslides rank among the most significant geologic hazards facing railroad infrastructure because even localized slope failures can interrupt transportation across a much longer stretch of the network.

Landslides affect railroad systems through several distinct mechanisms depending on the type, size, and location of the slope failure. Debris flows can transport soil, rock, and vegetation onto railroad tracks quickly, blocking rail operations and damaging drainage systems (Figure 7.2). Rockfalls can strike tracks, bridges, tunnels, or even passing trains with little advance warning, particularly along steep rock cuts. Earth slides and rotational slumps move more slowly but can still seriously deform embankments and undermine track foundations [149, 155]. Any of these failure types can trigger immediate inspection, speed restriction, emergency stabilization work, or temporary service suspension until conditions are confirmed safe.

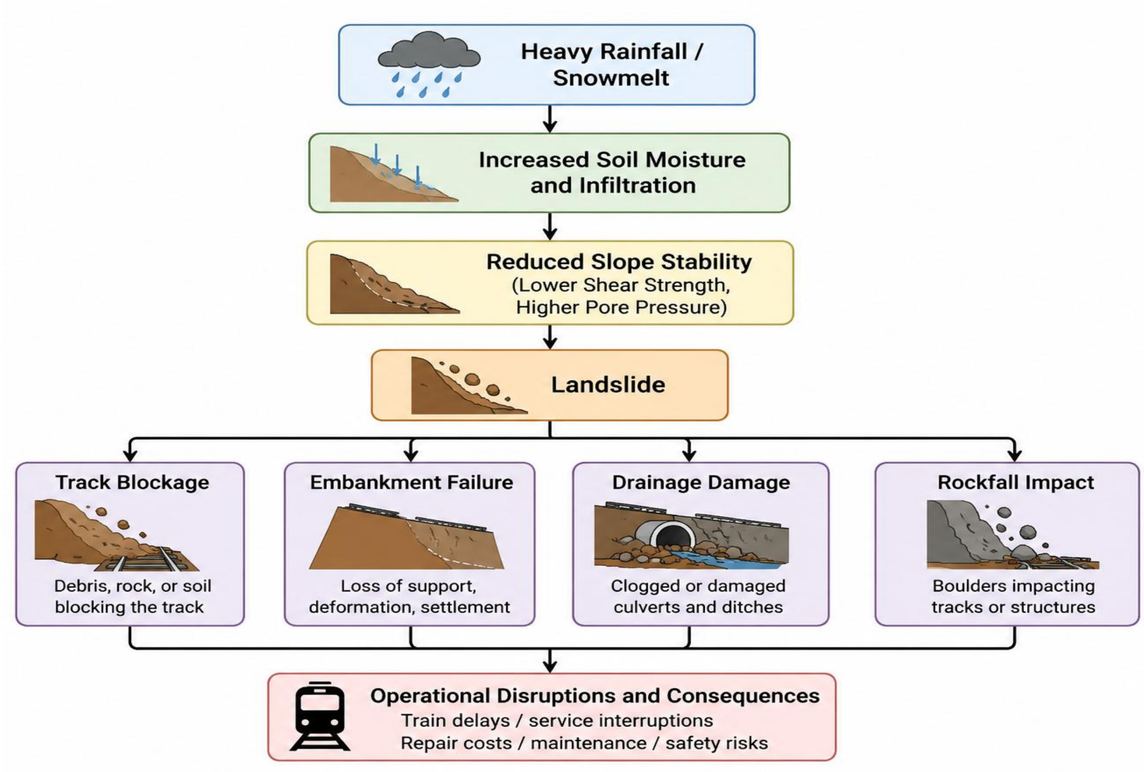


Figure 7.2. Conceptual pathways through which landslides affect railroad infrastructure

Hydrologic conditions play a critical role in triggering many railroad related landslides. Prolonged rainfall, short but intense storms, rapid snowmelt, and inadequate drainage all increase soil moisture, reduce effective shear strength, and elevate pore-water pressure in slopes adjacent to railroad corridors [148, 157]. Recent atmospheric river events on the West Coast have shown how extreme precipitation can set off widespread shallow landslides, debris flows, and embankment failures that disrupt transportation infrastructure over large geographic areas [158]. Given climate projections pointing toward more frequent intense precipitation events, understanding historical landslide exposure along railroad corridors becomes increasingly relevant going forward.

The consequences of these events extend well beyond the physical damage itself. Track damage from slope failures interrupts freight movement, delay passenger services, drive up maintenance costs, and ripple through regional supply chains. Emergency repairs typically involve debris removal, slope stabilization, drainage improvements, and rebuilding damaged embankments, all of which require substantial funding and downtime [159]. Because many critical freight corridors have limited alternative routings, even a relatively small landslide can carry on outsized economic and logistical costs.

Managing landslide hazards effectively requires transportation agencies to identify locations where historical slope failures have occurred and prioritize corridors for closer investigation accordingly. Site-specific geotechnical investigations give the most detailed picture of slope stability, but running them across an entire national railroad network isn't practical or economically feasible. National-scale geospatial screening offers a more efficient starting point by combining historical landslide inventories with railroad infrastructure data to identify locations where documented landslides occur near rail corridors. These preliminary assessments help agencies direct field investigations, monitoring programs, and engineering studies toward the areas with the greatest concentrations of documented landslide activity while supporting long-term infrastructure resilience planning.

7.2.3 GIS-Based Proximity and Buffer Analyses in Hazard Mapping

Geographic Information Systems (GIS) have become a core tool for hazard assessment because researchers are able to integrate, visualize and analyze large geospatial datasets within a single spatial framework. In landslide research, GIS has been applied widely to inventory development, susceptibility mapping, hazard assessment, and infrastructure planning by bringing terrain, environmental, and infrastructure datasets together to evaluate how slope failures are spatially distributed [147,160]. That ability to combine multiple datasets lets researchers and transportation agencies pinpoint exactly where documented landslide occurrences intersect, or sit close to, critical infrastructure.

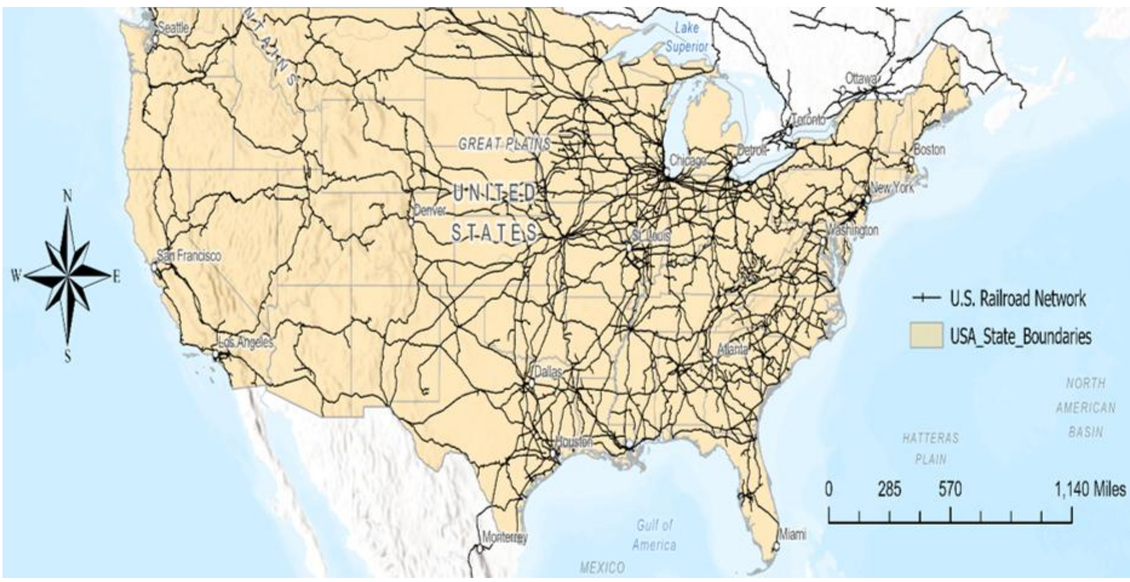


Figure 7.3. National railroad network used in the GIS-based landslide exposure analysis.

One of the most widely used GIS techniques for infrastructure assessment is proximity analysis, which evaluates the spatial relationship between hazards and exposed assets by measuring the distance between geographic features [159]. It has been applied extensively to transportation networks to identify roads, railroads, pipelines, and utility corridors located near natural hazards, offering an efficient way to screen large infrastructure systems for potential exposure [159]. Rather than predicting future landslide occurrence, proximity analysis identifies locations where documented historical landslides occur within a specified distance of infrastructure, supporting the prioritization of locations for subsequent investigation.

Buffer analysis extends this idea by creating zones around geographic features to define the spatial extent of an assessment. Buffers are commonly applied to linear infrastructure such as highways, railroads, and pipelines to identify hazards within a predetermined distance of the corridor [159]. The buffer distance itself functions as a screening threshold, one that allows consistent evaluation of infrastructure exposure across large geographic areas. By narrowing extensive geospatial datasets down to manageable subsets, buffer analysis supports regional and national infrastructure assessments while giving analysts a standardized way to compare exposure across different geographic regions.

GIS-based proximity and buffer analyses are efficient, but they do not directly measure landslide susceptibility, hazard probability, or infrastructure risk. Susceptibility assessments bring in terrain, geology, hydrology, and other conditioning factors to estimate the likelihood of a future slope failure, whereas proximity analysis is limited to evaluating the spatial relationship between documented historical landslides and existing infrastructure [147]. On top of that, the completeness of historical landslide inventories varies from region to region because of differences in mapping practices, reporting, and data availability [148]. Given both constraints, GIS-based proximity analysis is best understood as a preliminary exposure screening tool, one that identifies locations warranting more detailed environmental, geotechnical, and engineering investigation, not a quantitative measure of landslide risk on its own.

8.2.4 Limitations of Proximity-based screening

Geographic Information System (GIS) based proximity analysis is an efficient, reproducible way to identify documented landslide occurrences located near railroad infrastructure across large geographic areas [147]. By integrating historical landslide inventories with transportation networks, this kind of screening enables rapid evaluation of infrastructure exposure and helps flag corridors that may warrant a closer look. That makes it particularly useful for national-scale assessments, where detailed geotechnical investigations simply are not practical given the size and complexity of the railroad network. As a result, proximity analysis has become an effective preliminary screening tool for infrastructure planning and hazard management [147, 159].

That said, proximity-based screening does not evaluate landslide susceptibility, hazard, or quantitative risk. A documented landslide near a railroad corridor tells you about a historical spatial relationship [161]. It doesn't mean the adjacent slope is still unstable, and it doesn't mean a future failure is likely. Susceptibility assessments draw on terrain, geology, soil properties, hydrologic conditions, land cover, and climate to gauge slope stability; hazard assessments estimate the probability and magnitude of future events; and risk assessments extend this analysis further by considering the consequences of slope failure on infrastructure, operations, and public safety [147, 161, 162]. Therefore, the results of this study should be interpreted as measures of historical landslide exposure rather than predictions of future landslide occurrence or infrastructure failure.

A second limitation concerns the quality and completeness of historical landslide inventories. National landslide inventories are compiled from many federal, state, and local sources, each using different mapping techniques, reporting standards, and periods of record. That means the density of documented occurrences across the country reflects both actual landslide activity and how thoroughly different regions have been mapped [147]. The 328-ft (100-m) screening buffer used here gives national comparisons a consistency baseline, but it doesn't account for differences in landslide size, runout distance, topographic setting, or failure mechanism. Some landslides sitting just outside the buffer still affect railroad infrastructure, while some within the buffer may have little operational bearing at all.

Even with these warnings, proximity-based screening holds up well as a first-level assessment for a network this large. It gives agencies a consistent, reproducible framework for combining historical landslide inventories with railroad infrastructure, letting them identify where documented landslide exposure concentrates and prioritize locations for field investigation. It is not a substitute for detailed environmental, geotechnical, or engineering analyses; it is a decision-support layer that helps direct resources and guide infrastructure monitoring. Paired with site-specific investigations and additional environmental data, this kind of screening feeds into better corridor management, sharper maintenance planning, and stronger railroad resilience against landslide hazards.

7.3 Materials and Methods

7.3.1 Data Sources

This national railroad landslide exposure analysis relied on publicly available geospatial datasets covering documented landslides, railroad infrastructure, and administrative boundaries across the contiguous United States. The primary datasets were the U.S. Geological Survey (USGS) National Landslide Inventory and the Federal Railroad Administration (FRA) National Railroad Network from the National Transportation Atlas Database (NTAD). U.S. Census Bureau TIGER/Line state boundaries were used for state-level aggregation and regional analyses. Before analysis, every dataset was checked for completeness, spatial consistency, and coordinate reference system compatibility to ensure accurate spatial integration.

Table 7.1: Geospatial datasets used for national railroad landslide exposure screening.

Dataset	Source	Purpose in Study
National Landslide Inventory	U.S. Geological Survey (USGS)	Documented historical landslide locations and associated attributes
National Railroad Network	Federal Railroad Administration (FRA), National Transportation Atlas Database (NTAD)	Railroad centerlines used to generate the 328-ft (100-m) screening buffer and perform proximity analysis
U.S. State Boundaries	U.S. Census Bureau (TIGER/Line Shapefiles)	State-level aggregation and regional analysis
Basemap	Esri ArcGIS Online	Reference map for visualization only; not used in spatial analysis

7.3.2 Data Preparation

This study integrates the USGS National Landslide Inventory with the FRA National Railroad Network to build a standardized GIS-based framework for evaluating historical landslide exposure along U.S. railroad corridors. A 328-ft (100-m) bilateral screening buffer, combined with GIS proximity analysis, was used to identify documented landslides located adjacent to railroad infrastructure, giving a consistent basis for flagging corridors that need further investigation

The workflow ran in ArcGIS Pro following standard GIS data preparation procedures. Every dataset was checked for completeness, spatial consistency, and coordinate reference system compatibility prior to analysis. A 328-ft (100-m) bilateral buffer was generated around railroad centerlines to define the screening area and spatial overlay and proximity analysis identified documented landslides falling within that buffer, producing a national landslide–rail exposure inventory.

The resulting inventory was summarized at national, state, and regional scales to characterize where documented landslides occur near railroad infrastructure. Maps, summary statistics, and regional breakdowns supported a first-level screening process aimed at prioritizing future environmental, geotechnical, and engineering investigations. The workflow itself is transparent and reproducible, and it can be updated readily as new landslide inventories and railroad datasets become available.

Figure 7.4. GIS workflow used to identify and assess documented landslide exposure along U.S. railroad infrastructure. The workflow integrates the USGS National Landslide Inventory and the FRA railroad network to generate a 328-ft (100-m) railroad buffer, identify documented landslides through spatial intersection, summarize exposure by state and railroad ownership, and produce national maps, statistical summaries, and visualization products.

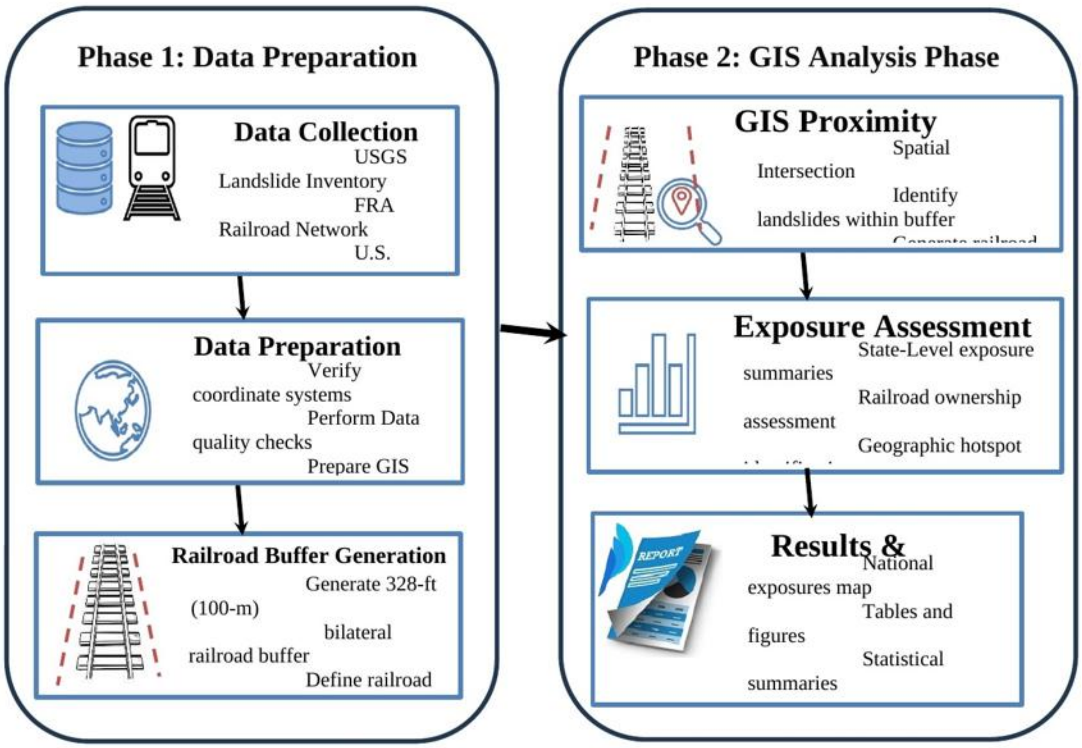


Figure 7.4. GIS workflow for evaluating historical landslide exposure along U.S. railroad infrastructure.

7.4 Results

7.4.1 National Landslide Railroad Distribution Near Railroad Infrastructure

The national GIS screening identified 1,455 documented landslide occurrences within 328 ft (100 m) of railroad infrastructure across the United States (Figure 7.5). The mapped distribution was geographically uneven, with landslide occurrences clustered along selected railroad regions rather than spreading evenly across the national network. The concentration sits primarily in the Pacific Coast and Appalachian regions, while much of the central and southern United States shows relatively few documented occurrences within the railroad screening zone.

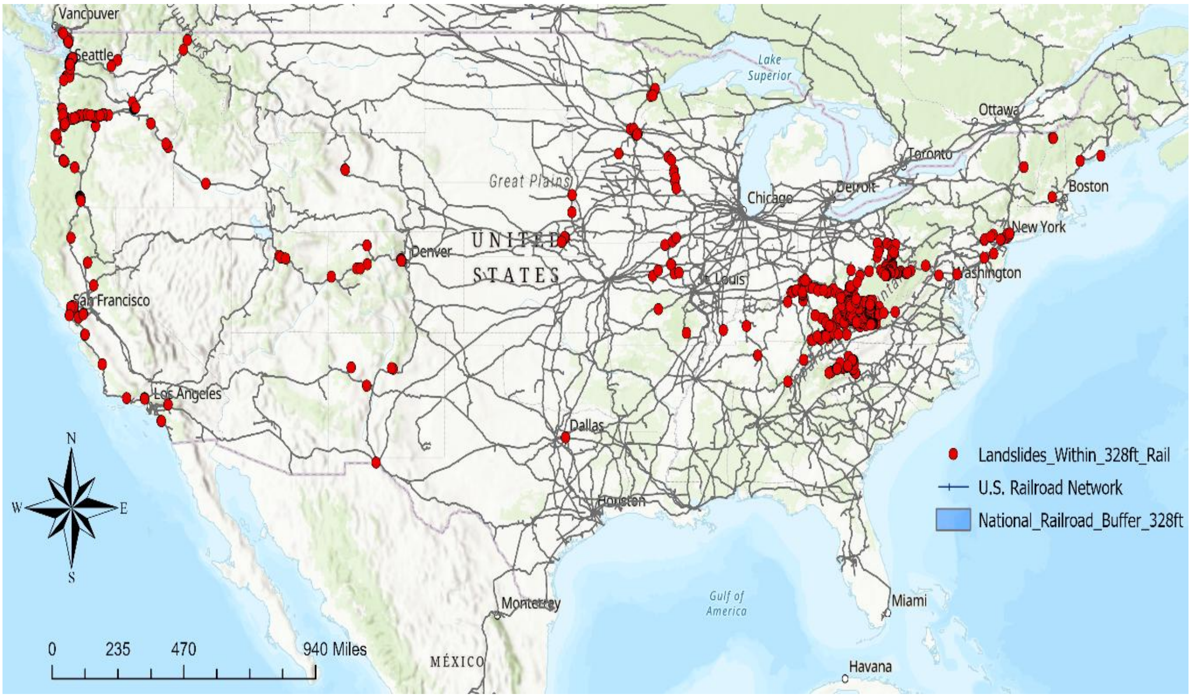


Figure 7.5. National distribution of documented landslides located within 328 ft (100 m) of railroad infrastructure.

The mapped distribution shows several prominent concentrations of documented landslides along railroad corridors. Washington, Oregon, California, West Virginia, Kentucky, and North Carolina hold the largest counts of landslides. In the western United States, landslide occurrences were especially evident along railroad corridors crossing mountainous terrain and coastal regions. In the eastern United States, substantial clustering runs throughout the Appalachian region, particularly in West Virginia and Kentucky. Together, these patterns point to documented landslide exposure near railroad infrastructure being concentrated in areas defined by steep terrain, complex topography, and a long history of landslide activity.

State-level aggregation makes concentration areas even clearer. Washington leads with 340 documented landslides within the screening zone, about 23.4% of the national total. West Virginia ranked second with 281 occurrences (19.3%), then Oregon with 248 occurrences (17.0%). California contained 168 occurrences, representing 11.5%, while Kentucky contained 155 occurrences, representing approximately 10.7% of the national inventory.

As Table 7.2 shows, the five highest-ranking states, Washington, West Virginia, Oregon, California, and Kentucky, account for roughly 81.9% of all documented landslide occurrences within 328 ft (100 m) of railroad infrastructure. The top ten states collectively accounted for approximately 94.7% of the inventory. This striking concentration that demonstrates documented railroad-adjacent landslides are not evenly distributed across states. Instead, a relatively small number of states contain most of the mapped occurrences.

The state comparison is presented graphically in Figure 7.6. A horizontal bar chart is appropriate because it clearly displays the ranking and magnitude of documented landslide occurrences while accommodating longer state names.

Table 7.2 -Top ten states with documented landslide occurrences within 328 ft (100 m) of railroad infrastructure.

Rank	State Name	Landslide Count	Percent of Total	Cumulative %
1	Washington	340	23.4%	23.4%
2	West Virginia	281	19.3%	42.7%
3	Oregon	248	17.0%	59.7%
4	California	168	11.5%	71.3%
5	Kentucky	155	10.7%	81.9%
6	North Carolina	90	6.2%	88.1%
7	Colorado	44	3.0%	91.1%
8	Pennsylvania	23	1.6%	92.7%
9	Minnesota	18	1.2%	94.0%
10	Missouri	11	0.8%	94.7%

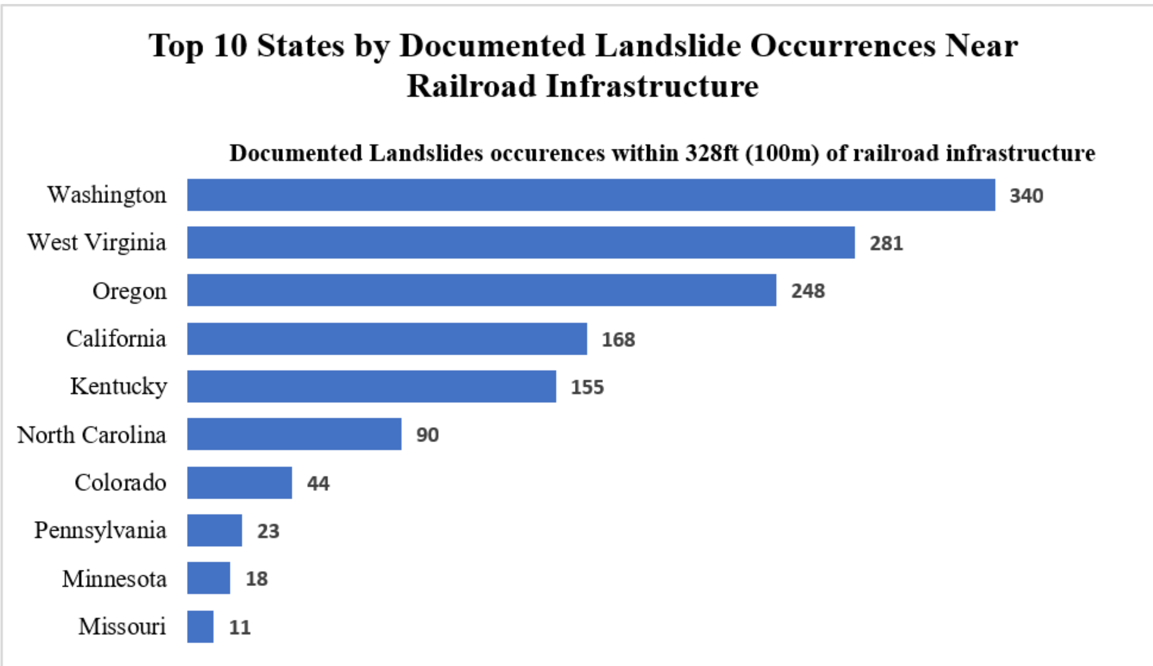


Figure 7.6. Top ten states with the highest number of documented landslide occurrences identified within 328 ft (100 m) of railroad infrastructure across the United States.

The horizontal bar chart works well here because it displays both the ranking and the magnitude of documented occurrences clearly, while still accommodating longer state names. Washington, West Virginia, and Oregon each contained more than 240 documented occurrences, and the counts declined considerably after Kentucky. North Carolina holds 90 occurrences, while each of the remaining states in the top ten contained fewer than 50. That steep decline reinforces the same conclusion that documented landslide exposure near railroad infrastructure is dominated by a limited number of geographically distinct regions.

These results represent historical landslide exposure within a defined proximity to railroad infrastructure, rather than the probability that a future landslide will occur. A high state-level count could reflect some combination of actual landslide activity, terrain conditions, railroad alignment, and differences in how complete each landslide inventory is. With that in mind, these findings are best read as a national screening inventory that points to locations where more detailed corridor-scale investigation may be worthwhile.

7.4.2 Geographic Concentration of Landslides Near Railroad Infrastructure

The national distribution in Section 7.1 shows that documented landslide occurrences near railroad infrastructure concentrated within a limited number of geographic regions rather than evenly distributed across the United States. Two dominant concentrations stand out: a western cluster along the Pacific Coast and an eastern cluster within the Appalachian region. Together, these two regions account for most of the documented landslides located within the 328-ft (100 m) railroad screening zone.

The Pacific Coast concentration spans Washington, Oregon, and California. Together, these three states hold 756 documented landslide occurrences, representing approximately 52.0% of the national inventory. Washington contributes 340, Oregon contributed 248, and California contributed 168. This concentration follows railroad corridors through coastal areas, mountain ranges, river valleys, and other topographically complex terrain, meaning more than half of the national railroad adjacent landslide inventory sits in just three western states.

Table 7.3. Regional concentration of documented landslide occurrences identified within 328 ft (100 m) of railroad infrastructure.

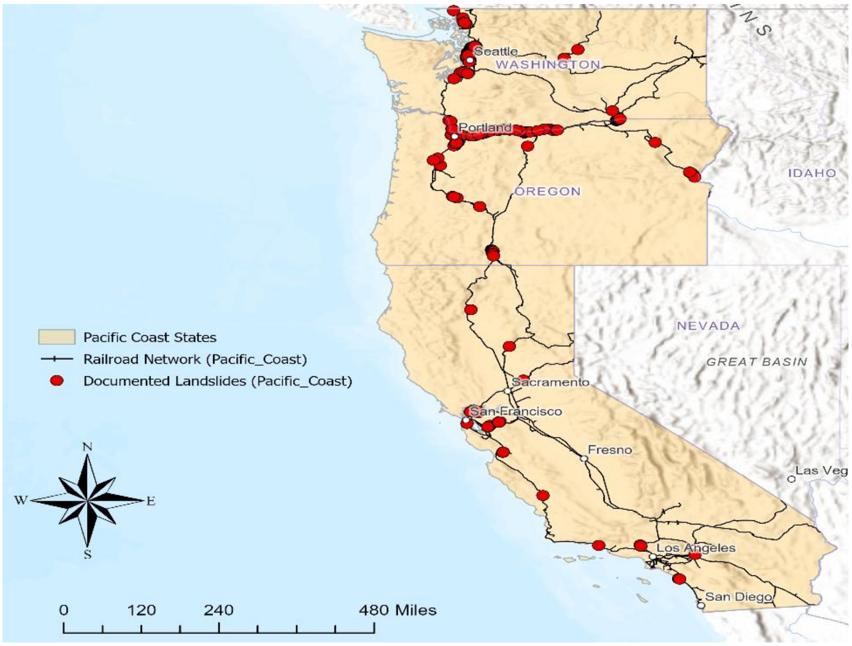
Geographic Concentration	Included States	Landslide Count	Percent of National Total (%)
Appalachian	West Virginia, Kentucky, North Carolina, Pennsylvania	549	37.7
Pacific Coast	Washington, Oregon, California	756	52.0
Other States	Remaining States	150	10.3
National Total	**United States**	**1455**	**100.0**

The Appalachian concentration covers West Virginia, Kentucky, North Carolina, and Pennsylvania. Collectively, these four states account for 549 documented landslide occurrences, representing approximately 37.7% of the national inventory. West Virginia carries the largest share within this region at 281 occurrences, followed by Kentucky with 155, North Carolina with 90, and Pennsylvania with 23. The mapped pattern showed substantial clustering along railroad corridors running through the Appalachian Mountains, particularly in West Virginia and eastern Kentucky.

Combined, the Pacific Coast and Appalachian concentrations contained 1,305 landslide occurrences, equivalent to approximately 89.7% of all documented landslides identified within the railroad screening zone. The remaining 150 occurrences, about 10.3% of the inventory, are scattered across other states. That degree of concentration indicates that national railroad landslide exposure is dominated by a relatively small number of geographic regions.

The Pacific Coast and Appalachian concentrations display different spatial patterns. Along the Pacific Coast, documented occurrences spread across several states and follow both coastal and inland railroad corridors. In the Appalachian region, landslide occurrences cluster more tightly along rail lines running through mountainous terrain and narrow valleys. In both cases, the mapped occurrences tend to trace the linear geometry of the railroad network itself, underscoring the value of corridor-based proximity screening for spotting where historical landslide activity intersects transportation infrastructure.

(a) Pacific Coast states



(b) Appalachian states

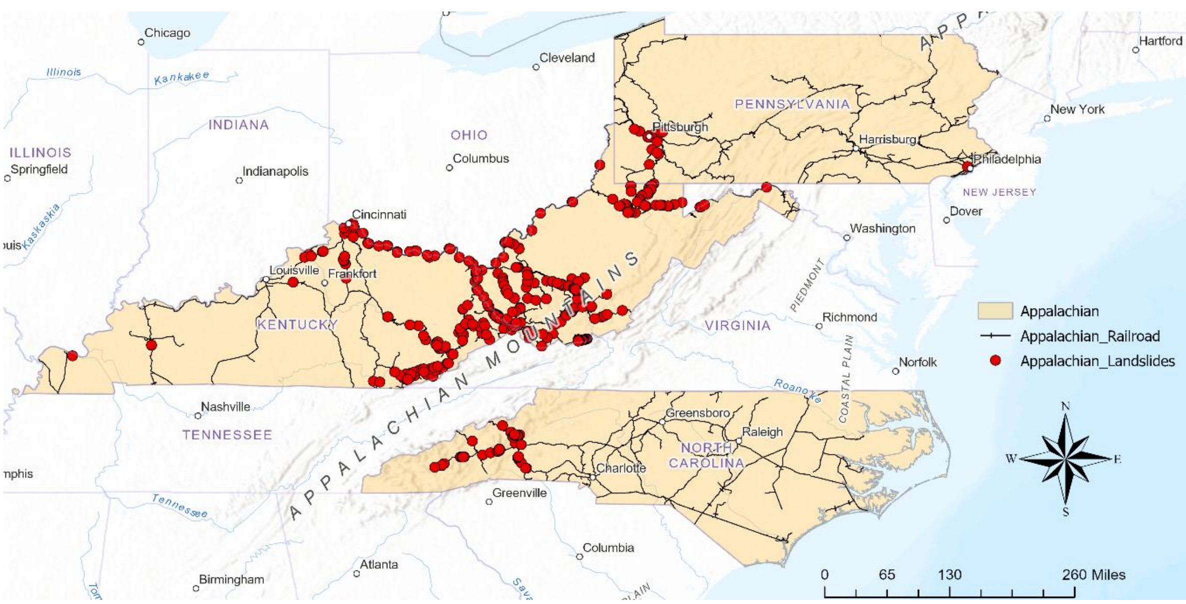


Figure 7.7. Regional concentrations of documented landslide occurrences identified within 328 ft (100 m) of railroad infrastructure: (a) Pacific Coast states and (b) Appalachian states

These concentration patterns are best read as evidence of documented historical exposure, not as a direct measure of future landslide probability or railroad failure risk. Regional differences may reflect real variation in landslide occurrence, railroad alignment, terrain, and how complete or detailed state and federal landslide inventories are. Even accounting for that uncertainty, the strong concentration of documented occurrences in the Pacific Coast and Appalachian regions marks these areas as reasonable priorities for more detailed corridor-scale geotechnical, hydrologic, and infrastructure assessment.

7.4.3 Railroad Ownership Exposure Assessment

A total of 1,455 documented landslides were identified within the 328-ft (100-m) screening buffer surrounding railroad infrastructure. Classification of the inventory by railroad ownership indicated a marked concentration of documented landslides along four major freight carriers. BNSF Railway exhibited the highest incidence of documented landslides (n = 475), followed by Union Pacific (n = 356), CSX Transportation (n = 348), and Norfolk Southern (n = 180). Collectively, these four operators accounted for 1,359 documented landslides, representing 93.4% of all documented landslides identified within the national screening buffer.

Figure 7.8 summarizes the distribution of documented landslides among four major freight railroad operators. BNSF Railway exhibits the highest number of documented landslides, followed by Union Pacific and CSX Transportation, with Norfolk Southern ranking fourth. Although these operators represent only a subset of the railroads included in the analysis, they collectively account for nearly all documented landslides identified within the screening buffer. This pattern reflects the extensive mileage of these Class I railroads through mountainous terrain and regions characterized by steep slopes, complex geology, and historically active mass movement.

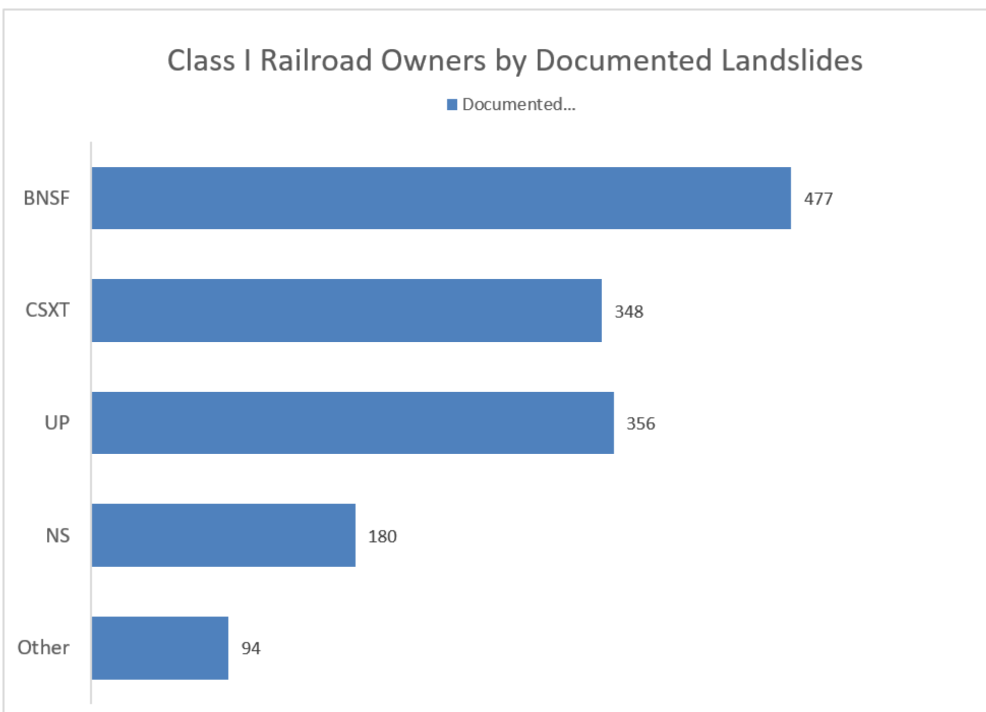


Figure 7.8. Distribution of documented landslides among the four major railroad operators within the national 328 ft screening buffer zone.

Table 7.4 demonstrates that documented landslides are highly concentrated among a small subset of Class I freight railroads. BNSF Railway accounts for the greatest number of documented landslides (n = 477), followed by Union Pacific (n = 356), CSX Transportation (n = 348), and Norfolk Southern (n = 180). These findings indicate that more than 93% of documented landslides within the national railroad buffer occur along four major railroad systems, underscoring a pronounced concentration of landslide occurrence among a limited number of operators.

Table 7.4 Distribution of documented landslides within the 328 ft (100 m) railroad buffer by owner and State

Railroad Owner	**State**	**Documented Landslides**
BNSF	California	96
BNSF	Idaho	2
BNSF	Iowa	4
BNSF	Minnesota	8
BNSF	Missouri	7
BNSF	Nebraska	11
BNSF	New Mexico	4
BNSF	Oregon	6
BNSF	Washington	337
BNSF	Wisconsin	2
BNSF	Wyoming	2
CSXT	Kentucky	130
CSXT	Maine	1
CSXT	Maryland	3
CSXT	Massachusetts	1
CSXT	North Carolina	53
CSXT	Pennsylvania	5
CSXT	Tennessee	2
CSXT	Virginia	2
CSXT	West Virginia	151
NS	Kentucky	24
NS	Missouri	2
NS	New Jersey	3
NS	North Carolina	22
NS	Ohio	5
NS	Pennsylvania	14
NS	Virginia	3
NS	West Virginia	107
UP	California	65
UP	Colorado	44
UP	Idaho	2
UP	Minnesota	4
UP	Missouri	1
UP	New Mexico	1
UP	Oregon	233
UP	Utah	6

***Distribution of documented landslides within 328 ft (100 m) screening buffer by railroad owner and State. Totals represent unique documented landslides intersecting buffered railroad infrastructure.

Figure 7.9 reveals a pronounced regional separation in documented landslide exposure among railroad operators. BNSF Railway and Union Pacific are predominantly affected in the western United States, with Washington, Oregon, and California accounting for the majority of documented landslides. Conversely, CSX Transportation and Norfolk Southern exhibit the highest documented landslide concentrations in West Virginia, Kentucky, and North Carolina, reflecting the influence of the Appalachian Mountains. The results demonstrate that documented landslide exposure is strongly controlled by regional physical geography and terrain rather than being uniformly distributed across the national railroad network.

From an infrastructure management perspective, these findings indicate that national landslide screening should prioritize geographic exposure rather than railroad ownership alone. While BNSF Railway and Union Pacific exhibit the largest number of documented landslides nationally, their exposure is concentrated within western mountain corridors, whereas CSX Transportation and Norfolk Southern are primarily affected within the Appalachian region. This regional distinction suggests that maintenance strategies, inspection priorities, and resilience planning should be tailored to the dominant geomorphic conditions encountered by each railroad system rather than adopting a uniform national approach.

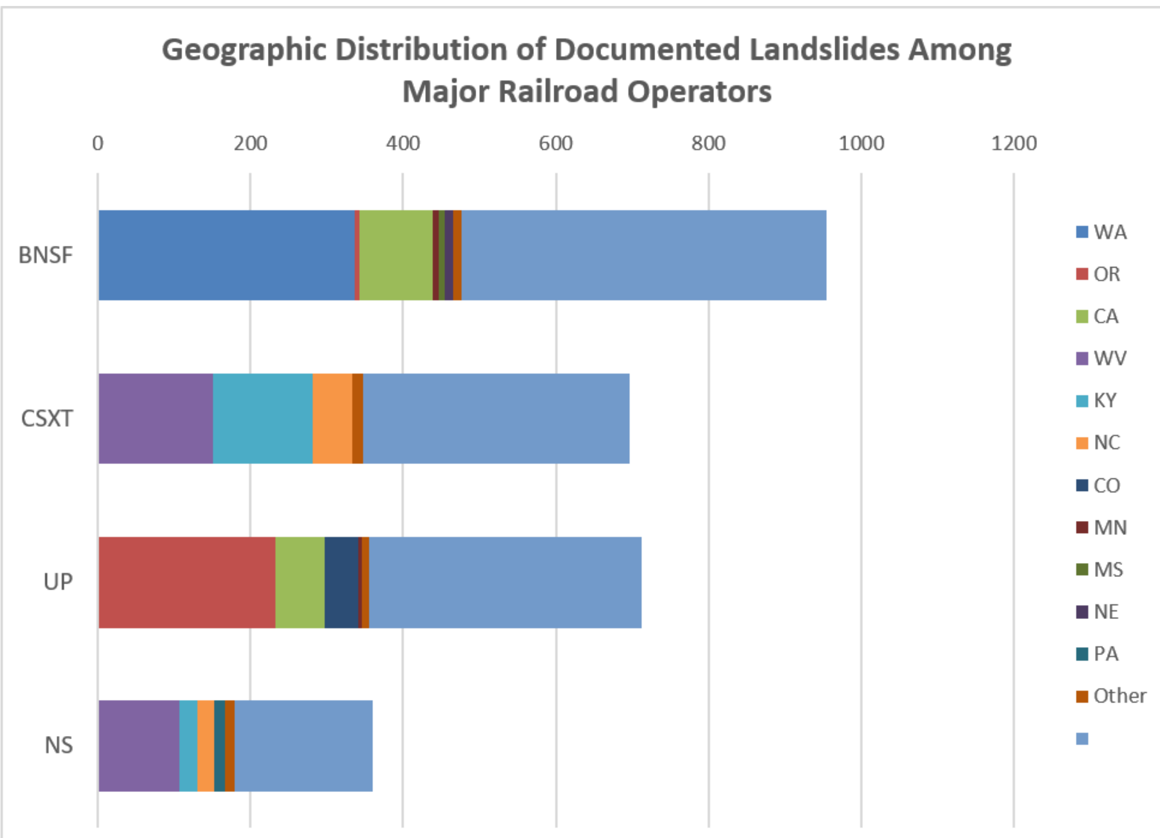


Figure 7.9. Geographic distribution of documented landslides among the four largest railroad operators. Bar segments indicate the contribution of individual states to each railroad's documented landslide inventory.

7.4.4 Implications for National Railroad Infrastructure Screening

The national inventory built in this study shows the value of Geographic Information Systems (GIS) as a rapid screening tool for identifying where documented landslide occurrences intersect railroad infrastructure. By combining the United States Geological Survey (USGS) National Landslide Inventory with the Federal Railroad Administration (FRA) railroad network, the analysis provides a consistent, reproducible framework for identifying historically exposed railroad corridors using publicly available geospatial data. Unlike detailed geotechnical investigations, which require extensive field surveys and site-specific analyses, the GIS-based approach offers a cost-effective first-level assessment that can be applied uniformly across the national rail system.

The results indicate that documented landslide exposure near railroad infrastructure concentrates heavily within a small number of geographic regions. Approximately 89.7% of all documented landslide occurrences identified within the 328-ft (100-m) railroad screening zone sit within the Pacific Coast and Appalachian regions. That level of spatial concentration suggests that transportation agencies can use national-scale inventory analysis to identify regions where corridor inspections, maintenance planning, and more detailed engineering evaluations are likely to deliver the most value. Rather than treating every railroad corridor the same, infrastructure managers can steer resources toward the areas where historical landslide exposure has been consistently documented.

The inventory also demonstrates why combining transportation and hazard datasets within a common GIS environment matters. Railroad infrastructure often traverses mountainous terrain, river valleys, and coastal environments, the same places landslides have historically occurred. Mapping the spatial relationship between documented landslide locations and railroad corridors lets agencies quickly flag potentially vulnerable segments without needing an immediate field investigation. That kind of information feeds directly into asset management, emergency preparedness, inspection scheduling, and long-term resilience planning by highlighting where historical slope instability has intersected transportation infrastructure.

The inventory is useful for screening, but it has a boundary that should not be crossed: it does not predict future landslide occurrences or railroad failure. The mapped locations represent documented historical landslide events located within a defined proximity to railroad infrastructure, and do not account for slope geometry, lithology, precipitation intensity, drainage conditions, vegetation, or other controlling factors that actually control landslide initiation. The inventory is most useful as a decision-support tool for identifying candidate corridors where more comprehensive geotechnical, hydrologic, or remote sensing investigations may be warranted.

The GIS workflow developed in this study is fully reproducible and can be readily updated as new landslide inventories or railroad datasets become available. Because the methodology relies on nationally available geospatial data and standard GIS proximity analysis, it can be adapted to other transportation assets such as highways, pipelines, transmission corridors, and utility networks. As more hazard inventories become available, the framework may also be expanded to cover other geologic hazards, such as debris flows, rockfalls, flooding, and coastal erosion. The ability to bring multiple spatial datasets together within one consistent analytical framework strengthens the case for GIS as a national infrastructure screening tool and gives transportation agencies a practical foundation for data-driven maintenance and resilience planning.

7.5 Discussion

7.5.1 Geographic Patterns of Railroad Landslide Exposure

The national inventory demonstrates that documented landslide occurrences near railroad infrastructure are not uniformly distributed across the United States but instead exhibit strong regional clustering. Approximately 89.7% of the documented landslides within the 328-ft (100-m) railroad screening zone sit within the Pacific Coast and Appalachian regions. That pattern lines up with prior research showing landslide occurrences driven by topographic relief, climate, and geological setting rather than being randomly distributed across transportation networks.

The Pacific Coast states, particularly Washington, Oregon, and California, contain extensive mountain ranges, steep river valleys, and coastal bluffs where heavy precipitation and complex geology combine to produce recurring slope instability. The Appalachian region tells a similar story, as West Virginia, Kentucky, North Carolina, and Pennsylvania are defined by rugged terrain, deeply incised valleys, weathered bedrock, and prolonged rainfall events, all of which promote landslide activity.

The regional concentration identified in this study highlights the importance of landscape characteristics when evaluating transportation hazards. This inventory was built solely from documented landslide occurrences near railroad infrastructure, yet the spatial patterns align with what is already understood about landslide prone environments, which is itself a useful confirmation that national geospatial inventories can effectively identify regions where transportation infrastructure has historically faced slope instability.

An additional perspective provided by this study is the relationship between regional landslide occurrence and railroad ownership. Although more than twenty railroad operators intersect documented landslides within the national screening buffer, exposure is overwhelmingly concentrated among four Class I freight railroads. BNSF Railway (475 documented landslides), Union Pacific (356), CSX Transportation (348), and Norfolk Southern (180) collectively account for 93.4% of all documented landslides identified within the national inventory. Importantly, the geographic distribution of these documented landslides mirrors the broader regional patterns observed in the state-level analysis. BNSF Railway and Union Pacific are primarily affected within the Pacific Coast states, whereas CSX Transportation and Norfolk Southern are concentrated within the Appalachian region. These findings demonstrate that railroad ownership does not alter the regional controls governing landslide occurrence but instead reflects where major rail networks intersect historically unstable terrain.

7.5.2 Railroad Network Size versus Landslide Exposure

One of the most significant findings of this study is that the size of a state's railroad network does not necessarily correspond to the number of documented landslides located near railroad infrastructure. Texas operates the largest railroad network in the United States, with more than 9,114 miles of railroad, yet relatively few documented landslide occurrences were identified within the 328-ft screening zone. In contrast, Washington and West Virginia, with 1,960 and 1,854 miles, respectively, have much smaller railroad networks but hold the highest counts of documented landslides adjacent to railroad infrastructure.

That contrast demonstrates that railroad mileage alone is not a reliable indicator of landslide exposure. Instead, the spatial distribution of documented landslides appears to be governed mainly by terrain, geology, and historical landslide occurrence. Railroad corridors constructed through mountainous regions naturally cross steeper slopes, less stable hillsides, and more active drainage systems where conditions are favorable for landslides. Extensive railroad systems on relatively flat terrain may see lower exposure despite containing thousands of miles of track.

The findings further demonstrate that documented landslide exposure is influenced by where railroad networks traverse mountainous terrain rather than by the total extent of railroad infrastructure. Although BNSF Railway, Union Pacific, CSX Transportation, and Norfolk Southern account for the overwhelming majority of documented landslides nationally, this reflects the geographic characteristics of their corridors rather than railroad ownership itself. Western Class I railroads traverse the Cascade Range, Sierra Nevada, Coast Ranges, and Rocky Mountains, while eastern operators cross the Appalachian Mountains, where steep slopes, complex geology, and prolonged precipitation have historically produced frequent slope failures. Consequently, exposure should be interpreted as a function of terrain and historical landslide occurrence rather than railroad mileage or ownership alone.

7.5.3 Utility of GIS-Based National Railroad Screening

The GIS-based workflow presented here shows how effectively public transportation and hazard datasets can be combined into a consistent national screening inventory. By pairing the USGS National Landslide Inventory with the FRA railroad network, the methodology rapidly identifies documented landslide occurrences located near railroad infrastructure without requiring extensive field investigations or site-specific geotechnical analyses.

Its biggest advantage is its reproducibility and scalability. Because the workflow depends on nationally available datasets and standard GIS proximity analysis, it can be updated as new landslide inventories become available or adapted for use by transportation agencies at national, regional, or state levels. The resulting inventory functions as a practical first-level assessment that identifies historically exposed corridors needing further investigation, rather than attempting to predict future landslide occurrence.

Beyond identifying regional hotspots, the GIS-based screening framework also provides valuable operational insight by linking documented landslides to individual railroad operators. This allows transportation agencies and railroad companies to identify not only where documented landslides have historically occurred but also which railroad systems experience the greatest historical exposure. Such information supports risk-informed maintenance planning, inspection prioritization, and infrastructure resilience by enabling agencies to focus resources on corridors where documented landslide occurrence has historically been concentrated. Because the workflow relies exclusively on publicly available geospatial datasets, it provides a transparent, scalable, and repeatable framework that can be updated as national landslide inventories or railroad datasets evolve.

7.5.4 Study Limitations

Although the inventory provides valuable information regarding historical landslide exposure near railroad infrastructure, several limitations should be acknowledged. First, the analysis is based exclusively on documented landslide occurrences contained within the USGS National Landslide Inventory. How to complete the inventory varies geographically, since landslide reporting practices, mapping efforts, and historical documentation differ from state to state. That means some regions almost certainly contain undocumented landslides that never show up in this dataset.

Second, the 328-ft (100-m) proximity threshold was selected as a standardized screening distance for identifying landslides adjacent to railroad infrastructure. It gives the analysis a consistent national framework, but proximity alone does not establish a direct geotechnical relationship between a documented landslide and railroad performance. Some mapped landslides inside the buffer may have no real effect on railroad operations, while hazards located outside the buffer could still influence infrastructure through debris movement, erosion, or drainage impacts.

Finally, this study does not evaluate slope geometry, lithology, rainfall intensity, groundwater conditions, vegetation, or other environmental factors controlling landslide initiation. Therefore, the inventory should be interpreted as a historical exposure assessment rather than a landslide susceptibility or hazard prediction model.

7.5.5 Future Research

The GIS framework developed in this study provides a foundation for more comprehensive railroad hazard assessments. Future research should bring in higher resolution topographic datasets, geological mapping, precipitation records, and remotely sensed terrain information to sharpen corridor-level hazard characterization. Incorporating LiDAR-derived elevation models, slope analyses, and hydrologic analyses would support more detailed evaluation of how terrain conditions influence railroad stability.

Future investigations should expand the railroad ownership analysis by incorporating railroad route mileage to develop standardized exposure metrics, such as documented landslides per 100 miles of railroad. Such normalization would enable more equitable comparisons among railroad operators with substantially different network lengths and would distinguish between exposure resulting from extensive railroad mileage and exposure associated with particularly hazardous terrain. Integrating these normalized metrics with maintenance records, traffic density, and service interruption data would further strengthen infrastructure risk assessments.

Additional work should also investigate exposure metrics that account for railroad mileage within individual states. Comparing documented landslide occurrence per unit length of railroad infrastructure could provide a standardized measure of exposure and facilitate more meaningful comparisons among states with substantially different railroad network sizes. Bringing in historical maintenance records, service disruption data, and field observations would further help validate GIS-based screening results and strengthen the practical case for national railroad hazard inventories.

7.6 Conclusion and Recommendations

This study developed a national Geographic Information System (GIS) based workflow to identify documented landslide occurrences located within 328 ft (100 m) of railroad infrastructure across the United States. By integrating the United States Geological Survey (USGS) National Landslide Inventory with the Federal Railroad Administration (FRA) railroad network, it establishes a standardized, reproducible screening methodology built entirely on publicly available geospatial datasets. The resulting inventory provides the transportation agencies a consistent framework for identifying railroad corridors that have historically been exposed to landslide activity.

The analysis identified 1,455 documented landslide occurrences within the defined railroad screening distance. Although documented landslides were identified across numerous states and railroad systems, their distribution was highly concentrated geographically and operationally. The Pacific Coast states of Washington, Oregon, and California accounted for approximately 52.0% of the national inventory, while the

Appalachian states of West Virginia, Kentucky, North Carolina, and Pennsylvania contributed an additional 37.7%, together representing nearly 90% of all documented landslides adjacent to railroad infrastructure. Furthermore, four Class I freight railroads—BNSF Railway, Union Pacific, CSX Transportation, and Norfolk Southern—accounted for 93.4% of all documented landslides identified within the national screening buffer, demonstrating that historical landslide exposure is concentrated within a relatively small number of geographically distinct transportation corridors. A key finding of this study is that railroad network size alone does not determine landslide exposure. Texas possesses the largest railroad network of 9,114 miles in the United States but exhibited comparatively few documented landslides near rail infrastructure. In contrast, Washington and West Virginia with 1,960 and 1,854 miles, respectively, have much smaller railroad networks but hold the highest counts of documented landslide occurrences adjacent to railroad corridors. Terrain, geology, climate, and the historical record of landslide activity clearly carry more weight in determining landslide exposure than railroad mileage does on its own. That means prioritizing infrastructure purely by network length risks overlooking corridors that actually experience substantially greater geologic risk.

By incorporating railroad ownership into the national screening framework, this study extends traditional proximity-based inventory analyses beyond geographic identification to include an operational perspective relevant to transportation asset management. The ability to associate documented landslide occurrences with specific railroad operators provides transportation agencies and infrastructure managers with actionable information for prioritizing inspections, allocating maintenance resources, and supporting long-term resilience planning. This ownership-based perspective complements regional hazard mapping and demonstrates how publicly available geospatial datasets can support strategic decision-making at national and regional scales.

The findings demonstrate that railroad network size alone does not determine documented landslide exposure. Texas maintains the nation's largest railroad network yet exhibits relatively few documented landslides within the screening buffer, whereas Washington and West Virginia contain substantially smaller railroad networks but considerably higher concentrations of documented landslides. Similarly, the high documented landslide counts observed for BNSF Railway, Union Pacific, CSX Transportation, and Norfolk Southern reflect the mountainous terrain traversed by their rail corridors rather than railroad ownership itself. These results confirm that terrain, geology, climate, and historical landslide occurrence are the primary controls governing national railroad landslide exposure.

Bibliography

1. Sheng, Z., Owolabi, O., Liu, Y. Final Report: Improve Highway Safety by Reducing the Risks of Landslides (Phase 1). Morgan State University, Baltimore, MD, 2025.
2. Sheng, Z., Owolabi, O., Liu, Y. Final Report: Improve Highway Safety by Reducing the Risks of Landslides (Phase 2). Morgan State University, Baltimore, MD, 2025.
3. Highland, L. M., and Bobrowsky, P. The Landslide Handbook - A Guide to Understanding Landslides. U.S. Geological Survey Circular 1325, 2008.
4. Hungr, O., Leroueil, S., and Picarelli, L. The Varnes classification of landslide types, an update. Landslides, 11, 167-194, 2014.
5. Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., and Savage, W. Z. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Engineering Geology, 102, 85-98, 2008.
6. Gannett Fleming, Inc. Stability Evaluation and Conceptual Mitigation Recommendations Letter Report - I-68 at Haystack Mountain, MDSHA Contract No. BCS 2014-03D. Maryland Department of Transportation State Highway Administration, 2018.
7. Torres, R., et al. GMES Sentinel-1 mission. Remote Sensing of Environment, 120, 9-24, 2012.
8. Massonnet, D., and Feigl, K. L. Radar interferometry and its application to changes in the Earth's surface. Reviews of Geophysics, 36, 441-500, 1998.
9. Rosen, P. A., et al. Synthetic aperture radar interferometry. Proceedings of the IEEE, 88, 333-382, 2000.
10. Berardino, P., Fornaro, G., Lanari, R., and Sansosti, E. A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. IEEE Transactions on Geoscience and Remote Sensing, 40, 2375-2383, 2002.
11. Lanari, R., et al. A small-baseline approach for investigating deformations on full-resolution differential SAR interferograms. IEEE Transactions on Geoscience and Remote Sensing, 42, 1377-1386, 2004.
12. Ferretti, A., Prati, C., and Rocca, F. Permanent scatterers in SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing, 39, 8-20, 2001.
13. Hooper, A., Zebker, H., Segall, P., and Kampes, B. A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. Geophysical Research Letters, 31, L23611, 2004.
14. Crosetto, M., Monserrat, O., Cuevas-Gonzalez, M., Devanthery, N., and Crippa, B. Persistent scatterer interferometry: a review. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 78-89, 2016.
15. Colesanti, C., and Wasowski, J. Investigating landslides with space-borne synthetic aperture radar interferometry. Engineering Geology, 88, 173-199, 2006.
16. Wasowski, J., and Bovenga, F. Investigating landslides and unstable slopes with satellite multi-temporal interferometry: current issues and future perspectives. Engineering Geology, 174, 103-138, 2014.
17. Notti, D., Davalillo, J. C., Herrera, G., and Mora, O. A methodology for improving landslide PSI data analysis. International Journal of Remote Sensing, 35, 2186-2214, 2014.
18. Cascini, L., Fornaro, G., and Peduto, D. Advanced low- and full-resolution DInSAR map generation for slow-moving landslide analysis at different scales. Engineering Geology, 112, 29-42, 2010.
19. Bekaert, D. P. S., Handwerger, A. L., Agram, P., and Kirschbaum, D. B. InSAR-based detection method for mapping and monitoring slow-moving landslides in remote

regions with steep and mountainous terrain: an application to Nepal. Remote Sensing of Environment, 249, 111983, 2020.
20. Glaser, J. D., and Brezinski, D. K. Geologic Map of the Cumberland Quadrangle, Allegany County, Maryland. Maryland Geological Survey, 1994.
21. Brezinski, D. K., and Conkwright, R. D. Geologic Map of Garrett, Allegany, and Western Washington Counties, Maryland. Maryland Geological Survey, 2013.
22. Horn, B. K. P. Hill shading and the reflectance map. Proceedings of the IEEE, 69, 14-47, 1981.
23. Casagli, N., et al. Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning. Geoenvironmental Disasters, 3, 9, 2016.
24. Carla, T., Intrieri, E., Di Traglia, F., Nolesini, T., Gigli, G., and Casagli, N. Guidelines on the use of inverse velocity method as a tool for setting alarm thresholds and forecasting landslides and structure collapses. Landslides, 14, 517-534, 2017.
25. Handwerger, A. L., Huang, M.-H., Fielding, E. J., Booth, A. M., and Burgmann, R. A shift from drought to extreme rainfall drives a stable landslide to catastrophic failure. Scientific Reports, 9, 1569, 2019.
26. Lacroix, P., Handwerger, A. L., and Bievre, G. Life and death of slow-moving landslides. Nature Reviews Earth & Environment, 1, 404-419, 2020.
27. Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. STL: a seasonal-trend decomposition procedure based on Loess. Journal of Official Statistics, 6, 3-73, 1990.
28. Box, G. E. P., and Jenkins, G. M. Time Series Analysis: Forecasting and Control. Holden-Day, 1976.
29. Hochreiter, S., and Schmidhuber, J. Long short-term memory. Neural Computation, 9, 1735-1780, 1997.
30. Yang, B., Yin, K., Lacasse, S., and Liu, Z. Time series analysis and long short-term memory neural network to predict landslide displacement. Landslides, 16, 677-694, 2019.
31. Zhou, X., Wen, H., Zhang, Y., Xu, J., and Zhang, W. A deep-learning method for landslide displacement prediction based on multi-temporal InSAR time series. Remote Sensing of Environment, 318, 114580, 2025.
32. Baltimore City Department of Transportation. 26th Street Wall Collapse, community presentation, 2018. https://s3.amazonaws.com/baltimorecity.gov.if-us-east-1/s3fs-public/2024-01/26th_street_wall_presentation_final.pdf.
33. RK&K. 26th Street Retaining Wall Emergency Repair, Baltimore, Maryland, project description, 2026. https://www.rkk.com/our-work/portfolio/26th-street-retaining-wall-emergency-repair-baltimore-maryland/.
34. Huntley, D., Rotheram-Clarke, D., Pon, A., et al. Benchmarked RADARSAT-2, Sentinel-1 and RADARSAT Constellation Mission change-detection monitoring at North Slide, Thompson River Valley, British Columbia: ensuring a landslide-resilient national railway network. Canadian Journal of Remote Sensing, 47, 635-656, 2021.
35. Schlogl, M., Widhalm, B., and Avian, M. Comprehensive time-series analysis of bridge deformation using differential satellite radar interferometry based on Sentinel-1. ISPRS Journal of Photogrammetry and Remote Sensing, 172, 132-146, 2021.
36. Pechnikov, A. PyGMTSAR: Python InSAR, open-source satellite interferometry software, 2025. https://github.com/AlexeyPechnikov/pygmtsar.
37. Fuhrmann, T., and Garthwaite, M. C. Resolving three-dimensional surface motion with InSAR: constraints from multi-geometry data fusion. Remote Sensing, 11, 241, 2019.

38. Intrieri, E., Carla, T., and Gigli, G. Forecasting the time of failure of landslides at slope scale: a literature review. Earth-Science Reviews, 193, 333-349, 2019.
39. Chen, T., and Guestrin, C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794, 2016.
40. van Natijne, A. L., Bogaard, T. A., Dehls, J. F., Lindenbergh, R. C., and van Leijen, F. J. Machine-learning-based nowcasting of the Vogelsberg deep-seated landslide. Natural Hazards and Earth System Sciences, 23, 3723-3740, 2023.
41. Zhou, C., Cao, Y., Gan, L., et al. A novel framework for landslide displacement prediction using MT-InSAR and machine-learning techniques. Engineering Geology, 334, 107497, 2024.
42. Hussain, M. A., Chen, Z., Zheng, Y., et al. Landslide susceptibility mapping using machine-learning algorithm validated by persistent scatterer InSAR technique. Sensors, 22, 3119, 2022.
43. Miao, F., Ruan, Q., Wu, Y., et al. Landslide dynamic susceptibility mapping based on machine learning and the PS-InSAR coupling model. Remote Sensing, 15, 5427, 2023.
44. Wei, Y., Qiu, H., Liu, Z., et al. Refined and dynamic susceptibility assessment of landslides using InSAR and machine-learning models. Geoscience Frontiers, 15, 101890, 2024.
45. Wooten, R. M., Witt, A. C., Miniat, C. F., Hales, T. C., and Aldred, J. L. Frequency and magnitude of selected historical landslide events in the Southern Appalachian Highlands of North Carolina and Virginia. In Natural Disturbances and Historic Range of Variation, Springer, 203-262, 2016.
46. Manunta, M., De Luca, C., Zinno, I., et al. The parallel SBAS approach for Sentinel-1 interferometric wide swath deformation time-series generation: algorithm description and products quality assessment. IEEE Transactions on Geoscience and Remote Sensing, 57, 6259-6281, 2019.
47. Lin, S., Chen, S.-E., Rasanen, R. A., et al. Landslide prediction validation in western North Carolina after Hurricane Helene. Geotechnics, 4, 1064, 2024.
48. Duncan JM, Wright SG. Soil Strength and Slope Stability. Wiley, 2005.
49. Abramson LW, Lee TS, Sharma S, Boyce GM. Slope Stability and Stabilization Methods. Wiley, 2002.
50. Fredlund DG, Krahn J. Comparison of slope stability methods of analysis. Canadian Geotechnical Journal. 1977.
51. Griffiths DV, Lane PA. Slope stability analysis by finite elements. Géotechnique. 1999.
52. Dawson EM, Roth WH, Drescher A. Slope stability analysis by strength reduction. Géotechnique. 1999.
53. Itasca Consulting Group. FLAC3D Theory and Background Manual. https://docs.itascacg.com/flac3d700/common/docproject/source/manual/program_guide/program_guide.html?node67 accessed in August 2026.
54. Bishop CM. Pattern Recognition and Machine Learning. Springer. 2006.
55. Bui D.T., Shahabi H., Shirzadi A., Chapi K., Alizadeh M., Chen W., Mohammadi A., Bin-Ahmad B., Panahi M., Hong H., et al. Landslide Detection and Susceptibility Mapping by AIRSAR Data Using Support Vector Machine and Index of Entropy Models in Cameron Highlands, Malaysia. Remote. Sens. 2018;10:1527. doi: 10.3390/rs10101527.
56. Breiman L. Random Forests. 2001.
57. Vapnik VN. Statistical Learning Theory. 1998.
58. Goodfellow I et al. Deep Learning. 2016.

59. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. NIPS, 2017.
60. Molnar C. Interpretable Machine Learning, 3rd ed. 2025.
61. Itasca Consulting Group. FLAC3D – Fast Lagrangian Analysis of Continua in 3 Dimensions, Version 5.0 User's Guide. Minneapolis, MN: Itasca Consulting Group, 2012.
62. Posner A, Georgakakos K (2016) An early warning system for landslide danger. Eos 97.
63. Guzzetti F, Reichenbach P, Ardizzone F, et al (2006) Estimating the quality of landslide susceptibility models. Geomorphology 81(1-2):166–184.
64. Iverson RM (2000) Landslide triggering by rain infiltration. Water resources research 36(7):1897–1910.
65. Liu Q, Su L, Zhang C, et al (2022) Dynamic variations of interception loss-infiltration-runoff in three land-use types and their influence on slope stability: An example from the eastern margin of the tibetan plateau. Journal of Hydrology 612:128218.
66. Zhang Z, Zeng R, Meng X, et al (2023) Effects of changes in soil properties caused by progressive infiltration of rainwater on rainfall-induced landslides. Catena 233:107475.
67. Baum RL, Godt JW, Savage WZ (2010) Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. Journal of Geophysical Research: Earth Surface 115(F3).
68. Coe JA, Michael JA, Crovelli RA, et al (2004) Probabilistic assessment of precipitation-triggered landslides using historical records of landslide occurrence, Seattle, Washington, vol 10, Association of Environmental & Engineering Geologists, pp 103–122.
69. Wieczorek G, Glade T, Jackob M, et al (2005) Debris-flow Hazards and Related Phenomena, Springer Berlin Heidelberg, Berlin, Heidelberg, chap Climatic factors influencing occurrence of debris flows, pp 325–362.
70. Yang H, Wei F, Ma Z, et al (2020) Rainfall threshold for landslide activity in dazhou, southwest china. Landslides 17(1):61–77.
71. Maturidi A, Kasim N, Taib KA, et al (2020) Empirically based rainfall threshold for landslides occurrence in cameron highlands, malaysia. Civil Eng Architect 8(6):1481–1490.
72. Rosi A, Segoni S, Canavesi V, et al (2021) Definition of 3d rainfall thresholds to increase operative landslide early warning system performances. Landslides 18(3):1045–1057.
73. Bhattacharya D, Ghosh JK, Boccardo P, et al (2013) Automated geo-spatial system for generalized assessment of socio-economic vulnerability due to landslide in a region. European Journal of Remote Sensing 46(1):379–399.
74. Xiao Y, Dai J (2020) Application and analysis on geological hazard monitoring and early warning system based on internet of things. In: Journal of Physics: Conference Series, IOP Publishing, p 032015.
75. Conrad JL, Morphew MD, Baum RL, et al (2021) Hydromet: a new code for automated objective optimization of hydrometeorological thresholds for landslide initiation. Water 13(13):1752.
76. Stanley TA, Kirschbaum DB, Benz G, et al (2021) Data-driven landslide nowcasting at the global scale. Frontiers in Earth Science 9:640043.
77. Vishnu C, Oommen T, Chatterjee S, et al (2022) Challenges of modeling rainfall triggered landslides in a data-sparse region: A case study from the western ghats, India. Geosystems and Geoenvironment 1(3):100060.

78. Caine N (1980) The rainfall intensity-duration control of shallow landslides and debris flows. Geografiska annaler: series A, physical geography 62(1-2):23–27.
79. Aleotti P (2004) A warning system for rainfall-induced shallow failures. Engineering geology 73(3-4):247–265.
80. Godt JW, Baum RL, Chleborad AF (2006) Rainfall characteristics for shallow landsliding in seattle, washington, usa. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group 31(1):97–110.
81. Brunetti MT, Peruccacci S, Rossi M, et al (2010) Rainfall thresholds for the possible occurrence of landslides in italy. Natural Hazards and Earth System Sciences 10(3):447–458.
82. Saito H, Nakayama D, Matsuyama H (2010) Relationship between the initiation of a shallow landslide and rainfall intensity—duration thresholds in Japan. Geomorphology 118(1-2):167–175.
83. Yubonchit S, Chinkulkijniwat A, Horpibulsuk S, et al (2017) Influence factors involving rainfall-induced shallow slope failure: numerical study. International Journal of Geomechanics 17(7):04016158.
84. Singh J, Thakur M, Dhiman RK, et al (2025) Development of rainfall threshold equation and bayesian probabilistic analysis for landslide prediction: A case study of shimla, northwestern himalaya, india. Natural Hazards Research 5(3):455–467.
85. Peruccacci S, Brunetti MT, Luciani S, et al (2012) Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central italy. Geomorphology 139:79–90.
86. Vessia G, Parise M, Brunetti MT, et al (2014) Automated reconstruction of rainfall events responsible for shallow landslides. Natural Hazards and Earth System Sciences 14(9):2399–2408.
87. Gariano SL, Sarkar R, Dikshit A, et al (2019) Automatic calculation of rainfall thresholds for landslide occurrence in chukha dzongkhag, bhutan. Bulletin of Engineering Geology and the Environment 78(6):4325–4332.
88. He S, Wang J, Liu S (2020) Rainfall event–duration thresholds for landslide occurrences in China. Water 12(2):494.
89. Lee WY, Park SK, Sung HH (2021) The optimal rainfall thresholds and probabilistic rainfall conditions for a landslide early warning system for chuncheon, republic of Korea. Landslides 18(5):1721–1739.
90. Guzzetti F, Melillo M, Mondini AC (2025) Landslide predictions through combined rainfall threshold models. Landslides 22(1):137–147.
91. Chleborad AF, Baum RL, Godt JW (2006) Rainfall thresholds for forecasting landslides in the Seattle, Washington, area: Exceedance and probability. US Geological Survey Open-File Report 1064:31.
92. Pennington C, Dijkstra T, Lark M, et al (2014) Antecedent precipitation as a potential proxy for landslide incidence in south west United kingdom. In: Landslide Science for a Safer Geoenvironment: Vol. 1: The International Programme on Landslides (IPL). Springer, p 253–259.
93. Crosta G (1990) A study of slope movements caused by heavy rainfall in Valtellina (Italy–July 1987). In: Proc. 6th Int. Conf. and Field Workshop on Landslides ALPS, pp 247–258.
94. Glade T, Crozier M, Smith P (2000) Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “antecedent daily rainfall model”. Pure and Applied Geophysics 157(6):1059–1079.
95. Teja TS, Dikshit A, Satyam N (2019) Determination of rainfall thresholds for landslide prediction using an algorithm-based approach: Case study in the Darjeeling Himalayas, India. Geosciences 9(7):302.

96. Menon V, Kolathayar S (2025) Empirical and machine learning-based approaches to identify rainfall thresholds for landslide prediction: a case study of Kerala, India. Discover Applied Sciences 7(3):203.
97. Zhao B, Zhang L, Gu X, et al (2025) How is the occurrence of rainfall-triggered landslides related to extreme rainfall? Geomorphology 475:109666.
98. Segoni S, Battistini A, Rossi G, et al (2015) An operational landslide early warning system at regional scale based on space–time-variable rainfall thresholds. Natural Hazards and Earth System Sciences 15(4):853–861.
99. Vaz T, Zˆezere JL, Pereira S, et al (2018) Regional rainfall thresholds for landslide occurrence using a centenary database. Natural Hazards and Earth System Sciences 18(4):1037–1054.
100. Guzzetti F, Peruccacci S, Rossi M, et al (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and atmospheric physics 98(3):239–267.
101. Berti M, Martina M, Franceschini S, et al (2012) Probabilistic rainfall thresholds for landslide occurrence using a bayesian approach. Journal of Geophysical Research: Earth Surface 117(F4).
102. Frattini P, Crosta G, Sosio R (2009) Approaches for defining thresholds and return periods for rainfall-triggered shallow landslides. Hydrological Processes: An International Journal 23(10):1444–1460.
103. Jaiswal P, van Westen CJ (2009) Estimating temporal probability for landslide initiation along transportation routes based on rainfall thresholds. Geomorphology 112(1-2):96–105.
104. Sarkar R, Dikshit A, Hazarika H, et al (2019) Probabilistic rainfall thresholds for landslide occurrences in Bhutan. International Journal of Recent Technology and Engineering.
105. Chinkulkijniwat A, Salee R, Horpibulsuk S, et al (2022) Landslide rainfall threshold for landslide warning in northern Thailand. Geomatics, Natural Hazards and Risk 13(1):2425–2441.
106. Dietrich WE, Reiss R, Hsu ML, et al (1995) A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. Hydrological processes 9(3-4):383–400.
107. Crozier MJ (1999) Prediction of rainfall-triggered landslides: A test of the antecedent water status model. Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group 24(9):825–833.
108. Baum RL, Savage WZ, Godt JW (2008) TRIGRS: a Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0. US Geological Survey Reston, VA, USA.
109. Pomeroy JS (1988) Map showing landslide susceptibility in Maryland. Tech. rep., US Geological Survey.
110. Clark WB (1918) The Geography of Maryland, vol 10. John Hopkins Press.
111. Ramanathan R, Aydilek AH, Tanyu BF (2015) Development of a GIS-base failure investigation system for highway soil slopes. Frontiers of Earth Science 9(2):165–178.
112. Hosseinizadeh A, Sheng Z, Liu Y (2025) High-resolution spatiotemporal mapping of surface soil moisture using convlstm model and sentinel-1 data. Water 17(22):3300.
113. Shepard D (1968) A two-dimensional interpolation function for irregularly spaced data. In: Proceedings of the 1968 23rd ACM national conference, pp 517–524.
114. Gumbel EJ (1954) Statistical theory of extreme values and some practical applications: a series of lectures, vol 33. US Government Printing Office.

115. Maryland Department of Transportation/State Highway Administration. (2022). MDOT SHA Geotechnical Asset Management Plan.
116. Terzaghi, K. (1950). Mechanism of landslides. In S. Paige (Ed.), Application of Geology to Engineering Practice. Geological Society of America.
117. Sidle, R. C., & Ochiai, H. (2006). Natural factors influencing landslides. In Landslides: Processes, Prediction, and Land Use (pp. 41–119). American Geophysical Union.
118. Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P. (2008). The rainfall intensity–duration control of shallow landslides and debris flows: An update. Landslides, 5(1), 3–17.
119. Segoni, S., Piciullo, L., & Gariano, S. L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. Landslides, 15(8), 1483–1501.
120. Huang, J., Wu, X., Ling, S., et al. (2022). A bibliometric and content analysis of research trends on GIS-based landslide susceptibility from 2001 to 2020. Environmental Science and Pollution Research, 29, 86954–86993.
121. Cruden, D. M., & Varnes, D. J. (1996). Landslide types and processes. In A. K. Turner & R. L. Schuster (Eds.), Landslides: Investigation and Mitigation, Transportation Research Board Special Report 247 (pp. 36–75). National Research Council.
122. Baum, R. L., & Godt, J. W. (2010). Early warning of rainfall-induced shallow landslides and debris flows in the USA. Landslides, 7, 259–272.
123. Aleotti, P. (2004). A warning system for rainfall-induced shallow failures. Engineering Geology, 73, 247–265.
124. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
125. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
126. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
127. Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11), 2673–2681.
128. Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. arXiv:1803.01271.
129. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. International Conference on Learning Representations (ICLR 2015).
130. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.
131. Lague, D., Brodu, N., & Leroux, J. (2013). Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z). ISPRS Journal of Photogrammetry and Remote Sensing, 82, 10–26.
132. Besl, P. J., & McKay, N. D. (1992). A method for registration of 3-D shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2), 239–256.
133. Chen, Y., & Medioni, G. (1992). Object modelling by registration of multiple range images. Image and Vision Computing, 10(3), 145–155.
134. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6), 381–395.
135. Rusu, R. B., Blodow, N., & Beetz, M. (2009). Fast point feature histograms

(FPFH) for 3D registration. Proceedings of the 2009 IEEE International Conference on Robotics and Automation, 3212–3217.
136. Qi, C. R., Yi, L., Su, H., & Guibas, L. J. (2017). PointNet++: Deep hierarchical feature learning on point sets in a metric space. Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 5099–5108.
137. Crosetto, M., Monserrat, O., Cuevas-González, M., Devanthéry, N., & Crippa, B. (2016). Persistent scatterer interferometry: A review. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 78–89.111
138. Kang, S. K., Sheng, Z., Liu, Y., & Owolabi, O. (2025). Numerical simulation-guided machine learning for landslide risk assessment. AGU Fall Meeting 2025.
139. Okegbola, M., Owolabi, O., Liu, Y., & Sheng, Z. (2025). Integrating multi-temporal LiDAR and AHP for landslide detection and susceptibility mapping in Prince George's County, Maryland. AGU Fall Meeting 2025.
140. Aladeokin, O., Hare, O., Sheng, Z., & Owolabi, O. (2025). Integrating GIS-based susceptibility mapping and machine learning for landslide prediction and early warning in Baltimore County, Maryland. University Transportation Center Safety-21 Deployment Partners Consortium Symposium.
141. Wu, J., Lü, Q., Liao, Z., Yu, Y., & Wang, L. (2025). Dataset for high-locality landslides in Zhuji, China: Hydrological monitoring data, terrain features, and soil thickness [Data set]. Zenodo. https://doi.org/10.5281/zenodo.16748046
142. Realsee. (2025). Galois M2: Technical Specifications. https://home.realsee.ai/en/specs-galois
143. Ghoreishi, B., Staes, B. M., Liu, C., & Wang, H. (2025). Mapping climate extremes and infrastructure vulnerability: A new database for U.S. railroads. Scientific Data, 12(1), 1932. https://doi.org/10.1038/s41597-025-06205-z
144. Liang, J., Qi, W., & Xu, C. (2025). Evaluation of geological hazards susceptibility along a key railway based on machine learning. Scientific Reports, 15, 42497. https://doi.org/10.1038/s41598-025-26496-x
145. Zhu, Y., Qiu, H., Cui, P., Liu, Z., Ye, B., & Yang, D. (2023). Early detection of potential landslides along high-speed railway lines: A pressing issue. Earth Surface Processes and Landforms, 48(15), 3302–3314. https://doi.org/10.1002/esp.5697
146. Huang, J., Wu, X., Ling, S., Li, X., Wu, Y., Peng, L., & He, Z. (2022). A bibliometric and content analysis of research trends on GIS-based landslide susceptibility from 2001 to 2020. Environmental Science and Pollution Research International, 29(58), 86954–86993. https://doi.org/10.1007/s11356-022-23732-z
147. Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically based landslide susceptibility models. Earth-Science Reviews, 180, 60–91.
148. Mirus, B. B., Belair, G. M., Wood, N. J., Jones, J. M., & Martinez, S. N. (2020). Landslides across the USA: Occurrence, susceptibility, and data limitations. Landslides, 17(7), 2271–2285. https://doi.org/10.1007/s10346-020-01424-4
149. Palese, D., Nisi, V., & Santini, M. (2022). Computer-vision and AI techniques for rail-corridor landslide risk indices. Remote Sensing, 14(24), 6246. https://doi.org/10.3390/rs14246246
150. Paunescu, N., & Narcisa, M. (2000). The landslides risk assessment in areas with transport infrastructures. In Second Asia Pacific Conference & Exhibition on Transportation and the Environment, Beijing, China, People's Communications Publishing House. 1371-1381.
151. Zang, L., Wu, Y., & Zhang, X. (2020). Rainfall-induced vulnerability curves for railway infrastructure in China. Natural Hazards, 104(1), 123–142.
152. U.S. Geological Survey. (2025). Landslide inventories across the United

States (ver. 3.0, February 2025) [Data set]. https://doi.org/10.5066/P14AJF8I
153. He, Y., Bin, Z., Xu, X., Yu, H., Zhang, Y., Li, N., & Li, M. (2025). Landslide risk assessment along railway lines using multi-source data: A game theory-based integrated weighting approach for sustainable infrastructure planning. Sustainability, 17(12), 5522. https://doi.org/10.3390/su17125522
154. Freeborough, K. A., Diaz Doce, D., Lethbridge, R., Jessamy, G., Dashwood, C., Pennington, C., & Reeves, H. J. (2016). Landslide hazard assessment for national rail network. Procedia Engineering, 143, 689–696. https://doi.org/10.1016/j.proeng.2016.06.104
155. Ferrero, A., & Migliazza, M. (2012). Consolidation works and protection measures along Italian transport corridors. Engineering Geology, 127–128, 45–58.
156. Federal Railroad Administration. (2025). Maps – Geographic Information System. U.S. Department of Transportation. https://railroads.dot.gov/rail-network-development/maps-and-data/maps-geographic-information-system/maps-geographic
157. Jakob, M., & Lambert, S. (2009). Climate change effects on landslides along the southwest coast of British Columbia. Geomorphology, 107(3–4), 275–284. https://doi.org/10.1016/j.geomorph.2008.12.009
158. Ralph, F. M., Dettinger, M. D., Cairns, M. M., Galarneau, T. J., & Eymann, J. (2019). Defining "atmospheric river": How the Glossary of Meteorology helped resolve a debate. Bulletin of the American Meteorological Society, 100(5), 837–855. https://doi.org/10.1175/BAMS-D-18-0023.1
159. Freeborough, K., Dashwood, C., Diaz Doce, D., Jessamy, G., Brooks, S., Reeves, H., & Abbott, S. (2019). A national assessment of landslide hazard from Outside Party Slopes to the rail network of Great Britain. Quarterly Journal of Engineering Geology and Hydrogeology, 52(4), 345–360. https://doi.org/10.1144/qjegh2018-029
160. Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T. (2012). Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, 112, 42–66. https://doi.org/10.1016/j.earscirev.2012.02.001
161. Corominas, J., van Westen, C. J., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., ... Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. Bulletin of Engineering Geology and the Environment, 73(2), 209–263.
162. Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M. (2006). Estimating the quality of landslide susceptibility models. Geomorphology, 81(1–2), 166–184.

SUPPLEMENTARY MATERIAL

Appendix A: Research Products for this Project

A.1 Journal Articles

1. Okegbola, M.O., Owolabi, O., Sheng Z. & Liu Y. (2026). Integrating Multi-Temporal LiDAR and AHP for Landslide Detection and Susceptibility Mapping in Prince George's County, Maryland, Journal of Safety Science and Resilience, doi: https: //doi.org/10.1016/j.jnlssr.2026.100311
2. Okegbola, M. O., Owolabi, O., Sheng, Z., Liu, Y., & Etuke, J. O. (2026). Detection and Susceptibility Mapping of Landslide Using Multi-Temporal Lidar Digital Elevation Models: Case Study of Prince George's County, Maryland. Transportation Research Record: Journal of the Transportation Research Board, 0(0). doi: https://doi.org/10.1177/03611981261441298
3. Hosseinizadeh, A., Sheng, Z., & Liu, Y. 2025. High-Resolution Spatiotemporal Mapping of Surface Soil Moisture Using ConvLSTM Model and Sentinel-1 Data. Water, 17(22), 3300. https://doi.org/10.3390/w17223300.

A.2 Conference Publications

1. Adeluola, A.O., Pryor, J. and Owolabi, O. 2026. Developing a Decision-Support Framework for Cost-Effective Landslide Mitigation of Transportation Infrastructure, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
2. Akinniyi, O. Griffin, P., Mehmandari, T., Khalifa F., Sheng, Z. 2026. An InSAR–Deep-Learning Framework for Directional Forecasting of Slope Movement and Landslide Detection, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
3. Aliyu, S., Olowa, F., Darby, D., Liu, Y., Sheng, Z. 2026. Laboratory Characterization and Classification of Soils from Maryland Landslide Inventory Sites Using Moisture Content, Atterberg Limits, and Grain Size Distribution Tests, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
4. Hosseinizadeh. A., Sheng, Z. and Liu, Yi. 2026. Interpretable Machine Learning Approach for Developing Soil-Specific Probabilistic Rainfall Thresholds for Landslide Initiation in Maryland, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
5. Olowa, F., Sheng, Z. and Liu Y. 20226. Evaluation of Stream Scouring and Embankment Washout Susceptibility along Highways using Geotechnical Properties. 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
6. Fadipe, S., Liu, Y. 2026. Development of a TRIGRS-Informed Landslide Early Warning Index Using Machine Learning: A Pilot Study in Maryland, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
7. Paul, S., Sheng, Z., Liu, Y., 2026. Soil–Structure Interaction and Water Effects in

Retaining Wall Failure: A Comparative Review of Gravity, Cantilever, and MSE Systems, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].

8. Sultan, R., Sheng, Z., Nyarko, K. 2026. Physics-Informed Multi-Sensor Fusion and Consumer-Grade LiDAR Monitoring for Rainfall-Induced Landslide Early Warning System: A Prototype Framework Validated on an Independent Field Dataset, 2026 School of Engineering Summer Research Symposium, August 4, Baltimore, MD [Oral Presentation].
9. Farquharson, T., Sheng, Z., Lui, Y., Owolabi A.O., 2026. A Transferable ArcGIS Integration of WEPP and RUSLE Screening for Erosion Risk and Associated Landslides in Union Pacific Railroad Corridors: California Case Study, Proc. Of World Environmental and Water Resources Conference, 935-949, ASCE, Mobile, Alabama, April, 25-29. [Paper, Presentation].
10. Okebola, M., Owolabi A.O., Sheng, Z., Lui, Y. 2026, Detection and Susceptibility Mapping of Landslide Using Multi-Temporal LiDAR DEMs: A Case Study of Prince George's County, Maryland, The 105th Transportation Research Board annual meeting, Washington DC, January 11-15 [Paper, Presentation].
11. Sheng, Z., Y. Liu, O. Owolabi, K. Nyarko, T. Mehmandari, S. Kang, Hosseinizadeh, A., Okegbola, M., Aladeokin O., Aliyu, S., Farquharson, T., Sultan R. 2026. Improving highway safety by reducing the risks of landslides. The USDOT National Safety Summit of University Transportation Centers. DC, March 19 [Poster].
12. Hosseinizadeh, A., Z. Sheng, Y. Liu. 2025 Mapping Soil Moisture using a Machine Learning Method, AGU Fall Meeting 2025, New Orleans, LA, December 15-19 [Poster Presentation].
13. Kang, SK, Z. Sheng, Y. Liu, O. Owolabi. 2025. Numerical Simulation-Guided Machine Learning for Landslide Risk Assessment, AGU Fall Meeting 2025, New Orleans, LA, December 15-19 [Oral Presentation].
14. Okegbola, M., O. Owolabi, Y. Liu, Z. Sheng. 2025. Integrating Multi-Temporal LiDAR and AHP for Landslide Detection and Susceptibility Mapping in Prince George's County, Maryland, AGU Fall Meeting 2025, New Orleans, LA, December 15-19 [Poster Presentation].
15. Aladeokin, O., O. Hare, Z. Sheng, O. Owolabi. 2025. Integrating GIS-Based Susceptibility Mapping and Machine Learning for Landslide Prediction and Early Warning in Baltimore County, Maryland, University Transportation Center – Safety 21 Deployment Partners Consortium Symposium, November 13, Pittsburgh, PA [Poster Presentation].
16. Aliyu, S.K., A.U. Bah, Y. Liu, Z. Sheng. 2025. Laboratory Investigation of Soil Properties for Landslide Susceptibility Assessment in Montgomery County, Maryland, University Transportation Center –Safety 21 Deployment Partners Consortium Symposium, November 13, Pittsburgh, PA [Poster Presentation].
17. Farquharson, T., F. Thoronka, Z. Sheng, Y. Liu. 2025. A Nationwide ArcGIS Inventory Framework to Assess Rail Corridor Soil Erosion and Landslide Hazard, University Transportation Center –Safety 21 Deployment Partners Consortium Symposium, November 13, Pittsburgh, PA [Poster Presentation].
18. Hosseinizadeh, A., K. Collymore, Z. Sheng, Y. Liu. 2025. Mapping Soil Moisture using a Machine Learning Method to use in Landslide Study, University Transportation Center –Safety 21 Deployment Partners Consortium Symposium, November 13, Pittsburgh, PA [Poster Presentation].
19. Sultan, R., C. Mushambo, K. Nyarko, Z. Sheng. 2025. Early Warnings, Safer

Roads: A Machine Learning Approach to Landslide Risk Mitigation Using Sensor Network. University Transportation Center –Safety 21 Deployment Partners Consortium Symposium, November 13, Pittsburgh, PA [Poster Presentation].

20. Okegbola, M.O., C. Mincey, O. Owolabi., Z. Sheng, Y. Liu. 2025. Landslides Detection and Mapping with Remote Sensing Time-Series LiDAR Data, University Transportation Center –Safety 21 Deployment Partners Consortium Symposium, November 13, Pittsburgh, PA [Poster Presentation].

A.3 Datasets

A comprehensive CaseBank consisting of 4,500 numerical simulation cases was established by considering representative combinations of soil properties, slope geometry, and groundwater conditions. The resulting database enables engineers to directly retrieve previously simulated cases with similar site conditions while preserving the corresponding numerical models and failure-pattern information. To further improve computational efficiency, multiple machine learning algorithms were trained using the CaseBank, and XGBoost demonstrated the highest predictive performance for estimating the factor of safety.

Appendix B: Summer Internship Program and Symposium Activities

UTC Safety21 | Morgan State University | Summer 2026

B.1 Purpose and Scope

This appendix documents the educational, mentoring, research-communication, and dissemination activities associated with the Summer 2026 internship program supporting the UTC Safety21 project Improve Highway Safety by Reducing the Risks of Landslides. Seven student research groups prepared technical contributions that were integrated into Chapters 2-8 of this report. The supplementary pages record how the work progressed from team formation and research planning through technical investigation, chapter preparation, symposium presentation, and final-report integration.

Documentation boundary. The materials in this appendix provide programmatic context and a representative record of internship and symposium activities. They do not introduce additional technical findings, modify the analyses in Chapters 2-8, or establish operational warning thresholds.

B.2 Program Context

The summer program brought students, postdoctoral researchers, faculty members, and transportation-research collaborators together around a shared objective: improving transportation safety through better detection, characterization, modeling, and communication of landslide and slope-instability risks. The work combined geotechnical engineering, remote sensing, GIS, laboratory testing, numerical simulation, and artificial intelligence/machine learning.

Each group developed a focused technical study while participating in a common reporting process. Regular mentoring and quality review supported alignment of methods, figures, conclusions, and chapter structure. The program culminated in technical presentations and a poster session at the Morgan State University Summer Research Symposium, followed by consolidation of the contributed chapters into this Phase 3 final report.

B.3 Summer Internship Research Workflow

Figure A.1 summarizes the coordinated workflow used to move the summer projects from initial scoping to final dissemination. Although individual research tasks differed, the common sequence provided a consistent framework for mentoring, technical review, and preparation of report-ready products.

Figure B.1. Summer internship research, mentoring, presentation, and reporting workflow.

B.4 Summer Research Symposium Highlights

The UTC Safety21 Summer Research Symposium was held on August 5, 2026, at Morgan State University. Approximately 30 undergraduate and graduate students, postdoctoral researchers, faculty members, and transportation-research collaborators participated in the program. The event highlighted interdisciplinary approaches for earlier hazard detection, stronger infrastructure resilience, and safer transportation systems.

The program opened with remarks from Dr. Zhuping Sheng, principal investigator at Morgan State University. Co-principal investigators Dr. Yi Liu, Dr. Kofi Nyarko, and Dr. Oludare Owolabi welcomed participants and recognized the student and researcher contributions. Ms. Selena Distler, Safety21 Program Manager at Carnegie Mellon University and a certified Project Management Professional, presented an overview of the national Safety21 center and its collaborative research and workforce-development mission.

Two plenary sessions featured research on rainfall-triggered landslide thresholds; InSAR, LiDAR, and deep-learning approaches for slope monitoring; laboratory soil characterization; highway embankment washout; railroad landslide-exposure mapping; transportation decision-support systems; and retaining-wall performance. A poster session, networking lunch, and group photograph created additional opportunities for exchange among students, faculty members, and collaborators.

Figure B.2. Summer Research Symposium participants and UTC Safety21 research collaborators at Morgan State University, August 5, 2026.

B.5 Newsletter Feature

The August 2026 newsletter feature presented the symposium as a research-to-action activity that connected student development, technical progress, and transportation-safety applications. The original feature is reproduced below as a program record; the photographs on the following pages provide a curated visual summary of opening remarks, plenary presentations, and the poster session.

SAFETY21 NEWS CARNEGIE MELLON UNIVERSITY

SUMMER RESEARCH SPOTLIGHT | AUGUST 2026

Morgan State Summer Symposium Highlights Safety21 Landslide Research

Students and researchers showcased interdisciplinary tools for earlier hazard detection, stronger infrastructure resilience, and safer transportation systems.

The UTC Safety21 Summer Research Symposium, held August 5, 2026, at Morgan State University, brought together 30 undergraduate and graduate students, postdoctoral researchers, faculty, and transportation research collaborators to showcase progress in landslide-risk reduction and infrastructure safety.

PROJECT FOCUS Improve Highway Safety by Reducing the Risks of Landslides - a multi-phase initiative integrating geotechnical engineering, remote sensing, GIS, and artificial intelligence.

The program opened with remarks from Dr. Zhuping Sheng, principal investigator at Morgan, together with co-principal investigators Dr. Yi Liu, Dr. Kofi Nyarko and Dr. Oludare Owolabi also welcomed participants and recognized the students and researchers whose summer work contributed to the UTC program. Ms. Selena Distler, Safety21 Program Manager at Carnegie Mellon University and a certified Project Management Professional, provided an informative overview of the national center and its collaborative network, connecting Morgan State's research with Safety21's broader transportation-safety mission.

Two plenary sessions featured research on rainfall-triggered landslide thresholds, InSAR and deep-learning approaches for slope monitoring, laboratory soil characterization, highway embankment washout, LiDAR-based early warning, railroad landslide-exposure mapping, transportation decision-support systems, and retaining-wall performance. A poster session, networking lunch, and group photograph created additional opportunities for students, faculty, and collaborators to exchange ideas.

FROM RESEARCH TO ACTION: Closing remarks recognized the participants' accomplishments and emphasized translating interdisciplinary research into practical information for transportation agencies and communities.

Besides USDOT support, the summer research is also in part sponsored by Morgan State University AI/ML Center, Transform Morgan Program, School of Engineering, National Science Foundation CREST-DPSI(S) program and U.S. Department of Education’s Elevate to R1 grant.

Link to access the report’s high quality photos
Link to view additional photos

Safety21 | USDOT National University Transportation Center for Safety safety21.cmu.edu

Figure B.3. Reproduction of the August 2026 UTC Safety21 Summer Symposium newsletter feature.

B.6 Photo Documentation

The selected photographs below document program orientation, technical presentations, and interdisciplinary discussion. Captions identify the activity shown without attributing individual technical claims beyond the accompanying newsletter record.

Figure B.4. The PI and Co-PI delivering opening-session remarks emphasizing collaboration, research development, and transportation safety outcomes.

Figure B.5.Co-PI delivering opening-session remarks emphasizing collaboration, research development, and transportation safety outcomes.

Figure B.6. Safety21 CMU Program Manager presenting an overview of the UTC Safety21 program, including its transportation safety mission, collaborative structure, and student research and workforce development opportunities.

Figure B.7.Postdoctoral scholar's report on summer internship mentoring and technical presentations on remote sensing and hybrid machine-learning methods for landslide monitoring and prediction.

Figure B.8. Presentation of soil-specific probabilistic rainfall thresholds developed for landslide-initiation assessment in Maryland using rainfall data and machine-learning methods.

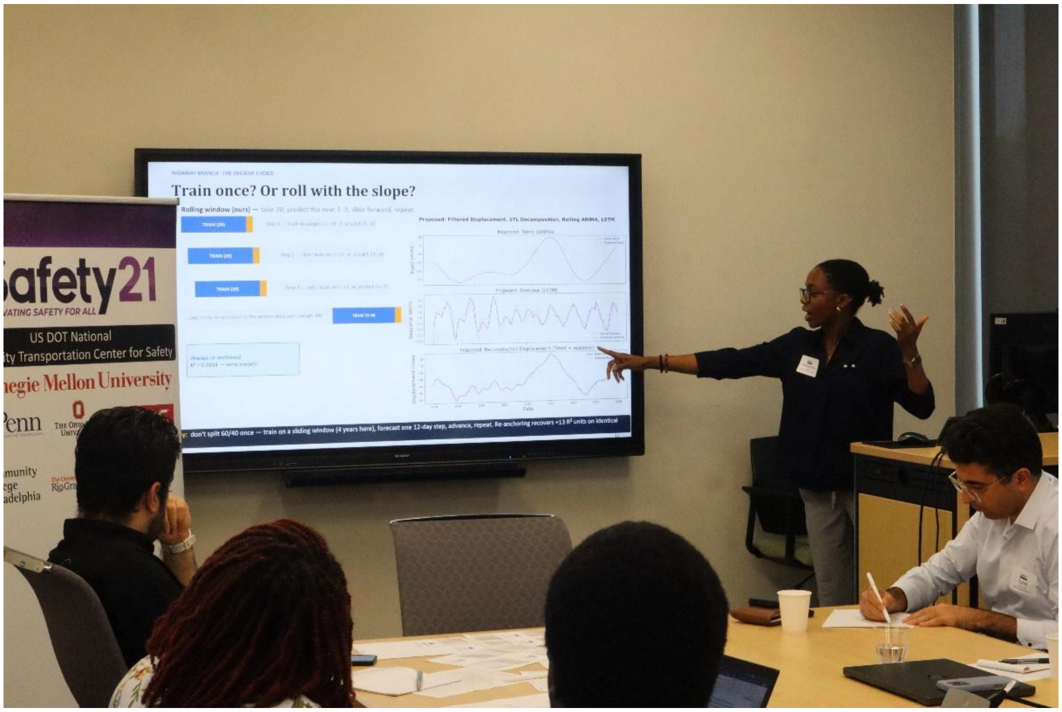


Figure B.9. Student researcher presenting machine-learning methods for railroad landslide monitoring, risk assessment, and early prediction.

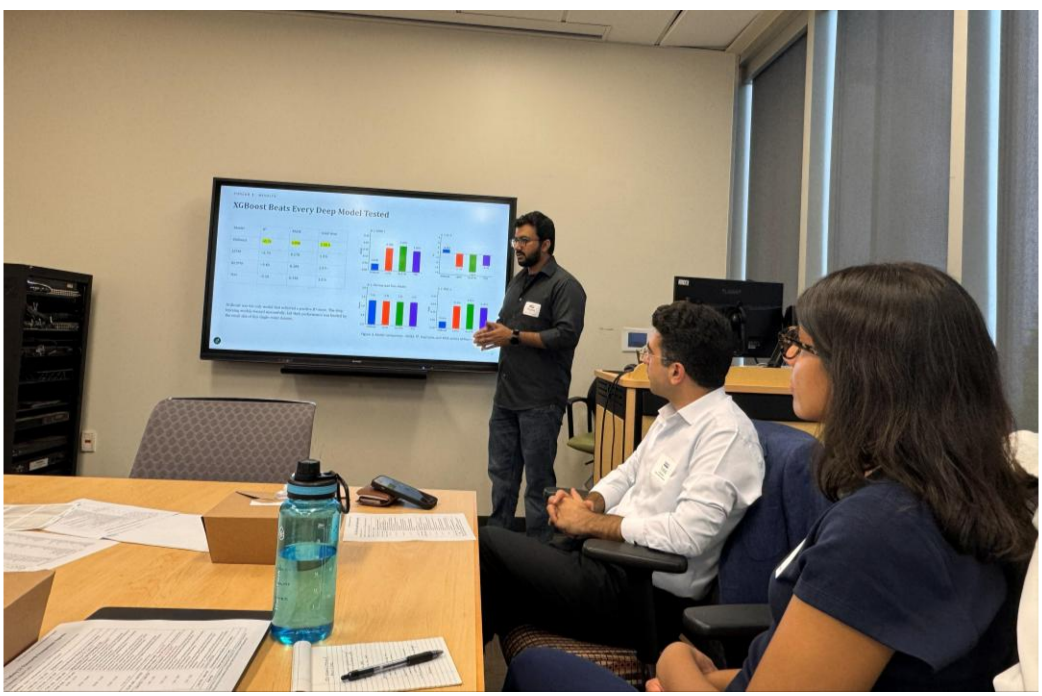


Figure B.10. Student researcher presenting the application of LiDAR imagery and geospatial analysis for landslide identification, monitoring, and risk assessment.

Figure B.11.Student researcher presenting the development of a national railroad landslide exposure inventory for hazard identification and transportation risk assessment.

Figure B.12. Student researcher presenting laboratory soil analyses, including liquid-limit and plastic-limit testing, for the geotechnical characterization of landslide-prone soils.

Figure B.13. Student researcher presenting an integrated cost-analysis framework for sustainable landslide management using life-cycle cost analysis, benefit–cost analysis, and multi-criteria decision analysis.

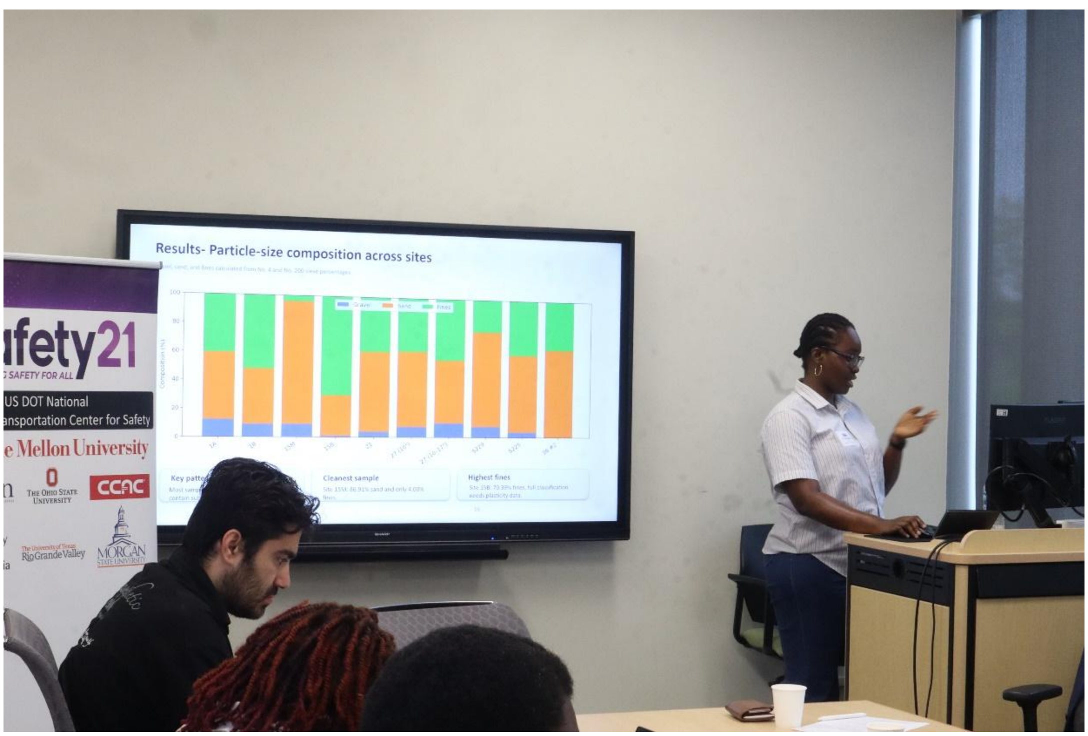


Figure B.14.Student researcher presenting the assessment of railroad washout hazards using hydrologic, geospatial, and infrastructure-risk analyses.

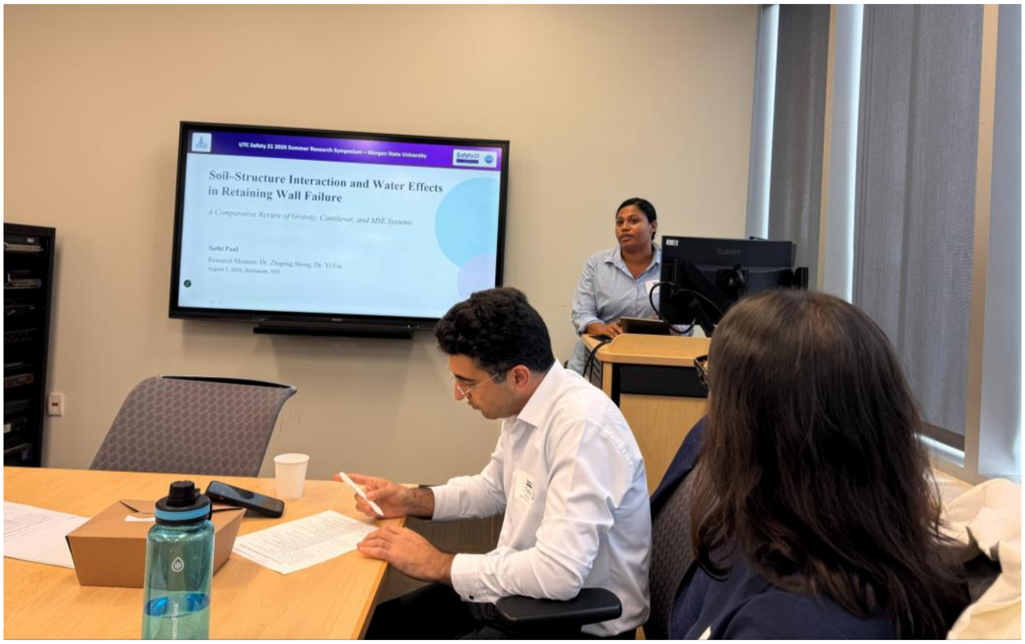


Figure B.15. Technical presentation on soil-structure interaction and water effects in retaining-wall performance.

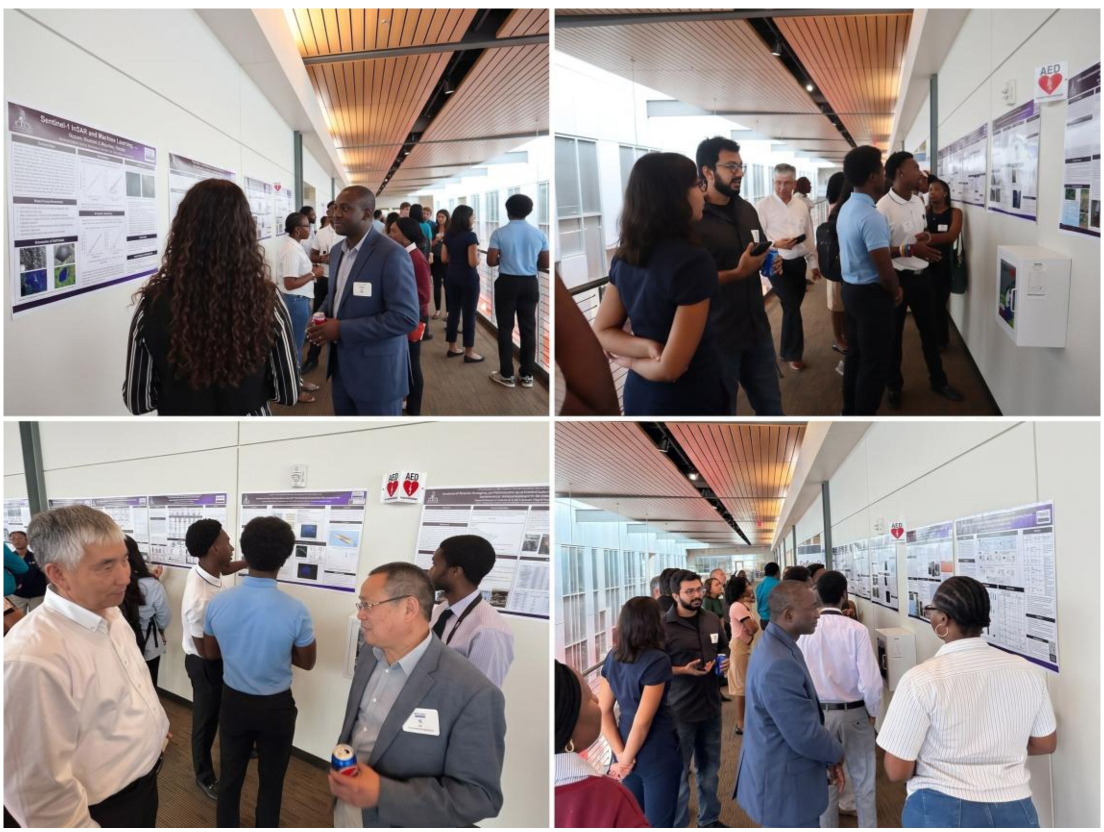


Figure B.16. Poster-session and networking activities supporting interdisciplinary knowledge exchange, research communication, and dissemination of the Summer 2026 internship outcomes.

B.7 Program Support and Dissemination

In addition to support from the U.S. Department of Transportation UTC Safety21 program, the summer research activities were supported in part by the Morgan State University AI/ML Center, Transform Morgan program, School of Engineering, National Science Foundation CREST-DPSI(S) program, and the U.S. Department of Education Elevate to R1 grant. The combined technical chapters, symposium presentations, posters, newsletter, and photographic record document both the research contributions and the student-development outcomes of the Summer 2026 program.