Contract BE939
DOT-RFP-20-9081-KW

A Review of Florida's FC-5 Raveling Condition Assessment and Measurement Methods

Final Report

Project performed by Pavement Analytics LLC in cooperation with the Georgia Institute of Technology

Florida DOT Project Manager: James Greene, P.E.
Co-Principal Investigators:
Yichang (James) Tsai, Ph.D., P.E., Georgia Tech
Bruce Dietrich, P.E., Pavement Analytics LLC
Wiley Cunagin, Ph.D., P.E., Pavement Analytics LLC

Researchers:
Yung-an Hsieh, Ph.D. Student, Georgia Tech
Lucas Yu, Ph.D. student, Georgia Tech

February 28, 2022

Disclaimer

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the Florida Department of Transportation.

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
A Review of Florida's FC-5 Raveling Condition Assessment and Measurement Methods	February 28, 2022
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Yichang (James) Tsai, Yung-An Hsieh, Bruce Dietrich, Wiley Cunagin, Lucas Yu	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Pavement Analytics LLC Georgia Institute of Technology	11. Contract or Grant No.
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Florida Department of Transportation, 605 Suwannee Street, MS 30 Tallahassee, FL 32399	Final Report March 18, 2020 to February 28, 2022
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

While open-graded friction course (OGFC) mixtures, called FC-5 in the Florida Department of Transportation (FDOT), provide several safety benefits, their lives are limited to approximately 14 years on average in Florida. The reduced life is generally due to raveling, which rapidly deteriorates once initiated. In FDOT, raveling is rolled up into the crack rating (CR). Therefore, it is difficult to distinguish what amount of the CR is contributed by raveling. This research aims at determining if raveling can be assessed in a more detailed manner, which considers survey approaches and methodologies to rate and separate raveling that can proactively target the raveling-only treatment needs at an early stage. A literature review was conducted on the practices of pavement rating computation and treatment decisions in state DOTs and automated raveling detection and severity classification methods. Then a feasibility study was conducted on using machine learning (ML) models for automated raveling severity classification using FDOT pavement data. Two main recommendations were made based on the technology review and ML feasibility study. First, it is recommended that FDOT leverage its existing 3D line laser imaging system to collect high-resolution pavement surface data, which can be used to objectively identify and locate raveling with a much better spatial resolution. Second, the feasibility study showed that the ML models are feasible for per-image raveling classification on FDOT's 3D pavement data. It is recommended that FDOT use the random forest (RF) model for future implementation given its high accuracy (86.6%) achieved in the feasibility study. New raveling treatment criteria and condition ratings are also proposed based on the reviewed literature and two cases of FC-5 only projects. It is recommended that more case studies be conducted in the future to determine suitable thresholds for applying FC-5-only resurfacing.

17. Key Word	18. Distribution Statement
OGFC, Raveling, FC-5 only, 3D pavement data, Machine learning	

19. Security Classification. (of this report)	20. Security Classification. (of this page)	21. No. of Pages	22. Price
		67	

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

Acknowledgments

The researchers would like to thank the Florida Department of Transportation (FDOT) for the financial support and assistance to carry out this research.

Executive Summary

As a method to increase friction and minimize hydroplaning potential, the Florida Department of Transportation (FDOT) places open-graded friction courses (OGFC) on multilane roadways with a design speed of 50 mph or greater (OGFC is called FC-5 in FDOT). While OGFC mixtures provide several safety benefits, their lives have been limited to approximately 14 years on average compared to nearly 20 years for dense-graded pavements in Florida. The reduced life is generally due to raveling, which can rapidly degrade pavement once initiated. In FDOT, a pavement crack rating (CR) is a score from 0 to 10 that reflects both the severity and the percent of pavement area affected by cracks, raveling, and patching. The subjectivity of estimating the extent and severity of raveling and the potential rapid rate of raveling can lead to difficulties in forecasting the life of pavements with an OGFC and quick reaction to localized OGFC treatment needs. Also, because raveling and patching are rolled up into the CR, it is difficult to distinguish how much of the CR is determined by cracking and how much by raveling. Therefore, in this research, the Pavement Analytics and Georgia Tech team investigated whether raveling can be assessed in a more detailed manner and appropriately accounted for in FDOT's pavement condition survey rating. The research considered survey approaches as well as methodologies to rate and separate raveling that can proactively target the raveling-only treatment needs at an early stage before the rapid deterioration of OGFC begins.

In this research, a literature review was conducted on practices of pavement rating computation and treatment decisions in FDOT and other state DOTs with a special focus on raveling. Four potential needs were identified for improvement in FDOT's raveling-related practices. First, a rating system separating cracking and raveling is needed, which can proactively target the raveling treatment needs before rapid deterioration of OGFC sets in. With the separated distress ratings, new treatment decision criteria separating cracking and raveling are also needed. In this case, more cost-effective treatments (e.g., FC-5 only treatment) can be applied. Second, the in-field visual inspection methods for raveling can be error-prone, time-consuming, and labor-intensive. There is a need to develop an automated raveling detection and classification method. Third, the appearance of raveling on digital images is susceptible to ambient lighting conditions. To overcome this issue, using 3D pavement data is a better alternative for capturing pavement surface texture. Finally, a fixed raveling survey spatial unit is not defined in FDOT's current practice, which causes difficulty in obtaining precise localization information for raveling. Currently, FDOT's 3D pavement image size is 12 ft (4 m) by 15 ft (5 m) in transverse and longitudinal directions, respectively. With the detailed 3D pavement data, a fixed and multiscale survey unit (e.g., 15 ft, 100 ft, 0.1 mile or 1 mile, etc.) can be generated with proper data aggregation methods.

Next, automated raveling detection and severity classification methods were investigated and recommended for FDOT. A literature review on existing automated methods was first conducted. Traditionally, research efforts on automated raveling detection and classification focused on nontrainable methods. These methods utilize nontrainable algorithms to calculate different indicators of raveling from the pavement data. However, these methods have several limitations, such as requiring certain assumptions about the pavement surface and lacking systematic validation and requiring frequent parameter tuning based on empirical experiments. In recent years, machine learning (ML) models for automated raveling detection and classification have been developed. By utilizing ML techniques to train a more robust model using real-world 3D pavement data that performs both raveling detection and classification, Tsai et al. (2021) overcame most of

the problems in nontrainable methods. Given its good performance achieved in a systematic validation, this ML method is recommended for further investigation.

Based on the literature review, a feasibility study on using ML models for automated raveling severity classification on FDOT's pavement data was conducted. The ML method in Tsai et al. (2021) was applied in the feasibility study, which includes the macrotexture features extracted from 3D pavement images and three traditional ML models: the Support Vector Classifier (SVC), random forest (RF), and Adaptive Boosting (AdaBoost). The 3D range images that FDOT had already collected using the Laser Crack Measurement System (LCMS) were used in this study. The raveling severity of each range image was annotated by FDOT engineers as the ground truth to train and evaluate the ML models. Through the feasibility study, two important conclusions are developed. First, the quality of annotation on raveling severity is critical for improving the performance of ML models. We found that all ML models achieved a better performance by training with the data that have a better annotation quality of raveling ratings. Second, with the provided quantized range images and around 8.5 miles of data, the outcomes of the study show that the ML models are feasible for per-image raveling classification on FDOT's pavement data with suitable ML classifiers. RF classification has the best potential to be used for implementation in the future given its high accuracy achieved in the feasibility study. The RF classifier achieved a testing accuracy of 86.6% in this feasibility study.

Based on the technology review and ML feasibility study, the preliminary recommendations for FDOT were developed. First, for the data collection devices, it is recommended to use FDOT's existing 3D line laser imaging system to collect high-resolution pavement surface data. A more reliable automated raveling detection and classification method can be developed using the 3D pavement data. Therefore, resources can be optimized by utilizing the 3D pavement data that FDOT already collected using the LCMS. Second, the feasibility study of ML models shows promising outcomes in automated raveling classification on 3D pavement data. It is recommended that FDOT use the RF model for future implementation given its high accuracy achieved in the feasibility study. With automated raveling severity classification using 3D pavement images, this research outcome will enable FDOT to cost-effectively identify and locate raveling with a much better spatial resolution (e.g., 15 ft, 100 ft, 0.1 mile). Recommendations for future implementation are listed at the end of this report.

Finally, raveling treatment criteria and condition rating are proposed for FDOT. The proposed treatment decision table consists of two levels of criteria: the overall pavement condition deduct and the individual distress deduct (such as raveling, cracking, rutting, etc.). The initial treatment criteria in the treatment decision table are designed based on the current rating practices in FDOT and the case study on two FDOT FC-5-only projects. Then, the proposed initial raveling deduct table is provided. The initial deduct values in the table are proposed based on the designed treatment criteria, the affected area and severity level of raveling, and the two FC-5 only projects. Although the initial treatment criteria and deduct values of raveling have been established based on two cases of FC-5-only project and engineering judgment, more cases (both successful and unsuccessful ones) should be studied for determining the right timing for applying FC-5-only treatment.

Table of Contents

List of Figures

List of Tables

1. Introduction

As a method to increase friction and minimize hydroplaning potential, the Florida Department of Transportation (FDOT) places open-graded friction courses (OGFC) on multilane roadways with a design speed of 50 mph or greater (OGFC is called FC-5 in FDOT). OGFC provides a skid-resistant surface that also reduces splash and spray by allowing water to drain horizontally through the open structure rather than on the surface. Due to these safety features, roadway designers also have the option of placing OGFC on high-speed curbed roadways with a history of wet weather crashes. While OGFC mixtures provide several safety benefits, their lives have been limited to approximately 14 years on average compared to nearly 20 years for dense-graded pavements in Florida. The reduced life is generally due to raveling, which can rapidly degrade pavement once initiated. As part of FDOT's pavement management program, the State Materials Office (SMO) rates the condition of asphalt pavements annually to compute the overall pavement condition rating (PCR). The PCR is derived from three ratings, which are crack, rut, and ride. A pavement crack cating is a score from 0 to 10 that reflects both the severity and the percent of pavement area affected by cracks, raveling, and patching. The crack rating is the only component of the pavement condition survey that is conducted by visual estimation. Due to the subjectivity associated with estimating the extent and severity of cracking and raveling, both are classified according to the following broad categories.

- Severity: Light, moderate, or severe
- Affected area: 1% to 5%, 6% to 25%, 26% to 50%, and greater than 50%

The large ranges of the affected area of raveling, subjective visual rating, and potential rapid rate of raveling can lead to difficulties in forecasting the life of pavements with an OGFC and quickly reacting to localized OGFC treatment needs. Also, because raveling and patching are rolled up into the crack rating, it is difficult to distinguish how much of the crack rating is determined by cracking and how much by raveling.

In this research study, the Pavement Analytics and Georgia Tech team conducted an investigation to determine if raveling can be assessed in a more detailed manner and appropriately accounted for in Florida's pavement condition survey rating as well as in subsequent pavement performance forecasting. The research considered survey approaches as well as methodologies to rate and separate raveling that can proactively target raveling-only treatment needs at an early stage before the rapid deterioration of OGFC begins. The research included the following tasks:

Task 1 – Literature Review: The Pavement Analytics and Georgia Tech (PA/GT) team conducted a comprehensive literature review to determine the state-of-the-art practice, challenges, and critical issues for assessing, rating, and forecasting performance of pavements with OGFC. The team also reviewed FDOT's pavement condition survey and forecasting methods to understand current practices and procedures.

Task 2 – Raveling Condition Assessment: The PA/GT team investigated and recommended automated methods to assess the extent and severity of raveling. The methods considered current pavement condition data collection devices such as an inertial profiler and 3D data collection systems as well as other potential devices and methodologies. Recommended methods were

investigated for repeatability and accuracy. Survey practices, analysis methods, and reporting procedures were fully documented.

Task 3 – Rating and Managing Pavements with Raveling: The PA/GT team identified the best practices determined from the literature review in consultation with the FDOT Project Team and the findings from Task 2 and evaluated those practices/methodologies to determine those that are most appropriate for use in rating the raveling condition of a roadway. The methodology weighed the benefits of reporting raveling as individual distress versus combining it with the other surface distresses. The impact on maintenance and rehabilitation strategies were also considered. In addition, raveling thresholds to trigger maintenance and rehabilitation were evaluated and discussed with the FDOT Project Manager.

Task 4 – Draft Final and Closeout Teleconference: The PA/GT team has submitted a draft final report to the FDOT Research Center. The draft final report includes the literature review, a description of critical issues and challenges, automated procedures for assessing the condition of raveling, methods to predict raveling, and thresholds to trigger rehabilitation based on raveling. The PA/GT team Co-Principal investigators will schedule a closeout teleconference. The Co-Principal investigators have prepared a comprehensive PowerPoint presentation.

This report is organized as follows. Chapter 1 presents the introduction and the tasks of this project. Chapter 2 presents the literature review, including the practices of pavement rating computation and treatment decisions in FDOT and other State DOTs with a special focus on raveling, and the existing methods for automated raveling detection and severity level classification. Chapter 3 presents the feasibility study of machine learning (ML) for automated raveling classification using FDOT's pavement data. Chapter 4 presents the design of raveling condition rating and treatment trigger for FDOT. Finally, Chapter 5 presents the conclusions of this project and recommendations for future works.

2. Literature Review

2.1 FDOT's Practices on Pavement Survey, Rating, and Treatment Decision

2.1.1 Pavement Section Selection (FDOT, 2017)

In FDOT's practice, the pavement survey unit (a section) is the same as project termini with varying distances/lengths. The project termini are typically the spatial unit used for pavement treatment (maintenance and rehabilitation). The beginning and end points of the project termini are typically determined based on the following factors:

1. County line
2. County section or subsection
3. Construction limits
4. Significant changes in pavement condition.
5. Structures in excess of 0.25 miles.
6. Rigid pavement in excess of 0.50 miles within a flexible pavement section.
7. Changes in the number of lanes (2 - 3 lanes, etc.)

The survey and the recorded pavement rating are based on the whole section or project termini. During the survey, the operator keeps track of cracks. Usually, the rating is not split up unless there is a dramatic change between the small sections. Pavement sections less than 0.50 miles are not rated separately but are combined with adjacent sections having the most similar condition.

2.1.2 Pavement Condition Rating (FDOT, 2017)

FDOT's overall pavement condition rating (PCR) is derived from three ratings - which are Crack, Rut, and Ride.

1. **Crack Rating (CR)** is a combination of **cracking, raveling, and patching**. Three cracking classes are considered in CR and the definition and measurement of the affected area are shown in Table *1*. Note that both patching and raveling are included in Class III. To calculate the CR, the total percent affected area and the predominant crack class has to be determined. The calculation is done by the following steps:

    a. Calculate the affected area of raveling by combining the affected area of all severity levels. (e.g., 5% light, 10% moderate, and 5% severe = 20% Raveling (Predominate severity level: moderate).)

    b. Calculate the affected area of Class III by combining the affected area of raveling, patching, and Class III cracking. (e.g., 15% Class III crack, 5% patching, 20% raveling = 40% Class III.)

    c. Calculate the total affected area by combining the affected area of all three crack classes. (e.g., 10% 1B, 12% II, 40% III = 62% total affected area.)

    d. Determine the predominant class by selecting the class with the highest percent affected area. (e.g., 10% 1B, 12% II, 40% III = Class III is the predominant class.)

e. Determine the deduct value based on the total affected area (e.g., 62%) and predominant class (Class III) from Table *2*. Note that the wheel path (CW) and non-wheel path (CO) are considered separately. The cracking on the wheel path (CW) has a higher deduct weight than the one on the non-wheel path as shown in Table *2*. (e.g., Assume Class III with 62% total affected area in CW, then from Table *2*, giving a deduction code of L (deduction value of 7.0).)

f. Calculate the CR based on the deduct value of CW and CO:

$$CR = 10 - (CW\ deduct + CO\ deduct)$$

2. **Rut Rating** is determined by the rut depth. Rut depths are collected using a laser profiler. The laser profiler measures rut depths at highway speeds and record the average rut depth of the two-wheel paths for each section evaluated. The rut depth is then assigned a deduct value as shown in Table *3*.

3. **Ride Rating** is based upon a scale of 0 (very rough) to 10 (very smooth). The International Roughness Index (IRI) is used to determine Ride Rating. The longitudinal laser profile of each wheel path is measured at highway speeds by a non-contact inertial laser profiler. Longitudinal laser profile data are collected at the smallest sample interval possible, usually, less than one inch, depending on data collection frequency and driving speed. The data are then processed using a longitudinal profile distance of 6 inches, a moving average of 12 inches, and 300-foot wavelength filtering. The longitudinal profile data are used to calculate the IRI.

After the Crack, Rut, and Ride Ratings are calculated, the PCR is determined by selecting the minimum value among these three ratings.

Table 1. FDOT's Crack classes.

	Class IB	**Class II**	**Class III**
Definition	Hairline cracks < 1/8" wide in either the longitudinal or transverse direction.	Cracks with 1/8" to ¼" wide. Also includes alligator cracking.	Cracks > ¼" wide. Includes Raveling and Patching. Includes severely spalled Class II cracks.
Measurement of the affected area	Cracks are estimated individually for the total linear **length**. The width of the affected area is considered 1 ft (0.30 m).	Cracks are considered **rectangular**, and the total affected area in square feet is counted.	Cracks are considered **rectangular**, and the total affected area in square feet is counted.

Table 2. Deduction codes for Crack Rating (FDOT, 2017).

NUMERICAL DEDUCTIONS FOR CRACKING METHOD

PERCENT OF PAVEMENT AREA AFFECTED BY CRACKING	CONFINED TO WHEEL PATHS (CW) *PREDOMINANT CRACKING CLASS*					
	1B CRACKING		II CRACKING		III CRACKING (Including RAV & PT)	
	CODE	DEDUCT	CODE	DEDUCT	CODE	DEDUCT
00 -- 05	A	0.0	E	0.5	I	1.0
06 -- 25	B	1.0	F	2.0	J	2.5
26 -- 50	C	2.0	G	3.0	K	4.5
51+	D	3.5	H	5.0	L	7.0

PERCENT OF PAVEMENT AREA AFFECTED BY CRACKING	OUTSIDE OF WHEEL PATHS (CO) *PREDOMINANT CRACKING CLASS*					
	1B CRACKING		II CRACKING		III CRACKING (Including RAV & PT)	
	CODE	DEDUCT	CODE	DEDUCT	CODE	DEDUCT
00 -- 05	A	0.0	E	0.0	I	0.0
06 -- 25	B	0.5	F	1.0	J	1.0
26 -- 50	C	1.0	G	1.5	K	2.0
51+	D	1.5	H	2.0	L	3.0

Table 3. Calculation of Rut Rating (FDOT, 2017).

PROFILER RUTTING VALUES

RUT DEPTH (IN)	RUT DEPTH (MM)	RANGE (IN)	RANGE (MM)	DEDUCT	RUT RATING
0	0	0.00 – 0.06	0.00 - 1.59	0	10
1/8	3.18	0.07 – 0.19	1.60 - 4.76	1	9
1/4	6.35	0.20 – 0.31	4.77 - 7.94	2	8
3/8	9.53	0.32 – 0.44	7.95 - 11.11	3	7
1/2	12.70	0.45 – 0.56	11.12 - 14.29	4	6
5/8	15.88	0.57 – 0.69	14.30 - 17.46	5	5
3/4	19.05	0.70 – 0.81	17.47 - 20.64	6	4
7/8	22.23	0.82 – 0.94	20.65 - 23.81	7	3
1	25.40	0.95 – 1.06	23.82 - 26.99	8	2
1 1/8	28.58	1.07 – 1.19	27.00 - 30.16	9	1
1 1/4 +	31.75	1.20 +	30.17 +	10	0

2.1.3 Raveling Survey and Rating (FDOT, 2017)

In FDOT practice, the condition rating for raveling is included in Crack Rating (CR). Raveling is recorded only if the survey section has at least one percent of its area raveled and is categorized into three severity levels:

1. **Light (L)**: The aggregate and/or binder has begun to wear away but has not progressed significantly, with some loss of aggregate.
2. **Moderate (M)**: The aggregate and/or binder has worn away, and the surface texture is becoming rough and pitted; loose particles generally exist; loss of aggregate has progressed.
3. **Severe (S)**: The aggregate and/or binder has worn away and the surface texture is very rough and pitted; loss of aggregate is very noticeable.

Figure *1* illustrates the raveling classification used by FDOT. During the survey, the **total percent affected area** and the **predominant severity level**, which has the highest affected area, are both recorded in FDOT's database using the codes listed in Table *4*. FDOT's raveling codes (FDOT, 2017). (e.g., L1, M2, S1). The pavement survey unit (a section) is the same as the project termini with varying distances/lengths. The project termini are typically the spatial unit used for pavement treatment (maintenance and rehabilitation). Among the recorded raveling information, the affected area is the only information used in the calculation of CR although the severity level of raveling and code (representing the percentage of raveling area) are both recorded in the database. It is observed that the percentage of raveling area and the corresponding severity level could potentially be used to evaluate raveling-only treatment and perform raveling forecasting.

(a) Light (b) Moderate (c) Severe

Figure 1. Raveling classification in FDOT.

With the visual inspection method used in the current FDOT raveling condition survey practice, it is extremely difficult to survey and record raveling with a detailed level of location reference resolution. Therefore, only the approximate percentage of raveling is recorded in a pavement project. For example, consider a 20-mile pavement project with 1-5% light raveling (i.e., code L1 in Table 4) being recorded by the manual survey. In this case, the project length for recording the distress is 20 miles; thus, it is difficult to locate the exact raveling location within the project limits. In addition, the current raveling distress has been combined with cracking and patching to yield a pavement Crack Rating within which it is difficult to differentiate and identify the impact of each distress type. Consequently, it is very difficult for the current protocol to support FDOT's identification of cost-effective pavement sections for resurfacing that replaces the FC-5 only.

Table 4. FDOT's raveling codes (FDOT, 2017).

RAVELING CODES

PERCENT OF PAVEMENT AREA AFFECTED BY RAVELING	RAVELING SEVERITY LEVEL AND CODE		
	LIGHT	MODERATE	SEVERE
01 -- 05	1	1	1
06 -- 25	2	2	2
26 -- 50	3	3	3
51+	4	4	4
Note: Code the Predominant severity level only			

2.1.4 Treatment Determination

In FDOT's current practice, PCR < 6.5 is the main threshold to trigger the need for pavement treatment. In an urban area, a lower speed < 50 mph, and a different threshold (ride rating < 5.5) are used. FDOT's funding protocol focuses on performing resurfacing on the deficient pavement. Therefore, resurfacing is the major treatment method and there is no comprehensive pavement preservation treatment decision tree. FDOT's main treatment is to mill and overlay regardless of the percentage of raveling and cracking. If the sections with raveling-only are identified, the only difference will be the depth of milling and overlay to remove and replace the OGFC only rather than milling into the structural course to remove cracks as well. Note that although PCR is used as the threshold, the majority (90%) of treatment is triggered by CR < 6.5. The other two ratings (rutting and ride ratings) triggered less than 10% of treatment.

2.1.5 Forecasting

FDOT forecasts the total mileage of roads that are deficient (i.e., PCR rating less than 6.5) over a 5-year horizon for budget preparation. The forecasting can be categorized as follows:

- 5-Year Forecast: Most concerned with total deficient lane miles. The deficient lane miles are translated into the total budget for the state.
- 4-Year Forecast: Concerned with total deficient lane miles and individual road predictions. The budget is fixed but district-level allocation can be relatively adjusted among different districts.
- 3-Year Forecast: Most concerned with individual lane predictions. The budget is fixed but district-level allocation can be adjusted.
- 1-Year and 2-Year Forecasts: There is no budget flexibility in the 1 and 2-year time horizon.

Note that the 3 to 5-year forecasts are the highest priority of the pavement condition forecasting as the year 1 and 2 budgets cannot be changed. The practice of pavement distress forecasting in FDOT is also shown symbolically in Figure *2*.

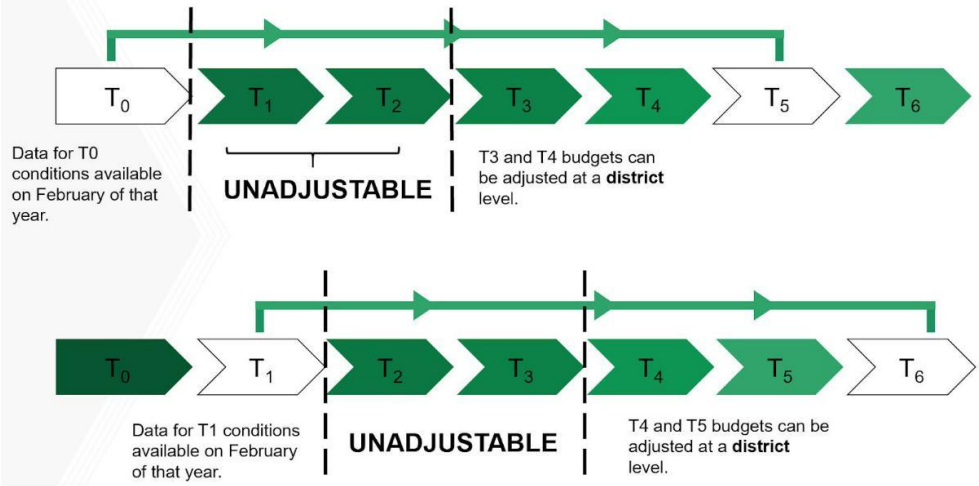


Figure 2. FDOT's pavement condition forecasting practice.

2.1.6 Current Issues and Needs for Improvement

Based on the review of FDOT's practices for its raveling survey and rating computation, the following are the potential issues and needs for improvement.

Need 1: A rating system separating cracking and raveling. Currently, raveling and patching are rolled up into the CR. Therefore, it is difficult to distinguish how much of the crack rating is impacted by cracking proper and how much by raveling. For example, when there is both light cracking and heavy raveling, the CR does not differentiate between cracking and raveling as they are combined based on their respective areas. In addition, raveling severity level is recorded but has not been used to compute CR. Cost-effective maintenance and rehabilitation, like raveling-only treatment, cannot be applied because of the above-mentioned issues. For example, if raveling of an open-graded friction course was the sole distress responsible for a deficient CR, milling and replacement of the open-graded friction course, which does not touch the structural course, may be sufficient. If extensive cracking is also present, deeper milling to remove cracks may be required. Thus, a rating system separating cracking and raveling deducts and rating needs to be developed, which can proactively target the raveling treatment needs before rapid deterioration of OGFC sets in. In this way, the rating computation results can be assessed and deducted as individual distress and can also be combined as a composite rating.

Need 2: Automatic detection and classification of raveling instead of a visual survey. The current visual assessment of raveling is not precise because the recorded raveling affected area has large ranges (1 to 5%, 6 to 25%, 26 to 50%, and greater than 50%; see also Table *4*). Therefore, there is a need to develop an automated raveling detection and classification process. The location and timing of raveling treatment are important. Cost-effective treatment to prevent rapid raveling deterioration of OGFC requires detailed measurement and recording to determine the location(s) of the raveling. The use of smaller intervals (e.g., 15-ft image size; 100-ft, 0.1 miles or 1 mile, etc.) to capture more precise trigger points for raveling treatment is desirable. However, since the raveling survey is conducted by visual estimation in FDOT's current practice, the recorded percent of raveling affected area is based on the entire project section and the precise raveling location cannot be attained. Therefore, a digital survey and automatic raveling detection and classification

process are urgently needed to obtain and record more detailed and accurate information. Additional studies of automatic detection and classification of raveling were conducted in Task 2.

Need 3: A consistent and fixed raveling survey unit. It is a challenge to ensure the visual survey data quality with the current FDOT survey practice since it does not have a fixed survey spatial unit. For example, visually surveying and accurately determining the percentage of raveling area for a long section (e.g., a 10 or 20-mile section) is difficult. Some state DOTs (e.g., Georgia DOT, GDOT) has a consistent condition survey unit of one mile to ensure its survey consistency and accuracy. As mentioned in Need 2, the detailed recording to determine the location of deficient raveling and reveling rating at smaller intervals is important for determining and triggering raveling treatment. In FDOT's current practice, the survey and the recorded pavement rating are based on the whole section, which has a variable length. Therefore, the length of a pavement section is usually too large to obtain precise localization information of raveling. Currently, the FDOT's 3D pavement image size is 12ft (4m) by 15ft (5m) in transverse and longitudinal directions respectively. With the detailed 3D pavement data, a fixed and multi-scale survey unit (e.g., 15ft/5m, 100ft, 0.1 miles or 1 mile, etc.) can be generated with proper data aggregation methods. For example, the raveling condition can be recorded in image-size intervals and summarizing them in 0.1-mile segments then to the section level. This technology will allow the acquisition of more detailed information to localize, measure, and rate pavement distresses using a smaller survey unit so FDOT can more quickly react to early raveling.

Need 4: A treatment criteria and threshold for triggering raveling-only treatment. Currently, CR < 6.5 is the primary threshold to trigger treatment in FDOT's practice. However, as mentioned in Need 1, including raveling and patching in CR makes the maintenance and rehabilitation decision difficult to make as well as assess the more cost-effective treatment. Therefore, it is recommended that a revised rating system be developed that considers cracking and raveling individually as well as new treatment decision thresholds separating cracking and raveling where:

1. Raveling is the only or predominant distress.
2. Raveling and cracking are both presented.

An example of separated treatment thresholds for cracking and raveling extent or severity levels is shown in Figure 2. With the detailed pavement survey data available from Need 2, information such as affected area, severity level, and trend of changes of raveling can be potentially used to determine the treatment decision threshold.

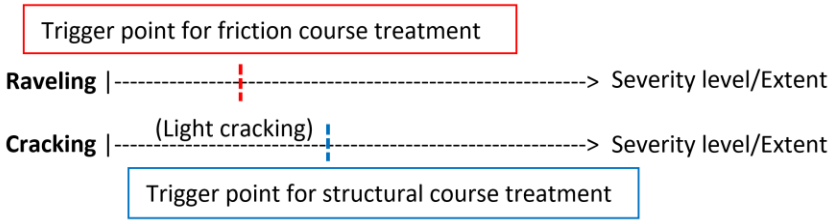


Figure 3. An example of a separated treatment threshold for cracking and raveling.

2.2 Raveling Classification and Measurement by Other State DOTs

2.2.1 Georgia Department of Transportation (GDOT) (GDOT, 2017)

In GDOT practice, raveling is classified into Severity Levels 1, 2, and 3 based on the following definitions of raveling conditions:

1. **Level 1**: Loss of a substantial number of stones.
2. **Level 2**: Loss of most of the surface.
3. **Level 3**: Loss of a substantial portion of the surface layer (>1/2 depth).

Figure *4* illustrates the raveling classification in GDOT. During field surveys, raveling is closely observed, and an estimate (to the nearest 5%) is made of the extent and the predominant severity of the distress within each relatively short, rated segment. The percent of the length of the rated segment (a mile or partial mile) that contains raveling is recorded along with the predominant severity level.

(a) Severity Level 1 (b) Severity Level 2 (c) Severity Level 3

Figure 4. Raveling classification in GDOT (GDOT, 2017).

By considering the condition rating of raveling along with the condition ratings of other pavement distresses (e.g., cracking, rutting, etc.), the treatment is then determined by the developed treatment decision tree (GDOT, 2017). In GDOT, micro-milling and thin overlay have been successfully developed and applied to cost-effectively treat raveling-only OGFC pavements in Georgia. The outcomes have been published in various journal papers (Gadsby and Tsai, 2020; Tsai et al., 2018; Tsai et al., 2016; Tsai et al., 2014b; Tsai et al., 2012; Lai et al., 2012).

In GDOT's pavement condition survey and pavement management system, a pavement project (say a 20-mile pavement project) is defined as a pavement section with consistent roadway characteristics that the same treatment can be applied. However, the pavement survey is in a finer unit which is one mile or less than one mile in the entire roadway network. This provides one mile of spatial location resolution to differentiate different pavement conditions. In addition, the current raveling distress is recorded as independent distress so that it can be more easily identified and queried from a database.

2.2.2 Oregon Department of Transportation (ODOT) (ODOT, 2019)

In ODOT practice, the severity levels of raveling are defined as follows:

1. **Low**: Aggregate has worn away resulting in noticeably rough or pitted pavement surface texture in the left wheel path, right wheel path, or center lane zone. Loss of chip seal rock should be rated as raveling, but this is the maximum severity for chip sealed surfaces.
2. **Moderate**: The surface texture is moderately rough and/or pitted with moderate loss of pavement surface aggregate in the left wheel path, right wheel path, or center lane zone. Loose aggregate particles may be present outside the traffic area.
3. **High**: The surface texture is very rough and/or pitted with severe loss of pavement surface aggregate in the left wheel path, right wheel path, or center lane zone. Flat bottom potholes may be present where there is a complete loss of aggregate.

Figure *5* illustrates the raveling classification in ODOT. During field surveys, raveling is identified by a roughened or pitted texture on the pavement surface. Mechanical abrasion from tire chains, studs, snowplows, or dragging equipment that results in significant loss of aggregate should be rated as raveling. Studded tire rutting which does not roughen up the texture significantly should not be rated as raveling. During the measurement, the linear feet of each severity level for the left wheel path, right wheel path, and center lane zones are recorded. The maximum quantity is 528 feet for each zone and 1,584 feet per 0.10-mile.

(a) Low (b) Moderate (c) High

Figure 5. Raveling classification in ODOT (ODOT, 2019).

2.2.3 Texas Department of Transportation (TxDOT) (TxDOT, 2015)

The classification of raveling severity levels is based on the percentage of raveled area in TxDOT practice, which is defined as follows:

1. Low: The percent of raveled pavement area is from 1% to 10%.
2. Medium: The percent of raveled pavement area is from 11% to 50%.
3. High: The percent of raveled pavement area is greater than 50%.

During the measurement, raveling that occurs anywhere in the lane is rated following the rating code shown in Table *5*. The rating code indicates the percent of the rated lane's total surface area. The rated lane is decided as the lane that has the most distress on each roadbed.

Table 5. TxDOT's raveling codes (TxDOT, 2015).

RATING CODE		AMOUNT (PERCENT AREA)
0	NONE	0
1	LOW	1 – 10
2	MEDIUM	11 – 50
3	HIGH	> 50

2.2.4 Alabama Department of Transportation (ALDOT, 2017)

In ALDOT practice, raveling is defined as the loss of bond between the asphalt binder and the aggregate through either a cohesion or adhesion failure, usually caused by the action of water. Raveling is not classified into different severity levels in ALDOT. Instead, the square feet of raveling are recorded per asphalt lane-mile. During the survey, the raters inspect the surface area for raveling, then measure and record the total square feet of raveling in all lanes.

2.2.5 Summary

Although raveling is defined in almost the same way by different DOTs, measurement and rating methods of its severity levels and extent change from agency to agency. This section describes the challenges and research needs for automated raveling detection and classification methods based on the reviewed practices of DOTs.

First, the survey and classification of raveling in major transportation agencies are conducted manually through in-field visual inspection methods. These methods are time-consuming and labor-intensive. Also, these methods can be error-prone since raveling can appear differently when it is observed from a moving vehicle or standing on the ground. The manual recorded raveling conditions can also be subjective since some raters may tend to rate heavier than others. Moreover, a manual survey on high traffic volume highways can be challenging and dangerous to the raters because of the heavy traffic volume. Therefore, there is a need to develop an automated raveling detection and classification method to overcome these problems.

Second, one of the main challenges of the raveling survey lies in the fact that raveling is the change of pavement surface texture, which causes its appearance on digital images, or 2D images, to be susceptible to ambient lighting conditions, especially for low-severity raveling. For example, under direct sunshine, it is hard to recognize lightly raveled surfaces. Since raveling develops quickly after it starts, DOTs must identify it in its early stage so that cost-effective treatments (e.g., OGFC replacement only) can be applied before it deteriorates to higher severity levels and requires more expensive corrective treatments. Therefore, the difficulty of low-severity raveling detection under natural lighting conditions can cause DOTs to be unable to perform cost-efficient treatments in a timely fashion. To overcome this issue, using 3D pavement data is a better alternative for capturing pavement surface texture because it is independent of ambient lighting conditions and

can be accurately collected at highway speed. Thus, a more reliable automated raveling detection and classification method can be developed using the 3D pavement data.

Finally, a fixed raveling survey spatial unit is not defined in some DOTs (e.g., TxDOT, FDOT). In this case, the pavement is rated based on a whole section, which has a variable length and is usually too large to obtain precise localization information for raveling. Currently, the FDOT's 3D pavement image size is 12ft (4m) by 15ft (5m) in transverse and longitudinal directions respectively. With the detailed 3D pavement data, a fixed and multi-scale survey unit (e.g., 15ft/5m, 100ft, 0.1 miles or 1 mile, etc.) can be generated with proper data aggregation methods. This technology will allow the acquisition of more detailed information to localize, measure, and rate pavement distresses using a smaller survey unit so DOTs can more quickly react to early raveling and apply localized treatment to optimize pavement asset management.

2.3 Automated Methods for Raveling Detection and Classification

Automated raveling detection and classification methods can be classified into two categories: nontrainable methods and machine learning (ML) models. Traditionally, research efforts on automated raveling detection and classification focused on nontrainable methods. These methods utilize nontrainable algorithms to calculate different indicators of raveling from the pavement data. In recent years, ML has become a popular technique in almost every field. By providing a sufficient amount of data, ML models can automatically learn the hidden structures or relationships in the data. This "data-driven" property makes ML models more robust and general. Following this trend, ML models for automated raveling detection and classification have also been developed.

2.3.1 Nontrainable Methods

Ooijen et al. (2004) developed the "Stoneway" algorithm to detect raveling on porous asphalt pavement. In this algorithm, raveling was detected by analyzing each longitudinal laser profile for gaps that were both above a length and depth threshold (Figure *6*), indicating a possible loss of aggregate. The severity of raveling is classified by the percentage of aggregate missing on the surface. Several challenges are observed for the Stoneway algorithm. First, it was found that this approach generally underestimated raveling severity, scheduling maintenance operations later than recommended by visual condition surveys. Second, the road surface is assumed to be flat in both longitudinal and transverse directions, which limits the algorithm's ability to generalize to diverse conditions of the pavements. Finally, this method analyzed longitudinal profiles 500 mm apart, which may be too sparse to obtain the representative condition of the road.

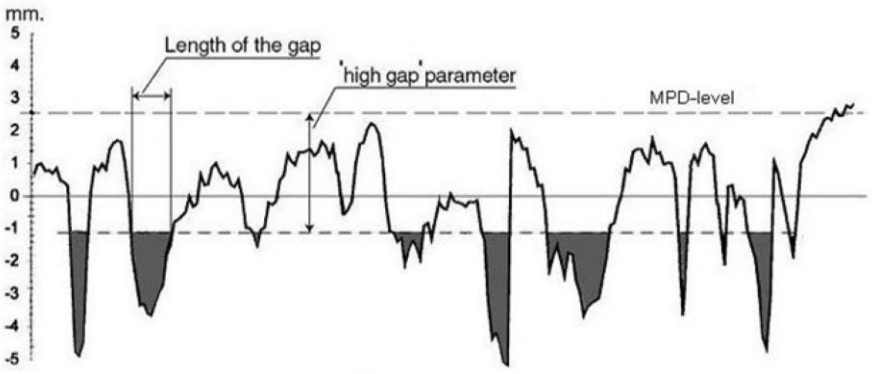


Figure 6. The Stoneway algorithm (Ooijen et al., 2004)

Another method based on a similar concept to the Stoneway algorithm was developed by Scott et al. (2008). This method calculates the Mean Profile Depth (MPD) of the laser profile. Locations that differ from the characteristic level by a sufficient depth and over a significant length are deemed to be raveled. Based on this measurement, the proportion of the road affected by raveling is reported. Similar to the Stoneway method, the MPD method only worked when the assumption that the pavement surface was flat (no incline) was held.

McRobbie and Furness (2008) and McRobbie et al. (2012) used a feature called Root Mean Square Texture (RMST) derived from 3D pavement data. Based on the RMST, a raveling detection algorithm that provides an estimate of the raveling condition is developed by comparing the distribution of RMST values in a small, or local, area against those from a much larger surrounding, or global, area. A group of sites totaling approximately 90 km was selected, representing a combination of different surface types (thin surface course, porous asphalt, hot rolled asphalt, etc.) and surface conditions. However, the pair of scales needed to be calibrated for each site makes this method not practical for implementation. Further research is needed to obtain the pair of scales that work for all pavement types. Another challenge for the RMST method is the determination of local and global areas, which should be different for different surfaces and road conditions.

Different from the MPD and RMST methods that estimate absolute measurements of raveling, McRobbie et al. (2015) proposed a relative manner for estimating raveling conditions by identifying the changes in surface conditions among successive collected data. In this method, the 3D profile data are aligned first in both longitudinal and transversal directions. Then, from 30 candidate parameters, seven are selected to detect changes in 3D pavement data. The validation of this method only utilized the ground truth with raveling conditions collected in the lab. Also, a large number of standard surface shape parameters need to be tested by empirical experiments to determine their usability for quantifying surface disintegration.

Laurent et al. (2012a; 2012b) proposed the Raveling Index (RI) to quantify raveling by measuring the volume of aggregate loss (holes due to missing aggregates) per unit of surface area (square meter). The volume of aggregate loss is measured automatically from the 3D pavement data using algorithms designed by the authors. However, no additional information is provided about these algorithms. The calculation of RI is defined as follows:

$$RI = V_{aggregate\_loss}/A_{total}$$

In the RI method, 3D line laser technology was used to collect high-resolution surface range data. Also, the density of 3D laser data used in this method was high enough to cover the entire lane transversely. However, the tests for raveling detection are very limited without systematic validation using a large-scale dataset. The general method used here is to perform raveling detection on the same road section repeatedly. The robustness of the raveling detection method is tested by observing if the results tend to be similar (Figure 7). Another validation conducted for the RI method is visually comparing RI on surfaces with different raveling severities, which is also not systematic.

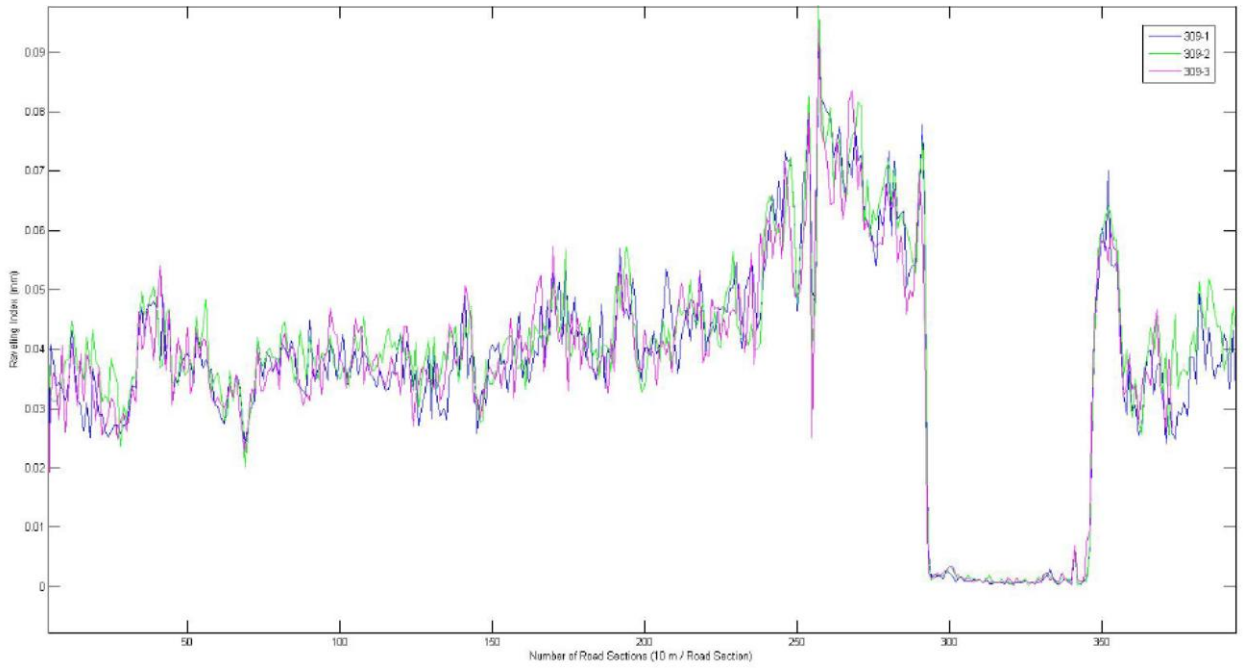


Figure 7. The repeatability test results of the RI method. Different colors indicate results from different passes (Laurent et al., 2012a).

Mathavan et al. (2014) presented a method to detect raveling from 3D pavement images (intensity and range). First, a texture descriptor method called Laws' texture energy measure is used in conjunction with the Gabor filter and other morphological operations to distinguish road areas from each other. Then raveled road areas are detected by estimating the standard deviation (STD) of the corresponding range data. By heuristically setting the thresholds for STD values, the raveling condition (within a limited grid) can be characterized as good, average, or bad. However, there is a lack of comprehensive validation in this paper. Detailed information on the validation dataset, such as the location of data collection and the distribution of raveling conditions in these data, are not mentioned in this paper. Moreover, the outcome of raveling quantification is not compared with the ground truth (e.g., visual survey results) specified based on transportation agencies' distress protocol (e.g. severity levels 1, 2, 3, or low, medium, and high).

Several other studies developed automated methods to detect pavement defects, including raveling. Hadjidemetriou & Christodoulou (2019) identify video frames that include any type of pavement defect using the entropy of images. Zhou et al. (2006) proposed a pavement defect detection criterion based on the statistical models in the wavelet domain. However, these methods

only determine if an image contains any pavement defect or not. That is, raveling is mixed with all other pavement defects (e.g., cracks, potholes, rutting, etc.) in the detection results, which is not sufficient for the objective of this research (i.e., to independently classify raveling based on its severity level).

2.3.2 Machine Learning (ML) Models

Hoang (2019) proposed an ML model based on the Stochastic Gradient Descent Logistic Regression (SGD-LR) model to automatically detect raveling on real-world digital images. In this model, image texture-based features are first extracted from the statistical properties of color channels and the Gray-level Co-occurrence Matrix. A set of extracted features are then used as the input variables to train the SGD-LR model to classify image samples into two categories, which are non-raveling and raveling. Two challenges are observed in this study. First, the model only classifies images into non-raveling or raveling, and the classification of raveling severity levels is not available. Second, the digital images used in this model are collected with a controlled lighting condition and the model is not validated with more diverse conditions of the pavement images.

A research team led by Professor Tsai at Georgia Tech had proposed a raveling detection and classification method using pavement surface data collected by 3D line laser technology and three ML models, the Support Vector Classifier (SVC), Random Forest (RF), and Adaptive Boosting (AdaBoost) (Tsai & Wang, 2015). This study was published in the International Journal of Pavement Research and Technology in 2021 (Tsai et al., 2021). We use the 2015 report as the reference in the remainder of this report since it has more details. This method consists of several steps (Figure *8*). First, the collected 3D pavement data are pre-processed to remove invalid data points and eliminate the cross slope of the pavement. Second, each full-lane-width, 5-m pavement section on a 3D range image is evenly divided into six subsections with three in each wheel path (Figure *9*), and the macro-texture features are calculated and extracted for each subsection. With the extracted macro-texture features, an RF model is then trained to classify each subsection into different severity levels of raveling based on GDOT practice. Finally, a raveling aggregation algorithm is developed to aggregate the classification results of the subsections to achieve a segment-level (~1 mile) raveling rating based on GDOT's standard. The method was validated using large-scale, real-world 3D pavement data collected on I-85 and I-285 near Atlanta Georgia.

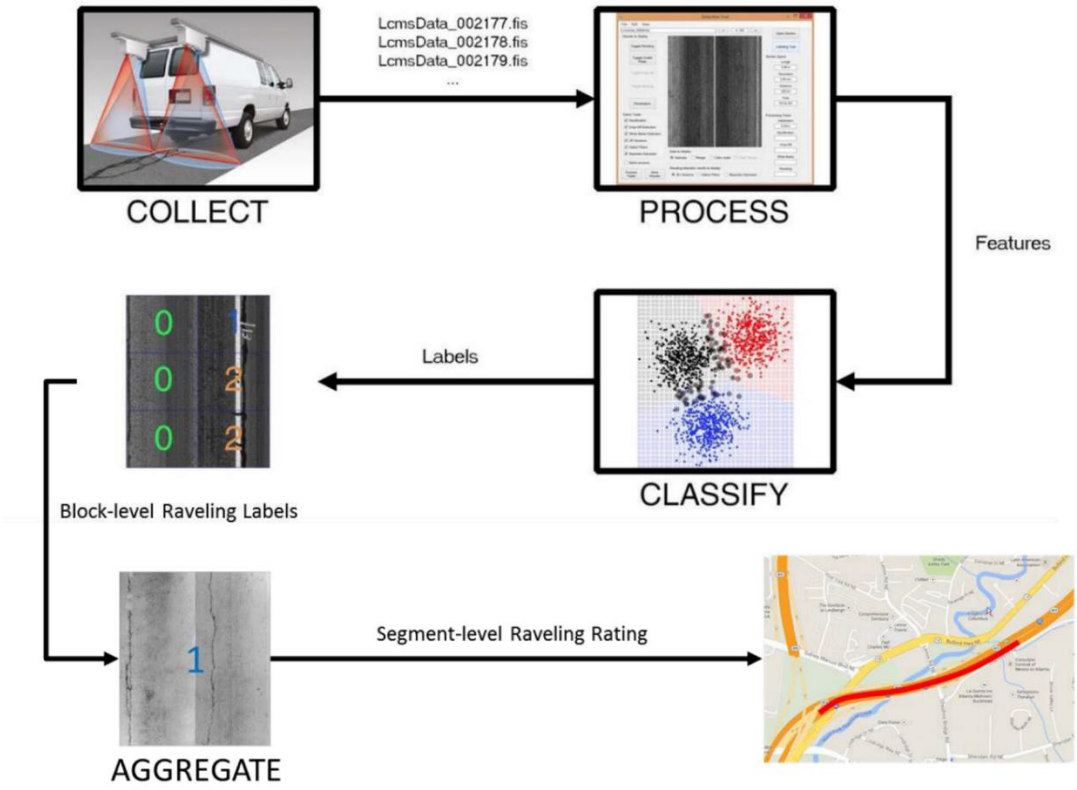


Figure 8. The raveling detection and classification method proposed by Tsai & Wang (2015).

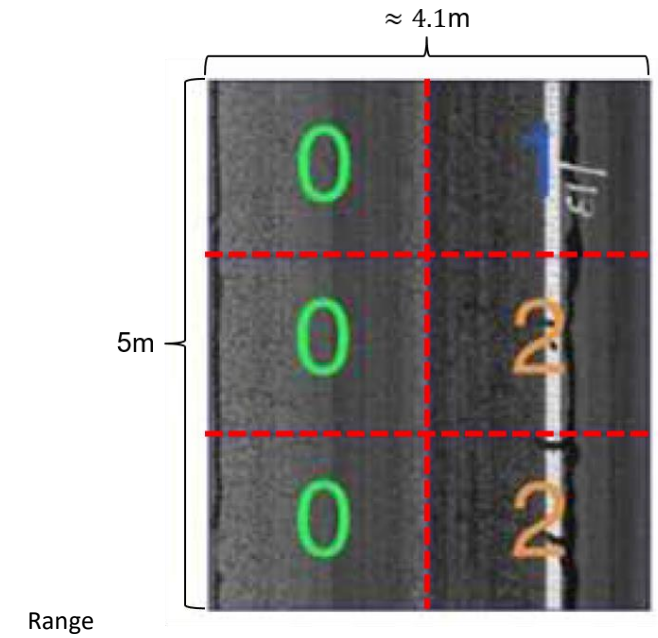


Figure 9. The six subsections of each 3D range image defined in Tsai and Wang (2015).

2.3.3 3D laser technology for automatic raveling detection and classification

To be able to record detailed information for raveling and further develop a new rating system and treatment thresholds, a digital survey and assessment method with high accuracy is an essential component. FDOT has been collecting interstate highway 3D pavement data using 3D laser technology, the Laser Crack Measurement System (LCMS), for three years.

In the U.S., 3D laser measurement has become a mainstream technology to collect high-resolution 3D pavement surface images. A survey reported in 2017 shows eighteen states in the U.S. have indicated their use of an automated 3D data collection system, and seventeen states have said they have a plan to use it within two years (Zimmerman, 2017). 3D pavement data have been studied and applied on asphalt and concrete pavements to automatically or semi-automatically detect and measure cracking (Tsai and Li, 2012; Jiang and Tsai, 2016), rutting (Tsai et al., 2013, 2015), potholes (Tsai and Chatterjee, 2018), and concrete joint faulting (Tsai et al., 2011; Geary et al., 2018). Advanced models based on statistics, computer vision, traditional machine learning, and deep learning have been successfully developed to utilize 3D pavement surface data to conduct automatic asphalt pavement crack detection (Zhang et al., 2017, 2019), crack classification (Tsai et al., 2014a; Wang et al., 2017; Li et al., 2020), crack deterioration analysis (Jiang et al., 2016), and raveling classification (Tsai and Wang, 2015), to cost-effectively and objectively manage pavements. The methods developed in these studies offered potential for use in this project to develop an automatic raveling detection and classification method for digital pavement condition assessment. Also, the resources could be optimized in this project by utilizing the LCMS data that FDOT already collected.

2.3.4 Summary

In summary, nontrainable methods for automated raveling detection and classification have several problems that constrain them from a broad application for different surfaces, different raveling conditions, or even different data sources. First, many methods are only capable of detecting raveling. The classification of the raveling severity is not available (Laurent et al., 2012a; Laurent et al., 2012b; Ooijen et al., 2004; McRobbie et al., 2015). Second, many of the indicators require certain assumptions about the pavement surface that might not apply to other cases (Ooijen et al., 2004; McRobbie & Furness, 2008; McRobbie et al., 2012; McRobbie et al., 2015). For example, RMST relies on the concept that raveled areas have a different texture pattern than non-raveled ones. When applying RMST on a long stretch of consistently raveled pavement, this indicator might fail to identify the raveled areas. Third, the validation of the methods was very limited and not systematic (Laurent et al., 2012a; Laurent et al., 2012b; Mathavan et al., 2014). Without a sufficient amount of pavement data with diverse conditions, it is not adequate to objectively reveal the true performance and limitations of the methods. Finally, many methods require frequent parameter tuning and adjustment based on empirical experiments (McRobbie & Furness, 2008; McRobbie et al., 2012; McRobbie et al., 2015). Thus, it is difficult for transportation agencies to implement these automated methods on real-world pavement surface data to perform both raveling detection and classification.

By utilizing ML techniques to train a more robust and general model using real-world 3D pavement data that performs both raveling detection and classification, Tsai & Wang (2015) overcame most of the above-mentioned problems. However, a major limitation of this method lies

in the fact that it is based on a traditional ML model. The main problem of traditional ML models is that a predefined feature extraction stage is required to reduce the complexity of the data and make patterns more visible to the trainable models. For image-based tasks, these handcrafted features limit the models to only learn from low-level features, such as edges or statistical properties. Without learning high-level features from the data, the performance of traditional ML models is limited even if more data are provided (Alom et al., 2019). Also, empirical experiments are still required to determine the suitable predefined features that are to be extracted from the data and train the models.

Different from traditional ML models, the features of the data are automatically learned through the training process of deep learning (DL) models. That is, DL models utilize the data-driven features learned from the training data instead of the handcrafted features, and a more accurate and robust model can be trained by providing more data (Alom et al., 2019). Convolutional neural networks (CNNs), a type of DL model widely used in computer vision, have already been studied on image-based, automated assessments of various pavement distress (Zhang et al., 2017; Maeda et al., 2018; Hsieh & Tsai, 2020a). However, the study of using CNNs and data-driven features to develop a more accurate and reliable automated raveling detection and classification model has been limited (Hsieh & Tsai, 2020b).

2.3.5 Recommendation

This research recommended the development of an automated raveling detection and classification method based on the ML models proposed by Tsai & Wang (2015) for two main reasons. First, by training the models to learn to make the decision using data with diverse conditions, ML models have been proven to achieve more accurate and robust performance not only for raveling detection and classification (Tsai & Wang, 2015; Hoang, 2019) but also for the detection and classification of many other pavement defects, such as cracks (Wang et al., 2017) and potholes (Hoang, 2018). Second, the ML models applied by Tsai & Wang (2015) have been systematically validated with large-scale, real-world pavement data using both raveling detection and severity levels classification following GDOT's standard. Since the raveling severity levels defined by FDOT and GDOT are interchangeable, the experiments and validation of the ML models would largely benefit the development of the automated raveling detection and classification method for FDOT. However, as mentioned in Section 3.4, the utilization of the traditional ML model has its disadvantages and limitations. Since using DL and data-driven features to develop a more accurate and reliable raveling detection and classification model has not been explored as yet, it was also recommended to explore how DL can benefit the development of the automated method. Also, other new quantitative measurement indicators, like the percent of loss of aggregates, could be explored to leverage the 3D pavement data.

Second, it was recommended to leverage the 3D pavement data already collected by FDOT. As mentioned in Section 2.6 Summary, 3D pavement data have the advantage of better capturing pavement surface texture because they are independent of ambient lighting conditions and can be accurately collected at highway speed. That is, a more reliable automated raveling detection and classification method can be developed using the 3D pavement data. Therefore, the resources could be optimized in this project by utilizing the 3D pavement data that FDOT already had collected using the Laser Crack Measurement System (LCMS). With the data already collected, not only

could a more accurate and robust ML model be trained, but also a systematic validation could be conducted.

Finally, for implementation in the future, it was recommended to also consider noise removal and aggregation of survey results into a fixed spatial unit, say 0.1 miles or 1 mile. As mentioned in Section 2.6, having a fixed and smaller survey unit is important to obtain detailed information to localize, measure, and rate pavement distresses so that timely and cost-efficient treatments can be performed. Based on the study from Tsai and Wang (2015), the data pre-processing steps for noise removal, and the results aggregation process are both important for practical applications of automated raveling detection and classification.

2.4 Pavement Rating Computation and Treatment Decision by Other State DOTs

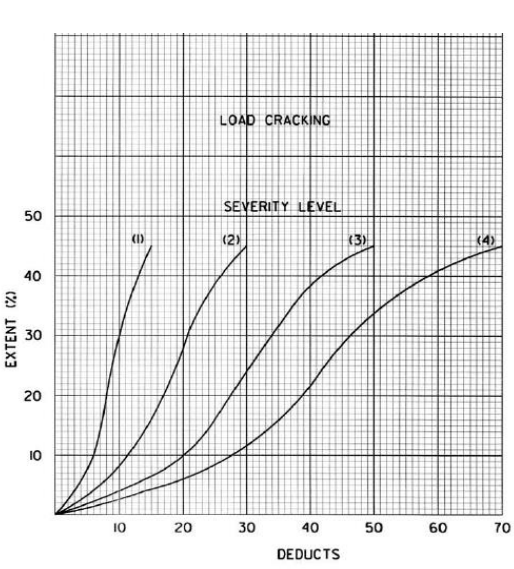
Here, the practices of three representative state DOTs were reviewed, including the Georgia Department of Transportation (GDOT), Texas Department of Transportation (Texas DOT), and Caltrans as they have similar environments but slightly different ratings and treatment designs. For each state DOT, its condition survey, rating, and treatment criteria was analyzed with a special focus on raveling related condition survey, rating computation, and treatment decisions. The discussion includes the condition survey spatial units as the pavement condition survey spatial units have changed from the visual survey to the automatic survey using automatically collected 2D and 3D pavement images. The resulting finer survey unit also creates the opportunity for spatial optimization of maintenance and rehabilitation decisions to save money for transportation agencies.

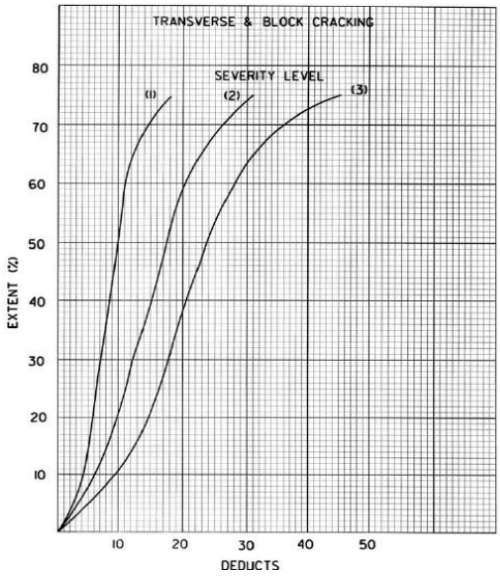
2.4.1 Georgia Department of Transportation (GDOT) (GDOT, 2017)

In GDOT, pavement condition has been surveyed based on a fixed spatial unit (1 mile) by selecting a 100-ft representative sample survey section in each one-mile segment. The defects noted for each rating segment (e.g. 1 mile) within a project (a length of roadway with a common pavement section, similar structural conditions, and logical beginning and ending points, say 10 miles) are then averaged to obtain the overall pavement condition rating for that project. A project rating is computed for the deduct values from individual distress types (10 types of pavement distresses), as shown below. Because the deducts for individual distresses are available in the GDOT deduct and rating computation, it is easy to apply an adequate and cost-effective treatment based on the following individual distress conditions (e.g. raveling-only).

1. Rutting (depth)
2. Load cracking (level 1-4)
3. Block cracking (level 1-3)
4. Reflection cracking (level 1-3)
5. Raveling (level 1-3)
6. Edge distress (level 1-3)
7. Bleeding/Flushing (level 1 and 2)
8. Corrugations/Pushing (level 1-3)
9. Loss of section (level 1-3)
10. Patches and pothole (counts)

For each distress type, the predominant severity level and percent extent are identified, and the deduction of each distress type is calculated based on Figure 10. With the separated deduction of each distress type, the rating of each segment (1 mile) is then deducted from 100. To calculate the project rating, the average extent and predominant severity level for the entire project are calculated. Then, the deduction of each distress type is determined by Figure 10 and the project rating is deducted from 100. With the project rating and the individual distress deduction values, the treatment is then determined by the developed decision tree shown in Figure 11.





Rutting Extent (inches)							
	0	1/8	1/4	3/8	1/2	5/8	3/4
Deducts	0	2	5	12	16	20	24

	Patches and Potholes Extent (# per mile)				
	1-2	3-6	7-10	11-15	>15
Deducts	2	5	10	17	25

		Corrugations/Pushing Extent (%)		
		1-10	11-25	>25
Severity	1	1	2	4
	2	2	4	7
	3	3	6	10

		Raveling Extent (%)					
		1-5	6-15	16-25	26-35	36-45	>45
Severity	1	2	5	6	8	10	13
	2	4	8	11	14	17	21
	3	6	12	16	20	25	30

		Edge Cracking Extent (%)			
		5-25	26-50	51-75	>75
Severity	1	1	2	3	4
	2	2	4	6	7
	3	3	6	8	10

		Loss of Pavement (%)			
		0-25	25-50	50-75	75-100
Severity	1	0	1	2	3
	2	2	4	6	8
	3	6	5	10	12

		Bleeding or Flushing Extent (%)		
		1-10	11-30	>30
Severity	1	2	5	8
	2	5	10	15

		Reflective Cracking (%)				
		5-15	16-30	31-45	46-50	>51
Severity	1	3	5	6	8	10
	2	6	8	11	14	20
	3	10	15	20	25	30

Figure 10. The distress deductions in GDOT's pavement rating computation (GDOT, 2017).

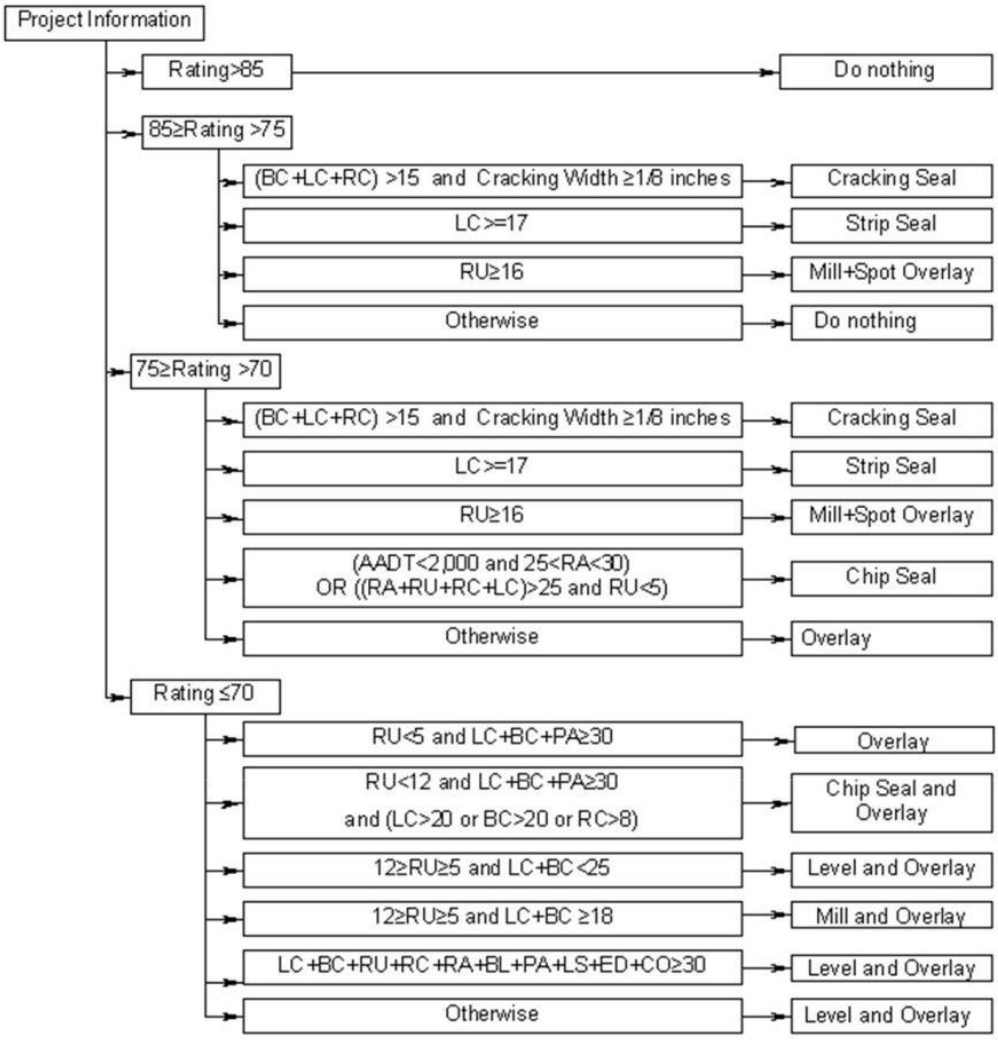


Figure 11. The treatment decision tree in GDOT's practice (GDOT, 2017). BC: block cracking deduct, LC: load cracking deduct, RC: Reflective cracking deduct, RA: raveling deduct, RU: rutting deduct, BL: bleeding deduct, PA: patches and potholes deduct, LS: loss of pavement, ED: edge cracking deduct, CO: corrugations deduct.

In GDOT, raveling is classified into Severity Levels 1, 2, and 3 based on the following definitions of raveling conditions:

1. **Level 1**: Loss of a substantial number of stones.
2. **Level 2**: Loss of most of the surface.
3. **Level 3**: Loss of a substantial portion of the surface layer (>1/2 depth).

By considering the condition rating of raveling along with the condition ratings of other pavement distresses (e.g., cracking, rutting, etc.), the raveling-related treatment is determined by the treatment decision tree in Figure 11. As shown in the figure, micro-milling and thin overlay

have been successfully developed and applied to cost-effectively treat raveling-only OGFC pavements in Georgia.

In addition to the treatment decision tree shown in Figure 2, Georgia Tech has worked with GDOT to perform a series of research projects to develop cost-effective pavement preservation technologies to treat predominantly raveling distresses on Georgia's interstate highways with open-graded friction course (OGFC) by applying (1) fog seal (Tsai, et. Al., 2021) at an early stage to slow down loss of aggregates and (2) micro-milling and thin overlay only for the OGFC layer at the right time. For micro-milling and thin overlay treatment only, studies include (a) characterizing micro-milling surface construction quality using newly established Ridge to Valley Depth (RVD) using 3D technology (Tsai, et. al. 2012; Tsai, et. al. 2014), (b) long-term performance of micro-milling and thin overlay (Tsai, et. al, 2016; Tsai, et. al. 2018; Tsai, et. al. 2019), and (c) its benefits on environmental impact and sustainability (Gadsby & Tsai, 2020).

There are two major differences between GDOT's and FDOT's rating systems. First, in GDOT, each distress has its individual deduct value. These individual deduct values are combined to obtain an overall deduct. With this two-level rating system, each distress can be more easily identified, and the corresponding treatments and trigger points can be designed in the treatment decision tree. Second, in GDOT's pavement management system, a pavement project (say a 10 miles pavement project) is defined as a pavement treatment section with consistent roadway characteristics to which the same treatment can be applied. However, the pavement survey is in a finer unit (i.e. one mile or less than one mile) in the entire roadway network. This provides one mile of spatial resolution to differentiate among pavement conditions so cost-effective pavement treatments can be applied. FDOT's pavement condition survey unit is the same as a treatment section (say 10 miles). This scale is too coarse to allow the use of pavement condition survey condition to support in-depth condition survey measures, like coring for the milling depth or making a finer and more suitable and optimal treatment decision spatially.

2.4.2 Texas Department of Transportation (TxDOT) (TxDOT, 2015)

The Pavement Management Information System (PMIS) in TxDOT contains more than 195,000 data collection sections. On average, the PMIS data collection sections are 0.5 mile in length, although some are longer, and some are shorter. When traveling across a PMIS section, the raters rate the lane that has the most distress on each roadbed. The lane being rated can change from section to section while traveling down the road. When deciding which lane to rate, the distresses from worst to best (generally) are failures (a localized section of pavement where the surface has been severely eroded, badly cracked, depressed, or severely shoved), alligator cracking, block cracking, patching, longitudinal cracking, and transverse cracking.

At the end of each PMIS section on flexible pavement, the raters enter their section ratings for all the flexible pavement distress types as shown below. The rating consists of a number for each of these ten distress types. The numbers indicate either the area or the amount of each distress that was observed.

1. Rutting- shallow (0.5 inch–1 inch): Area (0–100%)
2. Rutting – deep (1 inch–3 inch): Area (0%–10%)
3. Patching: Area (0% – 100%)

4. Failures: Total number (0~99)
5. Block cracking: Area (0~100%)
6. Alligator cracking: Area (0~100%)
7. Longitudinal cracking: Linear feet per station
8. Transverse cracking: Number per station
9. Raveling (0~3)
10. Flushing (0~3)

Different from FDOT and GDOT, the classification of raveling severity levels is only based on the percentage of raveled area by TxDOT, which is defined as follows:

1. **Low**: The percent of raveled pavement area is from 1% to 10%.
2. **Medium**: The percent of raveled pavement area is from 11% to 50%.
3. **High**: The percent of raveled pavement area is greater than 50%.

In TxDOT, the PMIS treatment decision support matrix for asphalt pavements is shown in Table *6*. In the table, the pavement distresses are classified into low, medium, and high extent levels based on Table 7. As shown in Table *6*, in addition to pavement distress, factors such as traffic and ride quality are also taken into consideration in the treatment decision. Note that for raveling, the recommended treatments in TxDOT are preventive maintenance. Also note that thin overlay treatment is recommended when raveling is presented but other distress, such as rutting and cracking, are not severe.

Table 6. PMIS treatment decision support matrix for asphalt pavements (Chang et al., 2014).

Level	Treatment	Climatic Zone				Traffic			Predominant Pavement Distress																								Ride Quality		
									Rutting				Patching	Failures	Block Cracking			Alligator Cracking			Longitudinal Cracking			Transverse Cracking			Raveling			Flushing					
		1	2	3	4	L	M	H	Shallow	Deep	Severe	Failure	-	-	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
Preventive Maintenance	Cape Seal	●	●	●	●	●	●	◑	●	●	◑		●	◑	●	●	●	●	●	◑	●	●	●	●	●	●	●	●	●	●	●	●	●	◑	
	Fog Seal	1*	1*	●	●	●	◑	◑					●		●	●	●	●	◑		●	●	●	●	●	●	●	●	●						
	Microsurfacing	●	●	●	●	●	●	●	●	●	◑		●					◑			◑			◑			●	●	●	●	●	●			
	Seal Coat	●	●	2*	2*	●	◑	◑	●	◑			●	◑	●	●	●	●	◑		●	●	●	●	●	●	●	●	●	●	◑	◑		◑	◑
	Thin Overlay 2" Thick or Less	●	●	●	●	◑	●	●	●	●	◑		●	◑	●	●	◑	●	◑		●	●	◑	●	●	◑	●	●	●	●	●	●	●	◑	
	Ultrathin Surface Course	●	●	●	●	◑	●	●	●	◑			●	◑	●	●	◑	●	◑		●	●	◑	●	●	◑	●	●	●	●	●	●	◑		
Light Rehabilitation	Base Repair and Seal	●	●	●	●	●	●	◑		◑	●	●		●				●	●	●													●	◑
	Cold In-Place Recycling	●	●	●	●	◑	●	◑		◑	◑	◑	●	◑		◑	●	●	◑	◑		◑	●		◑	●			●		◑	●	●	●	◑
	Hot In-Place Recycling	●	●	●	●	◑	●	●		◑	◑	◑	●	◑		◑	●	●	◑	◑		◑	●		◑	●			●		◑	●	●	●	◑
	Mill and Inlay; or Mill, Seal and Thin Overlay	●	●	●	●	◑	●	●		◑	●	◑	●	◑			◑	●	●	◑			◑			◑					◑	●	●	●	◑
	Overlay Greater Than 2" Thickness But Less Than 3"	●	●	●	●		●	●		◑	●	◑	◑	◑			◑	●	●	◑			◑			◑					◑	●		●	◑
Medium Rehabilitation	Base Repair, Spot Seal, Edge Repair and Overlay	●	●	●	●	◑	●	●		●	●	◑	◑	●			◑	●	●	◑			◑											●	●
	Level Up and Overlay	●	●	●	●	●	●	●	◑	●	●	◑	◑	●			◑	●	●	◑			◑											●	●
	Mill and Overlay	●	●	●	●	●	●	●		◑	●	◑		◑		◑	●		◑	◑			◑							◑	●			●	●
	Mill, Stabilize Base, and Seal	●	●	●	●	●	●	●		◑	●	●							◑	●														●	●
	Overlay Between 3" and 5"	●	●	●	●	●	●	●	●	●	●	●							◑	●														●	●
Heavy Rehabilitation	Full Depth Reclamation (Pulverization and Resurfacing)	●	●	●	●	●	●	●	●	●	●	●							◑	●															3*
	Mill, Cement Stabilize Base, and Overlay	●	●	●	●	●	●	●	●	●	●	●							◑	●															3*
	Reconstruction	●	●	●	●	●	●	●	●	●	●	●							◑	●															3*
	Thick Overlay Greater Than 5"	●	●	●	●	●	●	●	●	●	●	●							◑	●															3*

● = Recommended ◑ = Conditionally Recommended (may depend on other factors)

1* - Fog Seal not recommended for Medium and High Traffic sections
2* - Seal Coat not recommended for High Traffic sections

3*- Heavy Rehabilitation to correct poor ride quality (RS< 1.5 or IRI > 270 in/mi) is recommended when such treatment will also correct severe distortion and/or fracture in the existing pavement (such as in areas with expansive soils)

Table 7. Distress extent level definitions (Chang et al., 2014).

Distress	Units	Extent Level		
		Low	Medium	High
Alligator Cracking	Percent area	< 15	15 - 50	> 50
Block Cracking	Percent area	< 20	20 - 50	> 50
Longitudinal Cracking	Linear feet per station	< 100	100 - 175	> 175
Transverse Cracking	Number per station	< 6	6 - 10	> 10
Raveling	Percent area	≤ 10	11 - 50	> 50
Flushing	Percent area	≤ 10	11 - 50	> 50

2.4.3 California Department of Transportation (Caltrans) (Caltrans, 2017)

In Caltrans, pavement condition data are collected continuously by customized vehicles fitted with sensing equipment. These data are then aggregated into larger segments as shown in Figure *12*. The data collection element (DCE) is the smallest unit distresses are collected, with length that is usually 26.4 ft. The 0.1-mile and 1-mile segments are the two levels of aggregation of data collection elements for which pavement distresses and conditions are reported. The pavement data collection segment (DCS) is the smallest unit for which pavement distresses and conditions are stored in the PaveM system. DCS is 0.1 mile or 528 ft in length, measured in the longitudinal direction. A DCS is comprised of multiple DCEs. Finally, the pavement reporting segments (PRS) are used to report the condition of the road in segments of about one mile, which includes multiple DCSs. For each unit (DCE, DCS, and PRS), each severity of each distress is recorded by extent percentage.

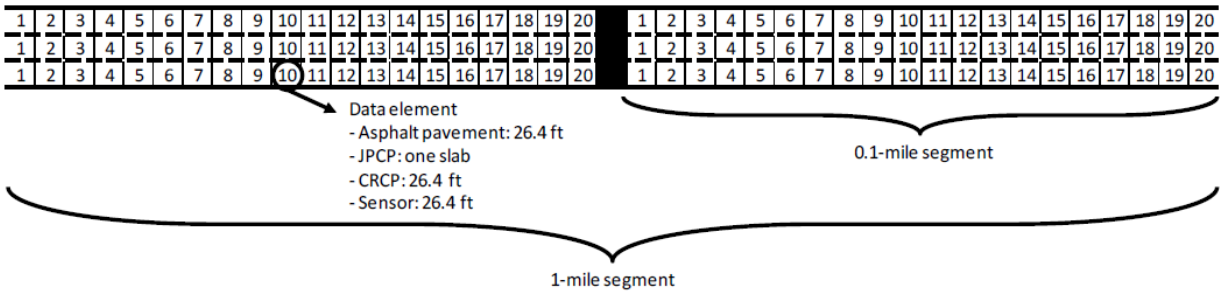


Figure 12. Condition report segments in Caltrans (Caltrans, 2017).

Like GDOT and TxDOT, Caltrans also uses individual ratings for each distress. Typically, the recorded numbers indicate either the length, area, or the amount of each distress that was observed. Each distress is also categorized into three severity levels (low, medium, and high) based on specific factors.

1. Alligator A crack: Length (ft) and average width (severity factor)
2. Alligator B crack: Length (ft) and average width (severity factor)
3. Block crack: Area (ft^2) and the average size of the block (severity factor).
4. Edge crack: Length (ft) and average width (severity factor)
5. Longitudinal crack: Length (ft), average width (severity factor), and the number of cracks.
6. Transverse crack: Length (ft), average width (severity factor), and the number of cracks.

7. Pothole: Area (ft^2), average depth, and the number of potholes (severity factor).
8. Bleeding: Length (ft) and length ratio (severity factor).
9. Raveling: Area (ft^2) and area ratio (severity factor).

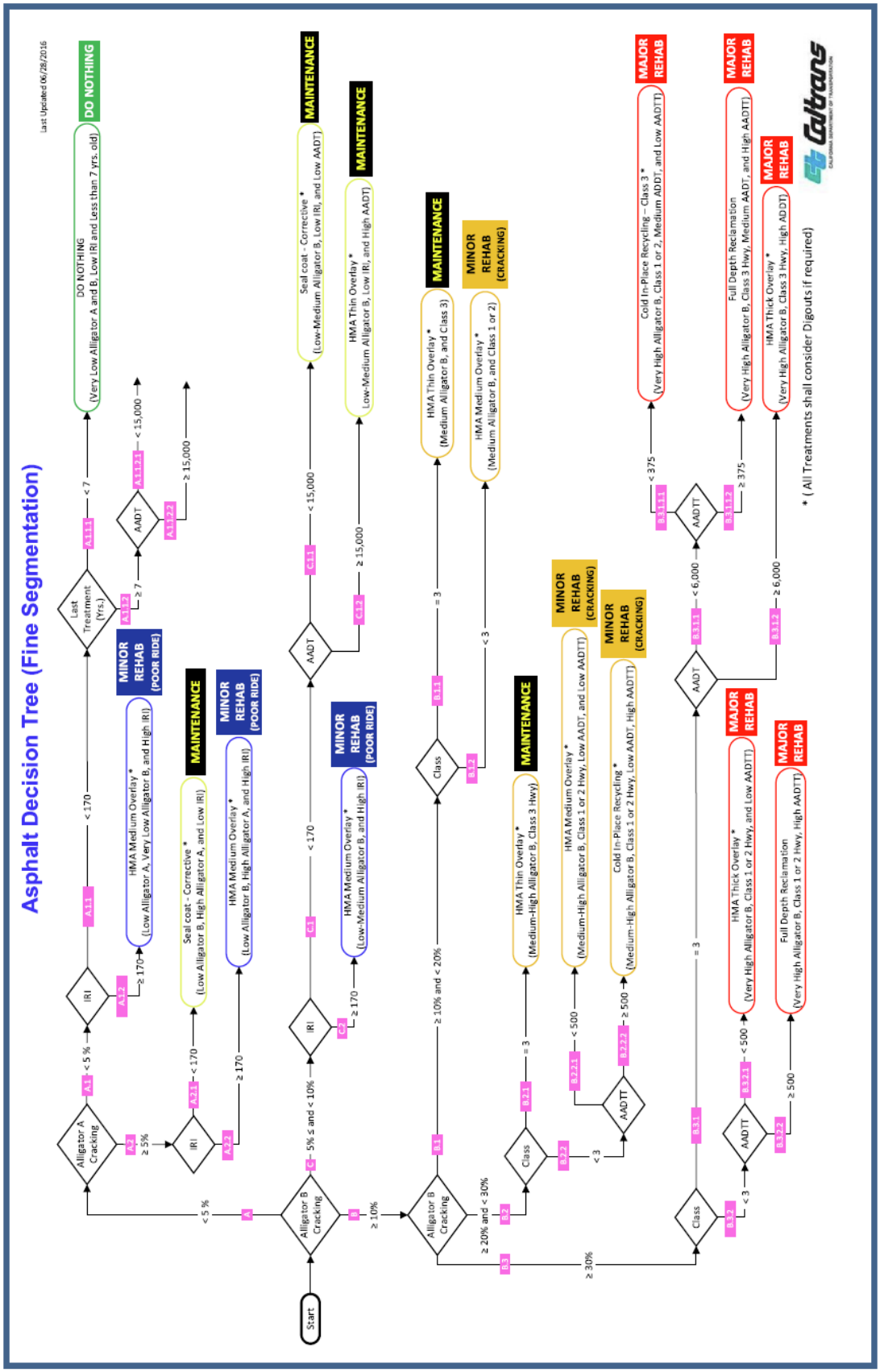
Similar to TxDOT, the classification of raveling severity levels in Caltrans are based only on the percentage of raveled area, which is defined as follows:

1. **Low**: The percent of raveled pavement area is from 1% to 10%.
2. **Medium**: The percent of raveled pavement area is from 11% to 50%.
3. **High**: The percent of raveled pavement area is greater than 50%.

In Caltrans, the treatment decision tree for asphalt pavements is shown in Figure 13. Note that the alligator A and B cracks are the two distresses used in this decision tree. Other important factors for treatment decisions in Caltrans include IRI, average daily traffic, and the roadway class defined by Caltrans.

In summary, Texas DOT and Caltrans classify raveling severity levels (low, medium, and high) based on only the percentage of raveled areas. This is different from GDOT and FDOT survey raveling conditions based on both the severity level (severity of aggregate loss) and the percentage of raveled areas.

Figure 13. The treatment decision tree for asphalt pavement in Caltrans (Wang and Pyle, 2019).



2.4.4 Raveling Treatment Alternatives

FDOT is seeking alternative treatments to cost-effectively treat raveling-only pavement distress (e.g. on FC5 pavement layer) in Florida's interstate highways. An alternative raveling-only pavement preservation method is presented below. It is the ***micro-milling and thin overlay***. It has been successfully developed and implemented by GDOT since it was applied on I-75 near Perry, Georgia in 2007. It has been applied on many interstate highway sections in Georgia. The micro-milling and thin overlay have been successfully developed and applied to cost-effectively treat raveling-only OGFC pavements in Georgia. The outcomes have been published in various journal papers (Gadsby and Tsai, 2020; Tsai et al., 2018; Tsai et al., 2016; Tsai et al., 2014b; Tsai et al., 2012; Lai et al., 2012). Based on these intensive research studies, the benefits of micro-milling and thin overlay are summarized below:

- **Economic Assessment**: Results showed that the micro-milling and thin overlay method's expected service life of 10-12 years is similar to the conventional method but will save Georgia $65,600 per lane mile over the conventional method. Its break-even service life is 5 years in comparison to the conventional method. It is also a good crack relief treatment, evidenced by having only 5% cracking occurring in 5 years on an I-95 project using 3D sensing technology.
- **Environmental and Social Assessment:** it produces 60% percent fewer greenhouse gases, uses 60% percent less water, and uses 60% percent less energy than the conventional method. The qualitative evaluation of social impacts, such as the reduced construction time and increased flexibility, shows improved safety for drivers and workers.
- **Overall**, micro-milling and thin overlay is a promising, sustainable pavement preservation alternative that will save money for transportation agencies if it is applied appropriately on pavements with sound structural conditions.

The following are the selected papers and reports related to a cost-effective alternative raveling-only treatment – "***micro-milling and thin overlay***".

Selected Papers:

- Gadsby, A. & Tsai, Y. (2020). "Environmental Impact of New Micro-Milling and Thin Overlay and Conventional Milling for Sustainable Pavement Preservation", International Journal of Pavement Research and Technology. June 2020
- Tsai, Y., Wu, Y. Geary, G. (2018). "Sustainable and Cost-Effective Pavement Preservation Method: Micro-Milling and Thin Overlay." *Journal of Transportation Engineering*, Part A: Systems, Volume 144, Issue 10, October 2018.
- Tsai, Y., Wu, Y., Gadsby*, A., Hanes, S. (2016). "Critical Assessment of the Long-term Performance and Cost-effectiveness of a New Pavement Preservation Method: Micro-Milling and Thin Overlay." Transportation Research Record, 2550: 8-14. Top 3%, 160 papers out of 5600 papers accepted for early publication.
- Tsai, Y., Wu, Y., and Lewis*, Z. (2014). "Full-Lane Coverage Micromilling Pavement-Surface Quality Control Using Emerging 3D Line Laser Imaging Technology." *Journal of Transportation Engineering*. Volume 140 Issue 2.

- Tsai, Y., Wu, Y., Lai, J., Geary, G. (2012) "Characterizing Micro-milled Pavement Textures Using RVD for Super-thin Resurfacing on I-95 Using A Road Profiler", Transportation Research Record, No.2306, pp.144-150.
- Lai, J., M. Hines, S., Wu, P.Y., and Jared, D. (2012). "Pavement Preservation with Micromilling in Georgia – Follow-Up Study*"*. *Transportation Research Record*. Vol. 2292, pp. 81-87.

Selected Reports:

- Tsai, Y, Wang, Z., Gadsby A. (2018) *Evaluation of the Long-term Performance and Benefit of Using An Enhanced Micro-milling Resurfacing Method* Draft final report, Georgia Department of Transportation.
- Tsai, Y, Wu, Y., Lai, J. S. (2012) *Validation of RVD-Based Micro-milled Pavement Surface Texture Quality Control* Draft final report, Georgia Department of Transportation.
- Lai, J. S. (2011) *Assessing Techniques and Performance of Thin OGFC/PEM Overlay on Micro-milled Surface*, Final Report, Georgia Department of Transportation.

3. Feasibility Study of Machine Learning for Raveling Classification

This chapter presents the feasibility study of ML for automated raveling severity level classification for FDOT. Three ML models were applied to determine their feasibility on FDOT's pavement data. These models were also used in Tsai et al. (2021). The pavement data that are used in this study are the 3D range images that FDOT already collected using the laser crack measurement system (LCMS). This chapter includes the introduction of the data used in this study, the applied ML models, the feasibility study of the ML models on FDOT's data, and the conclusions and recommendations.

3.1 FDOT's Pavement Data

3.1.1 Data description

The pavement data used in this study are 3D range images collected and provided by FDOT. The provided range images contain the quantized range value with a range of 0 to 255 at each pixel location. The provided data spreadsheet from FDOT includes the unique ID (including the Road ID and Image ID), GPS location, and crack rating, and the MPD and RI of each image data. These 3D pavement images were collected from seven different sections in six routes (each with a unique Roadway ID). These sections were selected to ensure the diversity of the severity levels of raveling. Table 8 shows the general information of these selected sections. The range images provided by FDOT for this feasibility study are approximately 8.5 miles in total, as shown in Table 8.

Table 8. Selected sections of the 3D pavement image.

Roadway ID	US	Rated Lane	PCS BMP	PCS EMP	PCS Crack Rating	PCS Raveling	PCS Patching	Image BMP	Image EMP	Num. of Images
26260000	75	R3	0	9.27	10	None	None	0.014	0.984	257
27090000	10	R2	8.942	11.1	4.5	L3	None	8.947	11.098	569
29170000	10	R2	16.323	20.69	4.5	M3	2	18.616	20.691	549
37120000	10	R2	5.861	15.099	9	S1	None	8.558	9.588	273
37120000	10	R2	15.099	25.508	4.5	S2	None	19.000	20.984	525
72001000	295	R3	13.2	14.2	10	None	None	13.201	14.193	263
72020000	95	R3	6.774	7.95	0	M3	None	6.776	7.920	303

3.1.2 Data annotation

Each pavement image was manually annotated with the raveling ratings by FDOT's rater, which has four levels based on FDOT's specification: None, Low, Medium, and Severe. The manual rating of the raveling severity served as the ground-truth for training and evaluating the ML models. FDOT has conducted two versions of raveling rating annotation for the data described in Section 5.1.1. The following discussion presents the difference between the two versions of annotation. The feasibility study in this chapter used Annotation Data Set 1.

Annotation Data Set 1: In this version, four FDOT raters participated in the rating of raveling severity, and each rater came up with an individual decision by looking at the 2D, 3D, and forward images (Figure 14) at the same time. The number of pavement images with each annotated predominant raveling severity is shown in Table 9. After the feasibility study, which is presented in Section 3.3.3, the following issues were identified:

1. Raters focused on different types and regions of the images during the annotations.
2. Different raters performed the rating individually and had different decisions, which caused the annotations to be inconsistent for some range images.
3. Scarring may affect the decision of the raveling rating annotation.

These issues can affect the quality of the ground-truth annotations for the training of ML models, which can cause the models to have degraded performance.

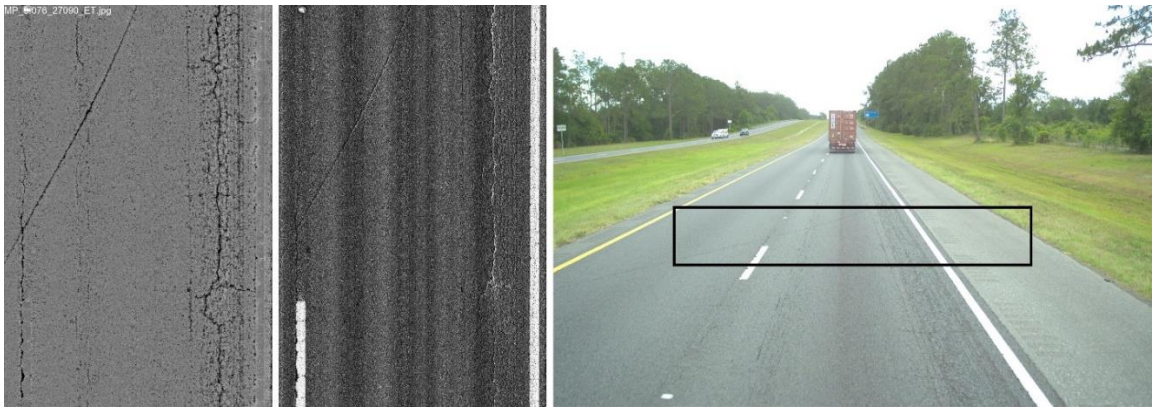


Figure 14. An example of the range: 3D image (left), 2D image (middle), and forward image (right) of the pavement.

Table 9. Statistics of the rating of predominant raveling severity in Annotation Data Set 1.

Roadway ID	None	Low	Medium	Severe	Total
26260000	257	0	0	0	257
27090000	20	51	329	148	548
29170000	20	33	194	301	548
37120000	16	250	224	308	798
72001000	263	0	0	0	263
72020000	23	0	54	201	278
Total	599	334	801	958	2692

Annotation Data Set 2: To resolve the issues in Annotation Data Set 1, multiple discussions were held with FDOT to identify a revised annotation practice that could better ensure the annotation quality of raveling ratings. The identified practice includes the following important items:

1. Record all raveling severity levels presented on each range image.
2. Record the decided predominant raveling severity level of each range image.
3. Focus mainly on the range image during annotation. This is due to using range images to train the ML models.
4. Instead of performing the rating individually, the raters collaborate and aggregate their decisions to come up with a final raveling rating annotation.
5. Following the practical survey process, scarring should not be considered as raveling when annotating the severity levels on range images.
6. Record if the predominant raveling severity level is difficult to decide for each range image. This can provide additional and useful information for model analysis.

The number of pavement images with each annotated predominant raveling severity is demonstrated in Table *10*. The results of the feasibility study using Annotation Data Set 2 are presented in Section 3.3.4.

Table 10. Statistics of the rating of predominant raveling severity in Annotation Data Set 2.

Roadway ID	None	Low	Medium	Severe	Total
26260000	257	0	0	0	257
27090000	27	197	324	0	548
29170000	22	482	44	0	548
37120000	137	347	279	8	798
72001000	263	0	0	0	263
72020000	23	0	224	31	278
Total	729	1053	871	39	2692

3.2 Methodology

This section introduces the applied ML models for the feasibility study. Specifically, given a 3D pavement image from FDOT, the ML models were trained to classify it into the severity of low, medium, or severe if raveling is presented. The model consists of two stages, the feature extraction stage and the classification stage based on the extracted features. The following subsections introduce the extraction of features and the ML classifiers.

3.2.1 Macrotexture Features

According to Tsai and Wang (2015), the macrotexture analysis of the 3D pavement images can provide good features for training ML models to perform raveling classification. These macrotexture features consist of two main categories, which are the statistical features of the whole image and the distributions of the statistical features on all small patches of an image.

Category 1 Features: The first category of the macrotexture features, which are the statistical features of the whole image, aims at capturing the changes in the characteristics of surface texture under different raveling severity levels. In this feasibility study, six macrotexture features were adopted from Tsai and Wang (2015), which are shown in Table 11.

Table 11. Macrotexture features for pavement images.

Feature	Formula
Arithmetic mean	$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$
Standard deviation	$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$
Root mean square	$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {x_i}^2}$
Skewness	$Sk = \frac{1}{N}\sum_{i=1}^{N} \frac{(x_i - \bar{x})^3}{\sigma^3}$
Kurtosis	$K = \frac{1}{N}\sum_{i=1}^{N} \frac{(x_i - \bar{x})^3}{\sigma^3}$
Interquartile range	$IQR = Q_3 - Q_1$

N: The total number of pixels.
x_i: The i^{th} pixel value of the flattened image.
Q_i: The i^{th} quartile of the pixel values.

Category 2 Features: The second category of the macrotexture features, which is the distributions of the statistical features on small patches of an image, aims at capturing the local characteristics of a raveled surface. For each range image, the six features (Table 11) are calculated for each small image patch with a size of 75 by 75 pixels, which is approximately 1 ft by 1 ft. For each feature, the distribution is captured by fitting a density function to the histogram of the feature values from all patches. The density function is then discretized by 100 data points. Thus, a feature vector with a size of 600 is extracted in this category.

The two categories of features are combined, which results in a total of 606 features for each range image. These features are then served as the inputs to the ML classifiers, which are introduced in the next section.

3.2.2 Raveling Severity Level Classifier

After the extraction of image features, the ML classifier is applied to perform the raveling severity level classification. Three ML models were applied, including Support Vector Classifier (SVC), Random Forest (RF), and Adaptive Boosting (AdaBoost) in the feasibility study because

of their previously demonstrated promising classification capability on pavement defects (Tsai et al., 2021; Wang et al., 2017). They are all supervised, shallow ML models.

Support Vector Classifier (SVC) Model: In SVC, training examples are mapped to points in space to maximize the width of the gap between the two categories. Testing examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall (Cortes and Vapnik, 1995).

Random Forest (RF) Model: In RF, multiple decision trees are constructed during training, and these decision trees are ensembled together to achieve more accurate and stable predictions (Ho, 1998).

Adaptive Boosting (AdaBoost) Model: In AdaBoost, a sequence of weak learners (i.e., models that are only slightly better than random guessing, such as small decision trees) are trained on repeatedly modified versions of the data. The predictions from all of them are then combined through a weighted majority vote to produce the final prediction (Freund and Schapire, 1997).

Given the 606 macrotexture features (Section 3.2.1) as the inputs, these three models were trained on the data with annotated raveling ratings (Section 3.1.2) to predict the raveling severity level of each range image. In this project, all models are implemented with Python and Scikit-learn, which is a free software machine learning library for Python, for research purposes.

3.3 Evaluation of the ML Models on FDOT's Data

The work described in this section evaluated the ML models introduced in Section 3.2.2 using FDOT's data to analyze its feasibility. This section first presents the preparation of data and the evaluation metrics. Then, the evaluation results and discussions are demonstrated.

3.3.1 Data preparation

Three ML models (SVC, RF, AdaBoost) were used for raveling classification in this feasibility study. The data provided by FDOT were used for this feasibility study, and the data were randomly split into training, validation, and testing datasets without duplication. The training dataset was used to train the ML models, the validation dataset was used to tune the hyperparameters of the models, and the testing dataset served as a set of unseen data to evaluate the performance of the models. To ensure that the training, validation, and test datasets contained a sufficient portion of each severity level, a stratified random splitting technique was applied with a distribution of 70%, 10%, and 20%, respectively. The same data splitting was applied to both Annotation Data Set 1 and Annotation Data Set 2. The number of range images for each severity level in the three datasets is shown in Table 12 and Table 13.

Table 12. The distribution of the datasets of Annotation Data Set 1.

Dataset	None	Low	Medium	Severe	Total
Train	419	233	561	670	1883
Validation	60	34	80	96	270
Test	120	67	160	192	539
Total	599	334	801	958	2692

Table 13. The distribution of the datasets of Annotation Data Set 2.

Dataset	None	Low	Medium	Severe	Total
Training	510	736	610	27	1883
Validation	73	106	87	4	270
Test	146	211	174	8	539
Total	729	1053	871	39	2692

3.3.2 Evaluation Metric

The ML models were evaluated by their performance on the testing dataset. Conventional metrics for multiclass classification were used for the evaluation. The accuracy of correctly predicting the severity levels gives a general performance measurement of the models, which can be calculated as follows:

$$\text{Accuracy} = \frac{\text{Total number of correct predictions}}{\text{Total number of testing samples}}$$

To have a closer analysis of the performance, the recall and precision of each target class $i\{\text{None}, \text{Low}, \text{Medium}, \text{Severe}\}$ was used to provide the performance measurement of the models on each raveling severity level, which can be calculated as follows:

$$Recall_i = \frac{TP_i}{TP_i + FN_i}$$

$$Precision_i = \frac{TP_i}{TP_i + FP_i}$$

Where TP_i, FN_i, and FP_i are the true positives (TPs), false negatives (FNs), and false positives (FPs) of severity level i, respectively.

3.3.3 Evaluation Results and Discussions on Annotation Data Set 1

Table 14 demonstrates the evaluation results of different ML models on Annotation Data Set 1 and Annotation Data Set 2. In this section, the results on Annotation Data Set 1 (i.e., the middle column of Table 14) are discussed. The table shows that both the SVC and RF achieved around 80% accuracy in classifying the predominant raveling severity of pavement images.

Table 14. The accuracy of different annotation data sets.

Model	Annotation Data Set 1	Annotation Data Set 2

SVC	77.9%	**84.6%**
RF	79.1%	**86.6%**
AdaBoost	57.0%	**66.2%**

Figure 15 shows the range images where the raveling severity levels were correctly predicted by the RF model. Among the three ML models, AdaBoost had the lowest accuracy of only 57%. This indicates that AdaBoost is not a suitable model for pavement raveling classification. On the other hand, RF achieved the best performance. One potential reason is that a total of 606 features were used and RF is especially effective when there is a large number of features. These results demonstrate that the ML models are feasible for FDOT's pavement data with suitable ML classifiers, such as RF and SVC.

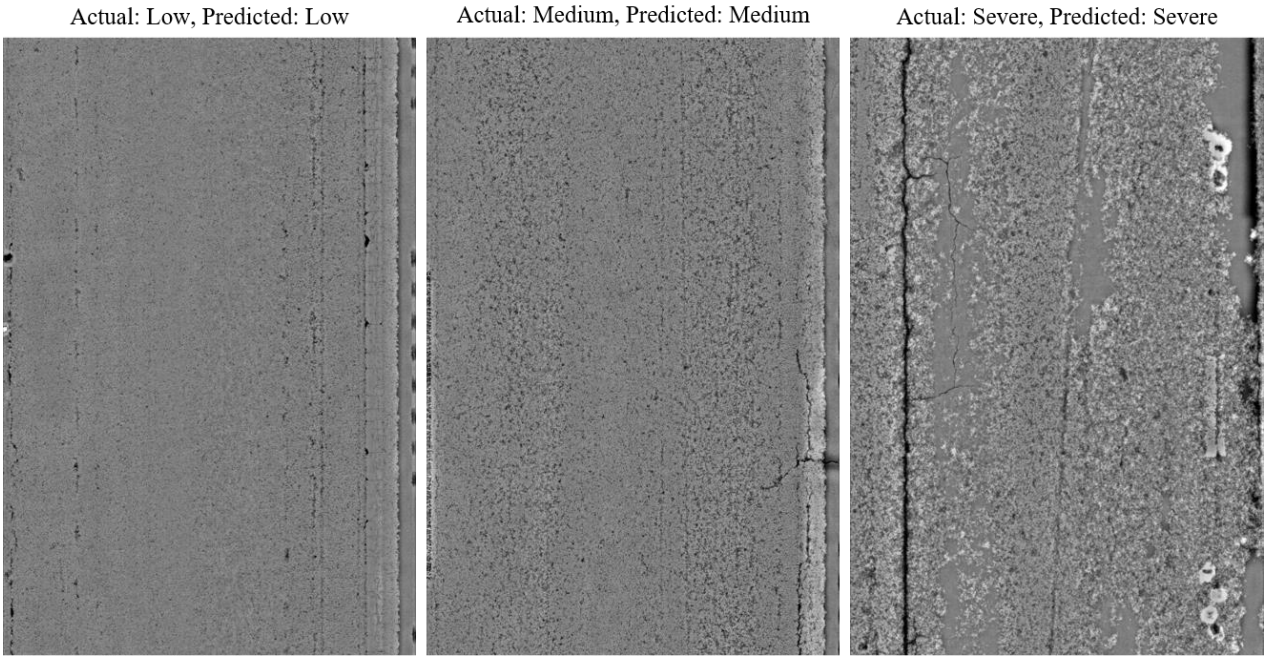


Figure 15. Cases where the RF model correctly predicted the raveling severity level.

To identify what should be done in the future for further improvement, a closer analysis was conducted for the RF model given that it had the best accuracy. Table 15 presents the confusion matrix of RF, along with the recall and precision values of each raveling severity level. From the table, it can be seen that a large portion of errors occurred in the confusion between Medium and Severe raveling. Therefore, the range images were evaluated that had the annotated predominant raveling severity as medium and severe. It was found that the major issue with these data is the inconsistency of the raveling rating annotation. That is, the standard for deciding the predominant level is not consistent among all range images and all raters. As shown in Figure 16, both range images were annotated with a raveling severity of severe while the pavement characteristics were very different. Figure 17 shows that the range images were annotated differently even when they had similar pavement characteristics.

Table 15. The confusion matrix of RF trained with Annotation Data Set 1.

Confusion Matrix						
Predicted / Actual	None	Low	Medium	Severe	**Recall**	**Precision**
None	117	3	1	0	0.97	0.94
Low	5	48	11	3	0.72	0.73
Medium	0	12	93	54	0.58	0.66
Severe	0	0	37	151	0.79	0.73

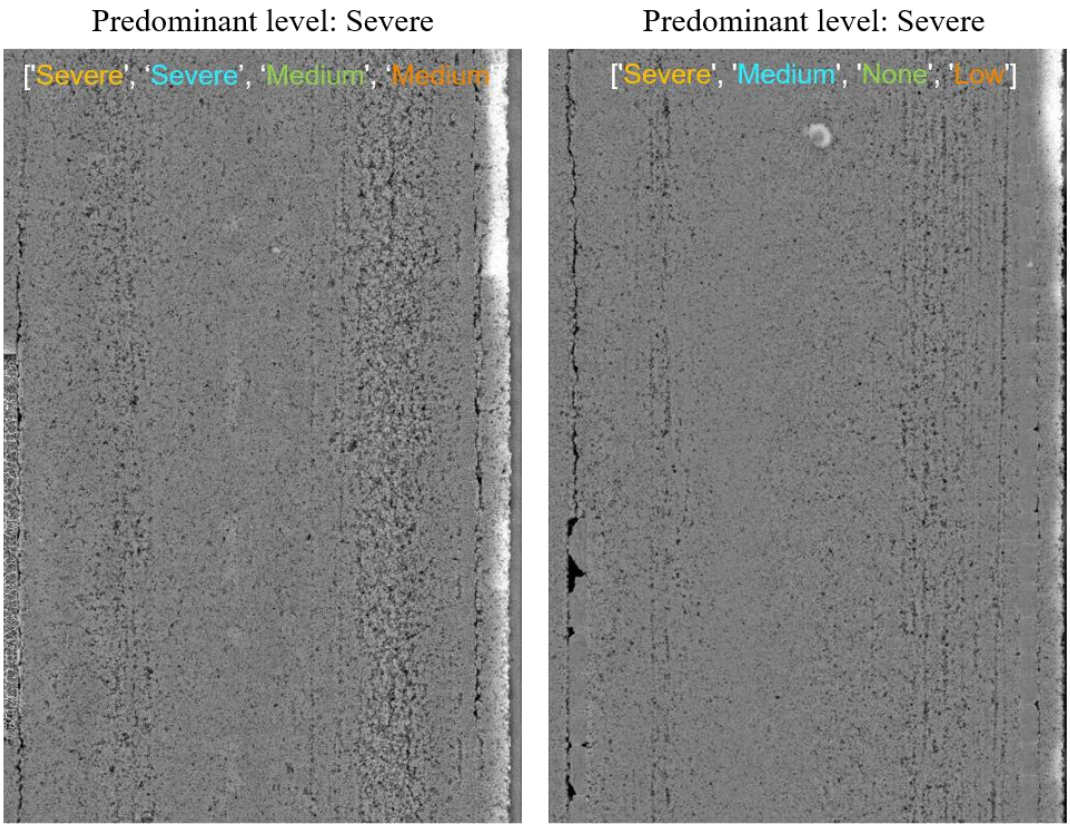


Figure 16. The inconsistency of raveling rating annotation in Annotation Data Set 1. The predominant level selected based on the majority annotations from four raters (shown with different colors on each image). If annotations from all raters are different, the predominant level follows the rating from the first rater.

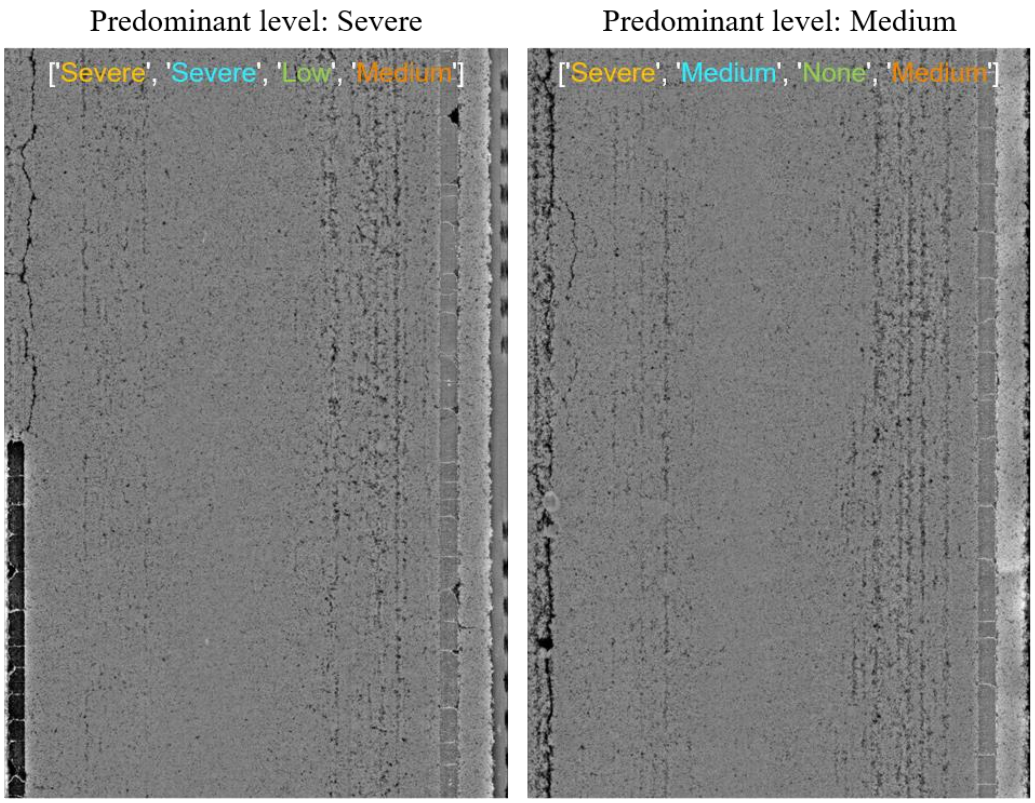


Figure 17. The inconsistency of raveling rating annotation in Annotation Data Set 1.

It was also noted that the inconsistency of annotation occurred more when the ratings from different raters were very diverse, such as the right image in Figure 16. These inconsistencies can be caused by the raters performing the rating individually by looking at different types and regions of the images shown in Figure 14. Furthermore, scarring on the pavement, which should not be considered as raveling, also affects the annotation of some images. An example of the scarring can be seen in Figure 18. The issues above can cause the degradation of annotation quality, which further leads to a degradation of the performance of ML models. Therefore, the most critical step for future improvement is to improve the annotation quality. Through a discussion with FDOT, a revised annotation practice was identified that can better ensure the annotation quality, as described in Section 3.1.2.

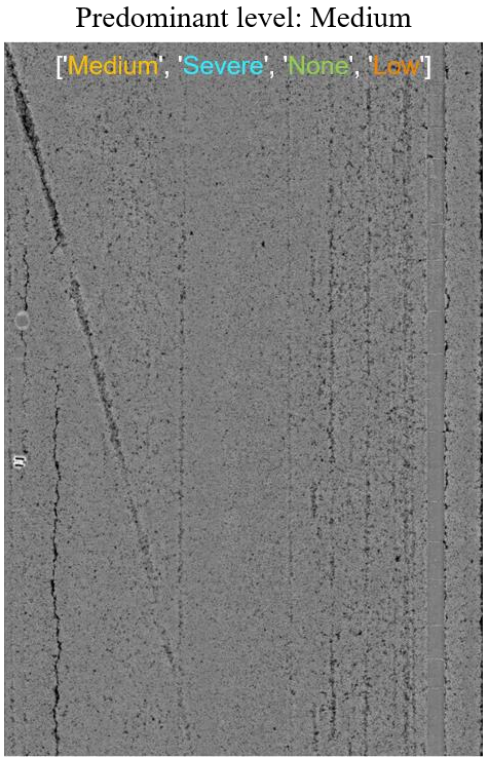


Figure 18. The effect of scarring on the raveling rating annotation in Annotation Data Set 1.

3.3.4 Evaluation Results and Discussions on Annotation Data Set 2

In this section, the evaluation results on Annotation Data Set 2 (i.e., the last column of Table 14) are discussed. Two important observations can be made from the table:

1. The quality of annotation largely affects the performance of the raveling classification models. From Table 14, it can be seen that with the revised annotation, all models trained with Annotation Data Set 2 achieved around 7% higher accuracy than models trained with Annotation Data Set 1.
2. Among the three models, RF achieved the best performance in both Annotation Data Set 1 and 2. This indicates that RF is the most suitable ML classifier for the automated raveling classification task based on FDOT's data.

Table 16 presents the confusion matrix for RF trained with Annotation Data Set 2. Figure 19 shows several examples that raveling severity levels were correctly predicted by the RF model. From the confusion matrix, it is clear that the model achieved better recall and precision values compared to the same model trained with Annotation Data Set 1. This demonstrates the importance of annotation quality in the classification of raveling severity levels. One major problem is the confusion between Low and its nearby severity levels, which is primarily caused by the ambiguity of the adjacent severity levels. Although Severe raveling is much easier to identify, the decision between None and Low, or Low and Medium, can be challenging through visual inspection on pavement images when the raveling condition is at the boundary of adjacent severity levels. Figure 20 shows several cases that the raveling severities are at the boundary of Low and Medium, and

the RF model did not generate correct predictions. Future studies can explore more systematic and quantitative labeling methods, such as incorporating calculated macrotexture features with visual inspections. Another future improvement is to include more data with the severe level or use over-sampling techniques to improve the model's performance on the raveling with severe severity.

Table 16. The confusion matrix of RF trained with Version 2 Annotation.

Confusion Matrix					**Recall**	**Precision**
Predicted / Actual	None	Low	Medium	Severe		
None	135	11	0	0	0.92	0.95
Low	7	181	23	0	0.86	0.83
Medium	0	27	147	0	0.84	0.84
Severe	0	0	4	4	0.50	1.00

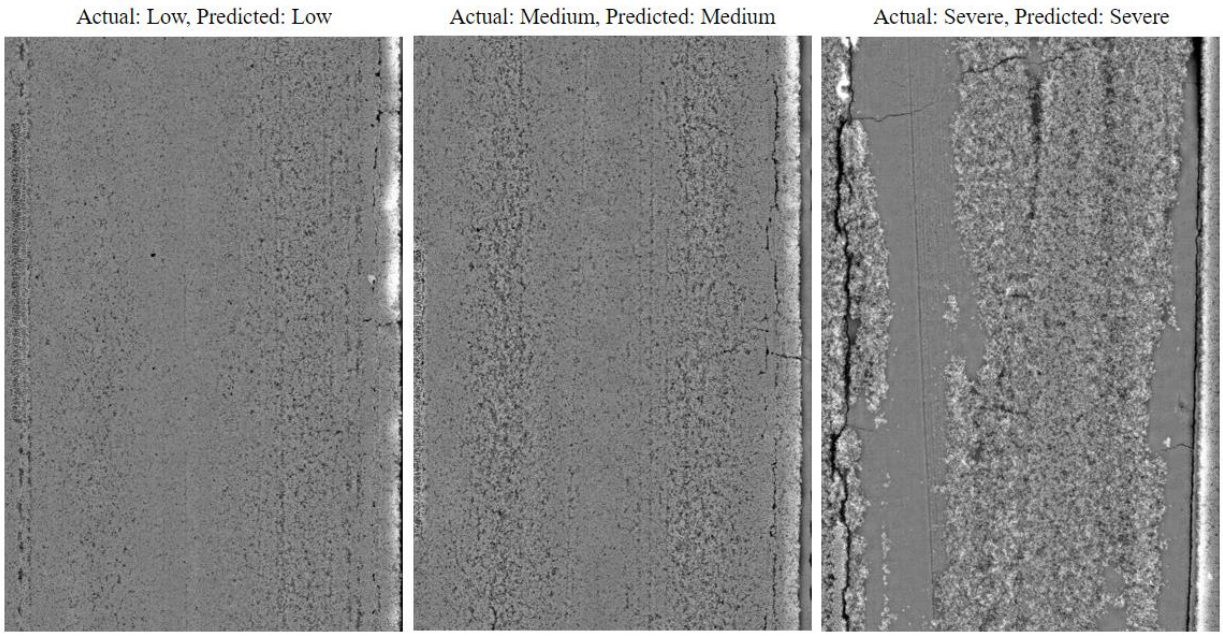


Figure 19. Cases where the RF model correctly predicted the raveling severity level.

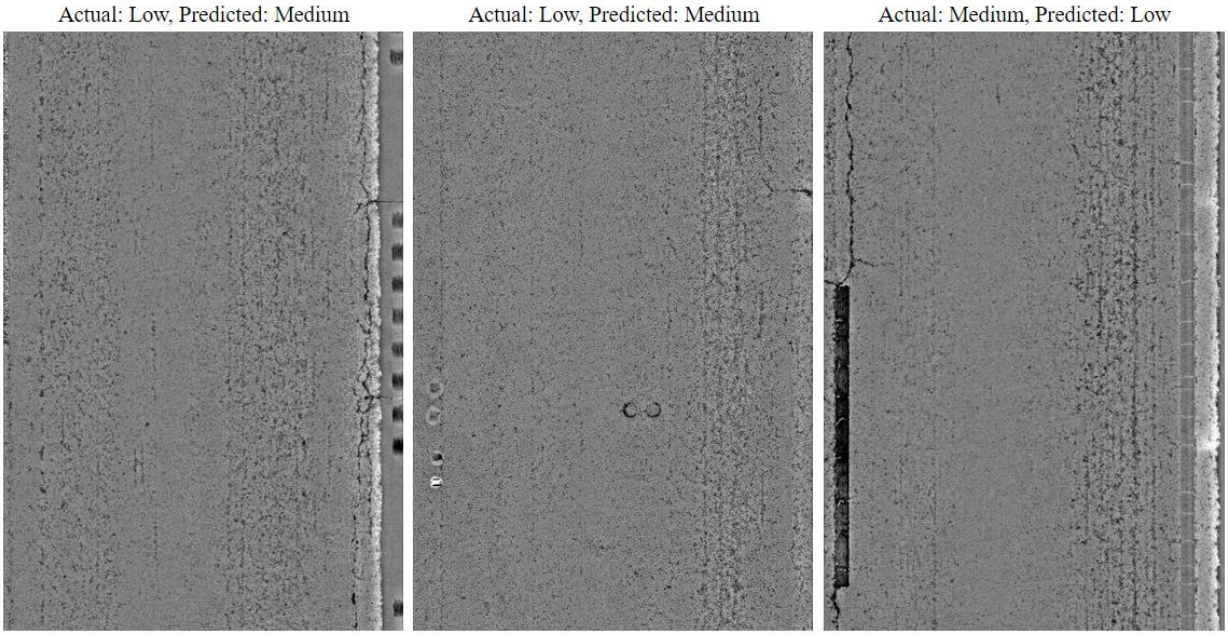


Figure 20. Challenging cases where the RF model did not generate correct predictions.

3.4 Preliminary Study on Deep Learning

This section describes preliminary studies on applying deep learning (DL) for FDOT raveling classification using Annotation Data Set 2. The same data splitting in Table 13 was used to train and evaluate the DL models.

3.4.1 Deep learning models

In this study, convolution neural networks (CNNs) were used to perform raveling classification. The major component of the CNNs is the convolutional (Conv) layer, which consists of a set of trainable kernels. The kernels of the Conv layers are similar to filters that extract features from the images. However, instead of having fixed weights that extract specific features, the kernels of the Conv layers consist of trainable weights that will be automatically adjusted to extract the most suitable features learned from the training data. Other than the Conv layers, the network also consists of pooling layers and activation functions. Here, the maximum pooling (Maxpool) layer and the rectified linear unit (ReLU) activation function were used.

Various architectures of CNNs have been developed in recent years. VGG (Simonyan and Zisserman, 2014) is an early developed network for image classification tasks. Although the VGG backbone may not achieve the current state-of-the-art performance, the VGG backbone is still popular because of its simplicity. The ResNet (He et al., 2016) largely improved the performance of CNNs. With a deeper backbone made possible by the residual blocks, the ResNet backbone can learn and extract more complex features to improve the performance. In this study, the 16-layer VGG (VGG16) and 50-layer Resnet (Resnet50) were trained and evaluated using Annotation Data Set 2 for raveling classification.

3.4.2 Training settings

To train the models, the Adam optimizer was used, with cross-entropy loss and batch size of 2 and 30 epochs. The Adam optimizer has an initial learning rate of 0.0001, and a step learning rate schedule was used, in which the learning rate was halved every 5 epochs. Early stopping was applied to select the weights of the model that obtained the best performance on the validation dataset during training. The selected weights were then used for the evaluation stage. Randomly flipping horizontally and vertically with a probability of 0.5 was used for data augmentation to prevent the models from overfitting the training data. Due to the high computation and memory requirements for training DL models, the input range images were downsized to 512 x 512 in this study.

3.4.3 Evaluation Results

Table 17 shows the accuracy of the two applied models. As expected, ResNet50 achieved a higher accuracy because of its deeper architecture and residual connections. Compared to shallow ML models in Table 14, ResNet50 did improve the accuracy, but the improvement is only around 1%. Comparing the confusion matrix of ResNet50 (Table 18) and RF (Table 16), it can be seen that the DL models have similar issues as the traditional ML models, which were the ambiguity of the adjacent severity levels and too few images with the severe level.

Future studies can be explored and designed using more advanced CNNs and finetuning techniques to further improve the performance of DL models. Some hyper-parameters that can be finetuned include the types of optimizers and loss functions, the learning rate and decay schedule, and different data augmentation methods. Oversampling techniques can also be explored to overcome the issue of having too few images with severe raveling severity during training. Furthermore, studies on advanced texture classification using deep learning can be sought and reviewed to adopt potential methods for improvement.

Table 17. The confusion matrix of RF trained with Version 2 Annotation.

Model	**Accuracy**
ResNet50	87.9%
VGG16	83.9%

Table 18. The confusion matrix of ResNet50 trained with Annotation Data Set 2.

Confusion Matrix						
Predicted / Actual	None	Low	Medium	Severe	**Recall**	**Precision**
None	145	1	0	0	0.99	0.96
Low	6	189	16	0	0.90	0.83
Medium	0	36	137	1	0.79	0.87
Severe	0	1	4	3	0.38	0.75

From the test outcomes, the best traditional ML model is RF with 86.6% accuracy and the best DL model is ResNet50 with 87.9% accuracy. Although DL has slightly better accuracy, the difference is not significant. Also, the application of DL models requires Graphics Processing Unit (GPU) resources, which further requires the installation and maintenance of necessary software, such as CUDA. During the evaluation, the DL models typically require around one minute to process the testing dataset (around 500 images) with a single NVIDIA GeForce RTX 2080Ti GPU, while the traditional ML models typically require around 10 seconds with CPU resources. In considering the above trade-off, it is recommended to use traditional ML, specifically the RF model, for the subsequent FDOT's automatic raveling classification implementation.

4. The Proposed Raveling Rating and Treatment Trigger for FDOT

In this section, the design of raveling condition rating and treatment trigger for FDOT is proposed. The recommended design rationale for the raveling rating and treatment trigger is described. Then, the procedure to set up the trigger ratings for treatment decisions is presented.

4.1 The Design Rationale

First, the Crack Rating (CR) used in FDOT's current practice was divided into three individual distress, i.e., Cracking, Raveling, and Patching, each with its own rating. Next, the raveling rating and treatment decision trigger were developed based on the concept of reverse engineering. Specifically, the treatment trigger was determined by considering the potential treatment decision. Then, the deduction table for the raveling condition was designed using the decided treatment trigger and other factors, such as raveling affected area and severity. The design presented in this report is structured to allow use of individual pavement distress deducts (e.g. raveling, cracking, etc.), reflecting individual distress condition, and the combined overall rating condition for making a more informed treatment decision. To obtain the overall deduct from the individual deducts, it is assumed that individual deducts are independent. This is the same as current FDOT practice which assume raveling area and cracking area are independent in deduct and rating computation. Therefore, the overall deduct can be obtained by either summing the individual deducts or selecting the maximum individual deduct.

4.1.1 Design the treatment decision and trigger

The proposed treatment decision and trigger table consist of two levels of criteria: the overall pavement condition deduct and the individual distress deduct. An example of the proposed treatment decision and trigger table is shown in Table 19. As shown in the table, the overall deduct serves as the first level of decision-making. Then, the individual deducts of each distress were taken into consideration to obtain the final treatment decision. Here, to make it similar to current FDOT practice, it is recommended that a deduct range of 0~10 and a rating < 6.5 (or a deduct > 3.5) be used as the main treatment trigger (i.e., pavement with > 3.5 deduct is considered deficient). The same design rationale can be flexibly applied to different rating scales (e.g., 0~100) and spatial units (e.g., 5m, 1mile, etc.) if FDOT plans to have finer rating scales in the future.

To determine the adequate treatment criteria in Table 19, the data of two projects with FC-5 only treatment provided by FDOT were utilized. The important information of these two cases is shown in Table 20. The first project is on I-75 in Alachua County (Roadway ID: 26260000), which is shown in the upper portion of Table 20. In this project, the overlay was applied in 2002 and the section was resurfaced again in 2017. This is a normal lifespan for most FC5 surfaces, so this would be considered a successful thin overlay. The second project is on I75 in Marion County (Roadway ID: 36210000), which is shown in the lower portion of Table 20. In this project, there was more extensive cracking than what was present in the Alachua County project. This overlay lasted 9-10 years so that it failed prematurely given the average life of FC5 surfaces is 14 years. The year that FC-5 only was applied is labeled in yellow in Table 20.

These cases provided the initial treatment criteria for FC-5 only in Table 19 based on the following criteria:

1. Following the overall deduct trigger (i.e., > 3.5 deduct), the raveling deduct for FC-5 treatment criteria is also set to > 3.5.
2. There should be low cracking and rutting for FC-5 only treatment. Since the FC-5 only project in Alachua County was more successful, the crack and rut ratings in that project were used. Specifically, the crack and rut deduct for FC-5 treatment criteria is set to ≤ 2.

*Note: If raveling is noted in years 2 or 3, the likelihood of an early failure due to raveling is high and the section should be closely monitored. In this case, the FC-5 only treatment can potentially be applied after close assessment, even if the raveling deduct is not greater than 4.

Table 19. An example of the treatment decision table.

Overall Deduct	Individual Deduct		Treatment
	Factor	Criteria	
≤ 3.5	-	-	No treatment
> 3.5	Raveling	> 3.5*	FC5 only
	Cracking	≤ 2	
	Rutting	≤ 2	
	...	...	
	Raveling	> 3.5	Resurfacing (with deeper milling depth than FC5 only)
	Cracking	> 2	
	Rutting	> 2	
	...	...	

* Please see the additional note on this criterion above the table.

Table 20. Two projects with FC-5 only treatment in FDOT.

Data Tested	Roadway ID	Crack Rating	Rut Rating	Ride Rating	Raveling	Pavement Age
1/11/2001	26260000	9.5	8	8.2	M1	7
1/10/2002	26260000	10	10	8.3	NULL	1
1/10/2003	26260000	10	10	8.4	NULL	2
1/7/2004	26260000	10	10	8.5	NULL	3
1/11/2005	26260000	10	10	8.3	NULL	4
1/10/2006	26260000	9.5	9	8.4	NULL	5
1/11/2007	26260000	8.5	10	8.3	NULL	6
1/9/2008	26260000	8.5	10	8.2	NULL	7
9/15/2009	26260000	8.5	9	8.3	NULL	8
8/18/2010	26260000	8.5	9	8.1	L1	9
8/16/2011	26260000	8.5	9	7.9	L1	10
1/12/2004	36210000	1	8	7.8	M2	NULL
1/12/2005	36210000	10	10	8.3	NULL	1
1/12/2006	36210000	10	9	8.2	NULL	2
1/12/2007	36210000	10	10	8.3	NULL	3
1/12/2008	36210000	9	9	8	NULL	4
12/1/2009	36210000	8	9	8.2	NULL	5
11/8/2010	36210000	7.5	9	8.1	L1	6
11/28/2011	36210000	7.5	8	7.9	S1	7
12/10/2012	36210000	6.5	9	7.9	S2	8
12/9/2013	36210000	6.5	8	7.7	S2	9
12/2/2014	36210000	4.5	8	7.5	S3	10

4.1.2 Design the deduct tables

For each distress, an individual deduct table can be set up based on its extent, severity (if applicable), and the trigger point designed in the treatment decision table. This is similar to the deduct table of Crack Rating in FDOT's current practice. An example of the deduct table for raveling is shown in Table 21. In this initial deduct table, the treatment trigger designed in the previous step (i.e., raveling deduct $>$ 3.5) is used with the cases shown in Table 20 to design the deduct values. The design rationale is as follows:

1. Low raveling with greater than 50% affected area should be considered as deficient. Therefore, a deduct of 4 is assigned to low raveling with 51%+ area.
2. For the project in Alachua County (Table 20), the raveling severity was M1 when FC-5 only was applied. Following this case, a deduct of 4 is also assigned in medium raveling with 1%~5% area.

3. In the deduct table, higher affected area and severity should both lead to a higher deduct value. Here, one deduct difference for one level higher in the affected area and three deduct difference for one level higher in the severity were applied.

Table 21. An example of the deduct table for raveling.

Percentage of the affected area	Raveling Severity		
	Low	Medium	Severe
01 ~ 05	1	**4**	7
06 ~ 25	2	5	8
26 ~ 50	3	6	9
51+	**4**	7	10

4.2 Refine the trigger ratings for treatment decision

Although the initial treatment criteria and deduct values were established based on the two cases of FC-5 only treatment and engineering judgement, it is recommended that more cases (both successful and unsuccessful ones) be studied for determining the more suitable thresholds for applying FC-5 only resurfacing. The following steps can be performed in future work to identify more reasonable treatment criteria and deduct values:

1. Identify several cases with the following different treatment decisions from FDOT:
    a. Isolated treatment (e.g., quick patch for safety reasons) when the extent of raveling is small but quite severe, or the raveling is not severe but has a large extent without other major distresses such as cracking and rutting.
    b. FC-5 resurfacing only where raveling is present but without other major distresses such as cracking and rutting.
    c. Traditional resurfacing where there are both raveling and other distress types.
2. Then, study the pavement distress conditions on these cases, including:
    a. Severity level
    b. Affected area
    c. Duration (i.e., how long does raveling occur)
3. Finally, establish the treatment decision and deduct trigger based on the above factors.

4.3 Design the spatial unit for pavement condition survey

Instead of using the current pavement condition survey unit the same as the project treatment unit (say, 10 miles), it is recommended that a finer pavement condition survey unit be used. With the advancement of sensor technology, FDOT can collect 2D/3D pavement images at a 15-ft (5m) interval. Thus, this 15ft value is recommended as the fundamental pavement condition survey spatial unit. All distresses would be automatically and semi-automatically extracted at this level. They can then be aggregated to different spatial units (e.g., 100-ft, 0.1 mile and 1 mile) to support different pavement maintenance and rehabilitation needs. For example, this 15-ft fundamental level pavement condition survey data can be mapped out to determine the appropriate coring locations for determining the appropriate milling depth. It can also be used to determine the optimal

maintenance and rehabilitation sections with homogeneous pavement conditions. For example, one may find one 200-ft section with poor condition requiring a deep patch while the remaining one mile is still in a good condition. Similarly, the pavement sections with raveling problems can be segmented for application of FC-5 resurfacing only to save maintenance and rehabilitation cost for FDOT. There are many other potential opportunities to take advantage of this high-resolution pavement condition survey data to optimize maintenance and rehabilitation for FDOT in the future.

5. Conclusions and Recommendations

This research first conducted a literature review on practices of pavement rating computation and treatment decisions in FDOT and other State DOTs with a special focus on raveling. The following were identified conclusions and research needs:

1. In FDOT, a rating system separating cracking and raveling is needed, which can proactively target the raveling treatment needs before rapid deterioration of OGFC sets in. Currently, CR < 6.5 is the primary threshold to trigger treatment in FDOT's practice. However, including raveling and patching in CR makes the maintenance and rehabilitation decision difficult. Therefore, new treatment decision thresholds separating cracking and raveling are also needed. In this case, more cost-effective treatments (e.g., FC-5 only treatment) can be applied.
2. The survey and classification of raveling in major transportation agencies are conducted manually through in-field visual inspection methods, which are error-prone, time-consuming, and labor-intensive. Therefore, there is a need to develop an automated raveling detection and classification method to overcome these problems.
3. The appearance of raveling on digital images is susceptible to ambient lighting conditions. To overcome this issue, using 3D pavement data is a better alternative for capturing pavement surface texture. A more reliable automated raveling detection and classification method can be developed using the 3D pavement data.
4. A fixed raveling survey spatial unit is not defined in some DOTs (e.g., FDOT, TxDOT). In this case, the pavement is rated based on a whole section, which is usually too large to obtain precise localization information for raveling. Currently, FDOT's 3D pavement image size is 12ft (4m) by 15ft (5m) in transverse and longitudinal directions respectively. With detailed 3D pavement data, fixed and multi-scale survey units (e.g., 15ft/5m, 100ft, 0.1 miles or 1 mile, etc.) can be generated with proper data aggregation methods. This will help DOTs to react to early raveling more quickly and apply localized treatments.

Next, a literature review on automated raveling detection and classification methods was also conducted. The results were as follows:

1. Nontrainable methods for automated raveling detection and classification have several limitations. First, many methods are only capable of detecting raveling but not its severity level classification. Second, many of the indicators require certain assumptions about the pavement texture. Third, the validation of the methods was very limited and not systematic. Finally, many methods require frequent parameter tuning and adjustment based on empirical experiments.
2. By utilizing ML techniques to train a more robust model using real-world 3D pavement data that performs both raveling detection and classification, Tsai & Wang (2015) and Tsai et. al (2021) overcame most of the problems in nontrainable methods. This method is based on traditional ML models, which require a predefined feature extraction stage to reduce the complexity of the data.

Based on the literature review, a feasibility study on using ML models (SVC, RF, AdaBoost) for automated raveling severity classification on FDOT's pavement data was conducted. Based on the feasibility study, the following conclusions were developed:

1. It was found that the annotation quality issue in Annotation Data Set 1 caused degraded performance in ML models. Therefore, Annotation Data Set 2 was developed based on a revised annotation practice that can better ensure the annotation quality of raveling ratings. The analysis showed that all ML models achieved a better performance by training with Annotation Data Set 2, which indicates that the quality of annotation on raveling severity is critical for improving the performance of ML models.
2. With the provided quantized range images and around 8.5 miles of data, the outcomes of the study show that the ML models are feasible for per-image raveling classification on FDOT's pavement data with suitable ML classifiers. RF has the best potential to be used for implementation in the future given its high accuracy achieved in the feasibility study. The RF classifier achieved a testing accuracy of 86.6% using Annotation Data Set 2. Future studies need to include more data with the severe classification level or use over-sampling techniques to improve the model's performance on raveling with severe severity.
3. A preliminary study was also conducted on using DL for raveling severity classification on FDOT's pavement data. Two popular CNNs were used in this study, which were the VGG16 and ResNet50 models. Although DL has slightly better accuracy (ResNet50 achieved 87.9% accuracy), the difference was not significant. Also, the application of DL models requires GPU resources and has a longer inference time. Therefore, it is recommended to use traditional ML, specifically the RF model, for subsequent implementation.

Based on the technology review and ML feasibility study, the following were the preliminary recommendations for FDOT:

1. For the data collection devices, it is recommended to use the 3D line laser imaging system, that FDOT already has, to collect high-resolution pavement surface data to leverage the already collected 3D pavement data collected by FDOT. FDOT can also add value to these already collected 3D pavement data to cost-effectively extract raveling. 3D pavement data have the advantage of better capturing pavement surface texture because they are independent of ambient lighting conditions and can be accurately collected at highway speed. That is, a more reliable automated raveling detection and classification method can be developed using the 3D pavement data. Therefore, the resources can be optimized by utilizing the 3D pavement data that FDOT already collected using the Laser Crack Measurement System (LCMS).
2. The feasibility study of ML models shows promising outcomes (86.6% testing accuracy in Annotation Data Set 2) in automated raveling classification on 3D pavement data. It is recommended to use the RF model for future implementation given its high accuracy achieved in the feasibility study.
3. With automated raveling severity classification using FDOT 3D pavement images, FDOT will be able to cost-effectively identify and locate raveling with a much better spatial resolution (e.g., 15-ft, 100-ft, 0.1 mile). This will enable FDOT to cost-effectively identify the roadway sections for resurfacing by replacing only the FC-5 to save substantial maintenance and rehabilitation cost.

In addition, the following directions were recommended for further improving the ML models and for future implementation:

1. The design and selection of pavement texture features are important for improving the accuracy of automated raveling classification models. Additional features, such as Gray Level Co-occurrence Matrix (GLCM) (Nanni et al., 2013), can be explored in the future. In addition, studying how each feature contributes to the prediction of raveling severities in ML classifiers could be a worthwhile future work.
2. Currently, the ML model predicts the predominant raveling severity for each range image. However, multiple severities of raveling can occur on a range image at the same time. The information for all presented raveling severities on the same 15-ft-long image could be used to modify and potentially improve the current models.
3. For future implementation, the following items are required:
    a. Currently, the ML models are trained and evaluated with a small set of data (8.5 miles) for research purposes. To better validate the feasibility of the ML models for implementation, a larger-scale validation dataset with diverse conditions will be required.
    b. In the feasibility study, Python libraries and codes already available were used for research purposes only. This tool is efficient and sufficient for this feasibility study to choose the right ML model. However, for future implementation, suitable programming language (e.g., C++), libraries, and code, that can be used for practical software development and deployment need to be explored.
    c. A spatial smoothing and clustering algorithm is required. After the ML models predict the raveling severity of each 15 ft 3D pavement image, the spatial continuity of the distribution of raveling should be considered to remove erroneous predictions or trigger a manual review for the quality control.
    d. Software developed for implementation, which consists of data preparation, processing, raveling prediction, spatial smoothing procedures, the QA/QC tool, and the reporting tool. The QA/QC tool is used to streamline the validation and correction of automatic raveling classification outcomes. In addition, how these processes should be integrated into FDOT's overall data analysis workflow need to be further considered.
    e. A function, using automatic raveling detection and classification outcomes along with other pavement distress conditions (e.g., cracking, rutting, etc.), to identify and segment the roadway sections that the FC-5-only resurfacing can be applied to save substantial maintenance and rehabilitation money for FDOT.

Finally, in this research, the authors have proposed raveling treatment criteria and condition ratings for FDOT. The proposed treatment decision table consists of two levels of criteria: the overall pavement condition deduct and the individual distress deduct (such as raveling, cracking, rutting, etc.). The initial treatment criteria in the treatment decision table (Table *19*) were designed based on the current rating practices in FDOT and the case study on two FC-5-only projects. Then, the initial raveling deduct table is proposed (Table *21*). The initial deduct values in the table are proposed based on the designed treatment criteria, the affected area and severity level of raveling, and the two FC-5-only projects. Although the initial treatment criteria and deduct values of raveling have been established based on two cases of FC-5 only project and engineering judgment, it is recommended that more cases (both successful and unsuccessful ones) be studied for determining more suitable thresholds for applying FC-5 only resurfacing.

References

ALDOT. (2017). *Level of Service Condition Assessment: Data Collection Manual*, Alabama Department of Transportation.

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. *Electronics*, *8*(3), 292.

Caltrans (2017). *Automated Pavement Condition Survey Manual*, California Department of Transportation.

Chang, C., Saenz, D., Nazarian, S., Abdallah, I. N., Wimsatt, A., Freeman, T., & Fernando, E. G. (2014). *TXDOT guidelines to assign PMIS treatment levels* (No. 0-6673-P1). Texas. Dept. of Transportation. Research and Technology Implementation Office.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

FDOT (2017) *Flexible Pavement Condition Survey Handbook*, Florida Department of Transportation.

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, *55*(1), 119-139.

Gadsby, A. & Tsai, Y. (2020) Environmental Impact of New Micro-Milling and Thin Overlay and Conventional Milling for Sustainable Pavement Preservation, *International Journal of Pavement Research and Technology*. https://doi.org/10.1007/s42947-020-0029-9

GDOT (2017) *Pavement Condition Evaluation System Manual*, Georgia Department of Transportation.

Geary, G. M., Tsai, Y., & Wu, Y. (2018). An Area-Based Faulting Measurement Method Using Three-Dimensional Pavement Data. *Transportation Research Record*, *2672*(40), 41-49.

Hadjidemetriou, G. M., & Christodoulou, S. E. (2019). Vision-and entropy-based detection of distressed areas for integrated pavement condition assessment. *Journal of Computing in Civil Engineering*, *33*(3), 04019020.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE transactions on pattern analysis and machine intelligence*, *20*(8), 832-844.

Hoang, N. D. (2018). An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable filter-based feature extraction. *Advances in Civil Engineering*, vol. 2018, Article ID 7419058, 2018.

Hoang, N. D. (2019). Automatic detection of asphalt pavement raveling using image texture based feature extraction and stochastic gradient descent logistic regression. *Automation in Construction*, *105*, 102843.

Hsieh, Y. A., & Tsai, Y. J. (2020a). Machine Learning for Crack Detection: Review and Model Performance Comparison. *Journal of Computing in Civil Engineering*, *34*(5), 04020038.

Hsieh, Y. A., & Tsai, Y. J. (2020b). Automated Asphalt Pavement Raveling Detection and Classification Using Convolutional Neural Network and Macrotexture Analysis. Accepted for presentation at the Transportation Research Board (TRB) 2021 Annual Meeting.

Jiang, C., & Tsai, Y. J. (2016). Enhanced crack segmentation algorithm using 3D pavement data. *Journal of Computing in Civil Engineering*,*30*(3), 04015050.

Jiang, C., Tsai, Y., & Wang, Z. (2016). Use of three-dimensional pavement surface data to analyze crack deterioration: Pilot study on Georgia State Route 26. *Transportation Research Record*,*2589* (1), 154-161.

Lai, J., M. Hines, S., Wu, P.Y., & Jared, D. (2012). Pavement Preservation with Micromilling in Georgia – Follow-Up Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2292, pp. 81-87

Laurent, J., Hébert, J. F., Lefebvre, D., & Savard, Y. (2012a). Using 3D laser profiling sensors for the automated measurement of road surface conditions. In *7th RILEM international conference on cracking in pavements* (pp. 157-167). Springer, Dordrecht.

Laurent, J., Hébert, J. F., Lefebvre, D., & Savard, Y. (2012b). High-speed network level road texture evaluation using 1mm resolution transverse 3D profiling sensors using a digital sand patch model. In *Proceedings of the 7th International Conference on Maintenance and Rehabilitation of Pavements and Technological Control, Auckland, New Zealand* (pp. 28-30).

Li, B., Wang, K. C., Zhang, A., Yang, E., & Wang, G. (2020). Automatic classification of pavement crack using deep convolutional neural network. *International Journal of Pavement Engineering*,*21*(4), 457-463.

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018). Road damage detection and classification using deep neural networks with smartphone images. *Computer-Aided Civil and Infrastructure Engineering*, *33*(12), 1127-1141.

Mathavan, S., Rahman, M. M., Stonecliffe-Jones, M., & Kamal, K. (2014). Pavement raveling detection and measurement from synchronized intensity and range images. *Transportation Research Record*, *2457*(1), 3-11.

McRobbie, S. & G. Furness (2008). *Automated Detection of Fretting on HRA Surfaces*, TRL Published Project Report 229, Transport Research Laboratory, Wokingham, Berks, U.K., 16 pp.

McRobbie, S., J. Iaquinta, A. Wright, P. Trumper, & J. Kennedy (2012). *Development and Validation of Algorithms for the Automatic Detection of Fretting Based on Multiple Line Texture Data*, Research into Pavement Surface Disintegration. Phase 2 Interim Report, Transport Research Laboratory, Wokingham, Berks, U.K.

McRobbie, S., C. Wallbank, K. Nesnas, & A. Wright (2015). *Use of High-Resolution 3-D Surface Data to Monitor Change over Time on Pavement Surfaces*, Research into Pavement Surface Disintegration. Phase 3, Transport Research Laboratory, Wokingham, Berks, U.K.

Nanni, L., Brahnam, S., Ghidoni, S., Menegatti, E., & Barrier, T. (2013). *Different approaches for extracting information from the co-occurrence matrix*. PloS one, 8(12), e83554.

ODOT (2019) *Pavement Data Condition Manual*, Oregon Department of Transportation.

Scott, P., K. Radband, M. Zohrabi, P. Sanders, S. McRobbie, & A. Wright (2008). Measuring Surface Disintegration (Raveling or Fretting) Using Traffic Speed Condition Surveys, *7th International Conference on Managing Pavement Assets*, Calgary, AB, Canada, June 23–28, 2008.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.*

Tsai, Y. C., & Chatterjee, A. (2018). Pothole detection and classification using 3D technology and watershed method. *Journal of Computing in Civil Engineering*,*32*(2), 04017078.

Tsai, Y. C., Jiang, C., & Huang, Y. (2014a). Multiscale crack fundamental element model for real-world pavement crack classification. *Journal of Computing in Civil Engineering*,*28*(4), 04014012.

Tsai, Y. C. J., & Li, F. (2012). Critical assessment of detecting asphalt pavement cracks under different lighting and low intensity contrast conditions using emerging 3D laser technology. *Journal of Transportation Engineering*,*138*(5), 649-656.

Tsai, Y. J., Li, F., & Wu, Y. (2013). A new rutting measurement method using emerging 3D line-laser-imaging system. *International Journal of Pavement Research and Technology*,*6*(5), 667.

Tsai, Y. J., & Wang, Z. (2015). *Development of an asphalt pavement raveling detection algorithm using emerging 3D laser technology and macrotexture analysis* (No. NCHRP IDEA Project 163).

Tsai, J. Y., Wang, Z., & Gadsby, A. (2019). *Evaluation of the Long-Term Performance and Benefit of Using an Enhanced Micro-Milling Resurfacing Method [GDOT 13-20]* (No. FHWA-GA-18-1320). Georgia. Department of Transportation. Office of Performance-Based Management & Research.

Tsai, J. Y. C., Wang, Z. H., & Li, F. (2015). Assessment of rut depth measurement accuracy of point-based rut bar systems using emerging 3d line laser imaging technology. *Journal of Marine Science and Technology*, *23*(3), 322-330.

Tsai, Y. J., Wu, Y., & Ai, C. (2011, January). Feasibility study of measuring concrete joint faulting using 3d continuous pavement profile data 2. In *Proceedings of the Transportation Research Board 90th Annual Meeting, Washington, DC, USA* (pp. 23-27).

Tsai, Y., Wu, Y., Gadsby*, A., Hanes, S. (2016). Critical Assessment of the Long-term Performance and Cost-effectiveness of a New Pavement Preservation Method: Micro-Milling and Thin Overlay Journal of The Transportation Research Record, National Academy of Sciences, Washington, DC., 2016 (2550): 8-14.

Tsai, Y., Wu, Y. Geary, G. (2018). Sustainable and Cost-Effective Pavement Preservation Method: Micro-Milling and Thin Overlay, Journal of Transportation Engineering, Part A: Systems, Volume 144, Issue 10, October 2018.

Tsai, Y, Wu, Y., Lai, J. S. (2012) Validation of RVD-Based Micro-milled Pavement Surface Texture Quality Control, Draft final report, Georgia Department of Transportation.

Tsai, Y., Wu, Y., Lai, J., Geary, G. (2012) Characterizing Micro-milled Pavement Textures Using RVD for Super-thin Resurfacing on I-95 Using A Road Profiler, Journal of The Transportation Research Record, No.2306, pp.144-150.

Tsai, Y., Wu, Y., and Lewis*, Z. (2014b). Full-Lane Coverage Micromilling Pavement-Surface Quality Control Using Emerging 3D Line Laser Imaging Technology. Journal of Transportation Engineering. Volume 140 Issue 2.

Tsai, Y. C. J., Zhao, Y., Pop-Stefanov, B., & Chatterjee, A. (2021). Automatically detect and classify asphalt pavement raveling severity using 3D technology and machine learning. *International Journal of Pavement Research and Technology*, *14*(4), 487-495.

TxDOT (2015). *Pavement Management Information System: Rater's Manual*, Texas Department of Transportation.

Van Ooijen, W., Van den Bol, M., & Bouman, F. (2004). High-speed measurement of raveling on porous asphalt. In *Symposium on Pavement Surface Characteristics [of Roads and Airports], 5th, 2004, Toronto, Ontario, Canada*.

Wang, S., Qiu, S., Wang, W., Xiao, D., & Wang, K. C. (2017). Cracking classification using minimum rectangular cover–based support vector machine. *Journal of Computing in Civil Engineering*,*31*(5), 04017027.

Wang, Z., & Pyle, T. (2019). Implementing a pavement management system: The Caltrans experience. *International Journal of Transportation Science and Technology*, *8*(3), 251-262.

Zhang, A., Wang, K. C., Fei, Y., Liu, Y., Chen, C., Yang, G., ... & Qiu, S. (2019). Automated pixel-level pavement crack detection on 3D asphalt surfaces with a recurrent neural network. *Computer-Aided Civil and Infrastructure Engineering*,*34*(3), 213-229.

Zhang, A., Wang, K. C., Li, B., Yang, E., Dai, X., Peng, Y., ... & Chen, C. (2017). Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering*,*32* (10), 805-819.

Zimmerman, K. A. (2017). *Pavement Management Systems: Putting Data to Work* (No. Project 20-05, Topic 47-08).