THE UNIVERSITY OF TEXAS AT AUSTIN
CENTER FOR TRANSPORTATION RESEARCH

TECHNICAL REPORT 0-7031-01-1
TxDOT PROJECT NUMBER 0-7031-01

Predicting Highway Friction on an Annual Basis on Texas Network

Christian Sabillon-Orellana
Danilo Inoue
Ruohan Li
Anna Hooton
Robing Huang
Hongbin Xu
Jorge Prozzi

August 2024

Published September 2026

https://library.ctr.utexas.edu/ctr-publications/0-7031-01-1.pdf

Technical Report Documentation Page

1. Report No. FHWA/TX-27/0-7031-01-1	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Predicting Highway Friction on an Annual Basis on Texas Network		5. Report Date Submitted: June 2024 6. Performing Organization Code
7. Author(s) Christian Sabillon-Orellana, Danilo Inoue, Ruohan Li, Anna Hooton, Robing Huang, Hongbin Xu, Jorge A Prozzi		8. Performing Organization Report No. 0-7031-01
9. Performing Organization Name and Address Center for Transportation Research The University of Texas at Austin 3925 W. Braker Lane, 4th Floor Austin, TX 78759		10. Work Unit No. (TRAIS) 11. Contract or Grant No. 0-7031-01
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Division P.O. Box 5080 Austin, TX 78763-5080		13. Type of Report and Period Covered Technical Report June 2024 14. Sponsoring Agency Code
15. Supplementary Notes Project performed in cooperation with the Texas Department of Transportation.		

16. Abstract

The research project developed a network-level model to predict highway friction at the network level on an annual basis. By analyzing pavement texture data and employing advanced artificial intelligence and machine learning techniques, the study established a quantifiable relationship between surface texture characteristics and friction. Extensive field data was collected using a specialized data collection prototype across various pavement types and climate conditions, leading to a robust predictive model for friction. Traditional methods, such as regression, showed limitations, including underfitting and the need for frequent recalibration due to surface deterioration. To address these issues, the researchers used a representative database of Texas pavement surfaces, ensuring the models' statewide applicability and employed machine learning combining unsupervised and supervised learning techniques. Cluster analysis grouped pavement sections with similar characteristics then, regression models were used to predict pavement friction. These clusters served as labels to train a classification algorithm, achieving an F1 score of 0.903 with the random forest model. An advanced ensemble linear regression model captured the complex relationships between texture indices, surface information, and pavement friction, achieving an R^2_{adj} of 0.767, a mean absolute error (MAE) of 0.050, and a root mean square error (RMSE) of 0.071. This model outperformed a deep neural network regression model, which had an R^2_{adj} of 0.669, an MAE of 0.066, and an RMSE of 0.085. The algorithms were integrated into the Texan Texture Friction Forecaster application, capable of processing raw laser sensor data, computing indices, and predicting friction using the developed models.

17. Key Words Pavement surface, friction, skid number, grip number, machine learning, artificial intelligence	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161; www.ntis.gov.

19. Security Classif. (of report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of pages TBD	22. Price

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

Predicting Highway Friction on an Annual Basis on Texas Network

Christian Sabillon-Orellana
Danilo Inoue
Ruohan Li
Anna Hooton
Robin Huang
Hongbin Xu
Jorge A Prozzi

CTR Technical Report:	0-7031-01-1
Report Date:	Submitted: June 2024
Project:	0-7031-01
Project Title:	Research and Technology Implementation: Develop Efficient Prediction Model of Highway Friction on an Annual Basis on Texas Network
Sponsoring Agency:	Texas Department of Transportation
Performing Agency:	Center for Transportation Research at The University of Texas at Austin

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration

Center for Transportation Research
The University of Texas at Austin
3925 W. Braker Lane, 4th floor
Austin, TX 78759

http://ctr.utexas.edu

Disclaimers

Author's Disclaimer: The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official view or policies of the Federal Highway Administration or the Texas Department of Transportation (TxDOT). This report does not constitute a standard, specification, or regulation.

Patent Disclaimer: There was no invention or discovery conceived or first actually reduced to practice in the course of or under this contract, including any art, method, process, machine manufacture, design or composition of matter, or any new useful improvement thereof, or any variety of plant, which is or may be patentable under the patent laws of the United States of America or any foreign country.

Engineering Disclaimer

NOT INTENDED FOR CONSTRUCTION, BIDDING, OR PERMIT PURPOSES.

Research Supervisor: Jorge A Prozzi

Acknowledgments

The authors want to acknowledge the Texas Department of Transportation (TxDOT) for providing continuous support during this project, access to the data and for their constant technical advice. In particular, the authors wish to express their appreciation to Dr. Jenny Li, Dr. Enad Mahmoud, Dr. Hui Wu, and Dr. Andre Smit for participating in the progress meetings, reviewing the technical memoranda, and providing guidance, valuable comments, and feedback. Furthermore, the authors would also like to recognize the contributions and support from Lacy Peter, Epigemio Gonzalez, Aldo Madrid, Melody Galland, Ed Goebel, and Dr. Andre Smit, who provided the vast majority of the data for this project. Our special gratitude also to Wade Odell, RTI Project Manager, for his constant managerial support during the duration of the project.

Table of Contents

List of Figures

List of Tables

List of Terms

2D	Two-Dimensional
3D	Three-Dimensional
AB	Adaptive Boosting
AGNES	Agglomerative Nesting
ASTM	American Society for Testing and Materials
AOI	Area of Interest
BA	Bootstrap Aggregating
CalLoad	Calibration Loaded
CM	Mean of Cross Width
CPI	Consumer Price Index
CSTD	Standard Deviation of Cross Width
CSV	Comma Separated Value
DEFCAM	Default Camera
DNN	Deep Neural Network
DS1	Ground Truth Dataset
DS2	Holdout Dataset
DT	Decision Tree
EXP	Exposure
FPC	Flexible Pavement Cluster
GB	Gradient Boosting
GLM	Generalized Linear Models
GN	Grip Number
GNB	Gaussian Naïve Bayes
HCME	Mean of Local Curvature
HCST	Standard Deviation of Local Curvature
HMA	Hot Mix Asphalt
HSTS	High-Speed Texture Sensor
ISO	International Organization for Standardization
KNN	K Nearest Neighbors
LOG	Logistic Regression
MPD	Mean Profile Depth
MTD	Mean Texture Depth
NASEM	National Academies of Sciences, Engineering, and Medicine
NOCAM	No Camera
NPV	Net Present Value
PCA	Principal Component Analysis
PCC	Portland Cement Concrete
PFC	Porous Friction Coarse
RAL	Autocorrelation Length
RDT	Maximum Absolute Gradient
RF	Random Forest
RKU	Kurtosis of Height
RMS	Root Mean Square
SMA	Stone Matrix Asphalt
SNCAM	Serial Number Camera

SN	Skid Number
SNN	Shallow Neural Network
SOFA	Sabillon-Orellana Filtering Algorithm
ST	Surface Treatment
SV2	Two-Point Slope Variance
SVM	Support Vector Machine
TSD	3D frame file
TXDOT	Texas Department of Transportation
UL	Unsupervised Learning
XGB	Extreme Gradient Boosting

Chapter 1. Experimental Design

The data collected during this project covered a diverse array of pavement surface types commonly found in Texas, including hot-mix asphalt (HMA), surface treatments (ST), and various plastic and hardened Portland cement concrete (PCC) pavement texturing methods. Researchers surveyed over 170 sites using a prototype data collection system. These sites were geographically distributed across the state, with the majority of sections located in six Texas Department of Transportation (TxDOT) districts: Atlanta, Austin, El Paso, Houston, Laredo, and Lubbock. Additionally, ten other TxDOT districts—Amarillo, Bryan, Fort Worth, Odessa, San Angelo, San Antonio, Tyler, Waco, Wichita Falls, and Yoakum—contributed locations with pavement surfaces of interest for this project.

1.1. Work Plan

The data collection plan aimed at creating a comprehensive database of various pavement surfaces in Texas, capturing key variables such as pavement type (flexible vs. rigid), texture (positive vs. negative), surface information (mix type, surface treatment, or texturing technique), materials (aggregate type or size), and age (different stages of service life). Outreach efforts to multiple TxDOT district offices across North, East, West, South, and Central Texas provided diverse highway section locations, including HMA, PCC, and various ST. The study prioritized newly constructed pavements to ensure the accuracy of surface information but also included older pavements that had not undergone maintenance.

Using an enhanced prototype data collection system developed in-house, researchers collected data and divided it into two subsets: the ground truth dataset (with known surface types provided by the TxDOT's personnel) and the holdout dataset (with unknown surface types). The ground truth dataset was employed in the development stages of the empirical models, while the holdout dataset was used to evaluate the model's performance.

A "testing site" was defined as a section of highway exceeding 0.4 miles, where the equipment operated continuously. A "pavement section" was defined as a 0.1-mile segment of the highway, designed to mimic the length of pavement sections used in Texas' pavement management system. Data collection spanned 400 miles of pavement texture, with at least 300 miles on flexible pavements and 100 miles on rigid pavements. Of the total data, 219 miles (54%) were included in the ground truth dataset for model development, and 186 miles (46%) were in the holdout dataset for model testing.

1.2. Prototype System for Data Collection

The research team collected data for this study using an enhanced prototype system that integrated a friction measuring device (i.e., GripTester), a line laser scanner, a high-speed digital camera, and a light source. This prototype measured the full spectrum of pavement macrotexture (wavelengths from 50 mm to 0.5 mm) and a portion of microtexture (wavelengths shorter than 0.5 mm) at speeds exceeding 45 mph along the left wheel path. The system simultaneously collected high-resolution images of the pavement surface and friction measurements in the same area scanned by the laser. Although the high-resolution images were not extensively used in this project, they served to confirm pavement surface type and condition.

1.2.1. Data Collection Prototype System Components and Assembly

Passenger vehicles, such as sedans and small pickup trucks, were initially used to transport the prototype system in this study. However, these were discarded due to significant vertical displacement issues while in motion, causing excessive bouncing, blurry images, abnormal sensor readings, and unwanted vibrations in the laser signal. Consequently, the researchers selected a heavy-duty pickup truck with a robust build, strong suspension, and reliable towing capacity for field testing. Despite being more than two feet above the ground, the stiff suspension of these trucks proved resilient against large vertical movements. However, the prototype system design had to accommodate the height variations of different vehicles, which ranged from one to two feet.

The solution involved a fully customizable mounting system, allowing users to attach the friction measuring trailer to measure the inner wheel path, right wheel path, or the centerline, as illustrated in **Figure 1.** The friction trailer was consistently affixed to measure the inner wheel path for safety reasons, despite a study by Goenaga et al. (1) indicating that the outer wheel path typically exhibits the lowest friction. This decision ensured the trailer remained secure and did not risk sliding off an embankment when the vehicle needed to stop on the shoulder or side of the road.

***Figure 1**: Mounting piece used to adjust the position of the friction trailer.*

The prototype incorporates 1.57 x 1.57 in. aluminum profiles, providing a versatile framework for vertically and laterally positioning the system's sensor components in relation to the vehicle. The system is secured to the test vehicle using the rear hitch of a steel mounting piece. A robust 2 x 2 in. steel square tubing is inserted into the hitch and firmly fastened with a heavy-duty bolt. On the opposite end, a vertical 2 x 2 in. square tube with a T-shape serves as the attachment point for the aluminum profile's main body. Error! Reference source not found. shows the black steel tubing entering the hitch receiver on the left and the aluminum profile's main body securely fastened to the horizontal part of the T-shaped steel tube on the right.

The horizontal aluminum profile serves as a versatile rail for mounting sensors, with each component, such as the line laser and the camera/light, having its own set of aluminum profiles for independent adjustments, as shown in **Figure 3**. Like the flexible black mounting piece in **Figure 1**, these profiles allow alignment across the vehicle's width and facilitate height adjustments of up to 3 feet, addressing the challenge of varying vehicle heights.

Precision is crucial in mounting the sensors to ensure accurate stand-off distances. The laser sensor requires a height of 17.5 in. from the camera window to the surface, while the camera and light systems need a height of 24 in. from the camera lens to the surface. This setup accommodates any vertical movement of the host vehicles. Sensor height measurements should be taken on a level, ideally paved, surface, and should account for the weight of passengers or additional load in the vehicle.

Figure 2: *Side and rear view of the mounting system.*

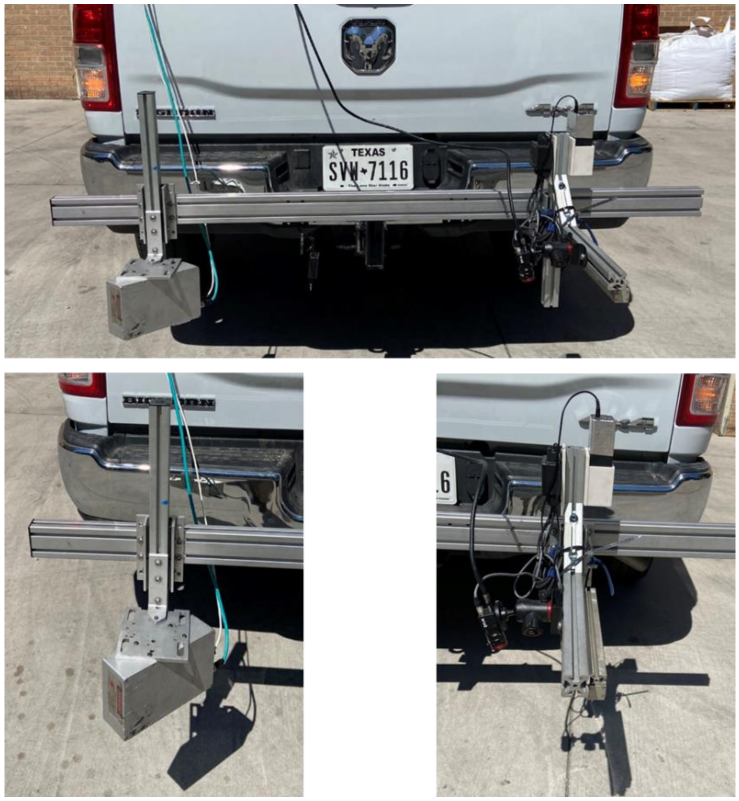


Figure 3: *System mounted onto the vehicle (top); laser sensor (bottom right); and camera and light system (bottom left).*

The electrical wiring of the system's components is tailored to each sensor's specific power and data communication needs. The line laser, which requires 12 V and 0.7 amps, is powered through the vehicle's plug and connected to the laptop via a 20-foot Cat 6E ethernet cable. The area camera operates on a 5 V supply and connects to the laptop with a 20-foot USB 3.1 cable, providing both power and data transmission. The lighting system uses a 24 V power source from a dedicated battery pack. **Figure 4** illustrates the wiring for the camera and light source, including their respective pins and color-coded wires. To ensure safety, a waterproof enclosure protects the wires, reinforced with specialized wire connectors.

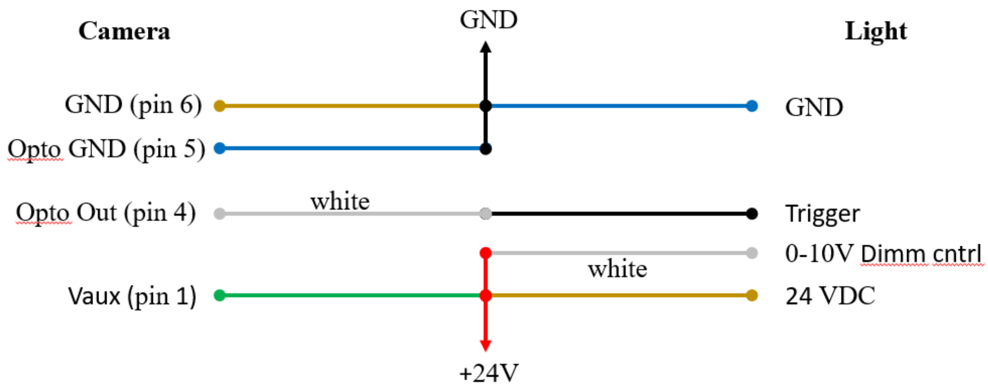


Figure 4: *Diagram to wire the camera and light source.*

Finally, synchronization between the area camera/light, the line laser sensor, and the friction trailer is achieved through the wheel encoder in the GripTester. This encoder, connected to the GripTester's driving wheel, is programmed to activate the laser and the camera/light sequentially, triggering approximately every 44 mm of forward travel. **Figure 5 s**hows the fully assembled prototype.

Figure 5*: Fully assembled prototype during daytime (left) and nighttime (right) testing conditions.*

1.2.2. Sensor Specifications

1.2.2.1. Friction Measuring Device

The GripTester was chosen for measuring friction due to its broad testing speed range of 3 to 80 mph, excellent repeatability and reproducibility, and complete operator control over water flow—crucial for achieving a specified water thickness between the test tire and the pavement surface. Following ASTM E2340 (2) standards, the device uses a smooth-thread measuring wheel with a 10 in. diameter, emulating a fixed wheel with a continuous 15 percent slippage. The measurement tire features an automatic watering system, ensuring precise water thickness as set by the operator. For this study, the water thickness was set at 0.5 mm. Once the water thickness was established, the system dynamically adjusted the water flow based on traveling speed. In this study, the speed was fixed at 50 mph, so no dynamic adjustment was required. The GripTester provided friction measurements in terms of the Grip Number (GN), ranging from a theoretical zero (no friction) to one (maximum friction).

1.2.2.2. Texture Measuring Sensors

This section outlines the minimum requirements for all texture-related sensors used during data collection. These specifications serve as baseline criteria for capturing future data with alternative sensors while maintaining high-quality standards. Derived from extensive testing of diverse hardware configurations, real-world data collection scenarios, and evaluations under various conditions, these parameters account for variations in vehicle heights, ambient factors, and road surface types. They are crucial for developing a system that operates seamlessly at speeds up to 80 mph, in both intense brightness and low-light conditions, across various pavement surfaces.

1.2.2.2.1. Line Laser Scanner

The in-house built line laser sensor in this study was assembled, calibrated, and tested to surpass commercially available options. An alternative laser sensor must have a working distance precisely set at 400 mm, with a tolerance of ± 75 mm to accommodate vehicle movement. To maintain accuracy, a stiff suspension system is optimal. With the working distance and Z-range set, the height resolution (vertical axis) is 3.8 micrometers, with any resolution below 5 micrometers considered acceptable. **Figure 6** illustrates a diagram showing the axis and direction conventions. The ideal field of view ranges from 142 to 190 mm, capturing a 100 mm length transversely at a 45-degree angle relative to the direction of travel. **Figure 6** illustrates the axis and direction conventions.

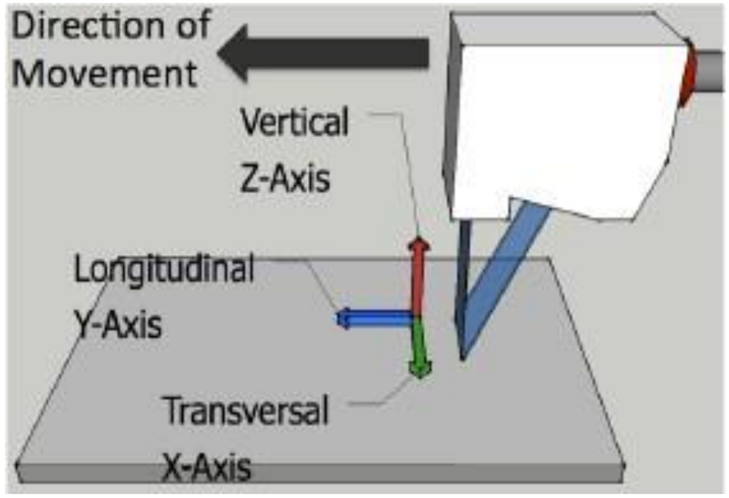


Figure 6: *Axis convention and direction of movement (3).*

The laser's wavelength is optimally set at 810 nm or higher to reduce interference from the solar spectrum, which peaks around 500 nm. A minimum optical power of 2.3 W is required to ensure discernible profiles on dark pavement surfaces, enabling the laser sensor's camera to operate with exposure times as brief as 3-5 microseconds. For optimal performance and precision, incorporating a minimum of 2,000 data points per scan (profile) is recommended.

An external trigger source is suggested to capture profiles at set intervals of distance traveled, regardless of vehicle speed. In this study, profiles were captured, on average, every 44 mm, with the transistor-transistor logic trigger sending a pulse to initiate profile collection. The trigger distance is adjustable and tied to the laser camera's sampling cycle, not the area camera, indicating the speed at which the camera captures profiles.

The laser sensor is designed to operate at speeds up to 80 mph, collecting data every 44 mm, requiring a sampling rate of 812 Hz. Data collection in this study occurred at 50 mph, equivalent to 508 Hz. The area of interest (AOI) is a defined window within the camera image focused on identifying the laser line, enhancing sampling cycles along with the Z-range. Maintaining a minimum of 608 pixels within the AOI is essential for optimal performance.

1.2.2.2.2. Area Camera

The area camera, Blackfly-S, was chosen for its high frame rate, low exposure time, and the global shutter. At a travel speed of 50 mph, the blur would be approximately 5.1 mils, resulting in well-defined images. Should a different camera be used, then the area camera chosen must have a high frame rate (≥ 50 FPS), low exposure time (≤ 10 microseconds), and a global shutter. At a travel speed of 50 mph, the blur should be no more than 6 mils to obtain well-defined images. The area camera should have a lens capable of obtaining an area of approximately 15 x 15 cm from a height of 50 cm. For instance, a 2/3" lens with a focal length of 12 mm, various iris stops (F1.4/F4/F8/F16) would be sufficient.

1.2.2.2.3. Light Source

The light source used was Metaphase's TxBAR, due to its practicality, great customer service, and synchronization with the area camera. Should a different light be chosen the light source used should provide enough light to be the main source of lighting for the main camera. This can be a continuous ON light source or a light burst synchronized with the area camera. Using bursts of light to achieve maximum luminosity is preferable because it enhances perceived brightness through temporal summation and perceptual contrast, making light appearing more intense than a constant light source. Additionally, bursts of light improve energy efficiency and heat management, extending the lifespan of the light source and reducing the need for extensive cooling systems.

Table 1 summarizes the specifications for all three types of sensors previously discussed.

Table 1: General Specifications for the sensors used in the data collection.

Line Laser	
Working Distance	400-450 mm
Z-Range (+/-)	75 mm
Resolution Height	< 5 micrometers
Field of View	142 - 190 mm
Wavelength	810 nm
Optical Power	2.3 W
Data Points	> 2000
Sampling Cycle	≥ 600 Hz
Resolution Lateral	< 100 micrometers
Trigger	TTL
AOI	608 pixels
Area Camera	
Resolution	≥ 1280 × 1024
Frame Rate	≥ 50
Megapixels	≥ 1.3
Chroma	Color
Operating Temperature	0° to 50°C
Readout Method	Global shutter
ADC	10-bit
Exposure Range	≤ 10.0 µs
Interface	USB 3.1 or Ethernet
Light Source	
Color	White, 6000K
Operating Temperature	0° to 40°C
IP Rating	IP64
Intensity	≥ 100,000 Lux

Chapter 2. Field Testing

The field testing program was developed to expand the dataset from TxDOT Project 0-7031 (4) by creating a more comprehensive and representative database of common Texas pavement surfaces. This new database captured factors influencing pavement friction, such as variations in texture, aggregate gradations, asphalt mixes, and concrete texturing methods. Including these types of surfaces enhances the robustness and accuracy of the friction models. Collaboration with various TxDOT district offices across Texas was crucial for timely data collection.

The researchers focused on new pavement sections for accurate surface type, age, and aggregate source information but also collected data from older pavements, over a decade old. Due to difficulties in obtaining accurate information on pavement age, aggregate source, asphalt binder content, water-cement ratios, and chemical admixtures, these factors were excluded from the study. However, their consideration is recommended for future models.

The data collection involved the use of the prototype system discussed in Section 1.2.1 (Data Collection Prototype System Components and Assembly), shown in **Figure 7**. This system measures friction continuously along the inner wheel path at 50 mph, while simultaneously collecting texture profiles and pavement surface images. Note that the system is also capable of nighttime and daytime data collection thanks to its lighting system.

Figure 7: *Data collection prototype system during daytime (left) and nighttime (right) conditions.*

The final database contains over 400 miles of texture and friction data. With these data, two data subsets were established: one with known pavement surface types, provided by TxDOT personnel, and a second with unknown surface types, serving as a holdout dataset for further testing and validation of the developed algorithms.

This approach not only aids in model development but also offers a preview of the potential real-world application of the friction prediction algorithm.

2.1. Logistics and Locations of Test Sites

The researchers contacted materials and maintenance personnel across multiple TxDOT districts to obtain accurate "ground truth" information on a range of pavement surfaces. A list of at least ten highway sections with distinct pavement surfaces was requested, including:

- **Hot and Warm Mix Asphalt Surfaces**: dense, gap-graded, and open-graded mixes.
- **Surface Treatments**: various applications such as seal coats of different grades, high friction treatments, microsurfacing, and thin overlay mixes.
- **Portland Cement Concrete Surfaces**: rigid pavements with different finishes such as carpet or burlap drag, longitudinal or transverse tining, diamond grinding, and grooving.

These categories represent the majority of Texas highway surfaces, providing a strong basis for developing empirical friction prediction models. Initially, the contract specified collecting data from 300 miles of different pavement sections across six districts (Atlanta, Austin, Bryan, Dallas, Houston, and San Antonio), with each district contributing about 50 miles. The project team surpassed this target by collecting over 400 miles across more than six districts, thanks to the enthusiasm and support of the TxDOT districts involved.

Below is a list of the six TxDOT districts that provided the most information and resources, along with the supporting personnel who assisted in data collection and identifying pavement sections with known surfaces.

- **Atlanta District:** Lacy Peters and Cody Fuller
- **Austin District:** Andre Smit, Andy Naranjo, and John Wirth
- **El Paso District:** Aldo Madrid, Mauricio Esquivel Lara, Monica Ruiz, Anthony Marquez, Christopher J. Weber
- **Houston District:** Melody Galland, Viet Pham, Juan Fuentes, John Zientek III
- **Laredo District:** Epigemio Gonzalez and Jesus Saavedra

- **Lubbock District:** Ed Goebel and Mike Stroope

The following list of districts and people indicates the other supporting districts that provided aid either with the data collection process or by providing locations of other pavement surfaces that could be of interest to the project:

- **Amarillo District**: Cody Harris
- **Bryan District**: Joel Withem
- **Fort Worth District**: Tom Brown
- **Odessa District**: Zane Honeyfield
- **San Angelo District**: Evan Jones and Lonnie Green III
- **San Antonio District**: Alejandro Miramontes and Jose Emilio Ramos
- **Tyler District**: Dustin Morgan
- **Waco District**: Jerrod Swift
- **Wichita Falls District**: Brian Moore
- **Yoakum District**: Bradley Polasek

The spread of the test sites is portrayed in **Figure 8** which shows three maps of Texas highlighting the TxDOT districts surveyed (left), the counties that had at least one site surveyed (middle), and the distribution of these sites on a map of Texas using blue circles (right).

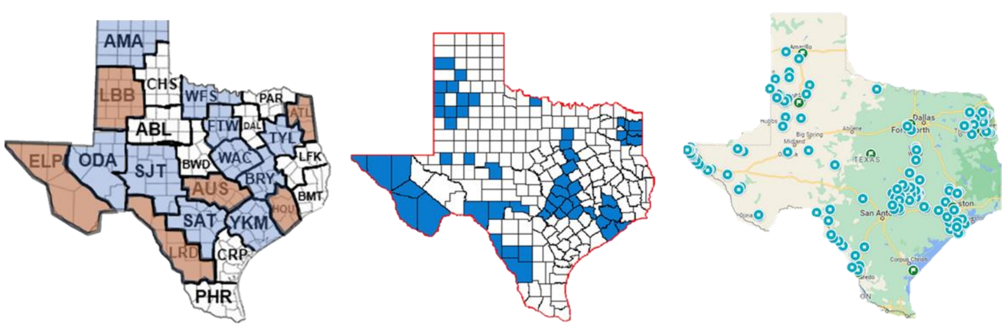


***Figure 8**: Testing sites showing TxDOT districts (left), counties (center) and specific locations (right).*

2.2. Data Collection Protocol

Before conducting fieldwork, the project team remotely surveyed designated sites using Google Street View and Pathweb video recordings. This strategy allowed accurate identification of pavement sections start and end points and assessment of visible distresses such as raveling, bleeding, cracking, patching, rutting, and punchouts. This preliminary information helped plan the survey route and coordinate with district personnel for scheduling and on-site support during data collection, such as refilling the water tank for friction measurement.

At the testing site, the survey vehicle was parked half a mile from the section's starting point. Researchers assembled the equipment, ensured all systems were properly wired and secured, activated hazard lights, and recorded the pavement's surface temperature (taking at least three measurements). The vehicle then accelerated to 50 mph in the rightmost lane. Upon reaching the desired speed and the pavement surface of interest, sensors were triggered to start data collection. The team monitored equipment performance until data collection was completed. Afterward, the vehicle was parked, and the equipment was dismounted for transport. Each day, all raw data were securely backed up to preserve information integrity for analysis.

The pavement sections analyzed for friction characteristics were not always flat and straight sections. Some data collection occurred on surfaces with varying cross-slopes and grades, under different lighting conditions, and without traffic control as data collection was conducted at highway speeds. This approach aimed at replicating real-world conditions encountered during pavement evaluations by the TxDOT's contracted vendors. Additionally, the GripTester used a smooth testing tire per ASTM E2340 (2) with a 12% slip and a water film thickness of 0.5 mm.

Texture measurements were taken along the inner wheel path of the rightmost lane using a laser sensor oriented at a 45-degree angle to capture features like tining, grinding, or grooving on rigid pavements irrespective of their orientation on the pavement, as depicted in **Figure 9**. The laser sensor, designed for high-resolution capture at 50 mph, produced raw profiles with 2,046 data points and an average point spacing of 0.18 mm, varying with sensor height. Profiles were spaced, on average, 44 mm apart, and the sensor was mounted at an initial height of 17.5 in., adjustable during movement. At this height, the sensor had a vertical resolution of 5 microns.

Figure 9: *Laser sensor shown from a top-view*

Photographs of the pavement surfaces were taken near the inner wheel path without interfering with the laser sensor. The camera was positioned about 12 in. to the right of the inner wheel path center, with the focus adjusted for optimal resolution. The camera, laser sensor, and GripTester were synchronized to ensure consistent and precise data collection over the same areas and timeframes. It should be noted that the photographs were not extensively used in this project, they served primarily to confirm pavement surface information.

2.3. Pavement Surfaces Surveyed

The researchers collected a wide variety of distinct pavement surfaces commonly found in Texas. **Table 2** summarizes the pavement types identified by the TxDOT personnel and tested by project team. These labels guided the development of the empirical models for friction prediction. In a few instances (less than 5% of the dataset), the surface type information provided did not match the actual site conditions. For example, a site expected to have a high friction surface treatment was found to have a rigid pavement with transverse tining. These mismatches were added to the holdout dataset for proper labeling afterward, rather than being discarded.

Table 2: Summary of all the pavement surfaces provided by TxDOT

Flexible Pavements	Rigid Pavements
Dense Graded Mixes (C, D)	AstroTurf or Carpet Drag
Thin Overlay Mix	Longitudinal and Transverse Tining
Superpave (C,D)	Fixed and Random Tining
Stone Mastic Asphalt (Rubber, D)	Conventional Diamond Grinding
Permeable Friction Coarse	New Generation Diamond Grinding
Thin Bonded Wearing Coarse	Exposed Aggregate
Chip Seals (Grades 3, 3S, and 4)	
Fog Seals	
High Friction Surface Treatments	
Microsurfacing	

2.4. Sections Mileage

This subsection provides a summary of the data collection in terms of mileage. **Table 3** details the mileage collected by pavement type. **Table 4** breaks down the mileage by dataset subsets: the ground truth and the holdout dataset. The ground truth dataset includes pavement sections where the surface matched TxDOT's information with visual inspection. The holdout dataset consists of sections with unknown surfaces, either due to mismatched information or sections identified as of interest by the project team.

Table 3: Breakdown of mileage for data collected in terms of pavement type.

Type	Mileage (percentage)
Flexible Pavement	313.7 (73.73%)
Rigid Pavement	111.8 (26.27%)
Total	425.5 (100.00%)

Table 4: Breakdown of mileage for data collected in terms of dataset.

Type	Mileage (mi.)
Ground Truth	238.4 (56.03%)
Hold-out	187.1 (43.97%)
Total	425.5 (100.0%)

Figure 10 summarizes the mileage collected broken down by the TxDOT district in which they were collected. The district where the most data was collected was

El Paso with almost 80 miles, whereas the district with the least amount of data collected was Tyler with 3 miles. **Figure 11** shows the mileage broken down by the type of pavement surface that was reported by each district. Note that different districts provided varying levels of detail when specifying the type of pavement surface. For instance, some districts that provided seal coat locations simply stated that the pavement at that site was a seal coat. Other districts specified the grade of the aggregates, their source, whether the aggregate was lightweight, or a combination of all of the above.

For simplicity, **Figure 11** has grouped all the pavements within the ground truth database into broader categories. Dragged surfaces include Astroturf, carpet and burlap drag. Diamond grinding includes conventional and new generation diamond grinding. Gap-graded mixes is composed of stone matrix asphalt (SMA) mixes. Open-graded mixes encompasses all the PFCs that were surveyed. Tinning accounts for transverse and longitudinal tining as well as fixed and random tining. Dense-graded mixes cover the coarse and fine graded mixes as well as Superpave mixes. Lastly, surface treatments encompass all flexible pavements consisting of seal coats, microsurfacing, high friction surface treatment and thin overlay mixes.

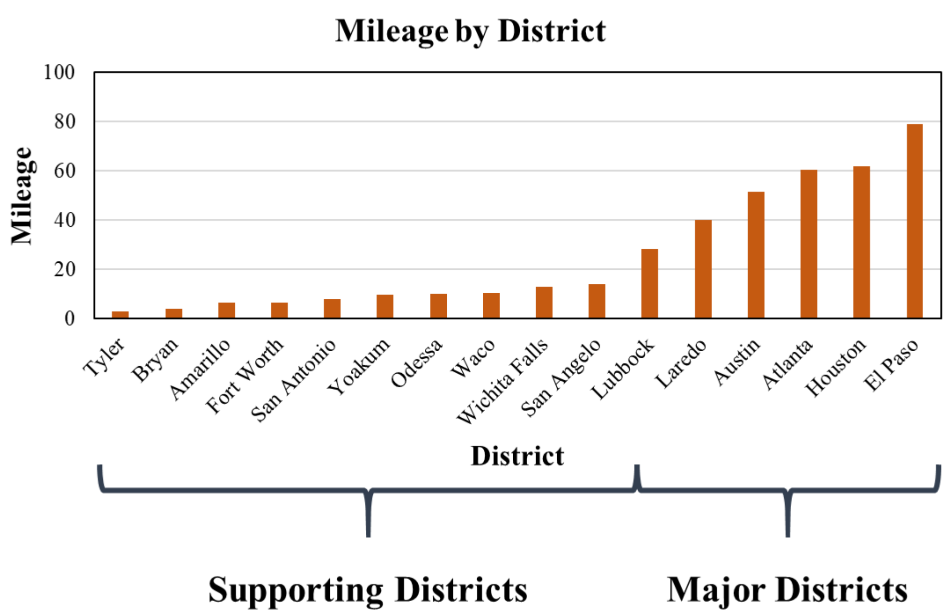


Figure 10: *Distribution of mileage surveyed broken down by TxDOT district.*

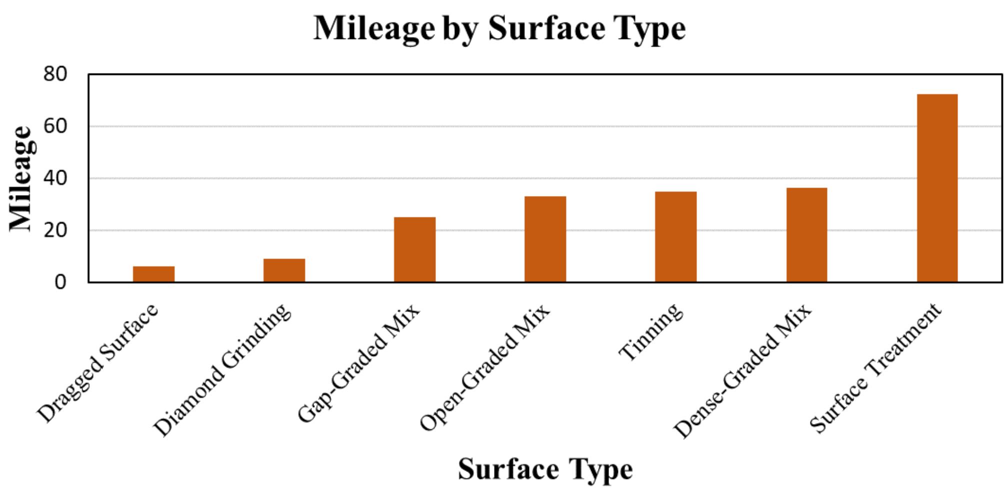


Figure 11: *Distribution of mileage surveyed broken down by pavement mix/surface type.*

2.5. Comparison between GripTester and Locked Wheel Skid Tester

The project team collected friction measurements using a GripTester, however, TxDOT uses the Locked Wheel Skid Tester (LWST) for their friction assessment of the highway network. Therefore, Grip Number (GN) measurements from the GripTester were compared to Skid Number (SN) measurements from the LWST. LWST data was provided by TxDOT's 2023 skid survey conducted at the same locations as the project data collection. Occasionally, there was a significant time lag between the LWST and GripTester measurements due to the coordination challenges, as the skid measuring equipment was often in use or being calibrated. Unlike the previous comparison in TxDOT project 0-7031, both the LWST and GripTester measurements were collected along the same wheel path this time. There were 201 sections of 0.1 miles with overlapping GN and SN data.

Table 5 shows the average skid readings from the GripTester and LWST across different rigid pavements, while **Table 6** shows the same information for flexible pavements of varying mix types and surface treatments. On average, the GN tends to be higher than the SN. This is likely because the LWST fully locks its measuring wheel, whereas the GripTester uses a 12% fixed slip, resulting in more grip. This trend holds for most pavements except for open-graded mixes, such as PFCs, and surface treatments like seal coats. It is possible that the LWST had difficulty fully locking its wheel on these pavements.

Table 5: Comparison of GN and SN measurements for rigid pavements

Surface	Highway	Average GN	Average SN
Transverse Tining	IH0010	0.34	0.20
	IH0010	0.30	0.13
	IH0010	0.39	0.18
	IH0010	0.50	0.35
	US0059	0.59	0.41
	US0059	0.53	0.33
	US0059	0.55	0.41
	US0287	0.38	0.16
	US0287	0.43	0.23
	IH0035	0.40	0.23
Longitudinal Tining	IH0069	0.41	0.19
	IH0069	0.47	0.25
Carpet Drag	US0287	0.38	0.28
Next Generation Diamond Grinding	IH0010	0.28	0.20
	IH0010	0.34	0.33
	US0290	0.42	0.45
	IH0010	0.34	0.26
Diamond Grinding	IH0010	0.30	0.20
	SH0035	0.44	0.38

Table 6: Comparison of GN and SN measurements for flexible pavements

Surface	Highway	Average GN	Average SN
Dense-Graded Mixes	SH0020	0.58	0.40
	SH0020	0.59	0.40
	FM2004	0.50	0.42
	IH0020	0.37	0.25
	SH0071	0.42	0.35
	SH0071	0.42	0.33
	SH0081	0.53	0.39
	SH0158	0.33	0.25
Gap-Graded Mixes	IH0035	0.32	0.31
	IH0035	0.32	0.29
	IH0035	0.33	0.26
	IH0035	0.37	0.28
Open-Graded Mixes	SH0035	0.34	0.41
	SH0035	0.35	0.46
	SH0035	0.42	0.48
	SH0035	0.36	0.50
	IH0035	0.39	0.48
	IH0020	0.33	0.35
Surface Treatments	FM2644	0.52	0.57
	US0277	0.47	0.49
	US0277	0.46	0.49
	US0385	0.41	0.48
	US0287	0.40	0.42

2.6. Conversion from GN to SN

The researchers conducted a detailed regression analysis between GN and SN. The data indicates a linear relationship between GN and SN that depends on the type of pavement surface. This suggests that the rate of increase of friction with texture (slope of the regression line) differs for various pavement surfaces.

Figure 12 illustrates four relationships identified in the analysis: two for rigid pavements and two for flexible pavements. While these could be modeled using four separate equations, the research team proposed a more concise approach. The researchers suggested using a single multiple regression equation to model these four best-fit lines, as shown in **Equation (1).**

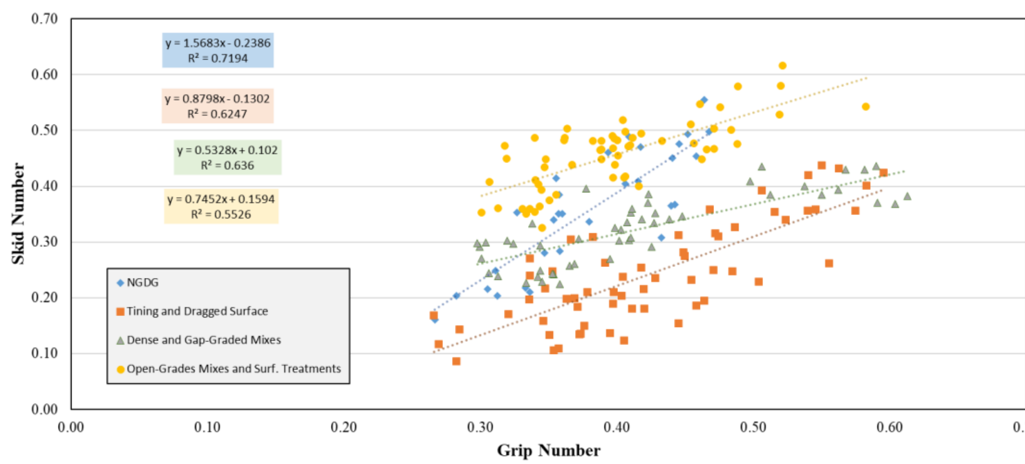


Figure 12: *Scatterplot of the average GN versus the average SN.*

$$\hat{Y} = \gamma_0 + \gamma_1 \boldsymbol{X_1} + \sum_{j=2}^{4} \gamma_{2,j} \boldsymbol{X_{2,j}} + \sum_{j=2}^{4} \gamma_{3,j} \boldsymbol{X_1 X_{2,j}} + \epsilon \qquad \textbf{(1)}$$

$$J \in \{1,2,3,4\}$$

$$X_{2,j} \begin{cases} 1 & if\ j = j \\ 0 & else \end{cases}$$

In this equation, $\hat{Y}$ represents the predicted SN, the parameters $\gamma_0, \gamma_1, \gamma_{2,j}, \gamma_{3,j}$ are all regression coefficients. J denotes the number of distinct pavement surface groups, with four groups in this dataset. Group 1, the baseline, includes dense and gap-graded mixes for flexible pavements. Group 2 comprises open-graded mixes and surface treatments for flexible pavements. Group 3 includes all diamond grinding texturing techniques for rigid pavements, while Group 4 encompasses all tining and dragged surface finishes for rigid pavements. The variable X_1 represents the measured GN. X_2, j is a dummy variable for the surface type, equal to one for a specific surface type and zero otherwise. The interaction term $X_1 X_{2,j}$ allows the slopes of different surfaces to vary, and ϵ is the error term in the regression. A summary of the regression analysis is provided in **Table 7**.

The parameter γ_0 represents aspects of SN not captured by the regression model and serves as the y-intercept for the best-fit line for Group 1, accounting for factors like vehicle wander or time lag differences. The parameter γ_1 captures the effect of a unit increase in GN on SN for Group 1. The parameter $\gamma_{2,j}$ represents the average change in SN for surface types in Group j relative to Group 1. Visually, the sum of $y_{2,j}$ and y_0 is the y-intercept for Group j. The parameter $y_{3,j}$ represents the differential effect of GN on SN for Group j relative to those of Group 1. Summing $y_{3,j} + \gamma_1$ gives the slope of the best-fit line for Group j's surface types.

The data indicates that surface types in Groups 1 and 2 have statistically identical y-intercepts, while both rigid pavement groups (Groups 3 and 4) have smaller y-intercepts than the flexible pavements. Additionally, all best-fit lines have different slopes, showing that SN changes at different rates depending on the pavement surface. On average, for every unit increase in GN, the SN increases by 0.48, 0.83, 0.88, and 1.57 units for Groups 1, 2, 4, and 3, respectively. This means that SN was lower than GN by these percentages for the first three groups, whereas, for Group 4, SN was 57% higher than GN. The adjusted R^2 value of 0.824 indicates a good fit to the data. A visualization of the predicted versus measured SN is shown in **Figure 13**.

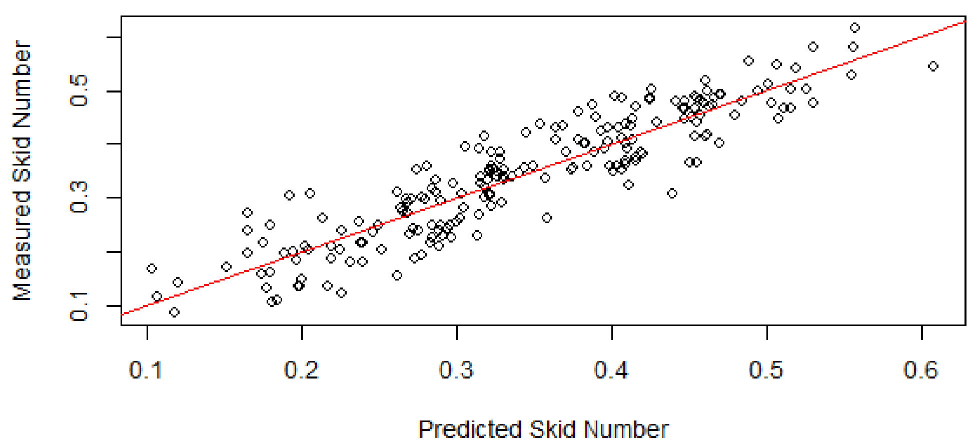


Figure 13: *Scatterplot shows the measured and predicted SN for the model presented in Table 2.2.*

Table 7 Linear regression summary analysis for transforming GN to SN

Regression Summary		
Variable	Coefficient (Std. Error)	P-value
Intercept	0.124 (0.025)	0.000*
GN	0.483 (0.060)	0.000*
Group^φ		
Group1 (Baseline)	~	~
Group2	~	~
Group3	-0.362 (0.065)	0.000*
Group4	-0.254 (0.041)	0.000*
Interaction^φ		
Group 1 (Baseline)	~	~
Group2	0.349 (0.023)	0.000*
Group3	1.085 (0.167)	0.000*
Group4	0.397 (0.097)	0.000*

Goodness-of-fit	
Metric	Numerical Value
Multiple R^2	0.829
Adjusted R^2	0.824
Residual Standard Error	0.048

* Indicates that the value is lower than 0.001

^φ Group 1 is dense and gap-graded mixes, Group 2 is open-grades mixes and surface treatments for flexible pavements, Group 3 is all diamond grinded surfaces, and Group 4 is all tinned and dragged and surfaces.

It should be noted that this linear transformation on the GN will not affect the goodness-of-fit statistics when regressing the predicted SN on texture and surface information parameters. This is because linear transformations on the predicted variable do not impact the correlation between variables, only changing the magnitude of the regression coefficients. However, the accuracy of the final friction prediction model, measured by the residual standard error, will be influenced by the error term of this model.

Chapter 3. Data Analysis

3.1. Preliminary Quality Control

Before processing the raw data, meticulous inspections were conducted to remove any invalid measurements. For the measured friction, the GripTester provides a summary file for each survey, including distance traveled, GN measurements every meter, and continuous measurement of load, speed and water flow. Based on preliminary testing, a GN was considered valid if:

- The load on the GripTester device was between 100 and 400 N.
- The water flow to the wheel was within ± 5 liters per meter from the mode.
- The surveying vehicle's speed was between 47 and 53 mph.

The first filter excluded measurements taken when the GripTester encountered potholes, debris, or other pavement distresses causing the device to bounce and the wheel load to fluctuate. The second filter ensured sufficient water flow to maintain a water thickness of 0.5 mm. The last criterion excluded measurements taken before the survey vehicle reached the appropriate testing speed. **Figure 14** shows a quality control graph where valid GN measurements meeting all criteria are highlighted in black.

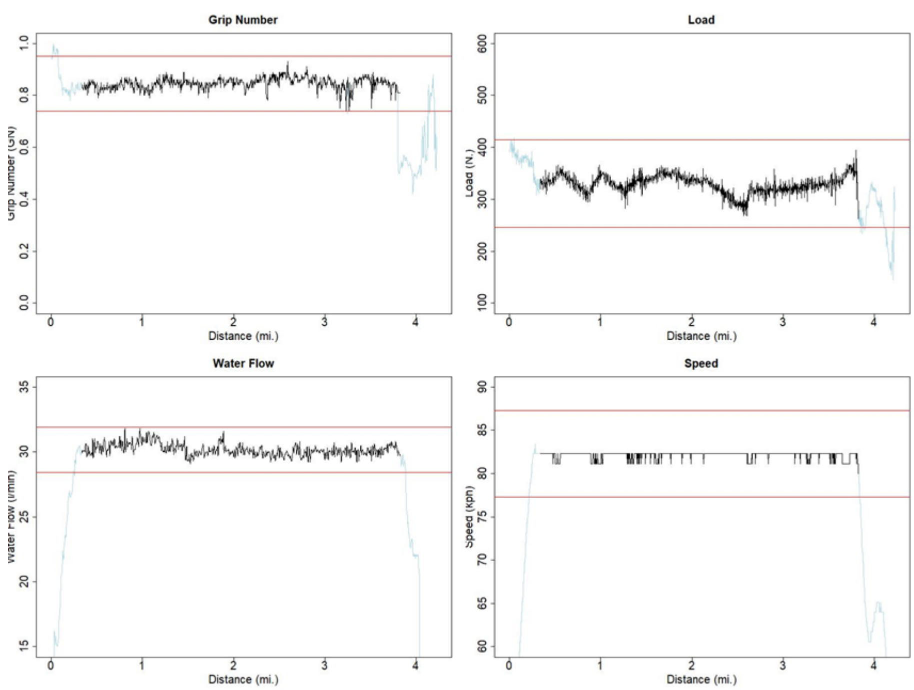


Figure 14: *Implemented Grip Tester quality controls.*

The texture data collected with the laser sensor were first filtered to retain only data corresponding to valid GN information. If friction data for a pavement section were invalid and filtered out, the corresponding texture data were also excluded. This process ensured the researchers worked only with valid friction and texture measurements. Next, the scan area was trimmed to isolate the center width of 150 mm, corresponding to the GripTester's wheel path, allowing for the computation of the mean profile depth (MPD). In terms of length, this trimming removed the leftmost and rightmost 89 mm of the cross-sectional profile.

3.2. Texture Data Processing

The researchers conducted an extensive review of texture processing and pavement surface characterization literature, revealing multiple methods for processing raw 2D surface profile data. A summary of these techniques is compiled in the diagram shown in **Figure 15**.

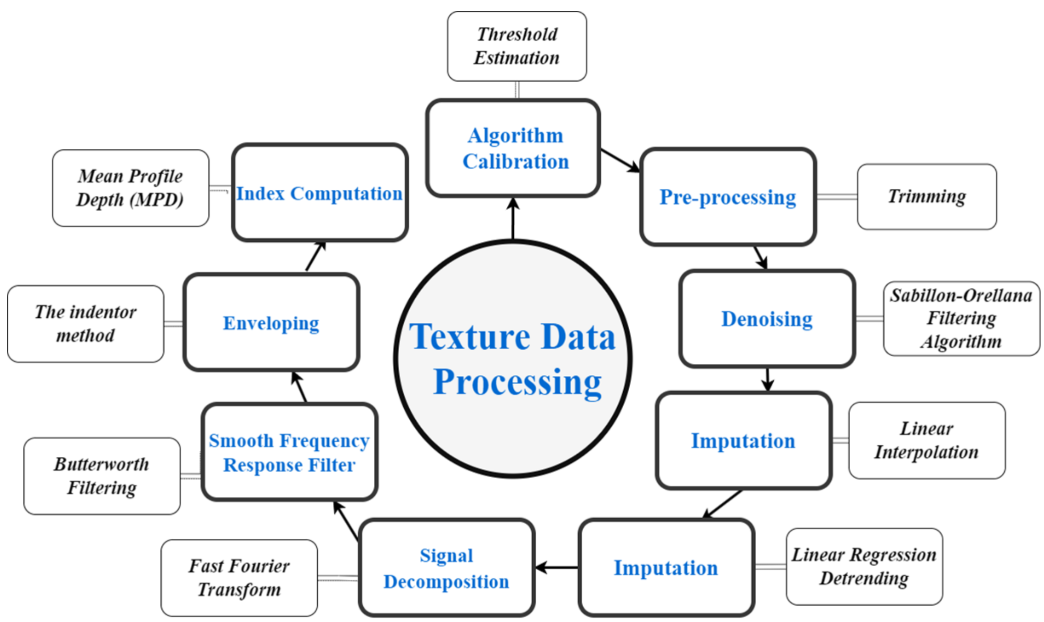


Figure 15: *2D pavement surface profile data processing schematic.*

The data processing begins with algorithm calibration, which varies depending on the chosen processing algorithm. This initial step involves estimating thresholds crucial for later removal of noisy data points from the signal (e.g., the maximum difference in height between adjacent points before a point is considered a spike). Subsequent steps are both algorithm and analysis dependent, typically including preprocessing tasks such as converting elevation units to millimeters, trimming profile edges, and applying pre-imputation methods to ensure the data is complete before further manipulation.

The next phase of analysis involves the detection and removal of noisy data points from the profile, commonly referred to as denoising. In this context, noise encompasses any disturbance such as spikes, flat signals, and white noise that negatively affect the measured signal (5). Methods for addressing these noisy data points include the Wavelet Methodology (6), ISO Standard 13473-1 (7), False Discovery Rate (8), ASTM Standard E178-21 (9), and the Sabillon-Orellana Filtering Algorithm (SOFA) (4).

Subsequently, an imputation method must be applied to replace any data removed during the denoising process. Linear interpolation is the most common imputation method used in the field and was proven by Sabillon-Orellana (5) to be the simplest, most efficient, and accurate imputation method for 2D surface profile data.

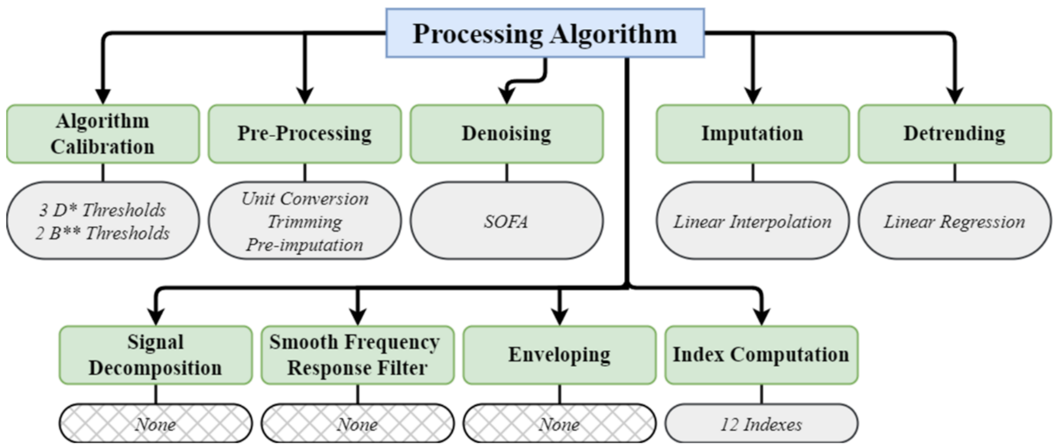
Following imputation, the surface profiles are detrended to eliminate any elevation biases and inherent trends. This step transforms the profiles to a nearly stationary and centered time series of elevation measurements. Both the ASTM Standard E1845-15 (10) and ISO Standard 13473-1 (7) recommend using a linear detrending method to center the data.

Subsequent steps, although not always utilized across research studies on the subject, include applying various mathematical transformations to the signal before computing a texture index. These transformations, as documented in ISO Standard 13473-4 (11) and NCHRP (National Cooperative Highway Research Program) Synthesis Report (12), may involve: i) signal decomposition through methods such as the Fast Fourier Transform (13) to break down the profile into wavelength components; ii) the application of smooth frequency response filters like a low-pass Butterworth filter (14) to isolate the macrotexture wavelengths; and iii) usage of enveloping methods, such as the indenter method (15), to accurately model the contact zone where the tire interacts with the pavement surface.

These transformations may be necessary depending on the goal of the research being conducted but they are not always required before quantifying the processed profiles into a summary index, such as the Mean Profile Depth (MPD), the Root Mean Square (RMS) among many others.

3.2.1. Implemented Algorithm

The research team developed **Figure 16** to illustrate the specific steps used to process the texture data, following the structure highlighted in **Figure 15**. The processing algorithm was calibrated by assessing at least 2,000 profiles to estimate five thresholds for identifying mild and extreme spikes in the data. Next, during the preprocessing stage, the algorithm converted the units of both the profile measurements and the thresholds to millimeters, trimming all measurements not within the center width of 150 mm and pre-imputing any missing data using the median height. For the denoising stage, the researchers implemented SOFA[1] (4). Linear interpolation imputation was then performed on profiles where the total number of removed data did not exceed 20% (profiles where noise exceeded 20% were discarded). Once imputed, a linear regression detrending subroutine was implemented to center the profile at an elevation of 0 mm.



* D stands for the difference between adjacent points
** B stands for boxplot filter

Figure 16: *Schematic illustrating a detailed description of the data processing algorithm.*

The researchers did not apply any signal decomposition, smooth frequency response filters, or enveloping methods to the profiles, as these methods require significant computational resources when processing large quantities of texture data, particularly at a network level. After full processing, a large representative sample of profiles from each tested site was visually inspected to ensure the algorithm performed as expected. **Figure 17** depicts a comparison between a raw profile and a fully processed profile, which were then used to compute a set of twelve texture summary indices.

[1] SOFA is an algorithm developed by The University of Texas at Austin which uses boxplot filters and the difference between adjacent points alongside the previously determined thresholds to detect and remove spikes or flat signals from the profile.

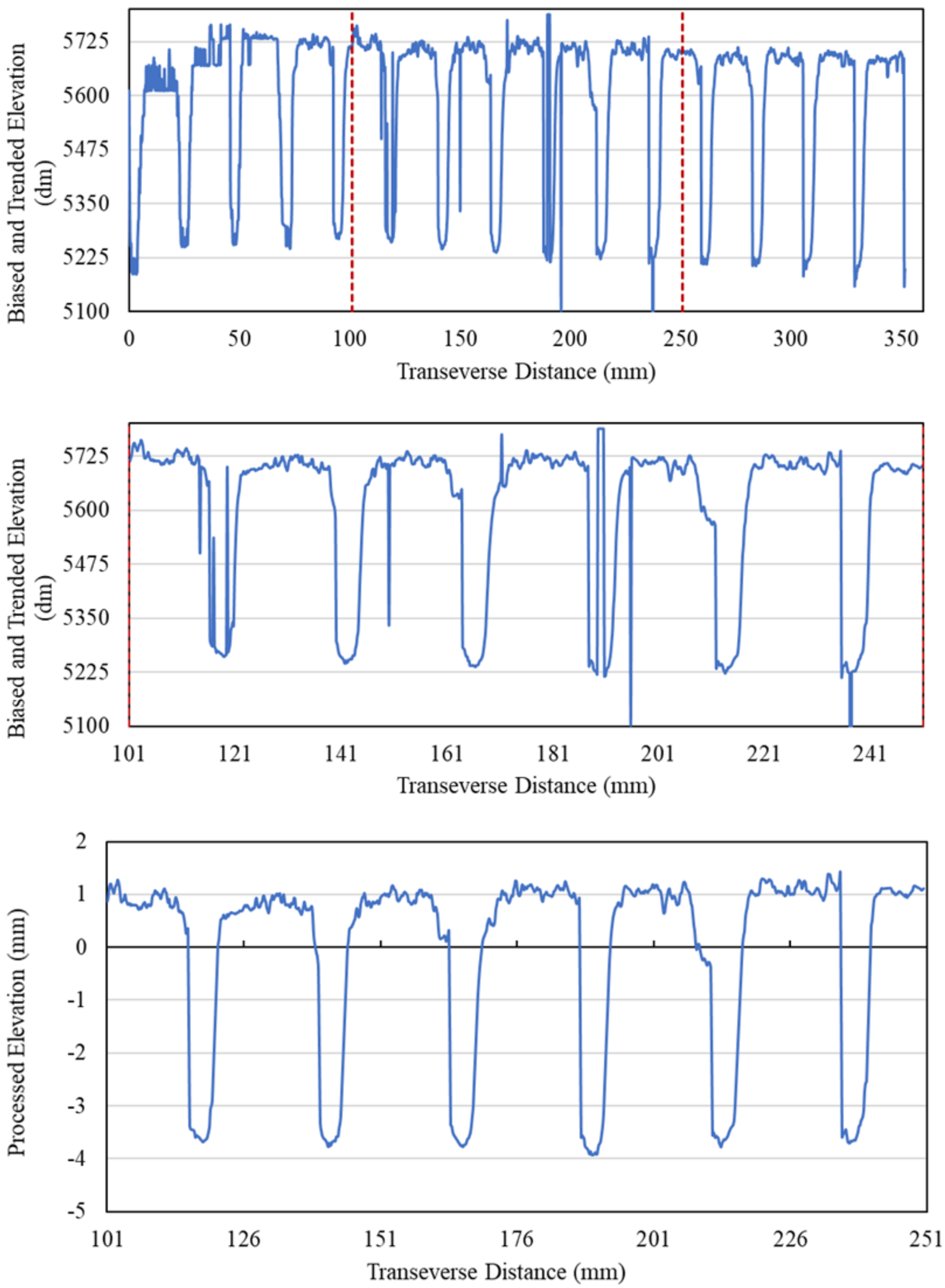


Figure 17: *Example of the SOFA on 2D texture profile for a pavement with new generation diamond grinded surface finish.*

3.2.2. Computed Indices

2D surface profiles primarily serve as visual representations of the pavement surface and lack inherent quantitative value or metrics that describe the surface's topology or morphology. Therefore, characterizing pavement texture necessitates meticulous quantitative analysis of these profiles. The literature on pavement surface characterization commonly uses interchangeable terms such as “statistic”, “parameter”, “metric”, and “index” to refer to a numerical summary characterizing a feature within the profile. This report will consistently refer to these numerical summaries as “indices”.

To develop a comprehensive understanding of how textural features are quantified, the researchers reviewed a wide range of research articles and standards from various engineering disciplines, sourcing material from ISO (16), Gadelmawla et al.(17), and Sabillon et al.(4). This extensive review initially yielded a list of 80 spatial texture indices. However, due to the complexity and lengthy computation time of some of those indices, particularly spectral ones, the list was refined by excluding those that were impractical for efficient processing.

Spatial texture indices were organized into five categories:

- ***Amplitude indices***: to characterize vertical deviations.
- ***Spacing indices***: to quantify horizontal characteristics.
- ***Hybrid indices***: to describe characteristics regarding the profile’s slope.
- ***Functional indices****:* derived more complex textural features by using auxiliary functions like the material ratio curve.
- ***Feature indices***: typically include amplitude, spacing, hybrid, and functional indices, calculated over specific profile length—such as 100 mm—rather than across the entire profile.

Significant linear correlations identified among these indices suggested that some simpler, well-understood indices could approximate more complex ones effectively. As a result, indices exhibiting over 60% correlation were excluded, except for essential ones that are widely recognized in pavement engineering, such as MPD and RMS of height. This correlation-based selection process reduced the list from 80 to 19 indices. A further reduction was made to prioritize indices that require minimal computation time, refining the selection to those most feasible for network-level analysis. This process refined the selection to 12 essential indices, ensuring their practicality for application and familiarity to industry professionals. These indices are detailed in **Table 8**.

Table 8: List of indices used in the study

Amplitude (Height) Indices	
Index	Equation
Root Mean Square Height	$RMS = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(h_i - \bar{h}\right)^2}$
Kurtosis	$R_k = \frac{n(n+1)}{(n-1)(n-2)(n-3)}\sum_{i=1}^{n}\frac{\left(h_i - \bar{h}\right)^4}{RMS^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$
Solidity Factor	$R_s = \frac{\min(h_i)}{\max(h_i)}$

Spacing Indices	
Index	Equation
Mean of Cross Width	$C_m = \frac{1}{n}\sum_{i=1}^{n} x_i$
Standard Deviation of Cross Width	$C_{std} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
Skewness of Cross Width	$C_s = \frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\frac{(x_i - \bar{x})^3}{C_{std}{}^3}$

Hybrid Indices	
Index	Equation
Maximum Absolute Gradient	$R_{dt} = \max\left(\left\lvert\frac{d\,h(x)}{dx}\right\rvert\right)$
Two-Point Slope Variance	$SV_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_i - h_{i+1}}{\Delta x}\right)^2}$

Functional Indices	
Index	**Equation**
Autocorrelation Function	$f_{ACF} = \frac{\frac{1}{l_e - \lvert t_x \rvert}\int_{l_0} (h(x) - \bar{h})(h(x + t_x) - \bar{h})\ dx}{\frac{1}{l_e}\int_0^{l_e} (h(x) - \bar{h})^2\, dx}$
Autocorrelation Length	$R_{al} = \min(\lvert t_x \rvert)$
Local Curvature	$\kappa(x) = \frac{d^2h(x)/dx^2}{(1 + (d\,h(x)/dx)^2)^{\frac{3}{2}}}$
Mean of Local Curvature	$Hcme = \frac{1}{n}\sum_{i=1}^{n} \kappa(x)_i$
Standard Deviation of Local Curvature	$Hcst = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} \left(\kappa(x)_i - \overline{\kappa(x)}\ \right)^2}$

Feature Index	
Index	**Equation**
Mean Profile Depth	$MPD = 0.5\left[\max\left(h_1, \dots h_{m/2}\right) + \max\left(h_{m/2}, \dots h_m\right)\right]$

Where h_i and $h(x)$ is the elevation at point i or x, respectively, $\bar{h}$ is the mean elevation, n is the number of datapoints, $h_{m/2}$ is the elevation value midway through segment, h_m is the elevation value at the end of segment, x_i is the horizontal spacing between inflection points along the profile, $\bar{x}$ is the mean spacing between inflection points, t_x is a spatial shift in the horizontal direction, l_e is the evaluation length, and l_0 is the overlap interval, $dh(x)/dx$ and $d^2h(x)/dx^2$ are the first and second derivatives, respectively, of the scale-limited profile h with respect to the position x.

3.3. Unsupervised Learning

Unsupervised Learning (UL) is a machine learning branch that trains algorithms using data without predefined labels or outcomes. UL includes three primary types: clustering, association rule learning, and dimensionality reduction. Clustering groups similar objects together, enhancing pattern recognition and classification. Association rule learning identifies significant relationships between variables in large databases, uncovering strong rules. Dimensionality reduction simplifies data by focusing on principal variables, reducing the number of variables considered without significant information loss, supporting data compression and feature extraction.

The researchers implemented two unsupervised learning techniques: dimensionality reduction and clustering. While higher dimensionality generally enhances machine learning performance, visualizing data effectively becomes challenging beyond three variables. To address this, the researchers used Principal Component Analysis (PCA), which revealed potential clustering patterns within the complex data. Following this, various clustering techniques were applied to objectively identify and analyze the clusters visualized through PCA.

3.3.1. Dimensionality Reduction

PCA is a dimensionality reduction technique that simplifies high-dimensional data while preserving essential trends and patterns. It creates a new set of variables, known as principal components, by forming linear combinations of the original data. The first principal component captures the most variance, with each subsequent component accounting for the maximum remaining variance. These components are uncorrelated and ordered, ensuring the initial few retain most of the variation from the original variables.

In clustering, PCA projects high-dimensional data into a lower-dimensional space with minimal information loss, enhancing interpretability. An effective way to visualize PCA's efficacy is by comparing scatterplot matrices of the first three principal components with those of the original variables. For instance, **Figure 18** shows a density scatterplot of the original data with six variables (top) and the first three principal components (bottom).

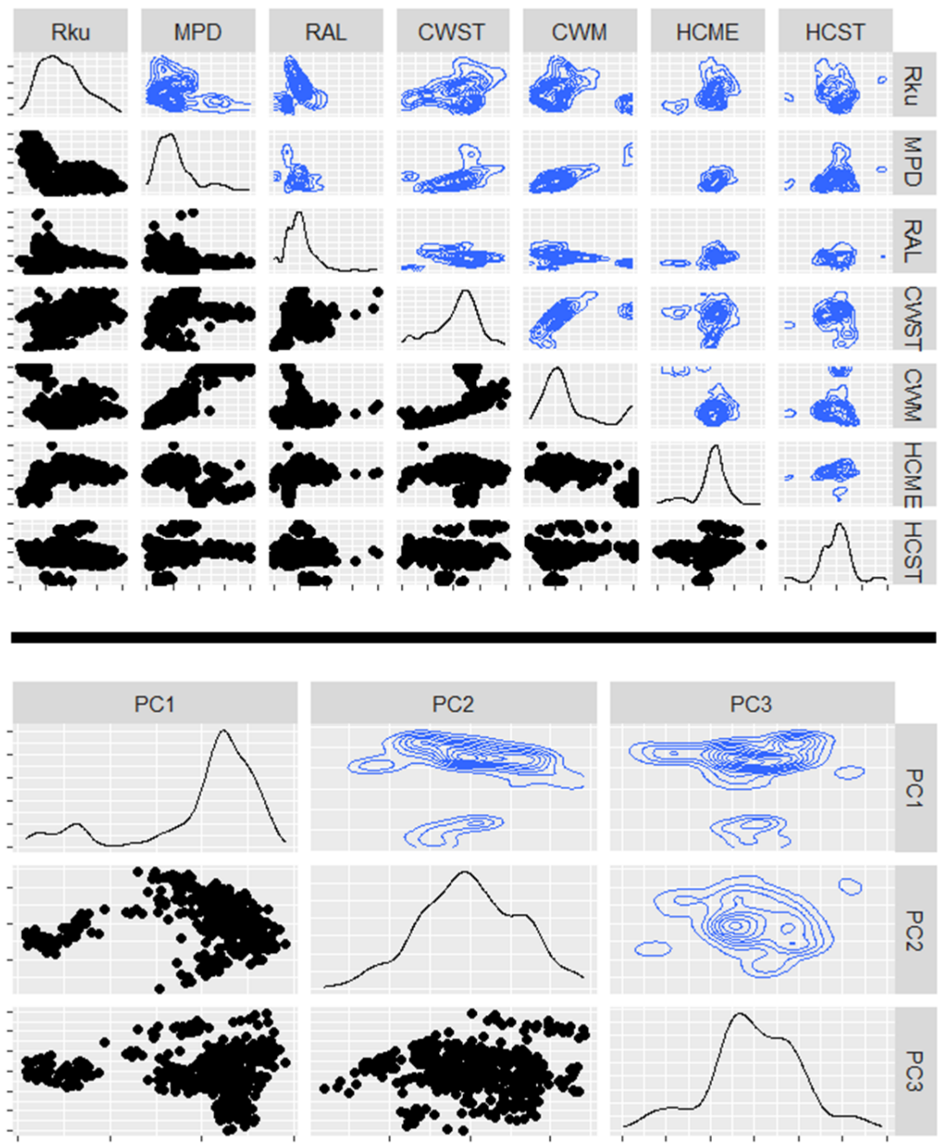


Figure 18: *Comparison between scatterplot of the original data (top) and the reduced data using the first three principal components (bottom)*[2].

The scatterplot matrices are divided into three distinct regions. The main diagonal shows the distribution of each variable, providing a straightforward visual of individual data trends. The bottom-left off-diagonal elements consist of scatterplots that showcase relationships between pairs of variables. The top-right off-diagonal elements contain density plots, which highlight areas of high density and identify clusters.

[2] Note that the numerical values on both axes were suppressed due to excessive overlap at this resolution which distracts from cluster visualization.

Both the original data scatterplot matrix (top) and the principal component scatterplot matrix (bottom) represent the same data. However, the original matrix, with more variables, is more complex and harder to interpret. The PCA scatterplot matrix, with its reduced dimensionality featuring only three variables instead of six, and provides a clearer and more interpretable visual. This simplification greatly enhances the understanding of the underlying data structure and relationships between variables.

Reducing the number of variables from six to three often suffices to reveal the data's clustering structure, especially when examining the density plot of the first and second principal components. These components typically capture most of the variance from the original variables. **Figure 19** zooms into the top middle element of the PCA scatterplot matrix to highlight the density plot for these two principal components. Visually, at least three distinct groups are easily identifiable: a large cluster at the top, a smaller cluster in the mid-right, and another cluster in the bottom middle. The clarity of these groupings in the density plot provides valuable insights that guide the application of clustering algorithms, leveraging the streamlined data structure achieved through PCA to enhance clustering accuracy.

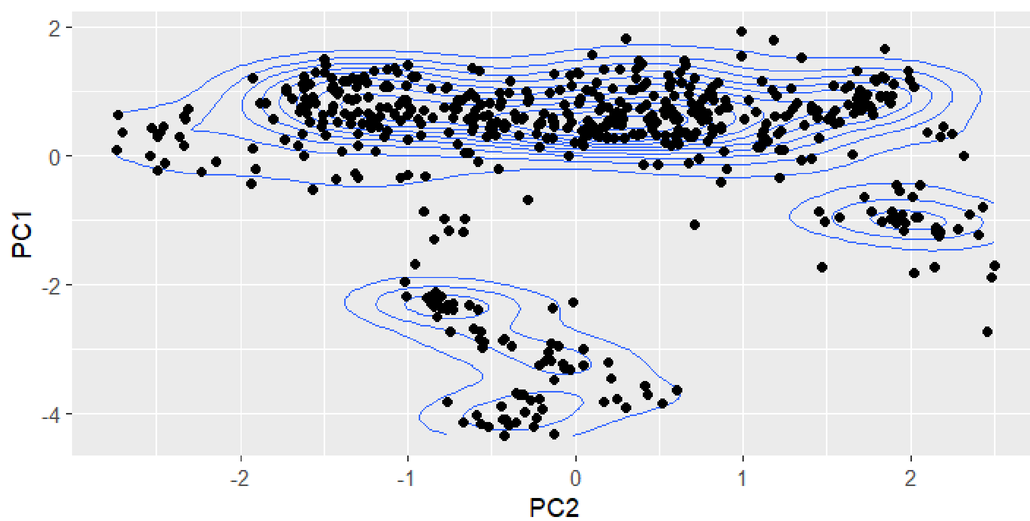


Figure 19: *Density and scatter plots of the first and second principal components.*

3.3.2. Clustering

Clustering methods differ from discrimination or assignment methods, which are types of supervised machine learning used for classification with predefined groups. Cluster analysis aims at identifying naturally occurring groups, or clusters, within the data. A cluster is generally characterized by the homogeneity of its members (internal cohesion) and their distinct separation from members of other clusters (external isolation). **Figure 20** illustrates these properties, showing clusters that are clear to most observers without a formal definition.

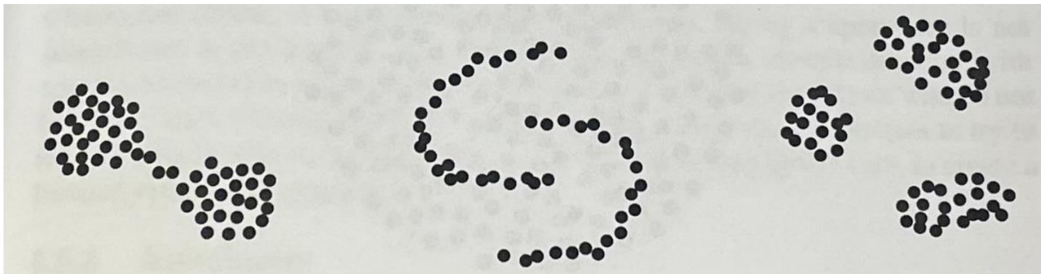


Figure 20: *Clusters with homogeneity and/or separation (18)*

Identifying clusters mathematically is more complex than visualizing them. The earlier qualitative definition of clusters highlights the challenges in rigorously defining concepts like homogeneity and separation. The difficulty in precisely defining these concepts and the uncertainty about how the human brain visually isolates clusters have led to a proliferation of clustering algorithms.

Given the vast number of algorithms, the researcher team focused on identifying one that could detect useful clusters with similar frictional-textural properties. The agglomerative nesting (AGNES) algorithm proved effective for this purpose. AGNES is a hierarchical clustering technique that groups data points based on their similarities using an agglomerative approach. Each data point starts as its own cluster, and clusters are gradually merged based on a specified linkage criterion, such as minimum, maximum, or average distance between clusters. This process results in a dendrogram, which is a tree-like diagram that displays the nested grouping of points and clusters (**Figure 21**).

The dendrogram is invaluable for analysts, as it helps them understand the data structure at different levels of granularity and identify the natural hierarchy in the data. AGNES is particularly useful when the number of clusters is unknown, allowing analysts to explore the hierarchical structure and determine an appropriate number of clusters by interpreting the dendrogram. To decide which clusters to combine, a measure of dissimilarity between sets of observations is required, typically using metrics like Euclidean distance and linkage criteria such as Ward's method, which minimizes the sum of squared differences within all clusters.

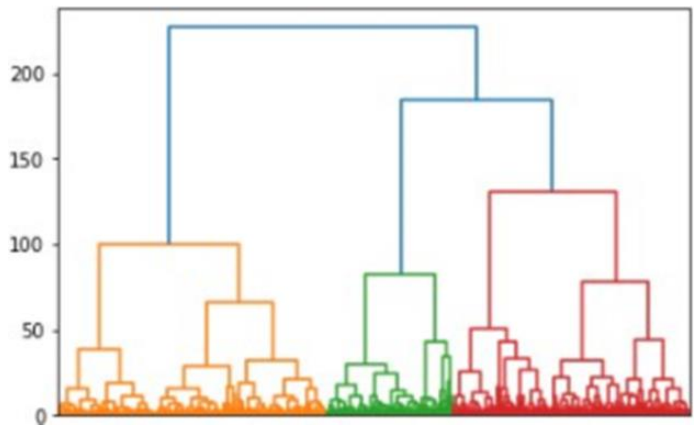


Figure 21: *Sample dendrogram.*

The mergers and splits in AGNES are determined using a greedy strategy, making locally optimal choices at each stage. While this approach may not always yield the global optimum, it often produces satisfactory results efficiently (19). However, the main drawback of AGNES is its computational time. Despite this, agglomerative hierarchical clustering with a specified metric and linkage criterion is guaranteed to converge to the same unique solution every time (20).

3.4. Supervised Learning

Supervised learning (SL) is a branch of machine learning where models are trained using a labeled dataset containing both input data and corresponding correct outputs. The primary objective of SL is to develop a model capable of making accurate predictions or decisions based on new, unseen data (testing dataset). This is achieved by learning the relationships and patterns present in the training dataset. SL requires a dataset with known outcomes for each entry, guiding the algorithm to understand how various inputs lead to specific outputs. Common tasks within SL include classification, where the output is categorical, and regression, where the output is a continuous value.

3.4.1. Classification: Predicting Clusters

The research team employed traditional SL classification algorithms on texture indices from 2D profiles. This section covers fundamental classification concepts and individual algorithms before presenting the results. Henceforth, the terms "features" and "label" refer to the input variables used by the models and the target categories they predict, respectively, ensuring proper SL terminology is used in discussions about classification algorithms.

Classification methods in SL teach models to map features to labels. The Researchers broke these methods down to five unique frameworks based on how each model addresses different predictive challenges with unique principles:

- ***Probability-based models***, like Naive Bayes and Generalized Linear Models (GLMs), estimate the probabilities of data points belonging to various classes based on input features. They are effective when there is a clear statistical link between features and labels and when prior probabilities are well understood.

- ***Tree-based methods*** use hierarchical models, such as Decision Trees, to segment data and make predictions. They work by splitting the dataset into smaller subsets through feature-based decisions, forming a tree structure with nodes as decision points and leaves as predictions.

- ***Ensemble methods*** improve predictive accuracy by combining multiple "weak learners" into a "strong learner". Techniques like bootstrap aggregating, boosting, and stacking reduce errors and provide more reliable predictions.

- **Network-based** such as neural networks, model complex, nonlinear relationships through layers of interconnected neurons. Each neuron processes inputs and passes outputs to subsequent layers, with backpropagation adjusting weights to minimize prediction errors.

- **Distance-based methods**, including k-nearest neighbors (KNN) and Support Vector Machines (SVMs), classify instances based on proximity to labeled instances. KNN uses distance measurements to predict classes based on nearest neighbors, while SVMs create a hyperplane to separate classes in the feature space.

The researcher team identified a suite of 12 traditional SL classification algorithms, capable of utilizing texture indices as input to predict the texture-friction pavement surface clusters to be identified during the cluster analysis. These algorithms encompass a diverse range of methodologies.

1) **K-Nearest Neighbor (KNN)** classifies a data point based on how closely it resembles other data points in the training set. It assigns a class by a majority vote of its *k* nearest neighbors, measured by a distance function like Euclidean distance.

2) **Gaussian Naïve Bayes (GNB)** applies Bayes' theorem with the assumption of independence between every pair of features, using Gaussian distribution

to handle continuous data. It is particularly effective for large datasets and provides a probabilistic prediction output.

3) **Logistic Regression (LOG)** predicts the probability of a binary outcome based on one or more predictor variables, using a logistic function. It is widely used for binary classification tasks but it can also be generalized to multi-class classification problems.

4) **Decision Tree (DT)** classify data by learning simple decision rules inferred from the feature values of the data points. They are easy to interpret and can handle both numerical and categorical data but are prone to overfitting.

5) **Random Forest (RF)** improves decision trees by building multiple trees with randomized data and features and averaging their predictions to improve accuracy and control overfitting. It is robust against overfitting and is effective for a wide range of classification and regression tasks.

6) **Bootstrap Aggregating (BA)**, also known as bagging, reduces variance in predictive models by building multiple versions of a prediction, usually decision trees, and using average voting (for classification) or averaging (for regression) as an ensemble method. It uses random samples of the training dataset with replacement to train each model.

7) **Adaptive Boosting (AB)** is an ensemble technique that combines multiple weak classifiers to form a strong classifier by iteratively updating the weights of incorrectly classified instances to focus learning on harder cases. It adjusts the weights of incorrectly classified instances so that subsequent classifiers focus more on difficult cases.

8) **Gradient Boosting (GB)** builds an ensemble of weak prediction models, typically decision trees, in a stage-wise fashion by optimizing a differentiable loss function. It is highly effective for predictive tasks with heterogeneous features and complex data structures.

9) **Extreme Gradient Boosting (XGB)** is an optimized distributed gradient boosting library designed to be highly efficient, flexible, and portable. It enhances the speed and performance of gradient boosting, making it effective in tackling large-scale and complex machine learning tasks.

10) **Support Vector Machine (SVM)** finds the hyperplane that best separates two classes by maximizing the margin between the closest points of the classes, which are known as support vectors. It is powerful for high-

dimensional data spaces and is effective for classification and regression with a clear margin of separation.

11) **Shallow Neural Network (SNN)** includes one hidden layer of nodes between the input and output layers, allowing it to model non-linear relationships more effectively than a perceptron. It uses a variety of activation functions to learn complex patterns in data, making it suitable for both classification and regression tasks that require more than simple linear decision boundaries.

12) **Deep Neural Network (DNN)** consists of multiple layers of interconnected nodes, capable of learning complex patterns and relationships in large datasets. They are extensively used in areas such as image recognition, natural language processing, and speech recognition, thanks to their deep architecture and powerful learning capability.

A detailed explanation of how each of these algorithms work is beyond the scope of this research report. The reader should refer to the final research report for TxDOT Research Project 0-7139 for a more comprehensive explanation of how these algorithms work. The researchers compared the performance of each of these methods and chose the best one of them all.

3.4.2. Friction Prediction

Regression is used to predict a continuous output variable based on one or more input features. It forms the backbone of many predictive modeling tasks and is fundamental to quantitative analytics. Unlike classification, which predicts categorical outcomes, regression models are essential for scenarios where precision in prediction is crucial.

Linear regression, the simplest form of regression, is effective and straightforward when the system of equations of unknown parameters is linear. For more complex relationships, nonlinear regression models can be used to solve non-linear problems. Time series regression is used for data points indexed in temporal sequences, such as financial analysis or weather forecasting, accounting for trends and seasonal variations over time.

For complex or noisy data structures, alternative regression methods provide robust solutions. Ensemble methods like Random Forests and Gradient Boosting Machines aggregate multiple decision trees to improve accuracy and stability. Neural networks offer a flexible architecture for modeling intricate patterns through layers of interconnected nodes, ideal for capturing nonlinear relationships. Distance-based methods like k-nearest neighbors (KNN) and Support Vector

Machines (SVMs) also provide unique approaches. KNN predicts values based on proximity to nearest neighbors, while SVMs (known as SVR for regression) fit a wide margin in the feature space, making them powerful for tasks with clear margins of separation.

These methods extend regression analysis capabilities to handle various challenging prediction problems across multiple fields. This project utilized multiple linear regression model, an extension of linear regression. Multiple linear regression is an extension of simple linear regression that allows for the prediction of a dependent variable based on two or more independent variables. The general form is presented in **Equation (2)**.

$$\hat{Y} = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \varepsilon \quad (2)$$

Where $\hat{Y}$ is the prediction label, x_i are the features, β_0 is the intercept, β_i are the coefficients of the independent variables, and ε represents the error term, capturing the deviation of the observed values from the line of best fit.

The strength of multiple linear regression lies in its capacity to reveal how different factors collectively influence an outcome and the extent to which each factor contributes when other factors are held constant. This is particularly valuable in fields with multiple influencing variables, such as economics, where multiple linear regression can predict consumer spending based on income, wealth, and interest rates, or in healthcare, where it can forecast patient recovery durations based on age, treatment type, and pre-existing conditions.

Regression coefficients in multiple linear regression are typically estimated using the method of least squares, which minimizes the sum of squared differences between observed and predicted values, ensuring high goodness of fit. Additionally, multiple linear regression analysis offers various statistical measures to assess the model's reliability and validity, such as R-squared, F-tests, and t-tests, which address the goodness-of-fit and the significance of individual predictors.

3.5. Results and Discussion

This section is divided into four parts: i) justification for using a complex machine learning approach, ii) cluster analysis, iii) cluster classification, and iv) regression analysis. The first part discusses the limitations of simple regression methods in developing empirical models and underscores the need for both unsupervised and supervised learning. Cluster analysis identifies groups of pavement sections with similar frictional-textural characteristics, aiding the development of regression models to predict pavement surface friction. These clusters serve as ground truth labels for training a classification algorithm to detect the same clusters in new datasets. The predicted cluster groups from this classification model are then used as inputs for regression models to predict pavement friction using texture indices.

3.5.1. Prediction Approaches

At this stage, one might question the necessity of conducting three complex statistical analyses (cluster, classification, and regression analysis) to predict surface friction on paved roads. Is such complexity essential for obtaining accurate and useful estimates of pavement surface friction? To address this concern, it is useful to present several results that highlight the limitations of simpler, more traditional approaches.

3.5.1.1. Predicting Friction using Multiple Linear Regression with Texture Indices

An analyst would typically begin modeling friction using texture with multiple linear regression, as it is the most interpretable and common method for modeling continuous variables. In this example, **Equation (3)** is specified where $\hat{f}$ is the estimate of friction in terms of GN; MPD, RAL and HCST are all texture indices used to characterize a 2D surface profile, and ω_i are regression coefficients:

$$\hat{f} = \omega_0 + \omega_1 MPD + \omega_2 RAL + \omega_3 HCST \tag{3}$$

The statistical significance across all three predictors in this equation is significantly high (p-values less than 2×10^{-16}) given the large sample size in the data, however this model is not implementable as it achieves a goodness-of-fit of 0.208 in $R^2$and a standard error of 0.106[3]. **Figure 22** shows a scatterplot of the predicted versus measured GN showing that the model's performance is poor at moment.

[3] Recall that the scale for Grip Number (GN), the friction indicator used, runs from 0 to 1

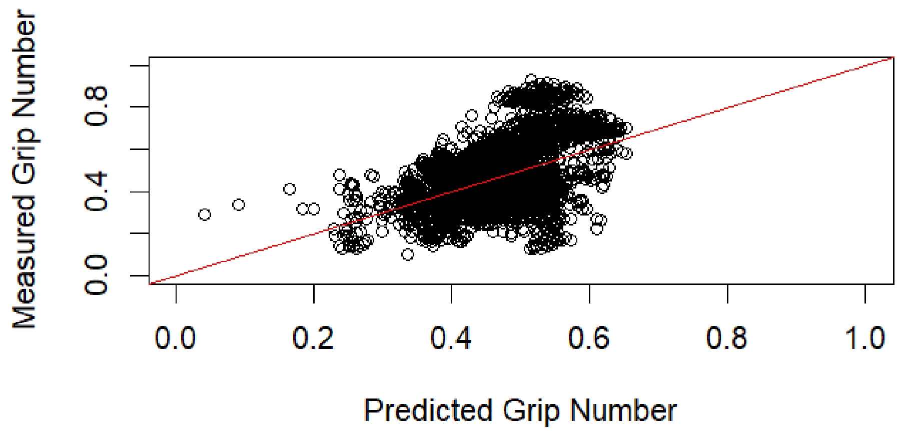


Figure 22: *Predicted and measured friction overlaid by the line of equality in red*

The limited performance of multiple linear regression in this context is evident from scatterplots showing friction against texture indices, revealing more than one relationship between the variables. For instance, consider **Figure 23**, a scatterplot of MPD versus GN, which shows multiple phenomena occurring simultaneously. The scatterplot suggests no clear relationship between MPD and GN, as groups of points trace horizontal lines. However, for MPD values lower than 1.5 mm, unique relationships are visible, though isolating these relationships is challenging due to the heterogeneity in surface characteristics across different pavement mixes or texturing techniques. These observations indicate that multiple linear regression is insufficient for predicting friction using only texture indices, as no single, clear linear relationship exists. The heterogeneity observed suggests a clustering structure within the data.

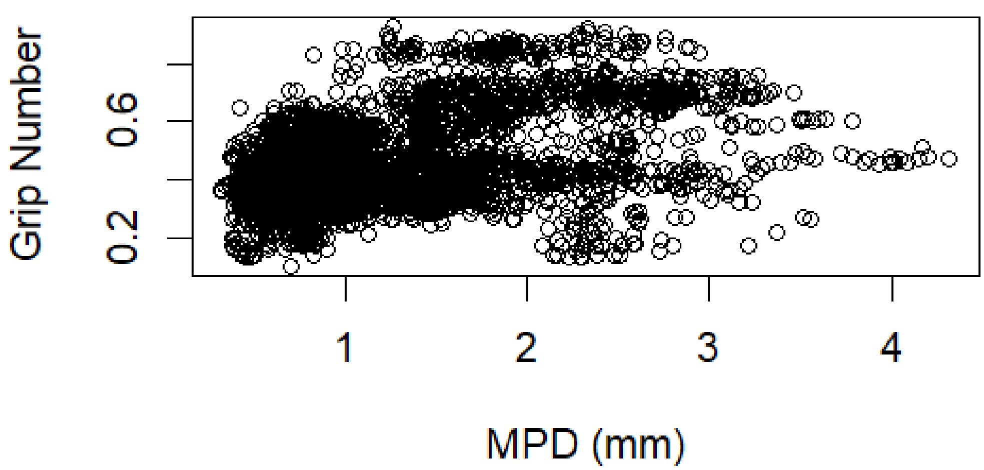


Figure 23 *Scatterplot of MPD versus Grip Number for all the data*

3.5.1.2. Predicting Friction using Multiple Linear Regression with Texture Indices and Surface Information

The next logical step is to isolate clusters with clear relationships between texture and friction, grounded in engineering principles to ensure reproducibility with different datasets. This can be achieved by linking clusters to the macrotexture properties of the pavement, such as aggregate gradation, mix design, and texturing method. Even if this information is not available at a network level, analysts can predict a proxy by assessing surface profile characteristics.

Past studies on flexible pavements show that surface information alone accounts for 50% of the variance in friction as measured by the GN. When combined with texture, the goodness-of-fit increases to approximately 80%, as measured by R_{adj}^2 (21). However, accurate surface information is often known only at the project level, not the network level. Therefore, to make friction prediction feasible at a network level, reliable estimates of the surface information must be obtained first.

Predicting Surface Information from Texture Indices

To bridge this knowledge gap, a past study used a dataset of approximately 500 flexible pavement sections in Austin, each 0.1-mile long, where surface information was known in advance. The study developed a machine learning algorithm based on decision trees to predict pavement surface types using texture indices as input (22). The decision tree model achieved an F1 score accuracy of around 90% when predicting six classes of flexible pavements: two gradations of chip seals, open-graded mixes, open-graded mixes with asphalt binder exudation, dense-coarse mixes, and dense-fine mixes. This project demonstrated that pavement surface information could be predicted by assessing texture indices from 2D surface profiles. However, there are limitations to using this method with a multiple linear regression model.

The first limitation involves multicollinearity and the “double-dipping” of texture indices. Predictions made by the decision tree algorithm are based on a set of texture indices, which are then used as independent variables alongside other texture indices to predict friction. The second limitation is the geographic restriction of the results, as the study focused on flexible pavement surfaces near Austin. For network-level implementation, models need recalibration using a larger dataset that includes all types of pavements across Texas, including rigid pavements. The third limitation is that the model in a snap-shot in time; it would require periodic recalibration as the surface deteriorates. Over time, a chip seal could start to look like a hot-mix asphalt surface, each with different tribological properties.

These limitations highlight areas where a friction prediction model needs improvement to be truly useful. First, the model must be robust to multicollinearity or use uncorrelated variables. Second, it should be trained on a larger database that includes both flexible and rigid pavements from across Texas. Finally, the surface information used in the regression model must reflect the current condition of the pavement surface, not just the original mix type or texturing technique. This approach ensures that as pavement deteriorates and transitions to more polished surface type, the model can still accurately predict the friction-texture relationship.

This rationale supports the proposed complex three-layer approach in the present study. The cluster analysis aims at finding patterns in texture and friction, while the classification algorithm maps texture indices to these clusters for prediction with new data. Finally, these cluster predictions can be used to predict an indicator of friction for a given pavement.

3.5.2. Cluster Analysis

The first step of cluster analysis is to find uncorrelated variables. Thus, a pairwise correlation matrix was computed to isolate those pavement texture indices whose Pearson correlation coefficient was lower than 60 percent. **Figure 24** depicts the features (texture indices and friction measurements) used for the clusters analysis of rigid and flexible pavements. These features include: the height kurtosis (RKU), mean profile depth (MPD), autocorrelation length (RAL), standard deviation of cross width (CWST), mean of local curvature (HCME), the standard deviation of local curvature (HCST), and the grip number (GN).

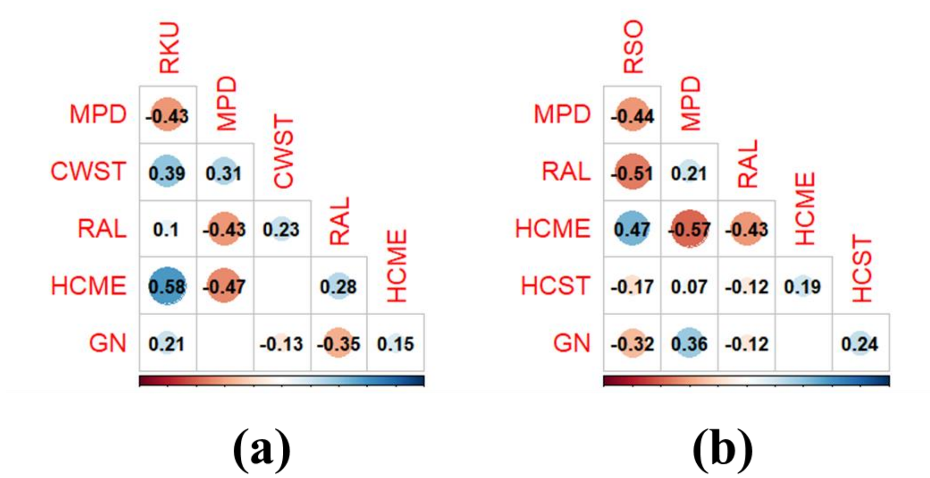


Figure 24: Pearson correlation coefficient matrix for the final selection of five indices: (a) rigid pavemenst and (b) flexible pavements.

Once the right features were selected for the cluster analysis, the researchers normalized the features by scaling them according to **Equation (4)**. Normalizing the features ensures that any identified clusters are both scale invariant and resistant to outliers.

$$X_R = \frac{x_i - \tilde{x}}{IQR} \tag{4}$$

Where, X_R is a robust scaled index, x_i is the original value of index i, $\tilde{x}$ is the median value of the index, and IQR is the interquartile range of the index or the difference between the third and first quartiles of the index.

The indices are then clustered using the AGNES clustering algorithm using the Ward linkage method, a method of measuring the distance between clusters that minimizes within-cluster variance. The Ward linkage method systematically identifies the pairs of clusters that, when merged, result in the smallest increase in total within-cluster variance, thereby ensuring that the most similar groups are combined to form more cohesive cluster structures.

A brief description of the pavements comprising each cluster, along with representative median values for six texture indices of the pavements within those clusters, are presented in **Table 9** and **Table 10**.

Table 9: Summary of rigid pavement cluster analysis

Cluster 1: *New Generation Diamond Grinding*					n: 200 (18%)
MPD	RKU	SV2	RAL	log| HCME	HCST
1.41	-0.22	1.12	0.71	-5.98	2.28
Cluster 2: *Conventional Diamond Grinding*					n: 99 (9%)
MPD	RKU	SV2	RAL	log| HCME |	HCST
0.87	-0.29	0.40	0.37	-7.41	2.21
Cluster 3: *Tining with deep grooves*					n: 419 (38%)
MPD	RKU	SV2	RAL	log| HCME |	HCST
0.87	1.54	0.42	0.96	-5.43	2.36
Cluster 4: *Tining with shallow grooves*					n: 179 (16%)
MPD	RKU	SV2	RAL	log| HCME |	HCST
0.62	0.95	0.33	1.11	-5.88	2.10
Cluster 5: *Exposed Aggregate, dragged, and polished surfaces*					n: 196 (18%)
MPD	RKU	SV2	RAL	log| HCME |	HCST
0.51	0.39	0.30	1.34	-5.97	2.17

Table 10: Summary of flexible pavement cluster analysis

FPC1: *Polished and flat open-graded mixes and seal coats*					n: 335 (11%)
MPD	Rku	Cwst	Cws	Ral	Hcme
0.82	0.79	4.78	1.89	1.51	-6.71
FPC2: *Porous friction and thin bonded wearing courses*					n: 281 (9%)
MPD	Rku	Cwst	Cws	Ral	Hcme
1.54	1.41	4.74	1.46	0.89	-5.06
FPC3: *Dense-graded mixes*					n: 935 (31%)
MPD	Rku	Cwst	Cws	Ral	Hcme
0.87	1.56	3.19	1.80	0.75	-4.72
FPC4: *High friction pavement surfaces*					n: 152 (5%)
MPD	Rku	Cwst	Cws	Ral	Hcme
1.87	-0.36	5.27	1.62	1.06	-4.79
FPC5: *Grade 4 seal coats*					n: 563 (18%)
MPD	Rku	Cwst	Cws	Ral	Hcme
2.13	-0.24	5.21	1.44	1.04	-5.56
FPC6: *High macrotexture surfaces with dense gradation*					n: 137 (4%)
MPD	Rku	Cwst	Cws	Ral	Hcme
1.38	1.35	4.70	1.50	0.96	-5.35
FPC7: *Polished surfaces with high macrotexture*					n: 580 (19%)
MPD	Rku	Cwst	Cws	Ral	Hcme
1.35	0.65	4.99	1.65	1.08	-5.56
FPC8: *High macrotexture surface with uniform gradation*					n: 78 (3%)
MPD	Rku	Cwst	Cws	Ral	Hcme
2.42	-0.48	5.99	1.24	1.14	-5.23

The number of pavement sections in each cluster and their corresponding percentage out of the total, in parenthesis, is shown in the right side of the table. **Figure 25** and **Figure 26** display these clusters on the corresponding first two principal components on the rigid and flexible pavement dataset, respectively.

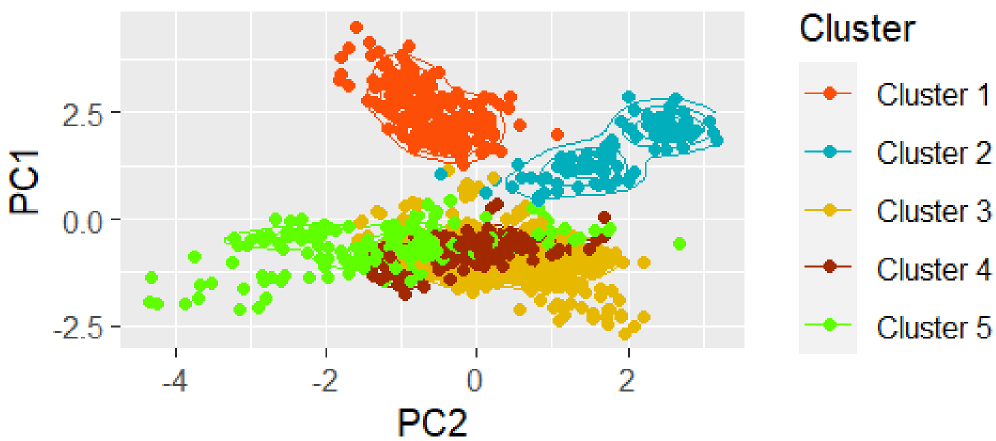


Figure 25: *Texture-friction clusters for the rigid pavement dataset.*

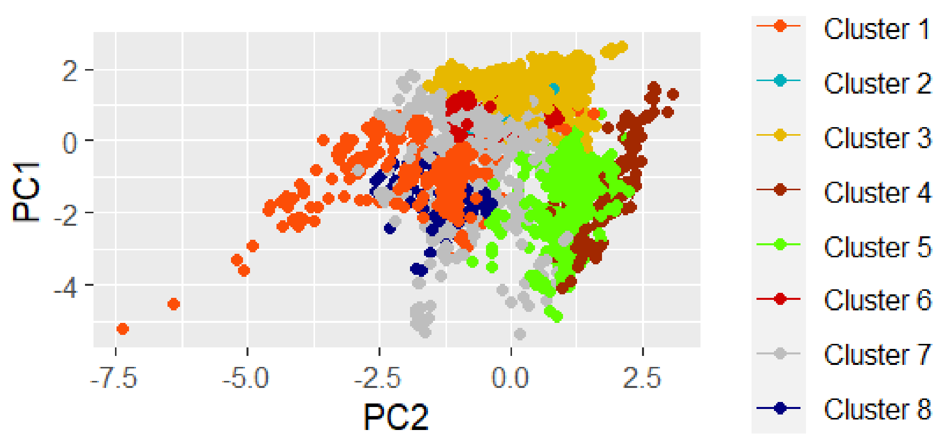


Figure 26: *Texture-friction clusters for the flexible pavement dataset.*

3.5.3. Classification of Clusters

The classification of the friction-texture pavement surface clusters was executed using the 12 ML algorithms outlined in Section 3.1. The performance of these ML algorithms was evaluated based on the F1 score of the weighted average model, which is a metric that calculates the harmonic mean of the precision (proportion of properly classified pavements over all classifications performed) and recall (ratio of properly classified pavements over the total number pavements predicted for that class). The F1 score ranges from 0 to 1, where values close to one signify near-perfect classification and scores near zero indicate pervasive misclassification. The weighted average F1 score was chosen over the standard F1 score to account for varying sample sizes across different pavement types.

In machine learning, data are typically divided into three distinct datasets before model development: training, validation, and testing. The training dataset is used to teach the model to recognize patterns and learn from them. The validation dataset is used to finetune the model's parameters and select the best version by providing feedback on its performance during training. The testing dataset evaluates the final model's performance in an unbiased manner after training and validation. This structured division ensures that the model can generalize effectively to new, unseen data, rather than merely memorizing patterns within the training data. Note that all F1 scores reported in this report were derived from the testing dataset.

The features used to train these models were the set (or a subset in some instances) of 12 texture indices, described in Section 1.2.2, and a dummy variable indicating whether the pavement is rigid or flexible. The labels of the model were the clusters detected using the AGNES algorithm in Section 4.1. The dataset used to predict the texture-friction clustered comprised the ground truth dataset (DS1) and the holdout dataset (DS2). DS1, with a total of 2,346 0.1-mile-long pavement sections, employed an unconventional 30/70 percent training/validation split. This specific division was chosen because pavement sections from the same site might be included in both training and validation datasets, and such double inclusions could negatively affect model performance in the testing dataset. By limiting the training data to 30 percent, the algorithms were compelled to generalize patterns effectively, preventing overfitting to overly site-specific patterns. In contrast, DS2, consisting of 2,019 0.1-mile-long pavement samples, was exclusively used to evaluate (test) the performance of the trained models.

The hyperparameters for all ML classification models, where applicable, were optimized to maximize the testing F1 score while ensuring that the training and validation F1 scores did not significantly differ from the testing F1 score. This approach helps preventing the overfitting of the models by keeping the differences between these scores negligible for practical purposes. Once the models were trained for a binary classification between rigid and flexible pavements, they were tested against the data in DS2, and their testing F1 scores were recorded, as presented in **Figure 27**.

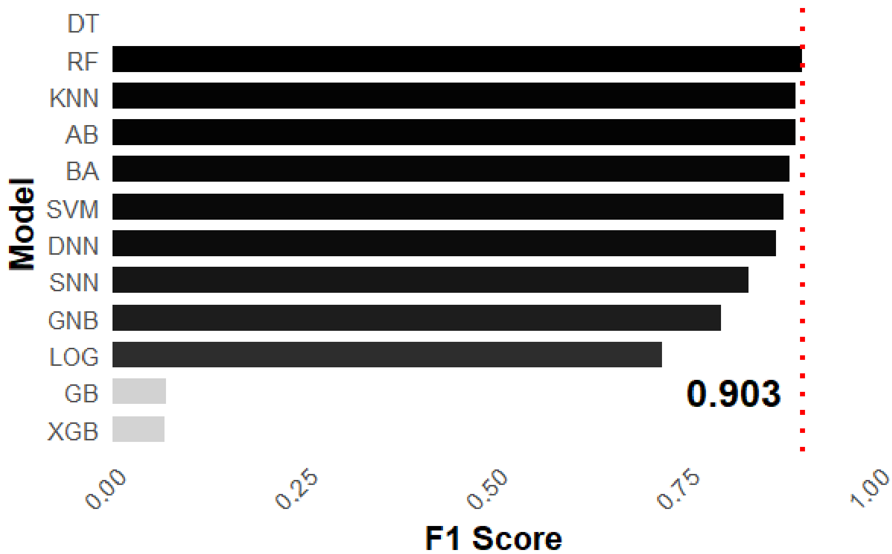


Figure 27: *F1 scores across all twelve ML classification models.*

As presented in the bar chart (**Figure 27**), the random forest (RF) achieved the highest performance, with an F1 score of 0.903, closely matched by the next three highest scores. It is important to note that these high results were achieved with the inclusion of pavement type information as an input variable. Without this input, the maximum F1 score for the model dropped to around 0.750. This highlights the challenges in accurately identifying specific pavement surfaces using solely texture indices without prior knowledge of the pavement type.

Gradient boosting models populated the lower end of the chart, with the XGB model recording a notably low F1 score of 0.080. This poor performance is attributable to the inherent limitations of the gradient boosting framework, where an error by a weak learner, given the structure of the provided data, proves difficult for subsequent weak learners to correct. Moreover, the decision tree model (DT) did not register an F1 score. This absence resulted from the model's inability to accurately predict any instance of clusters 6 and 8 within the flexible pavement clusters. This limitation is due to its simplistic framework and insufficient support for that particular surface type in the dataset. This outcome was anticipated, given the decision tree model's inadequacy for handling complex classification tasks like those required for this task. **Table 11** presents the classification report for the best model (RF) at classifying the friction texture clusters.

Table 11: Classification report for the RF model

Class	Testing Sample Size	Testing F1 Score	Weighted Average F1 Score
RC1	99	0.995	0.902
RC2	49	0.990	
RC3	209	0.961	
RC4	89	0.863	
RC5	97	0.942	
FC1	167	0.920	
FC2	139	0.873	
FC3	467	0.939	
FC4	75	0.770	
FC5	281	0.873	
FC6	67	0.790	
FC7	289	0.832	
FC8	38	0.907	

3.5.4. Ensembled Regression

Finally, the researchers proceeded to train one multiple linear regression for each of the thirteen clusters that have been identified and predicted. Each of these models were developed using DS1 and DS2, i.e., the entirety of the database. **Table 12** provides a shortcut notation for the naming of the texture indexes when specifying the regression equation of each cluster. The regression equations for each cluster are summarized in **Table 13**.

Table 12: Notation for the texture indices

Variable	Texture Index	Variable	Texture Index
$\boldsymbol{X_1}$	MPD	$\boldsymbol{X_7}$	CWS
$\boldsymbol{X_2}$	RMS	$\boldsymbol{X_8}$	SV2
$\boldsymbol{X_3}$	RKU	$\boldsymbol{X_9}$	RDT
$\boldsymbol{X_4}$	RSO	$\boldsymbol{X_{10}}$	RAL
$\boldsymbol{X_5}$	CWM	$\boldsymbol{X_{11}}$	HCME
$\boldsymbol{X_6}$	CWST	$\boldsymbol{X_{12}}$	HCST

Table 13: Summary of friction regression equations

Model	Equation
RS1	$f = \alpha_0 + \alpha_1\boldsymbol{X_3} + \alpha_2\boldsymbol{X_4} + \alpha_3\boldsymbol{X_{12}} + \alpha_4\boldsymbol{X_4}\boldsymbol{X_{12}}$
RS2	$f = \beta_0 + \beta_1\boldsymbol{X_4} + \beta_2\boldsymbol{X_{11}} + \beta_3\boldsymbol{X_{12}}$
RS3	$f = \gamma_0 + \gamma_1\boldsymbol{X_5} + \gamma_2\boldsymbol{X_6} + \gamma_3\boldsymbol{X_8} + \gamma_4\boldsymbol{X_{12}} + \gamma_5\boldsymbol{X_6}\boldsymbol{X_8}$
RS4	$f = \delta_0 + \delta_1\boldsymbol{X_5} + \delta_2\boldsymbol{X_6} + \delta_3\boldsymbol{X_{11}} + \delta_4\boldsymbol{X_5}\boldsymbol{X_{11}}$
RS5	$f = \eta_0 + \eta_1\boldsymbol{X_2} + \eta_2\boldsymbol{X_8} + \eta_3\boldsymbol{X_{11}} + \eta_4\boldsymbol{X_{12}} + \eta_5\boldsymbol{X_2}\boldsymbol{X_{11}} + \eta_6\boldsymbol{X_8}\boldsymbol{X_{12}}$
FS1	$f = \theta_0 + \theta_1\boldsymbol{X_3} + \theta_2\exp(-\boldsymbol{X_{10}}) + \theta_3\boldsymbol{X_{11}} + \theta_4\boldsymbol{X_{12}}$
FS2	$f = \iota_0 + \iota_1\boldsymbol{X_2} + \iota_2\boldsymbol{X_6} + \iota_3\boldsymbol{X_8} + \iota_4\boldsymbol{X_{12}} + \iota_5\boldsymbol{X_2}\boldsymbol{X_6}$
FS3	$f = \kappa_0 + \kappa_1\boldsymbol{X_4} + \kappa_2\boldsymbol{X_9} + \kappa_3\boldsymbol{X_{10}} + \kappa_4\boldsymbol{X_{12}} + \kappa_5\boldsymbol{X_{10}}\boldsymbol{X_{12}}$
FS4	$f = \lambda_0 + \lambda_1\boldsymbol{X_4} + \lambda_2\boldsymbol{X_{11}} + \lambda_3\boldsymbol{X_4}\boldsymbol{X_{11}}$
FS5	$f = \nu_0 + \nu_1\boldsymbol{X_3} + \nu_2\boldsymbol{X_4} + \nu_3\boldsymbol{X_5} + \nu_4\boldsymbol{X_{10}} + \nu_5\boldsymbol{X_{11}}$
FS6	$f = \xi_0 + \xi_1\boldsymbol{X_5} + \xi_2\boldsymbol{X_8} + \xi_3\boldsymbol{X_{11}} + \xi_4\boldsymbol{X_5}\boldsymbol{X_8}$
FS7	$f = \pi_0 + \pi_1\boldsymbol{X_1} + \pi_2\boldsymbol{X_5} + \pi_3\boldsymbol{X_8} + \pi_4\boldsymbol{X_{10}}$
FS8	$f = \rho_0 + \rho_1\boldsymbol{X_2} + \rho_2\boldsymbol{X_7} + \rho_3\boldsymbol{X_{12}} + \rho_4\boldsymbol{X_2}\boldsymbol{X_{12}}$

Where RS and FS stand for "rigid surface" and "flexible surface", respectively; f stands for the friction indicator; $\alpha, \beta, \gamma, \dots, \rho$ are all regression coefficients; and **X's** stand for the texture indices as shown in **Table 12**. All variables presented in the regression equation had a t-statistic higher than 3.29 (i.e., p-value less than 0.001). **Table 14** contains the numerical values for all coefficients shown in **Table 13**.

Table 14: Summary of regression coefficients

	0	1	2	3	4	5	6
α	-1.158	0.047	-4.681	0.637	1.969	~	~
β	1.208	0.047	-22.78	-0.295	~	~	~
γ	0.326	-0.056	0.078	1.997	-0.236	-0.142	~
δ	0.680	0.003	-0.052	-114.3	28.43	~	~
η	-0.058	1.133	-0.317	109.5	0.348	-368.8	-0.897
θ	0.242	0.100	0.436	-7.241	-0.022	~	~
ι	1.403	-1.556	-0.251	0.794	-0.077	0.271	~
κ	-0.167	-0.058	0.022	0.516	0.207	-0.205	~
λ	0.767	-0.095	7.136	8.270	~	~	~
ν	0.796	-0.067	-0.086	-0.020	-0.091	-2.895	~
ξ	0.806	-0.139	-0.831	6.290	0.284	~	~
π	0.264	0.072	-0.034	0.088	0.102	~	~
ρ	-0.511	0.468	-0.435	0.605	-0.242	~	~

The performance of these regression models was then measured in terms of goodness-of-fit and accuracy. Goodness-of-fit was measured in terms of the coefficient of determination (R^2), while accuracy is measured by both the root mean square error (RMSE) and the mean absolute error (MAE). R^2 is a statistical measurement that represents the proportion of variance in the dependent variable that is explained by the independent variables in the regression model. The range of R^2 is from 0.0 to 1.0. An R^2 value of 1.0 indicates that the model explains all the variance in the dependent variable, showing a perfect fit. Conversely, an R^2 value of 0.0 means that the model does not explain any of the variance in the dependent variable, implying no relationship between the independent and dependent variables. In addition, the coefficient of multiple determination $\left(R_{adj}^2\right)$ is used whenever more than one independent variable are used in the regression equation. R_{adj}^2 punishes the model for every additional feature on the regression that does not significantly contribute to explaining the variance in the dependent variable.

Note that neither R^2 nor R_{adj}^2 are not a measure of accuracy or predictive power. It assesses how well the model fits the data, but it does not indicate how well it will perform on new data or any indication of its prediction error. Thus, the researcher team also reported the RMSE and MAE, as measure of accuracy and predictive power.

RMSE is calculated as the square root of the average of the squared differences between the actual and predicted values. It measures the standard deviation of the residuals and provides an indication of the average size of the prediction errors. RMSE is sensitive to large errors due to the squaring of differences. MAE calculates the average of the absolute differences between the actual and predicted values. It measures the average size of the prediction errors without considering their direction. MAE is less sensitive to extreme values compared to RMSE. Both RMSE and MAE have a range from zero to infinity, where lower values indicate a better fit of the model to the data. **Table 15** summarizes the goodness-of-fit and accuracy metrics for each model.

Table 15: Summary of goodness-of-fit and accuracy of the various models

Model	R^2_{adj}	MAE	RMSE
RS1	0.302	0.033	0.044
RS2	0.648	0.036	0.049
RS3	0.615	0.039	0.051
RS4	0.405	0.049	0.060
RS5	0.569	0.042	0.059
FS1	0.672	0.046	0.061
FS2	0.697	0.025	0.030
FS3	0.142	0.060	0.074
FS4	0.270	0.016	0.022
FS5	0.348	0.035	0.044
FS6	0.839	0.027	0.034
FS7	0.299	0.048	0.062
FS8	0.764	0.041	0.049

An important note when assessing Table 15 is that comparing the R^2 of different models is not statistically correct, as each model was developed using a different subset of the data. Thus, the total square variance required to compute the R^2 is not the same across different models making the comparison a logical fallacy. As aforementioned, low R^2_{adj} is not an indication of poor prediction models. For instance, the FS1 model had a R^2_{adj} that would be considered high, but its MAE was 0.046, indicating that, on average, any prediction made by this model are likely to be $\pm$ 0.046 units of friction away from the true value. Conversely, the FS4 model had one of the lowest R^2_{adj} and simultaneously one of the lowest MAE of 0.016. That is because its predictor variable (**X's**) in the regression equation did not account for much of the variance in friction. Nonetheless, the friction of pavements in that cluster did not vary too much from the mean value, making the mean prediction one of the best predictions that minimizes the error. Overall, the researchers aimed to obtain MAEs and RMSE below 0.05.

Once all the data were clustered and assigned to its corresponding regression equation, the entire set of predicted friction was compared against the observed friction using a scatterplot (**Figure 28**).

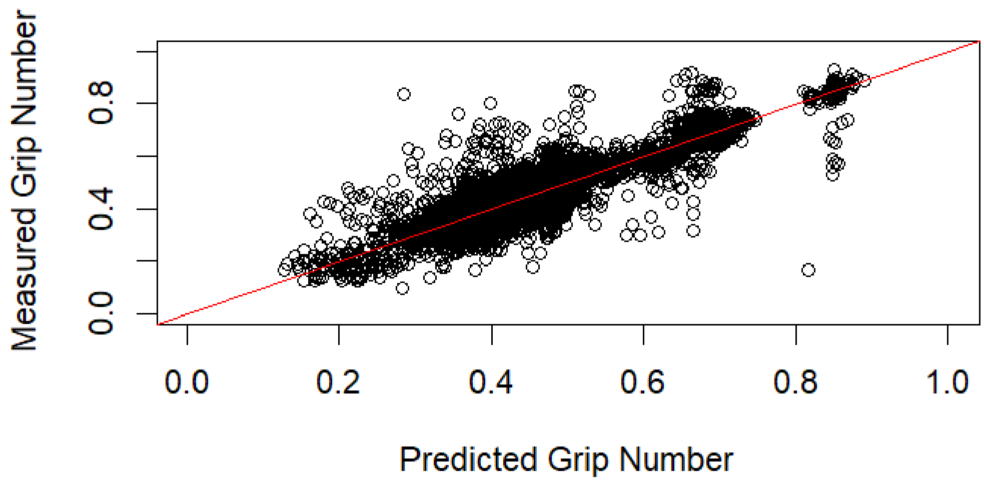


Figure 28: *Predicted and measured friction overlaid by the line of equality.*

Table 16 summarizes the goodness-of-fit and accuracy of the data at a global level. The global level refers to the instance where all the models were applied to the data in ensemble, and their goodness-of-fit and accuracy were recomputed.

Table 16: Summary of goodness-of-fit and accuracy for all the data

R^2	MAE	RMSE
0.767	0.050	0.071

When combined, all the regression models achieve an $R^2$of 0.767. The definition R^2 holds even when using multiple regression models in ensemble. Nonetheless, the reader should be wary of that value as there is no penalty for the complexity of the ensemble model. In a multiple regression model, one could use the coefficient of multiple determination $\left(R^2_{adj}\right)$, which punishes the model for having irrelevant features that will increase the R^2 regardless, but such metric has no analogy in ensemble methods.

A better assessment of model's accuracy performance is through the MAE and RMSE. Overall, the MAE and RMSE at global level were 0.050 and 0.071, respectively. Both metrics indicated that, on average, the error of the ensemble model was between $\pm$ 0.05 and $\pm$ 0.071 units of friction, depending on the metric used to measure the error. The lower boundary of that range aligns with the expectations of the researcher team.

This framework to predict friction at a network-level is the first of its kind, and builds upon the knowledge of past research studies (3, 4, 22, 23). The true novelty of this model, compared to all the previous model and variations available in this field, is that this framework predicts friction accounting for macrotexture information and surface information, but it does no longer constraint the surface information. This model now accounts solely for the expected deterioration of textural properties in the pavement with the passage of time and traffic. For instance, a brand-new porous friction course (PFC) or a high friction surface treatment (HFST) would originally cluster with the pavements of its kind, with high friction values. However, as the surface texture of these pavements deteriorates, these surface types will start to cluster with pavements having the same textural characteristics and lower friction values.

3.5.5. Utilizing a Neural Network to Predict Friction

A model of comparable complexity was developed to compare its performance to that of the ensemble multiple regression approach with clustering and classification. Specifically, a deep neural network (DNN) with two hidden layers was trained, validated and tested to predict pavement friction with the highest possible accuracy and goodness-of-fit.

Figure 29 depicts the architecture for the DNN used for the task of predicting pavement friction. The DNN was trained and validated on a stratified (30 and 10 percent, respectively) sample of the total data and then tested on the remaining 60 percent of the data. The DNN uses seven uncorrelated statistics[4] to predict Grip Number (GN). The reason for which surface information is not provided to the neural network is because a deep neural network has a built-in method to detect that different surface types will have different relationships through its multiple hidden layers and additional neurons.

[4] Pearson correlation coefficient is lower than 60%

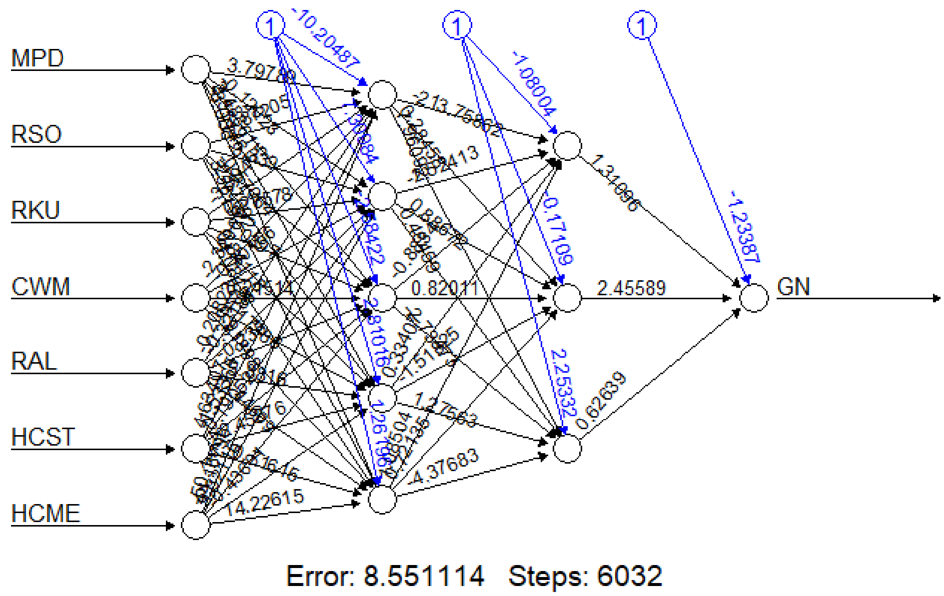


Figure 29 *Architecture for DNN model to predict GN*

The DNN was able to achieve results that were better than random chance. The model was able to achieve a goodness-of-fit in the order of 65% in the coefficient of determination and an accuracy in the order of $\pm$ 0.066 and $\pm$ 0.085 in MAE and RMSE, respectively. **Table 17** summarizes the goodness-of-fit and accuracy metrics for the DNN model.

Table 17: Summary of goodness-of-fit and accuracy for a deep neural network on the entire data

R^2	MAE	RMSE
0.669	0.066	0.085

It is noteworthy that when assessing a plot of the predicted versus measured GN for all the sections when using the DNN (**Figure 30**), it is clear that the DNN model is only good at predicting GN in the vicinity of 0.50. Otherwise, the model tends to overpredict the friction at high values and underpredict the friction at lower values.

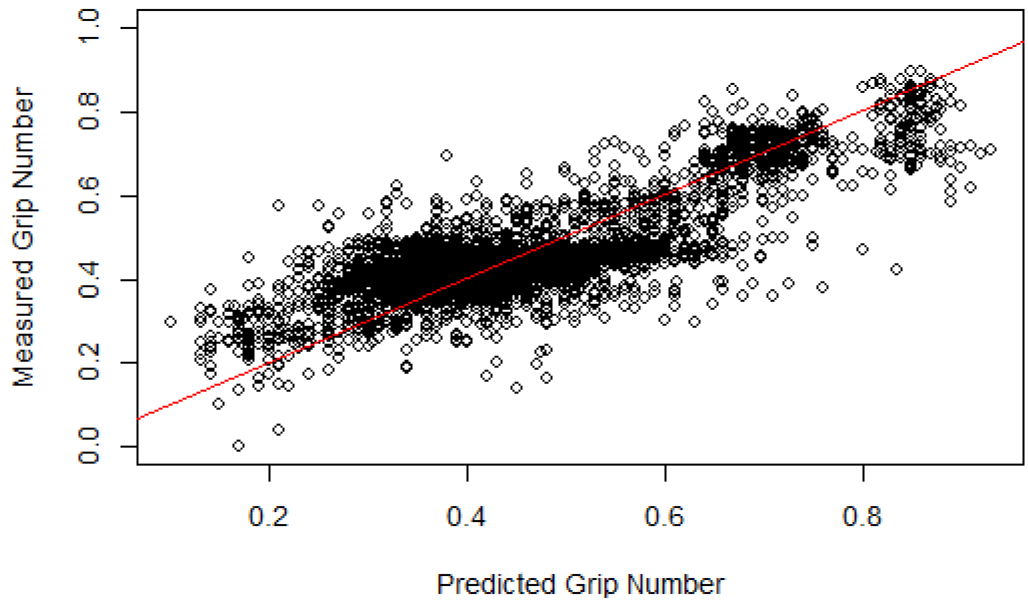


Figure 30 *Predicted and measured friction overlaid by the line of equality in red for DNN model*

Chapter 4. User Manual and Demo

This chapter details the development and implementation of a highway friction prediction model across Texas, serving as a user manual and demonstration guide for the software and hardware systems. It covers methodologies for collecting, processing, and analyzing pavement texture data to predict surface friction, providing step-by-step instructions to ensure accurate pavement assessments.

The chapter is divided into three sections. The first section, the technical manual, outlines the hardware and software setup for data collection, including vehicle prototype assembly, sensor mounting, data collection protocol, and UT_Texture software integration. The second section, the user manual for the TexTex FricFore application, guides users through processing raw data, computing 2D texture indices, and predicting pavement friction with detailed instructions for each module. The third section describes a video demonstration summarizing the technical memorandum, offering an additional resource to help users understand the procedures and applications discussed.

4.1. Technical Manual

4.1.1. Assembly of Texture Data Collection

A vehicle with a robust build, strong suspension, and reliable towing capacity should be used for field testing due to its resilience against large vertical bounces. To accommodate different vehicle heights, a customizable mounting system should be used to enable vertical and lateral positioning of the components on the wheel path or location of interest, as shown in **Figure 31**.

Figure 31: *Side and rear view of the mounting system.*

Accurate measurement of sensor heights is crucial to set them at the correct stand-off distances, with the laser sensor requiring a height of 17.5 in. from the camera window to the surface. Measurements should be taken on a level surface, like concrete, and account for the weight of passengers and/or extra load in the vehicle. The line laser requires a 12V power supply from the vehicle's cigarette lighter, consuming 0.7 amps, and data transmission is handled by a 20-foot (6 m) Cat 6E Ethernet cable connecting the sensor to a laptop in the front passenger seat. The final set up of the testing vehicle should like the one shown in **Figure 32**.

Figure 32: *Laser system mounted onto vehicle*

4.1.2. Data Collection Protocol

To collect data effectively, the user should begin by conducting a remote survey of the provided locations using Google Street View and Pathweb (or similar video recording software). This pre-survey allows for pinpointing the exact start and end points of pavement surfaces and checking for visible distresses such as raveling, bleeding, cracking, patching, rutting, and punchouts that might affect data quality. Planning the routes to travel between testing sites and coordinating with local district personnel to inform them of the data collection schedule and request field assistance if needed is also essential.

When ready for data collection, the user should set up by parking the surveying vehicle about half a mile from the start of the pavement site. The equipment should be mounted, quality control checks performed, and overhead amber lights activated. The vehicle should then travel in the rightmost lane, accelerating to a speed of 50 mph, and cruise control should be activated to maintain this speed. Data collection should be conducted along the inner wheel path of the rightmost lane, with the laser sensor oriented at a 45-degree angle to detect patterns of tining, grinding, or grooving on rigid pavement. Data collection should end a few yards from the end of the pavement surface of interest or if there is an unexpected obstacle or pavement change. After data collection, the vehicle should be parked in a safe spot, sensors dismounted for safe transport, and all raw data files backed up at the end of the day.

4.1.3. Laser Sensor: High-Speed Texture Sensor

The High-Speed Texture Sensor (HSTS) employs advanced single-line range technology to accurately measure surface elevations. Designed for versatility, users can mount the sensor on a moving vehicle, enabling efficient data acquisition on the go. Complementing the sensor is UT_Texture, a comprehensive software suite developed to integrate seamlessly with the High-Speed Texture Sensor.

This guide offers detailed instructions on how to effectively use UT_Texture. It explains the procedures for controlling the sensor during data capture, ensuring optimal performance in various conditions. Additionally, it provides guidance on adjusting sensor settings to meet specific requirements, visualizing the captured data for better analysis. By following this guide, users can maximize the capabilities of both the HSTS and UT_Texture software, ensuring accurate and reliable results in their surface elevation measurements.

4.1.3.1. Laser Safety Warning

The HSTS emits a class 3B laser line onto the pavement surface at a 60-degree fan angle, with a wavelength of 810 nm. This laser projects a single beam on the ground that is approximately 450 mm long and 0.1 to 0.3 mm wide.

Direct exposure to this laser, especially during installation and calibration, can cause vision damage or serious injury. To ensure your safety, please follow these precautions:

- **Avoid Direct Gaze**: Never look directly into the laser window from closer than three feet. Prolonged exposure to the laser beam can be harmful to eyesight.

- **Wear Protective Eyewear**: When the device is powered on and requires inspection, always wear suitable protective eyewear specifically designed for laser safety. This precaution is essential to protect eyes from potential harm.

- **Proper Orientation**: Before powering on the device, ensure that the laser window is directed downward towards the ground. This helps prevent unintended laser exposure to personnel and surroundings.

Never look directly into a live laser beam or its window without appropriate eye protection!

4.1.3.2. Software and Main User Interface

Double-click the UT_Texture.exe icon, shown in **Figure 33** to launch the application. This versatile program serves dual purposes: enabling both real-time sensor control and data acquisition, as well as offline data reloading and pre-processing.

Figure 33*: Icon for the UT_Texture software executable*

The main user interface, shown in **Figure 34**, divided into five regions: four display sections (A-C) and one control section (D).

- Section (A), at the top, plots a sample 2D profile from the scan area.
- Section (B), in the bottom, presents a black-and-white image of 256 accumulated profiles, or a frame, in either intensity or range.
- Section (C), the right-side vertical window, provides a view of accumulated frames, supporting up to 60 frames per saved file.

These sections facilitate the visualization of live data streamed directly from the sensor or from a file loaded into the software. The control section, (D), on the left side, offers easy access to sensor parameters, control options, and functional controls, enhancing user interaction and operational efficiency. **Figure 35** demonstrates the user interface during data collection.

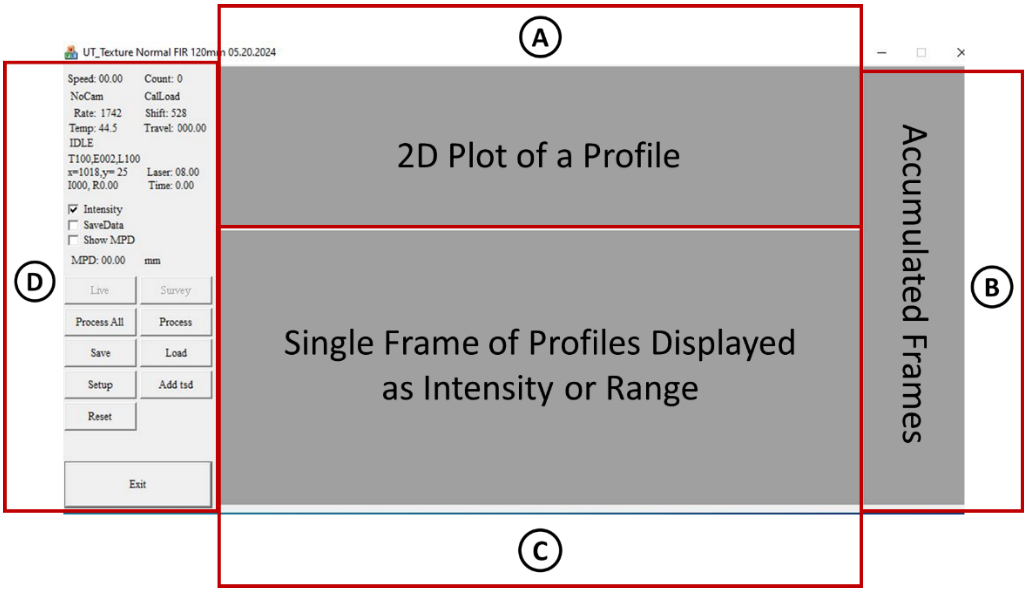


Figure 34: *Main areas within the graphic user interface prior to data collection*

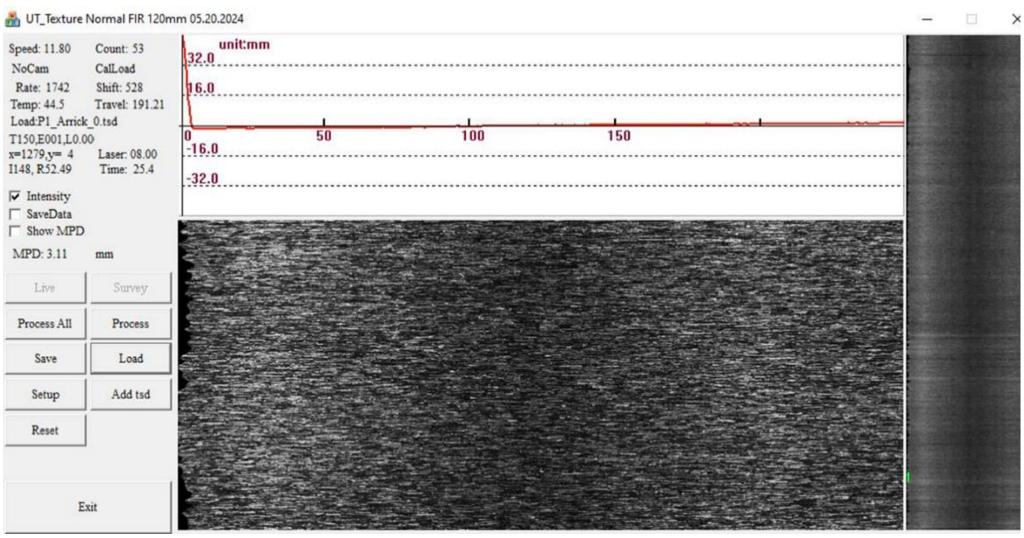


Figure 35: *Example of main graphic user interface during to data collection*

Given that the Control Area in Section (D) of the interfaces is the one where the user interacts the most with the software to monitor the status of the laser and adjust certain data collection parameters, a detailed review of every button and label of this section is provided in the subsequent subsections of the report.

4.1.3.2.1. Control Area

To make a more detailed description of each of these labels and buttons, the control area will also be broken down into 4 sections, A-D.

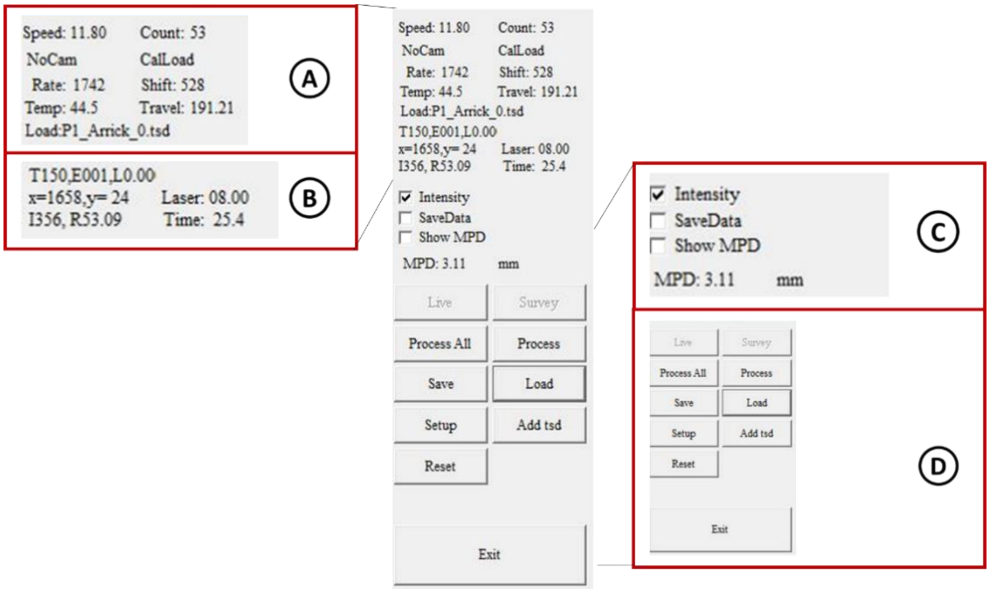


Figure 36: *Breakdown of the control area*

In Section A of the control interface, various parameters are displayed to monitor and manage the sensor system:

- **Speed**: Shows the travel speed in mph, accessible only in trigger mode and may not update in default time mode.
- **Count**: Indicates the number of frames collected.
- **DEFCAM/NoCam/SNCAM**: Reflects the sensor status. "DEFCAM" means a sensor is connected with a default initial file loaded, "NoCam" indicates no sensor connection or proper initial file, and "SNCAM" signifies a sensor connected and initialized with a corresponding serial number file.
- **CalLoad/No Cal**: Displays "No Cal" if no calibration file is loaded, indicating non-conforming range data accuracy. "CalLoad" shows that calibration is loaded and applied.
- **Rate**: Represents the number of profiles per second transmitted from the sensor. In time mode, it is controlled and calculated with the sensor cycle time. In trigger mode, the profile rate is governed by an external trigger signal, which may not display accurately as the rate may not remain constant.
- **Shift**: Currently inactive or not utilized in the software.

- **Temp**: Displays the sensor's internal camera temperature in Celsius, which must stay below 55 degrees. If it exceeds 55 degrees due to high ambient conditions, the sensor should be turned off.

- **Travel**: Shows the traveled distance in meters, applicable only when using an encoder.

In Section B of the control interface, various parameters and values are displayed to provide detailed information about the sensor and image data:

- **TXXX, EXXX, LXXX**: These denote internal sensor parameters. "T100" specifies a threshold of 100 for elevation (range) measurement, "E002" indicates an internal exposure time of 002 microseconds, and "L100" is reserved for presenting the laser light level.

- **x= XXX, y= XXX**: Display the mouse's x and y coordinates within the display image. Users can position the mouse over specific image locations to inspect intensity and range values, enhancing data exploration.

- **Laser**: Indicates the laser's power level.

- **IXXX, RXXX**: Represent the image pixel information at the mouse location. "I000" shows the intensity value, and "R000" provides the corresponding range value.

- **Time**: Displays the time in milliseconds required to process an image frame's data. This serves as a measure of the CPU's power consumption for image processing. If this time exceeds the image frame acquisition time, frames may be lost, indicating the need for a more robust computer.

In Section C of the control interface, several options and parameters are available for managing data display and saving, using a checkbox system. The other information in this section displays a "raw" MPD measurement. Implying that the profile is still raw when the index is computed. The options in the checkboxes include:

- **Intensity**: Allows users to toggle between intensity and range displays in the profile frame windows. The sensor captures two types of data in each profile sample: pavement surface brightness (intensity), shown in **Figure 37**, and surface elevation (range), shown in **Figure 38**. The intensity image consists of intensity data in profiles, used to verify sensor functionality and control laser line image brightness in the camera. The range image comprises range data in profiles, aiding in joint location detection and faulting value calculation.

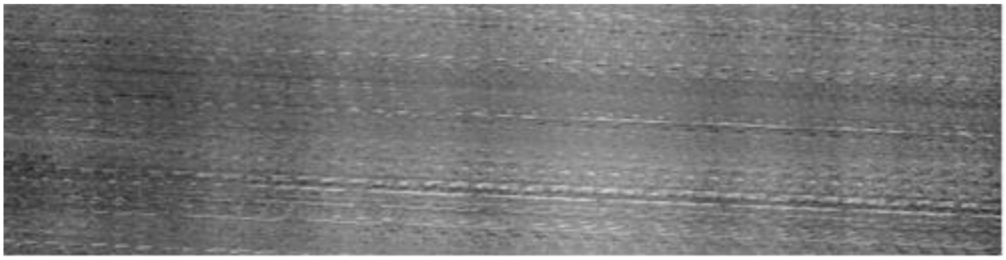


Figure 37*: Image of intensity*

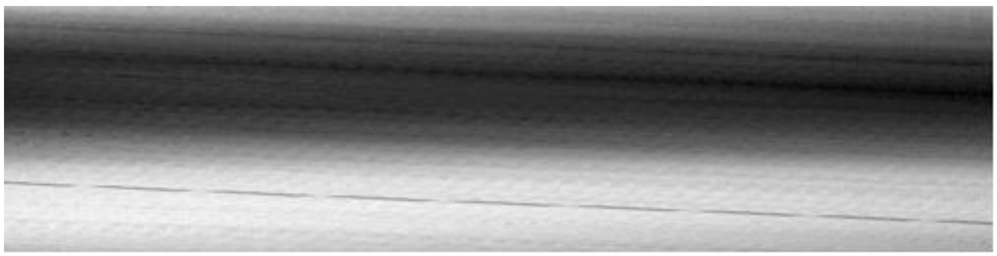


Figure 38: *Image of range*

- **SaveData**: Controls the preservation of raw data during collection. When a data file name is selected and the save raw data option is activated (refer to the Setup tool for details), checking this box commits all subsequent data to the already opened file. Unchecking it suspends the data-saving operation.
- **Show MPD**: Displays the calculated Mean Profile Depth.

In Section D of the control interface, various buttons and features are available for managing sensor data acquisition and processing:

- **Stop/Live**: When the sensor is connected and activated, live mode is enabled automatically. This button allows users to start and stop sensor data acquisition.
- **Survey**: Press this button when in trigger mode and ready to start data acquisition.
- **Process All**: Reprocesses all the buffer image data.
- **Process**: Reprocesses the current buffer image data.
- **Load**: Allows users to load saved frame image files (.3fd or .tsd) or buffer data (.3sd) from a disk. A .3fd file contains a single profile image frame for all three AOIs, while a .3sd file includes all profile image frames in the ring buffer.

- **Setup**: Opens a new dialog box for setting various parameters and controls for data collection and processing. Detailed operations are provided in the subsequent chapter.
- **Add tsd**: Loads a saved 3D frame file.
- **Reset**: Resets the buffer.
- **Save**: Allows users to save the current profile image frame data to a .3fd disk file or the entire buffer data to a .3sd file.
- **Exit**: Closes all open files and exits the application.

4.1.3.3. Mouse Control Functions

Within the main user interface, various functions are executed through mouse movements and button clicks. By positioning the mouse in the right-side vertical pavement image window and scrolling the mouse wheel forward or backward, users can respectively increase or decrease the buffer index, refreshing the profile frame within the entire data buffer. Two green bars on either side of the pavement image window move with the mouse wheel to indicate the currently displayed data frame.

Additionally, moving the mouse within the profile frame display updates the mouse location (x, y) and image pixel data (intensity and range values) on the left side of the main UI, providing real-time feedback and enhancing data interaction.

4.1.3.4. Parameter Setup

Several parameters and processes associated with HM Texture can be configured through the Parameter Setup. Clicking the Setup button in the main interface opens the Setup dialog box, shown in **Figure 39**. This dialog box organizes information and control areas by function. For clarity in explaining each component of the parameter setup screen, the screen is divided into five areas labeled A-E.

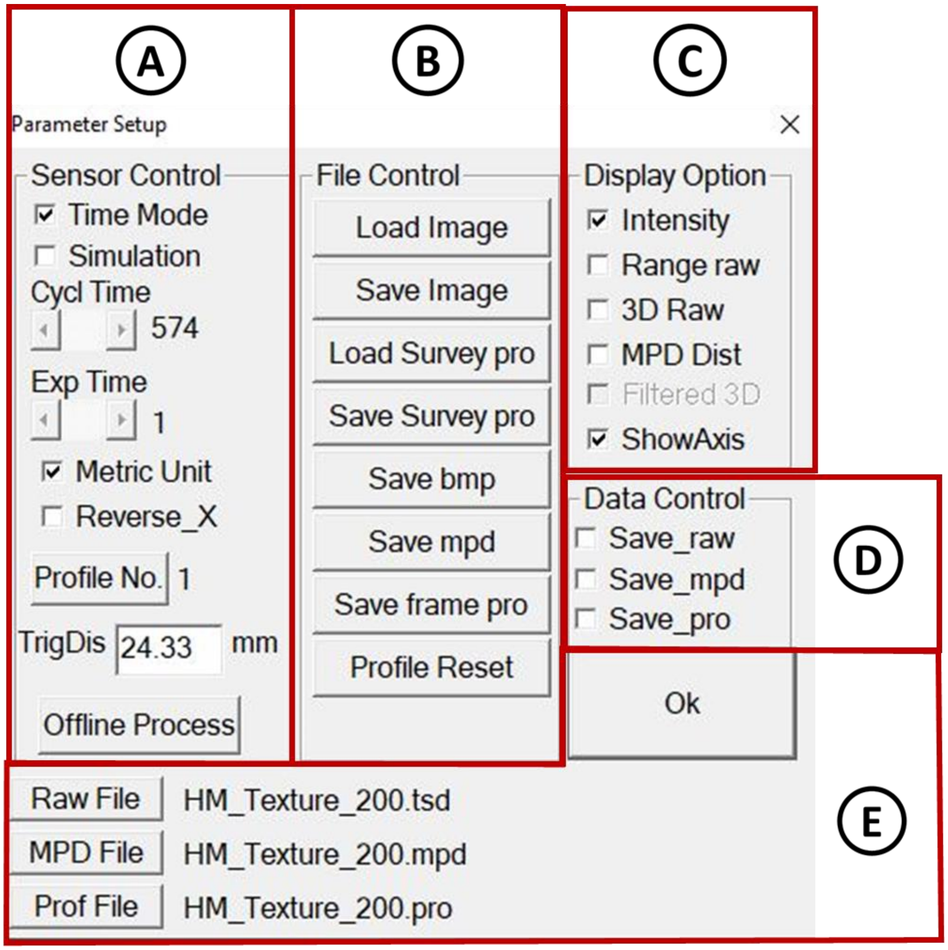


Figure 39*: Parameter setup dialog box*

Section A, Sensor Control, contains the settings that allow for configuration of the operation mode, cycle time, and exposure time, as well as options for reversing the X-axis and displaying units in metric.

- **Time Mode**: Allows the sensor to operate in either time mode or trigger mode. In time mode, the sensor's scan rate (Sf) is determined by the internal cycle time (Tc) and can be calculated using the formula: Sf = 1000000 / Tc, where Tc is in microseconds and Sf is in Hz. In trigger mode, an external TTL signal is required, with the trigger frequency not exceeding the internal cycle time's Sf.

- **Cycle Time**: Controls the sensor's scan speed in time mode. Adjusting this value changes the scan rate.

- **Exp Time**: Manages the sensor's exposure time. A higher exposure time results in a brighter laser line image but can cause motion blur at high speeds. It is recommended to keep the exposure time under 5 microseconds for speeds up to 70 mph to minimize motion blur.

- **Metric Unit**: Toggles between displaying units in metric or standard.
- **Reverse_x**: Reverses the X-axis of the profile.
- **Offline Process**: Used by the developer for offline processing.

Section B, File Control, focuses on managing the loading and saving of various data files, including options for saving images, survey data, and frame-specific data.

- **Load Image**: Loads a saved data file, which can be a single profile image frame (.3fd) or a stream profile image (.3vd).
- **Save Image**: Saves the current profile images to a .3fd file.
- **Load Survey Pro**: Loads previously saved PRO files.
- **Save Survey Pro**: Saves the current data as a PRO file.
- **Save bmp**: Saves the displayed profile images as a bitmap image file, containing either intensity or range data based on display options.
- **Save mpd**: Saves the Mean Profile Depth (MPD) data.
- **Save frame pro**: Saves only the current frame as a PRO file.

Section C, Display Options, enables users to toggle display settings for different types of data, such as intensity, range, 3D data, and MPD distribution.

- **Intensity**: Toggles the display between intensity and range data.
- **Range raw**: Displays raw range data.
- **3D Raw**: Displays raw 3D data.
- **MPD Dist**: Shows the MPD distribution.
- **Filtered 3D**: Shows filtered 3D data.
- **ShowAxis**: Displays axes on the profile.

Section D, Data Control, includes options for managing the saving of raw and processed data, and providing developer tools for calibration and offline processing.

- **Save_raw**: Saves only the current frame as a RAW file.
- **Save_mpd**: Enables saving MPD data for collection.
- **Save_pro**: Saves all profiles as a .csv file.

Section E, Additional Functions, focuses on options for loading and saving camera calibration files, and specifying file locations and names for texture, MPD, and profile data.

- **Raw File**: Allows users to choose the file location and name for texture data in .tsd format.
- **MPD File**: Allows users to choose the file location and name for MPD information in MPD format.
- **Prof File**: Allows users to choose the file location and name for profile data in PRO format.
- **OK**: Exits the Parameter Setup dialog and applies any change made to the software parameters.

4.1.3.5. Quick Operation Guide for the HM Texture Software

To effectively use the HM Texture software, follow these streamlined steps for setting up and initiating data collection. Ensure all settings are accurately configured to guarantee optimal performance during the survey process.

1) Launch the HM Texture software.
2) Confirm that the top right control area shows "DEFCAM" and "CalLoad."
3) Open the "Setup" menu.
4) In the "Setup" menu, adjust the "Time Mode" setting: Uncheck the box if using an encoder; otherwise, check "Time Mode."
5) In the "Setup" menu, select "Raw File" to specify the data saving location.
6) Enable the "Save_Raw" option in the "Setup" menu by checking the corresponding box.
7) Return to the main user interface.
8) In the control area, check the box next to "SaveData."
9) Click "Survey" to start data collection.

10) Click “Stop” or exit to end data collection.

Note that the software generates one .tsd file per batch of data. A “batch” consists of 15,360 profiles, the maximum number of profiles the data collection software can store in a single file.

4.1.3.6. Converting the Files to .csv Format

The UT_Texture software generates .tsd files by default, which can only be properly read and manipulated within the UT_Texture software. However, the the friction prediction application developed for this project requires data to be in a .csv format to operate correctly. To facilitate this, a separate application called “Convert CSV,” represented by the icon in **Figure 40**, is used to convert .tsd files to .csv format.

Figure 40*: Icon for the “Convert CSV” software executable*

The user must double-click on the “Convert CSV” executable and then the screen shown in **Figure 41** will appear on screen.

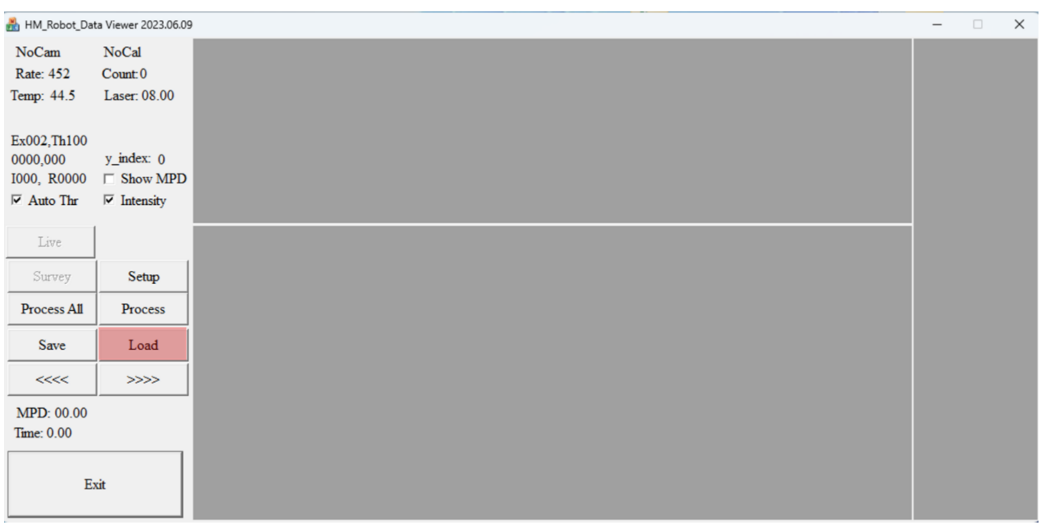


Figure 41*: “Convert CSV” main user interface window, with the “Load” button highlighted in red*

At this stage, the user must click on the “Load” button (highlighted in red) and select the .tsd file to be converted. The software can handle only one file at a time.

Once the file is loaded, the main screen will display the data, as shown in **Figure 42**.

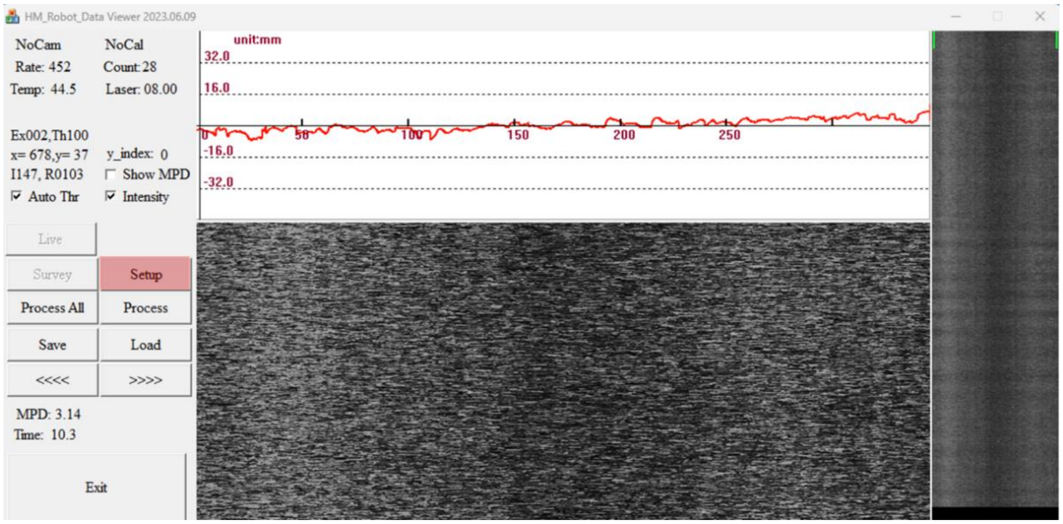


Figure 42: *User interface after loading a TSD file, with the "Setup" button highlighted in red*

Next, the user should click on the "Setup" button, highlighted in **Figure 42** . This action will bring up a window similar to the one shown in **Figure 43**.

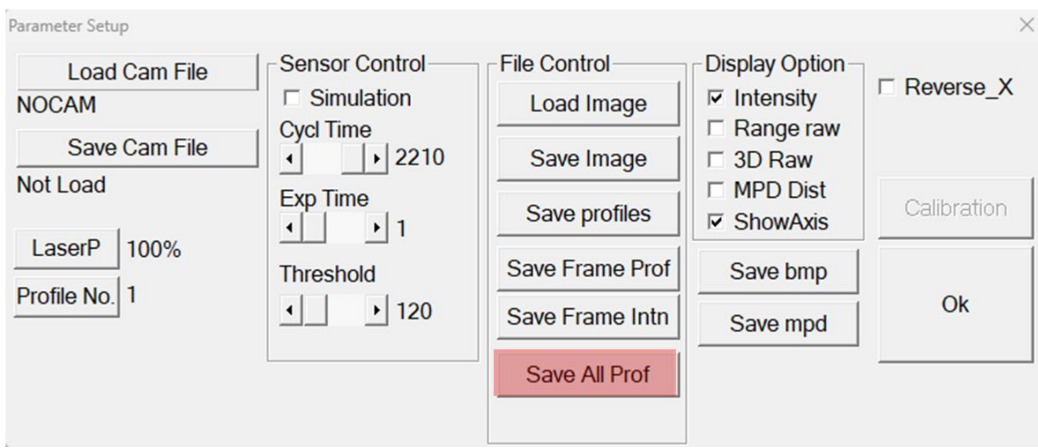


Figure 43: *Parameter setup window, with "Save All Prof" button highlighted in red*

The user should then click on the "Save All Prof", highlighted in red in **Figure 43**. This action will open a final dialog window (**Figure 44**) where the user can specify the name and location for the converted .csv file.

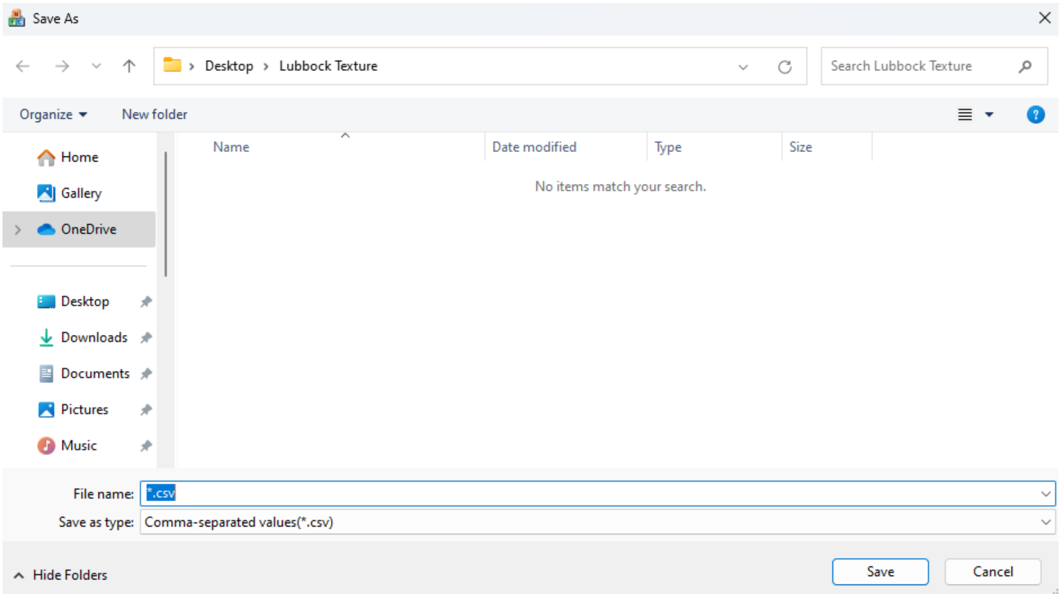


***Figure 44**: "Save as" dialog window for exporting the converted .csv file*

After clicking on the "Save" button, the software generates a .csv file where profiles are organized as row vectors. The first column indicates the frame index, the second column indicates the profile index within a frame, and the subsequent columns contain height values. The numeric values are whole integers; multiplying them by 0.01 will convert them to millimeters.

4.2. User Manual for Data Processing Application: TexTex FricFore

TexTex FricFore stands for The Texan Texture Friction Forecaster. This an application capable of processing raw field measurements of 2D profiles, compute a set of texture indices and then run a series of machine learning algorithms to output a prediction of surface friction in terms of both the Grip Number (GN) and Skid Number (SN).This user manual provides detailed instructions on how to navigate and use the combined application, including processing raw data, computing 2D texture indices, and predicting SN.

4.2.1. Installation and Setup

To use the TexTex FricFore, users should ensure they have a Windows Operating System with the .NET Framework installed. Additionally, R and the R.Net library must be installed on the system. Once these prerequisites are met, users can install the application executable file, by double clicking on the installer, as shown in **Figure 46**.

***Figure 45:** TexTex FricFore installer icon*

Upon double clicking on the installer, a window like the one shown in **Figure 46** will pop up. The user at this point clicks on the "Next >" button.

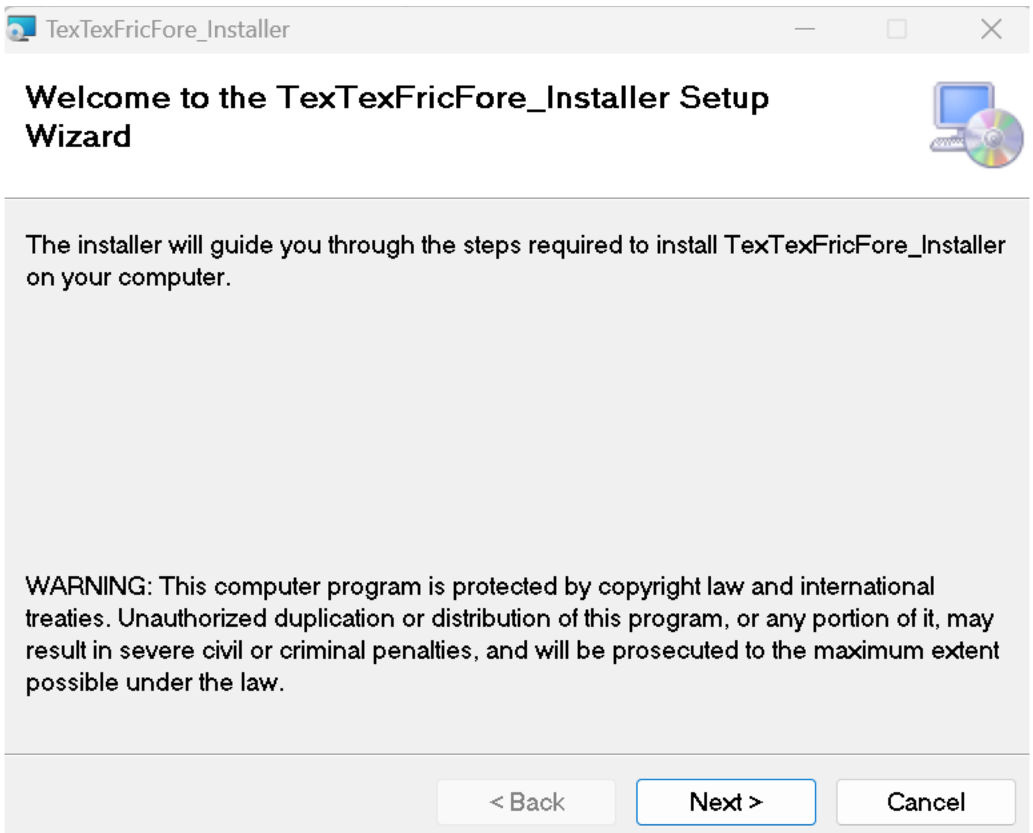


***Figure 46**: TexTexFricFore main installation window*

Next, the user will be taken to the select installation folder window (**Figure 47**), where the user specifies where the files for the application will be saved. The default installation path for the software will be in Program Files folder of the C drive of the computer. The user can also select whether the software will be installed to all user in the computer. The "Just me" option is chosen by default. The user will then click again on the "Next >" button.

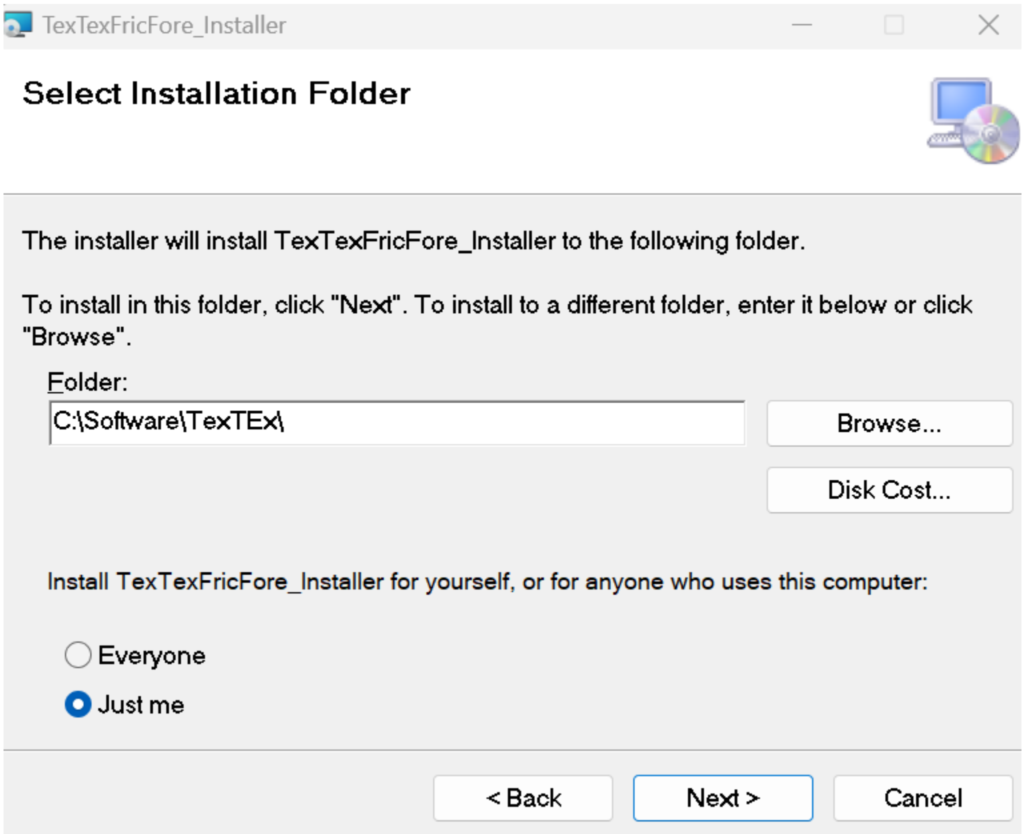


Figure 47*: Textex FricFore select installation folder window*

Once the user selects the installation folder, a confirmation window like the one on **Figure 48** will show up. The user can then click on the "Next >" to start the installation. Once the software is successfully installed a pop up will appear confirming that the installation was successful, at which point a desktop application shortcut will also be created with an icon like the one shown in **Figure 49**.

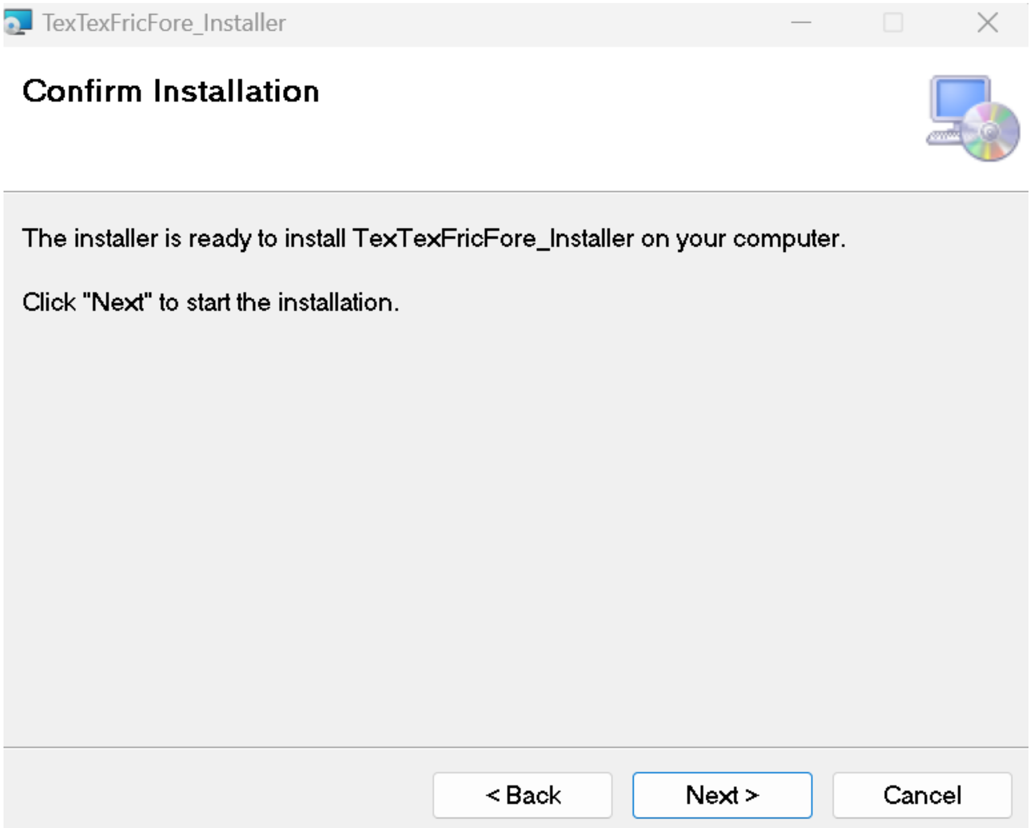


Figure 48*: Textex FricFore installation confirmation window*

Figure 49*: Textex FricFore desktop shortcut icon*

4.2.1. Launching the Application

To start the TexTex FricFore Application, users should double-click the application icon, shown in **Figure 49**. This action will open the main window of the application, which displays various buttons for navigating between different modules within the application, as shown in **Figure 50**.

Figure 50: *TexTex FricFore main window*

The main window comprises three main button which lead the user into the three modules of the application:

- **Process Raw Data Button**: This is the first module which is in charge of processing raw data from the sensor in .csv format.
- **Compute 2D Texture Indices Button**: This is the second module of the application and is in charge of computing a set of 2D texture indices, based on the output file from the first module.
- **Predict SN Button**: This is the third and final module which predicts the friction of the test section.

A more detailed description of each module is provided in the subsequent sections.

4.2.1.1. Processing Raw Data

The Processing Raw Data is the first module of the TexTex FricFore application designed for processing raw 2D profile data. The main window (**Figure 51**) of this module contains several buttons and controls for interacting with the data and performing various operations.

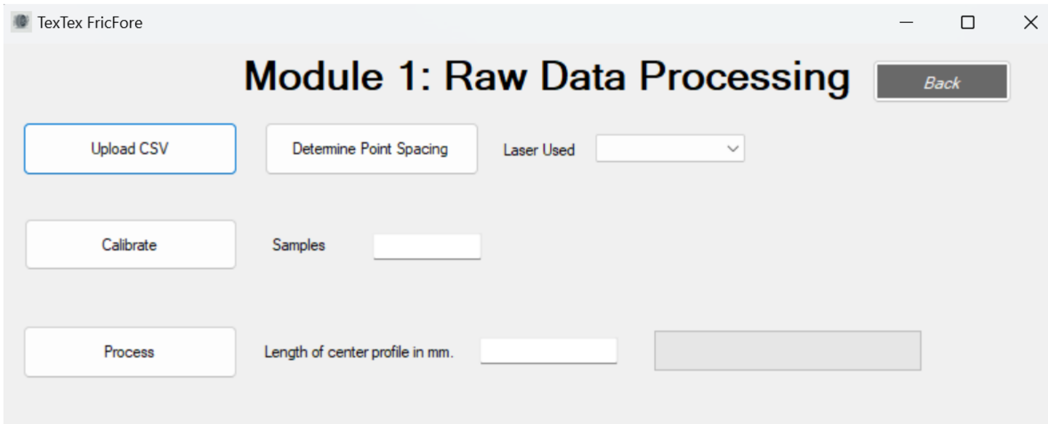


Figure 51*: Main window for the first module of the application used to process raw 2D profile data in .csv format.*

The key elements are:

- **Upload CSV Button:** This button enables users to upload a .csv file from their computer, containing the raw data to be processed.
- **Determine Point Spacing Button:** This button calculates the spacing in millimeters between data points in the measured profiles, using empirical curves determined through lab testing. To use this button, the user must first select a laser type from the dropdown on the right.
- **Laser Type Dropdown:** his dropdown allows users to select the type of laser used in data collection. If the laser sensor provided to the "Receiving Agency" is being used, select the "Current" option. The "Test" option is for a second laser sensor used during the initial stages of system calibration.
- **Back Button:** This button hides the current form and takes the user back to the main window.
- **"Samples" Fill-in Textbox**: Users must specify the number of sample profiles from the .csv file to be used for calibrating the profile processing algorithm. The Researchers recommends using at least 2,000 profiles for calibration. A larger sample size will provide more accurate calibration parameters but will take longer to process.
- **Calibrate Button:** This button runs a calibration subroutine to prepare the profile processing algorithm, using the number of samples specified in the textbox. It will not function until a valid number of profiles is inputted.
- **Process Profiles Button:** This button processes the raw profile data using the calibration parameters and spacing previously determined. It will not

execute until both the calibration and spacing subroutines have been completed.

- **“Length of center profile in mm.” Fill-in Textbox**: Users must specify the length of the processed profile after it has been trimmed by the processing algorithm. The Researchers recommends using 150 mm as a default value, but users can choose any value between 101 and 400.

- **Progress Bar:** This bar shows the progress of data processing operations.

4.2.1.1.1. Step-by-step to Properly Process Raw Data

The following is a step-by-step procedure for using the first module of this application:

- **Upload CSV File:** The user uploads an appropriate .csv file.

- **Select Laser Type:** The user selects “Current” as the laser type and clicks on the “Determine Point Spacing” button to compute the spacing between adjacent points in the measured profiles.

- **Define Number of Profiles:** The user specifies the number of profiles (samples) to be used for calibrating the algorithm; the recommended value is 2,000.

- **Set Profile Length:** The user defines the length of each processed profile in millimeters, with a recommended value of 150 mm.

- **Execute Algorithm:** The algorithm executes, and a progress bar displays the progress of the data processing. Once the algorithm is complete, the user specifies a location on the computer to save a .csv file with the processed data.

4.2.1.1.2. Output File

The output file generated by the TexTex’s first module includes three initial columns containing essential metadata alongside pavement surface elevations:

- **Spacing (mm)**: This column records the distance between adjacent points in millimeters.

- **Distance**: This index begins at zero and increments by one for each subsequent row, representing the cumulative count. Future modules will convert this index to distance in miles.

- **NA Counter (%)**: This column displays the percentage of data points within each profile that were identified as noise and subsequently removed.

Following these metadata columns, all other columns contain the fully processed pavement surface elevations, measured in millimeters. A sample view of the output files is shown in **Figure 52**.

A	B	C	D	E	F	G	H	I	J	K
Spacing	distance	na_counte	V3	V4	V5	V6	V7	V8	V9	V10
0.191334	0	2.45023	-0.62114	-0.55946	-0.51778	-0.42611	-0.48443	-0.26275	-0.07107	-0.1294
0.191334	1	3.828484	-0.10566	-0.07424	-0.03282	-0.18139	-0.21997	-0.23855	-0.12712	-0.0757
0.191334	2	4.287902	-0.43225	-0.33131	-0.27038	-0.15944	-0.27851	-0.13757	-0.22663	-0.2357
0.191334	3	3.828484	0.259527	0.149971	-0.11959	-0.34914	-0.3987	-0.51825	-0.50781	-0.49737
0.191334	4	4.441041	-0.25934	-0.04883	-0.08832	-0.20781	-0.2373	-0.24679	-0.17628	-0.28577
0.191334	5	3.981623	-0.79589	-0.63508	-0.49427	-0.52346	-0.43265	-0.41184	-0.39103	-0.54022
0.191334	6	5.053599	-0.15204	-0.03089	0.040257	0.001407	-0.04744	-0.19629	-0.10514	-0.02399
0.191334	7	5.513017	-0.00525	0.055985	0.077224	0.098462	0.089701	-0.00906	0.012178	0.033417
0.191334	8	4.287902	-0.18971	-0.12799	-0.16626	-0.23453	-0.1328	-0.38107	-0.45935	-0.42762
0.191334	9	3.675345	-0.11022	-0.0887	-0.18719	-0.19567	-0.08415	-0.12264	0.028875	-0.14961

Figure 52: *Sample view of the first module's output file*

4.2.1.2. Computing 2D Texture Indices

The Compute 2D Texture Indices is the second module of the TexTex FricFore application designed for computing a set of texture indices that characterize the texture in 2D surface profiles. The main window (**Figure 53**) of this module contains several buttons and controls for interacting with the data and performing various operations

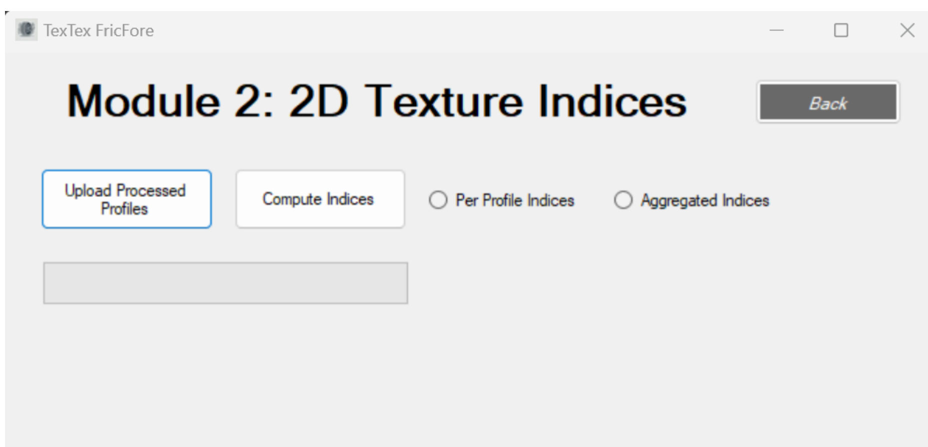


Figure 53*: Main window for the second module of the application used to compute 2D texture indices.*

The key elements are:

- **Upload Processed Profiles Button:** This button allows users to upload the .csv file containing the processed profiles, which is the output of the first module.

- **Compute Indices Button:** This button calculates the set of texture indices for further analysis. It will only execute after the user selects the required type of indices.

- **Index Type Selection:** This selection determines how the texture indices are computed. They can be calculated for each profile individually ("Per Profile Indices") or aggregated over every 0.1 miles ("Aggregated Indices"). If "Aggregated Indices" is selected, the software will first compute the indices for each profile and then report the ***median*** of each index every 0.1 miles. Moreover, the next module on predicting friction *requires* the aggregated indices.

4.2.1.2.1. Step-by-step to Properly Compute Texture Indices

The following is a step-by-step procedure for using the second module of this application:

- **Upload CSV File:** The user uploads the processed profiles .csv file generated by the first module of this application.

- **Select Index Type:** The user selects "Aggregated Indices" as the type of index to be computed. "Per Profile Index" was added as an option to study the distribution of the texture across the pavement section tested.

- **Execute Index Computation:** The algorithm runs, and a progress bar indicates the progress of the index computation. Upon completion, the user specifies a location on the computer to save a .csv file containing the 2D texture indices.

4.2.1.2.2. Output File

The output file generated by TexTex's second module, which provides aggregated indices, includes three initial columns with essential metadata, followed by a series of aggregated 2D texture indices:

- **n**: This column records the number of profiles included in the 0.1-mile pavement section.

- **Mile_group**: This variable represents the 0.1-mile segment since the sensor began collecting data. For this measurement to be accurate, the vehicle must be traveling at 50 mph.

- **Spacing**: This column records the distance between adjacent points in millimeters.

Following these metadata columns, all other columns contain a set of twelve 2D texture indices, specified in previous technical memorandums. A sample view of the output files is shown in **Figure 54**.

A	B	C	D	E	F	G	H	I	J
n	Mile_group	Spacing	RMS_med	Rku_med	Rso_med	MPD_med	Cwm_med	Cwst_med	Cws_med
20514	0.1	0.191334	0.168964	0.310913	-0.94745	0.411541	0.992915	1.567066	3.68208
20517	0.2	0.191334	0.173153	0.311259	-0.90646	0.488752	0.997548	2.302986	3.801083
20517	0.3	0.191334	0.172392	0.315521	-0.93442	0.418088	1.736757	1.844345	4.405983
20517	0.4	0.191334	0.169236	0.312955	-0.90971	0.456118	1.381153	1.859426	3.739393
20518	0.5	0.191334	0.170738	0.313574	-0.89828	0.488077	1.692441	1.806614	4.292692
20518	0.6	0.191334	0.172182	0.313314	-0.92196	0.493828	1.373428	2.439067	3.744226
20518	0.7	0.191334	0.172615	0.314147	-0.90582	0.418545	1.757074	2.081982	4.508459
20516	0.8	0.191334	0.171647	0.312675	-0.92765	0.474238	1.325857	1.613078	4.124104
20516	0.9	0.191334	0.16976	0.313967	-0.91146	0.45906	1.666625	2.004672	4.540264
20517	1	0.191334	0.168999	0.314697	-0.90381	0.48827	1.651191	1.77414	4.311439

Figure 54: *Sample view of the second module's output file*

4.2.1.3. Predicting SN

The Predict SN module is the third and final component of the TexTex FricFore application, designed to determine the friction of the test pavement section. When the user clicks the "Predict SN" button in the main window of TexTex FricFore, a pop-up window will confirm that both the classification and regression models required for predicting roadway friction have been successfully loaded. The main window (**Figure 55**) of this module contains buttons and controls for interacting with the data.

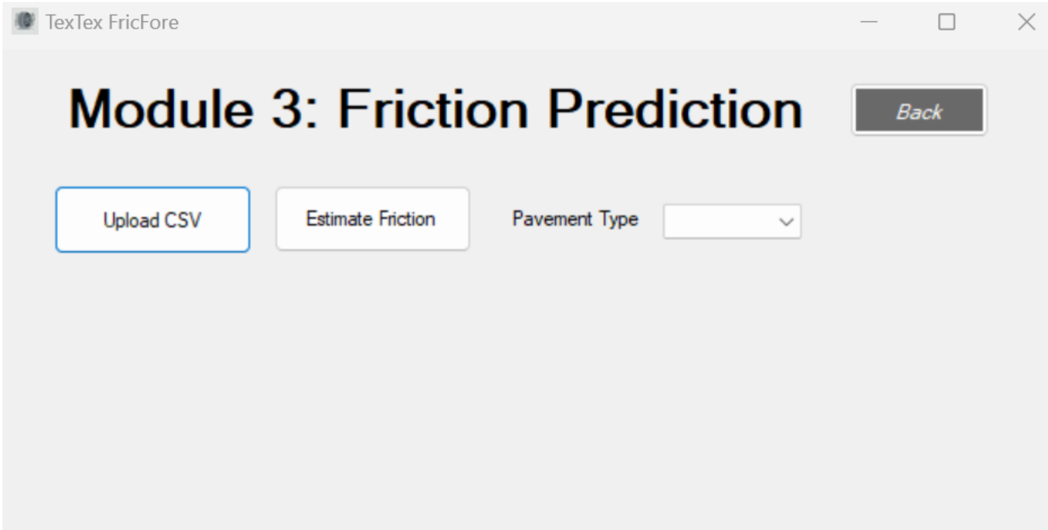


***Figure 55**: Main window for the third module of the application used to compute predict SN.*

The key elements are:

- **Upload Aggregated Texture Indices:** This button allows users to upload the .csv file containing the aggregated texture indices, which is an output of the second module.

- **Estimate Friction Button:** This button runs the classification and regression models to estimate friction on the test section. It will only execute once the user has selected a pavement type.

- **Pavement Type Selection:** This dropdown allows users to select the type of pavement tested, with options for "Flexible" and "Rigid." Users must ensure that the test section does not include a change from flexible to rigid pavement that exceeds 0.1 mile. Changes within the same pavement type, such as from a porous friction course (PFC) to a dense mix or from a diamond grinding surface to transverse tining, are acceptable.

4.2.1.3.1. Step-by-step to Properly Compute Texture Indices

The following is a step-by-step procedure for using the second module of this application:

- **Upload CSV File:** The user uploads the aggregated texture indices .csv file generated by the second module of this application.

- **Select Pavement Type:** The user selects either "Flexible" or "Rigid" as their pavement type from the dropdown menu.

- **Execute Algorithm:** The algorithm executes. Once the algorithm is complete, the user specifies a location on the computer to save a .csv file with the predicted friction.

4.2.1.3.2. Output File

The output file generated by the TexTex's third module includes three initial columns containing essential metadata alongside friction indices for the pavement section:

- **Mile_group**: This variable represents the 0.1-mile segment since the sensor began collecting data. For this measurement to be accurate, the vehicle must be traveling at 50 mph.

- **MPD (mm.)**: This column records the median MPD value seen across the pavement section, in units of millimeters.

- **Surface Description**: This variable outputs the predicted pavement surface type based on the cluster analysis conducted in this study. The clusters will be consistent with the Pavement Type originally specified by the user.

- **GN**: This variable is the predicted Grip Number, as measured by the GripTester. The lower and upper bounds for its predicted value are also provided, and those are based on the standard error of prediction.

- **SN**: This variable is the predicted Skid Number, as measured by the Locked Wheel Skid Tester. The lower and upper bounds for its predicted value are also provided, and those are based on the standard error of conversion from GN to SN, using the relationships developed in this study.

Following these metadata columns, all other columns contain the fully processed pavement surface elevations, measured in millimeters. A sample view of the output files is shown in **Figure 56**.

A	B	C	D	E	F	G	H	I
Mile_group	MPD (mm.)	Surface Description	GN_lower_bound	GN	GN_upper_bound	SN_lower_bound	SN	SN_upper_bound
0.1	1.26	New Generation Diamond Grinding	0.32	0.37	0.42	0.292	0.34	0.388
0.2	1.27	New Generation Diamond Grinding	0.32	0.37	0.42	0.292	0.34	0.388
0.3	1.32	New Generation Diamond Grinding	0.32	0.37	0.42	0.292	0.34	0.388
0.4	1.32	New Generation Diamond Grinding	0.31	0.36	0.41	0.282	0.33	0.378
0.5	1.28	New Generation Diamond Grinding	0.32	0.37	0.42	0.292	0.34	0.388
0.6	1.17	New Generation Diamond Grinding	0.33	0.38	0.43	0.312	0.36	0.408
0.7	1.16	New Generation Diamond Grinding	0.33	0.38	0.43	0.312	0.36	0.408
0.8	1.2	New Generation Diamond Grinding	0.31	0.36	0.41	0.282	0.33	0.378
0.9	1.16	New Generation Diamond Grinding	0.31	0.36	0.41	0.282	0.33	0.378
1	1.21	New Generation Diamond Grinding	0.29	0.34	0.39	0.242	0.29	0.338
1.1	0.76	Tining with deep grooves	0.35	0.4	0.45	0.172	0.22	0.268

Figure 56: *Sample view of the third module's output file*

4.2.2. Navigation and User Interface

The TexTex FricFore Application provides a seamless interface for navigating between different modules. Each module opens in a new window, and the main form is hidden during module operation to reduce clutter. To return to the main form from any module, users can click on the “Back” button within each of the modules. The main form will automatically reappear, allowing users to select another module or exit the application.

4.2.3. Handling Errors

The application handles errors gracefully by displaying appropriate messages for different issues. If any error occurs while using a module, an error message will indicate the problem. Each module has its own error-handling mechanism as described in their respective user manuals.

4.2.4. Exiting the Application

To exit the application, users should close the main window or any open module windows. The application ensures that all resources are properly disposed of upon closing.

4.2.5. Contact and Support

For further assistance, users can contact the Researchers’s principal investigator at prozzi@mail.utexas.edu, and carbon copy the supporting researchers at christian.sabillon@utexas.edu, and danilokeniti@utexas.edu.

This manual provides a comprehensive guide to using the TexTex FricFore Application effectively. If users encounter any issues or have further questions, they should not hesitate to reach out to the Researchers.

4.3. Video Demonstration

The Researchers has also created a 10-to-15-minute video summarizing the contents within this Technical Memorandum which will be electronically submitted to TxDOT RTO Office.

Chapter 5. Conclusions and Recommendations

5.1. Conclusions

As part of this project, a network-level statistical model for estimating friction in terms of Grip Number (GN) or Skid Number (SN) was developed, targeting the diverse pavement surfaces commonly encountered on Texas highway network. By rigorously analyzing pavement texture data and employing advanced artificial intelligence and statistical techniques, the researchers established a quantifiable relationship between surface texture characteristics and friction. Using a specialized data collection prototype, extensive field data was captured across various pavement types and climate conditions. This comprehensive approach enabled the creation of a robust predictive model for friction, paving the way for future models to significantly enhance pavement safety management.

The project highlighted the limitations of traditional methods like linear regression in developing empirical models using network-level data and heterogeneous pavement surfaces. These limitations included underfitting due to the inherent heterogeneity of different pavement types and the need for frequent recalibration due to surface deterioration. Traditional methods also require either a truly representative sample of every pavement surface or modeling the entire population, which is impractical.

To address these issues, the project team utilized a comprehensive database of pavement surfaces across Texas, ensuring the models' applicability to diverse pavement types statewide. Furthermore, the researchers employed a sophisticated machine learning framework that combined unsupervised and supervised learning techniques. The process began with cluster analysis to group pavement sections with similar characteristics, which informed the development of regression models for predicting pavement friction. These clusters then served as labels to train a classification algorithm for identifying similar clusters in new datasets. Finally, an advanced ensemble linear regression model was employed to capture the complex relationships between texture indices, surface information, and pavement friction, outperforming a deep neural network (DNN) regression model.

The clustering analysis identified eight distinct clusters for flexible pavements and five for rigid pavements. For flexible pavements, clusters ranged from polished and flat open-graded mixes to high-friction surfaces and high-macrotexture surfaces with uniform gradation. These groups highlighted the diverse textural and frictional properties across different types of flexible pavements, emphasizing that cluster differences stem from a combination of pavement mix design, gradation, and surface polishing and aging levels. For rigid pavements, clusters included new generation diamond grinding, conventional diamond grinding, and various tining techniques, reflecting significant differences in frictional-textural relationships based on texturing methods.

The classification analysis utilized twelve different machine learning algorithms to categorize the previously identified texture-friction clusters. The random forest (RF) model achieved the highest performance, with an F1 score of 0.903, demonstrating superior accuracy in identifying and classifying pavement clusters. This effectiveness is attributed to its robust mechanism of building multiple decision trees and averaging their predictions, enhancing accuracy and reducing overfitting. Conversely, the extreme gradient boosting (XGB) model recorded the lowest performance with an F1 score of 0.080, due to its limitations in correcting errors made by weak learners given the data structure. Additionally, the decision tree (DT) model failed to compute an F1 score as it could not predict any instances of clusters 6 and 8 within the flexible pavement clusters, highlighting its inadequacy in handling complex classification tasks.

The final ensemble multiple regression model, tailored for different texture-friction clusters, achieved an overall R^2 of 0.767, with mean absolute error (MAE) and root mean square error (RMSE) values of 0.050 and 0.071, respectively. In comparison, an alternative DNN model showed a goodness-of-fit of 0.669, with MAE and RMSE values of 0.066 and 0.085. While the DNN model also provided robust predictions, the ensemble multiple regression models demonstrated higher goodness-of-fit and lower error rates. This makes the ensemble approach more desirable for practical applications, as it combines the strengths of multiple models to deliver more precise and reliable predictions, crucial for effective pavement maintenance and safety management in Texas.

Lastly, the researchers integrated these algorithms into a standalone application called the Texan Texture Friction Forecaster (TexTex FricFore). TexTex FricFore is compatible with any Windows operating system that has the R programming language installed. The software can process raw data from the laser sensor in .csv format, performing data processing, index computation, and predict friction in terms of SN using the model classification and regression models outlined in this project.

5.2. Recommendations

The research identified several areas for future work to enhance pavement friction prediction and safety management. A key element would consist of improving the application's computational efficiency to enable real-time friction prediction as the data collection vehicle gathers texture measurements. Additionally, investing in advanced laser technologies with higher resolution and sampling rates is essential. These enhancements will allow for the capture of finer surface details, including microtexture, thereby increasing the predictive accuracy of the models.

Future work should aim at simplifying the approach by reducing the number of clusters or simplifying the regression equations. A sensitivity analysis should be conducted to understand the impacts of transitions between clusters on friction predictions. Incorporating pavement age, traffic volume, and type would enhance model accuracy and goodness-of-fit as well as extending the analysis into performance modelling as opposed to condition prediction.

The implications of this work are significant for pavement engineering and public safety. The proposed modeling approach could significantly contribute to the prevention of road crashes at the network level, particularly under wet-weather conditions, thereby reducing the economic impact of traffic incidents and potentially saving lives. Additionally, the environmental benefits of reducing resource consumption of water and tires for traditional friction testing methods mark a shift towards more sustainable infrastructure management practices.

Appendix A: Value of Research

This appendix provides a detailed analysis of the economic benefits derived from TxDOT Research Project 0-7031-01. The analysis evaluates the long-term effects of sponsoring this research project on a macroeconomic level, employing a net present value (NPV) cost-benefit analysis template provided by TxDOT to calculate monetary savings.

Motor-vehicle accidents can be attributed to numerous causes, which generally fall into three broad categories: driver-related, vehicle-related, and highway condition-related (24). While a state highway agency cannot fully prevent accidents due to driver-related, vehicle-related, or environmental factors, it can mitigate crashes in all three categories by improving pavement surface friction (25–28).

The primary goal of this project is to enhance safety across Texas' highway network by annually assessing friction conditions, allowing for timely maintenance scheduling whenever friction levels fall below a specified threshold. The reduction in skid incidents due to wet-weather conditions is of particular interest, as enhancing pavement road surface friction reduces the chances of hydroplaning and has the potential to prevent up to 70% of wet surface accidents (29–31). In the United States, wet-weather crashes contribute significantly to annual economic losses and can result in severe injuries or fatalities. Ensuring adequate friction is therefore crucial for public safety.

The macroeconomic analysis in this appendix estimates the average economic loss associated with motor-vehicle accidents in Texas and predicts the monetary savings achievable through comprehensive annual assessments of friction levels. This data allows for better allocation of funds to preventive maintenance projects, restoring skid resistance to serviceable levels and ultimately reducing the number of crashes.

Average Economic Loss Associated with Motor Vehicle Accidents

The National Safety Council in the United States has provided a guide for calculating the average costs associated with motor-vehicle injuries based on accident severity (32), as outlined in Table 18. These estimates include wage and productivity losses, medical expenses, administrative costs, motor-vehicle damage, and employer's uninsured expenses. However, these figures are conservative as they exclude the value of lost quality of life, which requires empirical calculation (32). Originally based on 2022 dollars, these costs have been converted to 2024 dollars using data from the Bureau of Labor Statistic's CPI Inflation Calculator.

This tool estimates a 7.3% price index adjustment to account for cumulative inflation trends from 2022 to 2024 (33).

Table 18: Average Economic Cost by Injury Severity or Crash in 2022 (32)

Crash Severity	Ave. Economic Cost in 2019 (in millions of dollars)	Ave. Economic Cost in 2024 (in millions of dollars)
Death	$1.869	$2.000
Disabling	$0.162	$0.173
Evident	$0.042	$0.045
Possible	$0.026	$0.028
No Injury	$0.007	$0.007
Property Damage	$0.006	$0.006

To estimate the total monetary costs of motor-vehicle accidents in Texas, the research team utilized the TxDOT 2021 report on "Rural and Urban Crashes and Injuries by Severity". As the current cost for crashes in Texas is unknown, the 2021 number of crashes were used for this analysis. The total expected cost of motor-vehicle accidents in Texas, across all levels of accident severity, was calculated by multiplying the average economic cost in 2024 dollars by the number of accidents for each severity level. Table 19 summarizes the number of crashes in Texas by severity and the expected monetary cost for each accident.

Table 19: Summary of Urban and Rural Crashes and Injuries by Severity for 2021 in Texas

Crash Severity	Number of Crashes in 2021	Expected Monetary Cost (in millions of 2024 dollars)
Death	4,028	$ 8,911
Disabling	15,771	$ 2,733
Evident	60,621	$ 2,724
Possible	87,105	$ 2,437
No Injury	360,976	$ 2,703
Property Damage	24,145	$ 155
Total Expected Cost (in millions of 2024 dollars)		**$19,665**

Assuming the number of crashes in Texas across each category remains relatively unchanged from 2021, the total expected cost incurred by both the state and its constituents is estimated to be approximately $19.665 billion.

Potential Outcome of Project Implementation

Consider the case where the products from this project are implemented, enabling TxDOT to obtain reliable friction estimates across the entire highway network. With comprehensive friction information, the maintenance division can identify low skid resistance hotspots and address these issues by applying high-friction chip seals to improve road friction. According to the results for TxDOT Project 0-7031 (4) applying a fresh seal coat (referred to in the study as “high macrotexture chip seal”) to a seal coat that is drastically polished distressed (referred to in the study as “low macrotexture chip seal”) results in an average improvement of approximately 0.13 in SN. This conservative estimate suggests that a seal coat on a low friction road would increase the SN by 13 points, although the actual improvement could be higher.

To translate this improvement in skid number to a reduction in crashes, the study by Burchett and Rizenbergs (34) was referenced. This study quantified the effect of skid resistance on wet weather accidents, finding a significant increase in wet pavement accidents when SN70R (skid number measured at 70 km/h using a ribbed tire) drops below 27. The fitted model from this study, shown in **Figure 57**, was used to estimate the percentage of wet weather accidents expected based on the skid number. The percentage of wet weather accidents before and after treatment was calculated to be 33.6% and 23.8%, respectively. This indicates that applying a seal coat to improve skid resistance could result in an average reduction of 9.7% in wet weather accidents on treated pavement sections. Conservatively, assuming an average reduction of 2.5% in wet weather crashes due to all maintenance work across the network is a prudent estimate, which is about a quarter of what the model in **Figure 57** predicts.

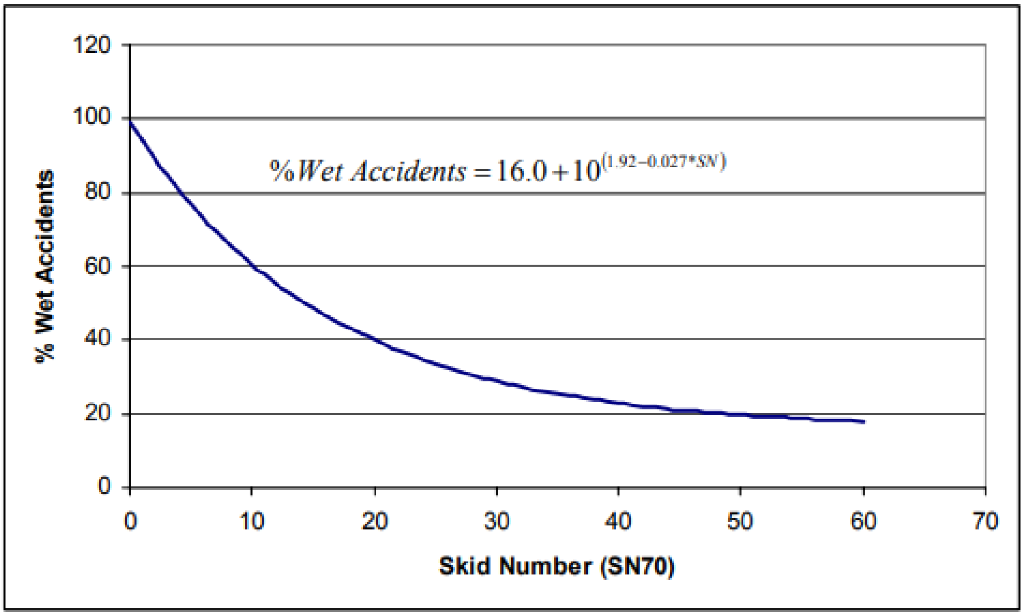


Figure 57: *Percentage of Wet Crashes Varying with SN70R (34)*

Data from TxDOT's Crash Statistics database indicates that out of the total crashes that occurred in Texas in 2021, about 9% occurred during wet weather conditions (35). Meaning that out of the total of 552,646 crashes 49,491 were wet weather-related crashes. It is now assumed that dry weather crashes remain unchanged. Given these assumptions, the number of wet and total accidents after TxDOT treats low skid sections are 48,254 and 551,409, respectively. All that remains is to match the proportion of crashes across each category to the ones observed in 2021 and the expected monetary cost after treatment can be computed. Table 20 summarizes all the information needed to compute the total expected monetary cost after implementation.

Table 20: Summary of information to compute total expected monetary cost after implementation

Parameter	Value
Number of accidents in 2021 (36)	552,646
Wet weather accidents[5] (35)	49,491
Avg. reduction in wet weather accidents	2.5%
Wet weather accidents after treatment	48,254
Number of accidents after treatment	551,409

Accident Severity	Proportion of Total	Predicted Number of Crashes	Expected Monetary Cost (in millions of 2024 dollars)
Death	0.01	4,019	$ 8,037
Disabling	0.03	15,736	$ 2,728
Evident	0.11	60,485	$ 2,718
Possible	0.16	86,910	$ 2,418
No Injury	0.65	360,168	$ 2,698
Property Damage	0.04	24,091	$ 155
Total Cost (in millions of 2024 dollars)			$18,753

After maintenance improvements, the total cost for the state and its constituents is $18.753 billion, resulting in savings of approximately $900 million at a macroeconomic statewide level. These substantial savings are primarily due to the high costs associated with vehicular crashes, particularly any injuries that evident, disabling or result in death. It is important to note that construction costs related to the maintenance work were not included in these estimates. To avoid overestimation, the Researchers will use only 0.5% of the total savings going forward, amounting to $4.5 million.

NPV Cost Benefit analysis

The final economic analysis was conducted by using the Excel template provided by TxDOT. This template performs a net present value (NPV) cost-benefit analysis by considering:

[5] Considered wet weather if conditions during crash involved: rain, sleet, hail, or snow

1. *Project budget*: the total amount of money allocated to finance this research project, measured in dollars,

2. *Project duration*: the agreed upon timeframe for project completion, measured in years,

3. *Expected value per year*: An estimation of the annual savings incurred by TxDOT after implementing the project's products, measured in dollars,

4. *Expected value duration*: the timeframe over which this economic analysis is conducted, measured in years, and

5. *Discount rate*: the interest rate used in discounted cash flow analysis to determine the present value of future cash flows, measured as a percentage.

Many of the inputs were dictated by TxDOT or could not be varied as they were based on values from the contract; however, there are two terms, Exp. Value (per Yr.) and Expected Value Duration (Yrs.), which the Researchers had full freedom to vary. The values associated with those two terms (highlighted in yellow at the top of **Figure 58**) governed the outputs of the economic analysis.

Inputs for the Economic Analysis

The project budget was set to $357,134.51. This value was the agreed upon budget as stipulated in the project's contract. The input project duration was 2.00274 years. The University Handbook (37) states that the project duration is not rounded. The project commenced on September 1, 2022, and the termination date is August 31, 2023. There are 731 days from the start date of the project to the end date (with the end date included), which equates to 2.00274 years. The expected value per year was $4.5 million. This input is the total savings that were computed for the macro analysis. As mentioned in the previous sections, most of the values used were as conservative as possible. The expected value duration of the project was assumed to be 10 years to reflect a potential timeframe between the inception of this project and the time it takes for it to be fully implemented. Finally, the input for the discount was 5% as recommended by the University Handbook (37). The inputs and outputs of this economic analysis can be seen at the top of **Figure 58** and a graphical representation of the NPV measured in millions of dollars over the timeframe of the economic analysis can be seen at the bottom of **Figure 58**.

	Project #	0-7031-01		
	Project Name:	Research and Technology Implementation: Develop Efficient Prediction Model of Highway Friction on an Annual Basis on Texas Network		
	Agency:	CTR	Project Budget	$ 357,135
	Project Duration (Yrs)	2.00	Exp. Value (per Yr)	$ 4,500,000
Expected Value Duration (Yrs)		10	Discount Rate	5%
Economic Value				
Total Savings:	$ 44,642,865		Net Present Value (NPV):	$ 36,698,142
Payback Period (Yrs):	0.079363		Cost Benefit Ratio (CBR, $1 : $___):	$ 103

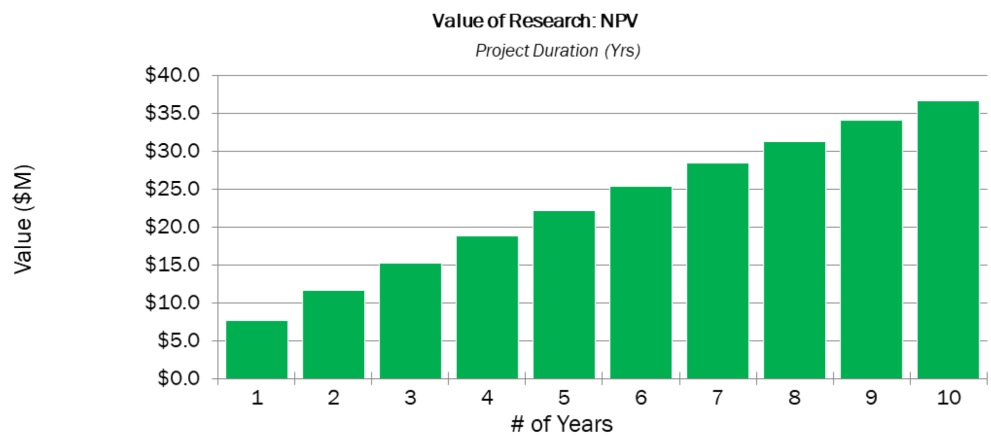


Figure 58*: Inputs and outputs of NPV cost-benefit analysis (top), graphical representation of NPV over the course of ten years (bottom)*

Conclusion

The expected economic savings generated by this project for road users and the Receiving Agency are estimated to be approximately $35 million over ten years. From a purely mathematical perspective, the project pays for itself in just a small fraction of a year (0.079) and boasts a cost-benefit ratio of 103. Such high values are common for projects focused on driver safety. The recommended technology from this project enables continuous friction measurements at any distance along the Texas' highway network, providing comprehensive information on friction. This allows TxDOT to identify and apply corrective measures more effectively, resulting in substantial cost savings for the Receiving Agency and improved safety for road users.

References

1. Goenaga B, Underwood BS, Castorena C, Rogers P. Early Friction and Texture Evolution After an Asphalt Overlay. Transp Res Rec J Transp Res Board. 2023;19–26.

2. ASTM. ASTM E2340: Test Method for Measuring the Skid Resistance of Pavements and Other Trafficked Surfaces Using a Continuous Reading, Fixed-Slip Technique. ASTM International; 2015. Available from: http://www.astm.org/cgi-bin/resolver.cgi?E2340E2340M-11R21

3. Zuniga-Garcia N. Prediction Friction with Improved Texture Characterization. Austin: The University of Texas at Austin; 2017.

4. Sabillon C, Hernandez J, Li R, Huang R, Prozzi J. Efficient Model for Predicting Friction on Texas Highway Network. Austin: Texas Department of Transportation; 2023.

5. Sabillon-Orellana. Evaluation of Data Imputation Techniques in Pavement Texture Processing. (Master Thesis). Austin: The University of Texas at Austin; 2020.

6. Katicha S, Mogrovejo DE, Flintsch GW, De Leon Izeppi E, Carrasco DM. Latest Development in the Processing of Pavement Macrotexture Measurements of High Speed Laser Devices. 9th International Conference on Managing Pavement Assets; 2015.

7. ISO. ISO 13473-1: Characterization of Pavement Texture by Use of Surface Profiles - Part 1: Determination of mean profile. Geneva: International Organisation for Standardisation; 2009.

8. Katicha SW, Mogrovejo DE, Flintsch GW, De Leon Izeppi ED. Adaptive Spike Removal Method for High-Speed Pavement Macrotexture Measurements by Controlling the False Discovery Rate. Transp Res Rec J Transp Res Board. 2019;2525:100–10.

9. ASTM. ASTM E178-21: Standard Practice for Dealing with Outlying Observations. West Conshohocken: American Society for Testing and Materials International; 2021.

10. ASTM. ASTM E1845-15: Standard Practice fro Calculating Pavement Macrotexture Mean Profile Depth. West Conshohocken: American Society for Testing and Materials International; 2015.

11. ISO. ISO 13473-4: Characterization of Pavement Texture by Use of Surface Profiles – Part 4: One Third Octave Band Spectral Analysis of Surface Profiles. Geneva: International Organisation for Standardisation; 2024.

12. NASEM. Protocols for Network-Level Macrotexture. Washington, DC: The National Academies Press; 2021.

13. Cooley J, Tukey J. An Algorithm for the Machine Calculation of Complex Fourier Series. Math Comput. 1965;297–301.

14. Butterworth S. On the Theory of Filter Amplifiers. Exp Wirel Wirel Eng. 1930;536–41.

15. Gottaut C, Goubert L. Texture-based Descriptors for Road Surface Properties and How They can be Used in the Appropriate Standards. Vienna: Rosanne; 2016. Report No.: P7-SST-2013-RTD-1.

16. ISO. ISO 21920-2: Geometrical Product Specifications (GPS) - Surface Texture: Profile - Part 2: Terms, Definitions and Surface Texture Parameters. Geneve: International Organisation for Standardisation; 2021.

17. Gadelmawla ES, Koura MM, Maksoud TMA, Elewa IM, Soliman HH. Roughness Parameters. J Mater Process Technol. 2002;123:133–45.

18. Gordon AD. Classification. 1st ed. London: Chapman and Hall CRC; 1980.

19. Nielsen F. Introduction to HPC with MPI for Data Science. Cham: Springer International Publishing; 2016 [cited 2024 Jun 14]. Available from: http://link.springer.com/10.1007/978-3-319-21903-5

20. Sibson R. SLINK: An optimally efficient algorithm for the single-link cluster method. Comput J. 1973 Jan 1;16(1):30–4.

21. Sabillon-Orellana C, Prozzi JA, Hernandez J. Cobertura total de la friccion de pavimentos asfalticos a nivel de red. Asf Paviment. 2023;8(51):13–21.

22. Sabillon-Orellana CA, Hernandez JB, Bassil M, Prozzi JA. Classification of Flexible Pavements Based on Texture Data. Transp Res Rec. 2022;2676:140–51.

23. Sabillon-Orellana CA, Li R, Hernandez JB, Prozzi JA. Evaluation of the Relationship between Friction, Texture, and Noise. Okemos: United States Department of Transportation; 2021.

24. Vester T, Fourie E. The Good, the Bad and the Ugly of South African Fatal. South Afr J Sci. 2018;1–7.

25. Cairney P. Skid Resistance and Crashes - A Review of the Literature. Vermont South: ARRB Transport Research; 1997.

26. Giles CG, Sabey BE, Cardew KHF. Development and Performance of the Portable Skid-Resistance Tester. ASTM International; 1962.

27. Rizenbergs RL, Burchett JL, Napier CT. Skid Resistance of Pavements. Frankfort: Department of Highways Commonwealth of Kentucky; 1972.

28. Viner H, Sinhal R, Parry T. Review of UK Skid Resistance Policy. In Toronto; 2004.

29. NASEM. Accident Modification Factors for Traffic Engineering and ITS Improvements. Washington D.C.: National Academies Press; 2008.

30. De Leon Izeppi E, Katicha SW, Flintsch GW, McGhee KK. Pioneering use of Continuous Pavement Friction Measurements to Develop Safety Performance Functions, Improve Crash Count Prediction, and Evaluate Treatments for Virginia Roads. Transp Res Rec. 2016;81–90.

31. Merritt DK, Lyon CA, Persaud BN, Torres HN. Developing Crash-Modification Factors for High Friction Surface Treatments. Mclean: Federal Highway Administration; 2020.

32. NCS. Costs of Motor-Vehicle Crashes [Internet]. Injury Facts. 2022 [cited 2024 Jun 28]. Available from: https://injuryfacts.nsc.org/all-injuries/costs/guide-to-calculating-costs/data-details/

33. BLS. Bureau of Labor Statistics. 2024 [cited 2024 Jun 28]. CPI Inflation Calculator. Available from: https://data.bls.gov/cgi-bin/cpicalc.pl?cost1=1&year1=201901&year2=202405

34. Burchett J, Rizenbergs R. Frictional Performance of Pavements and Estimates of Accident Probability. In: Pavement Surface Characteristics and Materials. ASTM, 1982. p. 73–97. Available: https:/asmedigitalcollection.asme.org/astm-ebooks/book/1398/chapter/27801487/

35. TxDOT. Weather Conditions for Crashes. Texas Department of Transportation; 2021. Available from: https://www.txdot.gov/content/dam/docs/trf/crash-reports-records/2021/24.pdf

36. TxDOT. Rural and Urban Crashes and Injuries by Severity. Austin: Texas Department of Transportation; 2021.

37. TxDOT. University Handbook: The Research and Technology Implementation Division. Austin: Texas Department of Transportation; 2022.