GEORGIA DOT RESEARCH PROJECT 23-04

Final Report

DEVELOPMENT OF AN ML-BASED GEORGIA PAVEMENT STRUCTURAL CONDITION EVALUATION SYSTEM

Office of Performance-based Management and Research
600 West Peachtree Street NW | ATLANTA, GA

July 2026

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.:
FHWA-GA-26-2304</td><td>2. Government Accession No.:
N/A</td><td colspan="2">3. Recipient's catalog No.:
N/A</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle:
Development of an ML-based Georgia Pavement Structural Condition Evaluation System</td><td colspan="2">5. Report Date:
July 2026</td></tr>
<tr><td colspan="2">6. Performing Organization Code:
N/A</td></tr>
<tr><td colspan="2">7. Author(s):
Yichang (James) Tsai, Ph.D., P.E. https://orcid.org/0000-0002-6650-2279
Zhongyu Yang, Ph.D. https://orcid.org/0000-0002-7535-0924
Bingjie Lu https://orcid.org/0009-0001-2381-0394</td><td colspan="2">8. Performing Organ. Report No.:
23-04</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address:
Georgia Institute of Technology
790 Atlantic Drive,
Atlanta, GA 30332-0355</td><td colspan="2">10. Work Unit No.:
N/A</td></tr>
<tr><td colspan="2">11. Contract or Grant No.:
PI# 0019906</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address:
Georgia Department of Transportation
Office of Traffic Operations
600 West Peachtree St. NW
Atlanta, GA 30308</td><td colspan="2">13. Type of Report and Period Covered:
Final Report
(January 2024 - July 2026)</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code:
N/A</td></tr>
<tr><td colspan="4">15. Supplementary Notes:
Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.</td></tr>
<tr><td colspan="4">16. Abstract:
Georgia's aging interstate highway network requires more reliable pavement evaluation to support maintenance and rehabilitation (M&R) planning. Current network-level practices rely primarily on surface condition data, which can lead to project-level cost overruns when structural deficiencies are discovered only after budgets have been defined from surface indicators. Traffic Speed Deflection Device (TSDD) technology enables continuous structural response measurement at highway speeds, but Georgia-specific TSDD-based screening rules calibrated against laboratory performance references have not been established. In addition, an integrated framework for combining structural and surface condition in treatment decision-making has not yet been implemented. This study develops Georgia-specific pavement structural condition classification methods and an integrated surface-structural treatment decision framework. The analysis uses TSDD, Ground Penetrating Radar (GPR), Three-Dimensional (3D) pavement surface data, and Hamburg Wheel-Track Testing (HWTT) results from I-59, I-285, and I-575. A quality-controlled dataset of 85 flexible pavement sections was constructed through missing data analysis, pavement-type screening, and a nearest-neighbor spatial joining process. Threshold-based and machine learning (ML) methods were evaluated under the same repeated stratified cross-validation (CV) protocol. The single-variable SCI_8 threshold achieved a mean F1 score of 0.806 and a mean Area Under the Curve (AUC) of 0.888 across 50 validation splits, providing a simple and interpretable candidate baseline for pilot structural screening. The SCI_8 + AC thickness duo-variable threshold achieved the highest mean F1 score among all methods, with F1 = 0.834 and AUC = 0.887, showing that GPR-derived asphalt thickness improves structural detection. Among the ML classifiers, Random Forest (RF) performed best, with F1 = 0.822 and the highest AUC of 0.891. A staged pilot implementation pathway is recommended: SCI_8 threshold screening for near-term pilot use, SCI_8 + AC thickness screening when reliable GPR data are available, and RF as a scalable ML engine as labeled data expand. The treatment framework translates structural and functional condition into four candidate M&R categories: No Action, Milling and Resurfacing, Close Monitoring, and Major Rehabilitation. Corridor-level maps demonstrate that the framework can distinguish surface-driven needs, structure-driven monitoring needs, and combined deterioration patterns, establishing an initial Georgia-specific research-scale foundation for incorporating TSDD-based structural screening into GDOT's pavement management system (PMS), subject to further corridor-level validation.</td></tr>
<tr><td colspan="2">17. Key Words:
Traffic Speed Deflection Device, Pavement Structural Condition, Ground Penetrating Radar, Threshold-Based Classification, Machine Learning, Pavement Management System</td><td colspan="2">18. Distribution Statement:
No Restriction</td></tr>
<tr><td>19. Security Classification (of this report): Unclassified</td><td>20. Security Classification (of this page): Unclassified</td><td>21. No. of Pages: 172</td><td>22. Price:
Free</td></tr>
</table>

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

GDOT Research Project No. 23-04

Final Report

DEVELOPMENT OF AN ML-BASED GEORGIA PAVEMENT STRUCTURAL CONDITION EVALUATION SYSTEM

By

Yichang (James) Tsai, Ph.D., P.E.
Professor of Civil and Environmental Engineering

Zhongyu Yang, Ph.D.
Research Engineer

Bingjie Lu, Ph.D. Student
Graduate Research Assistant

Georgia Tech Research Corporation

Contract with
Georgia Department of Transportation

In cooperation with
U.S. Department of Transportation
Federal Highway Administration

July 2026

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. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

3D	Three-Dimensional
AC	Asphalt Concrete
AUC	Area Under the Curve
CV	Cross-Validation
FHWA	Federal Highway Administration
FWD	Falling Weight Deflectometer
GDOT	Georgia Department of Transportation
GPR	Ground Penetrating Radar
HWTT	Hamburg Wheel-Track Testing
IRI	International Roughness Index
M&R	Maintenance and Rehabilitation
ML	Machine Learning
OMAT	Office of Materials and Testing
PCC	Portland Cement Concrete
PES	Pavement Evaluation Summary
PMS	Pavement Management System
RF	Random Forest
RFECV	Recursive Feature Elimination with Cross-Validation
SCI	Structural Curvature Index
TSD	Traffic Speed Deflectometer
TSDD	Traffic Speed Deflection Device
VIF	Variance Inflation Factor

EXECUTIVE SUMMARY

Georgia's interstate pavement network is under increasing pressure from aging infrastructure, heavy traffic demand, and growing maintenance and rehabilitation (M&R) needs. Current network-level pavement management practices rely primarily on surface condition indicators, such as cracking, rutting, and roughness. Although these indicators are important for evaluating functional pavement condition, they do not always represent the underlying structural capacity of the pavement. This limitation is especially important for pavements that have experienced repeated resurfacing, where the surface may appear acceptable while the underlying asphalt layers or supporting structure have already weakened. When structural deficiencies are discovered only after project scopes and budgets have been developed, agencies may face cost overruns, delayed project delivery, and late-stage treatment revisions. Therefore, Georgia Department of Transportation (GDOT) needs a practical network-level structural screening approach that can identify underlying structural risks earlier in the planning process.

Traffic Speed Deflection Device (TSDD) technology provides a practical way to address this need by collecting dense, continuous pavement structural response data at highway speeds, making it more suitable than traditional project-level structural testing for network-level pavement evaluation. To support GDOT pavement management, TSDD response indicators must then be calibrated to classify pavement sections as structurally Good or Poor under Georgia pavement conditions. This calibration is necessary because TSDD indicators are affected by asphalt thickness, base support, traffic loading, and environmental conditions, so thresholds developed elsewhere should not be directly adopted for Georgia without local validation. The resulting structural condition information must also be integrated with surface condition data because most existing GDOT pavement management workflows remain primarily surface-condition based and do not explicitly combine structural and functional condition for treatment screening.

The objectives of this project are: 1) to use TSDD-derived structural indicators, Ground Penetrating Radar (GPR)-derived layer thickness information, and Hamburg Wheel-Track Testing (HWTT) laboratory performance outcomes to develop and evaluate Georgia-specific pavement structural condition classification methods, including threshold-based and machine learning (ML)-based approaches, under a consistent validation framework; and 2) to develop an integrated surface–structural treatment screening framework that translates structural and functional condition information into Pavement Management System (PMS)-oriented M&R screening categories. The following are the major conclusions and implementation implications of this research project:

1) **A high-quality but limited coring-centered analytical dataset containing 85 records was developed to support model development.** The dataset was developed from three GDOT-provided asphalt interstate corridors, I-285, I-575, and I-59, where Pavement Evaluation Summary (PES) coring records and HWTT results were available as project-level reference data. For each record, the coring location was used as the spatial anchor, and the Layer 1 HWTT Pass/Fail result was used as the laboratory performance label. Nearby TSDD and GPR measurements were assigned to each coring location using nearest-neighbor spatial matching, while 3D surface condition data were also linked for later surface–structural interpretation. Records were manually reviewed and retained only when the coring location and matched field measurements represented the same flexible

mainline pavement section. This strict screening and data quality control process supports a reliable training and validation dataset, but it limits the sample size and corridor diversity available for model development. GDOT is recommended to continue collecting high-quality and diverse coring/HWTT reference data, paired with TSDD and GPR measurements, to support future model refinement and improve robustness and generalization.

2) **A simple SCI_8 threshold provides the most practical near-term starting point for GDOT pilot structural screening.** Among all evaluated single-variable threshold rules, SCI_8 provided the strongest standalone classification performance, with a mean F1 score of 0.806 and a mean AUC of 0.888 across 50 repeated stratified cross-validation splits. Based on the current 85-record dataset, a preliminary SCI_8 threshold of approximately 0.585 mils was identified. Because this rule requires only TSDD data, uses a single decision boundary, and is easy to interpret, it is the most practical candidate for near-term pilot use, subject to further validation across additional corridors.

3) **The SCI_8 + AC thickness duo-variable threshold method can improve structural screening when reliable GPR data are available.** The method accounts for the fact that the same SCI_8 value may have different structural meanings in thin and thick asphalt pavement sections. It achieved the highest mean F1 score among all evaluated methods, with F1 = 0.834 and AUC = 0.887. Compared with the single-variable SCI_8 threshold, it improved recall and reduced missed structural failures while maintaining similar overall discrimination. These results suggest that GPR-derived AC thickness provides important context for interpreting TSDD response, although the calibrated AC thickness split point

and group-specific SCI_8 thresholds should be updated as more labeled sections become available.

4) **ML is a promising future prediction engine, but it should not replace threshold-based screening at the current dataset scale.** Among the four ML classifiers evaluated, Random Forest (RF) achieved the strongest ML performance, with a mean F1 score of 0.822 and the highest mean AUC of 0.891. However, RF did not outperform the SCI_8 + AC thickness duo-variable threshold in F1 score, and the dominant predictors were a compact set of physically meaningful variables, including D0, SCI_8, and AC thickness. Therefore, RF should be retained as a future retrainable ML option as GDOT collects more high-quality and diverse labeled data, rather than used as the immediate replacement for transparent threshold-based screening.

5) **An integrated surface–structural treatment screening framework has been developed to translate condition classification results into candidate PMS-oriented M&R screening categories.** The framework combines predicted structural condition with functional surface condition derived from 3D pavement surface indicators. Four candidate screening categories are defined: No Action, Milling and Resurfacing, Close Monitoring, and Major Rehabilitation. The Close Monitoring category is especially important because it identifies sections with acceptable surface condition but potential underlying structural concerns that may not be captured by a surface-only PMS workflow.

6) **A staged pilot implementation strategy is recommended for GDOT.** First, GDOT should pilot the SCI_8 single-variable threshold as a near-term structural screening candidate because it is simple, stable, interpretable, and requires only TSDD data. Second,

where reliable GPR-derived AC thickness data are available, GDOT should use the SCI_8 + AC thickness duo-variable threshold as an enhanced screening option, with periodic recalibration as more labeled data become available. Third, GDOT should maintain RF as a future retrainable ML-based structural prediction engine as additional labeled data and independent validation become available. To support implementation, GDOT should expand paired TSDD, GPR, 3D surface, and HWTT data collection to additional corridors, standardize coring/HWTT data formats, and validate the treatment decision categories through project-level field investigation, coring, drainage review, and cost analysis.

Overall, this study establishes an initial Georgia-specific, research-scale foundation for incorporating TSDD-based structural screening into network-level pavement management. The results show that TSDD and GPR data can support interpretable structural screening rules calibrated against HWTT outcomes, that ML provides a scalable future modeling pathway, and that structural predictions can be translated into PMS-oriented candidate treatment screening recommendations when combined with surface condition data. These findings should be advanced through staged pilot implementation, continued reference-data collection, and additional validation before broader operational deployment.

CHAPTER 1. INTRODUCTION

1.1 Research Background

Georgia's aging interstate highway network is facing growing maintenance and rehabilitation (M&R) needs, making reliable project scoping, treatment selection, and budget planning increasingly important. Georgia's pavement management system (PMS) has historically relied on pavement surface distress indicators, such as cracking, rutting, and roughness, to support M&R project selection and programming (Elgamal et al., 2025). These indicators are cost-effective, operationally accessible, and well suited for network-level surveys. However, they do not always reflect the underlying structural capacity of the pavement, especially on pavements that have undergone multiple resurfacing cycles, where the surface may appear acceptable while the underlying structural layers have weakened (Zhang et al., 2024a). Some GDOT projects have experienced substantial cost increases when structural deficiencies were identified late in the project development process. These deficiencies may have been identified earlier if structural condition evaluation had been incorporated into the network-level planning stage. Therefore, relying primarily on surface condition data can lead agencies to underestimate structural rehabilitation needs during project scoping and budgeting.

The challenge is not whether structural information is useful, but how it can be collected and interpreted at the network level. Traditional structural evaluation using the Falling Weight Deflectometer (FWD) provides detailed site-specific deflection data, but it is difficult to scale for network-level screening because it requires stop-and-go measurements, approximately one to two minutes of stationary time per test location, and traffic control for safe operation (Schmalzer, 2006).

Traffic Speed Deflection Device (TSDD) technology provides a practical alternative by collecting deflection bowl measurements and structural curvature indices continuously at highway speed. This allows pavement structural information to be collected at the spatial density and operating speed needed for network-level management (Xiao et al., 2021). Transportation pooled fund research has also emphasized the value of incorporating pavement structural information into network-level pavement asset management (Katicha et al., 2022). When incorporated into planning and programming, TSDD-derived structural condition information can help identify structural needs earlier and reduce the likelihood of late-stage scope, budget, or treatment revisions caused by previously unidentified structural deficiencies (Weitzel et al., 2024).

Despite these advantages, several technical and implementation challenges still limit the routine incorporation of TSDD data into GDOT's PMS. Four key challenges have been identified in prior research and practice:

Challenge 1: Data quality. TSDD measurements can be affected by environmental, operational, device-related, and pavement-surface factors, including pavement temperature, survey speed, vehicle dynamics, sensor calibration, sensor stability, and surface roughness or cracking. These factors may introduce noise, bias, or instability into measured deflection velocity, slope, and derived deflection-basin indicators (Wang et al., 2025; Zofka & Sudyka, 2015). Therefore, quality assurance, calibration review, signal screening, and data filtering are necessary before TSDD-derived indicators are used for network-level structural screening.

Challenge 2: Data processing. TSDD measurements collected from multiple Doppler laser sensors require substantial post-processing before they can be used for structural evaluation. Because the primary measurements are deflection velocity or slope rather than direct static deflection, processing may include sensor filtering, speed and temperature correction, conversion to deflection-basin or curvature-based indicators, integration with Ground Penetrating Radar (GPR)-derived layer thickness information, and, where needed, backcalculation or mechanistic interpretation. Existing pooled-fund work provides guidance for TSDD data collection and interpretation, but no single universally adopted processing standard currently governs the full chain from raw sensor measurements to agency-ready structural condition metrics (Sergeson, 2024). As a result, differences among vendor, agency, and research workflows may lead to inconsistencies in how TSDD-based structural metrics are derived, compared, and implemented (Huang et al., 2022a).

Challenge 3: Structural condition classification. After TSDD-based structural indicators are derived, agencies still need a practical classification method to translate those indicators into structural screening categories, such as Good or Poor, that can support planning decisions. This translation is not universal because TSDD-derived metrics, including deflection-basin parameters and structural curvature indices, are affected by pavement structure, layer thickness, material properties, and subgrade support. Therefore, threshold values or classification models developed in one state should not be directly transferred to another state without local calibration and validation (Huang et al., 2022a; Jiang et al., 2025).

Challenge 4: Surface-structural condition integration. Most existing PMS workflows rely primarily on surface condition metrics and do not explicitly account for underlying structural adequacy. As a result, pavements with acceptable surface condition but poor structural capacity may be overlooked during network-level planning. Recent studies, including work by the Virginia Department of Transportation (VDOT), have explored decision frameworks that incorporate

TSDD-based structural condition information into PMS treatment selection by combining structural and surface condition information. However, these approaches are still emerging rather than routine in many agency PMS workflows. A practical framework is therefore needed to translate combined surface and structural information into candidate M&R screening categories that can support planning, prioritization, and project-level follow-up review (Katicha et al., 2020b; Murekye et al., 2024).

This study focuses on Challenge 3 and Challenge 4: structural condition classification and surface–structural condition integration. Challenge 1 and Challenge 2, related to TSDD data quality and data processing, are acknowledged as important implementation issues but are outside the primary analytical scope of this study. The analysis assumes that the TSDD data collected by the GDOT-hired vendor and the structural indicators extracted from those data are sufficiently reliable for model development. Using these TSDD-derived structural indicators, GPR-derived pavement thickness, and HWTT reference labels, this study develops Georgia-specific structural condition classification methods and then incorporates the resulting structural classifications into an integrated surface–structural treatment screening framework.

1.2 Research Objectives and Work Tasks

The objectives of this research are: (1) to develop and evaluate Georgia-specific pavement structural condition classification methods using threshold-based techniques and ML algorithms with TSDD-derived structural indicators, GPR-derived pavement thickness, and HWTT reference labels; and (2) to develop an integrated surface–structural treatment screening framework that jointly considers pavement surface and structural conditions to support GDOT's M&R planning. To accomplish these objectives, the research is organized into five major work tasks:

Work Task 1: Literature Review. Review prior research on pavement structural evaluation, surface condition assessment, TSDD technology, GPR layer-thickness measurement, 3D pavement surface condition data, and PMS-oriented surface–structural decision frameworks. This task establishes the technical background, implementation challenges, and research gaps that motivate the study.

Work Task 2: Data Collection and Preparation. Collect, organize, screen, and integrate GDOT-provided coring/HWTT, TSDD, GPR, and 3D pavement surface condition datasets. This task prepares the data foundation needed for structural condition model development, model evaluation, and surface–structural treatment screening.

Work Task 3: Threshold-Based Method Development. Develop threshold-based structural screening methods using TSDD-derived structural indicators and GPR-derived pavement thickness information, with HWTT outcomes used as laboratory reference labels. This task focuses on creating transparent and interpretable screening rules that can support pilot-level structural condition classification.

Work Task 4: Machine Learning Model Development. Develop and evaluate ML-based structural condition classification methods using TSDD-derived indicators and GPR-derived pavement thickness information. This task examines whether ML models can provide a scalable and retrainable prediction option as additional labeled data become available.

Work Task 5: Treatment Decision Framework. Develop an integrated surface–structural treatment screening framework that combines predicted structural condition with functional surface condition derived from 3D pavement surface condition data. This task translates separate structural and surface condition information into PMS-oriented candidate M&R screening categories for GDOT planning and follow-up review.

1.3 Report Organization

This report is organized into the following chapters and appendices:

Chapter 2: Literature Review reviews prior research on pavement structural evaluation, TSDD technology, GPR, 3D pavement surface condition data, and PMS-oriented surface–structural decision frameworks. The chapter first summarizes the evolution from traditional structural testing to traffic-speed structural evaluation, then discusses TSDD operating principles, structural indicators, data quality, data processing, and multi-sensor integration. It also reviews threshold-based and ML-based structural classification methods and identifies the research gaps addressed in this study.

Chapter 3: Data Collection and Preparation describes the GDOT-provided datasets and the procedures used to construct the final analytical dataset. The chapter introduces the Type I network-level sensing data and Type II project-level evaluation data, then presents the missing-data analysis and corridor selection process. It also documents the spatial integration procedure, distance diagnostics, map-based quality control, coring-centered refinement, final dataset composition, and descriptive statistics for the retained corridors.

Chapter 4: Experimental Design and Threshold-Based Structural Condition Evaluation establishes the analytical framework for threshold-based structural condition classification. The chapter defines the Layer 1 HWTT Pass/Fail outcome as the laboratory reference label, describes the candidate TSDD and GPR indicators, and explains why 3D pavement surface distress indicators are excluded from structural prediction. It then presents the repeated stratified cross-validation protocol and evaluates both the SCI_8 single-variable threshold rule and the SCI_8 + AC thickness duo-variable threshold framework.

Chapter 5: Machine Learning-Based Pavement Structural Condition Evaluation develops and evaluates ML-based structural condition classification methods using the same candidate feature space and validation protocol as Chapter 4. The chapter describes the selected classifiers, training-only feature selection procedure, model configuration, and cross-validation design. It then presents feature selection results, model performance, precision-recall behavior, error analysis, and the comparative role of ML relative to threshold-based screening.

Chapter 6: Treatment Decision Framework presents the integrated surface–structural treatment screening framework. The chapter explains how predicted structural condition from TSDD and GPR data is combined with functional condition derived from 3D pavement surface indicators. It introduces the four-category M&R screening matrix, demonstrates the workflow on a representative I-285 segment, and presents corridor-level treatment maps for I-285, I-575, and I-59.

Chapter 7: Conclusions and Recommendations summarizes the major findings of the study and provides recommendations for GDOT implementation. The chapter organizes the conclusions around data construction, threshold-based structural classification, ML-based prediction, and integrated surface–structural treatment screening. It also discusses staged implementation, future data expansion, model updating, treatment decision validation, and economic evaluation needs.

Appendix A: TSDD Data Dictionary defines the TSDD variables used in this study, including location and section referencing fields, structural curvature indices, deflection bowl measurements, slope indicators, survey environment and loading variables, and raw diagnostic fields.

Appendix B: 3D Surface Data Dictionary defines the 3D pavement surface variables used to characterize functional pavement condition, including cracking, rutting, roughness, surface texture, pothole, joint and faulting, and roadway geometry variables.

Appendix C: GPR Data Dictionary defines the GPR-derived pavement layer variables and interpretation fields, including route and lane referencing information, AC thickness, PCC thickness, base thickness, data availability indicators, and pavement-type interpretation notes.

Appendix D: Coring, HWTT, and Spatial Integration Records documents the standardized coring and HWTT fields used in this study, including roadway and coring location fields, HWTT outcomes by layer, coring-measured thickness fields, GPR reference fields, and spatial integration metadata.

Appendix E: Final Analytical Dataset Structure and Field Dictionary documents the complete field structure of the final coring-centered analytical dataset after spatial integration, map-based quality control, and coring-centered refinement.

CHAPTER 2. LITERATURE REVIEW

This chapter reviews prior research needed to support the two main components of this study: TSDD-based structural condition classification and surface–structural treatment screening for pavement management. Section 2.1 traces the evolution of pavement structural evaluation methods and establishes the motivation for traffic-speed structural assessment. Section 2.2 reviews TSDD technology, including global adoption, measurement principles, data quality and processing requirements, and key structural indicators. Section 2.3 covers complementary sensing technologies, including GPR and 3D pavement surface condition data, and discusses the role of multi-sensor integration. Section 2.4 reviews structural condition classification methods, including threshold-based and ML-based approaches. Section 2.5 discusses the integration of surface and structural condition information into PMS-oriented decision frameworks. Section 2.6 synthesizes the research gaps and explains how the present study addresses them. Section 2.7 provides a summary of the chapter.

2.1 Evolution of Pavement Structural Evaluation

Pavement evaluation generally includes two complementary components: functional assessment and structural evaluation (Zhang et al., 2024b). Functional assessment characterizes observable surface conditions such as cracking, rutting, and roughness, while structural evaluation assesses the pavement's load-carrying capacity and its ability to support traffic over the remaining service life (Elgamal et al., 2025). Because functional condition data can be collected efficiently at the network level, they have historically formed the backbone of PMS workflows. However, surface condition alone may not fully represent structural adequacy, especially on aging multilayered pavements where repeated resurfacing can preserve an acceptable surface appearance while underlying structural layers continue to weaken. In such cases, relying primarily on surface indicators may bias planning decisions toward surface-focused treatments even when structural deficiencies require more substantial intervention.

Early pavement deflection devices provided the foundation for structural evaluation but were limited in their ability to support continuous network-level screening. Devices such as the Benkelman Beam and the Lacroix deflectograph measured pavement deflection under static or slow-moving loading conditions (Smith et al., 2017). However, these systems relied on manual or semi-automated procedures, discrete measurement locations, and low operating speeds. Reported operating speeds were typically about 1.5 to 7 km/h, approximately 0.9 to 4.3 mph (Sjögren & Andrén, 2012). Although these methods were valuable for localized or project-level evaluation, their low productivity made large-scale structural surveys difficult to conduct efficiently.

The Falling Weight Deflectometer (FWD) later became a standard tool for project-level structural evaluation because it applies an impulse load to the pavement surface and measures the resulting deflection basin. The FWD has been widely used for more than 40 years and remains an important reference method for pavement structural assessment (Scullion et al., 2025). However, FWD testing still requires stop-and-go operation, with approximately one to two minutes of stationary testing time at each location. This point-based testing process can lead to low productivity, traffic disruption, and traffic control or lane closure requirements for corridor-wide applications (Mabrouk et al., 2022). In addition, discrete testing intervals may miss localized structural anomalies between test locations, and stationary impulse loading does not fully represent the dynamic response of pavement under moving traffic loads (Chen et al., 2025). These limitations restrict the practicality of FWD-based testing for routine network-level structural screening.

Traffic Speed Deflection Device (TSDD) technologies were developed to make pavement structural evaluation more feasible at the network level. Unlike conventional stop-and-go deflection testing, TSDDs measure pavement response under moving wheel loads at or near traffic speed. This technology class includes several device concepts, including the Rolling Wheel Deflectometer (RWD), the Rapid Pavement Tester (RAPTOR), and the Traffic Speed Deflectometer (TSD) (Katicha et al., 2022). Their development shows a technical progression from distance-based displacement measurement toward velocity-based deflection measurement.

The RWD was one of the earliest traffic-speed deflection systems developed for network-level pavement structural assessment. The device applies a moving wheel load and uses laser sensors to compare the distance to the deflected and unloaded pavement surface. This approach allows continuous deflection profiles to be collected within the traffic stream and provides much denser spatial coverage than FWD point testing. However, because RWD estimates pavement deflection from laser distance or displacement measurements, its accuracy depends on stable geometric referencing, vehicle motion compensation, pavement surface condition, and calibration. Surface texture, surface irregularities, and reference stability are therefore important considerations for RWD interpretation (Elseifi & Elbagalati, 2017).

RAPTOR represents a later distance-based traffic-speed deflection system. It was developed by Dynatest and the Technical University of Denmark and uses an array of line lasers to scan a strip of pavement near the moving wheel load. Compared with earlier distance-based systems such as RWD, the use of line lasers reduces the influence of pavement texture by averaging the scanned profile. However, the measurement is obtained at an offset from the wheel-load path rather than directly along the wheel-load path. Therefore, RAPTOR improves the stability of laser displacement measurement, but its interpretation must still account for wheel-load offset, surface texture, sensor referencing, vehicle motion, beam alignment, and calibration (Katicha et al., 2025).

Modern TSD systems represent a different measurement principle within the broader TSDD class. Rather than estimating pavement displacement from laser distance measurements, the TSD uses Doppler-shift laser sensors to measure the vertical velocity of pavement surface movement under a rolling wheel load. The measured vertical deflection velocity is divided by vehicle speed to obtain deflection slope. These slope measurements can be used directly as structural response indicators or converted into deflection-basin parameters through numerical integration or curve-fitting methods. This velocity-based principle reduces the direct influence of surface texture and roughness compared with distance-based displacement systems, although TSD data quality is still affected by temperature, vehicle speed, load configuration, beam stability, Doppler sensor alignment, and calibration.

The TSD has also evolved substantially in sensor configuration and data resolution. First developed in Denmark in the late 1990s and commercialized by Greenwood Engineering in 2004, early TSD systems used a limited number of Doppler laser sensors to measure pavement response near the loading area (Jiang et al., 2025). Later systems expanded the sensor array to better capture the deflection response at multiple offsets from the rear axle. Modern TSD systems may incorporate up to ten sensors, while newer devices such as the Measure deflectometer use thirteen sensors to provide enhanced spatial coverage of the deflection basin. Improvements in sensor density, measurement frequency, beam rigidity, motion compensation, and thermal stabilization have progressively improved the reliability and resolution of traffic-speed structural measurements (Katicha et al., 2022).

Overall, RWD, RAPTOR, and modern Doppler-laser-based TSD systems all support moving-load pavement response measurement, but their outputs should not be treated as directly interchangeable. RWD and RAPTOR infer deflection from laser distance or surface-profile measurements, while modern TSD systems measure vertical pavement velocity and derive deflection slope. These different measurement products require different processing and interpretation assumptions. In this report, TSDD is used as the broader term for traffic-speed pavement deflection measurement technologies, while TSD refers to the specific Doppler-laser Traffic Speed Deflectometer system. This distinction is used consistently because TSD is one specific device within the broader TSDD technology class.

2.2 TSDD Technology, Data Processing, and Structural Indicators

Following the historical evolution described in Section 2.1, this section focuses on the current implementation status, measurement principles, data processing requirements, and structural indicators of TSDD technology. The purpose is to clarify how traffic-speed deflection measurements are converted into structural information that can support network-level pavement evaluation.

2.2.1 Global Adoption and Implementation Status

As of 2025, approximately 22 TSD units are in operation worldwide across North America, Europe, Asia, and Australia (Jiang et al., 2025). In the United States, the FHWA Transportation Pooled Fund project TPF-5(385), led by the Virginia Department of Transportation (VDOT), has been a major driver of TSDD evaluation and implementation (Katicha et al., 2020a). The pooled-fund program has supported multi-agency demonstrations, data collection guidance, interpretation methods, and PMS integration studies for traffic-speed structural data. These efforts indicate that the main implementation question is no longer whether traffic-speed structural data can be collected, but how the data should be processed, calibrated, interpreted, and incorporated into agency decision workflows.

Several state DOTs have evaluated or adopted TSDD data for network-level and project-level pavement applications. Texas DOT began routine TSDD data collection in 2019 for both network-level and project-level uses. VDOT has conducted multi-year comparative studies and developed procedures for incorporating TSDD-derived structural condition into PMS decision-making. Idaho DOT has integrated TSDD outputs with ArcGIS for spatial visualization, while Mississippi DOT has used network-level TSDD data to support treatment project prioritization. Minnesota and Louisiana DOTs have evaluated TSDD or related traffic-speed deflection technologies for potential PMS integration. Georgia and Tennessee DOTs have also conducted pilot TSDD data collection programs, making their experiences directly relevant to the present study.

Internationally, TSDD technologies have also been adopted or evaluated in several countries. Denmark played a central role in early TSD development and commercialization. Germany has used integrated pavement evaluation systems that combine structural, surface, and subsurface sensing technologies. China has applied TSDD technologies to large-scale roadway assessments, while Australia has evaluated traffic-speed pavement structural condition through integrated platforms such as iPAVe. TSDD technologies have also been used or evaluated in the United Kingdom, Italy, Poland, Norway, Brazil, and South Africa.

Overall, the global and domestic adoption of TSDD shows that traffic-speed structural evaluation is becoming a practical component of modern pavement asset management. However, adoption alone does not make TSDD outputs directly transferable across agencies. TSDD-derived indicators must still be interpreted in relation to local pavement structures, layer thicknesses, materials, climate, loading conditions, and agency treatment practices. Therefore, the growing implementation of TSDD supports the need for Georgia-specific calibration and PMS-oriented integration, which are the main focus of this study.

2.2.2 Measurement Principles and Indicator Derivation

Modern TSD systems measure pavement structural response under a moving wheel load using Doppler laser technology. The system is mounted on a heavy vehicle that applies an axle load of approximately 100 kN through the rear dual tires. As the vehicle travels at traffic speed, multiple Doppler laser sensors mounted on a stabilized beam measure the vertical velocity of the pavement surface at different offsets from the loaded wheel path. The measured vertical deflection velocity, V_v, is converted to deflection slope, DS, using the relationship $DS = V_v/V_h$, where V_h is the instantaneous horizontal vehicle speed measured by a calibrated odometer (Jiang et al., 2025).

Unlike FWD testing, which applies a stationary impulse load and directly records a discrete deflection basin, TSDD captures pavement response continuously under a rolling load. Therefore, the primary TSD measurement is not absolute deflection but deflection velocity and the corresponding slope of the pavement response. This distinction is central to TSDD interpretation. Slope-based measurements can be used directly as structural response indicators, or they can be converted into estimated deflection-basin parameters for comparison with conventional FWD-based pavement evaluation frameworks.

When deflection-basin information is required, TSDD slope measurements can be converted into estimated deflection values through numerical integration or curve-fitting procedures. Common approaches include the Area Under the Curve (AUTC) method, Euler-Bernoulli beam-based formulations, Weibull curve fitting, the Greenwood proprietary method, and the AUTC-Composite Polynomial Model developed for stiffer pavement structures (Chen et al., 2025; Katicha et al., 2022). These methods allow TSDD measurements to be expressed in formats compatible with conventional structural indicators, including load-center deflection, deflection-basin shape, and curvature-based indices.

However, slope-to-deflection conversion is not a purely mechanical transformation. It depends on the assumed basin shape, integration procedure, loading representation, pavement response behavior, and correction procedures. For this reason, the reliability of converted indicators depends on appropriate calibration and validation, particularly when pavement structure, survey speed, and temperature conditions vary.

2.2.3 Data Quality and Processing Requirements

TSDD measurements are influenced by device-related, environmental, operational, and structural factors. These factors affect the reliability of deflection velocity, deflection slope, and derived basin indicators, and therefore require careful control during data collection, calibration, preprocessing, and interpretation.

Device-related factors include sensor accuracy, sensor spacing, sampling frequency, spatial resolution, laser focus, vehicle load, tire pressure, beam stability, and motion compensation. Calibration is a critical prerequisite for reliable TSDD data, but current practice still relies largely on vendor-specific procedures. These procedures typically address sensor geometry and angle, optical stability, vehicle loading, and beam motion correction. Certification and verification, referring respectively to system compliance checks and periodic performance checks under controlled conditions, remain important but are not yet supported by widely accepted standards (Beizaei et al., 2024).

Environmental conditions also affect TSDD response. Temperature is particularly important because asphalt viscoelasticity causes higher temperatures to reduce surface stiffness and increase measured deflections, especially for indicators such as D0 and SCI (Zhang et al., 2023). Temperature correction factors are therefore commonly used to standardize measurements to a reference temperature. In addition to pavement temperature, internal temperature gradients within the TSDD measuring beam can cause torsional distortion and sensor angle drift. Even small thermal gradients may introduce systematic measurement bias if not properly controlled. Surface moisture, pavement roughness, and crosswind conditions may further affect measurement stability.

Operational factors are mainly related to survey speed, traffic conditions, and vehicle loading configuration. TSDD systems generally perform most reliably at speeds between 50 and 80 km/h, approximately 31 to 50 mph. Data collected below about 25 km/h, approximately 16 mph, often occur near intersections, work zones, or congested urban areas and may exhibit unstable deflection characteristics. These records are commonly filtered out during preprocessing, which can create spatial gaps in the final dataset. In addition to speed effects, the multi-axle loading geometry of the TSDD vehicle introduces processing challenges. Although many interpretation algorithms simplify the load as a single rear-axle point load, the actual vehicle applies distributed loading through the front, middle, and rear axles. This loading configuration can create an asymmetric deflection response over a relatively long longitudinal influence zone, reported in some studies to exceed 18 m, approximately 59 ft. As a result, advanced correction methods may be required to separate the pavement response associated with the primary loading axle from the broader vehicle-induced deflection field (Chen et al., 2025).

Structural factors determine whether TSDD measurements can be interpreted reliably. Standard TSDD interpretation is most appropriate for flexible pavements, where the pavement response can be reasonably represented by continuous deflection behavior under a moving load. Inverted or semi-rigid structures may produce wide and shallow deflection basins that violate standard integration assumptions and can lead to biased structural capacity estimates if uncorrected (Jiang et al., 2025). Rigid pavements often generate very small deflections that approach sensor noise levels, reducing the reliability of traditional indicators such as D0 and SCI. Composite pavements can introduce stiffness discontinuities that produce abrupt slope changes, complicating basin reconstruction and indicator interpretation (Scavone et al., 2023). These structural effects cannot be fully removed through preprocessing and therefore require pavement-type-aware interpretation.

Data processing is needed to convert raw deflection velocity or slope measurements into reliable structural indicators. The first stage is typically signal preprocessing and denoising. Raw TSDD measurements may contain noise caused by environmental conditions, vehicle dynamics, surface roughness, sensor vibration, and local pavement irregularities. Classical methods such as wavelet transforms, smoothing splines, and moving-window filters have been used to remove high-frequency noise while preserving meaningful structural response patterns. More advanced approaches, including particle swarm optimization-variational mode decomposition (PSO-VMD) and Reweighted L1 Minimization, have been proposed to improve weak-spot detection and enhance signal stability at network scale (Scavone et al., 2023; Wu et al., 2024).

After denoising, slope-based measurements may be converted into deflection basins using numerical integration or curve-fitting procedures. From these reconstructed basins, standard structural indicators such as D0, SCI, DSI, BCI, and BDI can be computed. This step is important because many pavement evaluation frameworks and PMS decision rules are still organized around deflection-basin parameters rather than raw velocity or slope measurements. However, the accuracy of these derived indicators depends on assumptions used in basin reconstruction, pavement type, loading representation, and correction procedures.

Beyond indicator computation, TSDD data are increasingly used for pavement layer modulus estimation and structural characterization. Traditional backcalculation tools, including elastic programs such as EVERCALC and viscoelastic approaches such as ViscoRoute, can estimate layer moduli from deflection responses while accounting for pavement material behavior under dynamic loading. However, these methods can be computationally intensive and may suffer from non-unique solutions, particularly when layer thickness, temperature, or material properties are uncertain (Hamidi et al., 2024). To address these limitations, artificial neural networks, support vector machines, and other ML methods have been explored to estimate layer moduli directly from TSDD slope or deflection data (Abdelmuhsen et al., 2023). Physics-Informed Neural Networks represent a more recent direction by embedding physical constraints into the learning process, which may improve prediction stability under viscoelastic and structurally complex conditions (Kargah-Ostadi et al., 2024).

Temperature correction is another important component of TSDD data processing. Traditional correction methods often focus on adjusting center deflection, whereas recent approaches attempt to correct the full deflection basin or slope-based indicators. Neural network-based correction models trained with finite element simulation data have been used to estimate temperature correction factors across multiple basin positions (Katicha et al., 2022). Other correction frameworks incorporate time-temperature superposition principles to jointly account for vehicle speed and pavement temperature, improving the transferability of TSDD indicators across different environmental and pavement conditions (Zhang et al., 2023).

Although automated processing and AI-assisted correction methods have improved the efficiency of TSDD data analysis, engineering review remains necessary before processed indicators are used for structural evaluation. Quality control should verify whether denoising, basin reconstruction, temperature correction, and filtering procedures have produced physically reasonable response patterns. Records affected by missing data, low-speed operation, sensor anomalies, or pavement-type incompatibility should be flagged or excluded as appropriate. Therefore, a robust TSDD processing workflow should combine automated signal processing with engineering validation to ensure that the final indicators are suitable for structural condition classification and pavement management applications.

2.2.4 Key Structural Indicators

After TSDD measurements are processed, the resulting structural indicators must be interpreted according to the portion of the pavement response they represent. These indicators support network- and project-level pavement assessment by translating measured deflection slopes or reconstructed deflection basins into engineering parameters related to pavement stiffness and load-carrying capacity.

The Structural Curvature Index (SCI) is one of the most widely used TSDD-derived indicators. It is computed from the difference between deflection values at selected offsets from the load and is commonly used to characterize the near-surface structural response of asphalt layers. Depending on the sensor layout and agency convention, SCI may be reported at different offsets, such as SCI_200, SCI_300, SCI_8, or SCI_12. In general, higher SCI values indicate greater near-load curvature and weaker upper-layer structural response, while lower values suggest a stiffer pavement response.

Deflection at the load center, D0, represents the estimated pavement surface deflection at the nominal loading position. It is commonly used as an indicator of overall pavement flexibility and load-spreading capacity. Under moving-load TSDD conditions, however, D0 should not be interpreted as the absolute maximum deflection. The actual peak deflection may occur behind the load center, particularly for stiff or viscoelastic pavement structures. Therefore, D0 is more appropriately interpreted as a standardized deflection-basin parameter rather than the true peak deflection.

Other basin-derived indicators provide complementary information on pavement structural condition. The Deflection Slope Index (DSI) describes the spatial rate of change in deflection across measurement offsets and is sensitive to localized stiffness transitions. The Base Curvature Index (BCI) and Base Damage Index (BDI) are commonly used to characterize base and subbase support conditions. In some pavement evaluation frameworks, TSDD-derived deflection indicators may also be used to estimate Structural Number (SN), Effective Structural Number (SNeff), or Remaining Structural Life (RSL) through calibrated empirical or mechanistic models (Canestrari et al., 2023; Zhang et al., 2022).

Together, these indicators provide the structural response variables used in later stages of this study. However, their interpretation depends on data quality, processing procedures, pavement type, and pavement layer configuration. For this reason, later chapters combine TSDD-derived indicators with GPR-derived thickness information and HWTT reference labels to develop Georgia-specific structural condition classification methods.

2.3 Complementary Sensing Technologies and Multi-Sensor Integration

2.3.1 Ground Penetrating Radar

GPR is a high-speed, non-destructive sensing technology used to characterize pavement subsurface conditions. It operates by transmitting electromagnetic pulses into the pavement structure and recording the reflected signals generated at material interfaces with contrasting dielectric properties. These interfaces may include boundaries between asphalt concrete (AC), Portland cement concrete (PCC), base, and subgrade layers, as well as localized anomalies such as voids, moisture intrusion, stripping, debonding, and other internal defects. Previous review studies have shown that GPR is widely used for pavement diagnosis because it can detect subsurface features efficiently and non-invasively while reducing the need for destructive testing (Rasol et al., 2022).

Processed GPR data can provide continuous estimates of pavement layer thickness and pavement type along a roadway network. Compared with coring, GPR provides denser spatial coverage and can be collected at traffic speed, making it particularly useful for network-level pavement evaluation. In addition to layer thickness estimation, recent studies have explored ML and deep learning methods for automated GPR interpretation. For example, (Liu et al., 2024) proposed a Feature-enhanced Multiscale Vision Transformer (FeMViT) for road distress classification from GPR images, illustrating the growing role of AI-based interpretation in subsurface pavement assessment.

For TSDD-based structural evaluation, GPR provides essential structural context. TSDD deflection and curvature indicators are affected not only by material weakness but also by pavement layer geometry. A thick asphalt section and a thin asphalt section may produce different deflection responses even when their material condition is similar. Therefore, GPR-derived layer thickness helps interpret TSDD response and can support thickness-conditioned threshold development. GPR also helps distinguish flexible pavement sections from rigid or composite structures, which is important because TSDD interpretation is most reliable for flexible pavements.

2.3.2 3D Pavement Surface Sensing

3D pavement surface sensing provides complementary information on visible functional condition. Vehicle-mounted 3D laser scanning systems collect high-resolution pavement surface profiles at highway speed and generate detailed representations of pavement texture and surface geometry. These data are commonly used to quantify surface distresses and functional indicators, including cracking, rutting, longitudinal roughness, and macrotexture (Kumar, 2024; Plati et al., 2024).

Unlike TSDD and GPR, which provide structural response and subsurface information, 3D surface data describe the condition expressed at the pavement surface. This distinction is important because surface deterioration and structural deterioration do not always occur at the same time. A pavement section may show acceptable surface condition while having weak underlying structural support, or it may exhibit poor surface condition while retaining adequate structural capacity. Therefore, 3D surface data are necessary for distinguishing functional distress from structural weakness and for supporting treatment decisions that consider both dimensions.

In current pavement survey practice, integrated platforms can collect multiple sensing streams within a consistent spatial referencing framework. For example, ARRB's iPAVe platform integrates 3D laser sensors and TSDD technology to assess pavement surface condition and structural capacity at traffic speeds. Such platforms illustrate the practical feasibility of combining surface and subsurface measurements in one network-level survey system.

2.3.3 Benefits of Multi-Sensor Data Integration

The integration of TSDD, GPR, and 3D surface data enables a more complete characterization of pavement condition than any single sensing technology can provide. TSDD measures pavement structural response under a moving wheel load, GPR describes pavement layer geometry and subsurface features, and 3D surface sensing quantifies visible functional distress. Together, these data sources allow agencies to distinguish whether observed pavement problems are primarily structural, functional, or both.

This distinction is central to pavement management. Elevated TSDD curvature may indicate weak asphalt-layer response, but without GPR information it can be difficult to determine whether the response is associated with material deterioration or simply with a thinner pavement structure. Similarly, severe cracking, rutting, or roughness does not necessarily imply structural failure; it may indicate a functional surface problem that can be treated without major structural rehabilitation if the underlying pavement structure remains adequate. Multi-sensor integration reduces this ambiguity by linking structural response, layer thickness, pavement type, and surface condition within the same spatial framework.

In the present study, multi-sensor integration provides the data foundation for later modeling and decision analysis. TSDD-derived indicators characterize structural response, GPR-derived layer thickness provides structural context, and 3D surface data represent functional condition. This combination supports Georgia-specific structural condition classification and provides the input structure for the integrated treatment decision framework developed in later chapters.

2.4 Structural Condition Classification

2.4.1 Threshold-Based Structural Condition Classification

Threshold-based classification is the most direct and operationally transparent approach for converting TSDD-derived structural indicators into pavement condition categories. In this approach, one or more indicators are compared with predefined threshold values to classify pavement sections as structurally sound, marginal, or deficient. Because threshold rules are deterministic, easy to interpret, and compatible with existing agency rating systems, they are especially attractive for early-stage implementation of traffic-speed structural evaluation.

Several state-level studies have developed threshold systems using TSDD-derived structural indicators, and their reported values demonstrate that threshold selection is highly context dependent. VDOT developed a threshold-based classification system using SCI_300 and DSI, with thresholds stratified by road category and AC layer thickness. For interstate pavements with AC thickness greater than 9 inches, sections are classified as Poor when SCI_300 exceeds 3.7 mils and DSI exceeds 3.0 mils, and as Fair when SCI_300 ranges from 2.7 to 3.7 mils and DSI ranges from 2.2 to 3.0 mils; different threshold ranges are used for primary and secondary roads (Katicha et al., 2022). South Carolina DOT research used SCI_12, defined as D0 – D300, as the primary classification indicator, with Poor defined as SCI_12 greater than 3.3 mils, Fair as 1.6 to 3.3 mils, and Good as less than 1.6 mils (Ahmed et al., 2022). Research using U.S. National Park Service pavement data applied SCI_300 greater than 6 mils as a threshold for structural weakness (Bryce, 2024). Tennessee DOT has also developed a threshold-based structural evaluation approach using field experience and backcalculated moduli from TSDD data collected across its highway network (Huang et al., 2022b).

A consistent finding across threshold-based studies is that structural thresholds are not universally transferable. Appropriate threshold values depend on pavement type, asphalt thickness, base support, subgrade condition, climate, traffic loading, and agency-specific design practices. A threshold calibrated for one state or pavement family may not have the same meaning when applied to another network. This issue is especially important for TSDD interpretation because the same measured curvature value may indicate different levels of structural risk depending on layer thickness and pavement configuration.

This thickness dependency directly motivates the threshold-development strategy used in the present study. Rather than adopting published thresholds from other states, this study develops Georgia-specific threshold rules using the available TSDD, GPR, and laboratory performance data. In particular, GPR-derived AC thickness is used to examine whether thickness-conditioned thresholds can improve structural classification relative to a single global threshold. This approach preserves the interpretability of threshold-based screening while acknowledging that TSDD response must be interpreted in relation to pavement structure.

Another limitation of many existing threshold systems is that they are commonly based on expert judgment, historical FWD comparisons, backcalculated parameters, or field performance observations. These references are useful, but they may not provide a direct laboratory-based performance benchmark for asphalt-layer structural condition. In contrast, the present study uses HWTT outcomes as a ground reference for threshold calibration. This provides a more controlled basis for relating TSDD-derived structural indicators to measured asphalt-layer performance.

2.4.2 Machine Learning-Based Structural Condition Classification

ML-based classification provides an alternative to deterministic threshold rules by allowing multiple variables to be evaluated simultaneously. Instead of assigning pavement condition based on a single indicator or a small set of predefined cutoffs, ML models can learn relationships among TSDD-derived deflection indicators and GPR-derived layer thickness variables. This capability is useful when the boundary between structurally adequate and structurally deficient sections is nonlinear or when no single indicator provides a sufficiently clear separation between condition classes.

Recent studies have explored several ML methods for TSDD-related pavement structural evaluation, including both direct condition classification and related structural characterization tasks. (Ahmed et al., 2022) evaluated multiple supervised classifiers, including k-nearest neighbors, decision trees, RF, and gradient boosting, to predict pavement structural condition using TSDD-derived SCI indicators in the South Carolina DOT context. Their results showed that RF and gradient boosting models outperformed simpler classifiers and single-indicator threshold rules under cross-validated evaluation. (Mabrouk et al., 2022) applied artificial neural networks to backcalculate pavement layer moduli from TSDD deflection data, demonstrating the potential of data-driven models to estimate structural properties from traffic-speed measurements. (Abdelmuhsen et al., 2023) used support vector machine variants to estimate soil elastic modulus from TSDD slope data generated from numerical simulations, showing that ML models can generalize across pavement configurations when the training dataset covers a sufficiently diverse structural space.

However, ML-based classification also has practical limitations. Model performance depends strongly on the size, quality, and representativeness of the labeled dataset. This issue is especially important in pavement structural evaluation because ground-reference data, such as coring records and laboratory performance tests, are usually sparse, expensive, and unevenly distributed across pavement types. When the labeled dataset is limited, a well-calibrated threshold rule may perform similarly to a more complex ML model. In addition, ML models require careful feature selection, cross-validation, and leakage control to avoid overestimating generalization performance.

Therefore, threshold-based and ML-based approaches should be viewed as complementary rather than mutually exclusive. Threshold methods are attractive for immediate implementation because they are transparent, stable, and easy to communicate to agency engineers. ML methods are attractive because they can incorporate multiple predictors and may become more powerful as larger and more diverse TSDD, GPR, and laboratory datasets become available.

2.5 Integration of Surface and Structural Condition into PMS Decision Frameworks

Once pavement structural condition has been classified, the next step is to translate the classification output into M&R decision support. This step is different from structural sensing or structural classification. TSDD, GPR, and 3D surface data provide measurements; threshold-based or ML-based methods convert those measurements into structural condition categories; PMS integration determines how those categories modify treatment selection. Therefore, the purpose of an integrated decision framework is to convert structural screening results into actionable treatment recommendations that are compatible with agency pavement management workflows.

Existing studies have increasingly emphasized this implementation step. The Transportation Pooled Fund study TPF-5(385) was established to help agencies collect, interpret, and apply TSDD data, including integration into PMS decision-making processes (Katicha et al., 2022). A related implementation effort reported that 26 agency partners and FHWA participated in TSDD research and that structural condition data had been collected on more than 50,000 miles of U.S. roadways, indicating that the remaining challenge is no longer only data collection but also operational use in agency decision systems.

A practical approach is to use a treatment decision matrix in which structural condition adjusts or refines a surface-based treatment recommendation. VDOT provides one of the clearest examples. Earlier VDOT PMS procedures used decision matrices based mainly on surface condition indices; subsequent work incorporated structural condition when adequate structural information was available (Katicha et al., 2020a). Later TSDD-based studies evaluated augmented structural condition matrices to examine how TSDD-derived structural condition would affect recommended rehabilitation treatments and average maintenance cost per mile on interstate roads. This matrix-based approach is attractive because it does not require replacing the PMS. Instead, it adds a structural screening layer to determine whether a surface-based recommendation should remain unchanged, be downgraded to a lighter treatment, be upgraded to a heavier rehabilitation action, or be flagged for further investigation.

The economic value of this integration has also been evaluated. (Murekye et al., 2025) assessed approximately 2,000 lane-miles of Virginia primary roads and compared treatment planning based on functional condition alone with planning that incorporated structural condition. The study reported life-cycle cost reductions of up to 11.2% over a 30-year analysis period when structural condition was included in PMS decision-making. This finding supports the practical value of structural data not simply as an additional condition metric, but as a decision variable that can alter treatment timing, treatment type, and long-term cost outcomes.

The present study follows this implementation logic but adapts it to the Georgia interstate context. Structural condition is first classified using TSDD-derived indicators, with GPR-derived thickness information providing structural context. Surface condition is then paired with structural condition to produce treatment decision categories in Chapter 6. In this framework, the main purpose of PMS integration is not to repeat sensor-level diagnosis, but to translate classified surface and structural conditions into actionable decisions.

2.6 Research Gaps and Motivation for This Study

The preceding review shows that TSDD technology has advanced from experimental traffic-speed deflection measurement to a practical tool for network-level pavement structural evaluation. However, several gaps remain before TSDD-derived structural indicators can be reliably converted into pavement management decisions within the Georgia DOT context.

The first gap concerns the lack of Georgia-specific structural condition calibration. Existing threshold systems developed by agencies such as Virginia DOT, South Carolina DOT, Tennessee DOT, and the U.S. National Park Service demonstrate that TSDD-derived indicators can be used to classify pavement structural condition. However, their reported thresholds vary substantially by agency, pavement type, roadway category, asphalt layer thickness, and calibration basis. Directly adopting these thresholds for Georgia interstate pavements may be inappropriate because Georgia corridors may differ in asphalt layer thickness distribution, base structure, material properties, traffic loading, subgrade support, and environmental exposure. The present study addresses this gap by developing Georgia-specific structural condition classification methods using TSDD, GPR, and laboratory performance data.

The second gap concerns the selection of an appropriate structural reference for model calibration and validation. Prior TSDD threshold and classification studies have commonly relied on expert judgment, FWD comparisons, backcalculated structural parameters, or field performance observations as reference information. These references are useful, but they do not always provide a direct, repeatable laboratory-based measure of asphalt-layer performance. This is important because TSDD curvature indicators are particularly sensitive to near-surface asphalt-layer response. The present study addresses this gap by using HWTT outcomes as the structural performance reference for classifying asphalt-layer condition. Although HWTT does not represent the entire pavement structure, it provides a controlled and repeatable benchmark for rutting susceptibility and moisture-related asphalt-layer damage, making it suitable for calibrating the near-surface structural response emphasized in this study.

The third gap concerns the role of pavement layer thickness in interpreting TSDD response. Previous threshold studies show that structural cutoff values are sensitive to pavement configuration, especially AC thickness. However, many operational screening approaches still rely on global single-indicator thresholds that do not explicitly account for layer geometry. Because TSDD deflection and curvature responses are influenced by both material condition and pavement structure, the same SCI or deflection value may imply different levels of structural risk across thin and thick asphalt sections. This study addresses this gap by incorporating GPR-derived AC thickness into threshold development and evaluating whether thickness-conditioned rules improve classification performance compared with a single global threshold.

The fourth gap concerns the comparative evaluation of threshold-based and ML-based structural classification. Threshold rules are transparent, simple, and attractive for immediate agency implementation, while ML models can incorporate multiple TSDD and GPR variables and capture nonlinear relationships. However, the relative advantage of ML over well-calibrated threshold rules remains uncertain when labeled structural performance data are limited. Many ML studies in pavement structural evaluation focus on model development or related backcalculation tasks, but fewer studies compare threshold and ML methods under the same reference labels, feature space, and validation protocol. The present study addresses this gap by evaluating both method families under a consistent repeated stratified cross-validation framework, with careful feature selection and leakage control.

The fifth gap concerns the translation of structural classification outputs into PMS-oriented treatment decisions. Prior studies have demonstrated that TSDD-derived structural indicators can support network-level screening and, in some cases, project-level rehabilitation design. However, structural classification by itself does not directly specify M&R actions. For agency implementation, structural condition must be combined with surface condition to determine whether a pavement section requires no action, functional resurfacing, close monitoring, or major rehabilitation. Existing integrated decision frameworks remain largely agency-specific and have not been adapted to Georgia interstate pavements. The present study addresses this gap by linking structural classification outputs with 3D surface condition data to develop an integrated surface-structural treatment decision framework for Georgia interstate corridors.

Together, these gaps motivate the present study. The research develops Georgia-specific threshold-based and ML-based structural condition classification methods, uses HWTT outcomes as a laboratory performance reference, evaluates the added value of GPR-derived asphalt thickness, compares modeling approaches under a consistent validation framework, and translates structural screening outputs into an integrated treatment decision framework for network-level pavement management.

2.7 Chapter Summary

This chapter establishes the literature basis for using TSDD as the structural sensing component of an integrated pavement management framework. The reviewed studies show that TSDD can support network-level structural screening, but TSDD measurements must be interpreted with pavement layer information, surface condition data, and locally calibrated performance references before they can be translated into treatment decisions. Therefore, the literature supports the central logic of this study: use TSDD to characterize structural response, use GPR and 3D surface data to provide structural and functional context, calibrate classification methods with Georgia-specific HWTT outcomes, and convert the resulting condition labels into PMS-compatible treatment categories.

- **TSDD has become a practical network-level structural evaluation tool, but its indicators require local calibration before they can support agency decisions.** The literature shows that TSDD can collect continuous deflection-related measurements at traffic speed, making it more suitable for network-level screening than traditional stop-and-go structural testing. However, TSDD-derived indicators are affected by sensor configuration, temperature, survey speed, pavement type, layer thickness, and local structural conditions. Therefore, thresholds or models developed in other states should not be directly adopted for Georgia interstate pavements without local validation.

- **GPR and 3D surface data are needed to interpret TSDD response and connect structural evaluation to pavement management decisions.** TSDD indicators such as D0, SCI, DSI, and BCI describe pavement response under loading, but they do not fully explain whether a measured response is caused by material weakness, layer thickness, pavement type, or surface deterioration. GPR provides layer thickness and pavement-type information needed to interpret TSDD measurements, while 3D surface data describe visible functional condition. Together, these sensing sources support a more complete surface–structural evaluation framework.
- **Threshold-based and ML-based methods both have practical value for structural condition classification.** Threshold-based methods are transparent, easy to communicate, and suitable for near-term agency implementation when a small number of physically meaningful indicators dominate the structural signal. ML-based methods can incorporate multiple predictors and nonlinear relationships, making them useful as labeled datasets become larger and more diverse. For this reason, this study evaluates both approaches under the same validation framework rather than treating them as competing alternatives.
- **The literature review identifies five key gaps that motivate the analytical chapters of this study.** First, existing TSDD thresholds have not been calibrated for Georgia interstate pavements. Second, the selection of an appropriate structural reference remains a critical issue for model calibration. Third, the role of GPR-derived thickness in interpreting TSDD response requires further evaluation. Fourth, threshold-based and ML-based structural classification methods need to be compared under the same reference labels, feature space, and validation protocol. Fifth, structural classification outputs must be translated into PMS-compatible

treatment categories that combine structural and surface condition. These gaps motivate the following chapters, which construct a Georgia-specific multi-sensor dataset, calibrate structural condition classification methods using HWTT outcomes, compare threshold-based and ML-based approaches, and connect structural and functional condition results to an integrated treatment decision framework.

CHAPTER 3. DATA COLLECTION AND PREPARATION

This chapter documents the data collection, selection, processing, and spatial integration procedures used to construct the dataset for this study. Section 3.1 provides an overview of the data sources and candidate corridors. Section 3.2 characterizes each data source in detail. Section 3.3 presents the missing data analysis and study corridor selection. Section 3.4 documents the spatial integration methodology and coring-centered data refinement. Section 3.5 describes the composition of the final analytical dataset. Section 3.6 provides descriptive statistics and exploratory analysis across the retained corridors. Section 3.7 summarizes the chapter.

3.1 Overview of Data Sources

This study uses multi-source pavement data provided by GDOT to construct an integrated dataset combining TSDD measurements, 3D pavement surface data, GPR layer information, and PES records, including coring information and HWTT outcomes. The data originate from two distinct collection programs and are organized into two categories: Type I network-level sensing data and Type II project-level evaluation data.

Type I data consist of TSDD measurements and 3D pavement surface data collected simultaneously by ARRB Systems, formerly known as the Australian Road Research Board, during network-level pavement surveys conducted in 2023, 2024, and 2025.These data provide continuous corridor-level measurements at 0.01-mile longitudinal intervals. The TSDD captures pavement deflection response under traffic-speed loading, while the co-deployed 3D laser system records functional surface condition metrics such as cracking, rutting, roughness, and macrotexture.

Because the two sensing streams are collected in the same survey pass, they share consistent spatial referencing and survey timing.

Type II data include GPR layer-thickness data and PES records provided by the GDOT Office of Materials and Testing (OMAT). GPR data provide estimates of pavement layer thickness, including AC, PCC, and base layers where applicable. The PES records provide project-level pavement evaluation information, including coring locations, coring logs, and HWTT results. In this study, the HWTT results are used as laboratory-based performance reference data. According to AASHTO T 324-23 (Highway & Officials, 2017), HWTT evaluates the rutting and moisture susceptibility of compacted asphalt mixture specimens under repeated wheel loading in a water bath. Therefore, HWTT provides a controlled laboratory reference for assessing asphalt-layer performance relative to field-measured TSDD and GPR indicators.

The difference in spatial collection structure between Type I and Type II data is critical because it explains why spatial integration, rather than direct table merging, is required to construct the final dataset. Type I data are reported as continuous corridor-based records at fixed longitudinal intervals, whereas the key Type II performance reference is tied to discrete coring locations. Therefore, the dataset must be constructed through a coring-centered integration process, in which nearby TSDD, 3D surface, and GPR measurements are assigned to each HWTT location only when they represent the same physical pavement section.

Although 3D surface data are collected together with TSDD measurements, they are not used as the primary basis for structural performance prediction. Instead, they are retained to characterize functional pavement condition and to support the treatment decision analysis developed in later chapters.

Four GDOT corridors were initially considered in this study: I-24, I-59, I-285, and I-575. All four corridors had the required data sources available, including Type I sensing data and Type II project-level evaluation data. However, the existence of these data sources did not necessarily mean that a corridor could be used for the structural analysis. This study requires each HWTT outcome to be paired with nearby TSDD and GPR measurements that represent the same flexible pavement section. This requirement is important because TSDD deflection indicators are meaningful only when they reflect the same pavement structure tested by the core sample, and the TSDD measurement principle is not reliable over rigid or composite pavement sections. Therefore, each corridor was further evaluated in terms of pavement type, missing data patterns, and whether coring locations could be spatially matched to the mainline sensing records. The detailed screening process and final corridor selection decisions are presented in Section 3.3.

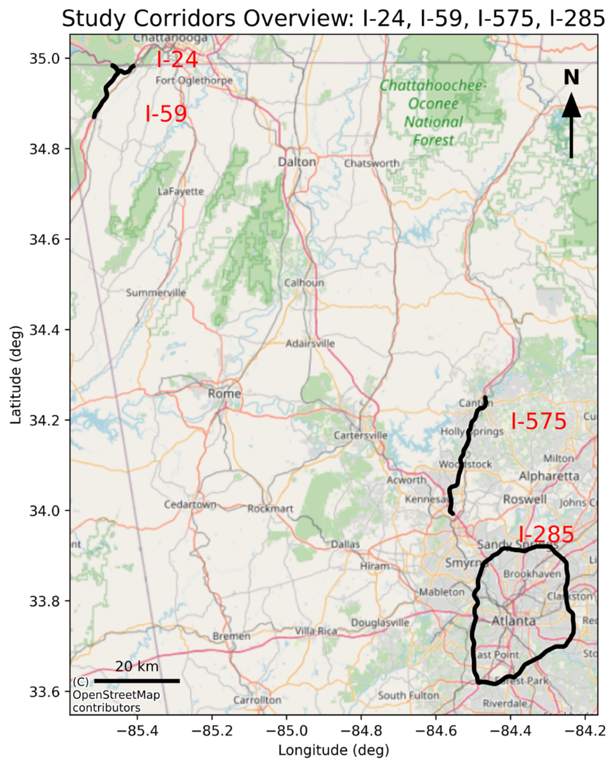

Figure 1. Geographic overview of the four candidate study corridors in Georgia.

Figure 1 presents the geographic distribution of the four candidate corridors across Georgia. I-24 and I-59 traverse rural and semi-rural areas in northwest Georgia, I-575 represents a suburban corridor north of Atlanta, and I-285 represents the Atlanta metropolitan beltway with heavy traffic demand and complex interchange geometry. Detailed descriptions of each data source are provided in Section 3.2, and the corridor screening and final inclusion decisions are presented in Section 3.3.

3.2 Data Description

3.2.1 TSDD Data

The TSDD dataset provides continuous measurements of pavement structural response along the candidate corridors. Each record corresponds to a 0.01-mile longitudinal segment, enabling quasi-continuous characterization of pavement behavior under traffic-speed loading. Data were collected by ARRB Systems in 2023, 2024, and 2025.

Each record is associated with location-referencing variables including a unique road section identifier (ROAD_SECTION_ID), route and direction labels, chainage information, and Global Positioning System (GPS) coordinates. The core structural information is the pavement deflection response induced by a moving rear axle load of nominally 100 kN. Deflections are reported in mils at the load center (D0) and at increasing offsets both upstream and downstream of the load, from D8 through D72, forming a deflection bowl analogous to that obtained from FWD testing.

Three Structural Curvature Indices (SCI) are derived from the deflection bowl: SCI_8 = D8 − D0 and SCI_12 = D12 − D0 capture near-surface curvature associated primarily with asphalt-layer stiffness; SCI_SUBGRADE = D60 − D36 characterizes deeper structural contribution and subgrade support. Slope-based indicators are provided at multiple offsets relative to the load. Raw deflection bowl measurements and raw slope values are retained for quality control. Survey-context variables including surface and air temperature, axle loads, and survey speed are included because all influence measured deflection response. Table 1 summarizes representative TSDD variables with example values from a record on I-59 Lane 1. A complete data dictionary is provided in Appendix A.

Table 1. Representative TSDD variables with example values (I-59 Lane 1).

Category	Example Fields	Units	Example Value*
Location & GPS	ROAD_SECTION_ID, ROUTE, DIRECTION, FROM_LATITUDE, FROM_LONGITUDE	—/deg	10000001, I-59_L1, 34.8696
Chainage	INSIGHT_CHAINAGE_FROM/TO, SECTION_SUB_CHAINAGE_FROM/TO	mi	0.000 to 0.010
Structural indices	SCI_8, SCI_12, SCI_SUBGRADE	mils	SCI_12 = 1.5; SCI_SUBGRADE = −0.3
Deflection bowl	D0, D8, D12, D24, D36, D60, D72	mils	D0 = −3.5; D12 = −1.9; D60 = −0.3
Slope indices	SLOPE-200, SLOPE-300, SLOPE-450, SLOPE130, SLOPE215	µm/m	SLOPE-450 = 68.5; SLOPE-300 = 84.5
Raw bowl & slopes	TD0, TD200, TD300; RawSlope-200, RawSlope-300	µm; µm/m	TD0 = 82.84; RawSlope-300 = −227.84
Temperature & load	SURFACE_TEMPERATURE, AIR_TEMPERATURE, STRAIN_GAUGE_LEFT	°F; lb	68.5°F; 84.5°F; 10,202 lb
Metadata	SURVEY_SPEED, SURVEY_DATE, SURVEY_TIME, SLOPE_QUALITY	mph/date	64.4 mph; 5/29/2023; Low

3.2.2 3D Pavement Surface Data

The 3D surface dataset is acquired simultaneously with the TSDD survey at 0.01-mile intervals using the co-deployed laser scanning system. Variables fall into five functional categories: surface distress, permanent deformation, longitudinal roughness, surface texture, and roadway geometry.

Surface distress indicators quantify the extent, severity, and spatial distribution of pavement cracking, including total cracking percentage, total crack length, high-severity cracking, and type-specific measures for alligator, longitudinal, and transverse cracking. Permanent deformation is captured by rut depth reported separately for the left and right wheel paths and in aggregated forms. Longitudinal roughness is quantified using the International Roughness Index (IRI). Surface texture variables include mean profile depth (MPD) and estimated texture depth (ETD). Additional variables include pothole counts and geometry, joint-related measurements, and roadway geometric attributes. Table 2 presents a representative summary of 3D surface variables by category. A complete data dictionary is provided in Appendix B.

In this study, 3D surface data are used primarily to characterize functional condition for the treatment decision framework in Chapter 6; the rationale for their exclusion from structural condition prediction is discussed in Section 4.4.

Table 2. Representative 3D pavement surface variables by category (I-59 Lane 1).

Category	Representative Variables	Unit	Example Value
Surface cracking	Total cracking, crack length, high-severity cracking, alligator cracking, longitudinal cracking, transverse cracking	%, ft	Total cracking = 18.3%; alligator cracking = 6.1%
Rutting	Left wheel-path rut depth, right wheel-path rut depth, average rut depth	in.	Average rut depth = 0.33 in.
Roughness	Left IRI, right IRI, average IRI	in./mi	Average IRI = 116 in./mi
Surface texture	MPD, ETD	in.	MPD = 0.12 in.
Joint and faulting indicators	Joint count, average faulting	count, in.	Average faulting = 0.04 in.

3.2.3 GPR Data

GPR data were provided as processed, engineering-interpretable outputs in which layer boundaries have been identified and converted into thickness estimates. Each record includes estimated thicknesses for AC, PCC, and base layers where applicable. A key feature of the GPR dataset is that layer-thickness fields are populated only when the corresponding pavement layer exists and can be reliably interpreted. Therefore, missing values do not always represent data loss. For example, missing PCC thickness may indicate a flexible pavement section with no concrete layer, while available PCC thickness indicates a rigid or composite structure. Some missing values may also result from uncertain radar interpretation where layer boundaries cannot be clearly identified. For multilane corridors such as I-285, GPR thickness estimates are reported separately by lane group because pavement structure can vary across the roadway width. Inner lanes, outer lanes, and auxiliary lanes may have different AC thicknesses or different underlying layers. Therefore, lane information is required to assign the appropriate GPR thickness to each coring location. Table 3 summarizes representative GPR variables with example values. A complete data dictionary is provided in Appendix C.

Table 3. Representative GPR variables with example values on I-575 Northbound (NB).

Category	**Example Fields**	**Units**	**Example Value***
Location & referencing	Route, Direction, Lane, Milepost (MP)/Chainage	—	I-575 NB, Lane 2
Asphalt layer	AC Thickness	in.	15.8
Base layer	Base Thickness	in.	25.5
Concrete layer	PCC Thickness	in.	— (absent; flexible pavement confirmed)
Data completeness	Missing indicator (layer-specific)	—	PCC missing confirms flexible pavement

3.2.4 HWTT Data

The ground reference for this study is derived from a coring-centered dataset integrating field coring records with HWTT outcomes and GPR layer thickness estimates at coring locations. Coring information was compiled from GDOT PES reports for the four candidate corridors. The HWTT, as specified in AASHTO T 324-23 (Highway & Officials, 2017), evaluates the rutting and moisture susceptibility of compacted asphalt mixture specimens under repeated wheel loading in water. For each coring location, HWTT outcomes are recorded separately by layer, including a Pass/Fail designation, number of wheel passes at failure, and maximum deformation in millimeters.

The Layer 1 HWTT outcome serves as the laboratory performance reference throughout this study, as discussed in Chapter 4.

Because coring data were historically collected in route-specific PES reports with varying formats, a dedicated standardization effort was undertaken. A unified data format and data dictionary were defined and applied across all routes to ensure consistent field naming, units, and definitions.

A critical aspect of preparation was the manual validation and correction of GPS coordinates. Milepost-based references in PES reports can introduce positional uncertainties on the order of hundreds of feet, particularly near complex geometry and interchanges. To address this issue, coring locations were validated using multiple independent visual sources, including PES field photographs, Google Street View imagery, and high-resolution 3D pavement surface intensity and range images. Figure 2 illustrates the manual GPS validation procedure for one representative coring location. The same visible cracking features, labeled A, B, and C, can be identified consistently across the PES photograph, Google Street View image, 3D intensity image, and 3D range image. This cross-source agreement confirms that the corrected GPS coordinates correspond to the intended physical coring site. Therefore, the manual validation step provided an important quality-control basis for subsequent spatial joining with TSDD, GPR, and 3D surface condition data. Table 4 summarizes representative variables in the standardized coring dataset. The complete data dictionary for the coring and HWTT Data is provided in Appendix D.

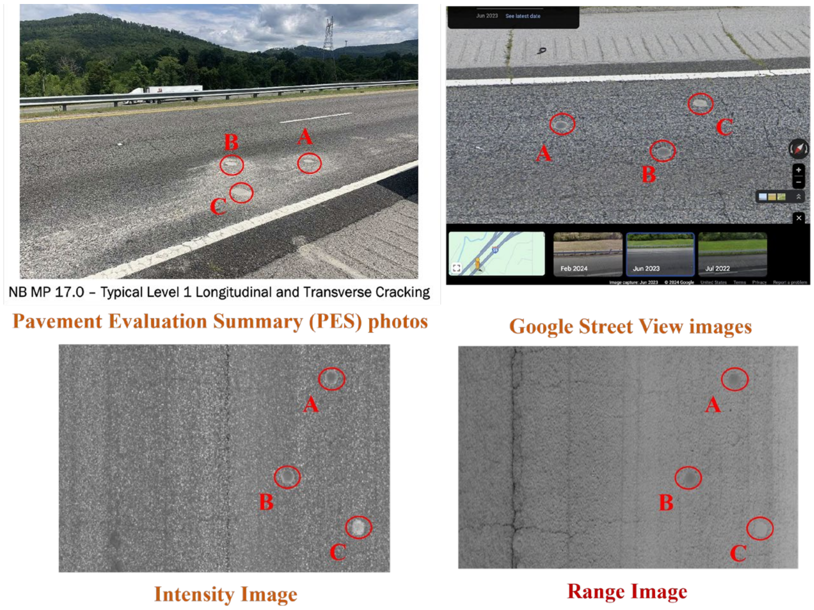

Figure 2. Cross-source visual validation of coring location.

Table 4. Representative coring and HWTT variables (I-575 NB).

Category	Example Fields	Units	Example Value*
Location & roadway	Route, Direction, Lane Number, Wheel Path	—	I-575, NB, Lane 2, Center
Position reference	Mile Post, GPS Latitude, GPS Longitude	mi/deg	MP 1.8; 34.02606; −84.55997
Sampling metadata	Coring Date	mm/yyyy	05/2023
HWTT — Layer 1	HWTT Result, Number of Passes, Max Deformation	— /cycles/mm	Passed; 20,001; 2.96
HWTT — Layer 2	HWTT Result, Number of Passes, Max Deformation	— /cycles/mm	Failed; 2,280; 12.5
GPR reference	GPR Milepost, Distance to GPR Point	mi/m	1.804; 0.49
Measured thickness	Actual AC Thickness, Actual Base Thickness	in.	11.26; 18.63

3.3 Missing Data Analysis and Study Corridor Selection

This study relies on TSDD deflection measurements as the primary structural indicator. TSDD measurements represent the pavement deflection response under a moving load, and the resulting deflection bowl reflects the structural condition of the pavement layers. This measurement principle is most suitable for flexible asphalt pavement structures, where load-induced deflection is continuous and structurally interpretable. For rigid and composite pavements, the presence of a stiff concrete layer changes the deflection response and may make TSDD-based structural indicators unreliable. Therefore, corridor selection cannot be based only on the nominal availability of TSDD, GPR, and coring data. It must also consider whether the available TSDD records are from pavement sections where deflection-based indicators have valid structural meaning.

A missing data analysis was conducted to evaluate whether TSDD, GPR, and coring records could be reliably integrated along each candidate corridor. The purpose was not only to quantify the percentage of missing TSDD records, but also to examine whether the missingness was corridor-wide or localized. Corridor-wide missingness may indicate structural incompatibility or systematic survey limitations, whereas localized missingness near bridges, ramps, merge areas, or interchange zones can usually be handled through spatial filtering and quality control.

Figure 3 presents the spatial distribution of TSDD record gaps for the four candidate corridors, and Table 5 summarizes the corresponding lane-level missing percentages. I-24 shows the most severe missing data problem, with missing rates of 67.4% in Lane 1 and 54.1% in Lane 2. These high missing percentages are not isolated record-level gaps but indicate extensive corridor-level incompatibility. Because these gaps are associated with rigid or composite pavement sections where TSDD deflection indicators are not structurally reliable, I-24 was excluded from the TSDD-based structural analysis.

In contrast, I-285, I-575, and I-59 show substantially lower missing percentages. All I-285 lanes have missing rates below 8%, ranging from 1.1% in Lane 5 to 7.5% in Lane 1. I-575 has the most complete TSDD coverage, with missing rates of 1.4% and 1.6% for Lanes 1 and 2, respectively. I-59 also retains acceptable coverage, with missing rates of 6.3% in Lane 1 and 3.5% in Lane 2. These missing records are mainly localized rather than corridor-wide and can be addressed through pavement-type screening, spatial filtering, and record-level quality control. Therefore, I-285, I-575, and I-59 were retained for subsequent structural analysis.

Overall, the missing data analysis supports the exclusion of I-24 and the retention of I-285, I-575, and I-59. The key distinction is not simply the amount of missing data, but the spatial and structural meaning of the missingness. I-24 exhibits extensive missingness associated with pavement structural incompatibility, while the remaining corridors show limited and localized gaps that can be handled during data integration and quality control. This screening step ensures that the final analytical dataset is built from corridors with sufficient structurally interpretable TSDD coverage.

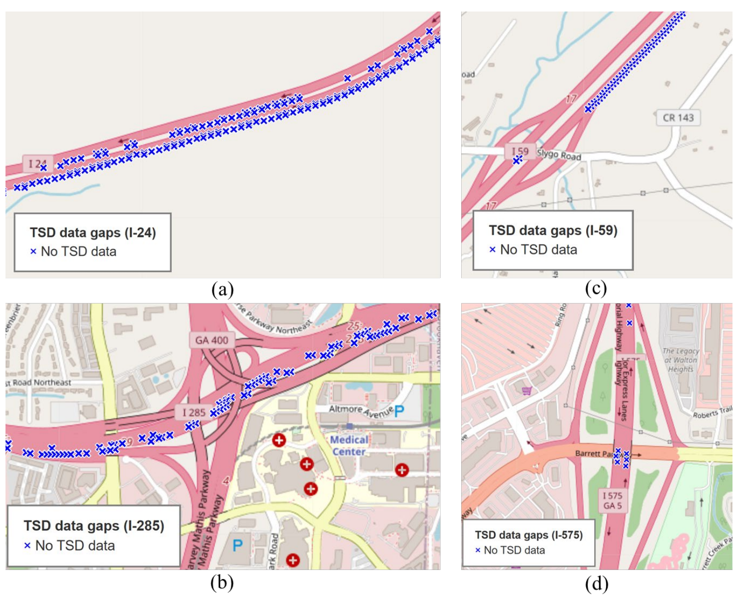

Figure 3. Map-based identification of TSDD missing data patterns across candidate corridors.

Table 5. Lane-wise TSDD missing data summary for candidate corridors.

Corridor	Lane	Total Records	Missing (%)
I-24	L 1	821	67.4%
I-24	L 2	821	54.1%
I-285	L 1	2,362	7.5%
I-285	L 2	2,617	3.2%
I-285	L 3	2,614	2.8%
I-285	L 4	1,908	1.6%
I-285	L 5	1,054	1.1%
I-575	L 1	4,025	1.4%
I-575	L 2	4,027	1.6%
I-59	L 1	1,710	6.3%
I-59	L 2	1,709	3.5%

3.4 Spatial Data Integration

After preprocessing, a spatial integration procedure was conducted to establish point-level correspondence among TSDD, GPR, and HWTT records. The three datasets differ substantially in spatial resolution, sampling density, and survey trajectory: TSDD and GPR data are collected continuously along mainline corridors at 0.01-mile intervals, whereas HWTT cores are located at sparse, discrete positions that may not lie on the mainline survey path. Direct record-level matching is therefore infeasible. A distance-based nearest-neighbor spatial join framework was adopted, comprising coordinate projection, a two-stage hierarchical matching procedure, quantitative distance diagnostics, map-based spatial quality control, and a final coring-centered refinement step.

3.4.1 Coordinate Projection

All datasets with recorded GPS coordinates were first converted into geospatial point layers. The original GPS coordinates were recorded in the WGS84 geographic coordinate system (EPSG:4326), which represents locations using latitude and longitude. Because latitude and longitude are angular units rather than linear distance units, they are not appropriate for directly evaluating whether two pavement records are within a few meters of each other. Therefore, all point layers were projected to the Georgia West State Plane coordinate system (EPSG:3460), a projected coordinate system designed for accurate distance calculations within Georgia. This projection converts the data into a planar coordinate system with linear units and reduces distance distortion within the study area. This step was necessary because the spatial integration procedure used small matching thresholds, including 3 m for TSDD–GPR matching and 9.2 m for HWTT–TSDD matching.

3.4.2 Two-Stage Spatial Join Framework

Stage 1: TSDD–GPR Integration. The first integration step matched each TSDD record to the nearest GPR record based on spatial distance. Because TSDD and GPR were collected through different survey programs and by different sensing systems, their records do not share identical sampling locations. A maximum matching distance of 3 m was therefore applied to allow for GPS positioning uncertainty and small differences between the TSDD and GPR survey paths. If multiple GPR points were located within 3 m of a TSDD point, the nearest GPR point was selected as the match. If no GPR point was found within 3 m, the TSDD record was left unmatched. The 3 m threshold provides a strict spatial tolerance for linking two continuous roadway datasets: it is small enough to exclude records that are clearly outside the local survey path, while still allowing valid matches affected by normal positioning and sampling differences. The resulting matching distance was retained as distance_gpr and used in the distance diagnostics described in Section 3.4.3. Available lane or lane-range information was also used during interpretation to verify that the assigned GPR thickness represented the appropriate pavement section. Records with questionable lane, pavement-type, or location consistency were further evaluated during the map-based quality control and coring-centered refinement steps described in Sections 3.4.3 and 3.4.4.

Stage 2: TSDD–GPR to HWTT Integration. The second integration step linked each HWTT location to the nearest record in the integrated TSDD–GPR dataset. Unlike TSDD and GPR, which are continuous roadway sensing datasets, HWTT results are tied to discrete coring locations. Therefore, a larger maximum search radius of 9.2 m was used to account for the sparse nature of coring points and possible offsets between the actual coring location and the mainline sensing trajectory. If multiple TSDD–GPR records were located within this radius, the nearest record was selected. If no TSDD–GPR record was found within 9.2 m, the coring location was left unmatched. The resulting matching distance was retained as distance_hamb, which was used as a quality-control variable to evaluate whether the core and the assigned TSDD–GPR record represented the same physical pavement section. As described in Section 3.4.3, large distance_hamb values were not treated as valid matches by default; instead, they were reviewed as potential evidence of geometric incompatibility, such as coring locations on ramps, auxiliary lanes, or merge areas rather than the mainline survey path.

3.4.3 Distance Diagnostics and Map-Based Quality Control

After the two-stage spatial join, matching distances were examined to evaluate whether the assigned records represented meaningful spatial relationships. For the TSDD-GPR join, the distance between each TSDD record and its assigned GPR record was retained as distance_gpr. For the HWTT-TSDD/GPR join, the distance between each coring location and its assigned integrated TSDD-GPR record was retained as distance_hamb. These distances were summarized using percentile plots, as shown in Figure 4.

The TSDD–GPR matching distances remained close to zero across nearly all percentiles, indicating that the two continuous roadway datasets followed closely aligned survey paths. This result supports the use of the 3 m threshold for linking TSDD and GPR records. In contrast, HWTT-to-TSDD/GPR matching distances showed a wider distribution. Many coring locations were close to their assigned TSDD–GPR records, but the upper percentiles increased sharply, with some distances exceeding 100 m. This long tail indicated that some coring locations were not located directly on the same mainline path as the nearest TSDD–GPR record. In such cases, the nearest-neighbor algorithm could still identify a closest available record, but that match might not represent the same physical pavement section.

Therefore, map-based visualization was used to inspect the spatial relationship between each coring location and its assigned TSDD–GPR record. Two matching categories were defined. Records were labeled LANE_MATCH when the coring location was located on or immediately adjacent to the mainline lane surveyed by TSDD and the assigned TSDD–GPR record appeared to represent the same pavement section. These records were considered spatially reliable and were carried forward to the next refinement step.

Records were labeled FALLBACK_DIST when the nearest-neighbor match required additional verification. These cases typically occurred when the coring location with HWTT results was not clearly aligned with the mainline TSDD–GPR survey path, or when the coring location appeared to be associated with a ramp, auxiliary lane, merge/diverge area, or other geometrically complex roadway segment. The FALLBACK_DIST label was used as a quality-control flag rather than an automatic exclusion rule. These records were carried forward for manual review during the coring-centered refinement step described in Section 3.4.4.

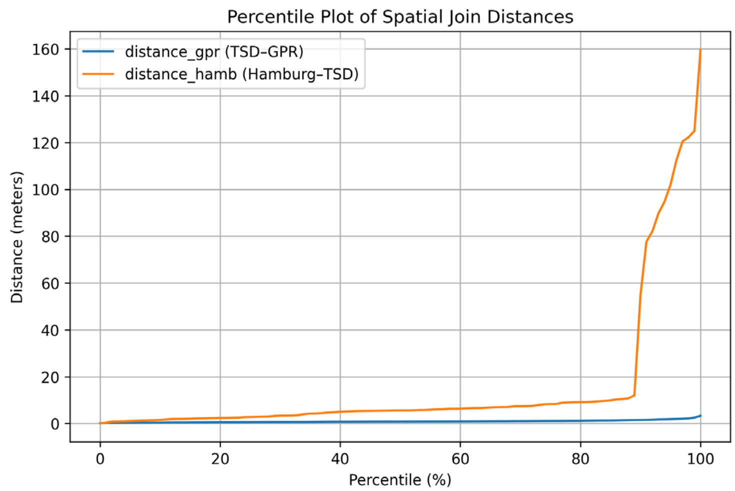

Figure 4. Percentile plot of spatial join distances for the two integration stages.

3.4.4 Coring-Centered Data Refinement

Following the distance diagnostics and map-based quality control, a final coring-centered refinement step was applied to determine which matched records should be retained in the analytical dataset. This step was necessary because spatial proximity alone does not guarantee structural consistency. A coring location with HWTT results may be close to a TSDD–GPR record but still correspond to a rigid, composite, ramp, auxiliary-lane, or other non-mainline pavement structure that is not appropriate for the flexible-pavement structural analysis in this study.

Each coring location was reviewed using coring logs, GPR-derived layer information, pavement-type indicators, roadway location context, and the map-based matching label assigned in Section 3.4.3. Records corresponding to rigid, composite, or other non-flexible pavement structures were excluded. This screening was especially important for I-285, where some locations were spatially close to the mainline sensing path but GPR confirmed the presence of a PCC layer. These records were removed because their TSDD deflection response would not be comparable to that of flexible asphalt pavement sections.

Coring locations on ramps, auxiliary lanes, merge/diverge areas, or other non-mainline segments were also reviewed. Records initially labeled FALLBACK_DIST were not automatically discarded. Instead, they were retained only when manual review confirmed that the coring location with HWTT results and the assigned TSDD-GPR record represented the same flexible mainline pavement section. If the location context, GPR layer information, or map-based inspection indicated that the coring location represented a different pavement structure or roadway segment from the assigned TSDD-GPR record, the record was excluded.

Together, the nearest-neighbor spatial join, distance diagnostics, map-based quality control, and coring-centered refinement formed the final data integration and quality-control pipeline. The resulting dataset retained only coring locations for which the HWTT outcome could be paired with TSDD and GPR measurements representing the same flexible mainline pavement section.

3.5 Final Dataset Composition

After spatial integration, map-based quality control, and coring-centered refinement, the final analytical dataset was organized at the coring-site level. Each row represents one retained coring site. For each site, the dataset combines the HWTT outcome, GPR-derived layer thickness, available coring-based thickness information, matched TSDD structural indicators, matched 3D surface condition variables, and spatial matching metadata. This structure allows the laboratory HWTT result at each coring location to be analyzed together with field measurements assigned to the same physical pavement section.

The final dataset includes only records that passed both spatial matching review and coring-centered structural consistency checks. In other words, records were retained only when the coring site with HWTT results and the assigned TSDD–GPR measurements could be interpreted as representing the same flexible mainline pavement section. Records initially labeled FALLBACK_DIST were included only if subsequent manual review confirmed this consistency; otherwise, they were excluded during refinement.

After applying all selection and refinement criteria, the final analytical dataset consisted of 85 retained pavement sections distributed across I-285, I-575, and I-59. Table 6 presents selected example records illustrating the structure of the final integrated dataset. The table shows how laboratory HWTT outcomes, GPR layer information, TSDD structural indicators, and 3D surface condition variables are combined into a single coring-centered analytical record. The complete field structure of the final analytical dataset is provided in Appendix E.

Table 6. Selected example records from the final integrated dataset.

Route	**Dir**	**MP**	**Lane**	**Wheel Path**	**Layer 1 HWTT**	**Layer 2 HWTT**	**GPR AC (in.)**	**GPR Base (in.)**	**TSDD Lane**	**SCI_8**	**SCI_12**	**D0 (mils)**	**Rut Avg (in.)**	**IRI Avg (in./mi)**
I-575	NB	1.8	2	CWP	Passed	Failed	18.17	11.26	Lane 2	0.84	0.85	−0.30	0.21	55
I-285	EB	28.6	1	LWP	Failed	Passed	16.81	9.95	Lane 1	1.00	1.01	−3.60	0.10	75
I-59	NB	14.0	2	RWP	Failed	Failed	10.72	18.28	Lane 2	2.43	2.44	−2.60	0.33	116

Note: CWP = center wheel path; LWP = left wheel path; RWP = right wheel path.

3.6 Descriptive Statistics and Exploratory Analysis

This section presents descriptive statistics for the final integrated dataset. The summaries compare corridor-level patterns across TSDD structural response, 3D surface condition, GPR-derived pavement structure, and HWTT-based laboratory performance. They also help determine whether the different data sources describe a consistent condition pattern across the retained corridors. For example, a corridor with poorer HWTT outcomes would be expected to show weaker TSDD

3.6.1 TSDD Structural Indicators

TSDD structural indicators were summarized to characterize the in-situ pavement response of the retained corridors. Because many TSDD variables describe related aspects of the deflection bowl, this section focuses on two representative indicators: SCI_12 and D0. SCI_12 describes near-surface curvature response and is commonly used to reflect asphalt-layer structural condition, while D0 represents the estimated deflection at the nominal loading position and reflects the combined contribution of the pavement layers and supporting foundation. Together, these two variables provide a concise description of both curvature-based and deflection-magnitude-based structural response.

Table 7 presents the lane-level descriptive statistics of SCI_12 and D0 for I-285, I-575, and I-59. I-575 shows the lowest and most stable TSDD response among the retained corridors, with mean SCI_12 values between 0.4 and 0.5 mils and mean D0 around −2.7 mils. I-285 shows moderate but more variable structural response, with mean SCI_12 values ranging from 0.4 to 0.9 mils across lanes and larger standard deviations than I-575. I-59 shows the highest curvature response and the largest deflection magnitude, with mean SCI_12 values around 2.0 mils and mean D0 around −4.6 mils.

These results indicate a clear corridor-level structural gradient in the retained dataset. I-575 represents the most stable flexible mainline condition, I-285 represents a more variable urban beltway condition, and I-59 represents the weakest structural condition among the retained corridors. This TSDD-based pattern provides structural context for the modeling and threshold-based analyses developed in subsequent chapters.

Table 7. Descriptive statistics of representative TSDD indicators by corridor and lane.

Corridor	Lane	Count	Missing (%)	Mean SCI_12 (mils)	Std SCI_12	Mean D0 (mils)	Std D0
I-285	L1	2,185	7.5	0.9	0.8	−3.0	2.7
I-285	L2	2,533	3.2	0.4	0.7	−3.0	2.7
I-285	L3	2,542	2.8	0.6	0.8	−3.0	2.7
I-575	L1	3,970	1.4	0.4	0.3	−2.7	1.5
I-575	L2	3,961	1.6	0.5	0.5	−2.7	1.5
I-59	L1	1,602	6.3	2.1	0.7	−4.6	1.9
I-59	L2	1,649	3.5	1.9	0.9	−4.6	1.9

3.6.2 3D Pavement Surface Indicators

The 3D pavement surface variables were summarized to characterize the functional condition of the final analytical dataset. Unlike the TSDD indicators, which describe pavement response under loading, the 3D surface indicators describe observable surface condition. In this section, the analysis focuses on cracking, rutting, and roughness measures that were retained in the integrated dataset.

Table 8 summarizes selected 3D surface indicators by corridor. I-575 shows the lowest values across cracking, rutting, and roughness measures, indicating the most favorable functional surface condition among the retained corridors. I-59 shows the highest cracking and rutting levels, with a mean total crack percentage of 40.20%, mean total crack length of 69.34 ft, and mean rut depth of 0.25 in. I-285 shows intermediate cracking and rutting levels, but it has the highest mean IRI among the three corridors, indicating greater roughness variation in the retained I-285 sections.

Overall, the 3D surface indicators show that the retained corridors represent different functional condition states. I-575 provides the most favorable surface condition group, I-59 represents the most distressed group in terms of cracking and rutting, and I-285 shows moderate surface distress with higher roughness. These results provide functional-condition context for the treatment decision analysis developed in later chapters.

Table 8. Summary of selected 3D pavement surface indicators in the final analytical dataset.

Corridor	**Retained Records**	**Mean Total Crack (%)**	**Mean Total Crack Length (ft)**	**Mean High-Severity Crack Length (ft)**	**Mean Rut Avg (in.)**	**Mean Rut Lane (in.)**	**Mean IRI Avg (in./mi)**	**Mean IRI Lane (in./mi)**
I-285	34	19.95	39.46	38.70	0.17	0.21	100.62	81.32
I-575	38	8.91	13.70	13.06	0.15	0.18	57.21	44.08
I-59	13	40.20	69.34	66.06	0.25	0.30	86.77	63.00

3.6.3 GPR Layer Thickness

GPR-derived layer thickness was summarized to characterize the pavement structural context of the final integrated dataset. Because the analysis focuses on flexible pavement sections, the primary variable of interest is AC thickness. Median thickness and interquartile range (IQR) were used instead of relying only on mean values because GPR thickness data may include local extremes near lane transitions, structural transitions, or interchange-related areas. The median represents the typical retained thickness level, while the IQR describes the main spread of thickness values across retained sections.

Table 9 summarizes the GPR-derived AC thickness statistics for the final analytical dataset. I-59 shows the thinnest and most uniform AC structure, with a median AC thickness of 9.82 in. and an IQR of 1.32 in. I-285 and I-575 both show thicker AC structures than I-59, with median AC thicknesses of 16.59 in. and 14.72 in., respectively. I-285 has the widest interquartile spread, indicating greater variability among the middle 50% of retained sections. I-575 has a similar IQR but includes the largest maximum AC thickness in the dataset.

Table 9. GPR-derived AC thickness statistics in the final analytical dataset.

Corridor	Retained Records	Median AC Thickness (in.)	P25–P75 (in.)	IQR (in.)	Min–Max (in.)	Structural Summary
I-285	34	16.59	13.22–18.26	5.04	10.69–21.10	Thick AC structure with the widest interquartile spread
I-575	38	14.72	13.76–18.08	4.31	11.54–23.46	Thick AC structure with several very thick retained sections
I-59	13	9.82	9.09–10.41	1.32	7.55–11.34	Thinner and more uniform AC structure

Overall, the retained dataset includes a meaningful range of AC thickness conditions. I-59 represents a thinner and more uniform flexible pavement group, while I-285 and I-575 represent thicker AC pavement groups with greater thickness variability. This variation provides structural context for interpreting TSDD response and supports the use of AC thickness as a conditioning variable in the threshold-based analysis developed in the next chapter.

3.6.4 HWTT Outcomes

HWTT outcomes were summarized for the final analytical dataset after spatial integration, map-based quality control, and coring-centered refinement. Because the final dataset is organized at the HWTT location level, each retained record has one Layer 1 HWTT outcome and, where available, one Layer 2 HWTT outcome. Layer 1 outcomes are reported for all 85 retained records, while Layer 2 HWTT outcomes are available for 77 records.

Table 10 summarizes the HWTT pass/fail distribution by corridor and layer. For Layer 1, the final dataset contains 48 passing records and 37 failing records. The distribution differs across corridors: I-575 contains 38 retained records, all of which passed Layer 1; I-285 contains 34 retained records, with 10 passing and 24 failing Layer 1; and I-59 contains 13 retained records, all of which failed

Layer 1. For Layer 2, the dataset contains 37 passing records, 40 failing records, and 8 records without reported Layer 2 HWTT outcomes.

Table 10. HWTT outcomes in the final analytical dataset.

Corridor	Retained Records	L1 Pass	L1 Fail	L2 Pass	L2 Fail	L2 Not Reported
I-285	34	10	24	11	21	2
I-575	38	38	0	26	12	0
I-59	13	0	13	0	7	6
Total	**85**	**48**	**37**	**37**	**40**	**8**

Because Layer 1 HWTT outcome is used as the primary performance reference in this study, subsequent modeling and analyses focus on the Layer 1 pass/fail result. Layer 2 HWTT outcomes are retained in the dataset where available and are reported here for completeness, but they are not used as the primary modeling label.

3.7 Chapter Summary

This chapter developed the data foundation for the structural classification and treatment decision analyses presented in the following chapters. Because the source datasets were collected through different programs, at different spatial resolutions, and for different purposes, the main task was not simply to compile available records. Instead, this chapter determined which TSDD, GPR, 3D surface, and PES/coring records could be meaningfully integrated to represent the same flexible pavement section.

- **The missing data analysis showed that corridor selection must be based on structural interpretability rather than data availability alone.** Although I-24, I-59, I-285, and I-575 all had relevant data sources, their usefulness depended on whether the TSDD measurements could be interpreted within a flexible-pavement structural framework and reliably linked to

GPR and HWTT records. I-24 was excluded because its missing TSDD data were dominated by rigid or composite pavement sections, leaving insufficient structurally compatible coverage. In contrast, I-285, I-59, and I-575 were retained because their data limitations were localized and could be addressed through pavement-type screening, spatial filtering, and location-level quality control.

- **The final dataset was built by matching continuous TSDD and GPR records first, then linking the integrated sensing records to discrete HWTT coring locations.** In the first matching stage, TSDD and GPR records were joined using a 3 m threshold to account for small spatial offsets between continuous sensing datasets. In the second stage, HWTT coring locations were matched to the integrated TSDD-GPR records using a 9.2 m threshold. The matched records were then reviewed through map-based quality control and coring-centered refinement to reduce the risk of combining measurements from different pavement structures, lanes, ramps, auxiliary lanes, or non-mainline segments.
- **The final analytical dataset consists of 85 flexible pavement sections, each linking one Layer 1 HWTT outcome with matched TSDD, GPR, and 3D surface measurements.** Each retained record includes a Layer 1 HWTT Pass/Fail label, TSDD structural indicators, GPR-derived pavement thickness, 3D surface condition variables, and spatial matching metadata. This coring-centered structure provides a consistent basis for testing whether field-measured structural response and pavement layer information can predict laboratory asphalt-layer performance.
- **The retained corridors provide three distinct condition groups that support model development and interpretation.** I-575 represents the favorable condition group, with

passing Layer 1 HWTT outcomes and lower TSDD structural responses. I-59 represents the poor-performance group, with failing Layer 1 HWTT outcomes and higher TSDD curvature and deflection responses. I-285 provides mixed HWTT outcomes and greater structural variability. This condition contrast supports the development, validation, and interpretation of both threshold-based and ML-based structural screening methods in the following chapters.

CHAPTER 4. EXPERIMENTAL DESIGN AND THRESHOLD-BASED STRUCTURAL CONDITION EVALUATION

Chapter 3 established the data foundation for this study by constructing an integrated dataset that links TSDD measurements, GPR layer thickness estimates, and HWTT outcomes for 85 pavement sections across three Georgia interstate corridors. This chapter builds on that foundation by defining the analytical framework used to evaluate structural condition classification methods and presenting the results of two threshold-based approaches.

The chapter is organized as follows. Section 4.1 defines the laboratory-based structural-performance reference used throughout the study and justifies the choice of the Layer 1 HWTT outcome as the binary classification target. Section 4.2 characterizes the final dataset and its class distribution. Section 4.3 describes the candidate structural indicator set and provides the physical rationale for each group of variables. Section 4.4 explains the intentional exclusion of 3D surface distress indicators from the structural screening framework. Section 4.5 documents the unified repeated stratified cross-validation (CV) protocol that is shared by all analytical methods in Chapters 4 and 5. Sections 4.6 and 4.7 develop and evaluate the single-variable SCI threshold rule and the duo-variable thickness-conditioned threshold rule, respectively. Section 4.8 synthesizes the comparative findings and provides deployment guidance that bridges to the ML evaluation in Chapter 5.

4.1 Ground Reference Definition

Every classification model requires a clearly defined target variable. In this study, the laboratory-based structural-performance reference is derived from the HWTT, specifically the Layer 1 performance outcome. Each pavement section is assigned a binary label (Pass or Fail) based on whether the extracted Layer 1 asphalt specimen exceeded the predefined deformation failure criterion under repeated wheel loading and controlled water immersion, as described in Chapter 3.

Three considerations motivate the use of Layer 1 HWTT outcome as the classification target. First, the selected explanatory indicators are most directly related to the upper asphalt layer. TSDD-based SCI measures mainly reflect near-surface stiffness changes, while GPR-derived AC thickness describes the geometry of the asphalt layer. Therefore, using the Layer 1 HWTT result as the ground reference keeps the target variable consistent with the physical meaning of the input indicators. In contrast, using a deeper-layer result, such as Layer 2, or a composite full-core outcome would link near-surface indicators to a performance reference from a different structural depth, reducing the physical interpretability of the classification.

Second, a binary classification framework matches the screening purpose of this study. The objective is not to predict a continuous HWTT deformation value, but to identify pavement sections that may have structural vulnerability based on network-level TSDD and GPR data. A Pass/Fail output is easier to translate into agency practice because GDOT engineers already use categorical condition ratings in M&R decisions. Therefore, the binary formulation allows the model results to be interpreted and applied within existing pavement management workflows.

Third, the HWTT provides a more standardized basis for threshold calibration than expert judgment or field observation alone. The HWTT applies repeated wheel loading and moisture conditioning to extracted asphalt specimens under controlled laboratory conditions, and the Pass/Fail result is determined using a predefined deformation criterion. This makes the ground reference repeatable and closely related to mixture performance under loading and moisture exposure. As a result, the calibrated threshold is tied to a mechanistic laboratory reference rather than to subjective condition ratings or indirect structural indices.

One limitation should also be noted. The HWTT outcome is a laboratory-based proxy for structural performance rather than a direct observation of field failure. In-service pavement performance is affected by construction quality, traffic history, environmental exposure, drainage, and other site-specific factors that are not fully represented in the laboratory test. Therefore, this study uses the Layer 1 HWTT Pass/Fail outcome as an operational laboratory-based performance reference for evaluating classification methods, while recognizing that it provides a repeatable and mechanistically meaningful benchmark rather than a complete description of field performance.

4.2 Dataset Overview and Class Distribution

The final analytical dataset consists of 85 pavement sections distributed across three interstate corridors. The Layer 1 HWTT ground reference assigns 48 sections as Pass and 37 as Fail, corresponding to a moderate class imbalance at the network level (43.5% Fail). Table 11 summarizes the distribution by corridor.

Table 11. Distribution of the final analytical dataset by corridor and Layer 1 HWTT.

Route	**Total Sections**	**Pass**	**Fail**	**% Fail**
I-285	34	10	24	70.6%
I-575	38	38	0	0%
I-59	13	0	13	100%
Total	**85**	**48**	**37**	**43.5%**

The Pass/Fail distribution differs markedly across the three corridors. I-285 includes both classes, with 10 Pass sections and 24 Fail sections, and is therefore the only corridor that supports route-specific discrimination between structurally adequate and inadequate sections. In contrast, all retained I-575 sections are classified as Pass, while all retained I-59 sections are classified as Fail.

This pattern reflects the practical context of the coring program. Coring locations are typically selected based on engineering needs rather than random sampling, so the resulting dataset tends to emphasize corridors with known condition characteristics. In this dataset, I-575 represents a corridor with generally favorable structural performance, whereas I-59 represents a corridor with consistently poor Layer 1 HWTT outcomes.

This class distribution affects how model performance should be evaluated. Metrics such as F1 score and AUC are meaningful only when both Pass and Fail sections are represented in the test subset. At the same time, the single-class corridors cannot be excluded because they are needed to preserve the limited sample size and the structural diversity of the dataset. For this reason, the stratified resampling strategy described in Section 4.5 is used. This strategy maintains the overall Pass/Fail proportion within each CV split, which produces more stable and interpretable performance estimates across methods.

4.3 Candidate Indicator Set and Physical Interpretation

To support network-level structural screening, the candidate indicators were selected to represent the primary pavement response mechanisms measured by TSDD and GPR. The TSDD variables characterize load-induced deflection response, including near-load curvature and deflection basin shape, while the GPR variables provide asphalt and base layer thickness information needed to interpret the measured response. Together, these indicators describe three engineering dimensions relevant to Layer 1 HWTT performance: near-surface stiffness, overall structural response, and layer configuration. The indicator set was intentionally kept compact to avoid redundant variables while retaining the key response and thickness measures needed for interpretable structural classification.

Figure 5 provides a schematic view of the TSDD loading configuration and the resulting deflection bowl under an approximately 100 kN axle load. The sensor offsets shown in the figure, including D0 at the load center and the offset measurements D8, D12, and D18, describe the basin shape relative to the pavement layer structure. Short-offset curvature measures derived from these deflections are used to characterize near-surface response, with SCI indicators being most sensitive to AC layer stiffness.

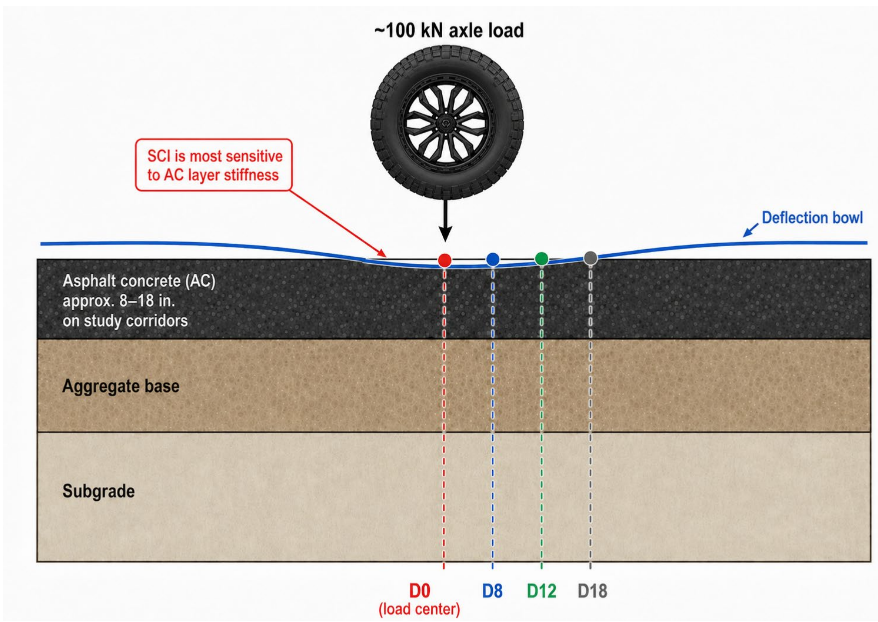

Figure 5. TSDD loading configuration, sensor offsets, and deflection bowl.

4.3.1 SCI-Based Curvature Indicators

Curvature-based indicators describe how rapidly the deflection bowl changes near the load center. Instead of using the absolute deflection at a single sensor, SCI measures use the difference between the load-center deflection and deflections measured at selected offsets. This makes SCI indicators useful for identifying localized stiffness changes in the upper pavement structure. Because Layer

1 HWTT performance is related to the condition of the upper asphalt layer, short-offset SCI measures provide a physically relevant input for structural classification.

Three curvature indicators are included. SCI_8 = D8 – D0 and SCI_12 = D12 – D0 represent short-offset curvature behavior associated primarily with asphalt layer response. SCI_SUBGRADE = D60 – D36 characterizes curvature over a longer effective distance, reflecting deeper structural contribution and underlying support conditions. Together, these three indicators provide a structured curvature descriptor set spanning shallow near-surface response to deeper composite structural behavior.

4.3.2 Representative Deflection Indicators

Absolute deflection indicators were included to describe the magnitude and shape of the measured TSDD deflection bowl. While SCI variables summarize curvature near the load center, individual deflection readings provide additional information on the overall compliance of the pavement structure. D0 represents the load-center deflection and reflects the primary basin magnitude. D-8, D-12, and D8 describe the near-center basin shape, while D-18 provides information from a farther offset and helps capture the basin tail. These representative offsets were selected to cover the main response zone without introducing excessive redundancy from closely spaced sensor measurements.

4.3.3 GPR-Derived Thickness Indicators

Deflection response must be interpreted in relation to pavement layer thickness because similar TSDD basin shapes can result from different structural configurations. A section with a thicker AC layer may show lower curvature and smaller deflection because the applied load is distributed through a deeper and stiffer structural system. In contrast, a thinner AC section may produce higher curvature under the same loading condition because the load is concentrated over a shallower asphalt layer. Therefore, deflection magnitude alone cannot distinguish whether a measured response is caused by material weakness, insufficient structural thickness, or the combined effect of both.

GPR-derived thickness variables were included to provide this structural context. AC thickness is directly related to near-surface load spreading, tensile strain development, and the stiffness contribution of the asphalt layer. It is also the thickness measure most closely aligned with the Layer 1 HWTT outcome because both describe the upper asphalt structure. Base thickness provides additional information on support conditions beneath the asphalt layer and contributes to the composite load-carrying capacity of the pavement system. Together, AC thickness and base thickness allow the analysis to account for both measured response and structural configuration. This makes it possible to evaluate whether TSDD response indicators alone are sufficient for classification, or whether thickness-conditioned interpretation improves the screening of Layer 1 HWTT Pass/Fail outcomes. The latter question is examined directly through the duo-variable threshold method in Section 4.7.

4.4 Rationale for Excluding 3D Surface Distress Indicators

Although 3D surface distress data, including cracking percentage, IRI, and rut depth, are important for pavement management, they were excluded from the structural screening models in this chapter. The objective here is to evaluate whether TSDD response measures and GPR layer thickness can identify structural vulnerability relative to the Layer 1 HWTT ground reference. For this purpose, the input variables should represent structural response and layer configuration rather than visible surface condition.

Surface distress indicators reflect the condition observed at the pavement surface, but they are not direct measures of structural capacity. Cracking, rutting, and roughness can be affected by traffic, aging, drainage, maintenance history, and localized repairs. They may be associated with structural weakness, but they can also occur for reasons that are not directly related to subsurface structural integrity. Including them in the Chapter 4 models would blur the distinction between structural screening and surface distress recognition.

Surface variables are therefore reserved for Chapter 6, where they are used as functional condition indicators in the treatment decision framework. This separation keeps Chapter 4 focused on structural signal extraction from TSDD and GPR data, while Chapter 6 combines structural and surface information for M&R recommendations.

4.5 Unified Evaluation Protocol

All threshold-based and ML methods evaluated in this study are compared under an identical evaluation protocol. This unified design is essential. If different methods were evaluated using different data splits, performance metrics, or information leakage policies, the comparisons would reflect procedural differences rather than true differences in predictive capability. As shown in Figure 6, the protocol uses a Repeated Stratified K-Fold design with five folds and ten repeats, producing 50 independent train-test splits. All threshold optimization, feature selection, and model fitting steps are performed only on the training subset. The held-out test subset is used only once, for final performance evaluation within each split.

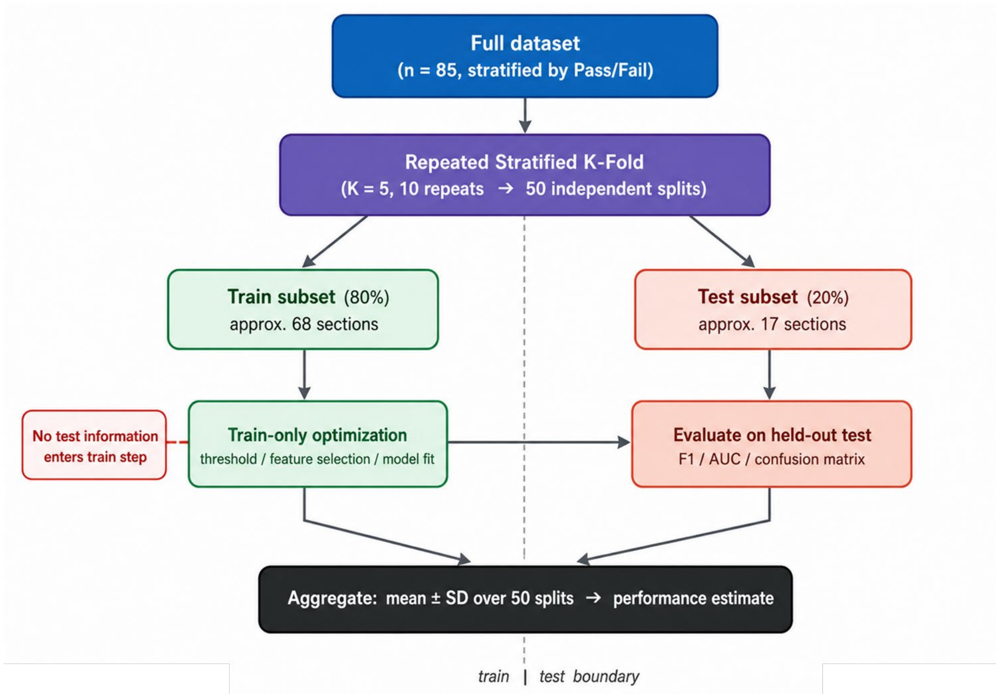

Figure 6. Unified cross-validation protocol.

4.5.1 Cross-Validation Design

A Repeated Stratified K-Fold CV design was used to evaluate all classification methods. The protocol used five folds and ten repetitions with a fixed random seed of 42, resulting in 50 train-test partitions. Stratification was applied to maintain the overall Layer 1 HWTT Pass/Fail proportion within each fold. This step is important because the dataset contains only 85 sections and has a moderate class imbalance, with Fail sections representing 43.5% of the data.

For each split, approximately 80% of the sections were used for training and the remaining 20% were held out for testing. The repeated design provides a more reliable performance estimate than a single train-test split because it reduces dependence on one particular data partition. It also allows the variability of model performance to be quantified across different resampled datasets.

Therefore, the mean and standard deviation across the 50 test results are reported as the primary performance summary throughout this chapter.

4.5.2 Training-Test Isolation

Training-test isolation was applied in every split to prevent information leakage and optimistic performance estimates. All rule selection, parameter calibration, feature selection, and model fitting were completed using the training subset only. The held-out test subset was used only for final performance evaluation.

For the single-variable threshold method, the rule direction and decision threshold were both determined from the training data. The rule direction was based on the Spearman correlation between the candidate indicator and the binary HWTT outcome. The threshold was then selected from a training-data candidate grid by maximizing the training F1 score. After calibration, the selected rule was applied directly to the held-out test subset without adjustment.

For the duo-variable threshold method, the AC thickness partition point and the two SCI thresholds for the thin and thick AC groups were optimized jointly on the training subset. The parameter combination with the highest training F1 score was selected and then applied unchanged to the test subset.

The same isolation principle was used for the ML models in Chapter 5. Correlation screening, VIF filtering, recursive feature elimination, and model fitting were all performed within the training subset. No test-set information was used during parameter selection, feature selection, model training, or performance optimization.

4.5.3 Performance Metrics

Model performance was evaluated using metrics appropriate for binary structural screening. In this study, Fail is treated as the positive class because the main screening objective is to identify pavement sections with potential structural vulnerability. The confusion matrix therefore includes four outcomes: true positives (TP), failed sections correctly classified as Fail; false positives (FP), passing sections incorrectly classified as Fail; false negatives (FN), failed sections incorrectly classified as Pass; and true negatives (TN), passing sections correctly classified as Pass.

F1 score was used as the primary performance metric because it balances precision and recall. Precision measures the reliability of failure flags, indicating how many sections classified as Fail are actually failed. Recall measures the ability to detect failed sections, indicating how many true Fail sections are successfully identified. In structural screening, both are important: low precision creates unnecessary follow-up investigation, while low recall misses structurally vulnerable sections. F1 summarizes this trade-off in a single metric and is more informative than overall accuracy when the dataset has moderate class imbalance.

AUC was reported as a secondary metric to evaluate the ranking ability of each method. Unlike F1, which depends on a selected decision threshold, AUC measures how well a method ranks failed sections as higher risk than passing sections across possible thresholds. For ML classifiers, this ranking is based on the predicted failure probability. For threshold-based rules, a continuous failure-risk score was derived from the underlying indicator value. The score direction was adjusted so that higher scores always represented greater failure risk. This allowed AUC to be compared consistently between threshold rules and ML classifiers.

All metrics were computed only on the held-out test subset in each split. The final results were summarized using the mean and standard deviation across the 50 CV partitions.

4.6 Single-Variable Threshold Method

The simplest possible structural screening rule classifies each pavement section based on a single measured indicator and a fixed decision boundary. This section first establishes which indicators carry the strongest structural failure signal through exploratory diagnostics, then evaluates the full candidate set under the unified CV protocol.

4.6.1 Methodology

For each candidate indicator, a single-threshold screening rule was calibrated using the training data in each CV split. The rule direction depended on the relationship between the indicator and the Layer 1 HWTT outcome. If higher indicator values were associated with failure, sections with values greater than or equal to the selected threshold were classified as Fail. If lower indicator values were associated with failure, sections with values less than or equal to the selected threshold were classified as Fail.

The rule direction was determined from the Spearman correlation between the candidate indicator and the binary HWTT outcome within the training subset. After the direction was set, a candidate set of threshold values was generated from the training data distribution. The optimal threshold was selected as the value that produced the highest F1 score on the training subset. The calibrated rule was then applied directly to the held-out test subset without any further adjustment. All reported performance metrics were computed only from the test subset, consistent with the training-test isolation policy described in Section 4.5.2.

4.6.2 Exploratory Indicator Diagnostics

Before evaluating the threshold rules, an exploratory correlation analysis was conducted to examine how each candidate indicator relates to the Layer 1 HWTT ground reference. This step helps identify which variables have the strongest direct relationship with structural failure and provides context for interpreting the cross-validated threshold results. It also shows whether multiple indicators contain overlapping information, which is important because highly correlated variables may add little new structural information when used together.

Figure 7 shows the Pearson correlation matrix for the candidate structural indicators and the binary Layer 1 HWTT outcome. The matrix is shown as a lower triangle to reduce repeated information. The HWTT outcome row is highlighted to focus attention on each indicator's direct association with the Pass/Fail ground reference. SCI_8 and SCI_12 show the strongest positive correlations with structural failure, both at approximately $r = 0.61$. This indicates that near-surface curvature has the clearest relationship with the Layer 1 HWTT failure classification among the tested indicators.

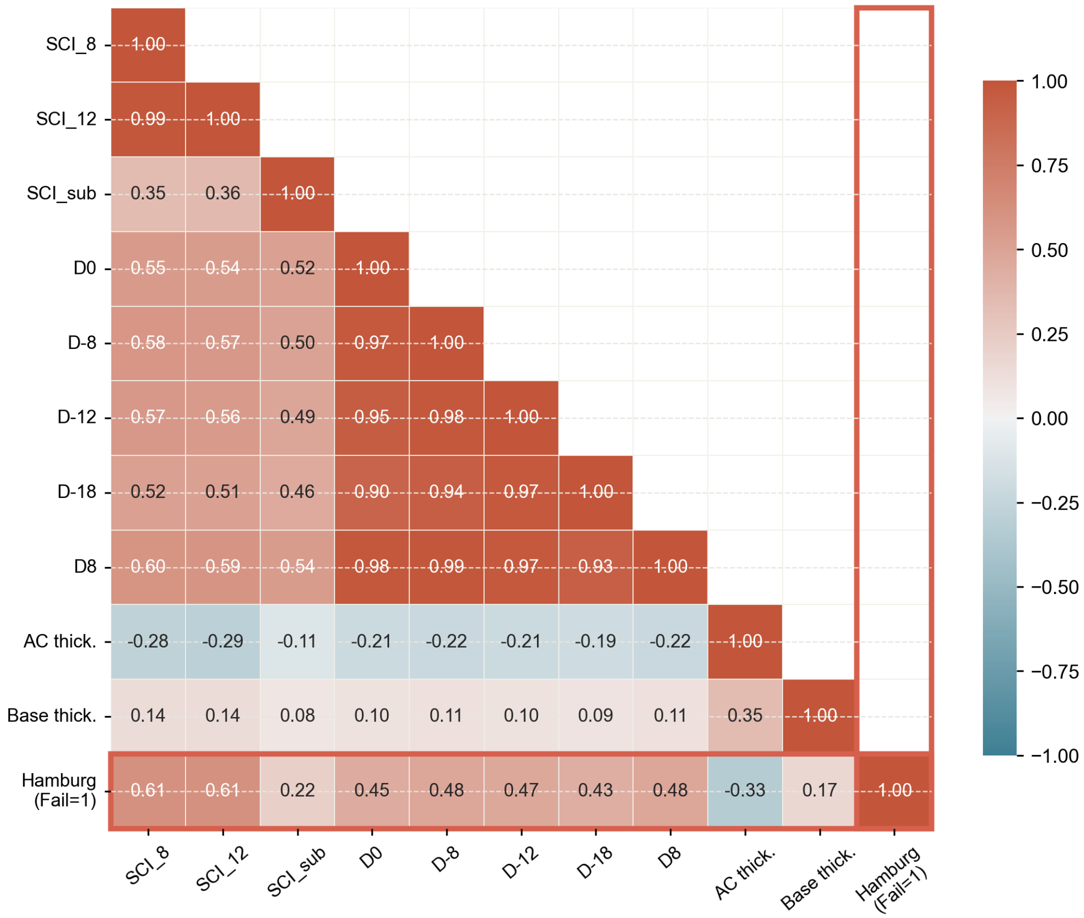

Figure 7. Pearson correlation matrix for the candidate structural indicators and the binary Layer 1 HWTT outcome.

AC thickness shows a moderate negative correlation with failure, at approximately r = −0.33. This means that thicker asphalt sections are generally associated with better Layer 1 HWTT outcomes, which is consistent with the load-spreading function of thicker AC layers. Base thickness has only a weak positive correlation, at approximately r = 0.17, suggesting limited standalone predictive value. Absolute deflection measures such as D0 and D-8 show moderate correlations of approximately r = 0.45 to 0.48. These relationships are weaker than those of the SCI indicators because absolute basin magnitude reflects both material stiffness and pavement layer configuration.

The matrix also shows substantial redundancy among several indicators. SCI_8 and SCI_12 are nearly perfectly correlated, with r ≈ 0.99, indicating that they represent almost the same near-surface curvature response. The absolute deflection indicators, including D0, D-8, D-12, D-18, and D8, also show mutual correlations greater than 0.90. This high collinearity indicates that adding multiple deflection magnitude variables is unlikely to provide much additional information. These results suggest that SCI_8 and SCI_12 should be expected to perform similarly, and that multi-variable models should be evaluated carefully to avoid relying on redundant predictors.

4.6.3 Cross-Validated Performance

Table 12 summarizes the cross-validated performance of the single-variable threshold rules. Each candidate indicator was evaluated using the same repeated stratified CV protocol. Values are reported as mean ± SD across 50 splits. The indicators are ordered by mean F1 score, and the rule direction indicates whether higher or lower values were associated with Layer 1 HWTT failure.

The results show that the SCI-based curvature indicators provide the strongest single-variable screening performance. SCI_8 achieves the highest mean F1 score of 0.806 and an AUC of 0.888. SCI_12 performs similarly, with a mean F1 score of 0.798 and an AUC of 0.875. These results are consistent with the correlation analysis in Section 4.6.2, where SCI_8 and SCI_12 had the strongest direct association with the Layer 1 HWTT outcome. Because the two indicators are nearly perfectly correlated, their similar performance is expected. In practical terms, either SCI_8 or SCI_12 can serve as the primary curvature-based screening indicator, depending on sensor availability and agency preference.

The absolute deflection indicators show weaker performance than the SCI indicators. D0, D-8, and D8 produce moderate F1 scores ranging from 0.614 to 0.678. Although these variables capture the overall magnitude of the deflection bowl, they are less effective as standalone classifiers because absolute deflection is influenced by both material condition and pavement layer thickness. As a result, a high or low deflection value does not always correspond directly to Layer 1 HWTT failure.

The thickness-only indicators also have limited standalone screening value. AC thickness achieves an F1 score of 0.585 and an AUC of 0.693, indicating that thickness contains some useful structural information but is not sufficient by itself to classify HWTT Pass/Fail outcomes. Base thickness has high recall but very low precision and AUC, which means it tends to flag many sections as Fail without reliably separating failed and passing sections. SCI_SUBGRADE shows a similar issue, with perfect recall but weak overall discrimination. These results indicate that some indicators can detect many failed sections but still perform poorly because they also misclassify many passing sections.

Overall, the single-variable results demonstrate that short-offset SCI curvature indicators are the most effective standalone predictors of Layer 1 HWTT structural failure. SCI_8 provides the best balance between precision, recall, and AUC, making it the strongest candidate for a simple threshold-based structural screening rule.

Table 12. Cross-validated performance of all single-variable threshold rules.

Indicator	Rule Direction	Threshold	F1	Precision	Recall	AUC
SCI_8	X ≥ t ⇒ Fail	0.585 ± 0.069	0.806 ± 0.103	0.846 ± 0.142	0.788 ± 0.130	0.888 ± 0.089
SCI_12	X ≥ t ⇒ Fail	1.012 ± 0.046	0.798 ± 0.113	0.867 ± 0.124	0.754 ± 0.148	0.875 ± 0.094
D0 (mils)	X ≤ t ⇒ Fail	−2.885 ± 0.211	0.678 ± 0.092	0.589 ± 0.105	0.820 ± 0.133	0.742 ± 0.128
D-8 (mils)	X ≤ t ⇒ Fail	−2.313 ± 0.265	0.624 ± 0.084	0.520 ± 0.079	0.799 ± 0.150	0.706 ± 0.137
D8 (mils)	X ≤ t ⇒ Fail	−2.137 ± 0.170	0.614 ± 0.097	0.495 ± 0.078	0.828 ± 0.171	0.668 ± 0.142
Base thickness (in.)	X ≤ t ⇒ Fail	20.69 ± 0.607	0.611 ± 0.028	0.446 ± 0.032	0.975 ± 0.056	0.160 ± 0.019
SCI_SUBGRADE	X ≤ t ⇒ Fail	−0.100 ± 0.000	0.598 ± 0.026	0.426 ± 0.027	1.000 ± 0.000	0.316 ± 0.081
D-18 (mils)	X ≤ t ⇒ Fail	−0.710 ± 0.534	0.594 ± 0.110	0.461 ± 0.071	0.857 ± 0.215	0.595 ± 0.142
AC thickness (in.)	X ≤ t ⇒ Fail	14.13 ± 2.571	0.585 ± 0.164	0.712 ± 0.215	0.558 ± 0.240	0.693 ± 0.150
D-12 (mils)	X ≤ t ⇒ Fail	−1.683 ± 0.495	0.571 ± 0.116	0.454 ± 0.083	0.796 ± 0.221	0.662 ± 0.147

4.6.4 Threshold and Parameter Stability

Strong predictive performance alone is not sufficient for an engineering screening rule. The calibrated threshold also needs to be stable. If a threshold performs well on average but changes substantially across different training samples, it cannot be confidently applied to new pavement sections.

Figure 8 shows the KDE distributions of the F1-optimal thresholds selected for SCI_8 and SCI_12 across the 50 repeated stratified CV splits. In each split, the threshold was calibrated using only the training subset. The distribution therefore shows how sensitive the selected threshold is to resampling. A narrow distribution indicates that the threshold is stable and not strongly dependent on one particular train-test partition.

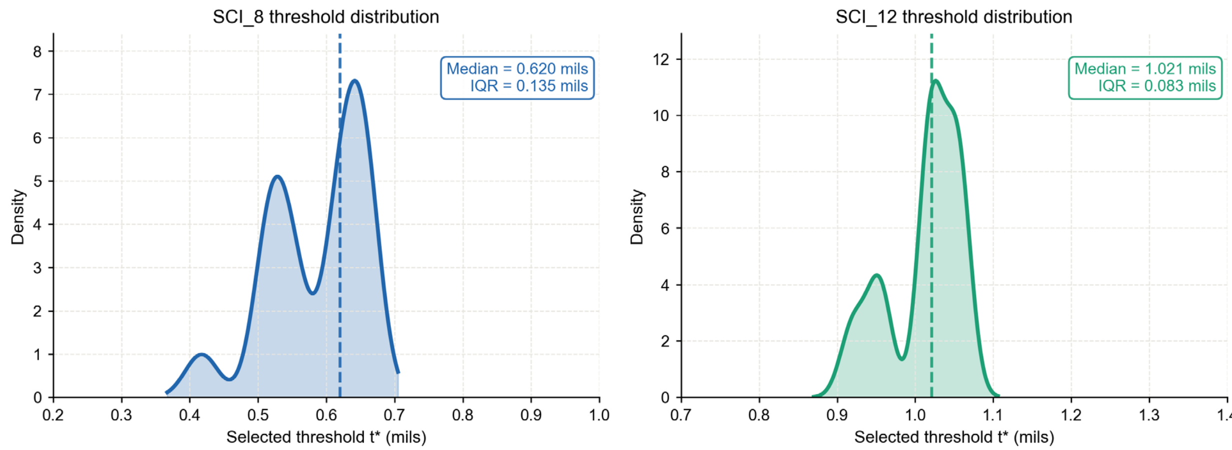

Figure 8. Stability of calibrated SCI thresholds.

Both SCI_8 and SCI_12 show concentrated threshold distributions. SCI_8 is centered near 0.62 mils, with an IQR of approximately 0.14 mils. SCI_12 is centered near 1.02 mils, with a narrower IQR of approximately 0.08 mils. These results indicate that the F1-optimal decision boundaries are relatively stable across the 50 splits. This stability is important for deployment because GDOT engineers need a calibrated threshold that can be applied consistently to future network survey data.

Both indicators provide stable calibration results. Therefore, the choice between SCI_8 and SCI_12 can be based primarily on sensor availability, agency convention, and implementation preference. Based on the CV results, the preliminary Georgia-specific threshold values, with mean SCI_8 of approximately 0.585 mils and mean SCI_12 of approximately 1.012 mils, can be considered reference values for GDOT network-level pilot screening and should be recalibrated as additional labeled data become available.

4.7 Duo-Variable Threshold Method

The single-variable analysis in Section 4.6 applies one global SCI_8 threshold to all 85 pavement sections. This approach is simple and easy to implement, but it assumes that the relationship between SCI_8 and Layer 1 HWTT failure is the same across all pavement structures. In practice, AC layer thickness affects how the TSDD load is distributed and how SCI_8 should be interpreted. A given SCI_8 value may represent different levels of structural concern in thin and thick asphalt sections. This section evaluates whether conditioning the SCI_8 threshold on AC thickness can improve the engineering logic of the screening rule while maintaining or improving predictive performance.

4.7.1 Motivation and Formulation

AC layer thickness affects both load spreading and near-surface strain development. A thicker asphalt layer can distribute the applied TSDD load over a larger structural depth and may produce lower SCI_8 values than a thinner asphalt layer under similar material conditions. Therefore, SCI_8 should not always be interpreted using one fixed global threshold.

Using a single threshold may lead to misclassification when pavement structures have different thickness configurations. A thick pavement section may be flagged as Fail if its SCI_8 value is slightly above the global threshold, even though the measured response may still be acceptable for its structural configuration. Conversely, a thin pavement section with structural vulnerability may be missed if its SCI_8 value falls slightly below the global threshold. In both cases, the problem is that the same SCI_8 value is interpreted without considering AC thickness.

The duo-variable framework addresses this issue by using AC thickness as a conditioning variable. The dataset is first divided into thinner and thicker AC groups, and a separate SCI_8 threshold is calibrated for each group. The rule is defined as follows:

If AC thickness $\leq t_{AC}$, classify as Fail when SCI_8 $\geq t_{thin}$.

If AC thickness $> t_{AC}$, classify as Fail when SCI_8 $\geq t_{thick}$.

Here, t_{AC} is the selected AC thickness split point, t_{thin} is the SCI_8 threshold for the thinner AC group, and t_{thick} is the SCI_8 threshold for the thicker AC group. In this formulation, SCI_8 measures the severity of the deflection curvature response, while AC thickness provides the structural context needed to interpret that response. The classification is therefore based on both measured response and pavement layer configuration.

4.7.2 Optimization Procedure

The duo-variable rule was calibrated using only the training subset in each CV split. Three parameters were selected: the AC thickness split point, the SCI_8 threshold for the thinner AC group, and the SCI_8 threshold for the thicker AC group.

The procedure first tested candidate AC thickness split points within the training data. Candidate split points were selected from percentile-spaced values between the 10th and 90th percentiles of AC thickness to avoid extreme partitions. Each candidate split divided the training sections into a thinner AC group and a thicker AC group.

For each candidate AC split point, separate SCI_8 thresholds were evaluated within the two thickness groups. The final parameter combination was the one that produced the highest F1 score on the full training subset. Candidate split points were excluded if they created a group with fewer than five training sections or a group containing only one class. This restriction prevents the optimization from selecting thresholds based on very small or non-discriminative subgroups.

After the best AC split point and the two SCI_8 thresholds were selected, the rule was applied directly to the held-out test subset without further adjustment. No test-set information was used during parameter selection, which keeps the evaluation consistent with the training-test isolation protocol described in Section 4.5.2.

4.7.3 Parameter Stability, Implementation Values, and Limitations

The duo-variable framework includes three calibrated parameters: the AC thickness split point, the SCI_8 threshold for the thinner AC group, and the SCI_8 threshold for the thicker AC group. Unlike the single-variable SCI_8 rule, the duo-variable method should not be interpreted as one fixed threshold value. Its parameters need to be examined to determine whether the thickness-conditioned rule is stable and whether the selected values have practical engineering meaning.

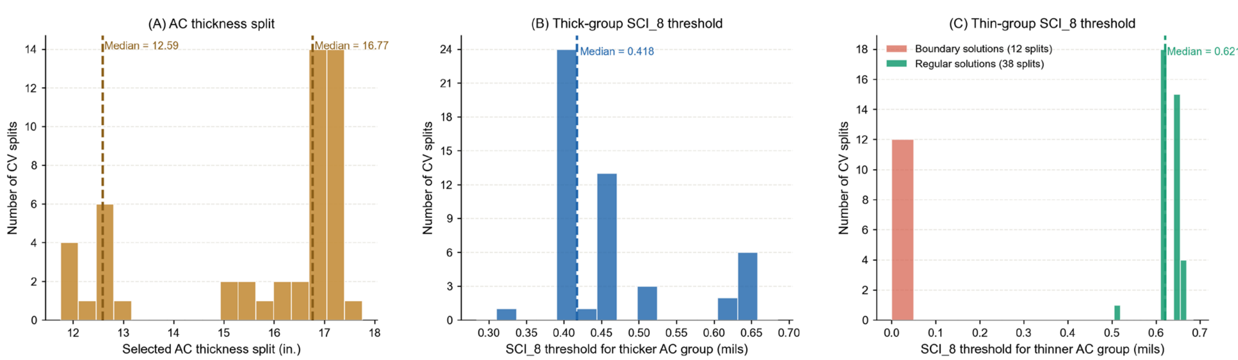

Figure 9 shows the parameter distributions across the 50 CV splits, and Table 13 summarizes the key distribution statistics. The AC thickness split point shows a clear bimodal pattern, with clusters near 12.6 in. and 16.8 to 17.0 in. This indicates that the optimization repeatedly identifies two practical thickness separation ranges rather than selecting random split points. From an engineering perspective, this suggests that the dataset contains distinct asphalt thickness regimes where the relationship between SCI_8 and Layer 1 HWTT failure may differ.

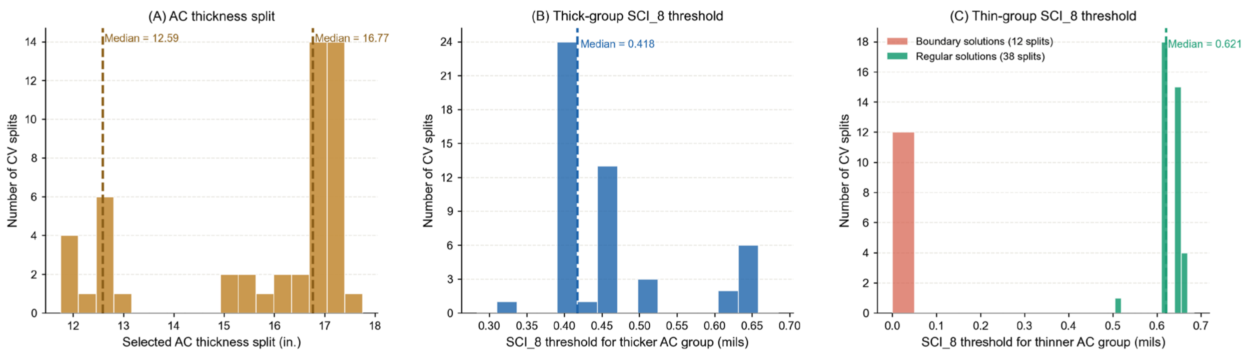

Figure 9. Stability of SCI_8 + AC thickness duo-threshold parameters across 50 CV splits.

Table 13. Summary of duo-variable threshold parameter distributions across 50 CV splits.

Parameter	**Mean ± SD**	**Median**	**P10 to P90 range**	**Interpretation**
AC thickness split	15.7 ± 1.9 in.	16.8 in.	12.6 to 17.2 in.	Bimodal distribution with clusters near 12.6 in. and 17.0 in.
SCI_8 threshold for thin AC group	0.48 ± 0.27 mils	0.62 mils	0.00 to 0.65 mils	Less stable; affected by boundary solutions in 12 splits.
SCI_8 threshold for thick AC group	0.47 ± 0.09 mils	0.42 mils	0.42 to 0.65 mils	More concentrated; indicates a relatively stable threshold for thicker AC sections.

Note: SD = Standard Deviation.

The SCI_8 threshold for the thicker AC group is relatively stable. Its median value is approximately 0.42 mils, and most selected values fall within a narrow range. This indicates that, for thicker asphalt sections, the SCI_8 failure boundary is reasonably consistent across resampling. A smaller secondary cluster near 0.65 mils appears in some splits, but the thick-group threshold remains more stable than the thin-group threshold overall.

The SCI_8 threshold for the thinner AC group is less stable. Its median value is approximately 0.62 mils, but 12 splits produce near-zero boundary solutions. These boundary solutions occur when the thinner AC group is small or has limited Pass/Fail variation in a training split. Under those conditions, the F1 optimization tends to classify nearly all sections in the thin group as Fail rather than identifying a well-separated SCI_8 boundary. This behavior reflects subgroup size limitations rather than a failure of the duo-variable concept.

If a reference rule must be reported from the current CV results, the most interpretable option is to use the stable median-based values from the main parameter clusters rather than the overall mean values. Based on the current results, the reference duo-variable rule can be stated as:

If AC thickness ≤ 16.8 in., classify as Fail when SCI_8 ≥ 0.62 mils.

If AC thickness > 16.8 in., classify as Fail when SCI_8 ≥ 0.42 mils.

These values should be treated as calibration-derived reference values, not universal design thresholds. Their main limitation comes from the small sample size and the uneven distribution of Pass/Fail outcomes within thickness subgroups. The thin-group threshold is especially sensitive to resampling because some training splits contain too few thin sections or limited class variation. In addition, the bimodal AC split distribution indicates that the current dataset supports more than one plausible thickness separation point. Therefore, the duo-variable threshold framework is best interpreted as a thickness-conditioned screening method that should be updated as more labeled TSDD, GPR, and HWTT-labeled data become available.

4.7.4 Cross-Validated Performance

Table 14 summarizes the cross-validated performance of the SCI_8 + AC thickness duo-variable framework across the 50 splits. Values are reported as mean ± SD across 50 splits. The SCI_8 single-variable threshold is included as the reference method because it was the best-performing single-variable rule in Section 4.6.

Table 14. Cross-validated performance of the duo-variable threshold framework compared with the single-variable reference.

Method	F1	Precision	Recall	AUC
Duo-variable, SCI_8 + AC thickness	0.834 ± 0.092	0.842 ± 0.123	0.838 ± 0.106	0.887 ± 0.083
SCI_8 single-variable reference	0.806 ± 0.103	0.846 ± 0.142	0.788 ± 0.130	0.888 ± 0.089

The duo-variable framework improves the mean F1 score from 0.806 to 0.834 while maintaining nearly the same AUC as the SCI_8 single-variable reference. This indicates that adding AC thickness as a conditioning variable improves the final Pass/Fail classification without reducing the overall ranking ability of SCI_8. The precision of the duo rule is slightly lower than the single-variable reference, 0.842 compared with 0.846, but the recall increases from 0.788 to 0.838. This means the duo rule identifies more truly failed sections, with only a minor increase in false failure flags.

Figure 10 shows the split-level distributions of F1, precision, recall, and AUC across the 50 CV splits. The distributions show that the duo-variable method performs consistently across resampled train-test partitions. The mean F1 and recall are above the SCI_8 baseline, while AUC remains nearly identical to the baseline. This result suggests that the thickness-conditioned rule improves failure detection while preserving the strong discrimination capability of SCI_8.

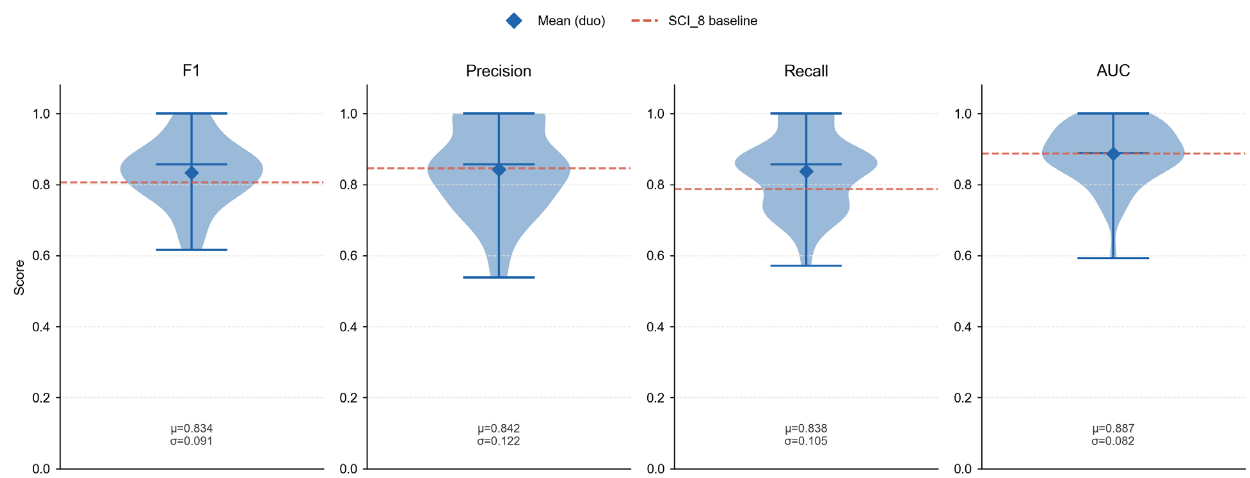

Figure 10. Performance distribution of the SCI_8 + AC thickness duo-variable threshold rule.

The aggregated confusion matrices in Figure 11 further clarify the change in error structure. Compared with the SCI_8 single-variable rule, the duo-variable rule increases true positives from 292 to 310 and reduces false negatives from 78 to 60. This reduction is important because false negatives represent failed sections classified as adequate, which is the highest-cost error in structural screening. The trade-off is a small increase in false positives, from 62 to 64. Overall, the duo-variable rule improves failure detection with only a limited increase in unnecessary failure flags.

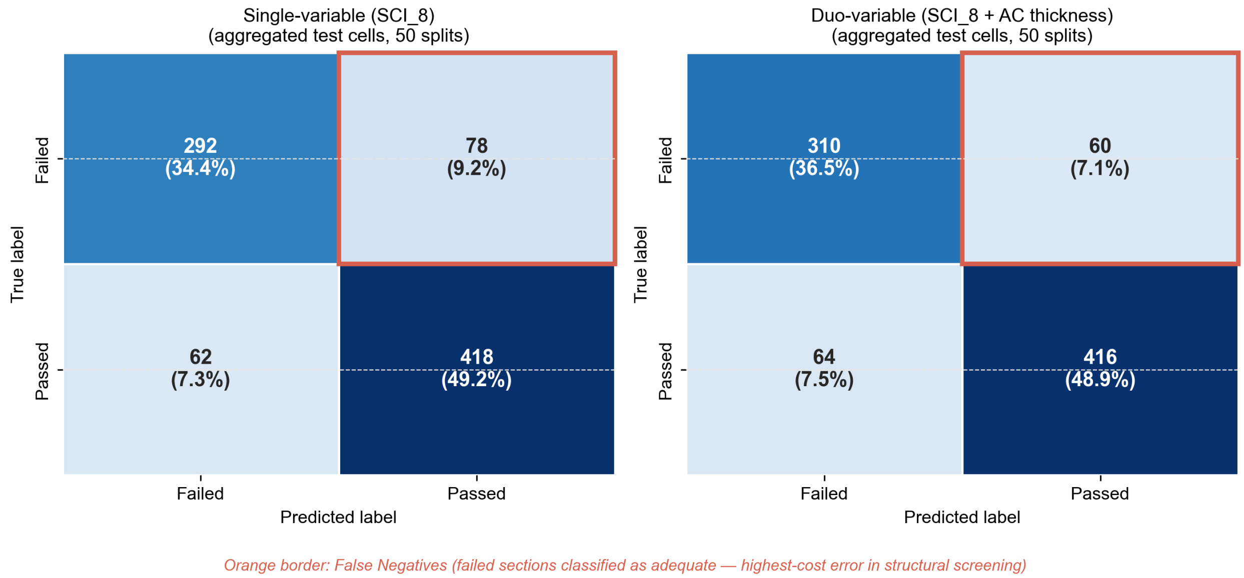

Figure 11. Aggregated confusion matrices for the SCI_8 threshold and SCI_8 + AC thickness duo rule.

Overall, the cross-validated results show that adding AC thickness to the SCI_8 rule improves F1 and recall compared with the global SCI_8 threshold. This improvement is reasonable from an engineering perspective because AC thickness affects how SCI_8 should be interpreted under TSDD loading. However, the duo-variable rule depends on how the data are divided into thinner and thicker AC groups, so its calibrated thresholds are more sensitive to the available sample composition than the single SCI_8 threshold. Therefore, the duo-variable method should be used as a recalibrated screening framework. The AC split point and the two group-specific SCI_8 thresholds should be updated when additional labeled sections become available, rather than treated as universal fixed thresholds.

4.8 Chapter Summary

This chapter evaluated whether TSDD and GPR data can be translated into interpretable threshold-based rules for Georgia-specific structural condition screening. Using the Layer 1 HWTT Pass/Fail outcome as the ground reference, the analysis compared single-variable structural indicators and a thickness-conditioned duo-variable threshold framework. The results show that SCI_8 provides a strong and interpretable candidate single-variable screening rule, while adding AC thickness can improve failure detection within the current dataset when reliable GPR data are available.

- **SCI_8 provided the strongest performance among the evaluated single-variable threshold rules and represents a practical candidate for near-term pilot structural screening.** Among all candidate single-variable indicators, SCI_8 achieved the best overall performance, with a mean F1 score of 0.806 and a mean AUC of 0.888 across 50 repeated stratified CV splits. The preliminary Georgia-specific threshold calibrated from the current dataset was approximately 0.585 mils. This result confirms that shallow deflection-basin curvature is closely related to Layer 1 HWTT performance. Because the SCI_8 rule requires only TSDD data, uses one decision boundary, and remained stable across CV splits, it is a practical near-term pilot screening option for GDOT, pending validation across additional corridors and pavement structures.
- **The SCI_8 + AC thickness duo-variable threshold improves failure detection but should be used as a recalibrated enhanced screening framework.** By conditioning SCI_8 on GPR-derived AC thickness, the method accounts for differences in pavement structural configuration rather than applying one global SCI_8 threshold to all sections. The duo-variable rule achieved the highest F1 score among the threshold methods, with a mean F1 of 0.834 and a mean AUC of 0.887. Compared with the single-variable SCI_8 rule, it improved recall and reduced false negatives while maintaining similar overall discrimination. However, its parameters are more sensitive to subgroup composition than the single SCI_8 threshold,

so the AC thickness split point and group-specific SCI_8 thresholds should be recalibrated as additional labeled data become available.

- **The threshold results establish an interpretable performance baseline for evaluating ML models in Chapter 5.** The single-variable SCI_8 rule provides a simple and stable baseline, while the SCI_8 + AC thickness rule represents a stronger thickness-aware threshold option. Together, these methods define the level of performance that more flexible multivariate ML models should be compared against under the same validation protocol. This comparison is necessary to determine whether ML provides enough added value beyond physically interpretable threshold logic for GDOT implementation.

CHAPTER 5. MACHINE LEARNING-BASED PAVEMENT STRUCTURAL CONDITION EVALUATION

5.1 Overview and Rationale

The threshold-based frameworks developed in Chapter 4 provide transparent and physically interpretable decision rules for pavement structural condition classification. The single-variable SCI_8 rule achieves a mean F1 score of 0.806 and an AUC of 0.888 under repeated stratified CV, demonstrating that a single deterministic curvature indicator captures a substantial portion of the structural failure signal embedded in the Layer 1 HWTT ground reference. The duo-variable thickness-conditioned framework further improves the SCI_8 threshold by incorporating AC thickness as structural context. Under the same CV protocol, it achieves a mean F1 score of 0.834 and an AUC of 0.887. This indicates that a physically interpretable two-variable threshold can outperform the single SCI_8 rule in final Pass/Fail classification, mainly by improving recall and reducing missed structural failures.

These results raise an important question: whether multivariate models can provide a more flexible and scalable structural prediction framework beyond fixed threshold rules. Although the threshold methods perform well, they still classify pavement sections using predetermined decision boundaries based on one or two indicators. In practice, structural failure may depend on the combined effects of shallow curvature response, deflection magnitude, basin shape, and layer thickness. ML classifiers are introduced in this chapter not simply to outperform the best threshold rule, but to evaluate whether multiple predictors can be integrated into a scalable structural prediction engine for future PMS-oriented implementation.

Therefore, this chapter does not position ML as an immediate replacement for the threshold-based screening methods developed in Chapter 4. Instead, it uses the same repeated stratified CV protocol to evaluate whether ML models can provide competitive predictive performance, identify stable multivariate feature patterns, and support a more expandable modeling framework as additional TSDD, GPR, and HWTT-labeled data become available. Four classifiers are evaluated: Logistic Regression, Decision Tree, RF, and Histogram-based Gradient Boosting.

The chapter is organized to follow the logic of model construction before model evaluation. Section 5.2 describes the methodology. Section 5.3 presents the feature selection results, establishing which predictors the models actually rely on before performance outcomes are reported. Section 5.4 presents cross-validated performance results and error structure. Section 5.5 synthesizes the comparative findings across all methods from Chapters 4 and 5 and provides deployment recommendations. Section 5.6 summarizes the chapter.

5.2 Methodology

5.2.1 Classifier Selection

Four classifiers are evaluated to compare whether increasing model complexity improves structural failure prediction beyond the threshold-based methods developed in Chapter 4. Logistic Regression is used as the simplest baseline because it estimates a linear decision boundary and provides interpretable coefficients. Its performance is useful for determining whether the current dataset can be adequately represented by a relatively simple linear relationship among TSDD and GPR predictors. If Logistic Regression performs similarly to more complex models, this suggests that additional nonlinear model structure provides limited benefit at the current dataset scale.

The Decision Tree is included because its rule-based structure is closest to the threshold logic used in Chapter 4. It classifies observations through a sequence of feature splits and therefore provides a useful bridge between deterministic threshold rules and more flexible ML models.

RF extends the Decision Tree by combining predictions from 600 trees. This ensemble structure reduces the instability of a single tree and can capture interactions among curvature indicators, deflection measurements, and layer thickness variables.

Histogram-based Gradient Boosting is included as a second nonlinear ensemble method. It builds trees sequentially and provides a stronger benchmark for testing whether additional nonlinear modeling capacity improves prediction performance. Together, these four classifiers allow the analysis to compare simple linear classification, single-tree rule learning, and nonlinear ensemble learning under the same CV protocol.

5.2.2 Feature Space and Cross-Validation Design

All ML models are trained using the same TSDD and GPR candidate variables introduced in Section 4.3. These variables include SCI-based curvature indicators, including SCI_8, SCI_12, and SCI_SUBGRADE; representative deflection basin measurements, including D0, D-8, D-12, D-18, and D8; and GPR-derived layer thickness variables, including AC thickness and base thickness. Using the same predictor pool as the threshold-based analysis ensures that the comparison between the two method families is based on how the decision boundary is represented, rather than on differences in available input information.

All preprocessing, feature selection, and model training steps are performed only within the training subset of each CV split. The held-out test subset is not used for correlation screening, VIF filtering, RFECV, scaling, model fitting, or model tuning. This design prevents data leakage and avoids overly optimistic performance estimates that would occur if feature selection or preprocessing were applied to the full dataset before splitting. The same Repeated Stratified 5-Fold CV protocol used in Chapter 4 is applied here, with five folds repeated ten times using a fixed random seed of 42. This produces 50 train-test partitions while preserving the pass/fail class proportion within each fold.

5.2.3 Training-Only Feature Selection

A structured three-stage feature selection procedure is applied independently within the training subset of each CV split. The purpose is to identify a compact and low-redundancy predictor set while ensuring that no information from the held-out test subset influences which variables are retained.

In the first stage, pairwise Pearson correlations among candidate features are computed using only the training data. When two features have an absolute correlation greater than 0.90, the feature with the weaker absolute association with the binary HWTT failure outcome is removed. This step reduces redundant predictors while retaining the variable that is more directly related to the structural failure label within each correlated pair.

In the second stage, multicollinearity among the remaining predictors is evaluated using the Variance Inflation Factor (VIF). The feature with the highest VIF is removed whenever the maximum VIF exceeds 10.0. This process is repeated until all retained predictors satisfy the VIF criterion. The first two stages therefore reduce the candidate feature pool to a more stable set of low-redundancy predictors.

In the third stage, Recursive Feature Elimination with Cross-Validation (RFECV) is used to determine the final feature subset within each training split. RFECV is not the final prediction model. Instead, it is a feature selection tool that repeatedly fits a base model, ranks the remaining features, removes the least useful feature, and evaluates the reduced feature set. In this study, class-balanced Logistic Regression is used as the base model for RFECV because it provides a stable and consistent way to rank features across all classifiers. Class-balanced weighting is used to reduce the influence of class imbalance between passed and failed sections.

Within each outer training subset, RFECV performs an internal five-fold stratified CV. For each candidate feature subset size, the training data are further divided into five internal folds while preserving the pass/fail class proportion. The feature subset that produces the highest mean F1 score across these internal folds is selected. A minimum of two retained features is enforced so that the selected subset does not collapse to a single-variable threshold-like solution. After RFECV selects the feature subset using only the outer training data, the corresponding held-out test subset is reduced to the same selected columns for final model evaluation. The test subset is never used to rank, remove, or select features.

5.2.4 Model Configuration

Fixed model configurations are used across all 50 CV splits to keep the comparison focused on model structure rather than hyperparameter tuning. Logistic Regression uses class-balanced weighting with standardized inputs. The Decision Tree is constrained to a maximum depth of 3 and a minimum leaf size of 3. RF uses 600 trees, a minimum leaf size of 2, and class-balanced weighting. Histogram-based Gradient Boosting uses a maximum depth of 3, a learning rate of 0.08, and 400 boosting iterations. All models are initialized with a fixed random seed of 42.

5.3 Feature Selection: What the Models Actually Use

Before examining model performance outcomes, it is useful to characterize the feature space on which the ML classifiers ultimately operate. The three-stage selection procedure described in Section 5.2.3 is designed to identify a stable, low-redundancy subset from the ten-variable candidate pool. Understanding which predictors are consistently retained and which are systematically eliminated provides the physical interpretation framework needed to contextualize the performance results that follow.

5.3.1 Feature Selection Frequency Across 50 Splits

Figure 12 shows the RFECV selection frequency for the variables that were selected at least once across the 50 CV splits. The original candidate pool included ten TSDD and GPR variables. However, several deflection-basin variables were removed during correlation screening or VIF filtering because they were highly redundant with stronger predictors such as D0 and SCI_8. Therefore, only six variables appear in the final selection-frequency summary. A feature is considered consistently predictive when it is selected in more than 50% of splits; sporadic selection below this threshold indicates that a predictor contributes marginal additional information once the dominant features are present.

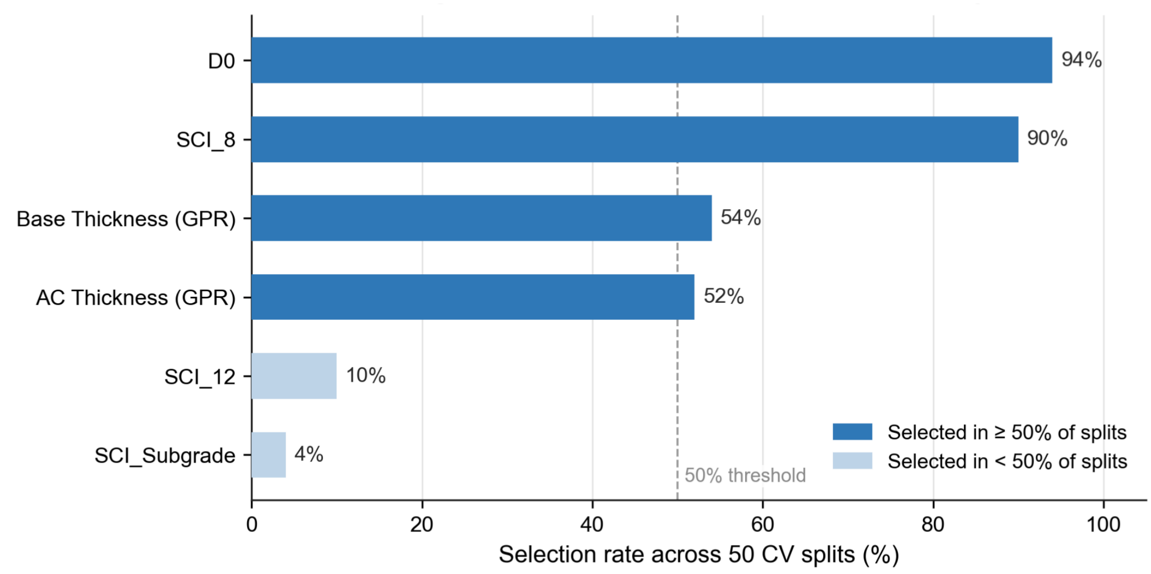

Figure 12. RFECV feature selection frequency across 50 CV splits.

D0 (center deflection at load point) and SCI_8 (the deflection difference between 8-inch and load-center offsets) are selected in more than 90% of splits, establishing them as the dominant structural failure predictors in this dataset. AC thickness and base thickness, derived from GPR measurements, are also selected at moderate-to-high rates, indicating that structural capacity proxies remain influential even when response indicators are present. Deeper SCI indicators, including SCI_12 and SCI_SUBGRADE, exhibit substantially lower selection frequencies. SCI_12 is correlated with SCI_8 at approximately r = 0.99, and the first-stage correlation pruning consistently removes the weaker of the two. SCI_SUBGRADE, which captures the deeper deflection bowl associated with subgrade response, provides limited incremental information once near-surface curvature and thickness are accounted for.

5.3.2 Physical Interpretation of the Dominant Predictors

The dominance of D0 and SCI_8 is physically consistent with the Layer 1 HWTT failure mechanism. The HWTT reflects the rutting resistance and moisture susceptibility of the upper asphalt layer under repeated loading and water exposure. A weakened upper layer produces larger center deflection, captured by D0, and stronger shallow-basin curvature, captured by SCI_8. Together, these two indicators describe both the magnitude and curvature of the TSDD response in the pavement zone most relevant to the Layer 1 HWTT criterion.

The repeated selection of AC thickness and base thickness indicates that pavement layer structure helps explain the TSDD response. A given deflection or SCI value does not have exactly the same meaning for all pavement sections. For example, a thick asphalt section may produce a lower SCI value because the load is distributed through a deeper structural layer, even if some material weakening is present. In contrast, a thinner section may show a stronger curvature response under similar material conditions. By including thickness variables, the ML classifiers can interpret D0 and SCI_8 in relation to pavement structure rather than using one uniform boundary for all sections. This is consistent with the thickness-conditioned threshold logic developed in Chapter 4.

5.3.3 Convergence to a Low-Dimensional Feature Space

Figure 13 presents the distribution of the total number of features selected per split across the 50 repetitions. The distribution is concentrated between two and four features, with a mean of approximately 3.04 and negligible variation. The consistent selection of approximately two to four features indicates that the prediction problem remains low-dimensional. In other words, the ML models do not rely on a large set of predictors to achieve their performance. Instead, they mainly use the same core structural indicators identified in Chapter 4, especially D0, SCI_8, and layer thickness. Therefore, the modest performance improvement from ML is likely due to a more flexible way of combining these variables, rather than the use of many additional predictors. The stability of the selected feature count across the 50 splits also suggests that these core predictors are not artifacts of a particular data split, but represent a consistent structural failure signal in the Georgia interstate dataset.

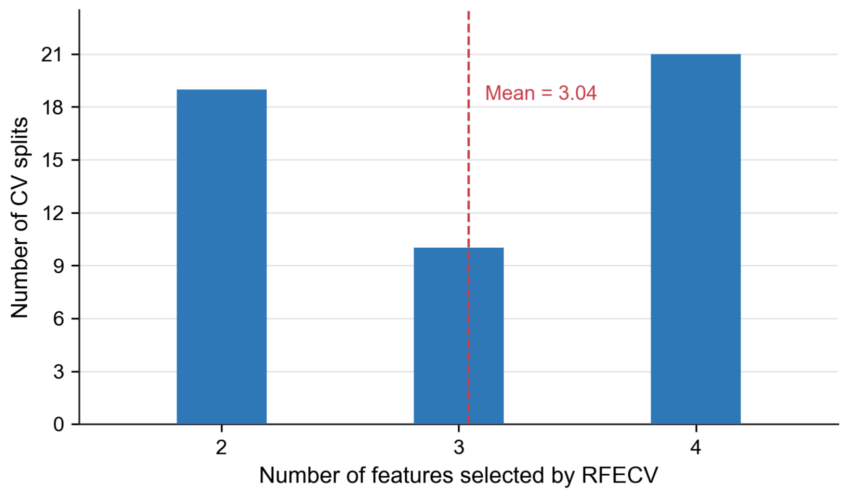

Figure 13. Distribution of the number of features selected by RFECV across 50 CV splits.

Based on the feature selection results, D0, SCI_8, and AC thickness form the most defensible compact feature set for ML deployment. D0 captures the overall deflection magnitude, SCI_8 captures shallow curvature response associated with upper-layer weakness, and AC thickness accounts for structural configuration effects on the deflection response. This three-feature set therefore preserves the dominant response and structural-context information while avoiding unnecessary redundancy from highly correlated deflection-basin variables.

5.4 Model Performance and Error Analysis

5.4.1 Overall Predictive Performance

Table 15 summarizes the cross-validated performance of the four ML classifiers across the 50 repeated stratified train-test splits. Values are reported as mean ± SD across 50 splits. Models are ordered by mean F1 score. The purpose of this section is to compare the behavior of the ML models under the same feature-selection and validation protocol, rather than to compare them with the threshold-based methods from Chapter 4. The broader comparison between threshold methods and ML classifiers is presented separately in Section 5.5.

Among the four classifiers, RF achieves the strongest overall performance, with the highest mean F1 score of 0.822 and the highest mean AUC of 0.891. Logistic Regression follows closely, with a mean F1 score of 0.817 and a mean AUC of 0.882. Histogram-based Gradient Boosting and Decision Tree produce lower but still comparable results, with mean F1 scores of 0.793 and 0.787, respectively.

The performance differences among the four ML models are modest. The F1 gap between RF and Decision Tree is approximately 0.035, and the AUC gap between RF and Logistic Regression is only 0.009. This indicates that increasing model complexity does not lead to a large performance gain under the current dataset scale. The competitive performance of Logistic Regression also suggests that the structural Pass/Fail boundary in the current dataset can be approximated reasonably well using a compact and relatively simple predictor space.

Overall, the ML results show that RF provides the best balance of classification accuracy and ranking ability among the evaluated models. However, the narrow performance spread also indicates that the current prediction problem is primarily controlled by a small number of dominant structural indicators, rather than by complex high-dimensional interactions.

Table 15. Cross-validated performance statistics for the four ML classifiers.

Model	F1	Precision	Recall	AUC
RF	0.822 ± 0.105	0.860 ± 0.126	0.802 ± 0.132	0.891 ± 0.077
Logistic Regression	0.817 ± 0.113	0.901 ± 0.101	0.761 ± 0.148	0.882 ± 0.092
Histogram-based Gradient Boosting	0.793 ± 0.088	0.843 ± 0.114	0.768 ± 0.137	0.862 ± 0.092
Decision Tree	0.787 ± 0.124	0.799 ± 0.164	0.791 ± 0.129	0.847 ± 0.103

5.4.2 Precision, Recall, and Model Behavior

The four classifiers show different balances between precision and recall, which is important for interpreting their operational behavior. In structural condition screening, precision represents the reliability of predicted failures, while recall represents the ability to identify truly failed sections. A model with high precision produces fewer false alarms, while a model with high recall misses fewer structurally deficient sections.

Logistic Regression achieves the highest mean precision, at 0.901, but also has the lowest recall, at 0.761. This indicates a conservative classification pattern. When Logistic Regression classifies a section as failed, that prediction is usually reliable; however, the model is more likely to miss some truly failed sections. This behavior is consistent with a relatively simple linear decision boundary that identifies failure only when the combined structural signal is strong.

RF provides the most balanced precision-recall profile, with a mean precision of 0.860 and a mean recall of 0.802. This balance explains why it achieves the highest mean F1 score among the four models. The ensemble structure allows RF to combine curvature response, deflection magnitude, and layer thickness variables more flexibly than a single linear classifier, while still avoiding the instability of a single decision tree.

Decision Tree shows a different pattern. Its recall is relatively high, at 0.791, but its precision is the lowest among the four models, at 0.799. This suggests that the Decision Tree is more willing to flag sections as failed, but this comes with a higher number of false positives. Histogram-based Gradient Boosting falls between these patterns, with moderate precision and recall but lower overall F1 and AUC than RF.

These results show that models with similar F1 scores can still have different practical implications. For structural screening, a model that minimizes missed failures while keeping false positives controlled is generally preferred. Based on this criterion, RF provides the most suitable balance among the four ML classifiers.

5.4.3 Confusion Matrix and Error Analysis

Figure 14 and Figure 15 provide a more direct view of model error behavior. Figure 14 presents the aggregated confusion matrix for RF, where TP, TN, FP, and FN counts are summed across all 50 held-out test partitions rather than reported for a single split. In this matrix, the diagonal cells represent correct classifications, including true positives, or correctly identified failed sections, and true negatives, or correctly identified passed sections. The off-diagonal cells represent errors, including false positives, where passed sections are conservatively flagged as failed, and false negatives, where failed sections are incorrectly classified as structurally adequate. Figure 15 compares the mean confusion matrix components across the four classifiers, with the false-negative group highlighted because missed failures represent structurally deficient sections that would be classified as adequate and therefore left without targeted intervention. This error type is more consequential than false positives, which mainly lead to conservative over-screening or additional inspection.

For RF, the aggregated results show 297 true positives and 427 true negatives, indicating that the model correctly identifies both failed and passed sections at a relatively high rate. The model produces 73 false negatives, where failed sections are classified as structurally adequate, and 53 false positives, where passed sections are conservatively flagged as failed. In a structural screening context, these two error types are not equivalent. A false negative allows a vulnerable pavement section to remain unidentified, while a false positive mainly leads to additional inspection or monitoring.

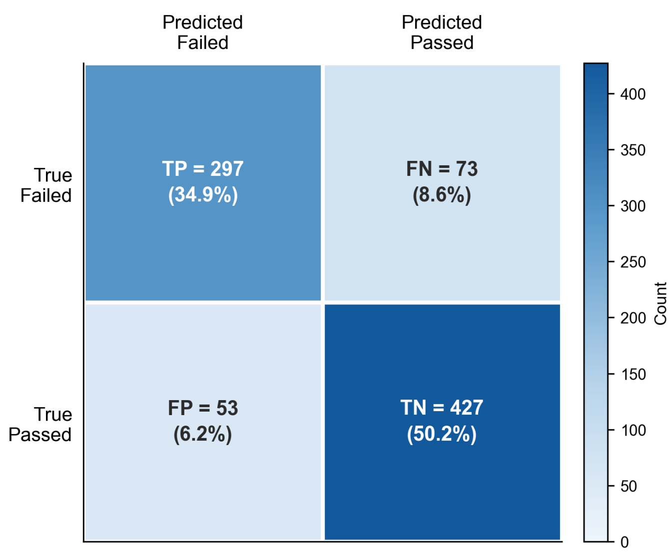

Figure 14. Aggregated confusion matrix for the RF classifier across all 50 CV test partitions.

The cross-model comparison in Figure 15 shows that RF provides the most favorable balance between missed failures and false alarms. Logistic Regression maintains the lowest false-positive count, consistent with its conservative precision profile, but this comes with a greater tendency to miss failed sections. Decision Tree achieves relatively high sensitivity but shows larger variability in both false-positive and false-negative counts, reflecting the instability of a single-tree structure under repeated resampling. Histogram-based Gradient Boosting shows intermediate error behavior. Overall, the error analysis supports the use of RF as the preferred ML classifier when missed failures are treated as the more consequential screening error.

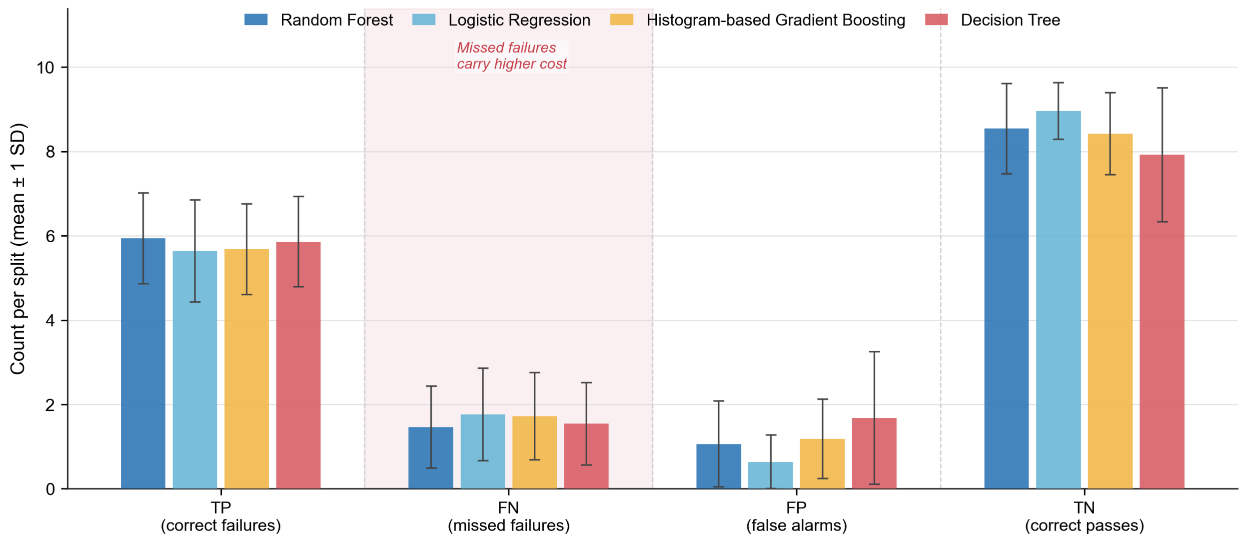

Figure 15. Mean confusion matrix components per CV split for all four ML classifiers.

Among the evaluated ML classifiers, RF is the preferred ML option because it achieves the highest aggregate F1 and AUC while maintaining the most favorable balance between missed failures and false alarms. This makes it more suitable than Logistic Regression when false negatives are treated as the more consequential screening error.

5.5 Comparative Evaluation and Deployment Recommendation

5.5.1 All-Method Performance Comparison

Table 16 presents a unified comparison of the threshold-based methods developed in Chapter 4 and the ML classifiers evaluated in Chapter 5. All methods were evaluated under the same repeated stratified CV protocol, which allows their performance to be compared on a consistent basis. Values are reported as mean ± SD across 50 splits. The comparison considers both predictive performance and practical implementation characteristics, including interpretability and deployment readiness.

Table 16. Unified performance comparison across all methods evaluated in Chapters 4 and 5.

Method	**Type**	**F1**	**AUC**	**Interpretability**	**Deployment readiness**
Duo-variable, SCI_8 + AC thickness	Threshold	0.834 ± 0.092	0.887 ± 0.083	High	Medium / High
RF	ML	0.822 ± 0.105	0.891 ± 0.077	Low	Medium
Logistic Regression	ML	0.817 ± 0.113	0.882 ± 0.092	Medium	Medium
SCI_8 single-variable threshold	Threshold	0.806 ± 0.103	0.888 ± 0.089	High	High
Histogram-based Gradient Boosting	ML	0.793 ± 0.088	0.862 ± 0.092	Low	Medium
Decision Tree	ML	0.787 ± 0.124	0.847 ± 0.103	Medium	Medium

The all-method comparison shows that the SCI_8 + AC thickness duo-variable threshold achieves the highest mean F1 score among all evaluated methods, while RF achieves the highest mean AUC. This distinction is important because F1 and AUC describe different aspects of model performance. The duo-variable rule provides the strongest final binary Pass/Fail classification under the selected operating threshold, mainly because it improves recall and reduces missed structural failures. RF, by contrast, provides the highest mean AUC among the evaluated methods, although the difference from the threshold methods is small. This suggests that RF can rank failed and passing sections effectively across possible decision thresholds, but the current dataset does not show a large ML advantage over the best threshold-based method.

The strong performance of the duo-variable rule is reasonable from an engineering perspective. SCI_8 captures shallow deflection-basin curvature, which is closely related to upper asphalt layer weakness, while AC thickness provides the structural context needed to interpret that curvature response. A single SCI_8 value may not carry the same structural meaning in thin and thick asphalt sections. By conditioning the SCI_8 threshold on AC thickness, the duo-variable rule directly encodes this physical relationship and improves the final classification of failed sections.

At the same time, the duo-variable result should be interpreted together with its calibration limitations. The method requires reliable GPR-derived AC thickness data, and its calibrated parameters are more sensitive to subgroup composition than the single SCI_8 threshold. As shown in Chapter 4, the AC thickness split point has a bimodal distribution, and the thin-group SCI_8 threshold is less stable because some CV splits produce boundary solutions. Therefore, although the duo-variable threshold achieves the highest F1 score in the current dataset, it should be implemented as a recalibrated screening framework rather than as a universal fixed-threshold rule.

The ML results should therefore not be interpreted as evidence that ML clearly outperforms threshold-based screening under the current dataset scale. RF achieves the best performance among the ML classifiers, but its mean F1 score remains slightly lower than the duo-variable threshold. This outcome is expected because the current dataset contains a strong dominant signal from SCI_8 and AC thickness, and the feature selection results in Section 5.3 show that the prediction problem converges to a compact set of dominant variables, especially SCI_8, D0, and layer thickness. Under this data condition, complex nonlinear models have limited room to improve binary classification performance.

However, RF still provides practical value as a forward-looking modeling option. It achieves the highest AUC, provides the strongest ML-based performance, and can combine multiple response and layer-configuration variables within a retrainable framework. This flexibility is not fully reflected in the current 85-section dataset, but it becomes increasingly important as additional TSDD, GPR, and HWTT-labeled data are collected across more corridors, pavement designs, material conditions, and structural configurations.

Overall, the comparison shows that threshold-based methods and ML models serve different implementation purposes. The threshold methods provide transparent and auditable screening rules for near-term use, with the duo-variable rule offering the strongest current F1 performance when reliable GPR thickness data are available. RF provides a scalable ML-based prediction engine for future PMS integration, especially as the labeled dataset expands and structural diversity increases. Therefore, ML should be viewed as a long-term extension of the threshold framework rather than an immediate replacement for it.

5.5.2 Deployment Trade-Offs and Recommended Hierarchy

The deployment implications should be interpreted as a staged hierarchy rather than a single-method recommendation. The results show that no one method is optimal for all implementation stages. Instead, each method has a different role depending on data availability, transparency requirements, and the maturity of the agency's structural screening workflow.

For near-term pilot implementation, the SCI_8 single-variable threshold may be used as a candidate baseline for network-level structural screening. It provides strong predictive performance, requires only TSDD-derived SCI_8, and uses a single calibrated decision boundary that can be easily communicated to pavement engineers. Its threshold is also relatively stable across CV splits, which makes it suitable for pilot testing within existing PMS workflows. Although it does not achieve the highest F1 score among all evaluated methods, it remains the simplest and most defensible starting point for routine GDOT structural screening.

When reliable GPR-derived AC thickness data are available, the SCI_8 + AC thickness duo-variable threshold should be used as an enhanced threshold-based screening framework. It achieves the highest mean F1 score among all evaluated methods and improves failure detection by accounting for pavement structural configuration. This makes it especially useful for corridors where asphalt thickness varies substantially and a single SCI_8 threshold may overgeneralize the relationship between TSDD response and structural condition. However, because its calibrated parameters are more sensitive to subgroup composition, the AC thickness split point and the two group-specific SCI_8 thresholds should be recalibrated when additional labeled sections become available. Therefore, the duo-variable method is best viewed as a recalibrated screening framework rather than a fixed universal rule.

For ML-based implementation, RF should be retained as the preferred structural prediction engine for future PMS integration. Although its mean F1 score is slightly lower than the duo-variable threshold in the current evaluation, it achieves the highest AUC and provides the strongest performance among the ML classifiers. More importantly, it can combine multiple structural response and layer-configuration variables and can be retrained as GDOT accumulates additional TSDD, GPR, and HWTT-labeled data. This makes RF more suitable for long-term implementation across structurally diverse networks than a fixed threshold rule.

Therefore, Chapter 6 uses RF not because it outperforms the updated duo-variable threshold in the current dataset, but because it represents the preferred forward-looking ML model for integrated treatment decision making. The threshold rules remain the most practical near-term screening tools, while RF demonstrates how a scalable ML-based structural prediction engine can be incorporated into future PMS-oriented workflows.

5.6 Chapter Summary

This chapter evaluated whether ML-based classifiers can provide a competitive and scalable alternative to the threshold-based structural screening methods developed in Chapter 4. The purpose was not to replace the threshold methods immediately, but to determine whether multivariate ML models can improve structural prediction, identify stable predictor patterns, and support future PMS-oriented implementation as GDOT's labeled dataset expands.

- **The ML models rely on a small group of physically meaningful predictors rather than a high-dimensional feature set.** Feature selection results show that D0 and SCI_8 were selected in more than 90% of the CV splits, confirming that near-load deflection magnitude and shallow deflection-basin curvature carry the dominant structural failure signal. AC thickness and base thickness were also frequently retained, indicating that pavement layer geometry helps interpret TSDD response. The selected feature space converged to approximately three variables per split, suggesting that the prediction problem is low-dimensional and controlled by a compact set of response and thickness indicators.
- **RF provided the strongest ML-based performance, but the performance gap among ML classifiers was modest.** Among the four evaluated classifiers, RF achieved the highest ML performance, with a mean F1 score of 0.822 and a mean AUC of 0.891. Logistic Regression performed closely behind, while Histogram-based Gradient Boosting and Decision Tree showed slightly lower performance. The relatively small performance differences indicate that

increasing model complexity provides limited improvement at the current dataset scale. This finding is consistent with the feature selection results, which show that most of the predictive signal is concentrated in a few dominant structural indicators.

- **The comparison with threshold-based methods supports complementary deployment roles rather than a single superior method.** RF slightly outperformed the single-variable SCI_8 threshold in F1 and achieved the highest AUC among all evaluated methods, while the SCI_8 + AC thickness duo-variable threshold achieved the highest mean F1 score. This means that threshold and ML methods each provide different strengths: the duo-variable threshold gives the strongest current F1-based screening performance, while RF provides the strongest ML-based ranking and prediction framework. However, the duo-variable rule is more sensitive to subgroup composition and should be treated as a recalibrated enhanced screening framework rather than a universal fixed-threshold rule.
- **The recommended deployment hierarchy should reflect both current performance and operational stability.** The SCI_8 single-variable threshold is the most practical immediate screening tool because it is simple, stable, interpretable, and requires only TSDD data. The SCI_8 + AC thickness duo-variable threshold is the preferred enhanced threshold option when reliable GPR thickness data are available because it improves failure detection and provides the highest current F1 score, but it should be recalibrated as additional labeled data become available. RF should be retained as the preferred future retrainable ML-based structural prediction engine, rather than as the current best screening rule, because it can combine multiple predictors, be retrained with additional labeled data, and support scalable corridor-level screening workflows after further validation.

- **Chapter 6 uses RF because it represents the most suitable forward-looking retrainable ML model for demonstrating integrated treatment screening.** RF does not exceed the updated duo-variable threshold in current F1 performance, but it provides a retrainable multivariate framework that is better suited for future expansion. Therefore, the threshold rules remain the near-term operational tools, while RF demonstrates how structural prediction can evolve toward a scalable PMS-oriented workflow as GDOT collects additional TSDD, GPR, and HWTT-labeled data across more corridors and pavement structures.

CHAPTER 6. TREATMENT DECISION FRAMEWORK

This chapter presents an integrated treatment decision framework for converting predicted pavement condition into network-level M&R prioritization recommendations. The framework combines two complementary condition dimensions: structural condition predicted from TSDD and GPR data, and functional condition derived from 3D pavement surface indicators. The paired condition outputs are then translated into four candidate treatment screening categories and visualized as corridor-level screening maps.

The chapter is organized as follows. Section 6.1 introduces the framework inputs, classification logic, and treatment decision matrix. Section 6.2 applies the framework to a representative I-285 segment to illustrate the transition from condition prediction to treatment recommendation. Section 6.3 presents corridor-level treatment decision maps for I-285, I-575, and I-59. Section 6.4 summarizes the chapter.

6.1 Purpose and Decision Logic of the Framework

The purpose of the treatment decision framework is to connect structural condition with functional condition. The need for this integration arises because structural and functional deterioration do not always occur simultaneously. A section with acceptable surface condition may still exhibit structural vulnerability, especially when the pavement has experienced repeated resurfacing or when structural weakening has not yet appeared as visible cracking or roughness. In contrast, a section with poor surface condition may still retain adequate structural capacity, in which case a surface renewal treatment may be more appropriate than structural rehabilitation. A surface-only treatment selection process cannot distinguish these cases.

Figure 16 summarizes the workflow of the integrated framework. The structural branch uses the RF model developed in Chapter 5. The selected structural predictors include TSDD-derived D0, SCI_8, and GPR-derived AC thickness. These variables represent complementary aspects of pavement response and structural configuration. D0 captures the near-load deflection magnitude, SCI_8 captures shallow curvature response associated with the upper asphalt layers, and AC thickness provides pavement structure information needed to interpret the measured response. The RF model classifies each evaluated interval into a binary structural condition label: Good or Poor.

The functional branch uses 3D pavement surface data to classify surface condition. In this framework, functional condition is determined using surface distress indicators, including percent total cracking and IRI. Rule-based thresholds are applied to assign a binary functional condition label. For demonstration purposes, a section is classified as functionally Poor when percent total cracking exceeds 30 percent or IRI exceeds 170 in./mi; otherwise, it is classified as functionally Good. These thresholds are used as transparent screening thresholds for illustrating the integrated framework and should be reviewed against GDOT's functional condition criteria before operational use. This rule-based functional classification is intentionally transparent and consistent with the screening-oriented purpose of the framework.

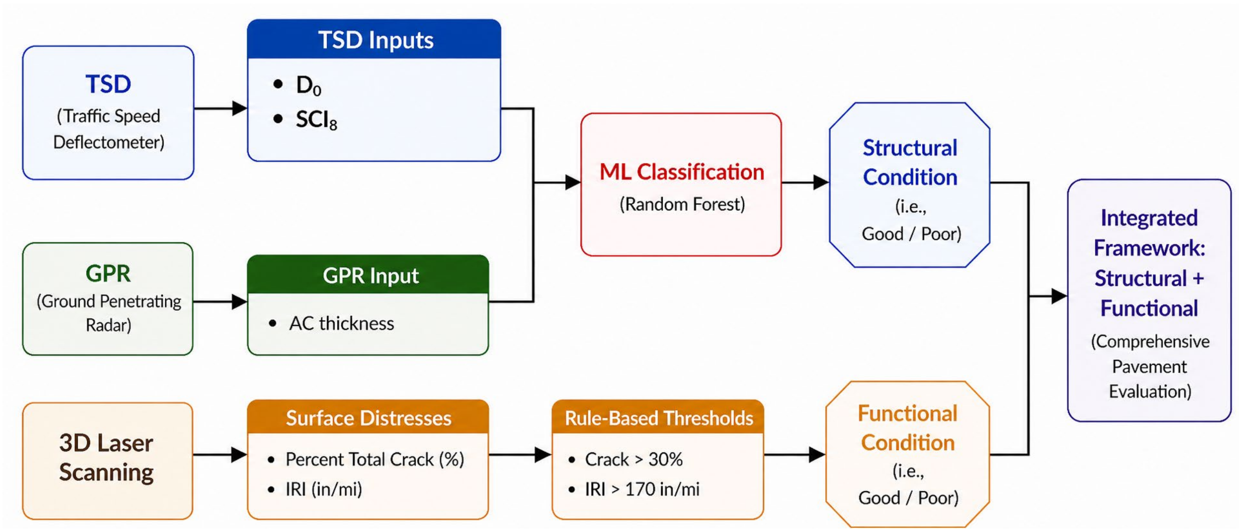

Figure 16. Workflow of the integrated structural–functional pavement evaluation framework.

After structural and functional condition labels are generated, the two condition streams are organized along the roadway referencing system using route, direction, lane, wheel path, and milepost information. Each condition stream is then spatially consolidated to reduce isolated point-level fluctuations and produce a more interpretable condition profile. The clustered structural and functional profiles are overlaid using their combined change points, so that each treatment decision segment has a consistent structural and functional label. The paired binary labels are then converted into treatment recommendations using the decision matrix shown in Table 17. For corridor-level visualization, the resulting treatment labels are further aggregated into continuous treatment segments to improve interpretability and support network-level screening.

Table 17. Treatment decision matrix based on structural and functional condition.

Functional Condition	Structural Condition	Treatment Decision
Good	Good	No Action
Poor	Good	Milling and Resurfacing
Good	Poor	Close Monitoring
Poor	Poor	Major Rehabilitation

No Action corresponds to sections with acceptable predicted structural and functional condition. Milling and Resurfacing corresponds to sections with poor functional condition but adequate predicted structural condition, where deterioration is primarily surface-related and a surface renewal treatment may be sufficient after project-level confirmation. Close Monitoring corresponds to sections with poor predicted structural condition but acceptable functional condition; these sections warrant additional structural verification, closer performance tracking, or consideration in future rehabilitation programming. Major Rehabilitation corresponds to sections where both structural and functional conditions are poor and therefore represents the highest-priority category for detailed engineering investigation and structural treatment evaluation.

This framework provides a direct link between sensing-based pavement evaluation and treatment-oriented network screening. The following section demonstrates the application of this workflow on a representative I-285 segment.

6.2 Representative Application on I-285

This section demonstrates how the framework in Figure 16 and the treatment matrix in Table 17 are applied to a roadway segment. A half-mile segment on I-285 EB, from MP 30.16 to MP 30.66, was selected as the representative example because it contains changes in predicted structural condition within a short distance while the surface condition remains mostly acceptable. This makes the segment useful for illustrating how the framework identifies structural concerns that may not be evident from surface condition alone.

6.2.1 Spatial Clustering

Figure 17 shows the structural condition prediction for the selected I-285 segment. The upper panel presents the interval-level predictions, where each evaluated location is classified as structurally Good or Poor. The lower panel converts these point-level predictions into a clustered structural condition profile.

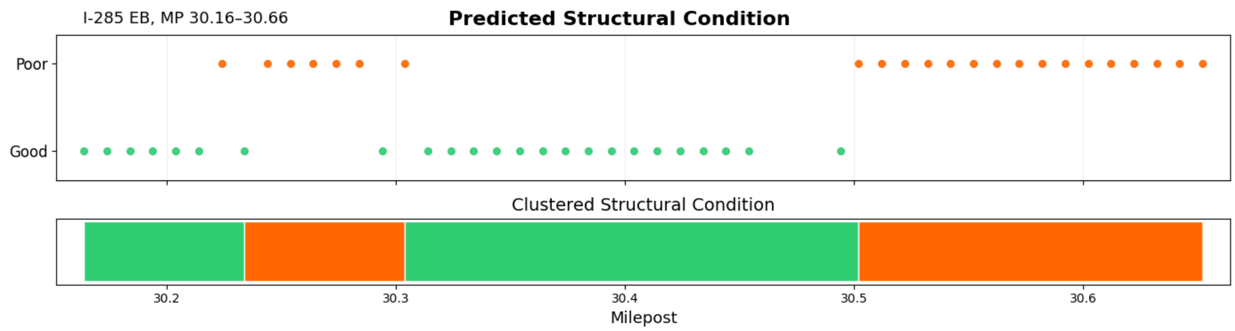

Figure 17. Predicted and clustered structural condition for I-285 EB, MP 30.16–30.66.

The interval-level predictions are first consolidated using a local majority rule. For each short roadway interval, the assigned structural condition is determined by the dominant Good or Poor prediction within its surrounding neighborhood. This majority-based consolidation reduces isolated label changes and produces a more stable structural condition profile for treatment screening. As shown in Figure 17, the consolidated structural profile divides the selected I-285 segment into four condition regions: Good, Poor, Good, and Poor. The result indicates localized structural variation within the half-mile window, rather than a uniform structural condition across the entire segment. The same clustering principle is used in the integrated profile in Section 6.2.2, where the structural and functional condition streams are first consolidated separately before being overlaid for treatment assignment.

6.2.2 Integrated structural–functional treatment decision

Figure 18 presents the integrated treatment decision profile for the same I-285 segment. The first bar shows the clustered structural condition, the second bar shows the clustered functional condition derived from 3D pavement surface data, and the third bar shows the final treatment decision. In this visualization, the structural and functional profiles are clustered independently so that each condition stream preserves its own spatial pattern. The treatment decision profile is then generated by overlaying the two clustered profiles and applying the decision matrix in Table 17. The gray vertical guide lines mark the locations where the final treatment decision changes.

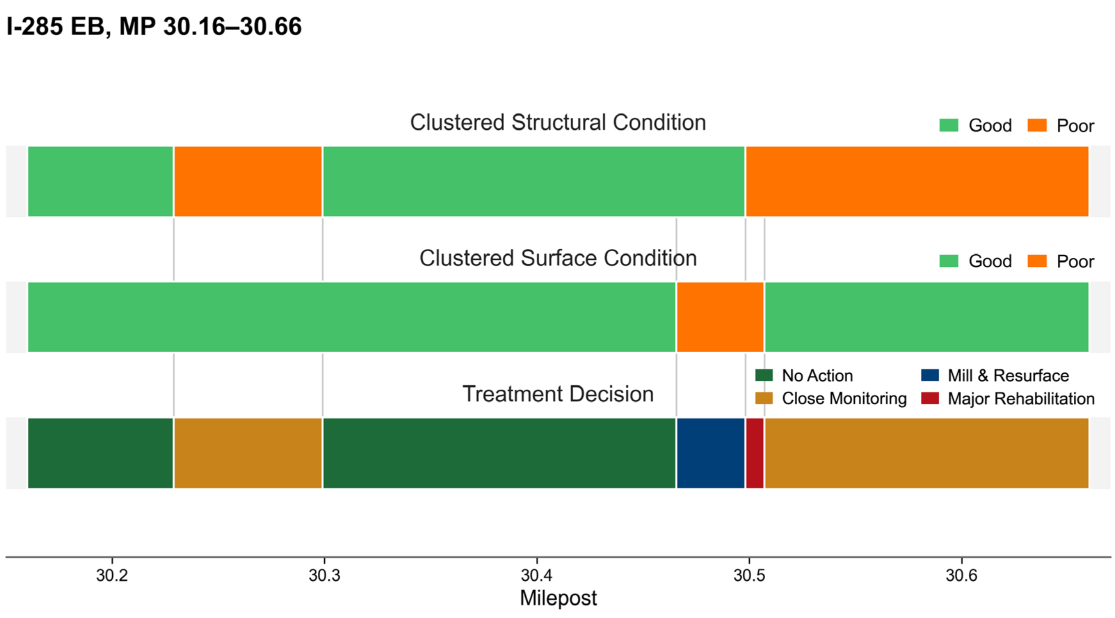

Figure 18. Integrated structural–functional treatment decision profile for I-285 EB, MP 30.16–30.66.

The structural and functional profiles show different spatial patterns within the selected half-mile segment. Structural condition alternates between Good and Poor, while functional condition remains mostly Good except for a localized Poor section near the middle of the segment. This mismatch demonstrates why structural and functional condition should be retained as separate inputs before treatment assignment. Where both structural and functional conditions are Good, the segment is assigned to No Action. Where structural condition is Poor but functional condition remains Good, the segment is assigned to Close Monitoring. These locations are important because they would likely appear acceptable under a surface-only workflow but are flagged by the structural prediction.

The localized functional Poor section produces additional treatment categories. Where the functional condition is Poor but structural condition remains Good, the decision becomes Milling and Resurfacing, indicating a surface-driven treatment need. Where both structural and functional conditions are Poor, the section is flagged for Major Rehabilitation review, indicating a more severe condition state requiring project-level structural verification. Overall, the I-285 example shows that the final treatment profile is not simply a restatement of either the structural or functional profile alone. Instead, it reflects the spatial interaction between the two condition dimensions and identifies different treatment needs within a short roadway distance.

6.3 Corridor-Level Treatment Mapping and Results

The representative example in Section 6.2 illustrates the interval-level overlay logic used to convert clustered structural and functional condition profiles into local treatment decisions. For corridor-level application, an additional aggregation step is needed because point-level or short-window clustered outputs remain too detailed for network-level engineering screening. Therefore, the interval-level treatment decisions are aggregated into 0.10-mile bins. Within each bin, the final treatment category is determined by majority vote among the interval-level treatment labels. The representative location of each bin is defined by the median milepost and median geographic coordinates of the records within that bin. This aggregation produces a corridor-scale decision unit that is more appropriate for network-level screening.

After bin-level aggregation, short isolated treatment segments are cleaned to reduce map fragmentation. If a short segment is located between two adjacent segments with the same treatment category and is shorter than the specified minimum length, it is reassigned to the surrounding treatment category. In this study, a primary threshold of 0.08 mile and a secondary threshold of 0.05 mile are used to remove very short isolated segments while preserving sustained changes in treatment category. Consecutive bins with the same treatment category are then merged into continuous line segments for Geographic Information System (GIS) visualization.

6.3.1 I-285

Figure 19 presents the corridor-level treatment decision map for I-285 in the westbound (WB) direction. The mapped section shows a clear sequence of treatment categories along the corridor rather than a single dominant recommendation. The western portion of the displayed route is primarily classified as No Action, indicating that both structural and functional condition are acceptable over that section. Moving eastward, the map shows a short Milling and Resurfacing segment near the northern part of the alignment, followed by Close Monitoring through the curved portion of the corridor. A localized Rehabilitation segment appears along the southeast-trending portion of the route, after which the corridor returns to No Action near the eastern end.

The I-285 result is consistent with the role of this corridor as a heterogeneous urban beltway. The treatment categories change over relatively short distances, which suggests that a uniform corridor-level treatment assumption would not be appropriate. For GDOT, this map can be used to separate sections that may only require routine review from sections that should be checked with additional structural information before treatment programming. In particular, the Close Monitoring and Rehabilitation segments would be reasonable candidates for follow-up review using project-level field inspection, coring, or deflection testing before final treatment selection.

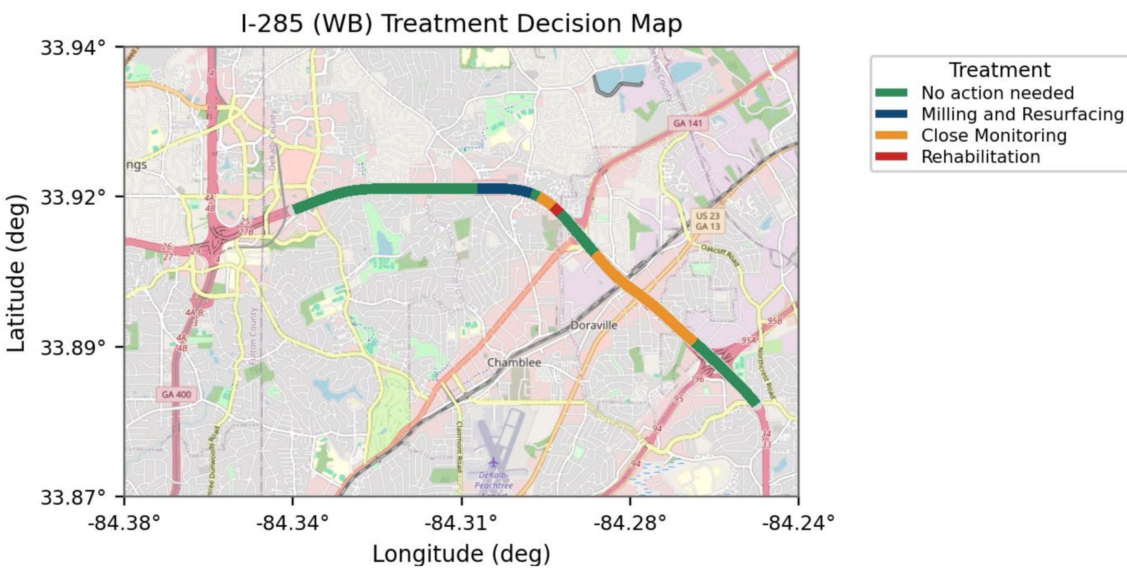

Figure 19. Corridor-level treatment decision map for I-285.

6.3.2 I-575

Figure 20 presents the corridor-level treatment decision map for I-575. Compared with I-285 and I-59, I-575 shows a more favorable treatment pattern. Large portions of the corridor are assigned to No Action, reflecting acceptable predicted structural and functional condition.

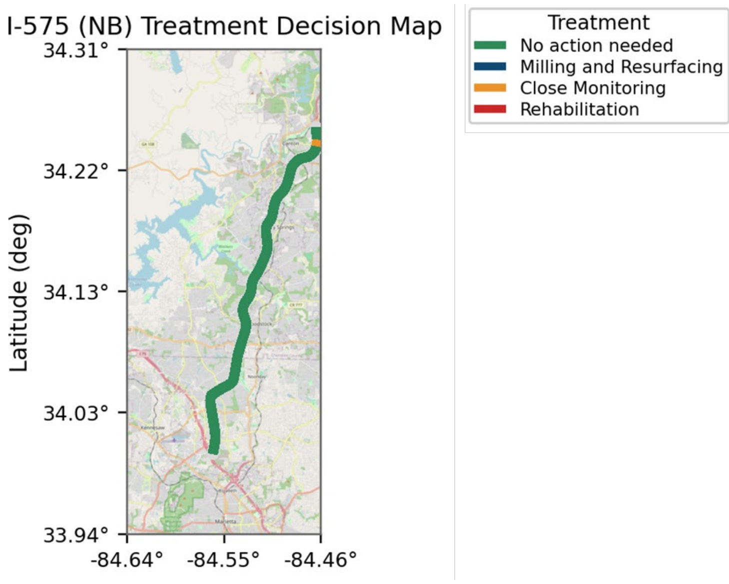

Figure 20. Corridor-level treatment decision map for I-575.

The I-575 map is consistent with the earlier data characterization, where this corridor showed favorable Layer 1 HWTT outcomes and relatively low surface distress levels. However, the map still identifies localized sections assigned to other treatment categories. These exceptions are important because they show that the framework preserves spatial variation even on a generally good corridor. For network-level use, the I-575 result can support confirmation of overall corridor condition while still flagging locations that may require additional review.

6.3.3 I-59

Figure 21 presents the corridor-level treatment decision map for I-59. In contrast to I-575, I-59 shows a more severe treatment pattern, with a higher concentration of sections assigned to Close Monitoring and Rehabilitation.

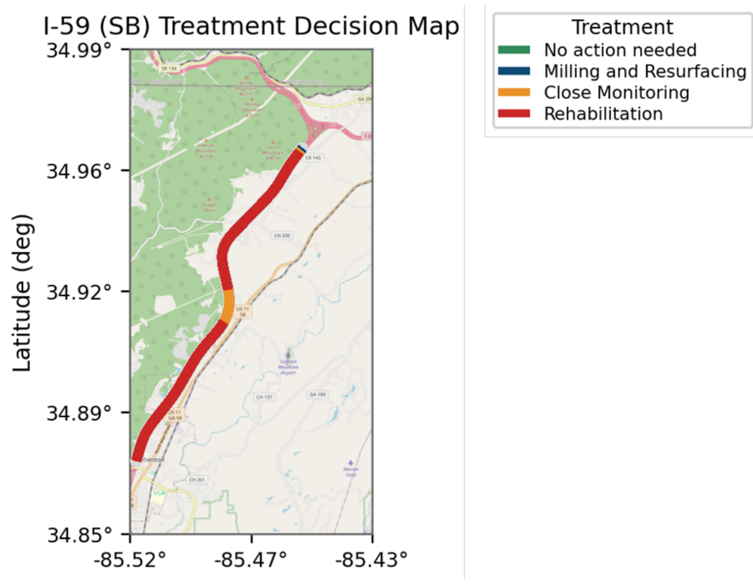

Figure 21. Corridor-level treatment decision map for I-59.

This result is consistent with the structural and laboratory characterization presented earlier in the report. I-59 had the weakest Layer 1 HWTT performance among the retained corridors and showed higher TSDD curvature and deflection responses. The treatment map reflects this condition pattern by assigning more intensive screening categories along the corridor. Sections flagged for Major Rehabilitation review indicate locations where both structural and functional conditions are poor, while Close Monitoring sections indicate structurally poor locations that may not be fully identified by surface condition alone.

Across the three corridors, the treatment maps demonstrate that the framework can distinguish different pavement condition patterns. I-575 is dominated by lower-intervention categories, I-285 shows the greatest spatial variability, and I-59 contains more intensive treatment categories. These differences indicate that the framework preserves spatially varying treatment needs and provides a practical basis for network-level screening and prioritization.

6.4 Chapter Summary

This chapter developed an integrated treatment decision framework that connects structural prediction and surface condition evaluation with network-level M&R screening. The framework is intended to support PMS-oriented decision-making by translating separate structural and functional condition labels into clear treatment categories that can be mapped and reviewed at the corridor scale.

- **The framework converts paired structural and functional condition labels into four treatment-oriented screening categories.** Structural condition is predicted from TSDD and GPR data, while functional condition is derived from 3D pavement surface indicators. These two condition dimensions are combined through a simple decision matrix to assign No Action, Milling and Resurfacing, Close Monitoring, or Major Rehabilitation. This structure allows the final recommendation to reflect both load-carrying adequacy and observable surface performance rather than relying on surface condition alone.
- **The Close Monitoring category is the main added value of the integrated framework because it identifies structurally weak sections that may still look acceptable at the surface.** These sections have acceptable functional condition but poor predicted structural condition. Under a surface-only PMS workflow, they may be classified as No Action because visible distress has not yet become severe. By flagging them separately, the framework provides GDOT with an early warning category for sections that may require structural verification, closer performance tracking, or inclusion in future rehabilitation programming.
- **The representative I-285 example shows that the framework can distinguish treatment needs within a short corridor segment.** In the selected segment, No Action and Close

Monitoring dominate, while short localized portions are assigned to Milling and Resurfacing or Major Rehabilitation where functional deterioration overlaps with the structural profile. This result demonstrates that the framework can separate surface-driven needs, structure-driven monitoring needs, and combined deterioration within the same local segment.

- **The corridor-level maps show that the framework preserves distinct treatment patterns across the three study routes.** I-285 exhibits spatially variable treatment needs, consistent with its heterogeneous urban corridor characteristics. I-575 is dominated by lower-intervention categories, reflecting generally favorable condition. I-59 shows more severe treatment needs, with more sections assigned to Close Monitoring and Major Rehabilitation. These differences indicate that the framework does not assign a single corridor-wide treatment category, but instead preserves spatial variation needed for network-level screening and prioritization.

CHAPTER 7. CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the major findings of the study, provides recommendations for GDOT implementation, and identifies limitations and future research directions. The conclusions are organized around four major components of the study: multi-source data construction, threshold-based structural condition classification, ML-based structural prediction, and integrated surface-structural treatment decision-making.

7.1 Conclusions

7.1.1 Data Construction and Multi-Source Integration

- **The missing data analysis was used to determine which corridors had structurally interpretable TSDD records for flexible-pavement analysis.** The analysis did not only quantify missing TSDD records; it also examined whether the gaps were caused by localized survey limitations or by pavement-type incompatibility. I-24 was excluded because most of its missing TSDD records were associated with rigid or composite pavement sections, where TSDD deflection-based indicators cannot be interpreted as valid flexible-pavement structural responses. I-285, I-575, and I-59 were retained because their missing data patterns were localized and could be addressed through pavement-type screening, spatial filtering, and coring-centered quality control.
- **The multi-source data were integrated through a coring-centered spatial matching process that links laboratory HWTT outcomes with nearby field measurements.** TSDD, GPR, 3D pavement surface, and HWTT data were combined across the retained Georgia interstate corridors using a two-stage nearest-neighbor procedure. In the first stage,

continuous TSDD and GPR records were matched using a 3 m threshold. In the second stage, discrete coring locations with HWTT results were matched to the integrated TSDD-GPR records using a 9.2 m threshold. This process produced a coring-centered dataset in which each retained HWTT record is paired with TSDD structural indicators, GPR-derived layer thickness, and 3D surface condition measurements representing the same flexible pavement section.

- **The final dataset contains 85 flexible pavement sections and captures meaningful variation in structural and functional condition across the retained corridors.** I-575 represents the favorable condition group, with all retained Layer 1 HWTT outcomes passing and generally lower TSDD structural responses. I-59 represents the poor condition group, with all retained Layer 1 HWTT outcomes failing and the highest TSDD curvature and deflection responses. I-285 provides mixed HWTT outcomes and greater spatial variability, making it useful for evaluating model discrimination and treatment decision logic. This corridor-level variation supports the development and evaluation of both threshold-based and ML-based screening methods.

7.1.2 Threshold-Based Structural Condition Classification

- **SCI_8 provided the strongest performance among the evaluated single-variable threshold rules and is the most deployable near-term candidate for pilot structural screening.** Among all candidate single-variable threshold rules, SCI_8 achieved the best overall performance, with a mean F1 score of 0.806 and a mean AUC of 0.888 across 50 repeated stratified CV splits. This result confirms that shallow deflection-basin curvature is closely related to Layer 1 HWTT performance and can serve as a practical network-level structural screening indicator. The preliminary threshold of approximately 0.585 mils provides a current dataset-based Georgia reference value tied to laboratory HWTT outcomes. Because this rule requires only TSDD-derived SCI_8, uses a single decision boundary, and is easy to interpret, it can be piloted within an existing PMS data workflow without specialized modeling software or complex feature processing.

- **The SCI_8 + AC thickness duo-variable threshold is the preferred enhanced screening option when reliable GPR-derived AC thickness data are available.** Instead of applying one SCI_8 cutoff to all sections, this method separates thinner and thicker asphalt sections and applies group-specific SCI_8 thresholds. This better reflects the fact that SCI_8 response depends partly on asphalt thickness. In the current 85-record dataset, the method achieved the highest mean F1 score among all evaluated methods, with F1 = 0.834 and AUC = 0.887. It improved recall and reduced false negatives compared with the single-variable SCI_8 threshold, meaning fewer HWTT-failed sections were missed. However, because the AC thickness split point and group-specific SCI_8 thresholds depend on the available calibration data, they should be treated as preliminary pilot-screening values and recalibrated as more paired TSDD, GPR, and HWTT data become available.

7.1.3 Machine Learning Model

- **Under the current 85-section dataset, RF provides the strongest ML-based performance and remains competitive with the threshold methods, but the results support complementary deployment roles rather than replacing interpretable threshold-based screening with ML.** Among the four ML classifiers, RF achieved the best ML performance, with F1 = 0.822 and the highest AUC among all evaluated methods, 0.891. The duo-variable threshold had the highest F1, while RF had the highest AUC, indicating that each method has a different practical strength. At the current dataset scale, threshold rules provide more transparent near-term screening, while RF provides a flexible and retrainable model structure for future PMS-oriented prediction.

- **Feature selection confirms that the ML models rely on the same core structural signals identified by the threshold analysis.** D0 and SCI_8 were selected in more than 90 percent of CV splits, while AC thickness and base thickness were also frequently retained. These predictors represent near-load deflection magnitude, shallow curvature response, and pavement layer geometry. Therefore, the ML models do not rely on a high-dimensional or opaque feature set; instead, they provide a more flexible way to combine the same physically meaningful structural indicators.

7.1.4 Treatment Decision Framework

- **The treatment decision framework translates structural and functional condition results into four candidate PMS-oriented screening categories.** Structural condition is predicted from TSDD and GPR data, while functional condition is derived from 3D pavement surface indicators. These two condition dimensions are combined to assign each pavement section to a candidate screening category: **No Action, Milling and Resurfacing, Close Monitoring, or Major Rehabilitation.** This structure allows GDOT to screen sections that may primarily require surface treatment, sections that warrant structural monitoring, and sections that should be prioritized for project-level review before major rehabilitation decisions are made.
- **The Close Monitoring category provides the main added value beyond a surface-only PMS workflow.** It identifies sections where surface condition remains acceptable but

predicted structural condition is poor. Under a surface-only workflow, these sections could be classified as No Action because visible distress has not yet become severe. By flagging them separately, the framework creates an early warning category for locations that may require structural verification, closer performance tracking, or future rehabilitation programming.

- **The corridor-level maps show that the framework produces treatment patterns consistent with the distinct condition characteristics of each interstate corridor.** I-575 is dominated by lower-intervention categories, consistent with its favorable HWTT and surface condition patterns. I-59 shows more severe treatment needs, with more sections assigned to Close Monitoring and Major Rehabilitation. I-285 shows the greatest spatial variability, consistent with its heterogeneous urban beltway characteristics. These results indicate that the framework can differentiate surface-driven needs, structure-driven monitoring needs, and combined deterioration patterns at the corridor scale.

7.2 Recommendations for GDOT

7.2.1 Near-Term Implementation

- **GDOT should pilot the SCI_8 single-variable threshold as a primary near-term structural screening candidate on selected interstate corridors.** The SCI_8 threshold is the simplest operational option because it requires only TSDD data, uses a single calibrated decision boundary, and produces results that are easy to explain and audit. The preliminary Georgia-specific threshold of approximately 0.585 mils provides a practical starting point for pilot testing TSDD-based structural screening within GDOT's PMS workflow.

- **Where reliable GPR-derived AC thickness data are available, GDOT should also pilot the SCI_8 + AC thickness duo-variable threshold as an enhanced screening option.** This method provides the highest current F1 score and improves failure detection by accounting for thickness-dependent differences in SCI_8 interpretation. However, the method should be implemented as a recalibrated framework rather than as a fixed statewide rule. The AC thickness split point and group-specific SCI_8 thresholds should be updated as additional labeled sections become available.
- **GDOT should pilot the integrated treatment decision matrix on the three study corridors before broader deployment.** The four-category decision structure can be embedded into existing PMS workflows as a transparent screening layer. Comparing the framework outputs with current programming decisions on I-285, I-575, and I-59 would allow GDOT to identify sections where surface-only decisions may understate structural needs or overstate the need for major rehabilitation.

7.2.2 Data Expansion and Organization

- **GDOT should expand paired TSDD, GPR, 3D surface, and HWTT data collection to additional corridors.** Priority should be given to routes with different AC thickness distributions, base materials, traffic loading, subgrade conditions, and pavement ages. A broader dataset will improve the reliability of threshold recalibration, reduce corridor-specific bias, and provide the diversity needed for ML models to demonstrate stronger generalization.
- **GDOT should adopt the standardized coring data format developed in this study as the template for future PES and HWTT data collection.** Consistent field names, units, layer definitions, GPS validation flags, and exclusion reason fields will reduce future data cleaning

effort and support continuous dataset expansion. A standardized database will also make future model updating more efficient and auditable.

7.2.3 Model Updating and Transition to ML-Based Screening

- **RF should be used as the future retrainable ML option, not as the immediate replacement for threshold-based screening.** Although the SCI_8 + AC thickness duo-variable threshold provides the highest current F1 score, RF provides the strongest ML-based performance, the highest AUC, and a model structure that can be updated as the labeled dataset expands. This makes RF more suitable for future PMS-oriented implementation, especially when additional TSDD, GPR, and HWTT-labeled data become available across more corridors and pavement structures.
- **For ML implementation, a compact feature set should be prioritized: SCI_8, D0, and AC thickness.** SCI_8 captures shallow curvature response, D0 captures deflection magnitude at the load center, and AC thickness provides the structural context needed to interpret the measured response. This compact input set preserves the dominant structural information while avoiding unnecessary redundancy from highly correlated basin variables.
- **GDOT should maintain threshold-based and ML-based screening in parallel during the next implementation stage.** The SCI_8 threshold and SCI_8 + AC thickness rule should serve as near-term pilot screening tools, while RF should be periodically retrained as new HWTT-labeled data become available. The performance gap between RF and the threshold benchmarks should be monitored on structurally diverse validation datasets. If RF achieves a sustained and practically meaningful improvement, it can be promoted from a parallel analytical model to the primary structural screening engine.

7.3 Limitations and Future Research Directions

This study is subject to several limitations related to data scope, ground reference definition, treatment decision validation, and economic evaluation. These limitations do not reduce the value of the proposed framework, but they define the main priorities for future research and implementation.

First, the analytical dataset is limited to 85 coring sections from three interstate corridors and includes only flexible pavement sections. Although the dataset was carefully constructed through spatial matching, pavement-type screening, and coring-centered quality control, the calibrated thresholds and ML models have not yet been validated across a broader range of Georgia pavement structures, traffic levels, construction histories, or environmental conditions. Future work should expand the HWTT-labeled dataset to additional corridors and pavement designs to evaluate the generalizability and stability of the proposed screening methods.

Second, this study uses the Layer 1 HWTT outcome as a laboratory-based structural-performance reference. This reference provides a standardized and repeatable laboratory-based benchmark for asphalt-layer rutting resistance and moisture susceptibility. However, it does not fully represent all forms of in-service structural deterioration, such as fatigue cracking, base deterioration, drainage-related failure, or subgrade deformation. Future research should examine the relationship between HWTT outcomes and long-term field performance records and should consider incorporating complementary structural reference indicators, such as additional coring results, FWD measurements, distress progression, or project-level rehabilitation records.

Third, the treatment decision framework provides network-level screening recommendations rather than final project-level treatment designs, and uncertainty in both

the structural and functional condition branches may affect the final decision output. In this study, the functional condition classification was rule-based and was not independently validated as a predictive model. Because each treatment category depends on the combined structural and functional labels, future work should evaluate the accuracy and sensitivity of the functional condition classifier and quantify how uncertainty propagates through the treatment decision matrix. Categories such as Close Monitoring and Major Rehabilitation should be interpreted as prioritization and review flags rather than final treatment prescriptions. Final treatment selection still requires project-level engineering confirmation, including field inspection, coring, drainage review, traffic assessment, constructability review, and cost evaluation. Future implementation should define how each screening category triggers follow-up project-level actions within GDOT's existing programming workflow.

Finally, this study does not quantify the economic benefit of the proposed framework relative to current GDOT practice. The framework is designed to reduce the risk of cost overruns by identifying structural concerns earlier in the planning process, but its life-cycle cost benefit has not yet been evaluated using Georgia-specific construction costs, treatment histories, and performance models. Future work should compare surface-only PMS decisions with the integrated surface-structural decision framework to estimate potential savings, treatment timing benefits, and reductions in late-stage scope revisions.

Overall, future implementation should proceed as a staged expansion rather than a one-time deployment. The SCI_8 threshold can support near-term pilot TSDD-based structural screening, the SCI_8 + AC thickness rule can be used as an enhanced pilot screening option where reliable GPR data are available, and RF can be maintained as a retrainable ML component as labeled data expand. Continued data collection, standardized coring and HWTT documentation, project-level validation, and economic analysis will be necessary to move the proposed framework from research-scale evaluation toward routine PMS integration.

APPENDIX A. TSDD DATA DICTIONARY

This appendix documents the TSDD variables used in this study. The TSDD dataset provides continuous pavement structural response measurements at 0.01-mile intervals. Variables are grouped into location and section referencing fields, derived structural condition indices, deflection bowl measurements, slope indices, survey environment and load measurements, and raw diagnostic variables.

A.1 Location and Section Referencing Variables

Field Name	**Description**	**Units / Format**
ROAD_SECTION_ID	Unique roadway section identification number	Text
ROAD_NO	Road identification code	Text
BLOCK_ID	Block identification number	Integer
DIRECTION	Survey direction	Text
ROUTE	Roadway name or route identifier	Text
FROM	Start section identifier	Text
TO	End section identifier	Text
INSIGHT_CHAINAGE_FROM	Distance from roadway origin to start of interval	Miles
INSIGHT_CHAINAGE_TO	Distance from roadway origin to end of interval	Miles
SECTION_SUB_CHAINAGE_FROM	Distance from section start to interval start	Miles
SECTION_SUB_CHAINAGE_TO	Distance from section start to interval end	Miles
HAWKEYE_URL	Link to Hawkeye Insight visualization	Hyperlink
HAWKEYE_URL_STRING	URL string for Hawkeye Insight visualization	Text

A.2 Derived Structural Condition Indices

Structural Condition Indices summarize the curvature of the calculated deflection bowl and provide indicators of pavement structural response.

Field Name	Description	Units
SCI_8	Structural Condition Index calculated as D8 − D0	mils
SCI_12	Structural Condition Index calculated as D12 − D0	mils
SCI_SUBGRADE	Subgrade curvature index calculated as D60 − D36	mils

A.3 Derived Deflection Bowl Measurements

Deflection values describe pavement surface response at multiple offsets relative to the applied wheel load. These values are calculated from the TSDD deflection bowl and reported in imperial units.

Field Name	Description	Units
D-18	Deflection at −18 in. offset	mils
D-12	Deflection at −12 in. offset	mils
D-8	Deflection at −8 in. offset	mils
D0	Deflection at load center	mils
D8	Deflection at 8 in. offset	mils
D12	Deflection at 12 in. offset	mils
D18	Deflection at 18 in. offset	mils
D24	Deflection at 24 in. offset	mils
D36	Deflection at 36 in. offset	mils
D48	Deflection at 48 in. offset	mils
D60	Deflection at 60 in. offset	mils
D72	Deflection at 72 in. offset	mils

A.4 Derived Slope Indices

Slope indices describe the spatial gradient of the deflection bowl and provide complementary indicators of pavement stiffness.

Field Name	Description	Units
SLOPE-450	Slope at −450 mm offset	µm/m
SLOPE-300	Slope at −300 mm offset	µm/m
SLOPE-200	Slope at −200 mm offset	µm/m
SLOPE130	Slope at 130 mm offset	µm/m
SLOPE215	Slope at 215 mm offset	µm/m
SLOPE300	Slope at 300 mm offset	µm/m
SLOPE450	Slope at 450 mm offset	µm/m
SLOPE600	Slope at 600 mm offset	µm/m
SLOPE900	Slope at 900 mm offset	µm/m
SLOPE1500	Slope at 1500 mm offset	µm/m

A.5 Survey Environment and Load Measurements

Field Name	Description	Units
SURFACE_TEMPERATURE	Pavement surface temperature during survey	°F
AIR_TEMPERATURE	Ambient air temperature during survey	°F
STRAIN_GAUGE_LEFT	Left wheel load measurement	lb
STRAIN_GAUGE_RIGHT	Right wheel load measurement	lb
SURVEY_SPEED	Vehicle speed during TSDD measurement	mph
SURVEY_DATE	TSDD survey date	Date
SURVEY_TIME	TSDD survey time	Time
SLOPE_QUALITY	Slope quality indicator	Text

A.6 Raw Diagnostic Variables

Raw deflection and slope variables are retained for diagnostic and quality-control purposes. These fields are not necessarily used directly in modeling but support data review and traceability.

Field Name	Description	Units
RawSci200	Raw SCI measurement at 200 mm offset	µm
RawSci300	Raw SCI measurement at 300 mm offset	µm
RawSciSub	Raw subgrade SCI measurement	µm
TD-450 to TD1500	Raw deflection bowl values at metric offsets	µm
RawSlope-450 to RawSlope1500	Raw slope values at metric offsets	µm/m

APPENDIX B. 3D SURFACE DATA DICTIONARY

This appendix documents the 3D pavement surface variables used to characterize functional pavement condition. The variables are grouped into cracking, rutting, roughness, surface texture, pothole, joint and faulting, and roadway geometry categories.

B.1 Surface Cracking Measurements

Field Name	**Description**	**Units**
PERCENT_TOTAL_CRACK	Percentage of pavement surface affected by cracking	%
TOTAL_CRACK_LENGTH	Total crack length across all crack types	ft
TOTAL_HIGH_SEVERITY_CRACK_LENGTH	Total length of high-severity cracking	ft
ALLIG_CRACK_TOTAL	Total area of alligator cracking	ft^2
LONG_CRACK_TOTAL	Total length of longitudinal cracking	ft
TRANS_CRACK_TOTAL	Total length of transverse cracking	ft
ALLIG_CRACK_Z1–Z5	Alligator cracking area by lateral zone	ft^2
LONG_CRACK_Z1–Z5	Longitudinal cracking length by lateral zone	ft
CRACK_Z*_LOW / HIGH	Crack density by zone and severity	ft/yd^2
ALLIG_W_Z*, LONG_W_Z*, TRANS_W_TOTAL	Crack width metrics	in.

B.2 Rutting Measurements

Field Name	**Description**	**Units**
RUT_RIGHT	Rut depth in the right wheel path	in.
RUT_LEFT	Rut depth in the left wheel path	in.
RUT_WIDTH_RIGHT	Rut width in the right wheel path	in.
RUT_WIDTH_LEFT	Rut width in the left wheel path	in.
CROSS_SECTION_RIGHT	Rut cross-sectional area in the right wheel path	in^2
CROSS_SECTION_LEFT	Rut cross-sectional area in the left wheel path	in^2
RUT_AVG	Average rut depth from left and right wheel paths	in.
RUT_LANE	Lane-level rut depth from the full-lane method	in.

B.3 Roughness Measurements

Field Name	Description	Units
IRI_RIGHT	IRI in the right wheel path	in./mi
IRI_LEFT	IRI in the left wheel path	in./mi
IRI_AVG	Average IRI from left and right wheel paths	in./mi
IRI_LANE	Lane-level IRI from the half-car method	in./mi

B.4 Surface Texture Measurements

Field Name	Description	Units
MPD_MACROTEXTURE_RIGHT	MPD in the right wheel path	mm
ETD_MACROTEXTURE_RIGHT	ETD in the right wheel path	mm
MPD_MACROTEXTURE_LEFT	MPD in the left wheel path	mm
ETD_MACROTEXTURE_LEFT	ETD in the left wheel path	mm
3D_MPD_ZONE1–ZONE5	3D macrotexture by lateral zone	mm

B.5 Pothole Measurements

Field Name	Description	Units
NUMBER_POTHOLES	Number of detected potholes within the interval	Count
POTHOLE_AREA	Total pothole area	ft^2
AVERAGE_DEPTH	Average pothole depth	in.
MAX_DEPTH	Maximum pothole depth	in.
POTHOLE_FHWA_SEVERITY	FHWA pothole severity classification	Text

B.6 Joint and Faulting Measurements

Field Name	Description	Units
JOINT COUNT	Number of joints within the interval	Count
MAX_FAULT_HEIGHT	Maximum joint fault height	in.
AVG_FAULT_HEIGHT	Average joint fault height	in.
MAX_JOINT_WIDTH	Maximum joint width	in.
AVG_JOINT_WIDTH	Average joint width	in.

B.7 Roadway Geometry Variables

Field Name	Description	Units
GRADE	Longitudinal grade of the interval	%
CROSS_SLOPE	Cross slope of the interval	%
HORIZONTAL_CURVATURE	Horizontal curvature	rad/mi
VERTICAL_CURVATURE	Vertical curvature	rad/mi

APPENDIX C. GPR DATA DICTIONARY

This appendix documents the GPR variables used in this study. GPR data provide pavement layer thickness estimates and are used to identify flexible, rigid, and composite pavement structures. The primary variable used in the structural analysis is AC thickness. PCC thickness is used mainly as a pavement-type screening indicator because the presence of PCC identifies rigid or composite structures that are not appropriate for TSDD-based flexible pavement structural analysis.

C.1 Location and Referencing Fields

Field Name	**Description**	**Units**
Route	Interstate route identifier	—
Direction	Travel direction	—
Lane / Lane Group	Lane number or lane group represented by the GPR record	—
Milepost / Chainage	Longitudinal position along the corridor	mi
Wheel Path	Left or right wheel path where reported	—

C.2 Pavement Layer Thickness Fields

Field Name	**Description**	**Units**
AC Thickness	Total asphalt concrete thickness	in.
Surface AC Thickness	Thickness of the upper asphalt concrete layer where separated	in.
Bottom AC Thickness	Thickness of the lower asphalt layer where separated	in.
Full AC Thickness	Total asphalt thickness where AC layers are not separated	in.
PCC Thickness	Portland cement concrete thickness where present	in.
Base Thickness	Base layer thickness beneath AC or PCC where reported	in.

C.3 Data Availability and Interpretation Notes

Field Condition	Interpretation
Thickness value present	Layer detected and interpreted in the GPR output
AC thickness present and PCC absent	Flexible pavement section, subject to further screening
PCC thickness present	Rigid or composite pavement section; excluded from flexible-pavement structural modeling
NA	Thickness not available or not reported
#DIV/0! or #NUM!	Spreadsheet calculation artifact; treated as missing
"all PCC"	PCC-dominated section with no meaningful flexible AC thickness for this study

C.4 Corridor-Level Thickness Interpretation

The GPR data show clear structural differences among the candidate corridors. I-24 includes substantial PCC presence and is dominated by rigid or composite pavement structures. This supports its exclusion from the TSDD-based flexible pavement analysis. I-59 is characterized by relatively consistent full-depth AC thickness, with typical values near 10 in. I-285 includes both flexible and composite sections, requiring pavement-type screening before integration. I-575 is primarily flexible and generally has thicker AC layers than I-59, although some interchange-related variability is present.

APPENDIX D. CORING, HWTT, AND SPATIAL INTEGRATION RECORDS

This appendix documents the standardized coring and HWTT data format used in this study. The dataset was compiled from GDOT PES records and organized to provide consistent roadway location information, coring metadata, HWTT outcomes, and associated pavement layer information. Each record represents one coring location.

Layer 1 HWTT results are used as the primary laboratory performance reference in this study because they represent the upper asphalt layer most directly associated with near-surface structural response. Lower-layer HWTT results are retained for completeness but are not used as the primary classification label.

Blank HWTT fields indicate that the corresponding layer was not tested or not reported. Blank PCC thickness fields generally indicate that no PCC layer was identified or reported for the coring location. Where PCC thickness is present, the record may represent a rigid or composite pavement structure and requires additional screening before use in flexible-pavement structural analysis.

D.1 Roadway and Coring Location Fields

Field Name	Description	Data Type	Notes
Road Section	Interstate route or roadway section where the core was extracted	String	Required
Direction	Travel direction of the roadway section	String	Required
Mile Post	Milepost location of the coring site	Float	Optional but useful for spatial referencing
Lane Number	Lane where the core was extracted; Lane 1 is the leftmost travel lane	Integer	Required
Wheel Path	Coring position within the lane, such as left, right, or center wheel path	String	Required
Coring Date	Date or month when the core was collected	Date / Text	Required
GPS Latitude	Latitude of the coring location	Float	Required when available
GPS Longitude	Longitude of the coring location	Float	Required when available

D.2 HWTT Fields

HWTT results are recorded by pavement layer. Layer 1 refers to the uppermost asphalt layer tested and is used as the primary laboratory performance reference in this study. Lower-layer results are retained where available for documentation and future analysis.

Field Name	Description	Data Type	Notes
HWTT result, Layer 1	Pass/fail outcome for the uppermost tested asphalt layer	String	Primary modeling label
Number of Passes, Layer 1	Number of wheel passes applied during Layer 1 HWTT	Integer	Used to document test termination condition
Max Deformation (mm), Layer 1	Maximum measured deformation for Layer 1	Float	Reported in millimeters
HWTT result, Layer 2	Pass/fail outcome for the second tested layer, where available	String	Retained for reference
Number of Passes, Layer 2	Number of wheel passes applied during Layer 2 testing	Integer	Blank if Layer 2 was not tested
Max Deformation (mm), Layer 2	Maximum measured deformation for Layer 2	Float	Blank if Layer 2 was not tested
HWTT result, Layer 3	Pass/fail outcome for the third tested layer, where available	String	Retained where reported
Number of Passes, Layer 3	Number of wheel passes applied during Layer 3 testing	Integer	Blank if Layer 3 was not tested
Max Deformation (mm), Layer 3	Maximum measured deformation for Layer 3	Float	Blank if Layer 3 was not tested
HWTT result, Layer 4	Pass/fail outcome for the fourth tested layer, where available	String	Retained where reported
Number of Passes, Layer 4	Number of wheel passes applied during Layer 4 testing	Integer	Blank if Layer 4 was not tested
Max Deformation (mm), Layer 4	Maximum measured deformation for Layer 4	Float	Blank if Layer 4 was not tested

D.3 GPR Reference and Layer Thickness Fields

These fields are included in the coring and HWTT dataset because they provide pavement layer information associated with each coring location. GPR-derived thickness values are based on the nearest available GPR measurement point, while actual thickness values are obtained from physical coring records where available.

Field Name	**Description**	**Data Type**	**Notes**
GPR MP	Milepost of the nearest GPR measurement point	Float	Used to link coring and GPR records
Distance (m)	Distance between the coring location and the nearest GPR measurement point	Float	Used to evaluate GPR match reliability
GPR AC Thickness (in.)	Asphalt concrete thickness estimated from GPR	Float	Reported in inches
GPR PCC Thickness (in.)	Portland cement concrete thickness estimated from GPR, where present	Float	Indicates rigid or composite pavement where populated
GPR Base Thickness (in.)	Base layer thickness estimated from GPR, where available	Float	Reported in inches
Actual AC Thickness (in.)	Asphalt concrete thickness measured from the core	Float	Used for comparison with GPR-derived thickness
Actual PCC Thickness (in.)	PCC thickness measured from the core, where present	Float	Indicates rigid or composite pavement where populated
Actual Base Thickness (in.)	Base thickness measured from the core, where available	Float	Reported in inches
Notes	Additional comments related to construction, materials, or data interpretation	String	Used for special cases and review notes

APPENDIX E. FINAL ANALYTICAL DATASET STRUCTURE AND FIELD DICTIONARY

This appendix documents the field structure of the final analytical dataset. The final dataset was created after spatial integration, map-based quality control, and coring-centered refinement. Each row represents one retained coring site that was successfully matched with relevant TSDD, GPR, and 3D surface condition information.

The final dataset contains 85 retained pavement sections from I-285, I-575, and I-59. Each record combines laboratory performance information, pavement layer thickness, coring-based thickness information, matched TSDD indicators, matched 3D surface condition variables, survey-context variables, raw diagnostic fields, and spatial matching information.

E.1 Dataset Organization

Dataset Level	Description
Row unit	One retained coring location
Spatial basis	Coring-centered record matched to nearby TSDD, GPR, and 3D surface data
Primary label	HWTT result, Layer 1
Structural predictors	TSDD deflection and SCI variables
Layer information	GPR-derived and coring-measured thickness variables
Surface condition variables	3D cracking, rutting, roughness, and geometry indicators
Matching metadata	Match type and associated spatial identifiers

E.2 Core Coring and HWTT Result Fields

Field Name	Description	Units / Type
route	Interstate route of the retained coring location	Text
direction	Travel direction	Text
mp	Milepost of the coring location	Miles
lane	Lane number of the coring location	Integer
wheel_path	Wheel path or coring position	Text
latitude	Latitude of the coring location	Decimal degrees
longitude	Longitude of the coring location	Decimal degrees
hamburg_result_layer1	HWTT Pass/Fail outcome for Layer 1	Text
hamburg_result_layer2	HWTT Pass/Fail outcome for Layer 2, where available	Text

E.3 GPR and Coring-Based Thickness Fields

Field Name	Description	Units
gpr_ac_thickness_(in.)	GPR-derived asphalt concrete thickness	in.
gpr_base_thickness_(in.)	GPR-derived base layer thickness	in.
actual_ac_thickness_(in.)	Asphalt concrete thickness measured from coring records	in.
actual_base_thickness_(in.)	Base thickness measured from coring records	in.

E.4 Matched TSDD Location and Referencing Fields

Field Name	Description	Units / Type
tsd_route	Route identifier from the matched TSDD record	Text
tsd_lane	Lane identifier from the matched TSDD record	Text
tsd_latitude	Latitude of the matched TSDD interval	Decimal degrees
tsd_longitude	Longitude of the matched TSDD interval	Decimal degrees
tsd_from	Start location or section reference of the TSDD interval	Text
tsd_to	End location or section reference of the TSDD interval	Text
tsd_insight_chainage_from (mi)	Start chainage of the TSDD interval	Miles
tsd_insight_chainage_to (mi)	End chainage of the TSDD interval	Miles
tsd_section_sub_chainage_from (mi)	Start sub-chainage within the roadway section	Miles
tsd_section_sub_chainage_to (mi)	End sub-chainage within the roadway section	Miles

E.5 TSDD Structural Indicator Fields

Field Name	Description	Units
tsd_sci_8	Structural Curvature Index calculated as D8 − D0	mils
tsd_sci_12	Structural Curvature Index calculated as D12 − D0	mils
tsd_sci_subgrade	Subgrade curvature index calculated as D60 − D36	mils
tsd_d-18 (mils)	Deflection at −18 in. offset	mils
tsd_d-12 (mils)	Deflection at −12 in. offset	mils
tsd_d-8 (mils)	Deflection at −8 in. offset	mils
tsd_d0 (mils)	Deflection at load center	mils
tsd_d8 (mils)	Deflection at 8 in. offset	mils
tsd_d12 (mils)	Deflection at 12 in. offset	mils
tsd_d18 (mils)	Deflection at 18 in. offset	mils
tsd_d24 (mils)	Deflection at 24 in. offset	mils
tsd_d36 (mils)	Deflection at 36 in. offset	mils
tsd_d48 (mils)	Deflection at 48 in. offset	mils
tsd_d60 (mils)	Deflection at 60 in. offset	mils
tsd_d72 (mils)	Deflection at 72 in. offset	mils

E.6 TSDD Survey Environment and Loading Fields

Field Name	Description	Units
tsd_surface_temperature (f)	Pavement surface temperature during TSDD survey	°F
tsd_air_temperature (f)	Air temperature during TSDD survey	°F
tsd_strain_gauge_left (lb)	Left wheel load measurement	lb
tsd_strain_gauge_right (lb)	Right wheel load measurement	lb
tsd_survey_speed (mph)	TSDD survey speed	mph

E.7 Matched 3D Surface Condition Fields

Field Name	Description	Units
tsd_percent_total_crack (%)	Percentage of pavement surface affected by cracking	%
tsd_total_crack_length (ft)	Total crack length	ft
tsd_total_high_severity_crack_length (ft)	Total high-severity crack length	ft
tsd_rut_avg (in)	Average rut depth	in.
tsd_rut_lane (in)	Lane-level rut depth	in.
tsd_iri_avg (in/mi)	Average IRI	in./mi
tsd_iri_lane (in/mi)	Lane-level IRI	in./mi

E.8 Roadway Geometry and GPS Quality Fields

Field Name	Description	Units / Type
tsd_grade (%)	Longitudinal grade of the matched TSDD interval	%
tsd_cross_slope (%)	Cross slope of the matched TSDD interval	%
tsd_horizontal_curvature (rad/mi)	Horizontal curvature	rad/mi
tsd_vertical_curvature (rad/mi)	Vertical curvature	rad/mi
tsd_to_latitude (deg)	Latitude at the end of the matched TSDD interval	Decimal degrees
tsd_to_longitude (deg)	Longitude at the end of the matched TSDD interval	Decimal degrees
tsd_to_altitude (ft)	Altitude at the end of the matched TSDD interval	ft
tsd_gps_position_is_calculated	Flag indicating whether the GPS position was calculated	Boolean / Text

E.9 Raw TSDD Slope Diagnostic Fields

These raw slope variables are retained for diagnostic and quality-control purposes. They are not necessarily used directly in modeling but provide traceability to the original TSDD measurement outputs.

Field Name	Description	Units
tsd_rawslope-450 (um/m)	Raw slope at −450 mm offset	µm/m
tsd_rawslope-300 (um/m)	Raw slope at −300 mm offset	µm/m
tsd_rawslope-200 (um/m)	Raw slope at −200 mm offset	µm/m
tsd_rawslope130 (um/m)	Raw slope at 130 mm offset	µm/m
tsd_rawslope215 (um/m)	Raw slope at 215 mm offset	µm/m
tsd_rawslope300 (um/m)	Raw slope at 300 mm offset	µm/m
tsd_rawslope450 (um/m)	Raw slope at 450 mm offset	µm/m
tsd_rawslope600 (um/m)	Raw slope at 600 mm offset	µm/m
tsd_rawslope900 (um/m)	Raw slope at 900 mm offset	µm/m
tsd_rawslope1500 (um/m)	Raw slope at 1500 mm offset	µm/m

E.10 Spatial Matching Field

Field Name	Description	Units / Type
match_type	Spatial matching or review label indicating how the coring location was matched to the corresponding sensing record	Text

ACKNOWLEDGEMENT

The research team would like to express its sincere appreciation to the GDOT for supporting this research project. The authors especially thank Kamari B. Jordan, Binh Bui, Ian Rish, Damon Frost, Brandon Clayton, Ryan Kellett, and Ariel N. Steele for their guidance, technical input, data coordination, and review comments throughout the project.

The authors also gratefully acknowledge Jordan Glass, Venkateshwarlu Balla, Sparsh Gupta, Mohsen Mohammadi, and Haolin Wang for their contributions as current or former members and collaborators of the research team. Their assistance with data preparation, technical discussion, analysis support, and report development helped advance the project and improve the final report.

The authors also acknowledge Greenwood Engineering and ARRB Systems for their technical support related to TSDD data collection, data interpretation, and device-specific information used in this study. Their assistance helped the research team better understand the TSDD measurement system and the associated data structure.

REFERENCES

Abdelmuhsen, A., Simonin, J.-M., Schmidt, F., Lièvre, D., Cothenet, A., & Ihamouten, A. (2023). On the variants of SVM method for the estimation of soil elastic modulus from TSD model: Numerical parametric study. *Transportation Engineering*, *13*, 100187.

Ahmed, N. S., Huynh, N., Gassman, S., Mullen, R., Pierce, C., & Chen, Y. (2022). Predicting pavement structural condition using machine learning methods. *Sustainability*, *14*(14), 8627.

Beizaei, M., Tirado, C., Rocha, S., Nazarian, S., & Rada, G. (2024). Qualitative assessment of Traffic Speed Deflectometer data collection practice. *International Journal of Pavement Engineering*, *25*(1), 2447482.

Bryce, J. (2024). Enhancing pavement performance prediction with traffic speed deflection data. *International Journal of Pavement Engineering*, *25*(1), 2343090.

Canestrari, F., Ingrassia, L. P., Spinelli, P., & Graziani, A. (2023). A new methodology to assess the remaining service life of motorway pavements at the network level from traffic speed deflectometer measurements. *International Journal of Pavement Engineering*, *24*(2), 2128349.

Chen, Z., Tong, X., Cheng, H., & Sun, L. (2025). Conversions of viscoelastic deflections of asphalt pavement under TSD load to elastic ones for pavement condition assessments. *Road Materials and Pavement Design*, *26*(1), 158–180.

Elgamal, E., Fawzy, H., Mohamady, A., & Basha, A. (2025). Pavement management system: Condition assessment, challenges, and future directions. *Journal of Contemporary Technology and Applied Engineering*, *3*(2), 93–102.

Elseifi, M. A., & Elbagalati, O. (2017). Assessment of continuous deflection measurement devices in Louisiana-rolling wheel deflectometer: final report 581.

Hamidi, A., Hoff, I., & Mork, H. (2024). Elastic and viscoelastic back-calculation of pavement layers' moduli using data obtained from traffic speed deflection devices. *Construction and Building Materials*, *447*, 138035.

Highway, A. A. o. S., & Officials, T. (2017). Standard method of test for Hamburg wheel-track testing of compacted asphalt mixtures. *AASHTO T324*, *94*.

Huang, B., Zhang, M., Gong, H., & Polaczyk, P. (2022a). *Evaluation of traffic speed deflectometer for collecting network level pavement structural data in tennessee*.

Huang, B., Zhang, M., Gong, H., & Polaczyk, P. (2022b). Evaluation of traffic speed deflectometer for collecting network level pavement structural data in tennessee.

Jiang, D., Wang, D., Pei, Z., Sun, Z., & Yi, J. (2025). A comprehensive survey of the new generation pavement structural condition assessment in pavement management system: traffic speed deflection device. *IEEE Transactions on Intelligent Transportation Systems*.

Kargah-Ostadi, N., Vasylevskyi, K., Ablets, A., & Drach, A. (2024). Physics-informed neural networks to advance pavement engineering and management. *Road Materials and Pavement Design*, *25*(11), 2382–2403.

Katicha, S., Flintsch, G., & Diefenderfer, B. (2022). Ten years of traffic speed deflectometer research in the United States: a review. *Transportation research record*, *2676*(12), 152–165.

Katicha, S. W., Flintsch, G. W., Murekye, A., Perrone, E., Amarh, E. A., & Diefenderfer, B. K. (2025). Evaluating the Potential Use of Traffic Speed Deflection Device-Based Pavement Structural Data for Asphalt Pavements in VDOT's Pavement Management Processes.

Katicha, S. W., Shrestha, S., Flintsch, G. W., & Diefenderfer, B. K. (2020a). Network level pavement structural testing with the traffic speed deflectometer.

Katicha, S. W., Shrestha, S., Flintsch, G. W., & Diefenderfer, B. K. (2020b). *Network level pavement structural testing with the traffic speed deflectometer*.

Kumar, A. V. (2024). Pavement surface condition assessment: a-state-of-the-art research review and future perspective. *Innovative Infrastructure Solutions*, *9*(12), 470.

Liu, Z., Wang, S., Gu, X., Wang, D., Dong, Q., & Cui, B. (2024). Intelligent assessment of pavement structural conditions: A novel FeMViT classification network for GPR images. *IEEE Transactions on Intelligent Transportation Systems*, *25*(10), 13511–13523.

Mabrouk, G. M., Elbagalati, O. S., Dessouky, S., Fuentes, L., & Walubita, L. F. (2022). Using ANN modeling for pavement layer moduli backcalculation as a function of traffic speed deflections. *Construction and Building Materials*, *315*, 125736.

Murekye, A., Katicha, S. W., Amarh, E. A., Flintsch, G. W., & Diefenderfer, B. K. (2025). Evaluating the Effect of Incorporating Network-Level Structural Condition on Pavement Lifecycle Cost. In *Airfield and Highway Pavements 2025* (pp. 163–175).

Murekye, A., Katicha, S. W., Urbaez, E. E., Flintsch, G. W., & Diefenderfer, B. K. (2024). Pilot Study to Incorporate Network-Level Structural Condition in Agency Pavement Management Practices. *Transportation research record*, 03611981241252152.

Plati, C., Gkyrtis, K., & Loizos, A. (2024). A practice-based approach to diagnose pavement roughness problems. *International Journal of Civil Engineering*, *22*(3), 453–465.

Rasol, M., Pais, J. C., Pérez-Gracia, V., Solla, M., Fernandes, F. M., Fontul, S., Ayala-Cabrera, D., Schmidt, F., & Assadollahi, H. (2022). GPR monitoring for road transport infrastructure: A systematic review and machine learning insights. *Construction and Building Materials*, *324*, 126686.

Scavone, M., Katicha, S. W., Flintsch, G. W., & Amarh, E. (2023). On the TSD deflection velocity measurements: a revision to the current state of the art and discussion over its applicability for concrete pavement assessment. *International Journal of Pavement Engineering*, *24*(2), 2138881.

Schmalzer, P. N. (2006). Long-term pavement performance program manual for falling weight deflectometer measurements.

Scullion, T., Dorsett, G., & Liu, W. (2025). Determine the Feasibility and Methodologies of Using Structural Data from the Traffic Speed Deflectometer in Network-and Project-Level Treatment Decision-Making.

Sergeson, P. (2024). *TPF-5(385) Pavement Structural Evaluation with Traffic Speed Deflection Devices (TSDDs)*.

Sjögren, L., & Andrén, P. (2012). Rolling deflectographs and equipment's for load capacity monitoring of roads, an overview. 4th European Pavement and Asset Management System Conference, September,

Smith, K. D., Bruinsma, J. E., Wade, M. J., Chatti, K., Vandenbossche, J., & Yu, H. T. (2017). Using falling weight deflectometer data with mechanistic-empirical design and analysis, volume I.

Wang, K., Wei, J., Hou, X., & Wu, C. (2025). Evaluation of Deflection Errors in Traffic Speed Deflectometer Measurements on Inverted Asphalt Pavement Structures. *Applied Sciences*, *15*(7), 4059.

Weitzel, N. D., Pierce, L. M., Carroll, E., & Thompson Jr, J. U. (2024). Use of traffic speed deflectometer data in project-level pavement rehabilitation design. *Transportation research record*, *2678*(6), 752–763.

Wu, C., Duan, Y., & Wang, H. (2024). Signal denoising of traffic speed deflectometer measurement based on partial swarm Optimization–Variational mode decomposition method. *Sensors*, *24*(12), 3708.

Xiao, F., Xiang, Q., Hou, X., & Amirkhanian, S. N. (2021). Utilization of traffic speed deflectometer for pavement structural evaluations. *Measurement*, *178*, 109326.

Zhang, A. A., Shang, J., Li, B., Hui, B., Gong, H., Li, L., Zhan, Y., Ai, C., Niu, H., & Chu, X. (2024a). Intelligent pavement condition survey: overview of current researches and practices. *Journal of Road Engineering*.

Zhang, A. A., Shang, J., Li, B., Hui, B., Gong, H., Li, L., Zhan, Y., Ai, C., Niu, H., & Chu, X. (2024b). Intelligent pavement condition survey: Overview of current researches and practices. *Journal of road engineering*, *4*(3), 257–281.

Zhang, M., Fu, G., Ma, Y., Xiao, R., & Huang, B. (2023). Speed and temperature superposition on traffic speed deflectometer measurements. *Transportation Geotechnics*, *40*, 100990.

Zhang, M., Gong, H., Jia, X., Jiang, X., Feng, N., & Huang, B. (2022). Determining pavement structural number with traffic speed deflectometer measurements. *Transportation Geotechnics*, *35*, 100774.

Zofka, A., & Sudyka, J. (2015). Traffic speed deflectometer (TSD) measurements for pavement evaluation. Proc. of the International Symposium on Non-Destructive Testing in Civil Engineering (NDT-CE), Berlin, Germany,