Report No. UT-26.16

MULTI-SOURCE COMPREHENSIVE PROGRAM TO ASSESS, MONITOR, AND REPORT RETROREFLECTIVITY OF PAVEMENT MARKINGS

Prepared For:

Utah Department of Transportation
Research & Innovation Division

Final Report
December 2025

RESEARCH
Keeping Utah Moving
Utah Department of Transportation - Research Division
4501 South 2700 West - P.O. Box 148410 - SLC,UT 84114-8410

DISCLAIMER

The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the U.S. Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore.

ACKNOWLEDGMENTS

The authors acknowledge the Utah Department of Transportation (UDOT) for funding this research, and the following individuals from UDOT on the Technical Advisory Committee for helping to guide the research:

- Abdul Wakil
- David Stevens
- Kelly Ash
- Chris Whipple
- Radhakrishnan Ukkiramapandian
- Ethan Turner
- Ryan Ferrin
- PJ Roubinet
- Benjamin Maughan
- David Schwartz
- Brad Loveless
- Dave Thomas
- David Bezzant
- Jason Richins
- Kyle Moncur
- Matt Lazanich
- Tim Predmore

TECHNICAL REPORT ABSTRACT

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
UT-26.16	N/A	N/A

4. Title and Subtitle
Multi-Source Comprehensive Program to Assess, Monitor, and Report Retroreflectivity of Pavement Markings

5. Report Date
December 2025

6. Performing Organization Code
N/A

7. Author(s)
Abbas Mohammadi, Juan C. Medina, Abbas Rashidi

8. Performing Organization Report No.
N/A

9. Performing Organization Name and Address
University of Utah
Department of Civil and Environmental Engineering
201 Presidents' Circle
Salt Lake City, UT 84112

10. Work Unit No.
5H094 28H

11. Contract or Grant No.
25-8077

12. Sponsoring Agency Name and Address
Utah Department of Transportation
4501 South 2700 West
P.O. Box 148410
Salt Lake City, UT 84114-8410

13. Type of Report & Period Covered
Final Report
Jul 2024 to Dec 2025

14. Sponsoring Agency Code
PIC No. UT24.205

15. Supplementary Notes
Prepared in cooperation with the Utah Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration

16. Abstract

This research investigates the feasibility of estimating retroreflectivity of durable longitudinal pavement markings using multi-source data to reduce UDOT's reliance on costly specialized field inspections. The study uses a combination of existing datasets, mobile LiDAR point clouds, street-level roadway imagery, and historical retroreflectivity measurements to develop robust predictive models for assessing pavement marking condition.

The research uses existing field retroreflectivity data across major Utah roadways, including I-15, I-80, I-84, and I-215. The study corridors are dominated by durable pavement markings, so the model development and validation presented in this report primarily reflect the performance of durable marking materials. LiDAR data provided geometry and intensity of light reflected back to the sensing unit, while roadway images were processed to extract visual features and assess the condition of lane markings. Both data streams (i.e., LiDAR and imagery) were analyzed independently and also in combination using machine learning methods, including Random Forest classifiers and convolutional neural networks (CNNs). A fusion model was developed to integrate LiDAR and image-based features for enhanced performance. Results demonstrated that image- and LiDAR-only models can accurately classify pavement markings as compliant or non-compliant with federal minimum retroreflectivity standards. The fusion model outperformed individual models, achieving over 95% accuracy and strong agreement with ground truth data at both segment and corridor levels. Regression models were also developed to predict continuous retroreflectivity values directly from image features. The methods developed in this research were implemented in a portable, user-friendly computer tool that enables UDOT to estimate and monitor pavement marking retroreflectivity via data fusion, LiDAR-only, or image-only models. This approach enhances operational efficiency, supports monitoring for federal compliance, and provides a scalable framework for statewide deployment. This study builds on a previous UDOT research phase that evaluated LiDAR-only retroreflectivity estimation and extends that work by incorporating street-level imagery and multi-source data fusion.

17. Key Words
Pavement marking retroreflectivity, LiDAR intensity, image processing, data fusion, machine learning, random forest classifier, convolutional neural network.

18. Distribution Statement
Not restricted. Available through:
UDOT Research & Innovation Div.
4501 South 2700 West
P.O. Box 148410
Salt Lake City, UT 84114-8410
www.udot.utah.gov/go/research

23. Registrant's Seal
N/A

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	72	N/A

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

UNIT CONVERSION FACTORS

SI* (MODERN METRIC) CONVERSION FACTORS				
APPROXIMATE CONVERSIONS TO SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa
APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. (Adapted from FHWA report template, Revised March 2003)

LIST OF ACRONYMS

CNN	Convolutional Neural Network
CSV	Comma-Separated Values
FHWA	Federal Highway Administration
GIS	Geographic Information System
LEL	Left Edge Line (Continuous Pavement Marking)
LiDAR	Light Detection and Ranging
LL	Lane Line (Dashed Pavement Marking)
mcd/m²/lux	Millicandelas per square meter per lux
MUTCD	Manual on Uniform Traffic Control Devices
RANSAC	Random Sample Consensus
REL	Right Edge Line (Continuous Pavement Marking)
ReLU	Rectified Linear Unit
RGB	Red, Green, Blue
RMSE	Root Mean Squared Error
SMOTE	Synthetic Minority Over-Sampling Technique
UDOT	Utah Department of Transportation

EXECUTIVE SUMMARY

Maintaining durable longitudinal pavement marking retroreflectivity at federally mandated levels is essential for nighttime roadway safety. Traditional methods for assessing retroreflectivity rely heavily on field-based data collection, which can be time-consuming, resource-intensive, and operationally burdensome. The Utah Department of Transportation (UDOT) currently uses mobile retroreflectometers to monitor longitudinal pavement markings, but increasing network size and regulatory pressures, particularly with the Federal Highway Administration's (FHWA) 2022 ruling requiring minimum retroreflectivity standards, create a need for more scalable, cost-effective assessment strategies.

This research addresses that need by developing and evaluating a multi-source retroreflectivity estimation tool that uses routinely collected datasets, specifically mobile LiDAR point clouds and street-level roadway imagery, to predict pavement marking performance. The approach significantly reduces reliance on direct field measurements by leveraging existing data while maintaining high accuracy and reliability. This work builds on UDOT's earlier LiDAR-only retroreflectivity research by adding image-based methods and integrated LiDAR-image fusion models.

Three key data sources were used:

- Field retroreflectivity measurements (ground truth),
- Mobile LiDAR data, collected by Pathway Services Inc.,
- Street-level roadway images, collected by Pathway Services Inc.

The research used the following methodologies for data pre-processing, machine-learning model development, and spatial integration:

- LiDAR-Based Estimation: Statistical summaries of LiDAR point cloud intensity and reflectance were used as input features for Random Forest models to classify compliance status.

- Image-Based Estimation: Roadway images were processed using a custom convolutional neural network with residual blocks (see Section 2.2.2). This architecture, trained from scratch on the UDOT imagery, predicts whether pavement markings are compliant (retroreflectivity >100 mcd/m²/lux) or non-compliant based on learned visual features.

- Fusion Models: Image-derived and LiDAR-based features were combined to train enhanced classification models, improving predictive performance.

Overall, model results demonstrated strong alignment with ground truth data:

- Image-only models achieved test accuracies above 84% across most highways.

- LiDAR-only models performed especially well where pavement markings were clearly captured in the point cloud (e.g., without occlusion along the REL), often outperforming image-based models in low-visibility conditions.

- Fusion models consistently outperformed individual models (i.e., only LiDAR or only images), with classification accuracies exceeding 95% in some corridors (e.g., I-84 REL Increasing). Cluster-level agreement (1-mile segments) between predictions and ground truth reached up to 97% in the best cases.

Beyond classification, a regression analysis was conducted to predict continuous retroreflectivity values from image features. This analysis supports granular degradation modeling and prioritization of maintenance activities. Results showed a strong correlation between predicted and true values, with R^2 scores reaching over 0.85 for whole corridors.

Although the image analysis results are promising, image data does not directly measure reflected light as LiDAR does. Instead, image-based models detect visual patterns, such as treatment condition or line integrity, that correlate with retroreflectivity. Image-only models should be interpreted with more caution, but they also carried valuable information to complement LiDAR in the fusion approach.

Importantly, the data processing pipelines developed during this project, covering image extraction, feature engineering, model training, and spatial aggregation, are fully automated and scalable.

A computer tool with all dependencies embedded in an executable format was also developed to implement the proposed methods and to provide future use of the models as new data is gathered over time. The tool generates outputs in user-friendly formats such as Excel tables, GIS shapefiles, and graphical reports, all intended for seamless integration into UDOT's existing workflows.

Key Findings:

- A machine learning–based tool can accurately classify and estimate pavement marking retroreflectivity using existing datasets, reducing the need for extensive field measurements.
- Fused models leveraging both LiDAR and imagery deliver the highest accuracy and reliability.
- Cluster-level aggregation of predictions provides actionable insights for maintenance planning.
- The methodology is scalable, portable, and compatible with UDOT's current data infrastructure.

In summary, this research demonstrates the feasibility of a scalable, data-driven approach to pavement marking retroreflectivity assessment, offering UDOT a robust alternative to traditional methods, one that supports compliance, safety, and operational efficiency.

1.0 INTRODUCTION

1.1 Problem Statement

Maintaining adequate retroreflectivity levels in longitudinal pavement markings is critical for roadway safety, especially during nighttime and low-visibility conditions (National Safety Council, 2024). Ensuring that pavement markings meet minimum retroreflectivity standards supports safe driving conditions and aligns with recent federal regulations, including the Federal Highway Administration (FHWA) ruling (87 FR 47921), which mandates minimum retroreflectivity levels to be maintained by transportation agencies (Federal Highway Administration, 2022). In this project, the focus is on durable longitudinal pavement markings, which constitute the majority of markings along selected corridors and are of particular interest for maintenance planning.

Currently, the Utah Department of Transportation (UDOT) relies on a combination of contracted services and internal data collection efforts to monitor pavement marking retroreflectivity (Utah Department of Transportation). This process requires the use of specialized equipment and workforce, as well as substantial financial resources (Satterfield, 2014). Furthermore, the need for recurring, time-intensive field data collection presents an operational burden, both in terms of costs and logistics. As UDOT's roadway network continues to grow and the FHWA compliance deadline of September 2026 approaches, there is a pressing need to develop more efficient and scalable methods for assessing and maintaining pavement marking retroreflectivity (Jackels, 2024; Pike et al., 2022).

This research seeks to address this challenge by evaluating whether existing datasets, such as mobile LiDAR scans and street-level imagery, routinely collected for other purposes, can be leveraged to estimate pavement marking retroreflectivity indirectly (Ai & Hennessy, 2022; Brinster et al., 2024; Che et al., 2019; Mohammadi et al., 2024; Zhang, 2016). By developing models that correlate LiDAR intensity and image-based features with retroreflectivity measurements, UDOT has the potential to reduce dependence on specialized field data collection. This approach promises to improve the efficiency of maintenance planning, enhance roadway safety, and ensure compliance with federal standards, while optimizing the use of available resources. This project represents the second phase of UDOT's retroreflectivity research program: The previous phase focused on LiDAR-only estimation, while the present phase evaluates complementary imagery-based methods and multi-source fusion.

1.2 Objectives

The primary objective of this research is to develop a multi-source methodology capable of estimating pavement marking retroreflectivity using existing LiDAR and street-level imagery data, supplemented by historical pavement marking maintenance records (Brinster, 2024; Manasreh et al., 2023). In addition, a user-friendly computer tool is also to be developed to implement the proposed methodology. This generates both current estimates of retroreflectivity, corresponding to the date of the imagery and LiDAR survey, and projections of retroreflectivity degradation over time (Idris et al., 2024). These outputs are intended to support UDOT's pavement-marking maintenance planning directly and to ensure compliance with federal retroreflectivity requirements. The initial development and calibration of the methodology are based on datasets dominated by durable pavement markings, and the findings are therefore most directly applicable to these marking types. The methodology developed here refines and extends LiDAR-only retroreflectivity models from a previous UDOT research phase by integrating imagery-based features and fusion techniques.

The tool will be delivered as a portable, self-contained application that requires no specialized installation or configuration. It will feature an intuitive interface that allows users to input processed LiDAR data, roadway imagery, and maintenance history to produce actionable outputs. The outputs will be available in commonly used formats such as spreadsheets (CSV or Excel), graphical charts, and GIS-compatible files, enabling integration into UDOT's existing workflows. The tool is in part funded by additional support from the CTIPS (Center for Transformative Infrastructure Preservation and Sustainability) university transportation center.

A secondary objective of this research is to equip UDOT with the capability to monitor and track pavement marking retroreflectivity efficiently, reducing reliance on large-scale field data collection. The tool's outputs will support data-driven maintenance scheduling by providing insights into when pavement markings are likely to fall below minimum retroreflectivity standards.

Additionally, this research will develop recommendations for limited, targeted field data collection to support ongoing calibration and validation of the retroreflectivity models. This will ensure the models remain accurate over time while minimizing the need for extensive future fieldwork. These recommended data collection efforts are intended to be lightweight, requiring significantly fewer resources compared to current practices.

1.3 Scope

The scope of this research includes a series of tasks focused on developing a reliable, data-driven tool for estimating the retroreflectivity of durable longitudinal pavement markings using existing datasets. The work involves compiling, processing, and integrating multiple data sources, along with developing analytical models to support retroreflectivity estimation and projection.

The primary datasets utilized in this research include:

- Mobile LiDAR survey data collected along state and federal aid routes, providing high-resolution spatial data and intensity values related to pavement markings.
- Street-level roadway imagery captured concurrently with the LiDAR data, offering visual information for identifying pavement markings and supporting image-based assessments.

The major tasks performed in this research include:

1. Data Compilation and Processing Pipeline Development: Collection and preparation of LiDAR and imagery. This task also includes conducting a literature review to identify best practices and relevant methodologies from similar studies. A data processing pipeline has been developed and tested using sample datasets to ensure scalability and reliability.

2. Model Development and Image Processing: Application of machine learning techniques to process street-level imagery for pavement marking identification and retroreflectivity classification. Existing LiDAR-based retroreflectivity estimation models are being incorporated and enhanced to improve accuracy.

3. Data Integration and Analysis: Development of methods to integrate LiDAR-based estimates with imagery-based assessments. This multi-source approach is designed to increase the reliability and robustness of the retroreflectivity estimates.

This final report documents the completed work, including data processing, model development, LiDAR- and image-based evaluation, multi-source fusion, validation results, and implementation of the methodology in a portable computer tool, and presents findings and recommendations for UDOT deployment.

1.4 Outline of Report

This final report is organized into six chapters and an appendix. Chapter 1 introduces the problem, objectives, and scope of the study. Chapter 2 describes the overall research approach and methods used to evaluate multi-source retroreflectivity and assess pavement marking performance. Chapter 3 summarizes the data sources, data collection, and data preparation steps used in the evaluation. Chapter 4 presents the analysis procedures and results, including compliance evaluation and the key findings derived from the multi-source data. Chapter 5 provides the study conclusions, including limitations and lessons learned. Chapter 6 presents recommendations and an implementation plan describing how the findings and proposed approach can be applied in practice. The appendix contains supporting technical details about the developed tool, supplemental tables/figures, and additional materials referenced in the main text. More specifically, the report is structured in chapters as follows:

1. Introduction
2. Research Methods
3. Data Collection
4. Data Evaluation
5. Conclusions
6. Recommendations and Implementation

2.0 RESEARCH METHODS

2.1 Overview

This chapter presents the technical methodologies applied to develop predictive models for pavement marking retroreflectivity. The LiDAR-based workflow builds directly on the LiDAR-only methods developed in the earlier UDOT research phase, whereas the image-based workflow and fusion framework are new contributions of the current phase. The research is centered around two primary computational workflows: one based on LiDAR point cloud data and another based on roadway imagery analysis (Guan et al., 2016). The chapter also covers the data integration framework that synthesizes results from both methods to produce a comprehensive assessment of retroreflectivity performance.

2.2 Image-Based Retroreflectivity Estimation

2.2.1 Image Processing Workflow

The image-based retroreflectivity estimation workflow begins with pre-processed pavement images. Each image undergoes several steps designed to isolate pavement markings and prepare the data for classification:

- Grayscale Conversion: Images are converted to grayscale to simplify feature extraction based on pixel intensity (Lee & Cho, 2023).
- Road Masking: A polygonal mask is applied to limit analysis to the pavement region and exclude background elements (Kong et al., 2022).
- Intensity-Based Thresholding: High-intensity pixels, which typically represent pavement markings, are isolated using an adaptive threshold set at the 99.85th percentile of pixel brightness within the masked area (Cheng et al., 2020).

- Lane Marking Localization: The centroid of the high-intensity region is computed, and a fixed-size square region is cropped around this point to serve as input for machine learning classification.

2.2.2 Deep Learning Model Development

A Convolutional Neural Network (CNN) with residual connections is used to classify whether the retroreflectivity of a pavement marking meets or falls below the required standard (Chun et al., 2020; Mamun et al., 2022). The model architecture includes:

- Convolutional layers with batch normalization and Rectified Linear Unit (ReLU) activation.
- Residual blocks to improve feature propagation and combat vanishing gradients.
- Max pooling and global average pooling for dimensionality reduction.
- Fully connected layers with dropout for regularization.

To enhance model robustness, data augmentation techniques such as rotation, zoom, translation, and brightness adjustments are incorporated during training. The output is a binary classification:

- Above threshold (compliant)
- Below threshold (non-compliant)

Model performance is evaluated using metrics such as accuracy, confusion matrices, and classification reports on validation and test datasets (Ibne Idris et al., 2024).

2.3 LiDAR-Based Retroreflectivity Estimation

2.3.1 Road Marking Extraction from LiDAR

The LiDAR processing workflow focuses on isolating pavement markings within the point cloud using a multi-stage filtering and geometric analysis process:

1. Sampling Area Definition: A rectangular region is defined within the LiDAR point cloud, centered around expected pavement marking locations.

2. Grid-Based Point Selection: A fine-resolution grid (0.2-foot spacing) is applied within the sampling area to create a detailed spatial cross-section.

3. Intensity Filtering: Points are filtered based on return intensity values. High-intensity points are likely to correspond to pavement markings (Cheng et al., 2022).

4. RANSAC Line Fitting: The RANSAC (Random Sample Consensus) algorithm is used to fit lines through high-intensity point clusters, effectively removing noise and isolating lane lines.

5. Iterative Alignment: Successive sampling areas are dynamically oriented based on the direction of previously detected lane lines, improving extraction accuracy across curved or complex road sections.

6. Continuous Presence Verification: The process checks for the consistent presence of lane markings in the sampling window before moving to the next segment (Cheng et al., 2024).

2.3.2 LiDAR Classification Model

A machine learning model, specifically a Random Forest Classifier, is employed to predict retroreflectivity condition based on LiDAR-derived features such as:

- Average LiDAR point intensity within the extracted marking.
- Average red-channel reflectance value.

The classification targets the same binary output as the image model:

- Above threshold (compliant)
- Below threshold (non-compliant)

To address class imbalance within the dataset, the Synthetic Minority Over-Sampling Technique (SMOTE) is applied during model training. The model produces point-level predictions that are later aggregated for spatial analysis.

2.4 Multi-Source Data Integration

2.4.1 Spatial Aggregation into Clusters

Both LiDAR-based and image-based predictions are spatially aggregated into uniform mileage-based clusters. A typical cluster represents a 1-mile roadway segment. For each cluster, the proportion of predicted "above threshold" markings is calculated to provide an interpretable metric for maintenance decision-making.

This aggregation simplifies the comparison of results across data sources and aligns with operational requirements for route-level assessments.

2.4.2 Cross-Source Comparison Framework

A comparison framework is employed to evaluate the consistency between:

- LiDAR-based predictions
- Image-based predictions

For each cluster, thresholds are used to assign condition categories:

- Green: >60% of markings predicted as compliant.
- Yellow: 40–60% compliant.
- Red: <40% compliant.

This system provides an intuitive visual and quantitative summary of pavement marking conditions across the network.

2.5 Summary

This chapter presented the computational methodologies for estimating pavement marking retroreflectivity using image-based deep learning models and LiDAR-based machine learning classifiers. The chapter also described how results from both sources are aggregated and compared to provide actionable insights for pavement maintenance planning. The data handling and acquisition processes are described separately in the Data Collection chapter.

3.0 DATA COLLECTION

3.1 Overview

This chapter describes the data collection processes undertaken to support the development of retroreflectivity estimation models. Three primary data sources were utilized: field-collected retroreflectivity measurements, mobile LiDAR point cloud data, and street-level roadway imagery. Each dataset was collected for the purpose of extracting features relevant to pavement marking condition assessment and for developing, training, and validating machine learning models, with an emphasis on durable longitudinal markings that make up the majority of the studied routes. Details regarding how, when, and why each dataset was collected are presented below.

It is important to mention that roadway construction and maintenance logs were reviewed to ensure that no new striping was installed on the sections included in this study between the dates the retroreflectivity and LiDAR/imagery datasets were collected, ensuring that the measurements were suitable for the analysis.

3.2 Field Retroreflectivity Measurements (Ground Truth Data)

3.2.1 Data Purpose and Collection Method

The retroreflectivity field measurements serve as the ground truth labels for both LiDAR-based and image-based models. These measurements were collected using a mobile retroreflectometer operated by Beck along selected routes (Lee, 2016). The mobile system measures the retroreflectivity of longitudinal pavement markings in $mcd/m^2/lux$, providing continuous measurements along the road markings (Benz et al., 2009; Migletz et al., 1999). The routes and sections selected for this project correspond primarily to durable longitudinal pavement markings, so the ground truth dataset is dominated by durable marking materials.

Data was collected during the 2023 maintenance inspection cycle, covering multiple routes including Interstate 15 (I-15), Interstate 215 (I-215), Interstate 80 (I-80), and Interstate 84 (I-84), both in increasing and decreasing directions.

3.2.2 Data Structure and Availability

The retroreflectivity data is organized into CSV files, with key columns representing both spatial location and measurement quality. The dataset captures pavement marking retroreflectivity along roadway segments, where each row corresponds to approximately a 0.1-mile segment. The structure allows for precise spatial alignment with complementary datasets such as LiDAR and imagery for further modeling and analysis.

The main columns in the dataset are as follows:

- Chain: Represents the mile marker location (mileage) along the route. This indicates the starting point of the segment for which measurements apply.
- Ave: The average retroreflectivity value for the segment, measured in mcd/m^2/lux.
- Valid: The number of valid measurement points collected within the segment. A higher number indicates more reliable data.
- Inv: The number of invalid measurement points that were excluded from the average. High invalid counts may indicate issues like occlusion, poor surface conditions, or sensor noise.
- Stdev: The standard deviation of retroreflectivity measurements within the segment. This indicates the variability or consistency of the measured retroreflectivity values across that segment.
- GPSLat / GPSLon: The geospatial coordinates (latitude and longitude) representing the approximate central point of the segment. This enables mapping and spatial analysis.
- Treatment: Describes any surface treatment applied to the pavement markings, such as Groove, which refers to markings applied in grooves to enhance durability and visibility.

- Surface Type: Indicates the pavement surface material, such as Concrete, which can influence retroreflectivity performance.

Each record in this dataset represents the average retroreflectivity performance over a short, 0.1-mile roadway segment, supplemented with indicators of data quality (valid, invalid counts, standard deviation) and geospatial context (GPS location, treatment, surface type). This structure ensures robust analysis for understanding the condition of pavement markings and for integrating with other spatial datasets. A sample data table is shown in Table 3.1.

Table 3.1 Field Retroreflectivity Data Sample

Chain	Ave	Valid	Inv	Stdev	GPSLat	GPSLon	Treatment	Surface Type
257.692	362	1055	527	57	40.12544833N	111.65350167W	Groove	Concrete
257.792	324	1606	0	23	40.12624500N	111.65192000W	Groove	Concrete
257.891	322	1627	8	101	40.12706167N	111.65039000W	Groove	Concrete
257.991	158	245	1353	37	40.12810167N	111.64904500W	Groove	Concrete
258.386	360	1543	0	66	40.13343500N	111.64658833W	Groove	Concrete
258.486	374	1544	0	30	40.13488500N	111.64659833W	Groove	Concrete
258.588	401	1519	0	31	40.13635833N	111.64658833W	Groove	Concrete
258.687	432	1492	0	24	40.13779333N	111.64658000W	Groove	Concrete
258.786	407	1474	9	80	40.13923167N	111.64657000W	Groove	Concrete
258.89	424	1481	0	27	40.14074333N	111.64656333W	Groove	Concrete

3.3 LiDAR Data Collection

3.3.1 Purpose and Description of LiDAR Surveys

LiDAR point cloud data was acquired through Pathway's periodic roadway scanning program. The LiDAR dataset consists of dense 3D point clouds with each point containing the following attributes:

- X, Y, Z: Spatial coordinates in a local or projected coordinate system, representing the three-dimensional position of each laser return.

- Intensity: The strength of the laser return signal, which correlates with surface reflectivity and is often indicative of material type or surface treatment (Hodaei et al., 2025).

- Red, Green, Blue: Color reflectance values obtained from imagery co-registered with the point cloud, representing visible spectrum reflectivity. In this study, the red channel is particularly useful for identifying and analyzing pavement markings.

3.3.2 Data Organization and Coverage

LiDAR data was pre-processed into segmented CSV files with each point linked to a milepost (Chain) value. For each mile marker, statistical summaries of pavement marking intensity and color reflectance were calculated, forming the input features for the LiDAR machine learning pipeline. LiDAR data covers the same sections as the retroreflectivity measurements, enabling side-by-side model training and validation. A sample of the LiDAR dataset structure is shown in Table 3.2.

Table 3.2 LiDAR Data Sample

X	Y	Z	intensity	red	green	blue
-228986594	64345389	-438818519	23904	15677	12850	11308
-228987115	64344806	-438817506	14739	15420	13107	11308
-228985704	64346275	-438818412	19174	15677	13364	11565
-228985229	64346432	-438817605	15408	17219	14649	12079
-228983748	64348046	-438818061	19031	16191	13878	11565
-228983439	64348234	-438817301	17055	16448	14135	12336
-228981883	64349716	-438818068	13225	16962	15163	12593
-228980376	64350635	-438817382	19807	17219	14906	13621
-228979872	64351397	-438817928	17197	15934	14392	11308
-228979328	64352145	-438818479	22696	16448	14135	12336

3.4 Street-Level Roadway Imagery Collection

3.4.1 Image Acquisition Process

Street-level imagery was obtained from UDOT's PathWeb platform, which archives roadway condition images captured during regular LiDAR surveys. Images are collected using vehicle-mounted panoramic cameras that record forward-facing and side-facing views at regular intervals along the highway network.

For this research, images were downloaded at intervals of 0.01 miles, focusing on right-edge lines corresponding to the retroreflectivity data. The image download process was automated using Selenium to systematically access PathWeb's web-based interface and retrieve base64-encoded image data for each desired chain mileage point. Table 3.3 summarizes the number of images for each highway.

Table 3.3 Number of Images per Highway

Highway	No. of Images
I-15	7,820
I-84	5,136
I-80	1,525
I-215	948
Total	15,429

3.4.2 Data Structure and Usage

The collected images are stored in organized folder structures based on route, direction, and retroreflectivity classification labels. Each image filename corresponds to its chain milepost for direct spatial correlation with LiDAR and field retroreflectivity data.

Images are processed to extract sub-regions centered on lane markings, which serve as input to the deep learning models described in the methodology. A sample of the image dataset is provided in Figure 3.1.

Figure 3.1 Image Data Sample

3.5 Summary

This chapter details the three primary datasets utilized in this research: field retroreflectivity measurements, mobile LiDAR point cloud data, and street-level roadway imagery. These datasets were collected from UDOT's vendors' existing survey programs, supplemented with automated extraction processes where necessary. Each dataset was structured to enable spatial alignment and joint analysis across data sources.

4.0 DATA EVALUATION

4.1 Overview

This chapter presents the evaluation and analytical methods used to assess the condition of pavement markings using image data, LiDAR point clouds, and their fusion. The objective of the data evaluation process was to explore whether retroreflectivity, a critical indicator of pavement marking visibility, can be reliably predicted using data sources already collected or easily obtainable by UDOT.

The chapter begins by detailing the image data pre-processing steps, which involve isolating and standardizing key visual components of pavement markings such as edge lines and dashed lane separators. These processed image segments were then used to train both regression and classification models (Golazad et al., 2024). Regression analysis aimed to estimate continuous retroreflectivity values, while classification models determined compliance with federal retroreflectivity thresholds.

Advanced statistical techniques and machine learning algorithms, including Random Forest regressors and deep convolutional neural networks (CNNs), were applied and validated using cross-validation strategies. Performance metrics such as RMSE, R^2, accuracy, precision, recall, and F1 score were used to evaluate model effectiveness.

The chapter further explores a data fusion approach, integrating both LiDAR-derived features and image-based deep features to improve predictive accuracy. Segment-level and spatial cluster-level analyses were conducted to align with operational decision-making needs. Visualization tools, including scatter plots and bar charts, were employed to present the results in an interpretable format.

Collectively, this evaluation framework provides a robust foundation for leveraging multimodal data to support efficient, data-driven maintenance of pavement markings across Utah's road network.

4.2 Image Data Processing

To standardize image inputs for classification, each raw roadway image was processed using a multi-step procedure designed to isolate the pavement markings and extract the most relevant visual content.

4.2.1 Processing Continuous Lane Markings (Edge Lines)

A dedicated image processing workflow was implemented to extract standardized visual segments around continuous pavement markings, primarily edge lines. Key steps included:

- Grayscale Conversion: Simplifies the image and focuses on light intensity, which corresponds to retroreflectivity.
- Road Area Masking: A polygon mask was dynamically applied to isolate the lower right portion of the image where edge lines typically appear.
- Percentile-Based Thresholding: A high-intensity threshold (99.85th percentile) was used to isolate bright lane markings.
- Binary Image Generation: A binary image was created to highlight high-reflectivity regions.
- Region Cropping: The centroid of the bright pixels was calculated, and a 15×15 square was extracted from the original image, centered on this point.
- Optional Contour-Based Cropping: For additional robustness, the largest contour was located, and a square region was extracted around it and upscaled to standardize resolution.

These processed image regions were used as inputs for a deep learning classifier. Visual samples were saved during each step to validate proper detection and alignment of the lane markings. Throughout the processing pipeline, intermediate outputs such as grayscale images, binary masks, and extracted squares were visually inspected to confirm that the correct features were being captured.

4.2.2 Processing Dashed Lane Markings

Dashed lines, such as lane separators, required a separate pre-processing strategy due to their distinct spatial position and intermittent nature. A modified version of the previous workflow was implemented with the following changes:

- Masking Shift: A polygon mask was applied to the lower left-center region of the image, where dashed markings typically occur, rather than the far right.
- Thresholding and Binarization: The same percentile-based method was used to isolate bright pixels.
- Centroid-Based Cropping: A small square (15×15 pixels) was extracted around the brightest cluster in the masked area.
- Contour-Based Extraction: The largest contour was found and used to define a square bounding box centered on the dashed marking.
- Upscaling: The extracted region was resized to double its original size using bilinear interpolation for improved model readability.

This process was applied to all images in the dashed-line dataset. Processed images were saved into categorized folders for use in downstream modeling tasks and further inspection. Figure 4.1 illustrates the step-by-step image processing workflow used to extract pavement marking features from roadway images.

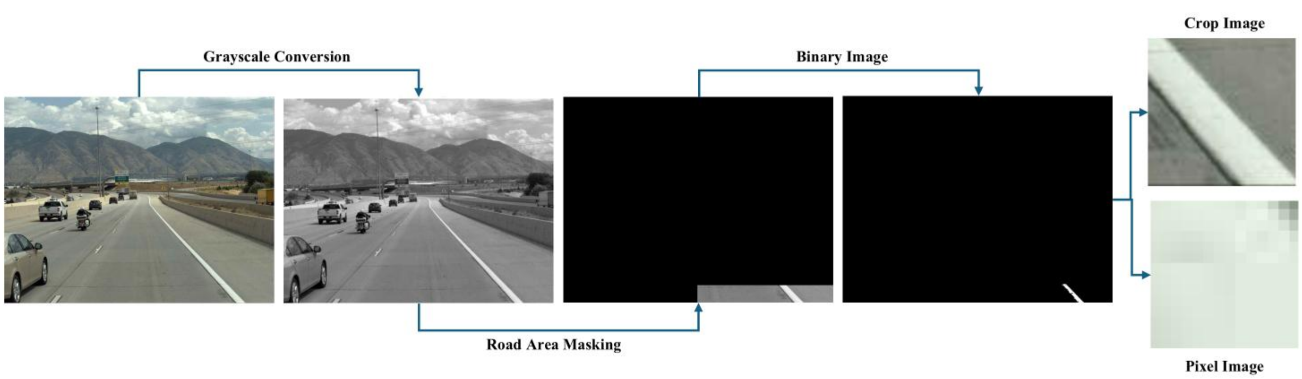


Figure 4.1 Image Processing Workflow

4.3 Predicting Retroreflectivity Value Using Images

This section describes a regression-based modeling approach developed to estimate continuous retroreflectivity values directly from image data. Unlike the binary classification tasks used earlier to identify compliant vs. non-compliant markings, this analysis focuses on predicting the actual average retroreflectivity (Ave) value for each pavement segment, measured in millicandelas per square meter per lux (mcd/m²/lux). The goal is to enable more granular estimation that can inform nuanced maintenance decisions and support retroreflectivity degradation modeling over time.

4.3.1 Feature Extraction Using ResNet-18

To convert raw image data into a format suitable for regression, a pre-trained ResNet-18 model was used as a feature extractor. Specifically:

- A custom function was implemented to extract deep features from cropped pavement marking images using the convolutional layers of ResNet-18 (excluding the final classification layers).
- Images corresponding to each 0.1-mile segment were identified using their filenames, which encode the mileage location.
- For each segment, all matching images were processed individually. The resulting 512-dimensional feature vectors were averaged to form a single representative embedding for that segment.
- A total of 512 image-based features were extracted per segment, along with a count of valid images, and appended to the master dataset that already contained retroreflectivity ground truth values.

This approach enables the representation of complex visual characteristics in a structured numerical format that can be used in regression modeling.

4.3.2 Model Training and Evaluation

A Random Forest Regressor was trained to predict the average retroreflectivity value (Ave) for each segment using only the extracted image features. The modeling workflow included the following steps:

- Data Filtering: To ensure reliability, the dataset was filtered to exclude segments with:
  - High proportions of invalid measurements (Valid > Inv)
  - High coefficient of variation (Stdev/Ave > 2× mean ratio)
  - Segments with no associated image data
- Train/Test Split: The filtered dataset was randomly split into training (80%) and testing (20%) sets.
- Model Configuration: A Random Forest Regressor with 100 estimators and default hyperparameters was used. This model was chosen for its robustness to non-linear relationships and insensitivity to feature scaling.
- Performance Metrics: Root Mean Squared Error (RMSE) was used to evaluate the model.

Table 4.1 summarizes the performance metrics of the regression models.

Table 4.1 Regression Results Summary

Road Marking	RMSE	R^2	Road Marking	RMSE	R^2
I-15 REL Inc	85	0.58	I-15 REL Dec	92	0.73
I-84 REL Inc	24	0.82	I-84 REL Dec	26	0.85
I-84 LL Inc	46	0.55	I-84 LL Dec	47	0.59
I-80 REL Inc	80	0.45	I-80 REL Dec	83	0.44

Figure 4.2 shows the predicted vs. true retroreflectivity values for the I-84 REL.

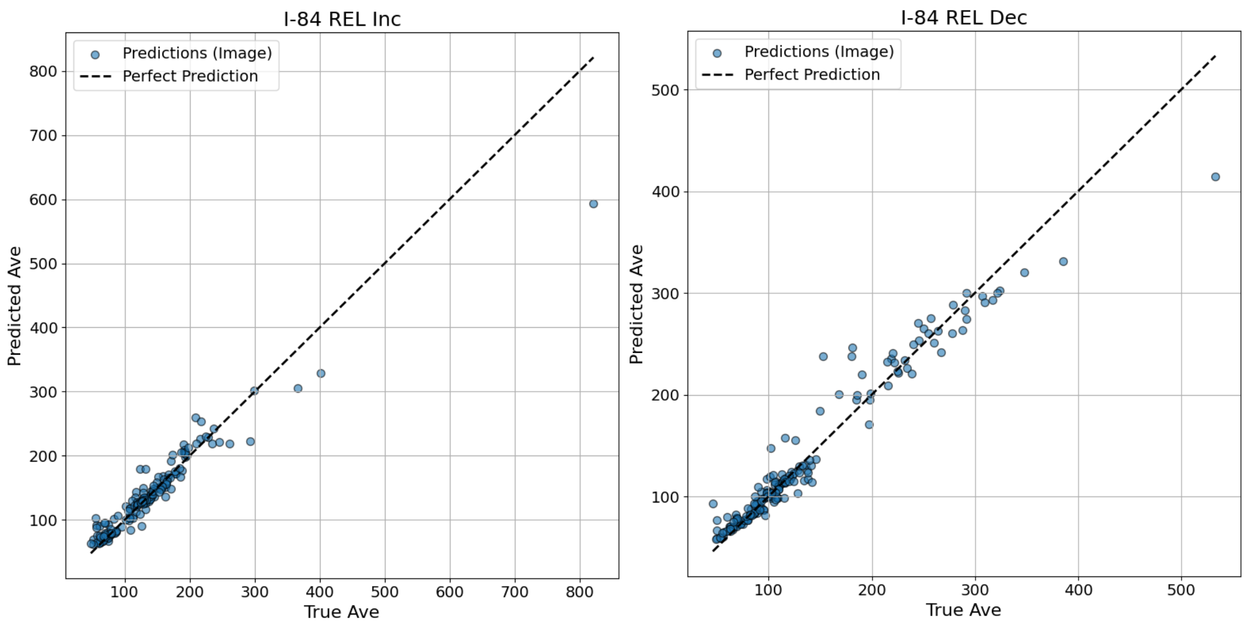


Figure 4.2 Predicted vs. True Retroreflectivity Values for I-84 REL

4.4 Image Classification

This chapter presents the analysis performed using deep learning models to classify pavement marking retroreflectivity based on image data. The objective of this analysis was to predict whether the retroreflectivity of a given lane marking exceeds the federally mandated threshold of 100 mcd/m²/lux. A convolutional neural network (CNN) was developed and trained to perform binary classification using cropped pavement marking images. The performance of the trained model was evaluated both at the image level and in aggregated spatial clusters to assess prediction accuracy across road segments (Mohammadi et al.; Mohammadi et al.).

4.4.1 Dataset Preparation and Class Balancing

Images used for training were pre-processed to extract square regions centered on pavement markings, as described in earlier sections. These image patches were labeled as either "above100" or "below100" based on matched field retroreflectivity measurements. To address class imbalance, data augmentation techniques such as rotation, translation, zoom, brightness adjustment, and horizontal flipping were applied using Keras' ImageDataGenerator. The augmentation increased the minority class sample count to match the majority class, ensuring a balanced dataset.

After augmentation, the dataset was shuffled and split into training (64%), validation (16%), and test (20%) sets using stratified sampling to maintain class balance across all subsets.

4.4.2 Model Architecture

The deep learning model was implemented using TensorFlow/Keras and is based on a convolutional architecture enhanced with residual blocks for better feature propagation. Key components include:

- Convolutional Layers with ReLU activation and batch normalization.
- Residual Blocks to allow gradient flow and mitigate overfitting.
- Global Average Pooling for dimensionality reduction.
- Fully Connected Dense Layers with dropout for regularization.
- Sigmoid Output Layer for binary classification.

The model was compiled using the Adam optimizer and binary cross-entropy loss function. Early stopping and learning rate reduction callbacks were employed to improve generalization and training efficiency.

4.4.3 Validation and Test Performance

After training the model on the combined dataset comprising images from all highways, performance was evaluated in two ways:

1. Aggregate Performance: The model was assessed on standard validation and test splits derived from the full dataset, which included a stratified mix of all highways. This reflects the model's general performance across the entire dataset.

2. Per-Highway Evaluation: The same trained model was then applied to subsets of images from individual highways to analyze performance variations across different road environments. This helps identify any bias, overfitting, or underperformance tied to specific highway conditions or markings.

Table 4.2 summarizes the aggregate validation and test results. Table 4.3 presents the performance metrics when the trained model is applied separately to each highway subset.

Table 4.2 Classification Performance of the Trained Model on the Overall Validation and Test Sets, Combining Data from All Highways.

Accuracy		Precision		Recall		F1 Score	
Validation	Test	Validation	Test	Validation	Test	Validation	Test
0.85	0.84	0.85	0.85	0.85	0.84	0.85	0.84

Table 4.3 Classification Performance of the Trained Model When Evaluated Separately on Each Highway's Dataset: Highlights Performance Variability Across Different Road Environments

	Accuracy		Precision		Recall		F1 Score	
Road Marking	Validation	Test	Validation	Test	Validation	Test	Validation	Test
I-15 REL Inc	0.88	0.86	0.88	0.85	0.88	0.86	0.88	0.86
I-15 REL Dec	0.73	0.73	0.74	0.73	0.73	0.73	0.69	0.70
I-15 LL Inc	0.72	0.73	0.71	0.72	0.72	0.73	0.71	0.73
I-15 LL Dec	0.73	0.70	0.73	0.70	0.73	0.70	0.73	0.70
I-84 REL Inc	0.90	0.89	0.90	0.89	0.90	0.89	0.90	0.89
I-84 REL Dec	0.84	0.85	0.84	0.85	0.84	0.85	0.84	0.85
I-84 LL Inc	0.79	0.77	0.79	0.77	0.79	0.77	0.78	0.76
I-84 LL Dec	0.69	0.76	0.67	0.75	0.69	0.76	0.67	0.75
I-80 REL Inc	0.91	0.88	0.89	0.79	0.91	0.88	0.89	0.84
I-80 REL Dec	0.85	0.85	0.74	0.73	0.85	0.85	0.79	0.78
I-80 LL Inc	0.95	0.97	0.98	0.98	0.95	0.97	0.96	0.97
I-80 LL Dec	0.96	0.98	0.98	0.98	0.96	0.98	0.97	0.98

4.4.4 Cluster-Level Validation and Spatial Aggregation

To evaluate model predictions in a way that aligns with roadway maintenance operations, image-level predictions were spatially aggregated into uniform cluster segments (e.g., 1-mile bins). For each cluster, the percentage of images predicted as "Above100" was computed, and clusters were color-coded based on compliance categories:

- Green: >60% of predictions indicate compliance
- Yellow: 40%–60% compliance (borderline)
- Red: <40% compliance

These predictions were compared against ground truth retroreflectivity data, which was similarly clustered and color-coded. Moreover, cluster-level agreement between predicted and true compliance categories was used as a performance indicator. For the example dataset:

Cluster Accuracy = proportion of clusters where predicted color matched true color

This method offers an intuitive and scalable way to assess how well image-based models support network-level maintenance decision-making. The cluster-level accuracies are summarized in Table 4.4.

Table 4.4 Cluster-Level Classification Summary

	Accuracy		
Road Marking	1-Mile	2-Mile	3-Mile
I-15 REL Inc	0.887	0.930	1.00
I-15 REL Dec	0.707	0.75	0.710
I-15 LL Inc	0.814	0.844	0.810
I-15 LL Dec	0.784	0.821	0.750
I-84 REL Inc	0.925	0.952	0.933
I-84 REL Dec	0.886	0.913	0.933
I-84 LL Inc	0.905	0.909	0.875
I-84 LL Dec	0.867	0.778	0.667

I-80 REL Inc	0.957	0.917	0.875
I-80 REL Dec	0.857	0.917	0.875
I-80 LL Inc	1.000	1.000	1.000
I-80 LL Dec	1.000	1.000	1.000

4.5 Data Fusion

To enhance the accuracy and reliability of retroreflectivity classification, LiDAR-derived features and image-based features were fused and evaluated jointly (Gadzicki et al., 2020; Yeong et al., 2021). The fusion methodology involved structured feature extraction from both data modalities, followed by supervised learning and spatial aggregation to assess prediction performance on roadway segments.

4.5.1 Dataset Construction

A combined dataset was created by merging LiDAR statistics with image-based deep feature embeddings. For each pavement marking segment (defined by milepost "Chain"), the following features were compiled:

- LiDAR Features: Median, mean, standard deviation, 10th percentile, and 90th percentile for each of four channels: intensity, red, green, and blue.
- Image Features: Using a pre-trained ResNet-18 backbone, each cropped image patch associated with a segment was converted into a 512-dimensional feature vector. These vectors were averaged across all images within the segment.
- Image Count: The number of valid images contributing to the average embedding was also recorded.

The result was a comprehensive feature set for each Chain-based segment, encompassing statistical summaries and semantic image features. Data cleaning steps were applied to remove outliers, invalid retroreflectivity records, and segments with no image data or low reliability (e.g., high coefficient of variation).

4.5.2 Classification and Cross-Validation

Three sets of models were trained using the following feature combinations:

- LiDAR-Only Model: Utilized the statistical features from LiDAR channels.
- Image-Only Model: Utilized the ResNet-extracted deep features from images.
- Fusion Model: Combined both LiDAR and image features into a unified feature space.

All models used a Random Forest Classifier with balanced class weights to handle label imbalance. A 5-fold stratified cross-validation approach was employed to compute key performance metrics, including accuracy, precision, recall, and F1-score. The true labels were derived from measured retroreflectivity values, with a threshold of 100 mcd/m²/lux separating "high" and "low" categories. Table 4.5 presents a summary of cross-validation metrics across the three models.

Table 4.5 Cross-Validation Classification Metrics

	LiDAR-Only				Image-Only				Fusion			
Road Marking	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score	Accuracy	Precision	Recall	F1 Score
I-15 REL Inc	0.944 ± 0.035	0.919 ± 0.086	0.838 ± 0.091	0.868 ± 0.083	0.918 ± 0.035	0.865 ± 0.100	0.787 ± 0.066	0.816 ± 0.077	0.922 ± 0.033	0.890 ± 0.104	0.790 ± 0.063	0.822 ± 0.071
I-15 REL Dec	0.819 ± 0.047	0.774 ± 0.084	0.725 ± 0.075	0.736 ± 0.079	0.870 ± 0.041	0.927 ± 0.020	0.739 ± 0.085	0.775 ± 0.090	0.889 ± 0.040	0.936 ± 0.021	0.775 ± 0.087	0.813 ± 0.085
I-84 REL Inc	0.915 ± 0.052	0.900 ± 0.060	0.914 ± 0.060	0.904 ± 0.060	0.927 ± 0.031	0.916 ± 0.032	0.915 ± 0.041	0.915 ± 0.036	0.958 ± 0.031	0.949 ± 0.032	0.953 ± 0.044	0.951 ± 0.038

I-84 REL Dec	0.890 ± 0.029	0.890 ± 0.030	0.888 ± 0.030	0.887 ± 0.030	0.873 ± 0.034	0.882 ± 0.042	0.863 ± 0.029	0.868 ± 0.033	0.879 ± 0.023	0.880 ± 0.025	0.871 ± 0.024	0.874 ± 0.025
I-84 LL Inc	0.845 ± 0.072	0.825 ± 0.090	0.803 ± 0.104	0.807 ± 0.099	0.850 ± 0.052	0.862 ± 0.045	0.787 ± 0.092	0.799 ± 0.092	0.867 ± 0.041	0.867 ± 0.041	0.809 ± 0.074	0.825 ± 0.070
I-84 LL Dec	0.859 ± 0.045	0.865 ± 0.046	0.817 ± 0.070	0.827 ± 0.062	0.871 ± 0.050	0.863 ± 0.063	0.839 ± 0.062	0.847 ± 0.060	0.865 ± 0.026	0.869 ± 0.039	0.820 ± 0.036	0.836 ± 0.034
I-80 REL Inc	0.822 ± 0.057	0.547 ± 0.175	0.606 ± 0.183	0.569 ± 0.171	0.807 ± 0.033	0.410 ± 0.014	0.491 ± 0.018	0.446 ± 0.010	0.791 ± 0.028	0.408 ± 0.013	0.482 ± 0.022	0.442 ± 0.009
I-80 REL Dec	0.804 ± 0.123	0.773 ± 0.162	0.755 ± 0.158	0.755 ± 0.166	0.793 ± 0.083	0.728 ± 0.212	0.727 ± 0.125	0.713 ± 0.168	0.793 ± 0.082	0.810 ± 0.102	0.723 ± 0.115	0.726 ± 0.128

4.5.3 Segment-Level Cluster Evaluation

To align with maintenance planning and practical application, predictions were aggregated over uniform 1-mile clusters. For each cluster, the proportion of segments classified as "high" was computed for all three models (LiDAR-only, image-only, and fusion). A color-coding scheme was applied to each cluster for visualization and accuracy scoring:

- Green: >60% segments predicted as "high"
- Yellow: 40–60% segments predicted as "high"
- Red: <40% segments predicted as "high"

The same logic was used to assign true color labels based on measured retroreflectivity values. Cluster-level accuracy was defined as the percentage of clusters where predicted colors matched the true colors. Table 4.6 presents a summary of cluster-level accuracies.

Table 4.6 Cluster-Level Accuracy

Road Marking	LiDAR-Only	Image-Only	Fusion
I-15 REL Inc	0.917	0.917	0.933
I-15 REL Dec	0.794	0.794	0.824
I-84 REL Inc	0.931	0.931	0.966
I-84 REL Dec	0.964	0.893	0.893
I-84 LL Inc	0.935	0.806	0.839
I-84 LL Dec	0.909	0.879	0.879
I-80 REL Inc	0.889	0.833	0.833
I-80 REL Dec	0.895	0.842	0.842

4.5.4 Visualization

The results were visualized using horizontal bar plots, showing chain mileage on the x-axis and predicted compliance categories on the y-axis (true, LiDAR, image, fusion). The visualization highlights spatial alignment (or disagreement) between models and ground truth data.

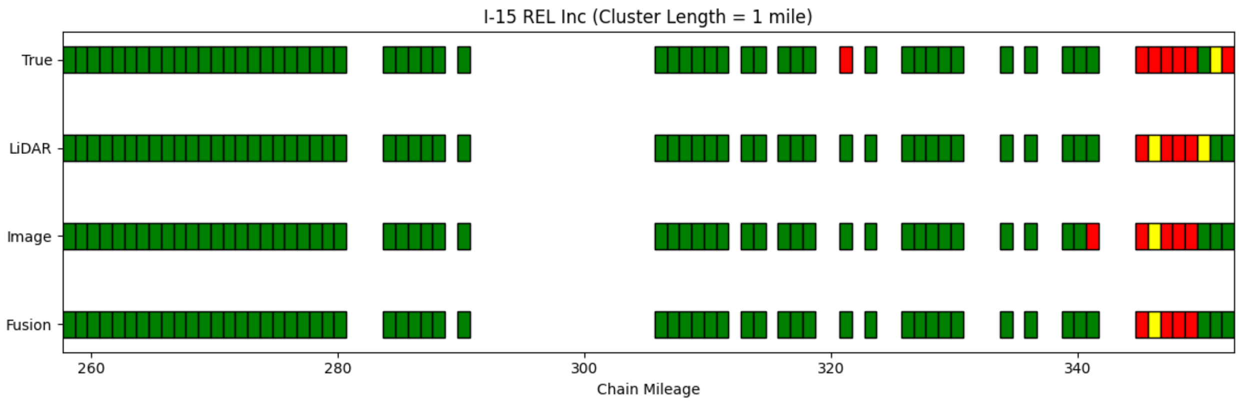


Figure 4.3 I-15 REL Inc Cluster-Level Classification

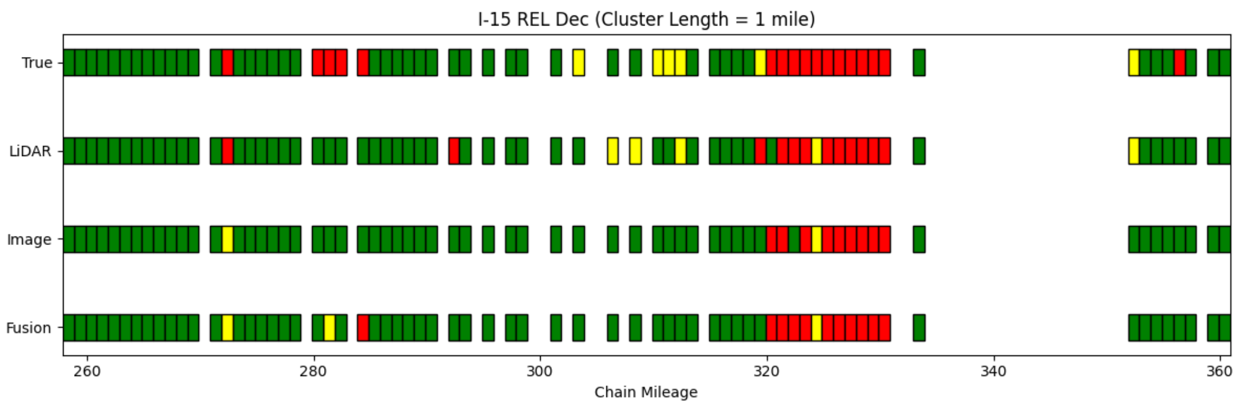


Figure 4.4 I-15 REL Dec Cluster-Level Classification

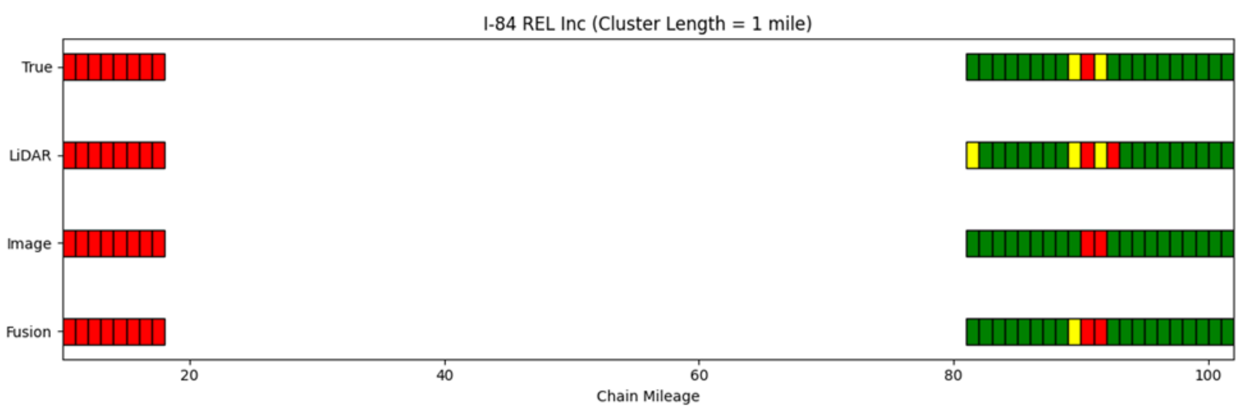


Figure 4.5 I-84 REL Inc Cluster-Level Classification

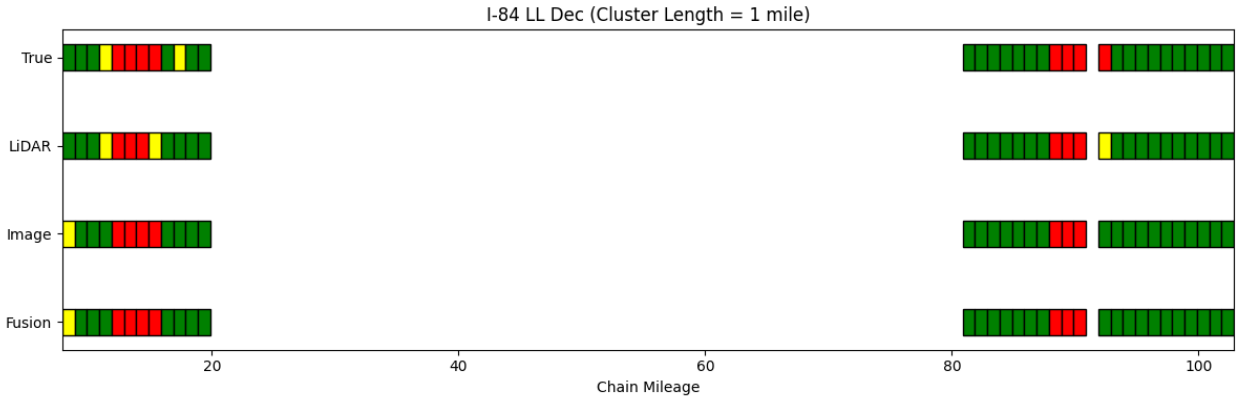


Figure 4.6 I-84 REL Dec Cluster-Level Classification

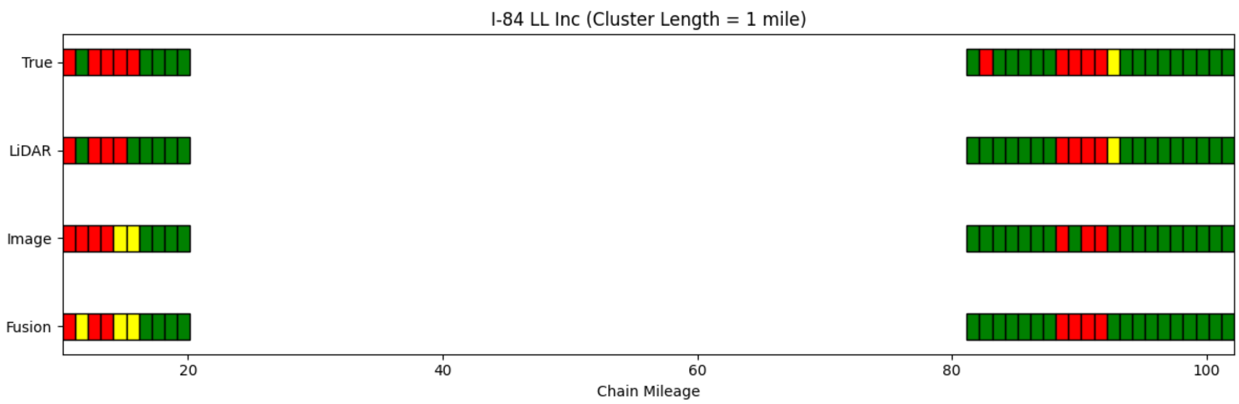


Figure 4.7 I-84 LL Inc Cluster-Level Classification

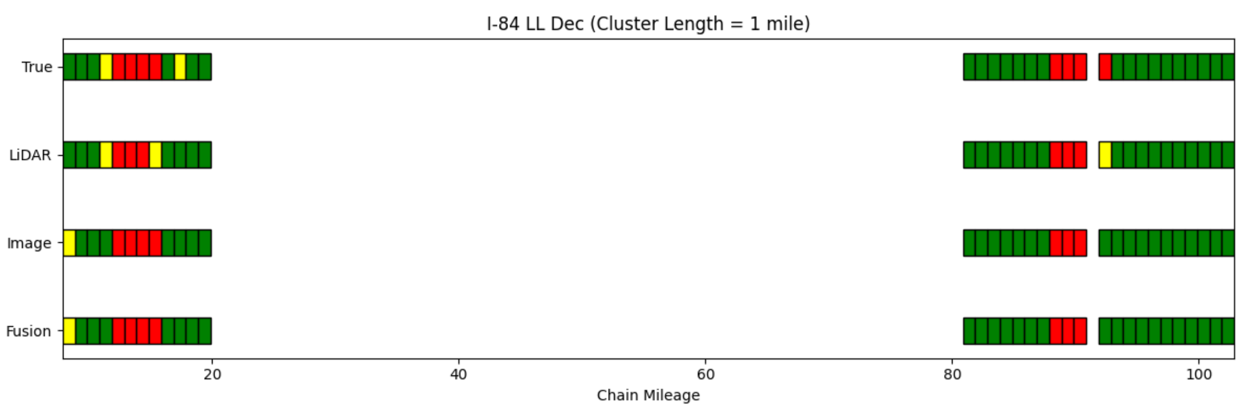


Figure 4.8 I-84 LL Dec Cluster-Level Classification

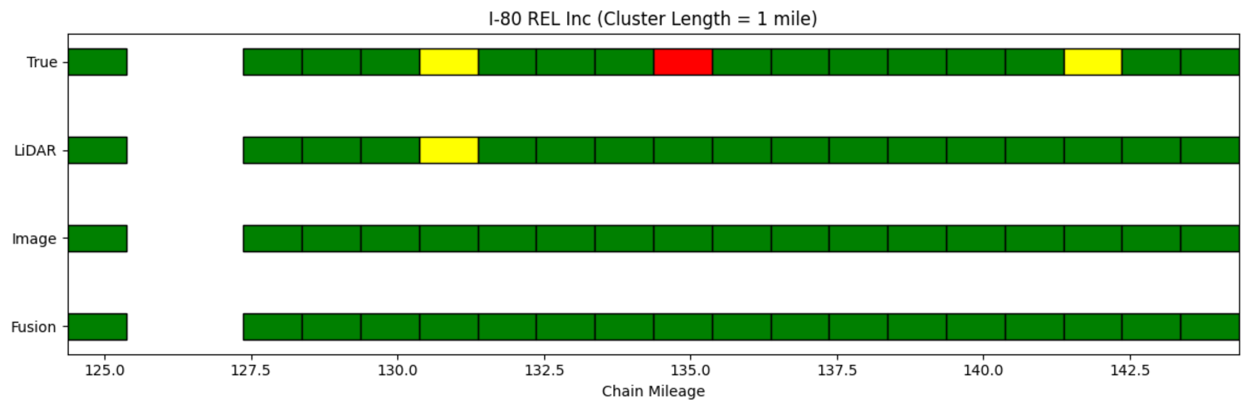


Figure 4.9 I-80 REL Inc Cluster-Level Classification

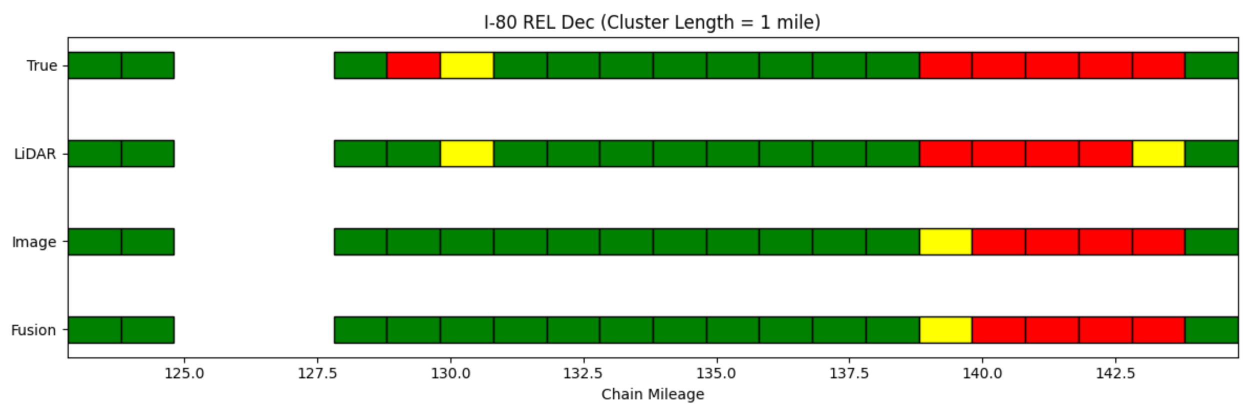


Figure 4.10 I-80 REL Dec Cluster-Level Classification

4.5.5 Summary

The proposed fusion approach demonstrated a significant improvement in predictive performance by combining complementary features from LiDAR and image data. The results showed that:

- The fusion model achieved the highest overall accuracy in both segment-level and cluster-level evaluations.
- Cross-validation provided robust metrics for model generalization.
- Spatial aggregation techniques yielded actionable insights for corridor-level retroreflectivity assessment.

This multimodal fusion framework offers a scalable and interpretable approach for statewide pavement marking monitoring and decision-making.

5.0 CONCLUSIONS

5.1 Summary

This research addressed the growing need for efficient, scalable, and cost-effective methods to monitor pavement marking retroreflectivity in support of roadway safety and federal compliance. This need is also driven by operational challenges faced by UDOT in conducting widespread retroreflectivity measurements using traditional field equipment, which is labor-intensive and resource-constrained. The work presented in this report builds upon a prior UDOT phase that demonstrated the feasibility of LiDAR-only retroreflectivity estimation, extending that foundation with imagery-based models and multi-source fusion. The analyses in this report are based on field data collected predominantly from durable longitudinal pavement markings, and the conclusions are therefore most directly applicable to durable marking systems.

The primary objective was to develop a multi-source, machine learning–based tool that can estimate pavement marking retroreflectivity using existing datasets, specifically mobile LiDAR point clouds and street-level roadway imagery. The tool is designed to assist UDOT in identifying non-compliant markings and in making informed maintenance planning decisions, without relying on full-scale field campaigns.

Three main data sources were used in the study:

- Field-collected retroreflectivity measurements serve as ground truth for model training and validation.
- Mobile LiDAR point cloud data, providing intensity and spectral (RGB) values of pavement surfaces.
- Street-level imagery, used to extract visual cues from pavement markings through deep learning.

Data were systematically processed and integrated. LiDAR data were cleaned and summarized into statistical descriptors. Imagery was cropped and pre-processed to isolate markings and extract deep feature embeddings using ResNet. These features were then used in both independent and combined (fusion) classification models to predict whether retroreflectivity values met or fell below the federally mandated threshold of 100 mcd/m²/lux.

Evaluation of the models included cross-validation for segment-level performance and spatial aggregation for cluster-level decision support. Key performance metrics (accuracy, precision, recall, F1-score) and visual analytics were used to assess how well the models aligned with the ground truth.

5.2 Findings

The study demonstrated that both LiDAR-based and image-based models can effectively predict retroreflectivity compliance, with the fusion of the two modalities yielding the most accurate results. Key findings include:

- Fusion Outperforms Individual Models: Across nearly all test cases, the fusion model, combining LiDAR statistics and image-derived features, achieved higher accuracy and F1-scores than either modality alone. For instance, the fusion model reached cross-validation accuracies as high as 95.8% on some datasets (e.g., I-84 REL Inc).

- Consistent Cluster-Level Accuracy: When predictions were aggregated into 1-mile clusters, the fusion model continued to outperform others. It achieved cluster-level accuracy improvements of up to 3–5% compared to LiDAR-only and image-only models. This is especially relevant for maintenance planning, where segment-level precision supports network-wide decision-making.

- Deep Learning Enables Robust Image Classification: The CNN model trained on cropped lane markings generalized well across different highways, achieving test accuracies above 84% in most cases. The model also maintained high performance across various surface types and pavement marking configurations.

- LiDAR Remains a Strong Independent Predictor: The LiDAR-based model alone performed well, particularly where pavement markings were geometrically prominent and cleanly captured by the point cloud. In some cases, it outperformed the image-only model, highlighting its value in low-contrast or low-resolution image contexts.

- Model Agreement with Ground Truth is High: When compared to ground truth compliance categories, the fusion model achieved high agreement rates at both the segment and cluster levels, demonstrating its practical utility for real-world applications.

- Data Processing Pipelines Are Scalable and Reproducible: Automated workflows for image downloading, LiDAR cleaning, feature extraction, and spatial alignment proved effective and scalable, indicating that the tool can be expanded statewide.

5.3 Challenges and Limitations

- Data Quality Variability: Differences in image resolution, lighting conditions, and LiDAR point density introduced variability in model performance across different corridors.

- Imbalanced Datasets: In some routes, a significant imbalance existed between compliant and non-compliant markings. This was addressed through data augmentation and oversampling, but it remains a consideration for future calibration.

- Underlying Mechanism for Image-Based Models: Even though image-based models provided strong performance in predicting retroreflectivity values and classification, it is important to recognize that the process applied to the images does not necessarily produce a direct measure of light reflected back to a source (unlike the LiDAR-based approach, using light intensity received back to the sensor unit). Rather, image-based models seem to identify patterns that correlate with retroreflectivity but are based on the treatment condition or the line integrity. Poor or degraded material, as well as material variations are the more likely indicators of poor retroreflectivity performance, and thus image-only models need to be taken with more reservations. Nonetheless, the analysis suggests that strong nuances in the imagery may complement the information from LiDAR-only models, resulting in overall best performances from fusion models.

- Edge Case Misclassifications: Models occasionally struggled with borderline retroreflectivity values or segments with inconsistent field measurements. Further refinement, including probabilistic classification or regression models, may help in future work.

- Focus on Durable Markings: The field datasets used for model development are dominated by durable longitudinal pavement markings on UDOT's interstate corridors. As a result, model performance metrics reported in this study are most representative of durable markings. Application to other marking materials would benefit from additional data collection and model calibration.

6.0 RECOMMENDATIONS AND IMPLEMENTATION

6.1 Recommended Use of the Computer Tool and Companion Guideline

This project's methods have been implemented in a portable, user-friendly computer tool that enables UDOT to estimate and monitor pavement marking retroreflectivity using data fusion (LiDAR + imagery), LiDAR-only, or image-only workflows.

The recommended operational use of the tool, including required inputs, output interpretation, and quality assurance steps, is documented in the companion guideline, "UDOT's Computer Tool to Assess and Report Pavement Marking Retroreflectivity."

Consistent with the scope of model development and validation in this report, the recommended initial deployment focus is on durable longitudinal markings (e.g., edge lines and lane lines) within the baseline model coverage documented in the guideline.

Recommended evaluation method selection (by data availability and decision criticality) is as follows:

- Fusion (LiDAR + imagery): Preferred when both data sources are available and properly aligned, as it leverages complementary information from intensity-based and visual cues to improve reliability for compliance screening and maintenance prioritization.
- LiDAR-only: Recommended as the primary fallback when imagery is incomplete or unavailable, and for corridors where LiDAR-derived features are known to be stable and consistent.
- Image-only: Recommended primarily as a rapid screening method or where LiDAR processing is infeasible, recognizing that fusion or LiDAR is generally preferred when available.

6.2 Recommended Outputs for Reporting and Maintenance Planning

The tool is designed to produce decision-ready outputs suitable for both technical reporting and operational planning. As described in the companion guideline, the outputs include:

- CSV tables with point- and segment-level location, ground truth (when available), LiDAR/image features, predictions, and classification outcomes;
- Plots showing segment-level condition and comparisons across evaluation methods; and
- GIS GeoPackage layers for corridor-wide visualization and integration into mapping workflows.

For implementation, segment-level assessments are strongly preferred over individual point outputs, because segments exploit the spatial correlation of adjacent observations and provide more stable and interpretable results for decision-making.

The guideline further provides a color-coded interpretation framework (Green/Yellow/Orange/Red) linked to recommended actions and prioritization for maintenance planning.

6.3 Suggested Annual Assessment Cycle

To support ongoing monitoring of durable pavement markings and to align outputs with operational maintenance cycles, the guideline recommends recurring assessments at least once per year, with a workflow that includes (i) gathering and organizing the most recent LiDAR and/or imagery inputs, (ii) running the tool by defined sections, and (iii) reviewing outputs visually and within GIS to identify priority segments and corridor-level patterns.

This annual cycle is intended to reduce reliance on full-network retroreflectometer surveys while still providing defensible, data-driven identification of non-compliant or at-risk segments requiring maintenance attention.

6.4 Model Maintenance, Calibration, and Future Retraining

While the results in this report demonstrate strong performance for the studied corridors, long-term deployment requires continued verification because LiDAR intensity and imagery can be influenced by collection settings and ambient conditions.

Accordingly, the recommended approach is to use limited, targeted field measurements to confirm performance over time and to trigger recalibration only when needed.

The companion guideline defines two levels of model maintenance:

a. Annual "light-touch" calibration (verification): A small sample of segments (e.g., ≤10%) is measured with a retroreflectometer to confirm that accuracy and error metrics remain consistent year-over-year. This sample is sufficient to detect systematic drift at the corridor level.

b. Recalibration / retraining (refresh): If drift is detected or error metrics degrade materially, models should be refreshed by adding new observations and retraining model coefficients/classifiers. Models can be refreshed annually when new LiDAR/imagery surveys are completed, and recalibration becomes necessary when performance degrades or data distributions shift.

In addition, the guideline recommends collecting additional ground truth when model outcomes systematically disagree over multiple consecutive segments (i.e., targeted spot checks), and maintaining a model registry with versioning, training data vintages, and performance metrics to support governance and auditability.

Together, these recommendations provide an implementation path that supports repeatable annual monitoring, defensible compliance screening, and a controlled strategy for retraining models as new data, sensors, or operational conditions evolve.

REFERENCES

Ai, C., & Hennessy, E. (2022). A pavement marking inventory and retroreflectivity condition assessment method using mobile LiDAR. *Massachusetts Department of Transportation,*.

Benz, R. J., Pike, A. M., Kuchangi, S. P., & Brackett, Q. (2009). Mobile retroreflectivity best practices handbook [Tech Report]. https://rosap.ntl.bts.gov/view/dot/16768

Brinster, G. L., Hodaei, M., Eissa, A. M., DeLoach, Z., Bruno, J. E., Habib, A., & Bullock, D. M. (2024). Evaluating Mobile LiDAR Intensity Data for Inventorying Durable Tape Pavement Markings. *Sensors*, *24*(20).

Brinster, G. L., Hodaei, M., Eissa, A. M., DeLoach, Z., Bruno, J. E., Habib, A., & Bullock, D. M. (2024). *Leveraging LiDAR Intensity to Evaluate Roadway Pavement Marking Retroreflectivity*. Indiana Department of Transportation.

Che, E., Olsen, M., Parrish, C., & Jung, J. (2019). Pavement Marking Retroreflectivity Estimation and Evaluation using Mobile Lidar Data. *Photogrammetric Engineering & Remote Sensing*, *85*, 573-583. https://doi.org/10.14358/PERS.85.8.573

Cheng, Y.-T., Lin, Y.-C., & Habib, A. (2022). Generalized LiDAR intensity normalization and its positive impact on geometric and learning-based lane marking detection. *Remote Sensing*, *14*(17), 4393.

Cheng, Y.-T., Patel, A., Wen, C., Bullock, D., & Habib, A. (2020). Intensity Thresholding and Deep Learning Based Lane Marking Extraction and Lane Width Estimation from Mobile Light Detection and Ranging (LiDAR) Point Clouds. *Remote Sensing*, *12*(9).

Cheng, Y.-T., Shin, Y.-H., Shin, S.-Y., Koshan, Y., Hodaei, M., Bullock, D., & Habib, A. (2024). Image-aided LiDAR extraction, classification, and characterization of lane markings from mobile mapping data. *Remote Sensing*, *16*(10), 1668.

Chun, C., Lee, T., Kwon, S., & Ryu, S. K. (2020). Classification and Segmentation of Longitudinal Road Marking using Convolutional Neural Networks for Dynamic Retroreflection Estimation. *Sensors (Basel)*, *20*(19). https://doi.org/10.3390/s20195560

Federal Highway Administration. (2022). *National standards for traffic control devices; The Manual on Uniform Traffic Control Devices for streets and highways; Maintaining pavement marking retroreflectivity*. https://www.federalregister.gov/documents/2022/08/05/2022-16781/national-standards-for-traffic-control-devices-the-manual-on-uniform-traffic-control-devices-for

Gadzicki, K., Khamsehashari, R., & Zetzsche, C. (2020). *Early vs Late Fusion in Multimodal Convolutional Neural Networks*. https://doi.org/10.23919/FUSION45008.2020.9190246

Golazad, S., Mohammadi, A., Rashidi, A., & Ilbeigi, M. (2024). From raw to refined: Data preprocessing for construction machine learning (ML), deep learning (DL), and reinforcement learning (RL) models. *Automation in Construction*, *168*, 105844. https://doi.org/https://doi.org/10.1016/j.autcon.2024.105844

Guan, H., Li, J., Cao, S., & Yu, Y. (2016). Use of mobile LiDAR in road information inventory: A review. *International Journal of Image and Data Fusion*, *7*(3), 219-242.

Hodaei, M., Eissa Aser, M., Hany, Y., Shin, S.-Y., Brinster Gregory, L., DeLoach, Z., Bruno Joseph, E., Li, S., Bullock Darcy, M., & Habib, A. (2025). Potential of Using Intensity Data from Mobile Multibeam Spinning LiDAR for Predicting Standard Retroreflectivity of Lane Markings. *JOURNAL OF INFRASTRUCTURE SYSTEMS*, *31*(3), 04025019. https://doi.org/10.1061/JITSE4.ISENG-2641

Ibne Idris, I., Mousa, M., Hassan, M., Carlson, P., & Greyson, E. (2024). Development of a Machine-Learning-Based Tool to Predict Retroreflectivity of Pavement Markings Across the U.S. *Transportation Research Record*, *2679*(3), 1021-1036. https://doi.org/10.1177/03611981241287196

Idris, I. I., Mousa, M., & Hassan, M. (2024). Modeling retroreflectivity degradation of pavement markings across the US with advanced machine learning algorithms. *Journal of Infrastructure Preservation and Resilience*, *5*(1), 3.

Jackels, J. P., Ethan. (2024). *FHWA Pavement Marking Retroreflectivity Requirements – Putting Research into Practice* CEAM Study Review Committee Meeting #1 – Kickoff,

Kong, W., Zhong, T., Mai, X., Zhang, S., Chen, M., & Lv, G. (2022). Automatic Detection and Assessment of Pavement Marking Defects with Street View Imagery at the City Scale. *Remote Sensing*, *14*(16).

Lee, S., & Cho, B. H. (2023). Evaluating Pavement Lane Markings in Metropolitan Road Networks with a Vehicle-Mounted Retroreflectometer and AI-Based Image Processing Techniques. *Remote Sensing*, *15*(7).

Lee, T. (2016). Advanced Methods for Mobile Retro-reflectivity Measurement on Pavement Marking Final Report [Tech Report]. https://rosap.ntl.bts.gov/view/dot/54370

Mamun, A. A., Ping, E. P., Hossen, J., Tahabilder, A., & Jahan, B. (2022). A comprehensive review on lane marking detection using deep neural networks. *Sensors*, *22*(19), 7682.

Manasreh, D., Nazzal, M., & Abbas, A. (2023). Feature-Centric Approach for Learning-Based Prediction of Pavement Marking Retroreflectivity from Mobile LiDAR Data. *Buildings*, *14*. https://doi.org/10.3390/buildings14010062

Migletz, J., Graham, J. L., Bauer, K. M., & Harwood, D. W. (1999). Field Surveys of Pavement-Marking Retroreflectivity. *Transportation Research Record*, *1657*(1), 71-78. https://doi.org/10.3141/1657-10

Mohammadi, A., Golazad, S., & Rashidi, A. (2024). A review of trends and practices in using visual data for construction-related machine learning models. 2024 Winter Simulation Conference (WSC),

Mohammadi, A., Medina, J., & Rashidi, A. Developing a Methodology to Estimate the Retroreflectivity of Longitudinal Pavement Markings using LiDAR.

Mohammadi, A., Medina, J. C., & Rashidi, A. Leveraging Machine Learning and LiDAR for Predictive Maintenance of Roadway Markings. In *Computing in Civil Engineering 2025: Computational and Intelligent Technologies* (pp. 196-205).

National Safety Council. (2024). *The most dangerous time to drive*. https://www.nsc.org/road/safety-topics/driving-at-night

Pike, A. M., Satterfield, C., Schertz, G., Cheung, J., & Falk, K. W. (2022). *Methods for Maintaining Pavement Marking Retroreflectivity* (FHWA-SA-22-028). Federal Highway Administration.

Satterfield, C. H., H. Gene Jr.; Schertz, Greg. (2014). *Assessment of Economic Impacts of SNPA Proposed Minimum Levels of Pavement Marking Retroreflectivity*. O. o. S. Federal Highway Administration.

Utah Department of Transportation. *Pavement Marking Projects and Studies*. Utah Department of Transportation. Retrieved July 10 from https://www.udot.utah.gov/connect/public/pavement-markings/pavement-marking-projects-and-studies/

Yeong, D. J., Velasco-Hernandez, G., Barry, J., & Walsh, J. (2021). Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review. *Sensors*, *21*(6), 2140.
Zhang, H. (2016). *Rapid inspection of pavement markings using mobile laser scanning point clouds* University of Waterloo].

APPENDIX: UDOT'S COMPUTER TOOL USER GUIDELINE TO ASSESS AND REPORT PAVEMENT MARKING RETROREFLECTIVITY

Overview

This guideline is intended to provide general information, instructions, and recommendations to enable the use of UDOT's LiDAR- and Image-Based Pavement Marking Retroreflectivity computer tool.

Outcomes from the computer tool support maintenance decisions leading to federally compliant longitudinal pavement marking retroreflectivity using a multi-source approach that:

1. Assesses retroreflectivity from routinely collected mobile LiDAR and street level imagery,
2. Aggregates results at operational user-defined segment-level scales,
3. Prioritizes maintenance based on data-driven results, and
4. Verifies and recalibrates models with limited, targeted field measurements.

Assessments are directed at durable pavement markings, as their in-service time is expected to be multi-year.

Data Scope and Coverage

- Lane Markings – Durable Markings:
    - Right edge line (REL).
    - Left edge line (LEL).
    - Dashed lines / Lane lines (LL).

- Baseline Model Coverage (Includes Most Durable Marking Mileage):

Table A1. Scope of Mileage Coverage in the Base Models by Freeway

Route	Increasing Direction MP	Decreasing Direction MP
I-15	257-290; 305-357	361-350; 330-301; 299-257
I-80	122-144	144-122
I-84	10-18; 81-102	102-81; 20-10
I-215	6-20	20-6

- Datasets:
    - 2023 mobile retroreflectometer measurements (≈0.1-mile segmentation).
    - 2023 mobile LiDAR (XYZ + intensity + RGB) - collected by Pathway Services Inc.
    - 2023 roadway imagery frames - collected by Pathway Services Inc.
- Evaluation Methods:
    - Fusion (LiDAR + Images).
    - LiDAR only.
    - Images only.

Methods & Pipeline

LiDAR-Only

- Ingest, index LiDAR files (.laz)
- Extract scaled XYZ, intensity, and RGB.
- Transform coordinates to WGS 84 to join with ground truth field data.

- Engineer features from LiDAR data (intensity and RGB/material properties).
- Train random forest model for classification and exponential/linear models for regression.
- Apply trained models to predict point and segment* level classification compliance (FHWA threshold).

Images-Only

- Ingest, index image files (.jpg, .jpeg, .png, .bmp).
- Auto crop patches around markings using polygon masks and percentile thresholding.
- Train Convolutional Neural Network (CNN) to assess classification considering class imbalance remedies (e.g., SMOTE) and validated by K Fold, and to support image-based regression.
- Apply trained models to predict point- and segment-* level classification compliance (FHWA threshold).

Fusion (LiDAR + Images)

- Ingest LiDAR- and image-derived predictions and/or features at the segment level.
- Pre-process and spatially align both sources on a common milepoint / segment ID and marking type.
- Train fusion model that combines LiDAR and image information.
- Apply trained models to predict point- and segment-* level classification compliance (FHWA threshold).

* Segments are units of aggregated data, where the level of aggregation is defined by the user (with a default value of 1 mile).

Classification Outcomes (Point or Segment Levels)

Before describing the classification outcomes, it is important to differentiate between sections, segments, and points, as defined in this guideline. The illustration below helps with understanding each concept:

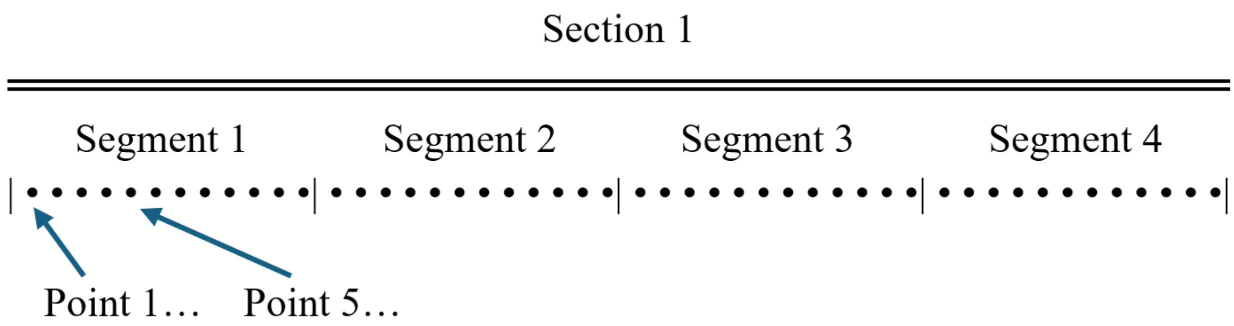


Figure A1. Schematic Representation of Sections, Segments, and Points

Sections are composed of segments (typically 1-mile long each), and segments contain multiple measurement points. Each point contributes a reading either from LiDAR or images, and there any many more points (about two orders of magnitude higher) in the LiDAR datasets compared to the image datasets. For example, a typical 1-mile segment uses about 10 images for the assessment of a segment, whereas 1000 LiDAR points could exist in that same segment.

While output files from the computer tool provide detailed quantitative point and segment assessments, more user-friendly classification outcomes and recommendations are also given in terms of the color-coding scheme shown in Table A2 below.

Table A2. Color-Coded Classification Definitions

Color-Coded Classification	Criteria (% of points at or above threshold*)	Interpretation	Recommendation
Green	≥ 80 %	Compliant	No Immediate Attention (Reassess next maintenance cycle)
Yellow	60 % to < 80 %	Mostly Compliant	Further Review (Limited field data collection / localized issues)
Orange	40 % to < 60 %	At Threshold	Treatment Required Before Next Survey (Likely not compliant before next survey)
Red	< 40 %	Non-Compliant	Treatment Required – Priority (Non-compliant / priority section)

* FHWA compliance threshold = 100 mcd/m²/lux (for high-speed longitudinal markings)

Note: Segment-level assessments are highly preferred over individual points. Segments take advantage of inherent correlation between adjacent points along the road, smoothing variability from individual LiDAR or image points.

Outputs

The computer tool produces the following outputs:

- CSV tables (separate tables for individual points and segments) containing:
    - Point and segment location (route, direction, milepoint/chain, coordinates).
    - Field-collected retroreflectivity.
    - LiDAR intensity and RGB values.
    - Predicted retroreflectivity and Root Mean Square Error (RMSE).
    - Classification outcome: color-coded classification.

- Plots with segment representation of classification outcomes including field-measured retroreflectivity vs. selected evaluation method: Fusion (LiDAR + Images), LiDAR only, Images only.
- GIS GeoPackage: Layer, tables with field-measured and predicted retroreflectivity, and classification outcomes for GIS-ready analysis.

Suggested Annual Assessment

Recurrent assessments of durable pavement markings are recommended at least once per year. This enables revaluation and planning of major maintenance efforts after a complete season cycle, capturing periodic performance changes on pavement markings. The steps below are suggested for this assessment:

- **Gather and Prepare Data Input:**
  - Obtain and organize recent mobile LiDAR survey files and/or associated imagery for roadway sections of interest. Separate folders for LiDAR and images are recommended.
  - LiDAR data can be extremely large, so consider handling smaller sections at a time (see estimated processing times below).
  - Images are much easier to manage than LiDAR data, and their processing time is only a small fraction of LiDAR's. However, Fusion or LiDAR are preferred methods over images, given the inherent stronger relationship of intensity to retroreflectivity.
  - Identify Sections (i.e., unique combinations of routes, lane markings, directions, and milepost ranges) to be assessed and organize them in a Task List to track progress as the analysis is conducted.
    - Data needs to be split into Sections based on the size and processing requirements of LiDAR datasets.

        - Plan for processing times accordingly, particularly for runs using LiDAR data. Estimated processing times for each Section using Fusion or LiDAR-only methods will vary depending on the Section length and the specifications of the computing machine used to run the tool, as follows:
            - Consumer-grade laptop: Every 1 mile of LiDAR ≥ 5 minutes.
            - Enterprise-grade workstation: Every 1 mile of LiDAR ≥ 2 minutes.
        - Evaluations using the Image-only method can contain long Sections without practical limitations, as the processing times are short and on the order of seconds or a few minutes.

- **Run Computer Tool and Record Initial Assessment:**
    - For each of the Sections in the Task List, complete a run using the tool (i.e., one run per Section).
    - Use separate, properly labeled folders to save the output files from each run.
    - For each folder, complete a high-level visual exploration of the output plots to identify sections that may require further attention:
        - Add a column to the Task List to record the Section classification outcomes.
        - For each folder (or Section), record in the Task List if all segments in the Section are "Compliant," or "Further assessment" otherwise (regardless of the number of segments different than Green).

- **Visualize Overall Segment Outcomes Using GIS:**
    - Using GIS software, load the output layers produced by the tool for all Sections in the Task List. This compilation of layers provides a visual overview of the study area and informs of possible efforts that can be combined into a single maintenance order to address issues along the same route.

- **Further Analysis and Maintenance Plans:**
    - Revisit the folder for each Section under the "Further Assessment" classification outcome in the Task List.
    - Identify potential maintenance efforts based on the degree of classification codes different than Green. Use the recommendation descriptions provided in Table A2 to prioritize Sections.
    - Develop a list of priority maintenance targets recording segment rankings (allowing ties), where each segment is identified by route, direction, and mile range.
    - Explore the spreadsheets with data summaries by chain (e.g., by segment) provided by the tool to further understand the prediction values and their variability.
    - Consider the following recommendations when developing maintenance plans:
        - Segments classified as Red ("Not Compliant") are likely to have upstream or downstream segments with similar classification, creating maintenance targets that extend beyond a single segment.
        - If a single isolated segment classification is different than "Compliant," with surrounding segments being compliant, a spot-check field review may be triggered to assess possible localized treatment issues.
- **Data Governance & Quality Assurance Considerations:**
    - Maintain a model registry (versioned model files, training data vintages, feature ranges).
        - In addition to a compilation of historical outputs in organized folders, it is also recommended to maintain a single GIS layer of all combined Sections for each annual assessment. This is useful for fast data exploration and to track changes in treatment performance over time.
    - Enforce feature clipping to the training domain.

        - Before running the models, ensure that input values (such as LiDAR intensity or RGB brightness) have not changed due to setup or sensor upgrades. For example, values in the data collected by Pathway Services Inc. range between 0 and 65,000 units. Therefore, unexpected values outside of this range may trigger a revision in the field data.

        - For new datasets and models, ensure that the range of values seen during model training is wider or equal to those points used during evaluation, preventing extrapolation (this is the case for the current setup using Pathway Services Inc., but it is important to keep in mind if the data collection setup is altered).

        - Identify potential data and model issues for LiDAR and image datasets by tracking trend shifting over time. Examples include: Unexpected systemic classification changes (both toward lower or higher estimates), increase in missing or inadequate data resulting in no classification outcomes, and overall changes in the rate of retroreflective estimates degradation on a year-over-year basis for the whole study area.

    - Keep logs from completed roadway-marking work orders (treatment date, type, material, and location). Such records can be linked to measured/predicted retroreflectivity over time to build degradation curves by material and/or corridor. These curves can support more precise projections to predict when markings are expected to drop below the threshold.

Ground-Truth Data Collection

Plan for limited, targeted field data collection using a retroreflectometer (manual or mobile setups) to verify models and maintain consistency over time, thereby reducing reliance on full-network mobile retroreflectometer surveys. Reserve systemwide ground-truth data collection for cases such as establishing a new baseline (e.g., every other year), major changes in LiDAR or imagery data collection setup (e.g., upgraded sensors or data postprocessing), or when model performance or data quality degrades materially.

Recommended Baseline Updates

Annual "Light-Touch" Calibration

- Use a manual or mobile retroreflectometer to sample a small percent (≤10%) of segments per route to collect readings from at least 30 locations or sample every ~1 mile per every 5 miles, whichever is larger. This "light-touch" calibration sample is large enough to detect systematic drifts in model predictions at the corridor level.

- Focus the sampling on Yellow or Orange classifications, with smaller frequency of locations classified as Red and Green. This strategy will strengthen the models near the threshold areas, improving decision making where accuracy is more critical.

Triggered Spot Checks

- Collect additional ground truth data when fusion vs. field retroreflectivity assessments systematically disagree over a tolerance level for at least five consecutive segments. A tolerance level may consider a threshold of >15 percentage points in the classification criteria at the segment level, and such threshold can be adjusted as needed.

Post-Treatment Verification

- For any treated segments, obtain retroreflectivity data from the verification runs after the treatment application is completed. Such measurements are expected to be available before the final receipt and closing of a task order. Keep track of treatments in historical logs to evaluate degradation on subsequent surveys.

Recalibration

- Recalibration refers to a more comprehensive effort than the "light-touch" calibration above, where LiDAR and image data is also included in the process and new observations are added to the models.

- Models can be refreshed annually (when new LiDAR/imagery surveys are completed).

- Annual recalibrations are optional, but they become a requisite if systemic prediction drift is detected or error metrics degrade materially.

- To simplify the recalibration efforts, identify and repeat the small set of data collection points used for the "light-touch" calibration to establish a consistent set of points per corridor/marking and anchor year-over-year continuity.

To perform the annual "light-touch" calibration, a few new field-measured retroreflectivity points are compared against prior years' field data and models to recompute key performance metrics. If performance remains at similar levels, the existing LiDAR, image, and fusion models are confirmed for continued use.

However, if drift or error metrics degrade materially, "recalibration" is expected, so that model coefficients and the underlying classifiers are retrained by adding observations to the model data files.

Implementation & Governance

- Model stewardship: Keep a registry of model versions, training data vintages, and performance metrics (accuracy, precision/recall, RMSE).

- Retrain or recalibrate when: (a) upgrades to LiDAR sensors, images, or data post-processing are introduced; (b) data distributions shift; (c) performance on the annual calibration sample degrades.

- Documentation: Archive inputs/outputs (e.g., CSV, GPKG, PNGs) and metadata from each run within a given assessment cycle.

- Integration. Consider pushing section-level classification outcomes and layers into UDOT's internal asset systems for work order creation and audit trails.

Limitations, Assumptions, and Risk Mitigation

- Sensor & illumination:

    - LiDAR intensity and imagery are influenced by collection settings and ambient conditions. This known limitation does not prevent the use of models for reliable approximations of retroreflectivity under repeatable data collection methods.

- However, mitigation is needed when data collection setups change, or when model prediction shifts are systematic. Model recalibration and basic feature scaling (where possible) are recommended in such cases, but annual light-touch calibration is sufficient to verify model performance.

- Data gaps & occlusions:
  - Moving or stationary vehicles, work zones, barriers, shadows, snowbanks, etc. can block the line of view of the LiDAR and/or imagery, leaving segments where the tool cannot reliably "see" the marking. Multiple consecutive segments without sufficient data should be identified as candidates for field verification (using a retroreflectometer or visual inspection) in such cases.

- Threshold policy changes:
  - Color-coding or performance thresholds may evolve over time, but these parameters are adjustable in the computer tool via recompilation of the original code. Qualified personnel with coding experience are expected to be able to perform such modifications and recompile the tool as a newer version. Documentation of versions is also part of the audit traces that must be tracked.

Computer Tool Running Instructions

Launch & Set Inputs

- Highway (I-15, I-80, I-84, I-215)
- Direction (Inc/Dec)
- Road Marking (REL, LEL, LL)
- Cycle (Year) for labeling
- Start/End Mileage
- Cluster Length (default set to 1 mile)

Choose Evaluation Method

- Fusion (LiDAR + Images) - Requires .laz + imagery
- LiDAR only - Requires .laz, no imagery.
- Image only - Requires imagery, no .laz files.
- Optional checkbox: Review/adjust auto crops before running image analysis.

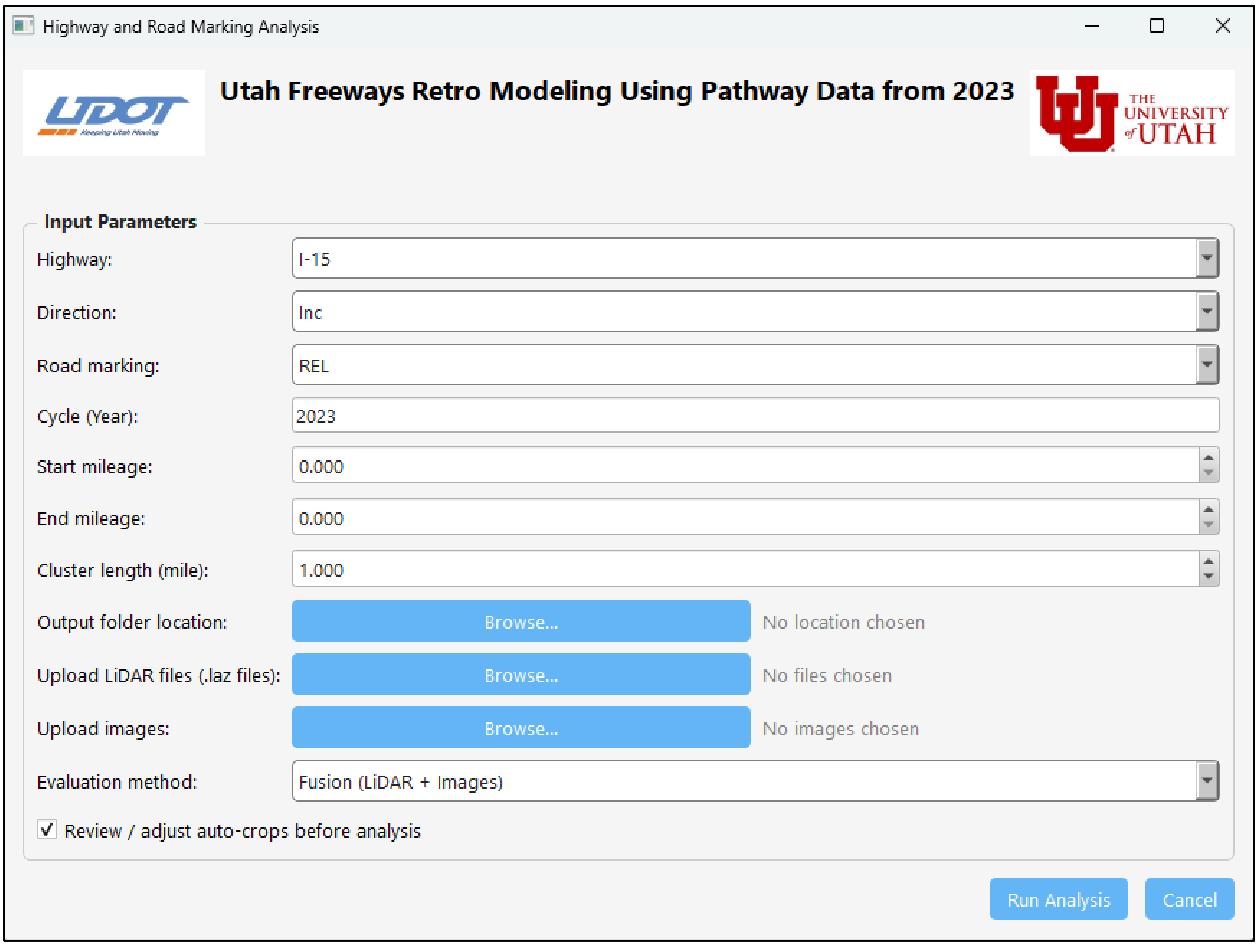


Figure A2. User Interface from the Computer Tool

Provide data

- Select and upload .laz files (LiDAR) and photos (if using imagery).
- Select Save location for outputs.

Run

- Click Run Analysis.
- This step will take a few minutes while the following operations are performed:
    - Read LAZ header and points; convert to WGS 84.
    - Load the correct models for the chosen corridor/marking.
    - Predict Class (Maintenance / No Maintenance) and Predicted Retro.
    - Aggregate by cluster length (default 1.0 mi).

Review Outputs

- Results table (first 1,000 rows preview).
- CSV with per chain summary (Section level stats + Green/Yellow/Orange/Red classification outcomes) and detailed point records.
- PNG plot of classification outcomes: Field-measured vs. LiDAR vs. Image vs. Fusion (according to the evaluation method).
- GeoPackage (.gpkg) for GIS.

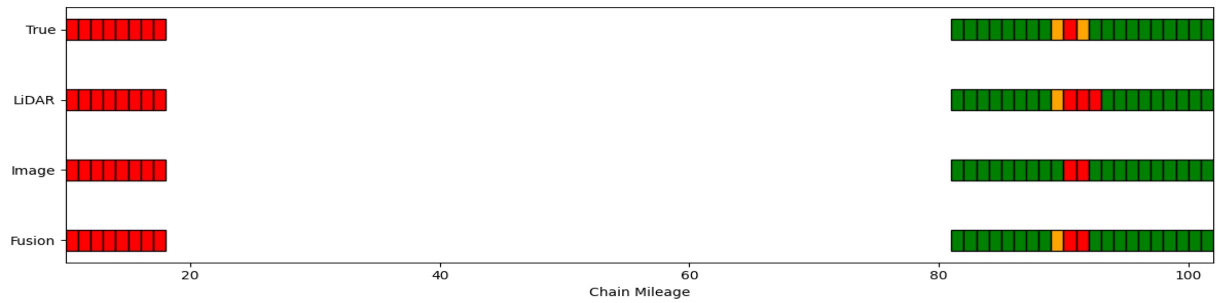
Assessment

- Follow instructions for Suggested Annual Assessment (Section 6) and Ground-Truth Data Collection (Section 7).
- Export maps/tables to support work orders and FHWA reporting.

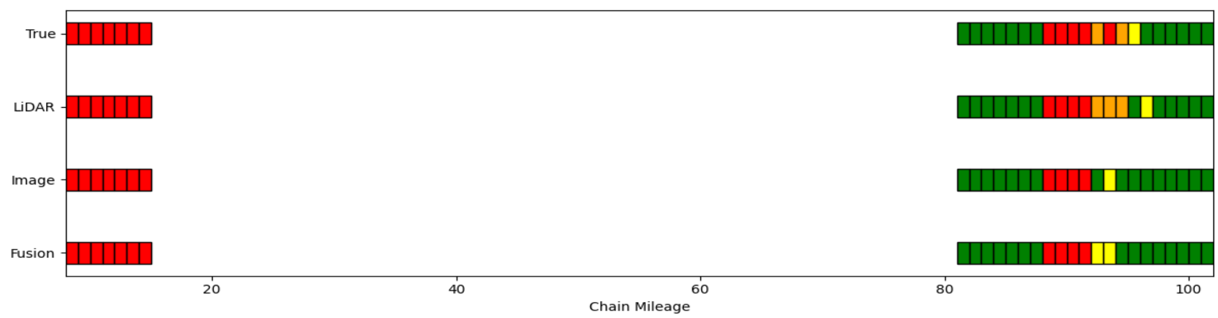
Example

Segment-level REL classification outcomes of two Sections (on each direction) using 2023 data (let's call them "West" and "East") are shown below in Figure A3, indicating close agreement between color codes for ground truth ("True" in the figure) and all three models for both increasing and decreasing directions. These results suggest that the tool predictions may be reliable enough to inform maintenance decisions at the 1-mile segment scale.

Now, if the LiDAR and image data from this example are assumed to be from 2024 without ground truth from the same year (but from 2023), the results from the "True" series will not be directly comparable but will serve as a template for overall expected trends (given some expected degradation between cycles).



I-84 REL - Increasing Direction



I-84 REL - Decreasing Direction

Figure A3. Color-Coding Outcomes for I-84 (Increasing and Decreasing Directions)

Running this scenario requires a significant amount of processing time using the tool, with 8 miles in the West section and 21 miles in the East section. According to the general estimation provided by the guideline, on a consumer-grade laptop the running time would be at least 2 hours and 25 minutes (29 miles x 5 minutes per mile / 60 minutes per hour) for each direction. However, these assessments are needed rather sparsely when new LiDAR or image data becomes available, so adequate planning is expected to be able to accommodate such time requirements.

Recall that image-only evaluations can be a valid, faster alternative for segment-level compliance predictions when LiDAR is unavailable. However, as noted previously, Fusion or LiDAR-only are preferred methods over images-only.

Moreover, the Fusion model can be generally taken as a default go-to approach when both LiDAR and image datasets are in close agreement with each other.

The following is a general assessment of the segments plots in Figure A3:

Increasing Direction:

- West Section:
    - Entirely "non-compliant," treatment is required and should be considered as high priority.
- East Section:
    - Most segments are classified as "Compliant," no immediate action needed for compliant segments.
    - Mileposts 89 – 92 classified either as "At threshold" or "Non-compliant," and immediate action is required.

Decreasing Direction:

- West Section:
    - Entirely "non-compliant," treatment is required and should be considered as high priority.
- East Section:
    - Slight majority of segments classified as "Compliant." No immediate action needed for compliant segments.
    - Mileposts 88 – 97 are classified either as "Mostly Compliant," "At threshold" or "Non-compliant," (Yellow, Orange, or Red) and immediate action is required for

segments starting near milepost 88. However, the approximate location where new treatment is needed may vary between milepost 94 (the east-most Yellow segment from the Fusion) and milepost 97 (the east-most Yellow segment from the LiDAR-only model).

In general, and for both directions, the exact location of the beginning and end of the section to be re-treated needs to be determined through a field inspection, where the contrast between the compliant and other classifications is expected to be recognizable by experienced personnel.

The tool output tables can also be used to verify the accuracy of the models when the data and grounds truth are collected in the same year. In this example, figures for both the increasing and decreasing directions are shown in Figure A4, where the agreement between field data and the image model achieves satisfactory outcomes, with $R^2 \approx 0.84–0.86$ and RMSE ≈ 22–67 (mcd/m²/lux). Similar plots can be generated for other models, in this case likely with similar outcomes.

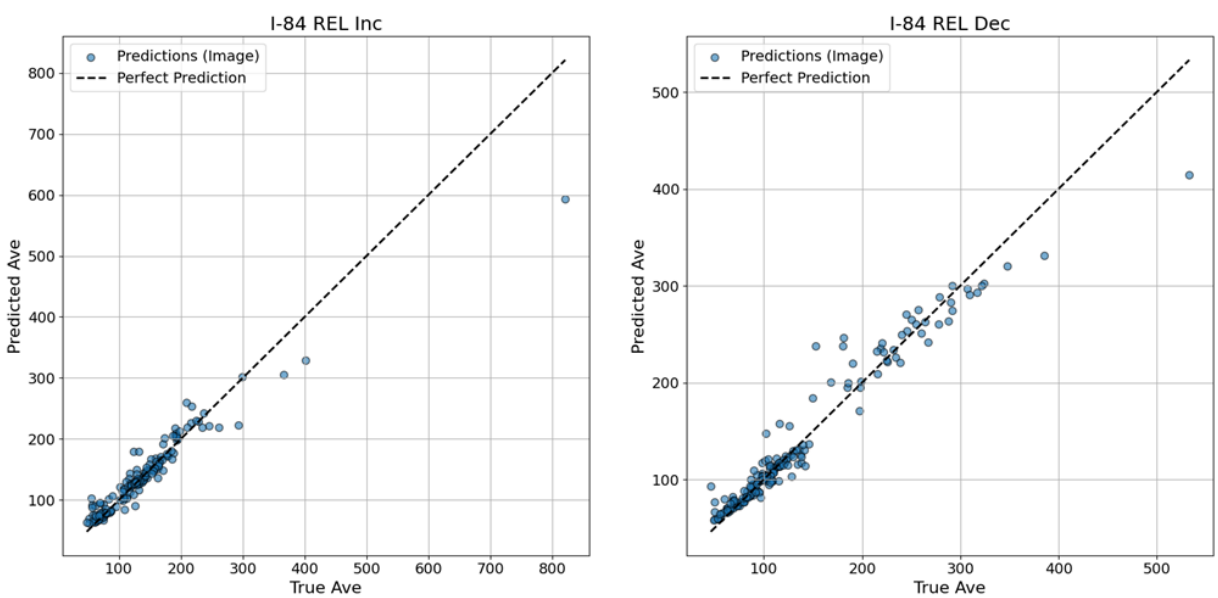


Figure A4. Color-Coding Outcomes for I-84 (Increasing and Decreasing Directions)

This example illustrates an assessment of two Sections in each direction along the I-84 corridor using all three evaluation methods (Fusion, LiDAR-only, Images-only) and describes possible interpretations of the classification and model regression outcomes in terms of pointers for maintenance decisions.

Additional exploration can be completed using the tool output files, including analysis on the data variability or percentiles of the predicted retroreflectivity to refine preliminary actionable items.

It is important to highlight that these assessments are intended to identify general areas of interest in terms of maintenance plans, but additional field-based observations through site visits are necessary to pinpoint the exact extent of the segments to be treated.