CIAMTIS

U.S. DOT Region 3 University Transportation Center

Landslide Risk Assessment in Cut Locations Using Artificial Intelligence Based on Right-of-Way Videos and Geophysical Data

June 30, 2022

Prepared by:
**M.E. Palese, Univ. of Delaware; T. Pei, Penn State Univ.;
T. Qiu, Penn State Univ.; C. Shen, Penn State Univ.;
A.M. Zarembski, Univ. of Delaware;
J.W. Palese, Univ. of Delaware**

r3utc.psu.edu

PennState College of Engineering

LARSON TRANSPORTATION INSTITUTE

Technical Report Documentation Page

1. Report No. CIAM-UTC-REG22	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Landslide Risk Assessment in Cut Locations Using Artificial Intelligence Based on Right-of-Way Videos and Geophysical Data		5. Report Date June 30, 2022	
		6. Performing Organization Code	
7. Author(s) Michael E. Palese, https://0000-0002-7685-3523; Te Pei, https://0000-0002-2154-8505; Tong Qiu, https://0000-0003-2516-6851; Chaopeng Shen, https://0000-0002-0685-1901; Allan M. Zarembski, https://0000-0002-4282-9330; and Joseph W. Palese, https://0000-0003-3946-3777		8. Performing Organization Report No.	
9. Performing Organization Name and Address Department of Civil and Environmental Engineering University of Delaware DuPont Hall, 127 The Green Newark, DE 19716		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No. 69A3551847103	
12. Sponsoring Agency Name and Address U.S. Department of Transportation Research and Innovative Technology Administration 3rd Fl, East Bldg E33-461 1200 New Jersey Ave, SE Washington, DC 20590		13. Type of Report and Period Covered Final Report 07/01/2020 – 06/30/2022	
		14. Sponsoring Agency Code	

16. Abstract

Sidehill and through cuts are often used in the construction of new railroad rights-of-way to limit the length, curvature, and grade of the route. However, rights-of-way that utilize cuts are susceptible to damage from falling debris driven by slope failure events such as shallow landslides and rockfalls. At-risk slopes, or geohazards, are traditionally analyzed using intensive field investigations and historical failure events to determine their likelihood of failure and the potential consequences of failure. Anticipating slope failures that may occur due to everyday weather events and other catalysts in the region helps protect railroad assets and employees, ensuring safe operations. Many rights-of-way have a large density of geohazards; thus, performing in-situ measurements to determine their failure likelihood requires extensive resources. In addition, installing infrastructure to detect or inhibit debris flow is expensive and often unrealistic for all geohazards. This study aimed to create a new slope stability risk framework for railroad cut sections by processing digital images of railroad rights-of-way recorded by inspection vehicles and related geophysical data. A geohazard-affected track section along the Harrisburg Line was used as the study area. Computer vision techniques were used to identify and quantify geohazard features that indicated slope instability. An object detection model based on deep learning (DL) was trained to detect these slope instability indicators and generate risk scores from rights-of-way inspection videos. Moreover, a landslide inventory was compiled, and a landslide susceptibility model was developed for the study area based on available geophysical data. The object detection model and the landslide susceptibility model were combined using a relative risk assessment framework to determine which sections were most at-risk of landslide, and results were compared with the railroad identified geohazard sections across the study area.

17. Key Words Railroad, slope failure, landslide, debris, slope instability, geohazard, risk assessment, sidehill, through cuts		18. Distribution Statement No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 72	22. Price

Form DOT F 1700.7   **(8-72)**   **Reproduction of completed page authorized**

DISCLAIMER

ADDITIONAL AUTHOR INFORMATION

Michael E. Palese, Ph.D. Candidate, Univ. of Delaware; Te Pei, Ph.D. Candidate, Penn State Univ.; Tong Qiu, Ph.D., P.E., Professor of Civil Engineering, Penn State Univ.; Chaopeng Shen, Ph.D., Associate Professor of Environmental Engineering, Penn State Univ.; Allan M. Zarembski, Ph.D., P.E., Professor of Civil Engineering, Univ. of Delaware; and Joseph W. Palese, M.C.E., P.E., Senior Scientist, Univ. of Delaware

TABLE OF CONTENTS

CIAMTIS
U.S. DOT Region 3 University Transportation Center

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

BACKGROUND

Roads and railroad rights-of-way are typically constructed with minimal inclines, declines, and curves to ensure safe and comfortable route traversal. While motor vehicles can handle steeper grades under ideal conditions, trains are much heavier and must reduce speed to traverse steep grades. It is atypical for freight trains to have a ruling grade steeper than 2% (i.e., the steepest gradient a train can climb) based on the motive power and train weight, and a ruling grade steeper than 3% for high-speed passenger trains. Since track systems must also minimize the use of tight curves to reduce large centrifugal forces on trains, mountainous and hilly terrain present unique construction challenges for new track systems. Building around elevated terrain usually requires train speed reduction due to the presence of tight curves. To avoid the use of tight curves, railroads often opt to use cuts or tunnels to create routes through elevated terrain, the benefits of which depend on the cost of the infrastructure versus the benefits to operational efficiency and safety of the alternate route. Since trains must reduce speed to navigate abrupt directional changes, track operations are made most efficient when a well-maintained tangent, level track is used.

PROBLEM STATEMENT

A cut in railroad engineering is an area where soil or rock is removed to construct a new right-of-way (ROW), often reducing the length, curvature, and grade of the route. Consequently, the use of cuts can introduce slopes whose base lies in the right-of-way and which vary in height and steepness depending on how much material was removed. Figure 1.1 shows an example of a through-cut where the right-of-way is cut such that geohazards (i.e., geological sites with features that may lead to widespread damage or risk) on both sides slope toward the track. A sidehill cut indicates a cut where only one side has a geohazard that slopes toward the track. Cuts in rights-of-way pose the risk of slope failures dumping debris onto the track system.

Although the study area used for this research effort only contains cuts with relatively short slopes near the railroad, slope failure events have occurred, dumping debris on the track and halting operations. Figure 1.2 shows a photograph of debris obscuring the track after a slope failure event occurred due to heavy rainfall. Landslides are typically major soil movements that are detectable from satellites due to their size and fresh soil exposure. Although a failure event like the one pictured would not be detectable in the same way, it was severe enough to halt operations and raise concern that track infrastructure was damaged.

Figure 1.1. Example of a through-cut.

Figure 1.2. A slope failure event near MP 40.5 occurred on 6/8/2021.

SCOPE AND OBJECTIVES

The research detailed in this report introduces a process to analyze geohazards present in ROW videos for relative failure likelihood by analyzing geohazards found adjacent to the railroad. The proposed algorithm includes only publicly available satellite images, historical landslide locations, and ROW recordings as source data. The geohazards shown in the videos were analyzed by

identifying and localizing slope features that indicate slope instability using an object detection algorithm. Additionally, important geographical and geological parameters were quantified using geospatial images of the study area. Several variables were calculated or extracted that summarize the geohazard's current and past condition. Each geohazard's relative health was investigated using a relative risk assessment matrix based on these values. The proposed algorithm serves as a preliminary slope stability analysis that informs engineers which geohazards are of primary concern in this study area.

REPORT ORGANIZATION

This report is organized into six chapters. A brief outline of each chapter is provided as follows:

 Chapter 1 outlines the background of this research and identifies the need for a more comprehensive risk assessment framework for slope failures along the railway cut sections. The main objectives of this report and the research program required to accomplish these objectives are briefly described.

 Chapter 2 details the literature review of slope failure causes, effects, and methods of the study performed to prepare for this research activity. Quantitative techniques commonly used to perform slope assessment methods are also discussed.

 Chapter 3 discusses all available datasets for this research effort and how they were prepared for analysis. This includes detailing what right-of-way images were available, what landslide inventory was used based on the location of the study route, and the origin of used geospatial parameters.

 Chapter 4 outlines all analytical methods applied to study the available data detailed in Chapter 3. Analysis techniques are covered generally and with respect to their application in this research effort.

 Chapter 5 presents all results obtained using the analysis techniques explored in Chapter 4. This includes testing a successfully trained object detector and validating a landslide prediction model. Furthermore, efforts to combine the results obtained by studying the right-of-way images and geospatial parameters to perform a relative risk assessment are discussed.

 Chapter 6 concludes the project report by summarizing the research detailed herein and discussing future research directions.

CHAPTER 2

Literature Review

INTRODUCTION

This chapter covers the scope of the literature review performed for this research activity. This included reviewing literature investigating the causes of slope failure events and their effects on railroad infrastructure and operational safety and efficiency. Furthermore, proposed methods of studying at-risk slopes using right-of-way recording and geophysical methods were reviewed. This included a review of landslide susceptibility modeling techniques and conventional machine learning (ML) methods.

SLOPE FAILURE CAUSES AND EFFECTS

When using cuts in railroad track systems, methods of mitigating infrastructure damage caused by slope failure events must be considered. Slope failures have caused an average of $3M in reportable damages per year over the past ten years to railroad infrastructure, as determined by the number of type-M101 accidents identified in the Federal Railroad Administration (FRA) safety database (Federal Railroad Administration, 2002). The actual cost of damage associated with these accidents is likely higher if cleanup costs and costs associated with fixing the right-of-way are included. Studies by the Association of American Railroads (AAR) suggest that the actual cost of an FRA-reported derailment is approximately double the reported amount. Geohazards must be periodically analyzed to determine if they are at a high risk of failure and what potential damages to infrastructure could be sustained from their failure to reduce the frequency of train accidents due to geohazard failures. Once at-risk geohazard sections are determined, mitigation efforts can be performed. Although it is often difficult to stop shallow, rapid landslides or rockslides from occurring, methods such as installing piles or retaining walls, changing the material composition of the slope, or modifying slope geometry to reinforce at-risk slopes will typically stop large debris flows from occurring (e.g., Chen et al., 2016; Ai et al., 2021). More often, railroads will opt to install structures to monitor at-risk slopes for failure events, which are significantly cheaper. Slide fences, for example, are fence structures built in front of geohazards that are wired into the railway signaling system to notify the railroad if a rockslide event has breached the fence. Thus, trains will not unknowingly pass over track sections covered in debris.

Several methods have been developed to determine where slope failure prevention and mitigation efforts should be prioritized. Soil geohazards are primarily assessed for risk by making geotechnical measurements to determine the slope's material composition, geometry, and internal stress conditions. Physics-based models have been developed to use geotechnical measurements to determine failure surfaces and the factor of safety against failure. One commonly used method of slope stability analysis, limit

equilibrium analysis, compares forces causing instability with forces resisting movement in a soil mass tending to slide toward the ground under the influence of gravity (Duncan, 1996; Hatheway, 1996). The method of slices is employed in many slope stability analysis tools and ranges from simplified to detailed analysis. The method divides the slope into several vertical slices whose free body diagrams are analyzed individually. Comparative analyses of the various methods have shown that the calculated safety factors vary little (Duncan and Wright, 1980).

Once a factor of safety is calculated for a geohazard section, further analysis can be performed to determine the likelihood of a failure event based on known or likely catalysts. Since most shallow landslide events are triggered by weather events, analyses often consider precipitation and earthquake events of varying severity. Rainfall threshold models, for example, consider a geohazard's composition and its risk for landslides based on the amount of rainfall in a given period (Pradhan et al., 2019; Wu et al., 2015). However, rockfalls and rockslides are governed by different models, since they fail under different mechanisms (Agliardi and Crosta, 2003; Alvioli et al., 2021; Dorren, 2003). The use of these models is challenging because an accurate factor of safety is not obtainable. Thus, on-site measurement of geohazards is often a crucial step in slope stability analysis such that a satisfactory factor of safety can be used to predict failure likelihood.

SLOPE ASSESSMENT USING RIGHT-OF-WAY RECORDINGS

This research was conducted to introduce a new method of assessing slope conditions to aid in determining where on-site measurement and remediation efforts should be prioritized. Specifically, videos of the right-of-way recorded by inspection vehicles were used to find indicators of geohazard instability using computer vision techniques. Inspection vehicle ROW videos are typically utilized for inspection-based assessment of the track superstructure, while sophisticated measurement systems collect data on the track's geometry. The videos were used to collect further numerical information about the track system using digital image processing techniques. While many inspection vehicles do not have high-resolution cameras installed, research beginning in the 1990s showed that ROW recordings are useful beyond inspective monitoring purposes (Velten et al., 1999; Resendiz et al., 2013). For example, artificial intelligence algorithms have been created to identify objects obstructing the right-of-way in real-time such as debris or trespassers (Zhang et al., 2018; Zaman et al., 2019; Boussik et al., 2021). Computer vision techniques have also been developed to monitor the health of specific track components (Guo et al., 2021a and b).

SLOPE ASSESSMENT USING GEOPHYSICAL DATA

Landslide Susceptibility Modeling

A slope failure or landslide is usually considered the movement of a mass of rock, debris, or earth down a slope. It occurs when the shear stress exerted by gravity or other forces exceeds the shear strength of slope materials. 2.4.1 Landslide Susceptibility Modeling (LSM) is often carried out in geoscience applications to produce a landslide probability map on a regional level. LSM predicts landslides based on existing landslide inventory. Statistical analyses are subsequently carried out to discover the hillslope condition at landslide locations. Areas with similar topographical patterns are determined to be landslide hazard areas.

LSM answers the question of where the landslides would occur (Guzzetti et al., 2005). The final product of LSM is a series of maps showing landslide probability at any geographical location in the study area. Methods used to determine areas susceptible to landslides can be broadly classified into physics-based and statistically based methods. Among these two categories, the physics-based methods rely on simplified soil mechanics models to determine the stability condition of a given slope through limit equilibrium analysis. The statistically based methods establish functional relationships between landslide-contributing factors and past/present landslide locations. For LSM, a few assumptions need to be satisfied despite the difference between these two methods: (1) landslides leave noticeable signs on the earth's surface that can be mapped into landslide inventories; (2) the occurrence of landslides obeys physical laws that can be evaluated both empirically and deterministically; and (3) conditions that caused past slope instability are likely to contribute to future landslides (Reichenbach et al., 2018). The complex nature of landslides has limited the application of physics-based approaches for LSM. So far, research attention has focused on conducting LSM through empirical methods worldwide (i.e., statistically based methods). More than a half-thousand research articles have been published on this topic in the past four decades (Reichenbach et al. 2018).

Workflow of Statistically-Based LSM

A comprehensive review by Reichenbach et al. (2018) identified six critical steps required for any statistically based LSMs. A comprehensive chart describing the workflow of LSM can be found in Figure 2.1. The steps are briefly summarized as follows:

1. Organizing landslide information (i.e., response variable): this step compiles a landslide inventory for the study area. The landslide inventory should contain the spatial (i.e., location of landslides) and temporal (i.e., date of the event) information. It is also essential to include the failure mode of individual landslides and their triggers.
2. Gathering thematic information (i.e., explanatory variable): this step identifies contributing factors that lead to potential slope instabilities, such as elevation, vegetation coverage, soil moisture, soil topography, and bedrock depth.
3. Selecting appropriate mapping units: this step determines mapping techniques suitable for the input and output data, such as grid-based analysis or slope-based analysis.
4. Determining appropriate statistical models: this step determines suitable models to establish functional relationships between input and output variables. When selecting models, one should consider the type of landslide inventories, the format of thematic information, and the extent of the study area.
5. Evaluating model performance: this step evaluates the robustness of the fitted model in terms of training accuracy, validation accuracy, and model uncertainties.
6. Model deployment: this step applies the landslide probability map to landslide protocols, which benefits hazard prediction, mitigations, and land planning.
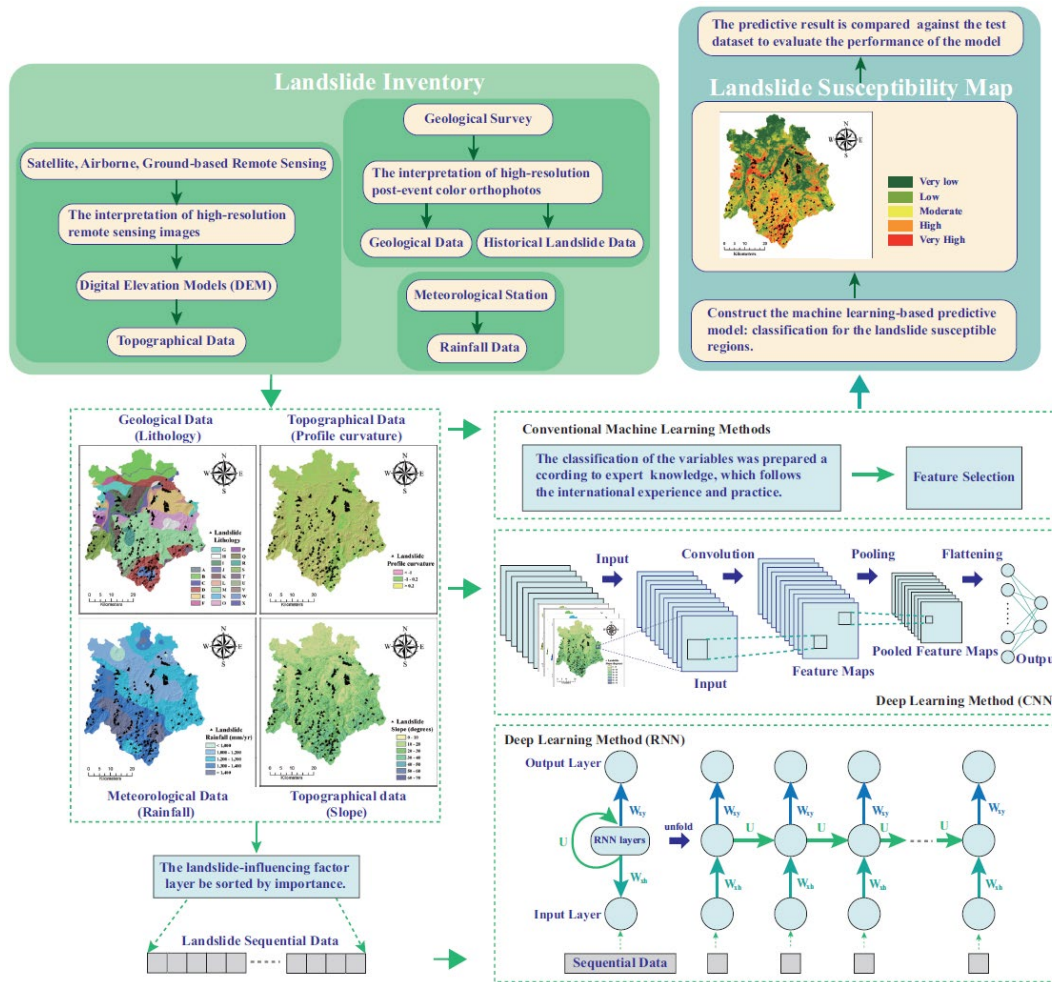
Figure 2.1. Summary of workflow and modeling framework for LSM (from Ma et al., 2021).

APPLICATION OF ML IN LSM

Analysis of landslide susceptibility can be considered a classification problem, where features of input variables (e.g., landslide contributing factors) are extracted and grouped based on their relationships and contributions to the response variable (i.e., landslide/non-landslide). Unlike traditional statistical methods (e.g., logistic regression), ML algorithms stand out in this type of application due to their excellent data clustering and classification capability. Recent advances in LSM with the application of ML methods are briefly summarized in this section.

Conventional ML Methods

The support vector machine (SVM) is a type of supervised ML method that has been widely adopted in various classification and regression analyses. The objective of the SVM algorithm is to find a hyperplane in N-dimensional space (e.g., groups of landslide contributing factors) that distinctly classifies the data points (i.e., landslides/non-landslides). For example, Lee et al. (2017) conducted an LSM for rainfall-induced landslides using SVM. In their study, 18 causative factors related to topography, geology,

soil, forest, and land use were first stacked into layers of 2D arrays and subsequently included in their model for training and validation. The results from two selected regions show that the developed SVM model is able to achieve a prediction accuracy of 81.36% for the training set and 77.49% for the test set.

K-means is a method of vector quantization that aims to partition N observations into K clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or cluster centroid), serving as a prototype of the cluster. Wang et al. (2017) conducted an LSM for the Three Gorges area in China using logistic regression with causative factors for landslides classified based on information theory (i.e., Shannon's entropy). The K-means clustering is subsequently applied to enhance prediction accuracy. Their developed model can boost the overall accuracy to 91.76%.

Decision tree (DT) is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g., a different range of slopes or different elevation intervals), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths from the root to the leaf represent classification rules. DT has been considered one of the most popular supervised classification algorithms due to its simplicity and interpretability. Tsangaratos and Ilia (2015) conducted LSM in Xanthi, Greece, using a modified DT classifier. A total of eight landslide contribution factors were analyzed and included in their study. Their trained model can achieve an overall accuracy of 96%.

DL Methods

DL methods are a subset of ML using algorithms inspired by the structure and function of the brain (i.e., artificial neural networks). The past decade has witnessed the rapid development of DL algorithms. DL methods have proven to be extremely powerful in dealing with tasks like image recognition and natural language processing. Recently, state-of-the-art DL methods have been applied to the area of LSM. DL algorithms consist of layered structures; each layer digests input quantities and transforms the data to increase the selectivity and the invariance of the representation (i.e., they gradually learn features from the input layer by layer). As a result, DL models are able to extract optimal features and establish indirect relationships between input and output.

Convolutional neural network (CNN) is a typical supervised DL method. It can extract and classify features from high-dimensional data by hierarchically composing simple local features into complex models; thus, CNN is ideal for learning features from raster data (Ma et al., 2021). Wang et al. (2019) evaluated the performance of different CNN architectures for LSM. In their study, landslide-contributing factors were gathered into raster formats. Techniques were subsequently used to convert these data into different representations: namely, 1D vector, 2D array, and 3D matrix. Three different CNN models were constructed based on the format of different data representations: 1D-CNN, 2D-CNN, and 3D-CNN. Among these different CNN architectures, 2D-CNN achieved the best prediction accuracy. The unique characteristics of different CNN architectures are summarized as follows:

- For 1D-CNN, it could exploit the local correction and gradually learn more intricate representations from the factor vectors.
- For 2D-CNN, it was able to extract the valuable hidden features sufficiently.
- For 3D-CNN, it not only learned factor representations but also extracted local spatial information.

Besides CNN, as a unique type of recurrent neural network (RNN), the long short-term memory (LSTM) network inherits RNN's capability of sequence learning. It is able to learn time-series data with long temporal dependency and automatically determines the optimal result by applying the gate control mechanism (Xiao et al., 2018). The gate control mechanism allows the network to only pass learned rules to the next time-step from the previous time step. Therefore, the LSTM networks are robust to noises and errors and can handle long time-series data. Xiao et al. (2018) presented an LSM framework that first extracted the earth's surface feature from high-resolution remote sensing images using CNN and combined it with other remotely sensed data. The classified features from CNN were subsequently combined with LSTM. Their proposed framework is able to handle dynamic evolution characteristics of mountains and landslide displacement, thus improving the prediction accuracy. Their results indicate that the developed CNN-LSTM model can outperform traditional ML techniques (e.g., SVM, DT) and assess landslide probability dynamically.

CHAPTER 3

Data Availability

INTRODUCTION

This research project aimed to assess a collection of geohazards in a selected railroad right-of-way using publicly available data and videos from track inspection runs. This chapter discusses the chosen study area, the scope of the available right-of-way recordings, and the publicly available data that had to be obtained and analyzed to complete this research task. This chapter exhaustively covers all data used in this research project.

DESCRIPTION OF STUDY AREA

A large population of geohazards exists in the Harrisburg railway line from milepost (MP) 83 (eastern end) to 95 (western end), which was the focus area of this study. Figure 3.1 shows a map of the studied track segment (in orange) illustrating the high density of geohazards (in blue). Table 3.1 lists the geohazards shown in Figure 3.1 by the index of the cut, starting milepost, ending milepost, and centerline milepost (S MP, E MP, and CL MP), type (R for rock slopes and S for soil slopes), side (B for both sides of the track and S for the south side of the track), height, and length. An additional map was created to demonstrate the large density of through-cuts in the study area. Figure 3.2 shows a slope map of the study area where the color of each pixel indicates the pitch of the slope (darker pixels indicate steeper slopes). Due to the large population of geohazards, the track section used in the present study is clearly visible on the map without being specifically indicated. The density of geohazards makes this track section an ideal area of study for assessing at-risk geohazards.

Figure 3.1. Map of Harrisburg Line (orange) with geohazard sections indicated (blue).
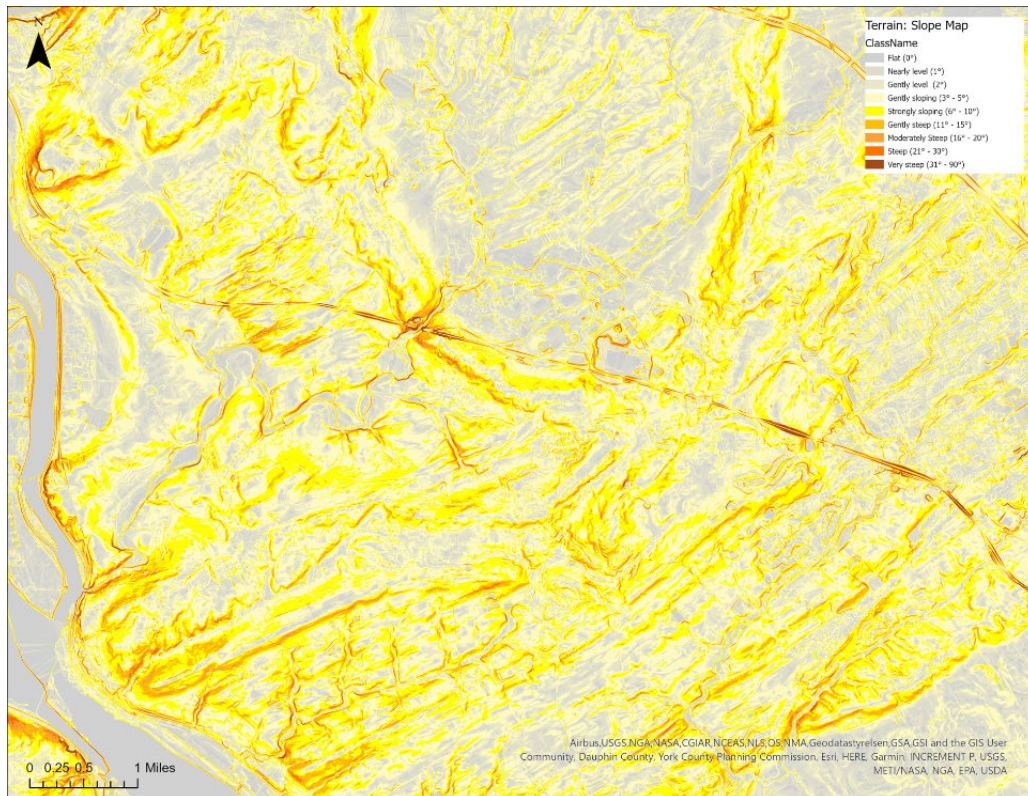
Figure 3.2. Slope map of the Harrisburg Line.

Table 3.1. Railroad identified geohazard sections.

Index	S MP	E MP	CL MP	Type	Side	Height (ft)	Length (ft)
1	84.80	85.26	85.03	R	B	40	2,430
2	85.66	85.98	85.82	R	B	70	1,690
3	87.00	87.37	87.19	R	B	40	1,950
4	87.63	87.75	87.69	R	S	20	630
5	87.90	87.95	87.93	R	B	20	260
6	88.17	88.44	88.31	R	B	20	1,430
7	89.54	90.14	89.84	S	B	20	630
8	90.30	90.35	90.33	R	B	20	260
9	90.79	91.02	90.91	S	B	20	1,210
10	91.24	91.56	91.40	S	B	30	1,690
11	93.03	93.46	93.25	R	B	40	2,270

DATA AVAILABILITY FOR OBJECTION DETECTION

Right-of-way videos were recorded in approximately 6-month intervals by an inspection vehicle on a mainline track segment in Pennsylvania. The geohazard inventory for the Harrisburg Line was primarily used to identify which ROW images show a geohazard, but most of the information can be recovered using digital image processing techniques. The available recordings were taken by a camera on the end of a track geometry inspection vehicle. The recording produced digital red-green-blue (RGB) frames with 480x720-pixel resolution. RGB digital images are colored using the RGB color model. Each color frame has integer pixel values ranging from 0 to 255, creating 256^3 unique colors when stacking the frames. Thirty frames were captured per second, meaning a new frame was recorded every 3 to 5 ft based on the speed of the inspection vehicle. To guarantee unique scenes in the image frames and to limit the size of the dataset, two frames were extracted per second. Figure 3.3 shows an example image of a geohazard section extracted from a ROW video. The information box at the top of the image details the location and speed of the inspection vehicle throughout its run. The current milepost of the vehicle was used to identify geohazards per Table 3.1 as the train passes them. Table 3.2 summarizes the available ROW videos by recording date, traffic direction, and track traversed upon.
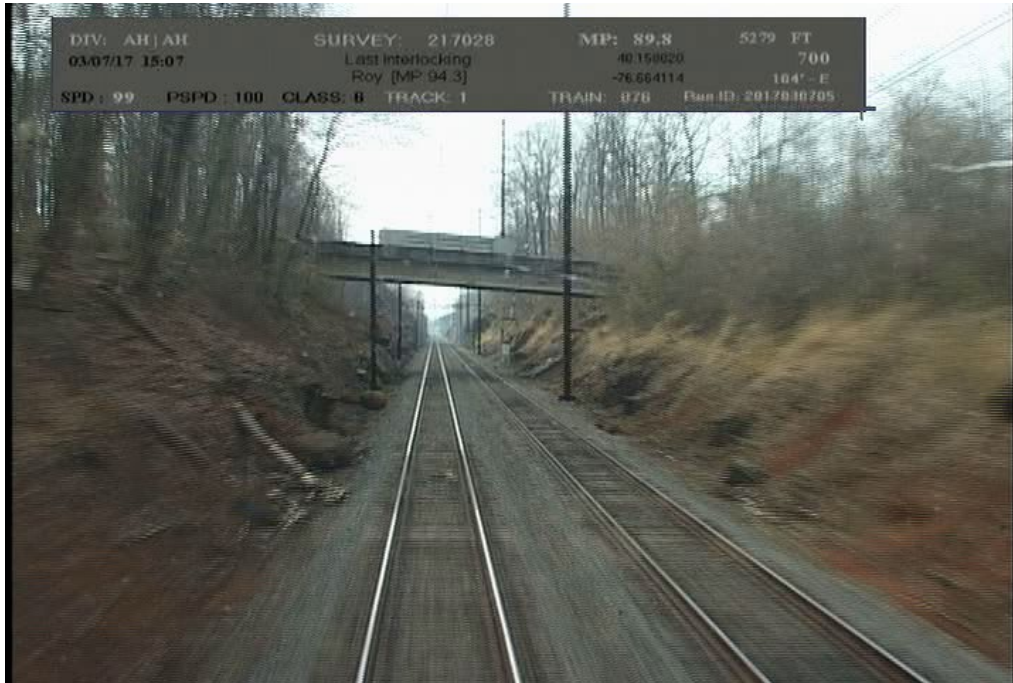
Figure 3.3. Example ROW image.

Table 3.2. Available right-of-way recordings.

Survey ID	Recording Date	Start MP	End MP	Traffic	Track
Rheems-Roy-217028	03/07/2017	94.6	83	East	North
Rheems-Roy-217191	12/12/2017	83.1	94.3	West	South
Rheems-Roy-218075	05/22/2018	94.5	83.8	East	North
Rheems-Roy-218223	12/04/2018	94.4	82.8	East	North
Rheems-Roy-219166	09/10/2019	84.5	94.4	West	North
Rheems-Roy-219259	12/17/2019	83.3	104.7	West	South
Rheems-Roy-220014	02/11/2020	83.5	104.6	West	South
Rheems-Roy-220194	12/01/2020	82.6	94.3	West	South
Rheems-Roy-221048	02/17/2021	83.3	94.4	West	South
Rheems-Roy-221048-TK4	02/17/2021	39.6	41	West	North
Rheems-Roy-221048-TK1	02/17/2021	41.8	39.8	East	South
Rheems-Roy-221064-TK4	03/30/2021	39.4	40.9	West	North
Rheems-Roy-221064-TK1	03/30/2021	41.5	39.7	East	South
Rheems-Roy-221084-TK4	06/03/2021	39.9	41.1	West	North
Rheems-Roy-221084-TK1	06/03/2021	41.7	39.6	East	South

DATA PREPARATION FOR LSM

Landslide Inventory

The United States Geological Survey (USGS) provides a nationwide landslide inventory for the United States. The database contains more than 300,000 landslide point and polygon records and provides related information for the landslide event, such as the extent (when available) and times (when available) of landslide occurrence (Mirus et al., 2020). As the USGS landslide inventory is compiled from many different sources, the accuracy and quality vary between landslide records. The USGS landslide inventory uses a semi-quantitative classification to rank the relative confidence in landslide occurrence and position for landslide records in their inventory. Five integer values are used to represent certainties of landslide records. These values are "1" (very low confidence), "2," "3," "5," and "8" (very high confidence) (Mirus et al., 2020). This study only focused on landslide point records with a confidence level of at least "5".

Unfortunately, the USGS landslide inventory does not contain enough high-confidence landslide records for Pennsylvania; it only reports 12 landslide records in Pennsylvania with a confidence level of "5" (11 around the Pittsburgh area and one around the Philadelphia area). For the present study, landslide records in New Jersey were considered to obtain sufficient landslide inventory for developing LSM models. New Jersey shares similar ecoregions with the southeast part of Pennsylvania (i.e., they share a similar hill slope environment and climate) (McMahon et al., 2001). Thus, landslide locations in New Jersey can be used as a complementary data source for the present study. The landslide inventory used in this study contains 12 landslide records in Pennsylvania and 154 landslide records in New Jersey (166 landslides in total). Figure 3.4 presents a map showing the location of landslide records used in this study and the location of our study area. It should be noted that these landslides are not closely located in the study area due to the limited landslide records in the USGS landslide inventory for the study area. However, the underlying mechanisms for landslide occurrences share similarities; it is hypothesized that the developed LSM models can be generalized to predict landslide risk for the study area.
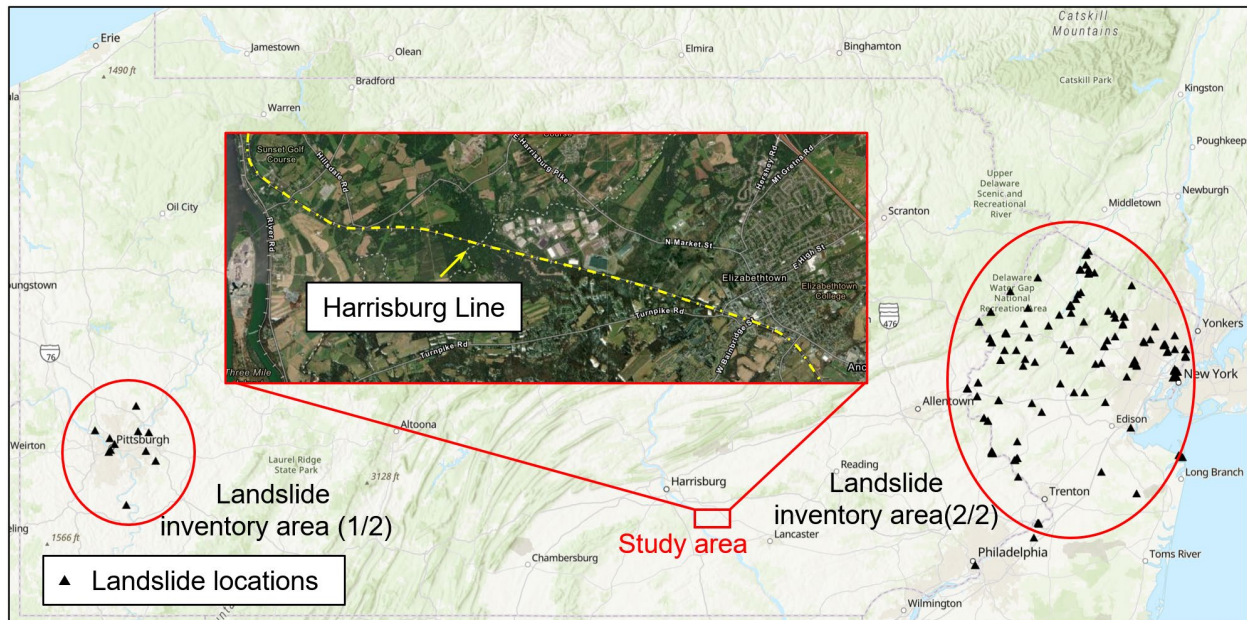
Figure 3.4. Summary of landslide locations used in the present study.

Landslide Contributing Factors

Quantitative evaluation of landslide hazards relies on analyzing the topographic features where landslides have occurred. The geomorphic requirement for potential slope instability is sufficiently steep slopes with material available for transport (Lu and Godt, 2013). This requirement can be stratified through different combinations of soil profiles, topographies, and vegetation coverages. Typically, the topography features can be obtained through geological maps. With recent advances in satellite, airborne, and ground-based remote sensing techniques, these data can also be remotely captured through satellites in raster format.

- **Terrain Morphology:** The terrain morphology information such as digital elevation model (DEM) data can be obtained from various sources, including shuttle radar topography mission (SRTM), advanced spaceborne thermal emission and reflection radiometer (ASTER), light detection and ranging (LiDAR), and synthetic aperture radar (SAR). The coverage and spatial resolution vary among different data sources.

- **Vegetation:** The vegetation on top of a slope controls water infiltration, and roots can mechanically reinforce soil layers. Therefore, it is essential to consider vegetation coverage when analyzing slope stabilities. A lack or shortage of vegetation can increase the susceptibility of slopes to landslides; on the other hand, slopes with decent vegetation coverage are less prone to landslides. Different vegetations have their unique spectrum characters, and these characters can be captured through multispectral images. The vegetation coverage, which is usually quantitively represented by the normalized difference vegetation index (NDVI), can be obtained by measuring the difference between near-infrared (NIR) bands from satellite images (e.g., Landsat-8 and Sentinal-2).

- **Soil and Bedrock:** Under rainfall events, water infiltrates into the hillslope and changes the stress state of soils. Soil hydrologic and mechanical properties have a significant impact on slope

stabilities. The permeability of soils varies between soils, and different soil types will have different responses to water infiltration and strength behavior. Therefore, it is essential to consider soil properties in any landslide analysis. The bedrock profile of a hillslope also significantly affects the hillslope response to rainfall as it is often considered a lower boundary for water infiltration. A recent study by Moradi et al. (2018) confirmed that the bedrock topography is an essential control for the spatial and temporal distribution of the potentially unstable zones in a hillslope. The soil and bedrock profile can be obtained from local geological maps. They can also be obtained as raster data on various online platforms. For example, SoilGrids (Hengl et al., 2017) provides global soil information and bedrock topography with 250 meters per pixel spatial resolution.

- **Precipitation:** The selection of appropriate precipitation data is crucial as it directly links to rainfall-induced landslides, especially for rainfall-threshold analysis for predicting landslides as precipitation is the primary input variable. Rainfall modifies stress conditions in hillslopes creating instabilities for landslides. Typically, precipitation data can be obtained from rain gauge measurement, radar, and satellite. Each source has its limitations. In general, it is challenging to measure precipitation precisely. Some of the most widely used precipitation products include NASA TRMM, NASA Daymet, and GSMaP.

Preparation of Geophysical Data

For the present study, all the landslide contributing factors were obtained from the Google Earth Engine (https://earthengine.google.com/). Google Earth Engine is a powerful online platform that combines a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities.

As the typical size of landslides in Pennsylvania is generally small, the resolution for topographic data should be high enough to reflect variations on the terrain surface. In the present study, USGS 3DEP was used to obtain high-resolution DEM data with a spatial resolution of 10 meters. These DEM data were used to calculate attributes like terrain slopes and aspects. The multi-scale topographic position index (mTPI) was also calculated based on the DEM for each location subtracted by the mean elevation within a neighborhood. The mTPI distinguishes ridge from valley forms. The soil information was obtained from SoilGrids (Hengl et al., 2017), which provides a global estimation of surface soil properties based on 150,000 soil profiles and remote-sensing data with a spatial resolution of 250 m. The soil information includes field capacities, sand/clay content, and soil bulk density. The annual average NDVI calculated based on Landsat-8 satellite images was used to represent vegetation coverage. The annual average rainfall was obtained from NASA Daymet. Table 3.3 summarizes the geophysical data used in the present study. It should be noted that these geophysical data were resized to the same spatial resolution as DEM for consistency prior to analysis. In addition, visualization of these geophysical data can be found in Figure 3.5.

Table 3.3. Summary of geophysical data.

No.	Factors	Resolution	Description
1	Elevation (m)	10 m	Terrain slope based on high-resolution, seamless raster elevation data produced and distributed by the USGS
2	Slope (deg)	10 m	Terrain slope based on high-resolution, seamless raster elevation data produced and distributed by the USGS
3	Aspect (deg)	10 m	Terrain aspect based on high-resolution, seamless raster elevation data produced and distributed by the USGS
4	mTPI (-)	270 m	Multi-scale topographic position index based on DEM that distinguishes ridges from valleys
5	NDVI (-)	30 m	Annual average NDVI based on Near-IR and Red bands using Landsat-8 images
6	Sand content (%)	250 m	Sand content for surface soils
7	Clay content (%)	250 m	Clay content for surface soils
8	Bulk density (10 kg/m^3)	250 m	Soil bulk density (fine earth) for surface soils
9	Field capacity (%)	250 m	Soil water content for 33 kPa suctions for surface soils
10	Rainfall (mm)	1000 m	Annual average precipitation based on daily surface weather data on a 1-km grid for North America
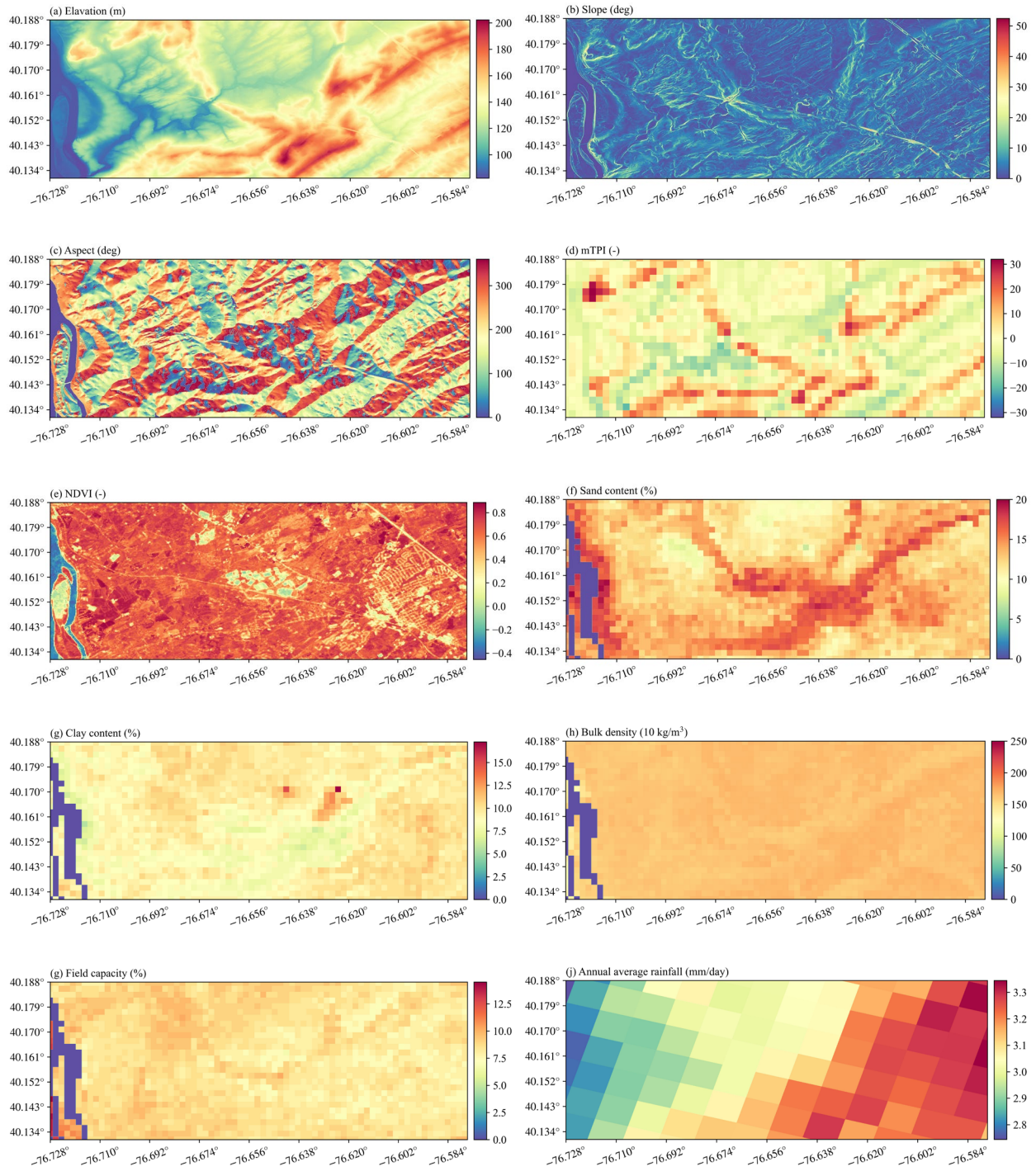
Figure 3.5. Visualizations of landslide contributing factors for the study area.

C H A P T E R   4

Analytical Methods

INTRODUCTION

This chapter covers analytical methods applied to study the data. This includes analyses focused on the railroad right-of-way images, historical landslide locations, and geospatial parameters.

EXPLORATORY DATA ANALYSIS

Digital image pre-processing techniques were applied to the extracted right-of-way image frames to study the data and clean them before developing an algorithm to assess slope conditions. This was done to further understand the data and to determine what pre-processing steps would be necessary to optimize an object detection algorithm. Several procedures were considered and tested throughout this research study. Simply, digital image processing techniques aim to treat images as a three-dimensional function $f(x,y,b)$ where each integer coordinate (x,y) is b layers deep based on the number of bands in the image (Pratt, 1994). For this research study, $b = 3$, since the extracted images from the videos were RGB images.

The first technique used to pre-process the images was filtering. Filtering is the process of convolving an image to calculate new pixel values from the original ones (Pratt, 1994). Filtering can be used to sharpen or blur images based on the intended use case. For the available data in this research study, two types of inconsistencies were identified in the images: artifacts on the camera lens (such as rain droplets or smears) and defective pixels in the images. A Gaussian filter was passed over the images to remove these inconsistencies. A Gaussian filter can be considered a low-pass filter, thus attenuating high-frequency signals (Haddad & Akansu, 1991). The larger the size of the Gaussian filter, the more blurring in the resulting image. Thus, a small 3x3 Gaussian filter was used, as shown in Figure 4.1. The filter was normalized such that the intensity of the image was not changed. Furthermore, since the filter is two-dimensional, it was passed over each color band individually. The filter was convolved over the ROW images, resulting in the blurring effect seen in Figure 4.2. Although the blurring reduces the sharpness of the image, several defective pixels seen in the image on the left were removed by convolving the Gaussian filter.

Figure 4.1. The 3x3 Gaussian filter for image blurring.
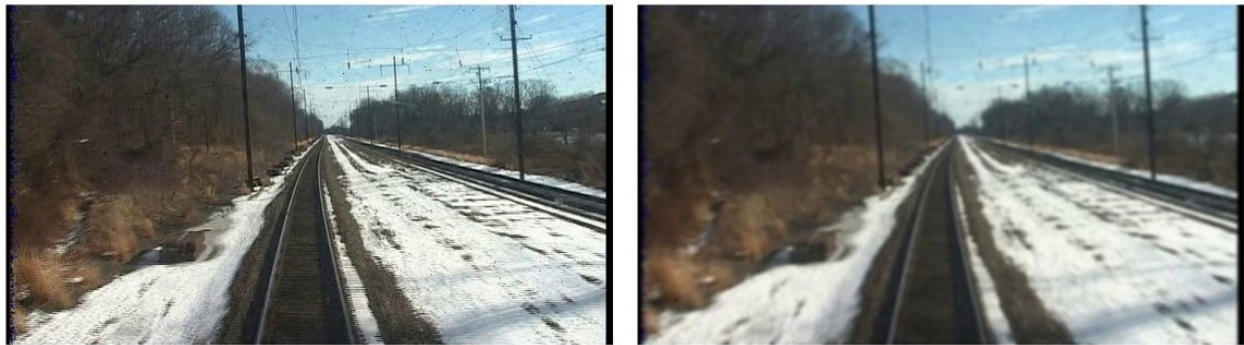
Figure 4.2. Raw image (left) and Gaussian filter image (right).

Average Image Intensities

Exploratory data analysis (EDA) was performed to investigate how the average pixel intensity in each image differed across color bands and recording dates. For each image, the average pixel intensity was calculated and normalized to have a value between zero and one. Figure 4.3 shows the results of this EDA for all westbound inspection runs, and Figure 4.4 shows the results for all eastbound inspection runs. Two plots were included in each figure to compare different pairs of calculated intensities by the color band. The primary observation made via this EDA was that average pixel intensity values by image vary both within the recording and across recording dates. The intensities are linearly correlated across color bands, indicating different luminosity in ROW images. Furthermore, certain inspection runs produced more significant variations in average pixel intensities, such as Rheems-Roy-218223 in Figure 4.4, which appears to be due to heavy shading of the right-of-way in certain parts of the recording due to the time of the recording and the weather. In summary, intensity values vary enough that the contrast of images must be normalized in some manner to remove the trend.
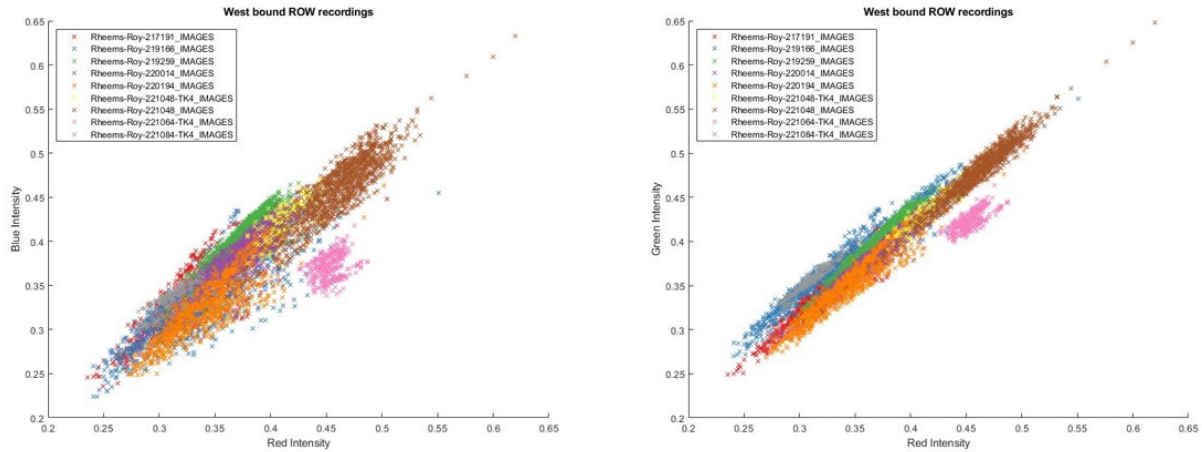
Figure 4.3. Average image pixel intensities by color bands (westbound).
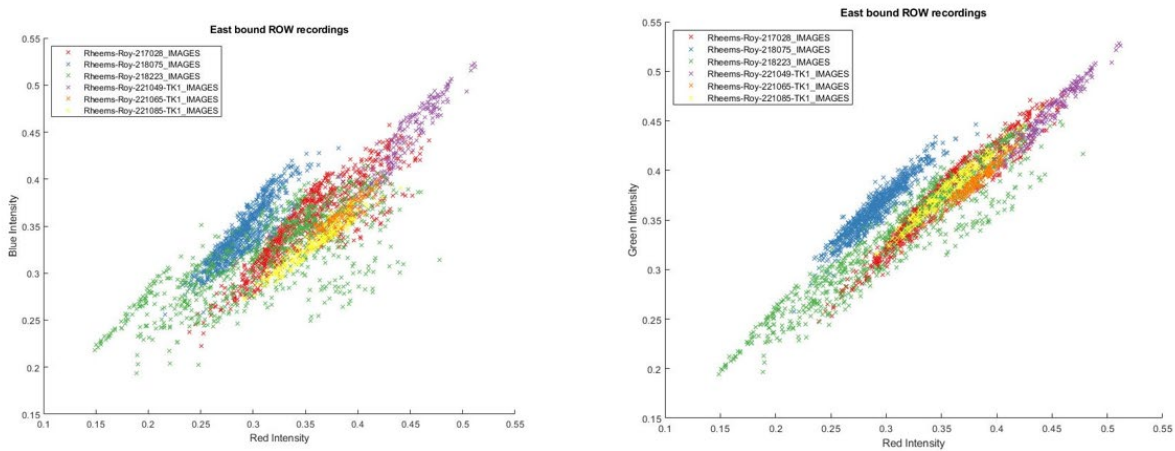
Figure 4.4. Average image pixel intensities by color bands (eastbound).

In addition to investigating average image intensity values by recording date, the values were also studied by season. Combining all the recordings, and producing a new legend based on the season of the recording date, resulted in the plots shown in Figure 4.5. Broadly, images recorded in the winter were more likely to be more luminous, while images recorded in the fall were more likely to be less luminous. This again supports the need for a contrast normalization algorithm to prepare the image for further study.
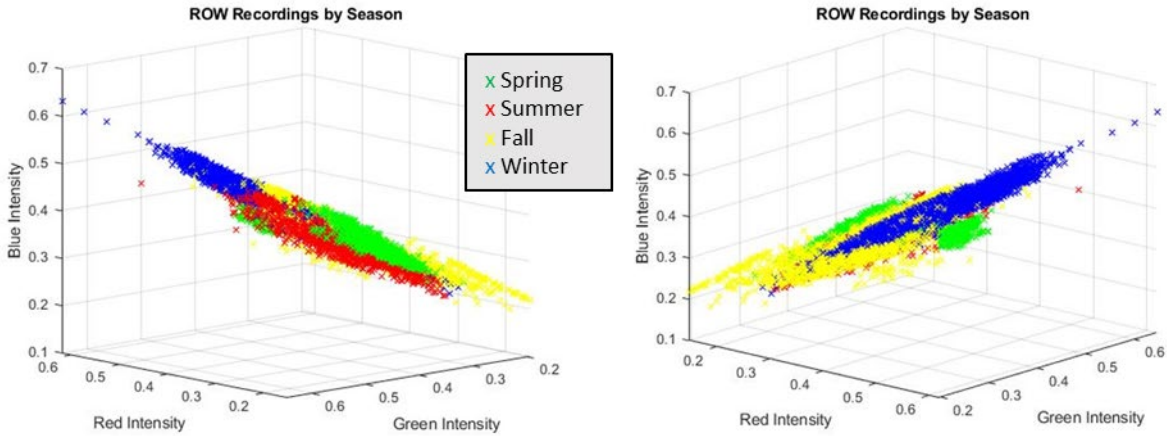
Figure 4.5. Average pixel intensities by color bands across seasons.

From the EDA, histogram equalization was the first considered algorithm to adjust the contrast of ROW images in each color band (Pratt, 1994); this process aims to better distribute pixel intensities in an image. This process is beneficial when comparing images recorded at different light levels due to the time of day or current weather, which was shown to be the case for this study's available data. Although the intensity values of pixels were found to be inconsistent across the three-color bands (as Figure 4.6 shows for a sample image) and across recording dates, an alternative method to adjust image intensity was used. Much like histogram equalization, normalization, or histogram stretching, was used to adjust all the ROW images to a similar intensity scale. This was done to correct the differences in light levels seen as described earlier. Each pixel's intensity value was adjusted per Eq. 4.1 (Pratt, 1994). This process was considered satisfactory to prepare the images for input into an object detection algorithm.
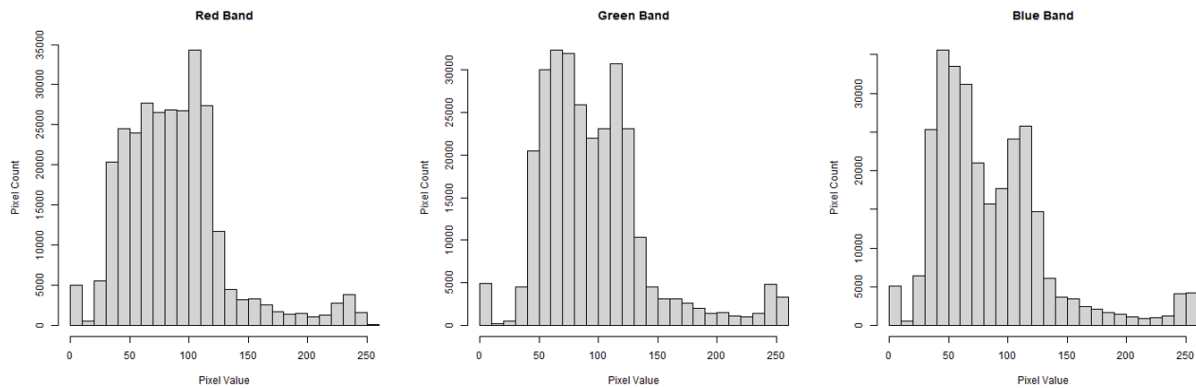
Figure 4.6. Histogram of pixel values for a sample image.

$$I_N = (I - MIN)\frac{nMAX - nMIN}{MAX - MIN} + nMIN \tag{4.1}$$

where:

I_N : New pixel value.

I : Original pixel value.

MIN / MAX : Minimum and maximum pixel values in original image.

nMAX / nMIN : Desired minimum and maximum pixel values in new image.

Edge Detection for Rail Identification

After testing data pre-processing techniques, the ROW images were further analyzed using digital image processing algorithms. First, the camera placement on the inspection vehicle and the resulting scene, and the location of geohazards and rail sections in the images was of interest. A simple edge detection algorithm was used to determine if the camera scene was consistent across all recordings. In digital images, edge detection models are used to identify "edges" or locations in an image where pixel values sharply change. Numerous techniques exist to perform edge detection, but all aim to find discontinuities in image brightness corresponding to variations in depth, surface orientation, material properties, or scene illumination (Ziou and Tabbone, 1998). For this research study, edge detection was primarily used in the images to identify rail sections and secondarily to identify infrastructure, such as bridges or catenary poles.

To perform edge detection, a second-order Gaussian derivative was approximated over each image in R (Barthelme, 2021; R Development Core Team, 2003). The algorithm was applied across both the vertical and horizontal axes to detect edges separately. The expected outcome of applying the algorithm was the rail sections being clearly identifiable when detecting edges across the horizontal axis. Figure 4.7 and Figure 4.8 show the edge detection results horizontally and vertically across a sample ROW image, respectively. Brighter pixel values (or pixels that are closer to white in color) indicate a significant change value between adjacent pixels. When comparing, the rail, as well as trees in the geohazard and catenary poles, have sharply defined edges when edges were detected horizontally across the image. Conversely, the bridge is much easier to see in Figure 4.8 than in Figure 4.7.
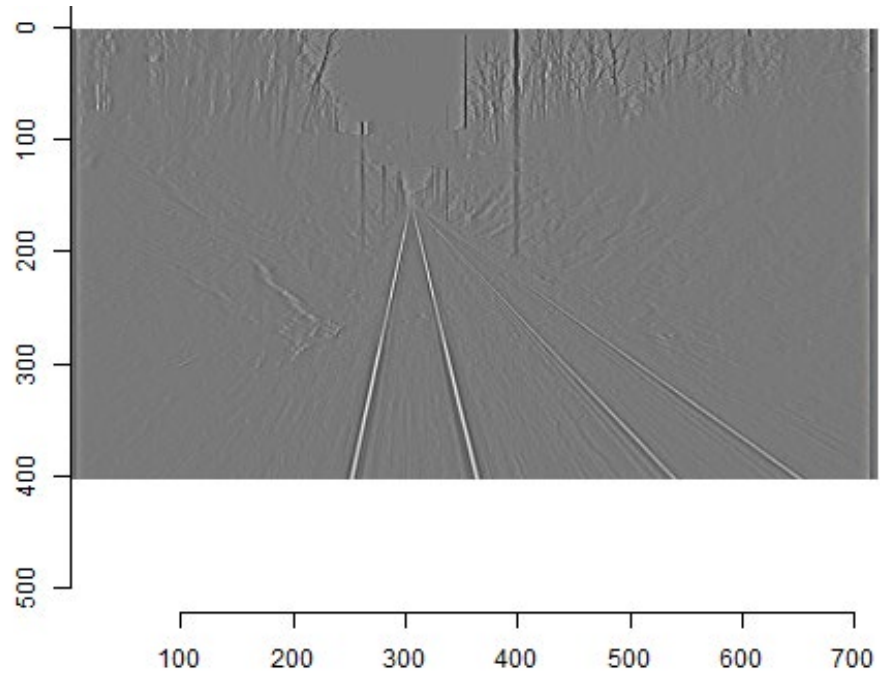
Figure 4.7. Edge detection horizontally.
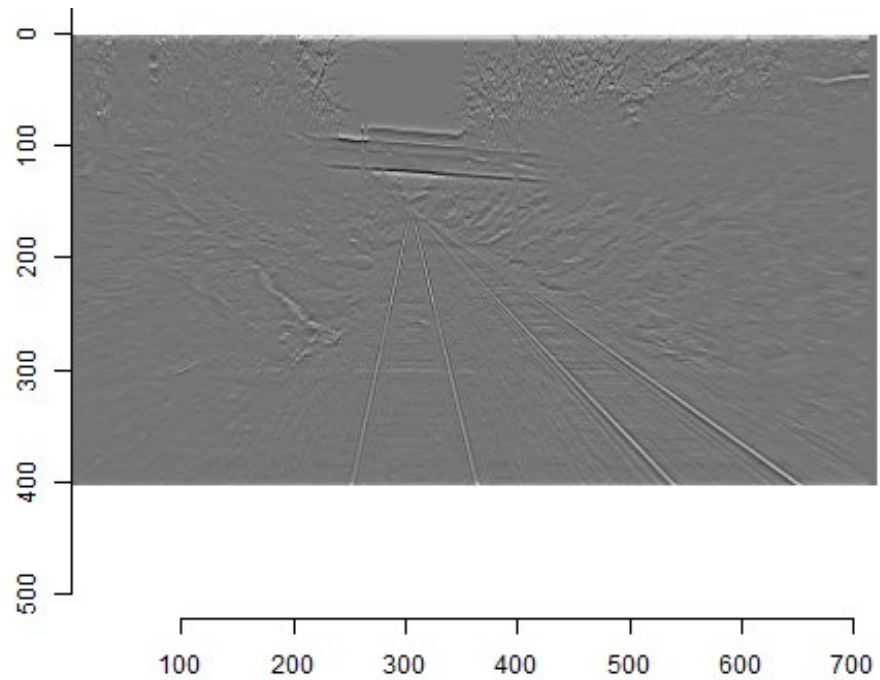
Figure 4.8. Edge detection vertically.

Since the rails produced clearly defined bright pixels, as shown in Figure 4.7, it was determined that these pixels could be isolated. By setting a threshold where pixel values above the threshold are white and below are black, followed by eliminating white pixels that had no other adjacent white pixels, the rail

was able to be isolated. Figure 4.9 shows isolated rail sections for two sample images; the left shows a section of tangent track and the right shows a section of curved track. Although additional pixels not representative of the railhead still exist in the images, the rail was still clearly identifiable.
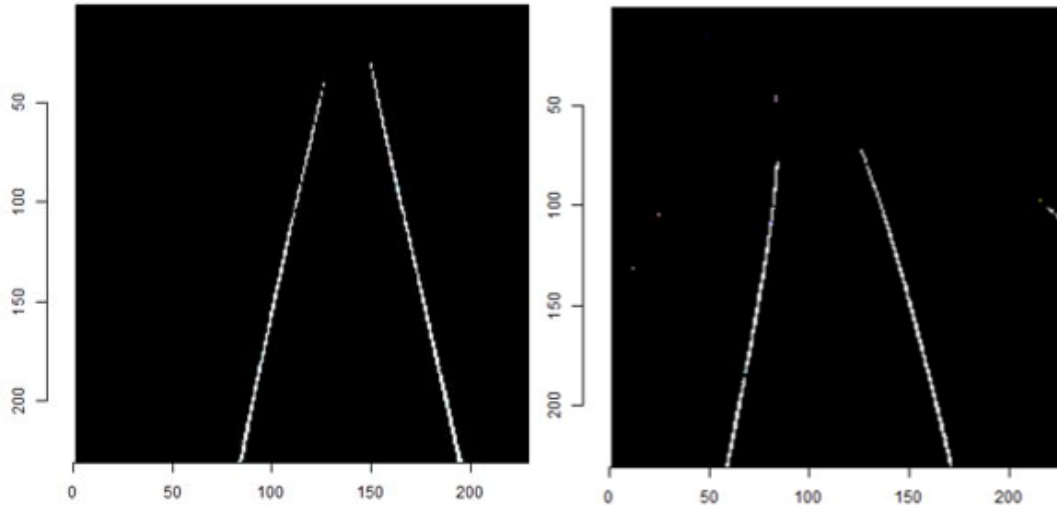
Figure 4.9. Isolated rail sections in ROW images.

Several analyses were performed using the isolated rail pixels. First, the rail sections were used to determine if the camera scene was identical throughout the ROW recordings. Via inspection of the collection of recordings, it was clear that the scene would not be identical for all the recordings. This is because, as columns four and five of Table 3.2 show, the videos were recorded on inspection runs traveling both east and west along the route as well as on the north and south tracks. Thus, it was not expected that the scene in all the videos would be identical. Instead, the scenes for sets of recordings with both the same inspection run direction and track were of interest. To determine if this was the case, the isolated rail pixels for images taken at the same location along the route but on separate dates were overlayed. Since the ROW images were previously georeferenced, this process was trivial. Figure 4.10 shows two images recorded at the same location along the route (one on 2/11/2020 and one on 12/1/2020), where rail pixels were isolated, and the track gauge was identified (red line at the bottom of the image). Although the track gauge calculated from the ROW images was identical on both dates (113 pixels), Figure 4.11 shows that the rail pixels do not align. This indicates that the camera had the same focus on both dates but was not centered on the same portion of the scene. In fact, to align the isolated rail pixels, one of the images had to be both shifted and rotated. Further analysis of isolated rail pixels calculated from images recorded on other dates indicated that the camera scene was inconsistent throughout the data collection period. This means that images must be aligned manually or automatically to investigate changes to the right-of-way or of geohazard features of interest over time to a fine detail.
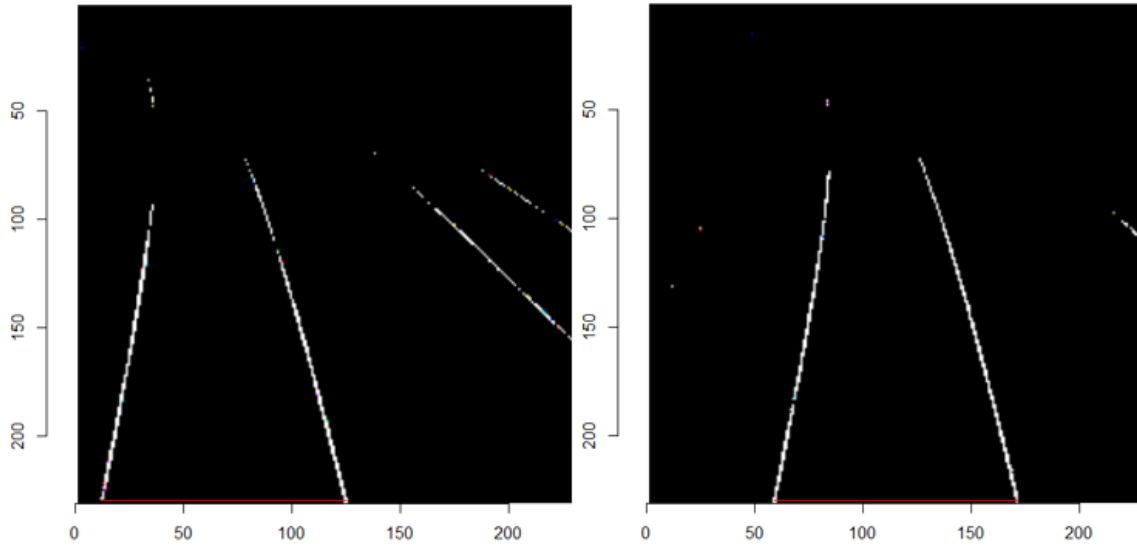
Figure 4.10. Isolated rail pixels for two images recorded at the same location.
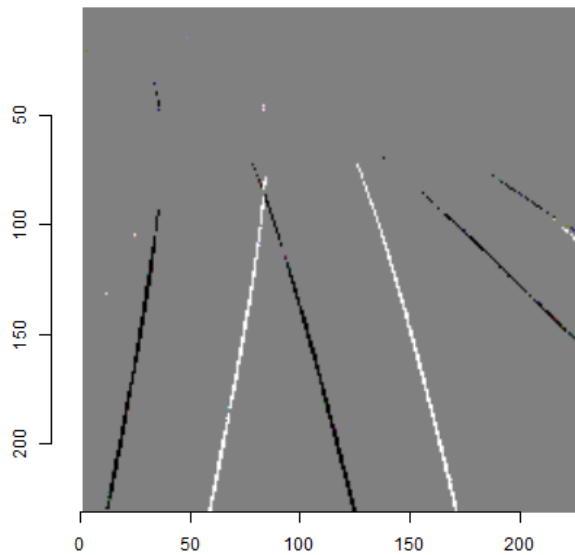
Figure 4.11. Overlayed rail pixels from images in Figure 4.10.

After inspecting the results, an additional use case for the isolated rail pixels was tested. As Figure 4.9 shows, isolated rail pixels from sections of tangent and curved tracks created different signatures. From this, it was hypothesized that the type of track could be determined automatically using the isolated rail pixels. To do this, the track centerline was calculated from the isolated pixels. This was done by calculating the average horizontal coordinate of every white pixel for each row of a ROW image. The expectation was that the centerline would be vertical and linear for the tangent track and skew for the curved track. Figure 4.12 and Figure 4.13 show, on the left, the isolated rail pixels after removing detected pixels that were not a part of the rail, and on the right, the calculated centerline for the isolated rail pixels for a tangent and curved track section, respectively. Although the pixel isolation process was not perfect, the difference

between the centerline for both scenarios is clear. The calculated centerline shown in Figure 4.12 is almost perfectly vertical, while the calculated centerline in Figure 4.13 is obviously curved.
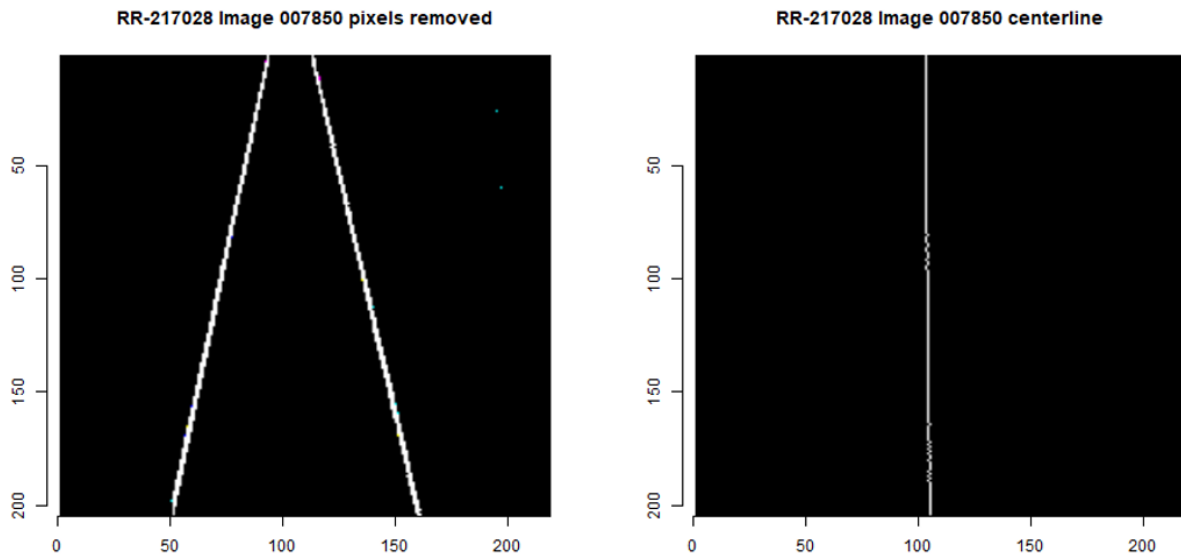
Figure 4.12. Centerline calculation sample for a tangent track.
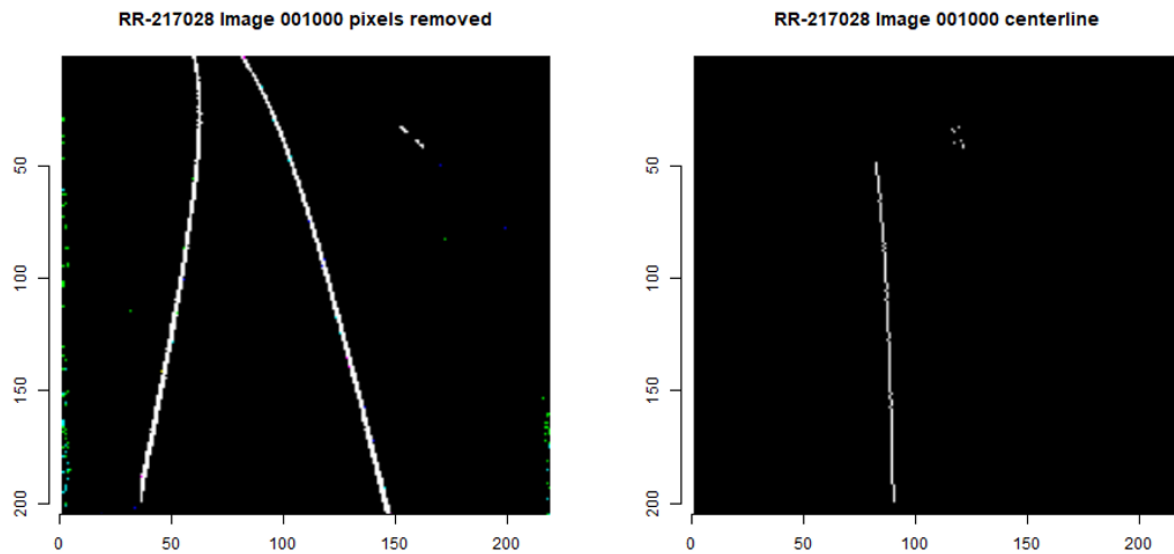
Figure 4.13. Centerline calculation sample for a curved track.

The above calculations were made for sample ROW images to see if curved and tangent track sections could be classified accurately. Using the centerline pixel coordinates, regression was used to fit a linear model and a second-order polynomial to each centerline. The expectation was that tangent track sections would produce low residuals for both fits, while curved track sections would produce low residuals for the polynomial fit only. Figure 4.14 was created for a small sample of ROW images and shows the squared residuals of the polynomial fit vs. the linear fit for each calculated centerline. The population of

data points with nominal residuals for both fits indicates tangent track sections where clean centerlines were able to be calculated. Since the centerline calculation technique was imperfect, some of the calculated centerlines were also imperfect. Thus, centerline residuals that skew from the line (indicating equal residuals for both fits) can be classified as not linear, although the behavior may be due to an imperfect centerline or the photographed track section being curved. Regardless, Figure 4.14 indicates that determining whether a photographed track section is tangent or curved is possible using this method with improvements to the pixel isolation technique.
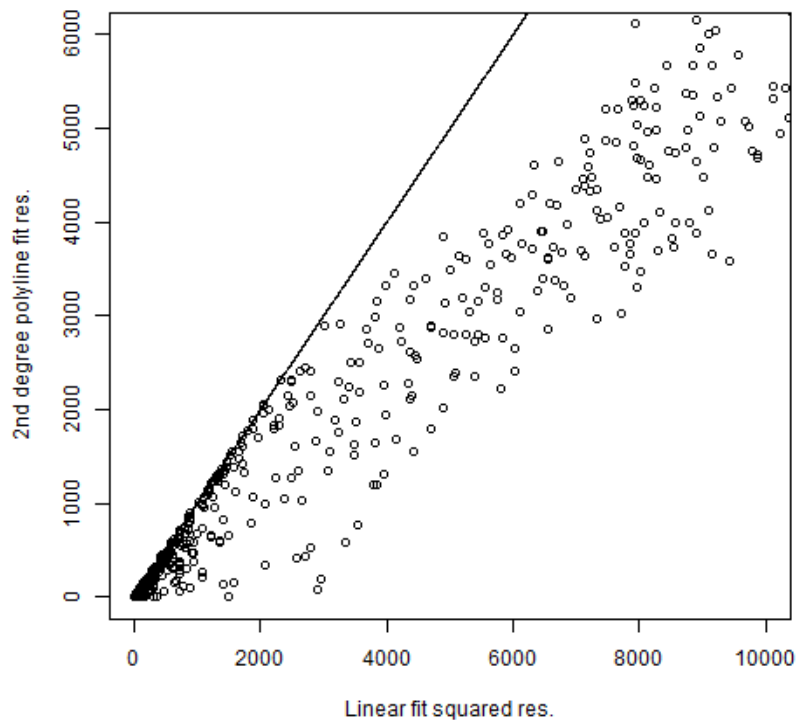
Figure 4.14. Track centerline fit residuals.

TRANSFER LEARNING FOR OBJECTION DETECTION

Feature Selection

Following exploratory work to learn more about the ROW images, initial steps were made to prepare them for use in an object detector. The goal was to use the extracted image frames to investigate the identified geohazards (shown in Table 3.1) for features that indicate potential slope instability. While rock and soil slopes have different failure mechanisms, unstable slopes generally have comparable forces driving slope movement and forces resisting slope movement. Using just the right-of-way videos, measurements cannot be made to determine these forces quantitatively. Instead, features that indicate past or potential slope movement are the key to determining slope stability. Figure 4.15 shows features that are of primary interest when performing slope reconnaissance (Li et al., 2002). These features indicate soil creep for soil slopes, meaning that the slope has already experienced progressive soil movement. Thus, a

catalyst such as a heavy rainstorm could be enough to trigger a large soil movement event. Figure 4.15 shows that there are leaning trees and hummocks present in an example geohazard image as well as some debris. These unstable slope indicators must be quantified to determine how many indicators are present at each geohazard and how that count changes over time.
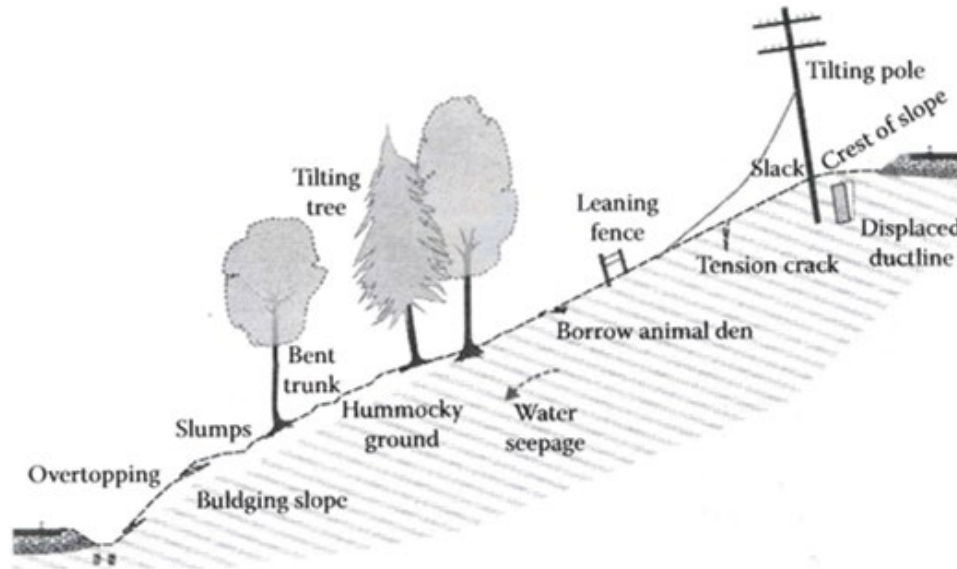
Figure 4.15. Surface features indicative of slope instability (Li et al., 2002).

Training Images Labeling

An object detection algorithm was created to identify and localize unstable slope indicators of interest in the right-of-way images. Object detection algorithms utilize CNNs to match patterns in testing datasets to existing patterns bounded in training data (Dasiopoulou et al., 2005). Like other ML algorithms, CNNs "learn" weights by studying training data. Unlike other neural networks, convolutional ones identify features found in a subset of the images using convolutional filters that slide across the image (Zhang et al., 1990). A training dataset was first created using a subset of the ROW images showing geohazards to facilitate accurate supervised learning for this use case.

With every image georeferenced to the study route, Table 3.1 was used to determine which images were recorded within a geohazard section. Of the total 12,000 images extracted from the ROW videos, 2,789 images were recorded in a geohazard section. Since slopes along the right-of-way only exist in geohazard sections, unstable slope features of interest are most likely to be found in these images. Thus, determining which images were captured in geohazard sections was an essential preliminary step in creating the object detection training dataset. To facilitate supervised ML, 50% of the geohazard images were randomly selected for use as training and validation data, while the remaining geohazard images were reserved for algorithm testing. For the training and validation images, unstable slope features were bounded and labeled manually such that the object detection algorithm can learn the image patterns that represent the chosen features. The six chosen features for the first pass of training and the color of their bounding boxes in ROW images are shown in Table 4.1.

Table 4.1. Legend of chosen unstable slope features

Feature	Color
Leaning tree	Lime green
Hummock	Red
Rock face	Cyan
Debris	Pink
Moisture	Violet
Tension crack	Dark blue

Features were identified and bounded using MATLAB's imageLabeler application (Mathworks, 2020). This application allowed rectangular bounding boxes to be drawn onto images in the chosen set. After identifying all features of interest in an image, the coordinates of each bounding box were automatically recorded. Although more sophisticated types of feature bounding exist, such as polygon and pixel bounding, rectangular bounding boxes were considered suitable as an initial feature identification method for this use case. To perform manual feature bounding, training and validation images were first resized to remove the text box at the top of the scene. This was done to create an image scene that only showed the right-of-way. Figure 4.16 shows an example of a training image with features of interest bounded.

Figure 4.16. Example of a training image with features bounded.

Several concerns were raised throughout the manual labeling process. First, due to the resolution of the images, it was difficult to define the borders of some features accurately. For example, the identified hummock in Figure 4.16 was bounded conservatively since the precise feature was hard to identify. Due to this, using more sophisticated bounding methods would have been challenging. Additionally, seasonal foliage often obscured features of interest, most often leaning trees. Although features of interest such as hummocks or leaning trees may exist on the right side of the right-of-way shown in Figure 4.16, they cannot

be easily identified and bounding. Due to these reasons, the training data labeling process may be re-investigated and improved in the future.

After inspecting all images in the training dataset for unstable slope features, the development of the object detection algorithm began. The first step in this process was to inspect the labeling results. Table 4.2 shows the number of instances of each object class bounded and the number of images that contained at least one instance of each object class. The latter shows that several images contained two or more instances of certain objects. Table 4.2 also shows that some object classes, specifically tension cracks, were not well represented in the study area. This could be due to bias in the labeling process but is more likely due to a few instances of tension cracks existing in the study area. Class imbalance in feature training examples can lead to learning bias in object detection algorithms (Oksuz et al., 2021). Consequently, only the first four object classes of the six in Table 4.2 were used to train the object detection algorithm. Omitting debris and tension crack objects yielded a more accurate object detection algorithm.

Table 4.2. Number of labeled objects and images present.

Feature	Count	Image Count
Leaning tree	536	416
Hummock	725	615
Rock face	785	489
Debris	214	190
Moisture	325	315
Tension crack	22	22

Model Selection

A suitable CNN algorithm and network structure were first determined to develop the object detection algorithm. The Faster RCNN structure was used for the sake of simplicity, since it provides accurate results in the shortest amount of computation time (Ren et al., 2017). Furthermore, rather than developing and training a novel network architecture, transfer learning—a research technique in which an ML solution to an existing problem is applied to a different but related problem—was used to adjust a pre-trained network architecture. This ensures that a network with a history of providing accurate results is used as a basis for learning. In short, if training accuracy is not satisfactory, it is much more likely that the dataset needs adjustments, not the network structure. The chosen pre-trained network structure for model training was the residual network (resnet18) architecture, which was trained on the ImageNet database and was built to have 1,000 output classes (Deng et al., 2009; He et al., 2016). The ImageNet visual database was designed for use in object recognition software and contains more than 14 million images containing 20,000 different categories with bounding boxes and labels included (Deng et al., 2009). Due to the extensivity of the database, neural networks trained on it have performed very well.

The resnet18 architecture is a relatively simple, publicly available network architecture that is easy to adjust for transfer learning. The network is only 18 layers deep, meaning transfer learning is not as computationally intensive as other network architectures (He et al., 2016). The resnet18 network

architecture used to train the object detector is shown in Figure 4.17. Since the object detector was to be trained on graphics hardware that is not the state-of-the-art, it was considered beneficial to use a network with fewer layers. This network was adjusted to have the correct number of output class labels (four) and for those class labels to correspond to a subset of a training image (region proposal) using MATLAB (Mathworks, 2020). To limit computation time and ensure that transfer learning would work as intended, the ROW images were resized from their original size of [720, 480, 3] to the design network input size of [224, 224, 3]. First, images were reduced from their original size to [720, 403, 3] to remove the text box. Though the aspect ratio of the images was not preserved by the resizing process, the geohazard features were still visible. Furthermore, since features were bounded prior to the second image resizing, the bounding boxes were analogously resized. Lastly, as described earlier in this report, the training and validation images were normalized to equalize the distribution of pixel intensity values.
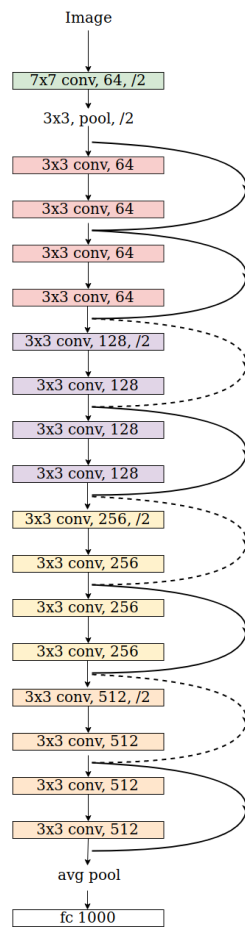
Figure 4.17. Adjusted residual network architecture.

LSM USING ML MODELS

LSM using ML methods can be considered a supervised classification problem where features of input variables (i.e., landslide contributing factors) are extracted and grouped based on their relationships

and contributions to the output variable (i.e., landslide/non-landslide labels in a binary classification). A complete workflow of LSM for the present study is summarized in Figure 4.18, which includes dataset preparation before developing LSM models, LSM model development and evaluation, and determination of the final landslide susceptibility map. The following sections will discuss details of the developed workflow for developing LSM models.
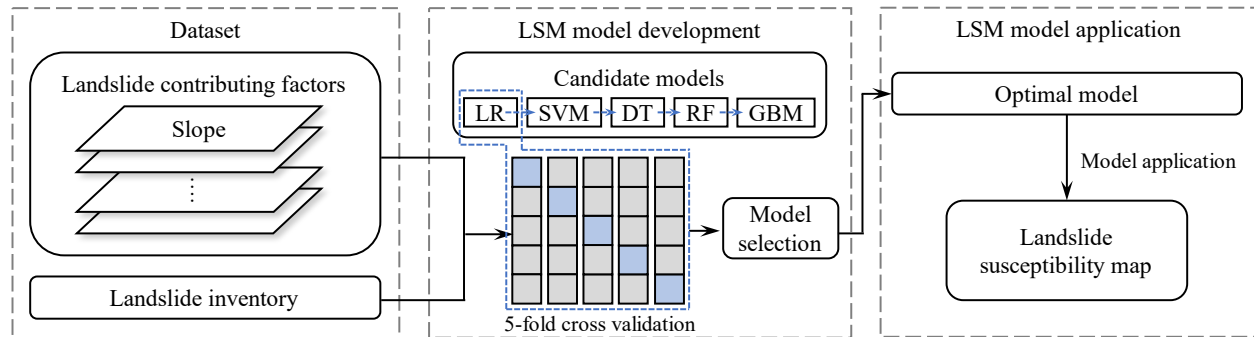
Figure 4.18. Developed LSM workflow for the present study.

LSM Model Description

The present study considered five commonly used ML models for slope stability classifications, including logistic regression (LR), SVM, DT, random forest (RF), and gradient boost machine (GBM). These models have been successfully used in multiple engineering applications, including slope stability predictions (e.g., Qi and Tang, 2018; Zhou et al., 2019). A detailed introduction of these models can be found in Kuhn and Johnson (2013), and a brief description of each model is given as follows.

- LR is a widely used and well-understood statistical model for binary classification problems. In LR, the conditional probability of the dependent variable is modeled as a logit-transformed multiple linear regression of the explanatory variables.
- SVM is a supervised ML algorithm based on the structural risk minimization principle (Vapnik, 1995; Burges, 1998). It can be used in both classification and regression problems. Given a dataset with N points, SVM finds a hyperplane that separates the data with the largest margin for linearly separable data. Nonlinear kernels are used for linearly inseparable data to allow the algorithm to fit the maximum-margin hyperplane in a transformed higher-dimensional feature space.
- DT is another commonly used supervised ML method for classification and regression. It is a non-parametric method with no presumed relationships between output and input variables (Qi and Tang, 2018). The learning objective for DT is to create a model that predicts target variables by learning simple decision rules inferred from input data. It creates a tree-like model to map from observations about a sample to conclusions about its target value.
- RF is an ensemble method in ML. It develops several DTs in parallel with bootstrap sampling. Various subsets of training data and different feature subsets are used to develop different DTs during model training. Each DT in the RF makes a class prediction, and the class with the most votes becomes the final model prediction. RF is considered a more robust model than DT, with better generalization capabilities.
- GBM is a powerful ensemble ML technique that has shown considerable success in many practical

applications (Natekin and Knoll, 2013); it can be used for classification and regression problems. Unlike RF, GBM combines a group of weak learners (e.g., DTs) in a stage-wise fashion; it consecutively reduces error with each additional weak learner until one final model is produced. GBM also allows the optimization of arbitrary differentiable loss functions; thus, it provides better flexibility during model development.

Database Preparation

For the present study, the 166 mapped landslide points in the study area and their corresponded values of landslide contributing factors were used as positive samples. For negative samples, non-landslide locations were randomly sampled 1 km within and 100 m outside of landslide locations. This approach to generating pseudo-negative samples was commonly used in LSM (e.g., Maxwell et al., 2021). These non-landslide points and their corresponding values of landslide contributing factors were used as negative samples. To create a balanced dataset, the number of sampled non-landslide points was equal to the number of landslide points. As a result, the dataset for developing LSM models for the present study consists of 332 samples, and each sample contains ten input features and one output/target (i.e., landslide or non-landslide). Tables 4.3 and 4.4 present the descriptive statistics for landslide and non-landslide samples in the database. As shown in Tables 4.3 and 4.4, the values of each landslide contributing factor cover a wide range of variations; thus, data standardization was used for scaling all the input features to ensure model training results.

Table 4.3. Descriptive statistics for landslide samples in the database.

	Elevation (m)	Slope (deg)	Aspect (deg)	mTPI (-)	NDVI (-)
Mean	114.9	13.1	181.0	-7.3	0.4
Std.	89.5	11.1	98.4	18.4	0.2
Min.	1.8	0.5	3.9	-82.0	-0.1
25th	36.4	3.7	97.9	-16.0	0.2
50th	96.8	9.8	171.5	-4.0	0.4
75th	178.3	20.8	265.0	4.0	0.6
Max.	333.0	43.2	358.0	32.0	0.8
	Sand Content (%)	Clay Content (%)	Bulk Density (10 kg/m3)	Field Capacity (%)	Annual Average Rainfall (mm/day)
Mean	44.6	18.5	153.6	21.5	3.4
Std.	7.0	4.3	8.1	3.7	0.2
Min.	26.0	10.0	119.0	14.0	2.8
25th	40.0	16.0	149.0	19.0	3.3
50th	45.0	18.0	155.0	21.0	3.4
75th	49.0	21.0	158.0	24.0	3.6
Max.	61.0	28.0	174.0	31.0	3.9

Table 4.4. Descriptive statistics for non-landslide samples in the database.

	Elevation (m)	Slope (deg)	Aspect (deg)	mTPI (-)	NDVI (-)
Mean	133.6	5.9	183.1	1.1	0.4
Std.	104.6	6.8	104.9	13.0	0.3
Min.	-1.3	0.0	0.0	-34.0	-0.5
25th	46.7	1.3	100.0	-6.0	0.3
50th	118.0	3.1	178.0	0.0	0.5
75th	209.0	8.1	282.5	8.0	0.6
Max.	399.0	34.8	359.0	46.0	0.8

	Sand Content (%)	Clay Content (%)	Bulk Density (10 kg/m3)	Field Capacity (%)	Annual Average Rainfall (mm/day)
Mean	43.6	19.3	156.0	21.5	3.4
Std.	8.3	4.5	6.7	3.6	0.2
Min.	24.0	10.0	135.0	14.0	2.8
25th	38.0	16.0	152.0	19.0	3.3
50th	44.0	19.0	157.0	21.0	3.5
75th	50.0	22.0	160.0	24.0	3.6
Max.	63.0	30.0	171.0	36.0	4.0

LSM Model Development and Hyperparameter Tuning

The performance of a trained ML model needs to be evaluated on new datasets to test its generalization capability. A stratified random sampling technique was used to split 80% of the database for training and the remaining 20% for testing. The *k*-fold cross-validation technique was used for dataset partition to reduce sampling bias in the present study (Stone, 1974). In cross-validation, the training dataset is divided into *k* folds, and the ML model is trained using *k*–1 folds and validated using the remaining one fold. This process repeats *k* times to allow each fold to be served as a validation fold, and the final model performance is the average model performance for each validation fold. Five-fold cross-validation was used for the present study, corresponding to an 80%/20% sample split. Figure 4.19 illustrates the cross-validation procedure for the present study.

In the present study, all the ML models were developed using Scikit-learn (Pedregosa et al., 2011). Each ML model considered in the present study contains hyperparameters; the values of these hyperparameters can be tuned to adjust model performance. However, minimum hyperparameter tuning was used in the present study to reduce potential overfitting and improve models' generalization capability, as the developed ML models need to be applied to the study area that was not adjacent to these landslide locations. The default values of hyperparameters provided by Scikit-learn were used for these ML models. For detailed information about these hyperparameters, readers can refer to help documents for Scikit-learn.
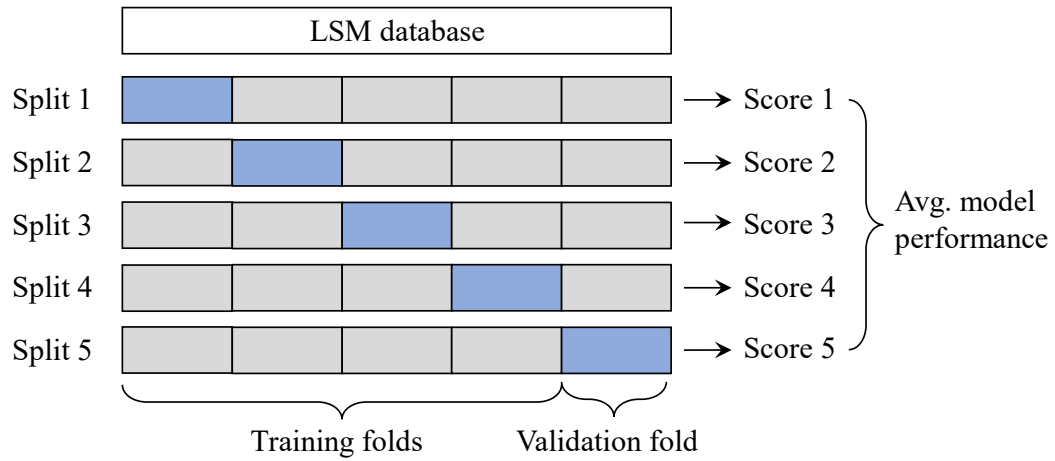
Figure 4.19. Fivefold cross-validation procedure for developing ML models.

CHAPTER 5

Results and Discussion

INTRODUCTION

This chapter covers all results obtained by applying the analytical methods introduced in Chapter 4. This includes the performance of an object detector and a landslide susceptibility map. Using these results, a relative risk assessment was conducted to compare the geohazard sections identified in the study area.

TRAINED OBJECT DETECTOR

Parameter Tuning

To train an object detector using transfer learning, several structural decisions had to be made. First, it was determined that using some of the training images to validate the algorithm would be beneficial. Table 5.1 shows the breakdown of image usage to train and validate the object detector. Validation data are not used to train the parameters during each training iteration but are instead used to tune the parameters at a predefined rate (Yegnanarayana, 1994). When validating the algorithm, improvements in error loss mean the hyperparameters are being appropriately trained and the algorithm's classification capability is improving. Thus, validation provides another metric to measure how well the object detector is performing as training progresses.

Table 5.1. Breakdown of image usage in transfer learning.

Image Set	Count
Training	1,132
Validation	150

Additionally, numerous training options had to be set to determine how the training procedure would function. The exhaustive list of training options is shown in Table 5.2. In brief, the algorithm processes "mini batch size" random images every training iteration. All images are processed in each epoch, meaning there were 187 iterations of training each epoch. The validation frequency instructs the algorithm to validate and tune the hyperparameters every 10 iterations. The initial learning rate is a factor that affects how much the hyperparameters can change each iteration. The learning rate was slowly reduced every five epochs, meaning as the algorithm's classification accuracy improves, the hyperparameters need to be tuned less and less. Since the algorithm is looking to classify portions of the image and not the image itself,

regions of the image are internally proposed to represent an object of interest. The number of regions and the number of regions sampled each iteration were numerically defined as well. Lastly, the positive and negative overlap ranges specify the network to treat region proposals as positive/negative samples by calculating the intersection over union of the region and the ground truth input. These parameters were adjusted to make both positive and negative samples more dependent on the intersection area.

Table 5.2. Object detector training options.

Option	Value
Mini Batch Size	6
Initial Learning Rate	4e-3
Max Epochs	20
Validation Frequency	10
Validation Shuffle	Every Epoch
Learn Rate Schedule	Piecewise
Learn Rate Drop Factor	0.5
Learn Rate Drop Period	5
# Strongest Regions	1500
# Regions to Sample	128
Negative Overlap Range	[0, 0.3]
Positive Overlap Range	[0.6, 1]

Algorithm Training and Testing

The object detector performed very well using the augmented resnet18 network architecture via transfer learning. After 20 epochs of training, the final model had a validation accuracy of 99.90% and an average validation root-mean-square-error (RMSE) of 0.10306, a little on the high side. The training results over each iteration, and the validation results over every 10 iterations, are shown in Figure 5.1. Training loss gradually shrunk over time and was relatively level when training was halted. Training time was roughly 40 minutes per epoch, meaning the total training time was about 13.5 hours. These results were the best results obtained following tuning of training options and removal of two of the six object classes labeled in the training images per Table 4.1.
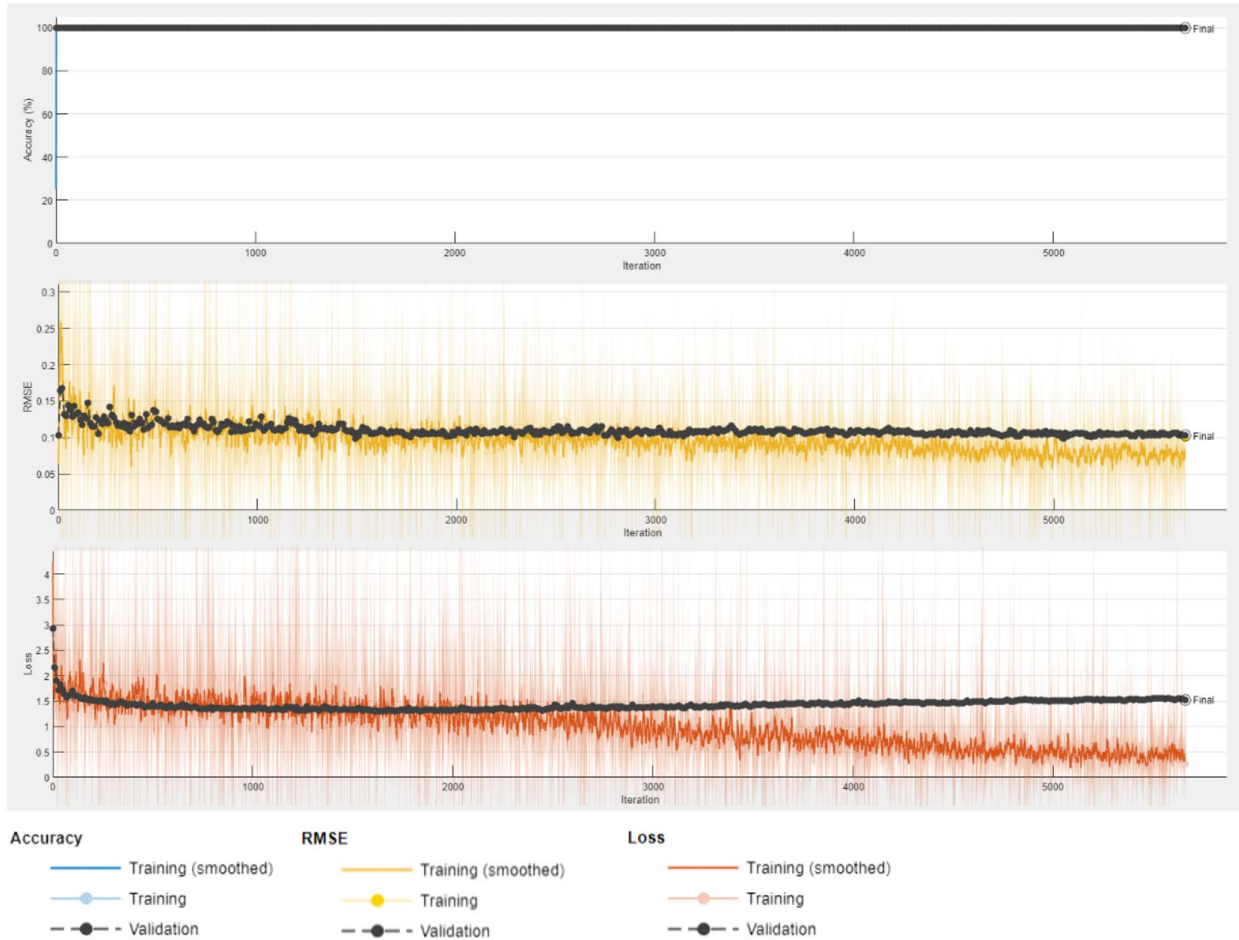
Figure 5.1. Training accuracy, RMSE, and loss.

To test the detector's performance, a precision-recall curve was created for each classified feature class. These plots are used to inspect the detector's capability by testing it on training images and measuring its precision and recall. Precision is the algorithm's positive predictive value, while recall is its sensitivity. Precision is defined as the true positive accuracy and recall is defined as the true positive rate (Powers, 2007). Precision and recall scores can be calculated as follows:

$$Recall = \frac{TP}{TP + FN} \tag{5.1}$$

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

where:
    TP : True positive.
    FN : False negative.
    FP : False positive.

For object detection algorithms, precision and recall are calculated to determine how accurately the detected objects in training images align with the training labels fed into the algorithm. Ideally, the detected objects are identical to the training objects. Figure 5.2 shows the precision-recall curves for the four labels the detector was trained to identify. The ideal algorithm would result in precision values of one for all recall values (a horizontal line at y=1). Considering this, the algorithm performed very well when identifying hummocks, rock faces, and moisture and performed less well when identifying leaning trees.
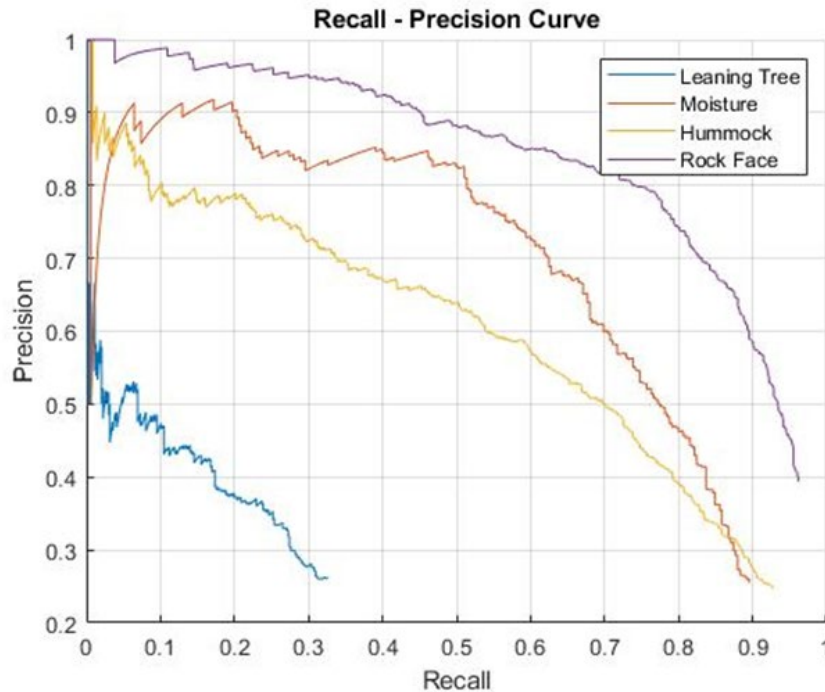
Figure 5.2. Recall-precision curve for trained object detector.

To visualize the algorithm's performance, training and test images were processed using the object detector to determine how well it was detecting objects of interest in the ROW images. Training images were used to see how well the detector was replicating the identified "ground truth" objects used to train the algorithm and testing images were used to visualize if the algorithm was detecting objects of interest in images that were not yet passed through the algorithm. Figure 5.3 and Figure 5.4 were randomly chosen to show the algorithm's ability to replicate ground truth labels. In both of these figures, the image on the left shows the "ground truth" (manually labeled) objects and the image on the right shows the objects detected by the algorithm. Numerous points of interest can be observed using these two figures. First, Figure 5.3 shows a section of moisture on the left side of the right-of-way that was properly identified, but interestingly, the detected bounding box (shown in the right image) is tighter to the section of moisture than the ground truth bounding box was. Although the detected box does not include all the moisture, it also does not incorrectly include a portion of the hill above the moisture like the ground truth bound box does. Similarly, the detector's ability is shown in Figure 5.4, where a rock face was correctly detected by the algorithm on the right side of the right-of-way that was not labeled. The same is true for the hummock identified on the

left side of the right-of-way. These examples indicate that the ground truth labels are not complete, and the labeling process could be improved.
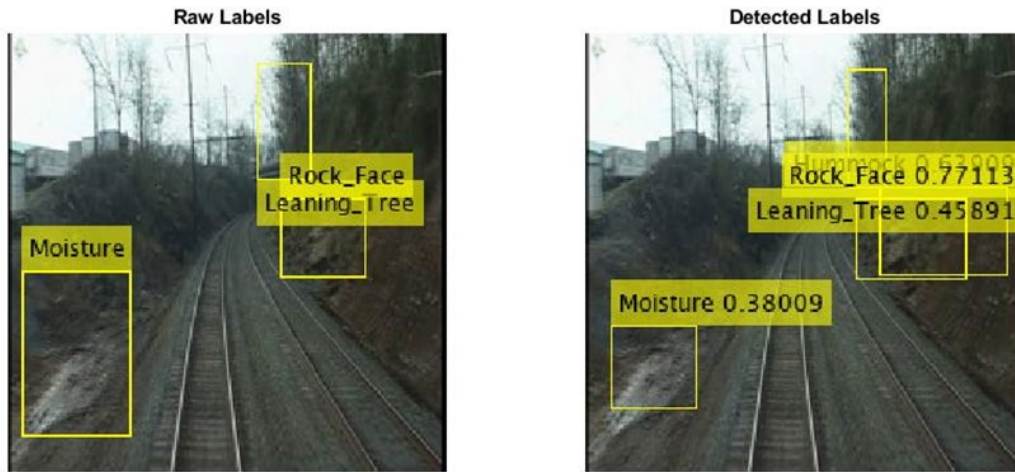
Figure 5.3. Detection results on first sample training image.
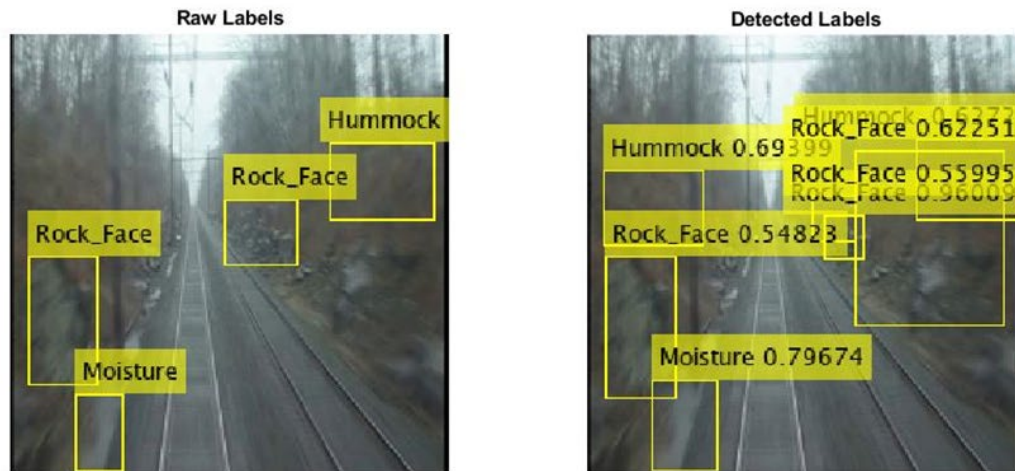
Figure 5.4. Detection results on second sample training image.

Although interesting, the algorithm's performance on test images is more important. Since the training images have already been seen by the algorithm, its predictive capability on them is considered biased. Figure 5.5 and Figure 5.6 show the algorithm's ability to identify objects of interest in sample test images. Objects of interest in the two sample images were mostly identified satisfactorily. Figure 5.5, for example, shows a leaning tree on the right side of the right-of-way that was identified properly but also shows a leaning tree on the left side that was not identified. Unsurprisingly, no leaning trees were identified in Figure 5.6 due to the prevalence of foliage. As discussed previously in this report, the density of foliage in some of the recorded ROW videos often obscures leaning trees. Similarly, the drape mesh seen on the right side of the right-of-way in Figure 5.6 obscures what could be either a hummock or a rock face. Since

sections of drape mesh were not labeled as part of the manual labeling process, the algorithm is not sure. Interestingly, the drape mesh was identified as a hummock, indicating that the algorithm's ability to identify hummocks is not biased to the color of the feature. Still, this observation indicates that the algorithm may have trouble differentiating hummocks from rock faces. Overall, the detector performed satisfactorily based on the quality of the images and the manual labels.

Figure 5.5. Detection results on first sample test image.
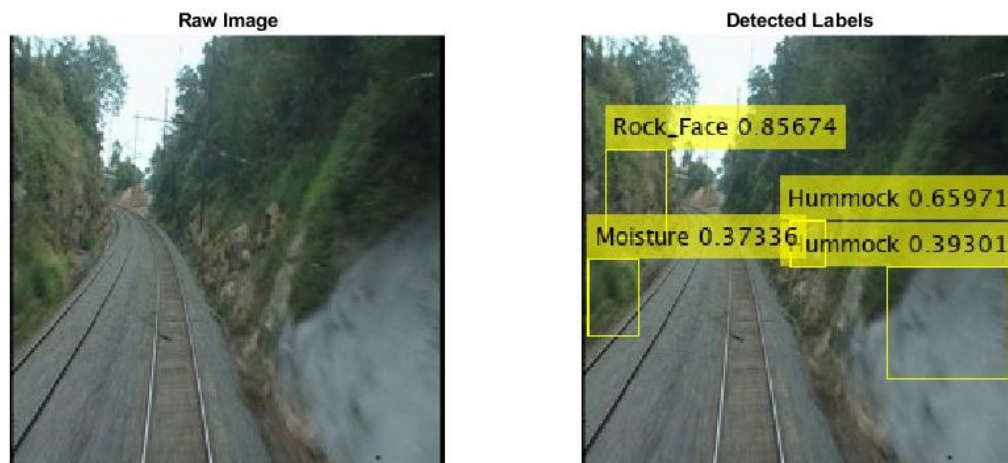
Figure 5.6. Detection results on second sample test image.

EDA of Identified Features

To further explore the quality of the ground truth labels, some exploratory data analysis was performed to determine why the algorithm's performance was skewed for one of the three object classes. Of interest was both the size and the location of the bounding boxes for each of the object classes. This was of interest, since the training images were not rotated or shifted before training the algorithm. Augmenting images for computer vision tasks has been shown to improve the algorithm's classification accuracy by removing spatial biases (Zoph et al., 2020). Figure 5.7 shows the location of bounding box centroids in the

CIAMTIS
U.S. DOT Region 3 University Transportation Center

ROW images. Expectedly, few objects were labeled in the middle of ROW images where the track was present. Also expectedly, leaning trees tended to be labeled near the top of training images, whereas moisture and debris objects tended to be labeled near the bottom. Lastly, hummocks and rock faces were identified in the middle of geohazard sections and spatially were found in similar portions of the images. Although this behavior was expected, it shows a pattern of object locations in the images that may have caused bias in the training process. Similarly, the bounding box's size can cause bias when training the algorithm. As Figure 5.8 shows, leaning trees had the smallest size variation, most often being tall and narrow. Interestingly, the sizes of bounding boxes representing rock faces were often much larger than those representing hummocks, indicating a different pattern than the one identified in Figure 5.7.
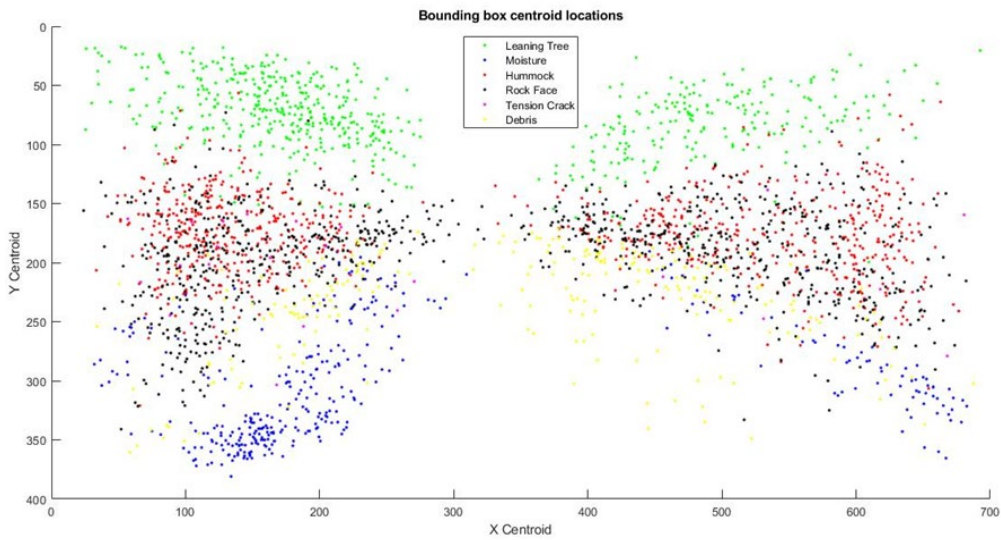
Figure 5.7. Ground truth bound box centroid locations.
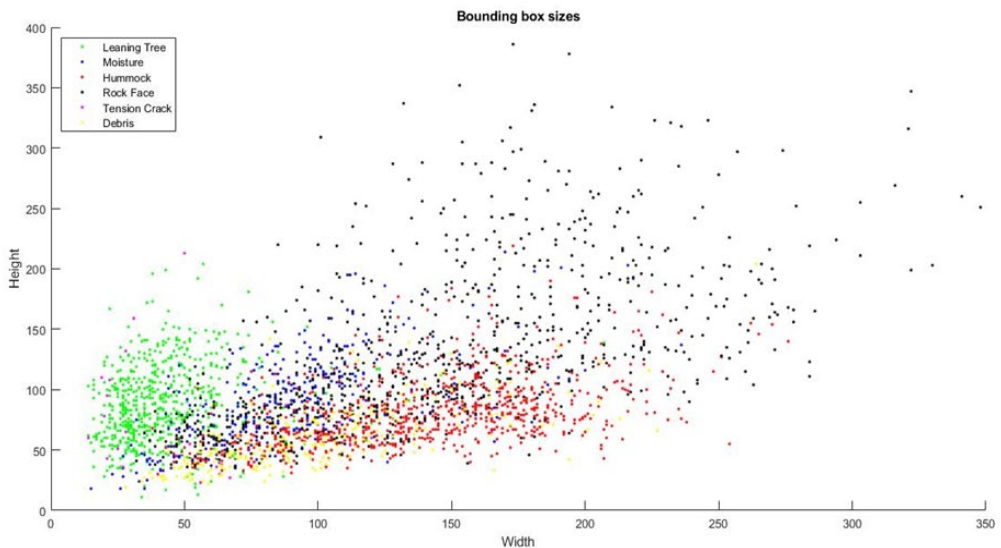
Figure 5.8. Ground truth bounding box sizes.

Using the trained object detector, all ROW images extracted from the recorded inspection run videos were assessed for features of interest. The algorithm tested random regions of the image to determine if an object was present. A low threshold of 0.3 was chosen such that more candidate features were classified as object instances. Due to the relatively small size of the training set, a lower threshold makes the detection procedure more lenient. This was done to determine both if features were primarily identified in geohazard sections versus non-geohazard sections and where features were identified in high density. The latter point of interest was used to relatively assess risk of failure of the geohazards in the study area. Several figures were created to illustrate the correlation between the location of detected objects in the study area and the identified geohazard sections. Figures 5.9 (a)-(d) show where each of the four features learned by the object detector were detected across the study area for a single inspection run and their correlation to the identified geohazard sections. The green boxes in the plots show which portions of the study area were identified to be geohazard sections. Object counts are plotted by number of occurrences in each image spatially across the study area and are shown as red lines. Furthermore, Figure 5.9 (b) also shows that leaning trees were detected less often than the other three object classes.

(a)

(b)

(c)

(d)
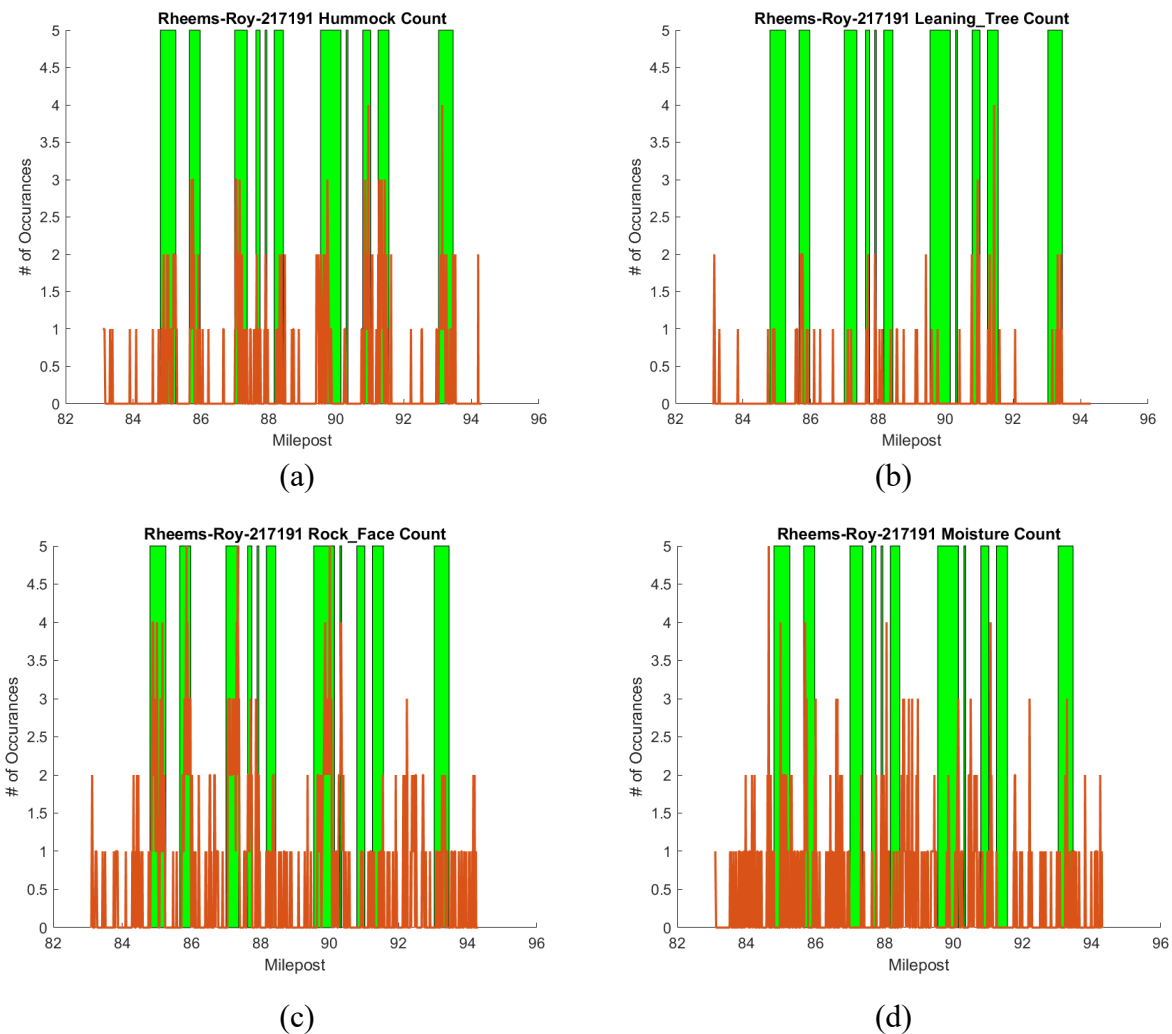
Figure 5.9. Identified geohazard objects across the study area (red: identified geohazard objects; green: railroad geohazard sections): (a) hummocks; (b) leaning trees; (c) rock faces; (d) moisture.

Figures 5.9 (a)-(d) show that most of the object class features were detected more frequently in geohazard sections than in non-geohazard sections for the example inspection run. To determine if this was the case across all inspection runs, Table 5.3 was created to illustrate the ratio of density of objects in geohazard versus non-geohazard sections across each inspection run in the primary study area. Eq. 5.3 was used to calculate each cell's value shown in Table 5.3. The calculated ratio is the density of detected objects in geohazard sections (summation of all jth class objects in geohazard sections divided by the number of geohazard images) divided by the density of detected objects in non-geohazard sections. Simply, the ratio represents how much more likely a detected object was found in a geohazard section than in a non-geohazard section. Looking at Table 5.3, barring detected moisture sections, features were much more likely to be found in geohazard sections than in non-geohazard sections for most inspection videos. Smaller calculated ratios were likely due to foliage-obscuring features and thus fewer detected objects in both geohazard and non-geohazard sections.

$$Ratio_j = \frac{\sum_{i=1}^{I_G} object_{i,j} \Big/ I_G}{\sum_{i=1}^{I_{NG}} object_{i,j} \Big/ I_{NG}} \tag{5.3}$$

where:

I_G : Number of geohazard images.

I_{NG} : Number of non-geohazard images.

j : Object classes.

Table 5.3. Ratio of detected object densities in geohazard and non-geohazard sections.

Survey ID	Leaning Tree	Moisture	Hummock	Rock Face
Rheems-Roy-217028	4.11	0.89	1.78	4.64
Rheems-Roy-217191	6.02	0.86	9.45	4.49
Rheems-Roy-218075	N/A	1.11	2.12	3.31
Rheems-Roy-218223	2.86	1.37	1.72	2.99
Rheems-Roy-219166	1.49	2.25	4.43	1.69
Rheems-Roy-219259	7.58	3.58	10.78	5.45
Rheems-Roy-220014	6.71	3.37	13.36	6.13
Rheems-Roy-220194	4.91	2.46	6.59	4.55
Rheems-Roy-221048	2.00	0.83	3.77	5.76

Although it is interesting to numerically show that more features were detected in geohazard sections than non-geohazard sections, the calculated ratio of densities carries no information about the number of detected objects throughout the inspection run. Furthermore, the number of geohazard and non-

geohazard images varies between inspection runs due to differences in inspection run distance and speed. Thus, Table 5.4 was created to show the number of detected objects in both geohazard sections and across the whole inspection video. Comparing Table 5.3 to Table 5.4 illustrates that a high-density ratio does not necessarily indicate many found features in geohazard sections. Inspection runs RR-219259 and RR-220014 are good examples of this observation when comparing the number of detected leaning trees to the density ratio. Without including observations drawn from Table 5.3, Table 5.4 shows which inspection videos yielded the most features and the distribution of those features across classes and across track section type. RR-217191 had detection results, since many features were found for each class and many of those features were found in geohazard sections. RR-221048 is satisfactory for similar reasons, but the disparity between the number of detected hummocks and rock faces indicates the detector may not have distinguished the two accurately.

Table 5.4. Count of detected objects in geohazard sections and in total.

Survey ID	Leaning Tree		Moisture		Hummock		Rock Face	
	Geoha.	Total	Geoha.	Total	Geoha.	Total	Geoha.	Total
Rheems-Roy-217028	40	65	88	341	128	313	303	471
Rheems-Roy-217191	62	89	82	332	209	267	298	472
Rheems-Roy-218075	0	0	144	466	183	398	311	545
Rheems-Roy-218223	21	40	121	350	162	405	193	360
Rheems-Roy-219166	3	8	76	160	269	420	227	560
Rheems-Roy-219259	24	49	241	772	344	596	209	512
Rheems-Roy-220014	34	77	210	739	321	525	224	534
Rheems-Roy-220194	51	79	140	293	230	324	260	414
Rheems-Roy-221048	63	140	204	802	97	160	631	899

While Table 5.3 and Table 5.4 summarize the detection results well, neither indicates the location along the study area where features were most identified. Figure 5.9 shows both the number and location of identified features across the study area satisfactorily, but only for one inspection run. Thus, a method to compare the number of identified features both over the study area and over time was hypothesized. A data cube was created to summarize all calculated summary statistics yielded from testing the object detection algorithm on ROW images. By extracting and grouping images by geohazard sections (all of which are shown in Table 3.1), calculated statistics are dependent on the recording date, the location along the track, and the object class, forming a hypercube of data. The number of images used to calculate statistics about the object detection predictions is dependent on how many extracted images show the geohazard of interest. Thus, the detected object density parameter used earlier was also used to investigate objects in individual geohazard sections. Primarily, calculated statistics for individual geohazards allow the geohazards to be compared to investigate changes in detected features over time.

Table 5.5 shows one slice of the proposed data hypercube. The table shows detected object densities (the number of object instances over the number of images in the geohazard section) for all 11 identified

geohazard sections (per Table 3.1) and for each object class over one inspection run (RR-217028). Table 5.5 only shows a single statistic calculated from one inspection run to illustrate results in two-dimensions, allowing easy demonstration. Numerically, it was expected that some geohazard sections would not contain any leaning trees due to poorer algorithm performance for that object class and due to foliage-obscuring instances. Regardless, Table 5.5 successfully details differences between object detection rates by geohazard section. For this inspection run, geohazard section 11 had far more identified leaning trees than other geohazard sections, and geohazard sections 1, 2, 3, and 11 had far more hummocks. This illustrates relative differences between the geohazard sections over a single inspection run.

Table 5.5. Rheems-Roy-217028 detected object densities by geohazard section index.

Geohazard Section Index	Leaning Tree	Moisture	Hummock	Rock Face
1	0.053	0.816	0.605	1.553
2	0.148	0.222	0.593	2.037
3	0.040	0.280	1.200	2.080
4	0.000	0.556	0.444	1.111
5	0.000	0.333	0.000	0.333
6	0.000	0.056	0.222	1.667
7	0.073	0.463	0.317	1.683
8	0.000	0.333	0.000	0.667
9	0.000	0.625	0.250	0.063
10	0.000	0.238	0.143	0.048
11	0.811	0.054	0.838	0.622

To analyze all inspection runs collectively, several statistics were calculated from the detected object density values. Table 5.6 shows the results for the two object classes of primary interest. Statistics were calculated using values from each inspection run for each geohazard section. For example, geohazard section 1 had a mean of 0.06 identified leaning trees per image across all inspection runs showing that geohazard section. Thus, the table summarizes object detection results across all inspection runs by geohazard section. Much like the results shown in Table 5.5, Table 5.6 shows that geohazard sections 1, 2, 3, 10, and 11 had the most identified hummocks per image. Furthermore, geohazard sections 2 and 11 contained the highest density of leaning trees. Observations of these summary statistics show that certain geohazard sections' representative images were more likely to contain slope features of interest across all ROW recordings. Table 5.7 shows the same statistics for moisture and rock face instances.

Table 5.6. Statistics of detected object densities by geohazard section (leaning trees and hummocks).

Geohazard Section Index	LTs Min	LTs Mean	LTs Median	LTs Max	Hs Min	Hs Mean	Hs Median	Hs Max
1	0.00	0.06	0.05	0.17	0.48	0.98	0.89	1.70
2	0.00	0.23	0.26	0.54	0.30	0.86	0.93	1.89
3	0.00	0.09	0.07	0.38	0.20	0.93	1.07	1.77
4	0.00	0.05	0.00	0.44	0.18	0.79	0.67	1.67
5	0.00	0.31	0.00	1.50	0.00	0.53	0.67	1.33
6	0.00	0.05	0.00	0.22	0.16	0.50	0.44	1.20
7	0.00	0.05	0.05	0.12	0.16	0.62	0.54	1.17
8	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.33
9	0.00	0.08	0.00	0.44	0.00	0.74	0.84	1.81
10	0.00	0.14	0.05	0.52	0.00	0.90	1.23	1.77
11	0.00	0.29	0.26	0.81	0.34	1.17	1.12	2.10

Table 5.7. Statistics of detected object densities by geohazard section (moisture and rock faces).

Geohazard Section Index	Moisture Min	Moisture Mean	Moisture Median	Moisture Max	Rock Faces Min	Rock Faces Mean	Rock Faces Median	Rock Faces Max
1	0.33	1.04	0.82	2.03	1.10	1.52	1.50	2.54
2	0.14	0.59	0.78	1.04	0.91	1.84	1.93	2.58
3	0.00	0.15	0.12	0.41	0.77	1.68	1.62	2.48
4	0.00	0.55	0.56	1.13	0.50	1.27	1.11	2.82
5	0.00	0.44	0.33	1.00	0.00	0.56	0.33	2.50
6	0.04	0.24	0.16	0.94	0.39	0.82	0.61	1.78
7	0.03	0.52	0.39	1.61	0.78	1.27	1.05	2.25
8	0.00	0.31	0.00	1.50	0.00	0.64	0.25	3.00
9	0.05	0.39	0.40	0.80	0.06	0.38	0.27	1.05
10	0.00	0.20	0.19	0.45	0.05	0.21	0.19	0.37
11	0.04	0.79	0.77	2.11	0.20	0.58	0.54	1.06

Three-dimensional plots were created for both leaning tree and hummock detected instances to visualize the values that were used to calculate the summary statistics shown in Table 5.6. This was done to provide another visualization of detection differences across geohazard sections in addition to inspection

runs. Figure 5.10 shows detected leaning tree densities, and Figure 5.11 shows detected hummock densities. These plots indicate more detected objects in the same group of geohazard sections identified earlier but additionally show detection differences between inspection runs. Geohazard section 11 again stands out as a high-risk section in Figure 5.10, since a higher density of leaning trees was identified than the average across most inspection runs. The same is true for detected hummocks, shown in Figure 5.11 to a slightly lesser extent.
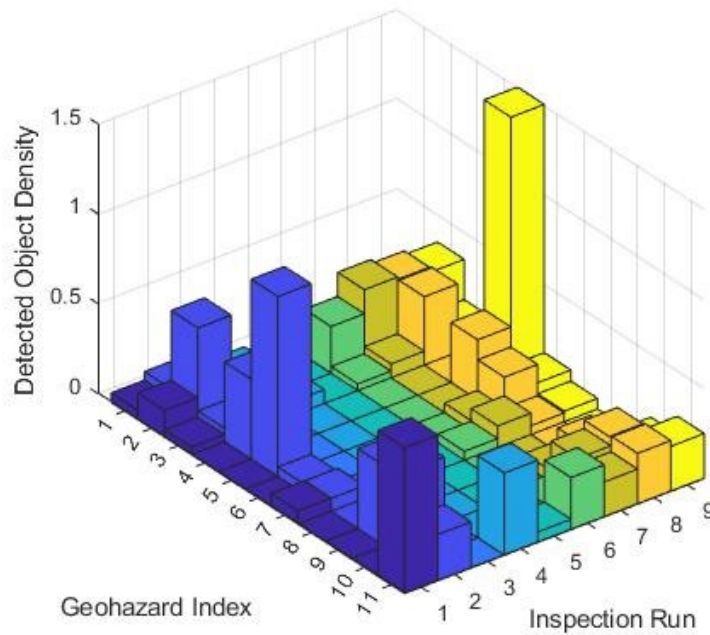

Figure 5.10. Detected leaning tree densities by geohazard section index.
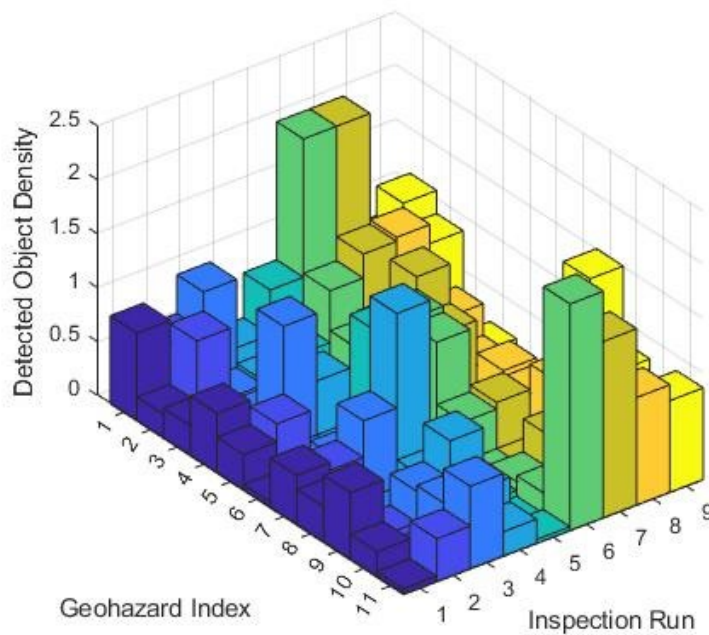
Figure 5.11. Detected hummock densities by geohazard section index.

DEVELOPED LSM MODEL

The performance of classification models is typically evaluated based on the confusion matrix and receiver operating characteristic (ROC) curve. The confusion matrix reports the four possible outcomes of model predictions: (1) true positive (TP), which represents the number of correctly predicted positive samples; (2) true negative (TN), which represents the number of correctly predicted negative samples; (3) false positive (FP), which represents the number of incorrectly predicted positive class; and (4) false negative (FN), which represents the number of incorrectly predicted negative class. Besides precision and recall, four additional performance indicators for classification tasks can be calculated using these four parameters, including accuracy, F_1, true positive rate (TPR), and false positive rate (FPR).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \tag{5.3}$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{5.4}$$

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{5.5}$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \tag{5.6}$$

Note that the F_1 score provides an aggregate measure of model performance score by combining the precision and recall into a single metric. The ROC curve is a 2D plot of FPR vs. TPR for all classification thresholds. The area under the ROC curve (AUC) can be calculated based on the ROC curve, providing an aggregate measure of model performance. AUC is often used as a single-value evaluation for classification models, as it measures the model's capability to distinguish two classes. A no-skill model (i.e., similar to random guessing) will have an AUC score of 0.5, whereas a perfect model will have an AUC score of 1.0. Based on all the evaluation metrics described above, five evaluation metrics were considered in the present study for a comprehensive model performance evaluation, including accuracy, precision, recall, F_1, and AUC.

For the present study, the output of five ML models was the probability of possible outcomes for each sample. A default threshold of 0.5 was used to split the model predictions into two categories for binary classification; subsequently, the model performance evaluation metrics were calculated according to these predicted binary outcomes. Table 5.8 presents the classification performance on the validation dataset based on the fivefold cross-validation procedure. As shown in Table 1, the five ML models achieved an average AUC score of 0.75 with an average accuracy of 0.69 and an average F_1 score of 0.69. These validation scores indicate that the developed ML models can reasonably correlate landslide contributing factors with landslide occurrences. In addition, it can be noted from Table 5.8 that the performance varies among ML models. For example, the DT and the SVM models exhibited the worst performance due to possible overfitting. The LR model and the GBM model exhibited intermediate performance. And the RF model achieved the highest classification performance with AUC scores of 0.8, which can be attributed to the ensemble modeling technique adopted by the RF model. Typically, a classification model with AUC scores above 0.8 can be considered a good classifier. Based on the AUC score, the RF and the GBM models were among the best-performing models, and they were selected as the optimal model for developing the landslide susceptibility map for the study area. Figures 5.12 and 5.13 present the landslide susceptibility map for the study area. It should be noted that applying the developed LSM model to predict landslide risk for the study area can be considered an out-of-sample scenario, as the landslide inventory used to develop these ML models was not adjacent to the study area. The trained ML models' generalization capability needs to be verified for the study area. By comparing Figure 5.12 and Figure 5.13, it can be noted that values of the predicted landslide risk for the study area based on the GBM model were more spread out from zero to one than those predicted by the RF model, which indicated that the GBM model was more suitable than the RF model for the study area. Therefore, the landslide susceptibility map produced by the GBM model was selected for the present study. As shown in Figure 5.13, the developed GBM model was able to identify several high landslide risk locations along the railway line. Figure 5.14 presents the landslide susceptibility along the track line for the study area vs. the identified geohazard sections. As shown in Figure 5.14, landslide risk based on the developed LSM model along the track section generally agreed with identified geohazard sections.

Table 5.8. Cross-validation results for ML models.

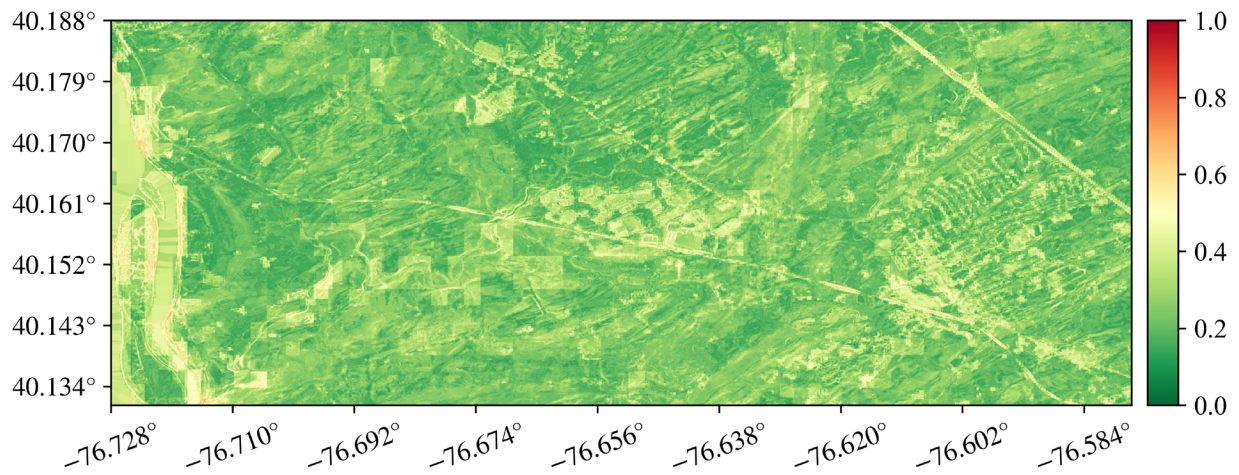| Model	Accuracy	Precision	Recall	F_1	AUC
LR	0.69	0.72	0.63	0.67	0.76
SVM	0.69	0.73	0.59	0.65	0.75
DT	0.66	0.65	0.68	0.66	0.66
RF	0.72	0.72	0.73	0.72	0.80
GBM	0.69	0.69	0.70	0.69	0.77
Avg.	0.69	0.70	0.67	0.68	0.75

Figure 5.12. Landslide susceptibility map for the study area based on the RF model.
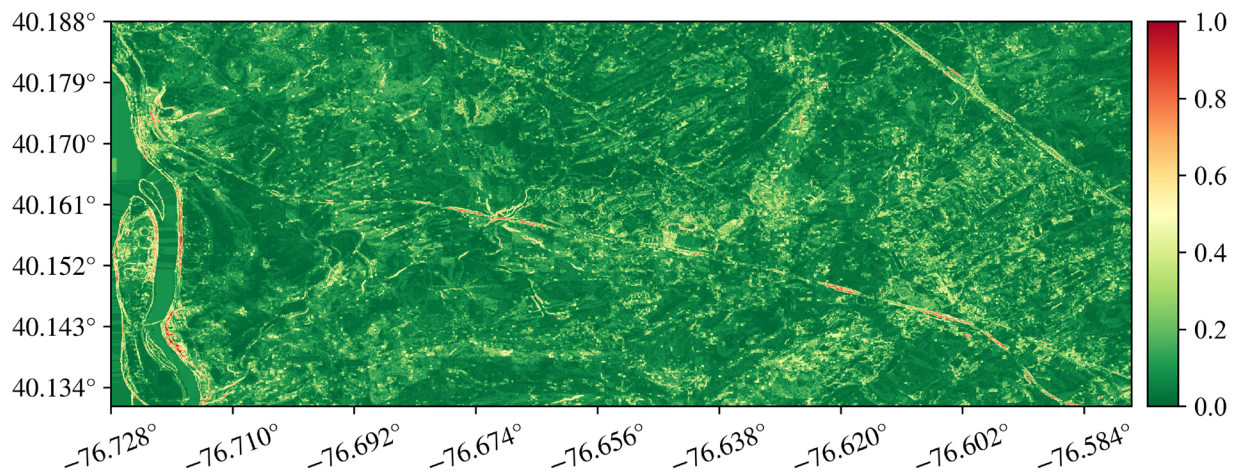

Figure 5.13. Landslide susceptibility map for the study area based on the GBM model.
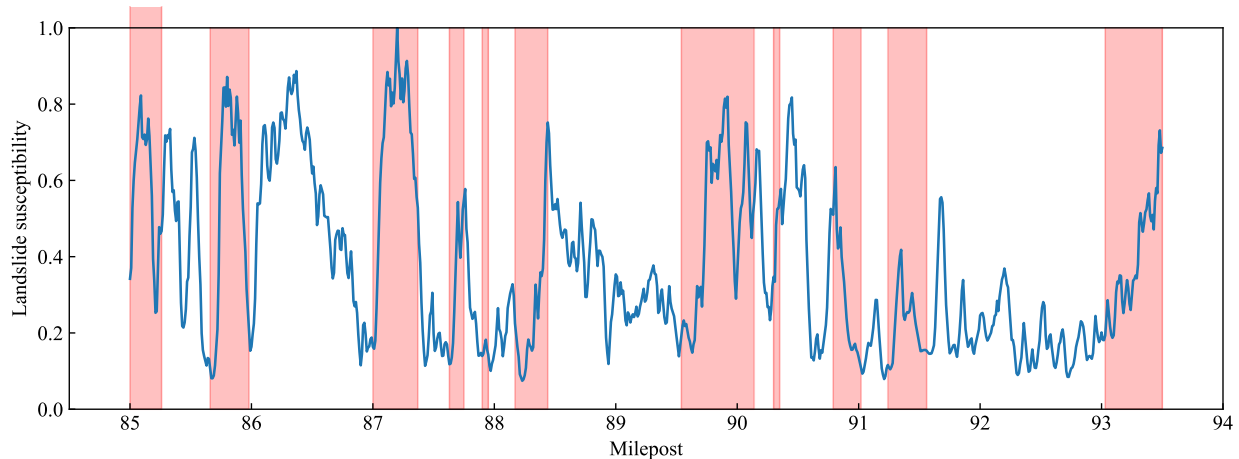
Figure 5.14. Landslide susceptibility along the railway track vs. identified geohazard sections (red shaded area).

RELATIVE RISK ASSESSMENT

In this study, to assess the risk of each identified geohazard section relatively, feature detection results using ROW images and landslide location likelihood results using geospatial variables and landslide location history were combined. Both parameters were unionized by taking the unweighted square sum, thus assuming both parameters equally indicated risks for a particular geohazard section. Prior to calculating the unionized relative risk index, both methods required a numerical result ranging from zero to one, thus indicating a scaled risk index individually. This approach results in a unionized relative risk index that is also scaled from zero to one.

Several approaches were considered to quantify relative risk using the right-of-way images. As shown in the previous chapter, the density of detected objects varied between geohazard sections and inspection runs. This was true for all object classes, but only leaning trees and hummocks were indicative of geohazard sections, as shown in Table 5.3. Thus, it was determined that relative risk should be calculated using detection results for these two object classes and not all trained object classes. Furthermore, slope stability analysis shows that observation of these two feature classes is more likely to precede slope failure (Li et al., 2002).

Detected object density calculations for both of these object classes show clearly that certain geohazard sections were more likely to contain features of interest. However, the scale of these differences varies by inspection run. Furthermore, no discernable pattern based on inspection dates was identified. Thus, combining each inspection run to provide a single metric that is indicative of relative risk per geohazard section was determined to be the best approach. Several statistics calculated from object density values were proposed in Table 5.6 and Table 5.7 and were considered to develop the risk assessment approach. In addition, other calculated metrics from the object detection results were also considered. The count of objects, for example, indicates how many objects are present in a geohazard section but has the negative impact of not scaling counts to the size of the geohazard section, as discussed earlier.

Additionally, the prediction accuracy from the detection procedure has not been discussed in depth

to this point. As introduced earlier, detected objects are identified by feature class, bounding box location, and prediction accuracy (the confidence of the algorithm that the proposed image region contains a feature of the specified class). A low prediction threshold was chosen to yield a larger population of detected objects considering the scope of the dataset used to develop the algorithm. A leaning tree predicted with 100% confidence is more interesting than the one with 40% confidence. However, future improvements to an object detection algorithm would bolster its predictive capabilities and thus would not require a low threshold. Detected features in this study are binary in nature, either present or not, and prediction accuracy does not represent severity in this use case. For these reasons, the accuracy metric should not be used as the primary assessment method.

To further investigate this, Figure 5.10 and Figure 5.11 were recreated to consider the object detector's predictive confidence. The object density parameter introduced earlier was adjusted using Eq. 5.7 such that each identified feature was multiplied element-wise with the detector's confidence score. Thus, Figure 5.15 and Figure 5.16 show the object density values scaled to the feature confidence scores (where features detected with higher confidence carry more weight). Though some differences were identified in several density values, the resulting plots look essentially the same as their counterparts. Due to this and the reasons mentioned earlier, predictive confidence was not used to assess geohazards for relative risk using the right-of-way images.

$$Ratio\_adjusted_j = \frac{1}{I_G} \sum_{i=1}^{I_G} score_{i,j} \cdot object_{i,j} \qquad (5.7)$$
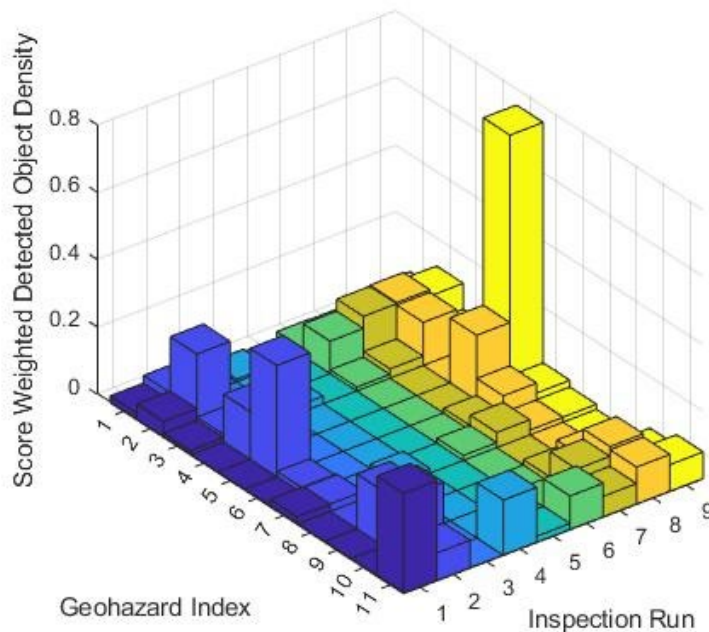

Figure 5.15. Detected leaning tree score-adjusted densities by geohazard section index.
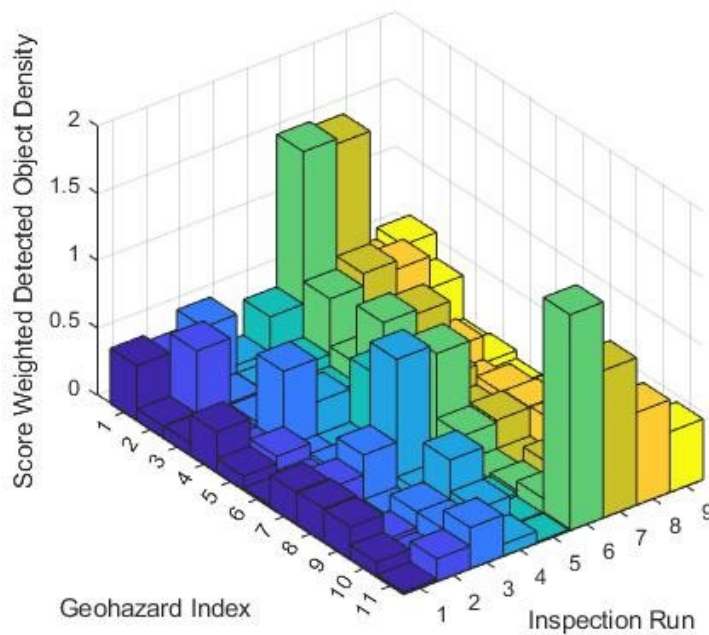
Figure 5.16. Detected hummock score-adjusted densities by geohazard section index.

After considering alternative metrics to assess the geohazards, it was determined that the detected object density metric is satisfactory. As previously mentioned, the risk was assessed across identified geohazard sections only and not over time. This is primarily due to inconsistent inspection run timing and identified differences in the inspection runs (such as the weather and seasonal effects dependent on the recording date) that resulted in varied object detector performance. Additionally, the research presented in this report did not contain a more rigorous investigation into change detection over time for individual objects or the number of objects. Thus, the evolution of features over time was not considered when developing the risk assessment approach. These points and others are discussed further in the final chapter of this report.

It was determined that detected object density values would be used to quantify relative risk. However, using the maximum calculated density value would put too much weight into the detector's predictions on one inspection run. Furthermore, while eliminating outlier inspection runs, the median calculated density value does not consider each inspection run equally. Thus, using the mean of the calculated density values was determined to be the best approach. The calculated mean values (shown in Table 5.6) were scaled to instead fall between zero and one to weigh both the detector's performance on hummocks and leaning trees equally. Afterward, the average of the two detected object density mean values was calculated across geohazard section indices. These resulting values define the relative risk of each geohazard section calculated by assessing the right-of-way images and are shown in Table 5.9. The calculated risk indices show apparent detection differences between the eleven identified geohazard sections. The risk index indicates geohazard section 11 is of the highest concern, while geohazard sections

2, 5, and 10 are also of concern. Interestingly, this collection of geohazard sections makes up some of the tallest in the study area per Table 3.1, providing some confirmation to this assessment method.

Table 5.9. Relative risk assessment of geohazard sections using ROW images.

Geohazard Section Index	Leaning Trees (scaled mean)	Hummocks (scaled mean)	Risk Index (average)
1	0.19	0.82	**0.51**
2	0.73	0.72	**0.72**
3	0.29	0.78	**0.53**
4	0.16	0.65	**0.40**
5	1.00	0.40	**0.70**
6	0.16	0.37	**0.27**
7	0.17	0.48	**0.32**
8	0.00	0.00	**0.00**
9	0.24	0.60	**0.42**
10	0.43	0.75	**0.59**
11	0.92	1.00	**0.96**

Based on the relative risk assessment of geohazard sections using ROW images and the object detection model, leaning trees and hummocks were identified as suitable indicators for geohazards along the railway section. Using Eq. 5.7, the score-weighted object densities for both leaning trees and hummocks were combined and calculated. Figure 5.17 presents the score-adjusted density vs. identified geohazard section. As shown in Figure 5.17, the score-adjusted density based on the object detection model generally agreed with the identified geohazard sections. The trained object detection and LSM models can be used to evaluate slope failure risk along the railway line. However, the object detection model using ROW images could not consider geophysical conditions related to slope stability, and the LSM model cannot produce detailed predictions along the railway line due to coarse spatial resolution. Therefore, a relative assessment for geohazards along the track section is necessary to combine the results from both the object detection model and the LSM model so that each method can complement the other method. For the present study, the combined risk scores for each milepost were obtained using outputs from the object detection model multiplied by outputs from the LSM model, as shown in Figure 5.18. It can be noted from Figure 5.18 that the combined risk scores were able to yield better agreement with identified geohazard sections compared to each model on their own (see Figures 5.16 and 5.17).
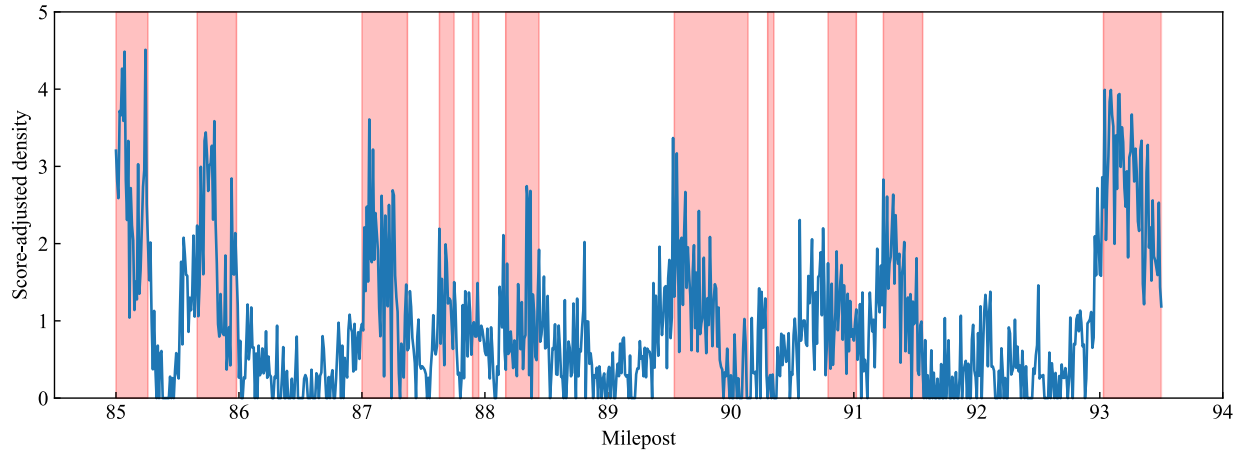
Figure 5.17. Score-adjusted density based on object detection model along the railway track vs. identified geohazard sections (red shaded area).
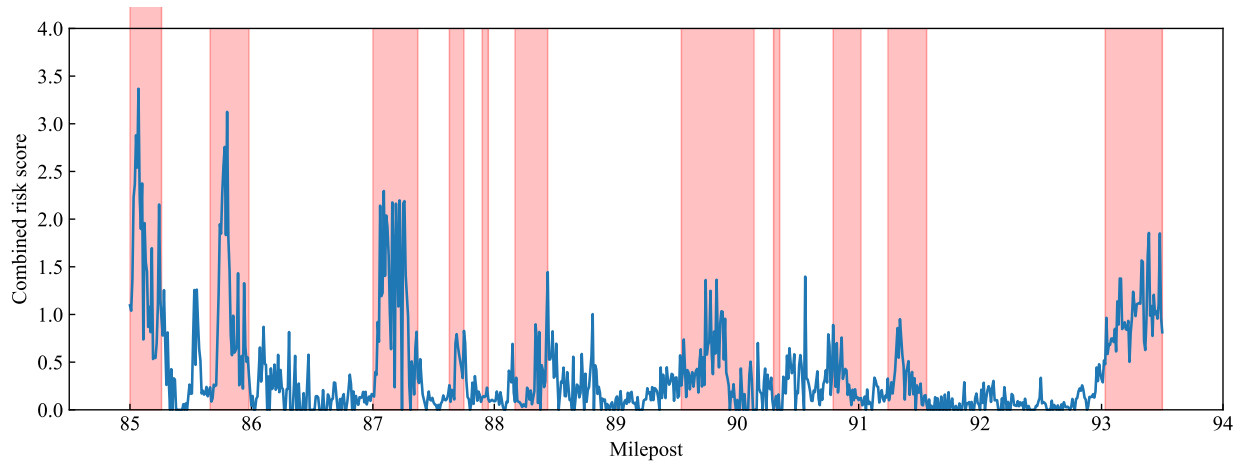

Figure 5.18. Combined risk score along the railway track vs. identified geohazard sections (red shaded area).

CHAPTER 6
Concluding Remarks

INTRODUCTION

The research detailed in this paper shows that geohazard sections along railways created by through and side-cuts can be assessed for relative risk using right-of-way videos and remote sensing data. Notably, risk assessment was performed without on-site measurements of the geohazards. The model framework proposed satisfactorily assessed geohazards in the study area, determining the most at-risk ones that should be prioritized for further measurement and potential remediation efforts.

SUMMARY OF RESEARCH

The research conducted herein aimed to investigate methods of assessing geohazards along railroad rights-of-way created via cuts using videos recorded by inspection videos and publicly available data. First, the right-of-way videos were pre-processed and explored to detail the scope of available data. It was determined that transfer learning could be used to train an object detection algorithm to identify unstable slope features of interest from extracted ROW images. Furthermore, a landslide susceptibility model was developed to identify slope failure risks by comparing geospatial parameters to the locations of historical landslides. Using the class, location, and time of detected features in the ROW images injunction with the identified slope failure risks allowed geohazards to be relatively assessed.

Without on-site measurements detailing the soil composition, stress-strain, and slope geometry of identified geohazards, conventional slope stability assessment methods could not be feasibly implemented to assess the risk of slope failure. Slope reconnaissance is typically used to identify slope features indicative of instability when assessing geohazards to support on-site measurement. This study performed slope reconnaissance by identifying features of interest in ROW images. This was done by encasing features of interest with bounding boxes, which allows the coordinates and class of features of interest to be encoded in a select set of ROW images. Using the labeled "training images," an object detection algorithm was trained to identify features of interest in unlabeled ROW images. Thus, slope reconnaissance was algorithmically performed on all ROW images.

For the present study, geospatial parameters for the study area were collected, representing hillslope geometries, surface hydrology, and soil conditions. In addition, a landslide inventory was developed based on historical landslides adjacent to the study area and within similar ecoregions. Based on the geospatial parameters and landslide inventory, an LSM model was developed using ML techniques to predict landslide susceptibility for the study area. Based on the result analysis, the landslide risk predicted by the developed LSM model was generally consistent with the identified geohazard sections.

Using both the detected features and geospatial parameters, a large dataset was created that summarizes the study area. Moreover, a relative risk assessment framework was developed to evaluate slope failure risk across the study area by combining the results from both the object detection model and the LSM model. It should be noted that every calculated or retrieved parameter is dependent on the time and location of measurement. Furthermore, object detection parameters are dependent on the list of possible class labels. Thus, each identified parameter must be considered in multiple dimensions to fully visualize its scope.

FUTURE RESEARCH

The research presented in the report details a framework to assess geohazards for risk relatively, showing promise for a limited dataset. Since the study area only covers about 15 miles of track, the object detection and landslide susceptibility models are biased on the landscape and topography of the area where data were recorded. Furthermore, the size of the dataset limits the scope of the presented results.

The second phase of this research effort will involve acquiring further data to refine the framework steps presented herein and to further develop the risk assessment methodology. The object detection model will be retrained using more ROW training and validation images to improve its accuracy. Furthermore, additional ROW images will require manual labeling. Since bounding boxes are not the most precise way to encase objects of interest, more sophisticated approaches such as semantic segmentation (in which the pixels associated with an object are identified rather than a region of an image) will be considered (Wang et al., 2018). As part of this process, slope reconnaissance will be revisited to determine if additional labels should be included to diversify the object detectors' capabilities. With a more accurate object detector, objects will be predicted with higher confidence and will additionally be identified with higher frequency.

For the landslide susceptibility models, the current methods only considered landslides that occurred adjacent to the study area, which may not be sufficient to represent landslides that occurred along the railway cut sections (e.g., some of the landslides happened on natural terrain). In order to improve LSM models' performance to predict landslide risk along the railway cut sections, additional landslide records need to be obtained. Ideally, these landslide locations should be along the railway lines. Moreover, the benefits of additional geospatial parameters on landslide risk prediction should be explored, such as collecting geospatial parameters containing additional features or different spatial resolutions.

Several other avenues to study the right-of-way images will be considered to develop the risk assessment framework further. First, additional information will be gathered from the right-of-way images beyond detecting features of interest. It has been hypothesized that the height of the slope in geohazard sections can be approximated. This can be used to provide more localized information about the study area beyond what is provided by geospatial data. Additionally, more comprehensive methods of studying identified features of interest have been proposed. While it is interesting to identify local features in geohazards, the severity of individual features has not been studied. Additionally, the evolution of features over time, which can be indicative of soil creep, for example, has not been explored. As more measurements of the geohazards are made, the risk framework will become more comprehensive and accurate in assessing slope stability.

After revising each portion of the framework proposed herein, the model will be tested on a wider

CIAMTIS
U.S. DOT Region 3 University Transportation Center

variety of geohazards to determine its accuracy and effectiveness. The authors are confident that the model is of sound design and will be an accurate tool for geohazard assessment in real-time using only publicly available data and ROW recordings. With this model, railroads will be able to optimize measurement and remediation efforts to only the most at-risk geohazards along their rights-of-way.

References

Agliardi, F., & Crosta, G. B. (2003). "High resolution three-dimensional numerical modelling of rockfalls." *International Journal of Rock Mechanics and Mining Sciences*, 40(4), 455–471. https://doi.org/10.1016/S1365-1609(03)00021-2.

Ai, X., M. Sheng, X. Su, S. Ai, X. Jiang, S. Yang, Z. Huang, and Y. Ai. (2021). "Effects of frame beam on structural characteristics of artificial soil on railway cut-slopes in southwestern China." Land Degrad. Dev., 32 (1): 482–493. https://doi.org/10.1002/ldr.3719.

Alvioli, M., Santangelo, M., Fiorucci, F., Cardinali, M., Marchesini, I., Reichenbach, P., Rossi, M., Guzzetti, F., & Peruccacci, S. (2021). "Rockfall susceptibility and network-ranked susceptibility along the Italian railway." *Engineering Geology*, 293 (106301). https://doi.org/10.1016/j.enggeo.2021.106301.

Burges, C.J. (1998). "A tutorial on support vector machines for pattern recognition." *Data Mining and Knowledge Discovery*, 2(2), pp.121-167.

Barthelme, S. (2021). *imager: Image Processing Library Based on "CImg". R package version 0.42.11* (R package version 0.42.7). https://cran.r-project.org/package=imager.

Boussik, A., W. Ben-Messaoud, S. Niar, and A. Taleb-Ahmed. (2021). "Railway obstacle detection using unsupervised learning: An exploratory study." *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE.

Chen, Z., W. Chen, C. Li, Y. Pu, and H. Sun. (2016). "Effects of polyacrylamide on soil erosion and nutrient losses from substrate material in steep rocky slope stabilization projects." *Sci. Total Environ.*, 554–555: 26–33. https://doi.org/10.1016/j.scitotenv.2016.02.173.

Duncan, J. M., and S. G. Wright. (1980). "The accuracy of equilibrium methods of slope stability analysis." *Eng. Geol.*, 16 (1–2): 5–17. https://doi.org/10.1016/0013-7952(80)90003-4.

Duncan, J. M. (1996). "State of the art: Limit equilibrium and finite-element analysis of slopes." *J. Geotech. Eng.*, 122 (7): 577–596. https://doi.org/10.1061/(asce)0733-9410(1996)122:7(577).

Dorren, L. K. A. (2003). "A review of rockfall mechanics and modelling approaches." *Prog. Phys. Geogr.*, 27 (1): 69–87. https://doi.org/10.1191/0309133303pp359ra.

Dasiopoulou, S., V. Mezaris, I. Kompatsiaris, V.-K. Papastathis, and M. G. Strintzis. (2005). "Knowledge-assisted semantic video object detection." *IEEE Trans. Circuits Syst. Video Technol.*, 15 (10): 1210–1224. https://doi.org/10.1109/tcsvt.2005.854238.

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. (2009). "ImageNet: A large-scale hierarchical image database." *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE.

Federal Railroad Administration. (2002). *Office of Safety Analysis Web Site*. United States Department of Transportation. http://safetydata.fra.dot.gov/OfficeofSafety/

CIAMTIS
U.S. DOT Region 3 University Transportation Center

r3utc.psu.edu

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., and Ardizzone, F. (2005). "Probabilistic landslide hazard assessment at the basin scale." *Geomorphology*, 72(1-4), 272-299.

Guo, F., Y. Qian, and Y. Shi. (2021a). "Real-time railroad track components inspection based on the improved YOLOv4 framework." *Autom. Constr.*, 125 (103596): 103596. https://doi.org/10.1016/j.autcon.2021.103596.

Guo, F., Y. Qian, Y. Wu, Z. Leng, and H. Yu. (2021b). "Automatic railroad track components inspection using real-time instance segmentation." *Comput.-aided civ. infrastruct. eng.*, 36 (3): 362–377. https://doi.org/10.1111/mice.12625.

Haddad, R. A., and A. N. Akansu. (1991). "A class of fast Gaussian binomial filters for speech and image processing." *IEEE Trans. Signal Process.*, 39 (3): 723–727. https://doi.org/10.1109/78.80892.

Hatheway, A. W. (1996). "Slope stability and stabilization methods." *Environ. Eng. Geosci.*, II (3): 447–449. https://doi.org/10.2113/gseegeosci.ii.3.447.

He, K., X. Zhang, S. Ren, and J. Sun. (2016). "Deep residual learning for image recognition." *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B. (2017). "SoilGrids250m: Global gridded soil information based on machine learning." PloS one, 12(2), e0169748.

Kuhn, M., and Johnson, K. (2013). *Applied predictive modeling*. Springer New York, New York, NY.

Li, D., Hyslip, J., Sussmann, T., & Chrismer, S. (2002). Railway geotechnics. In *Railway Geotechnics*. CRC Press Taylor & Francis Group. https://doi.org/10.1201/b18982.

Lu, N., and Godt, J. (2013). *Hillslope hydrology and stability*. Cambridge University Press, Cambridge.

McMahon, G., S. M. Gregonis, S. W. Waltman, J. M. Omernik, T. D. Thorson, J. A. Freeouf, A. H. Rorick, and J. E. Keys. (2001). "Developing a spatial framework of common ecological regions for the conterminous United States." *Environ. Manage.*, 28 (3): 293–316. https://doi.org/10.1007/s0026702429.

Moradi, S., Huisman, J., Class, H., and Vereecken, H. (2018). "The Effect of Bedrock Topography on Timing and Location of Landslide Initiation Using the Local Factor of Safety Concept." *Water*, 10(10), 1290.

Mathworks. (2020). Deep Learning Toolbox™ R2020a. In *Mathworks* (No. 2021a). www.mathworks.com.

Mirus, B. B., Jones, E. S., Baum, R. L., Godt, J. W., Slaughter, S., Crawford, M. M., Lancaster, J., Stanley, T., Kirschbaum, D. B., Burns, W. J., Schmitt, R. G., Lindsey, K. O., and McCoy, K. M. (2020). "Landslides across the USA: occurrence, susceptibility, and data limitations." *Landslides*, 17(10), 2271–2285.

Ma, Z., Mei, G., and Piccialli, F. (2021). "Machine learning for landslides prevention: a survey." *Neural Computing & Applications*, 33(17), 10881–10907.

Maxwell, A. E., M. Sharma, J. S. Kite, K. A. Donaldson, S. M. Maynard, and C. M. Malay. (2021).

"Assessing the generalization of machine learning-based slope failure prediction to new geographic extents." *ISPRS Int. J. Geoinf.*, 10 (5): 293. https://doi.org/10.3390/ijgi10050293.

Natekin, A., and Knoll, A. (2013). "Gradient boosting machines, a tutorial." *Frontiers in Neurorobotics*, 7(7) 1–21.

Oksuz, K., B. C. Cam, S. Kalkan, and E. Akbas. (2021). "Imbalance problems in object detection: A review." *IEEE Trans. Pattern Anal. Mach. Intell.*, 43 (10): 3388–3415. https://doi.org/10.1109/TPAMI.2020.2981890.

Pratt, W. K. (1994). Digital Image Processing. In *European Journal of Engineering Education* (Fourth, Vol. 19, Issue 3). Pearson. https://doi.org/10.1080/03043799408928319.

Powers, D. M. W. (2007). "Evaluation : From Precision , Recall and F-Factor to ROC , Informedness , Markedness & Correlation." *Journal of Machine Learning Technologies*, *2*(December), 37–63.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research: JMLR*, 12(85), 2825–2830.

Pradhan, A. M. S., Lee, S. R., & Kim, Y. T. (2019). "A shallow slide prediction model combining rainfall threshold warnings and shallow slide susceptibility in Busan, Korea." *Landslides*, *16*(3), 647–659. https://doi.org/10.1007/s10346-018-1112-z.

Qi, C., and Tang, X. (2018). "Slope stability prediction using integrated metaheuristic and machine learning approaches: A comparative study." *Computers & Industrial Engineering*, 118, 112–122.

R Development Core Team. (2003). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

Resendiz, E., J. Hart, and N. Ahuja. (2013). "Automated visual inspection of railroad tracks." *IEEE Trans. Intell. Transp. Syst.*, 14 (2): 751–760. https://doi.org/10.1109/tits.2012.2236555.

Ren, S., K. He, R. Girshick, and J. Sun. (2017). "Faster R-CNN: Towards real-time object detection with region proposal networks." *IEEE Trans. Pattern Anal. Mach. Intell.*, 39 (6): 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Reichenbach, P., Rossi, M., Malamud, B., Mihir, M., and Guzzetti, F. (2018). "A review of statistically-based landslide susceptibility models." *Earth-Science Reviews*, 180, 60-91.

Stone, M. (1974). "Cross-validatory choice and assessment of statistical predictions." *Journal of the Royal Statistical Society*, 36(2), 111–133.

Tsangaratos, P., and Ilia, I. (2015). "Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece." *Landslides*, 13(2), 305-320.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer New York, New York, NY.

Velten, J., A. Kummert, and D. Maiwald. (1999). "Image processing algorithms for video-based real-time railroad track inspection." *42nd Midwest Symposium on Circuits and Systems (Cat. No.99CH36356)*. IEEE.

Wu, Y.-M., H.-X. Lan, X. Gao, L.-P. Li, and Z.-H. Yang. (2015). "A simplified physically based coupled

rainfall threshold model for triggering landslides." *Eng. Geol.*, 195: 63–69. https://doi.org/10.1016/j.enggeo.2015.05.022.

Wang, Q., Wang, Y., Niu, R., and Peng, L. (2017). "Integration of Information Theory, K-Means Cluster Analysis and the Logistic Regression Model for Landslide Susceptibility Mapping in the Three Gorges Area, China." *Remote Sensing*, 9(9), 938.

Wang, P., P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell. (2018). "Understanding Convolution for Semantic Segmentation." *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE.

Wang, Y., Fang, Z., and Hong, H. (2019). "Comparison of convolutional neural networks for landslide susceptibility mapping in Yanshan County, China." *Science of The Total Environment*, 666, 975-993.

Xiao, L., Zhang, Y., and Peng, G. (2018). "Landslide Susceptibility Assessment Using Integrated Deep Learning Algorithm along the China-Nepal Highway." *Sensors*, 18(12), 4436.

Yegnanarayana, B. (1994). Artificial neural networks for pattern recognition. In *Sadhana* (Vol. 19, Issue 2). Clarendon Press ; https://doi.org/10.1007/BF02811896.

Zhang, W., K. Itoh, J. Tanida, and Y. Ichioka. (1990). "Parallel distributed processing model with local space-invariant interconnections and its optical architecture." *Appl. Opt.*, 29 (32): 4790–4797. https://doi.org/10.1364/AO.29.004790.

Ziou, D., & Tabbone, S. (1998). Edge detection techniques - an overview. *Pria*, *8*(4), 1–41. http://nyx-www.informatik.uni-bremen.de/1044/1/ziou_pria_98.pdf.

Zhang, Z., C. Trivedi, and X. Liu. (2018). "Automated detection of grade-crossing-trespassing near misses based on computer vision analysis of surveillance video data." *Saf. Sci.*, 110: 276–285. https://doi.org/10.1016/j.ssci.2017.11.023.

Zaman, A., B. Ren, and X. Liu. (2019). "Artificial intelligence-aided automated detection of railroad trespassing." *Transp. Res. Rec.*, 2673 (7): 25–37. https://doi.org/10.1177/0361198119846468.

Zhou, J., Li, E., Yang, S., Wang, M., Shi, X., Yao, S., and Mitri, H. S. (2019). "Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories." *Safety Science*, 118, 505–518.

Zoph, B., E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le. (2020). "Learning data augmentation strategies for object detection." *Computer Vision – ECCV 2020*, 566–583. Cham: Springer International Publishing.