An Automated 3D Crack Severity Assessment Using Surface Data for Improving Flexible Pavement Maintenance Strategies

Li, Zhe; Eskandari Torbaghan, Mehran; Zhang, Tuo; Qin, Xia; Li, Wenda; Li, Yongjian; Zhang, Jiupeng

DOI: 10.1109/TITS.2024.3379997

License: Other (please specify with Rights Statement)

Document Version
Peer reviewed version

Citation for published version (Harvard):

Publisher Rights Statement:

© 2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

General rights
Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

• Users may freely distribute the URL that is used to identify this publication.
• Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
• Users may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
• Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy
While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.
An Automated 3D Crack Severity Assessment Using Surface Data for Improving Flexible Pavement Maintenance Strategies

Zhe Li, Mehran Eskandari Torbaghan, Tuo Zhang, Xia Qin, Wenda Li, Yongjian Li, Jiupeng Zhang

Abstract—Evaluation of crack severity in flexible pavements predominantly centers around the analysis of cracks’ surface characteristics. However, this study highlights the critical importance of 3D crack parameters, including volume and depth, for comprehensive assessment. The objective here is to develop an autonomous crack severity assessment, by predicting the vertical parameters of cracks exclusively from their surface properties. To achieve this, a dataset of 3D parameters comprising 200 cracks from eight flexible pavements was acquired, and both linear and nonlinear correlations were conducted among these 3D parameters. Subsequently, five single-output and one multi-output machine learning models were developed to explore the potential of utilizing surface parameters to predict the vertical parameters of cracks. The outcomes validated the effectiveness of two specific methods, namely, Artificial Neural Network and Extreme Gradient Boosting models, in predicting crack volume based on surface parameters, with R2 scores of 0.832 and 0.748, respectively. Additionally, the multi-output machine learning model we developed achieved classification prediction of the crack damage penetration depth using surface parameters, yielding optimal precision, recall, and F1 scores of 0.790, 0.779, and 0.761, respectively. This study has introduced a crack damage evaluation index, based on a 3D assessment, that relates crack depth classification to severity. We provide suggestions that could pave the way for informed decision-making on maintenance strategies that could be adopted to extend asset life cycle.

Index Terms—flexible pavement, pavement cracks, severity assessment, machine learning, crack 3D parameters, crack depth, crack volume.

I. INTRODUCTION

CRACKING, as a prevalent and prominent form of distress in flexible pavements, significantly impacts pavement structure and performance [1]. As part of the common engineering practices, periodic pavement condition assessments are conducted to evaluate the severity of cracking. Appropriate maintenance scheme and strategies are then devised to preserve the structural and functional integrity of the pavement [1].

The characterization of individual cracks in flexible pavement is typically based on parameters such as cracks’ spatial positioning, length, orientation, morphology, and connectivity with neighboring cracks [2], [3], [4, p. 227], [5]–[8]. Subsequently, distinct thresholds have been established by various road agencies, outlined in Table I, to categorize the severity levels of various crack types.

It should be noted that the criteria outlined in Table I are expressly designed to quantify and assess the severity of cracks with a primary focus on pavement surface damage. Nonetheless, the formation of flexible pavement cracks is a multifaceted, three-dimensional (3D) phenomenon influenced by climatic conditions and traffic loads. These diverse 3D crack indicators exert distinct influences on the holistic urban mobility ecosystem encompassing road user, vehicle, and pavement structural and functional performance [9], [10]. While surface/spatial characteristics of cracks (i.e., width, length, and area) exert a direct influence on driving comfort, vehicle’s vibration, and noise levels. Crack depth is tied to pavement’s structural integrity, water resistance, and its associated maintenance costs [11]. The volume of cracks varies contingent upon the propagation of each individual crack. Even for cracks with equivalent volume, they may manifest in diverse shapes, ranging from broad and shallow to narrow and deep, hence with different negative impacts.

The existing methods for measuring the depth of pavement crack can be divided into two classes: destructive testing (DT) [12] and nondestructive testing (NDT) [13]. DT, which includes core sampling, test pits, and dynamic cone penetrometer testing, is time- and resource-consuming, which it damages the road. Besides, it is challenging to determine the appropriate location and extent of the coring. NDT can be categorized into surface contact methods, including those utilizing impact echo or ultrasound approaches, and non-
contact methods, which involve vision-based techniques (e.g., Area Array Camera, Stereo Vision, Structure-from-Motion etc.), optical thermographic techniques [15], laser-based techniques, and electromagnetic-based techniques (e.g., Ground-Penetrating Radar - GPR).

As pavement cracks are often obstructed by soil, gravel, vegetation, and other debris, contact-based measurement methods encounter limitations in their ability to penetrate this debris, e.g., vision-based techniques [16]–[21], optical thermographic techniques [15], [22], [23], and laser-based methods [24], [25]. In contrast, electromagnetic waves [13], [26] possess the capability to penetrate the structural layers of the pavement to a certain depth. Upon encountering boundaries between distinct media, these waves reflect back, and the resultant electromagnetic signals are captured by a receiving device. These signals encapsulate valuable information pertaining to the internal composition of the pavement. Nevertheless, its reliance on expert experience-driven and/or post-processing of the acquired signals with potential bias/error, should be noted as a drawback.

Presently, crack volume measurement methods are categorized into automated and manual approaches. Automated measurements predominantly hinge upon NDT for 3D reconstruction. While these methods hold significant promise for acquiring 3D crack data, they are not devoid of limitations. For instance, Feng et al. [24] employed laser scanning technology to acquire a point cloud dataset of cracks, and the real crack volume was automatically calculated. However, this method is limited by its inability to penetrate dust and debris within the cracks, resulting in incomplete data about the bottom of the cracks.

Similarly, Li et al. [21] employed a binocular 3D smart scanner to acquire the pavement cracks data including crack volume. Nonetheless, their approach primarily focuses on volumes from uncleansed cracks and overlooks structural details beneath the debris layer. To address the consideration of deeper crack damage, Tong et al. [27] utilized a convolutional neural network to extract the crack longitudinal features from GPR data. They subsequently executed a 3D reconstruction of the cracks and computed the volumes for ten representative cracks. Regrettably, the specific methodology for volume calculation remained undisclosed in the paper, and the robustness and precision of this approach were not validated.

In contrast to automated measurement techniques, manual methods, particularly the sand-filling method [28], [29], offer superior precision in replicating the authentic crack conditions through meticulous crack cleaning and sand-filling procedures, consequently enabling accurate crack volume determinations. Furthermore, the sand-filling method provides control and reproducibility, rendering it as a well-established and extensively validated approach in engineering applications. However, it is essential to highlight that a notable drawback of the sand-filling method lies in the time-consuming and labor-intensive nature of in-situ measurements.

Measurements of crack depth and volume in flexible pavements serve a dual purpose, by enabling comprehensive assessment of crack severity while also facilitating the exploration of their interrelation with surface characteristics. In related literature, for instance, Lu et al. [12] have demonstrated the feasibility of employing an Artificial Neural Network (ANN) model to statistically estimate crack depth. Their model incorporates a range of input data, including crack width, slopes, measurable depth, pavement attributes (type, age, materials), roadway function, and average daily traffic data. The accuracy of the model was substantiated through comparisons with measurements obtained using a contact-type NDT system, though the study did not provide specific metrics for model prediction accuracy and evaluation criteria.

Furthermore, other investigations have explored certain 3D parameters of cracks but have not delved into modeling their interrelationships. For instance, Lee et al. [30]’s study involved 14 measurement sets of crack type, length, width, depth, and overburden thickness on the pavement surface. Additionally, Zhang et al. [31] developed a Kinect-based approach for capturing crack characteristics, including length, width, and surface depth.

Compared to previous work presented in the literature, the main contributions of this paper are as follows:

- Field-based experimentation and data pre-processing are employed to obtain 3D parameters encompassing width, length, area, volume, and depth of pavement cracks.
- A variety of machine learning (ML) models for both regression prediction of crack volume and classification of crack damage penetration depth are used. Importantly, the developed methodology ensures that both volume and depth can be automatically estimated, even in cases where solely surface parameters are available.
An equation for quantifying the extent of crack damage based on the penetration depth of the cracks is proposed.

A set of recommendations for informing pavement maintenance decision-making based on the distribution of crack depth is provided.

The rest of the paper is organized as follows: Section II discusses the employed methodology, including data acquisition and pre-processing. Section III elucidates the Pearson correlation analysis conducted among all individual parameters. Section IV comprehensively details the machine learning model developed for this study and expounds upon its theoretical foundations. Section V is dedicated to the prediction results and a thorough discussion of the model, alongside an exploration of how the research findings are applicable to the assessment of crack severity and the decision-making process in maintenance, and Section VI concludes findings of this study.

II. METHODOLOGY

In this research, a comprehensive analysis was conducted on eight flexible pavement sections, encompassing a total of 200 cracks, with data acquisition carried out through field-based 3D measurements, followed by post-processing the data [32]. In addition to classifying the cracks into types (including transverse, longitudinal, and net-shaped categories), a dataset, as shown in Table II, was meticulously gathered for each individual crack.

<table>
<thead>
<tr>
<th>No of readings</th>
<th>parameter</th>
<th>No of readings</th>
<th>parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>length</td>
<td>1</td>
<td>average visible depth</td>
</tr>
<tr>
<td>10</td>
<td>width</td>
<td>1</td>
<td>volume</td>
</tr>
<tr>
<td>1</td>
<td>average width</td>
<td>5</td>
<td>damage penetration depth</td>
</tr>
<tr>
<td>1</td>
<td>area</td>
<td>1</td>
<td>penetration depth</td>
</tr>
<tr>
<td>10</td>
<td>visible depth</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A new parameter was measured in this paper called ‘damage penetration depth’, defined as the deepest point affected by the crack’s perpendicular propagation. Subsequently, a Pearson correlation analysis was employed to investigate potential linear relationship among the variables of area, length, average width, volume, and the average damage penetration depth across the observed cracks.

To assess the impact and magnitude of various easily measurable surface parameters on the accuracy of volumetric and visible depth prediction, alongside exploring the potential of utilizing easily measurable indicators for predicting volumetric and visible depth parameters, five single-output ML and one multi-output ML models were tested across discrete input features. These features encompassed diverse configurations, including singular crack width, coupled dimensions of crack width and length, as well as multifaceted parameters like crack width, length, and area combined.

In addition, aiming to predict crack damage penetration depth, its relationships with various crack parameters were explored in the study. Then, crack attributes including type, width, length, area, volume, and visible depth were combined into three input feature sets. For the output feature, crack damage penetration depth was categorized into three groups: surfacing layer, asphalt layer, and base layer. Due to limited base layer data (just 4% of the dataset), it was merged with the asphalt layer to create an ‘original layer’. Moreover, to enhance model performance, the crack damage penetration depth was simplified to a binary data. The multi-output ANN model was then trained using these binary classifications.

Among them, the dataset in all ML models was subjected to random division in an 8:2 ratio, allocating 80% for the training set and 20% for the testing set. Furthermore, a four-fold cross-validation technique was systematically employed during the training phase. Fig. 1 provided a flowchart summarizing the developed methodology.

A. Data Collection

To acquire the tridimensional crack parameters, a series of experiments were conducted on eight campus roads at the University of Birmingham. The precise procedural delineation of these experiments is encapsulated in Fig.2.

Cracks cleaning, to acquire more accurate measurements, was conducted by removing any vegetation, soil, gravel, and
debris from the gap using screwdriver, steel brush, soft brush, and air compressor. Then, each crack was divided into 10 equal parts and caliper was used to measure the visible depth of the crack, using the method provided in [2]. Later, a GPR antenna, with 2.2GHz frequency, was used perpendicular to the crack at five locations along the crack to determine the pavement subsurface structural damage. In the next step, ultra-fine silica sand, kiln dried with a grain size of approximately 0.1-0.3 mm and a beige color, was filled into the cracks. The volume of filling sand was recorded with measuring cylinder (unit scale 1 ml). Simultaneously, the width at identical locations with visible depth was measured using a caliper. Finally, photos from the sand-filling cracks were captured by a camera boasting a resolution of $3024 \times 3024$ pixels mounted on a tripod with a fixed height of 107.5 cm.

B. Data Processing

The raw crack data relevant to width, visible depth, and volume acquired by caliper and measuring cylinder were directly used. However, the GPR data relevant to crack damage penetration depth information, and the images, to determine crack’s area and length, were needed to be further processed, which are explained in the following subsections.

1) Crack Area Calculation

Area of cracks were calculated using Equation (1) [33]:

$$A = k \times N$$
$$N = \sum_{x,y} B(x,y)$$

(1)

In which, $A$ denotes the total area of a crack, $k$ denotes the real area for each pixel per image, in this experiment, the $k$ constant equals to 0.355mm, $N$ denotes the number of non-zero pixels of a crack, and $B(x, y)$ denotes the pixel value in the binary image [34].

To mitigate measurement errors resulting from automatic image segmentation and accurately calculate the variant $n$ in the Equation (1), in this study, manual segmentation was employed. Pixel-level annotation of dataset images was performed using the open-source LabelMe tool [35] for semantic annotation. The information in the .json file was converted into ground truths and label visualizations. Then, the cracks on the label visualization images were extracted by image binarization, accomplished by establishing a threshold as shown in Equation (2).

$$B(x, y) = \begin{cases} 
1, & \text{if } R(x, y) > T \\
0, & \text{otherwise} 
\end{cases}$$

(2)

In this context, $B(x, y)$ denotes the pixel values within the binarized image, while $R(x, y)$ represents the pixel values of the crack channels in the label visualization image. The variable $T$ signifies the selected threshold. Examples of manually labelled images, ground truths, visualized multinomial masks, binary images, and pixel numbers are presented in Fig. 3.

2) Crack Length Calculation

Medial axis transform algorithm [36] was initially used to extract the skeleton of the cracks. Let $\Omega$ represent the boundary of the closed shape $S$. Let $v(x, y, r)$ denote the circle with centre coordinates $(x, y)$ and radius $r$. The medial axis $M$ is the ensemble of circle centres in $S$. The equation is as follows in 2D images:

$$M : \{v(x, y, r) \mid v(x, y, r) \subseteq \Omega, \forall v' (x, y, r') \subseteq \Omega, v' \leq v\}$$

(3)

Then, discrete skeleton evolution algorithm [37] was used which filters out branches by evaluating the reconstruction weights of skeleton. The weight $w_i$ for each end branch as:

$$w_i = 1 - \frac{A(R(S - P_i, f(l))))}{A(R(S))}$$

(4)

Where function $A()$ represents the area function. Function $R(S)$ pertains the reconstruction of a skeleton $S$. $P_i(l)$, $f(l)$ corresponds to the branch from the endpoint $l$, to the nearest junction point $f(l)$. $l = (i = 1, 2, \ldots, N)$ denotes the endpoints of a skeleton $S$. $f(l)$ signifies the nearest junction point for each $l$. Following repeated training and adjustment, in this study, 0.005 is selected as the threshold for $w_i$.

Moreover, let $r(s)$ denotes the radius of the maximal disk $B(s, r(s))$ centred at a skeleton point $s$. The reconstruction of the skeleton $S$ is denoted as $r(s)$ and is defined as:

$$R(s) = \bigcup_{s \in s} B(x,r(s))$$

(5)
Finally, an unbiased detection algorithm of curvilinear structure proposed by Steger [38] was used to fit one-pixel width skeletonized cracks. Three optional parameters were determined by Equation (3) ~ (5), and the visualization process was shown in Fig. 4 and 5.

Fig. 4. The skeletonization and pruning process of a binary image

Fig. 5. Application of ridge detection algorithm for crack skeleton fitting outcomes

Line width \( w \): The line diameter in pixels. It estimates the mandatory parameter \( \sigma \) by:

\[
\sigma = \frac{w}{2\sqrt{3}} + 0.5
\]  

(6)

High contrast \( b_{\text{upper}} \): Highest grayscale value of the line. It estimates the mandatory parameter ‘Upper threshold’ \( T_U \) by:

\[
T_U = 0.17 \cdot \frac{2 \cdot b_{\text{upper}}}{\sqrt{2\pi \sigma^2}} \left( \frac{w}{7} \right)^2 e^{-\frac{(w)^2}{2\sigma^2}}
\]  

(7)

Low contrast \( b_{\text{low}} \): Lowest grayscale value of the line. It estimates the mandatory parameter ‘Lower Threshold’ \( T_L \) by:

\[
T_L = 0.17 \cdot \frac{2 \cdot b_{\text{low}}}{\sqrt{2\pi \sigma^2}} \left( \frac{w}{7} \right)^2 e^{-\frac{(w)^2}{2\sigma^2}}
\]  

(8)

Given the one-pixel width of the crack skeleton in this investigation, the parameters were set as follows: \( w=0.5 \) (half the width), high contract threshold \( b_{\text{upper}}=255 \), and low contract threshold \( b_{\text{low}}=0 \). Upon inserting these values into Equations (6), (7), and (8), the ensuing computations yield \( \sigma=0.64 \), \( T_U=29.92 \), and \( T_L=0 \).

b) Analysis of Crack Damage Penetration Depth

During the field measurements, the identification of crack locations was facilitated through the utilization of GPR, which triggers a marker indicative of the crack’s presence. This marker is discernible as an anomaly within the raw GPR data, exemplified in Fig. 7. Following this initial detection, the lateral coordinates denoting the cracks’ positions in both the processed GPR data and interpreted GPR data can be accurately derived from the raw GPR data through a consistent proportional transformation.

Upon determination of the crack surface location, the proposed criteria for assessing the crack damage penetration depth are as follows:

- The crack-affected region is defined as the area between two red dashed lines, each spanning 10 cm from the left and right sides of the crack.
- The deepest anomaly affected by the crack must be continuous with the anomaly at the surface crack.
- Within the same structural layer, the crack-affected anomalies must exhibit obvious differences from the surrounding unaffected region.

Based on the aforementioned principle, the processing results for the thickness of eight resurfaced layer are exhibited in Table III.

### TABLE III

<p>| Thickness Measurements of Eight Campus Road Resurfaced Layers |
|----------------------|------------------|------------------|------------------|------------------|</p>
<table>
<thead>
<tr>
<th>Road Thickness (cm)</th>
<th>Road 1</th>
<th>Road 2</th>
<th>Road 3</th>
<th>Road 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Road 5</td>
<td>4.2</td>
<td>3.5</td>
<td>3.9</td>
<td>3.7</td>
</tr>
<tr>
<td>Road 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Road 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The GPR-acquired data underwent processing within the ReflexW software platform [39]. The damage penetration depth of the cracks was then analyzed in combination with the experience of the field experiments, the raw data, and the processed results to determine the crack damage penetration depth for each image. In total 1000 depth data was acquired.

In accordance with conventional practice, GPR data can facilitate the extraction of location and thickness details pertaining to individual pavement layers [30], [40]. Fig. 6 illustrates the assessment of road structure conditions utilizing the GPR dataset. In Fig. 6, the road structure is categorized into three distinct layers, namely the resurfaced layer, original surface layer, and base layer.

Fig. 6. Flexible pavement structural layer distribution from raw GPR data (a) and interpreted GPR data (b)

Given the one-pixel width of the crack skeleton in this investigation, the parameters were set as follows: \( w=0.5 \) (half the width), high contract threshold \( b_{\text{upper}}=255 \), and low contract threshold \( b_{\text{low}}=0 \). Upon inserting these values into Equations (6), (7), and (8), the ensuing computations yield \( \sigma=0.64 \), \( T_U=29.92 \), and \( T_L=0 \).
Based on the aforementioned criteria, Fig. 7 respectively illustrate the raw, processed, and interpreted GPR imagery data of crack damage penetration depth distribution in the resurfaced layer, original surface layer, and base layer. Notably, the pink and blue lines maintain the same definition as the yellow and red lines in Fig. 6.

In Fig. 7 (c), a downward-opening hyperbolic anomaly is observed within the resurfaced layer of a designated ellipse, while below it there seems to be an intact region, marked by a yellow rectangle. This indicates that these surface cracks only penetrated the resurfaced layer, with their depth determined by the apex of the hyperbolic arc. In Fig. 7 (f), two yellow rectangles highlight anomalies 1 and 2. However, the green section remains intact, suggesting that Anomaly 2 in the base layer is unrelated to Anomaly 1. The deepest hyperbolic within Anomaly 1 determines its damage penetration depth. Fig. 7 (i) shows a vertical anomaly suggesting continuous crack propagation into the base layer due to surface cracking. The lowermost hyperbolic curve within the yellow rectangle represents the extent of this influence. Finally, in accordance with the aforementioned analytical approach, each crack damage penetration depth, and its corresponding pavement structural layer were documented for subsequent crack depth prediction and crack severity evaluation.

III. PEARSON CORRELATION ANALYSIS

To explore the possible interactions and dependencies between different 3D metrics of cracks, i.e., average width, volume, length, area, and damage penetration depth, Pearson correlation analysis was conducted in Fig. 8. Pearson correlation coefficient was calculated to represent the degree of correlation using Equation (9).

\[ r = \frac{N(\sum XY) - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}} \]  

In which, \( r \) denotes the Pearson correlation coefficient, \( X \) and \( Y \) denote the two variables being compared (e.g., average width and volume), and \( N \) denotes the number of data points.

As depicted in Fig. 8, the investigation of quantifiable parameters reveals noteworthy insights. An adverse correlation is discernible between length and mean width (\( r = -0.036 \)). This observation underscores the nearly negligible linear correspondence between pavement crack length and its mean width. This observation prompts speculation that the longer crack lengths might not necessarily be associated with wider widths. Plausibly, external variables could potentially disrupt the interrelation between these metrics. Conversely, a moderately positive correlation between area and mean width is discerned (\( r = 0.525 \)). This observation postulates that pavement cracks exhibiting larger areas are concurrently linked with more substantial mean widths within the dataset. A plausible explanation lies in the conjecture that an amplified area might trigger the expansion of cracks in terms of mean width. Moreover, a robust correlation between area and length (\( r = 0.762 \)), unveiling the propensity for pavement cracks with expanded areas to concomitantly display elongated lengths. This alignment implies a proximate connection between the area and length of cracks, conceivably attributable to their shared significance in road maintenance and safety considerations.

Further inference can be drawn from the outcome of these analyses, pointing toward the preeminence of the area variable in relation to these easily measurable parameters. This conclusion is borne out by the moderate positive correlation between area and mean width (\( r = 0.525 \)), indicating the potential for alterations in area to engender modifications in mean width. Additionally, the robust correlation between area and length (\( r = 0.762 \)) signifies that an escalation in area potentially stands as a primary driver for an increase in length.

Subsequently, we delve into the interplay between facilely quantifiable indicators and their more intricate counterparts. Broadly, the indicators that lend themselves to facile
measurement exhibit discernible connections with the parameter of volume. The correlation between volume and mean width (r=0.671) accentuates the potential linkage between augmented volume and increased mean width. Similarly, the moderately positive correlation between volume and length (r=0.417) underscores the notion that amplified volume might correspond to elongated length. Furthermore, the substantial correlation between volume and area (r=0.766) unveils the trend where cracks boasting greater volumes are frequently accompanied by expanded areas. Cumulatively, these findings imply the presence of discernible associations between easily measurable indicators and the measure of volume, thereby underscoring the significance of volume as a pivotal gauge of crack severity.

Lastly, drawing from the paramount significance attributed to the damage penetration depth metric within this investigation, a comprehensive scrutiny was conducted to elucidate the interconnections between damage penetration depth and all pertinent indicators. The findings elucidate that the linear associations between damage penetration depth and the easily quantifiable indicators manifest in an exceedingly feeble manner, and comparably, the linear correlation with depth remains relatively subdued. A modest, intermediate correlation emerges in relation to visible depth. This discernment underscores the intricate nature of predicting damage penetration depth through the utilization of linear models based on other indicators.

IV. ML MODELS

To comprehensively assess the efficacy and limitations of various ML models, with a primary focus on bolstering the analytical robustness and precision of regression and classification predictions when confronted with limited sample sizes, five individual output models, namely ANN, Extreme Gradient Boosting (Xgboost), Support Vector Regression (SVR), Random Forest (RF), and K-nearest Neighbors (KNN) were employed to predict the average visible depth.

A. Volume Regression Prediction Models and Evaluation Matrix

1) Xgboost Model

Xgboost, renowned for its ability to capture intricate inter-feature dynamics, particularly those of a nonlinear nature, offers a distinct advantage in modeling data attributes such as length, width, and area, as pertinent to this research [41]. Introduced by Chen and Guestrin [42], Xgboost is deliberately selected for its commendable robustness and diminished susceptibility to overfitting, courtesy of the employed boosting algorithm, as articulated below [43]:

\[
\hat{y}_i^{(1)} = \hat{y}_i^{(0)} + \sum_{t=1}^{T} f_t(x_i)
\]

(10)

\[
f_t(x_i) = \omega_t f(x_i)
\]

(11)

Concisely, the fundamental computational paradigm of Xgboost entails an iterative amalgamation of outcomes iterated T times, delineated in Equations (10) and (11). Herein, \(i\) signifies the sample index, \(T\) stands for the count of decision trees, and \(\hat{y}_i^{(T)}\) represents the ultimate forecasted value for the \(i\)-th sample within the \(T\)-th decision tree. The function \(f_t(x_i)\) signifies the computation formulation for the \(i\)-th sample within the \(T\) decision tree. \(\omega\) corresponds to the leaf node's weight vector, contingent on the feature vector's projection onto the decision tree's leaf node.

2) Single-output ANN Model

Aiming to capture non-linear patterns and interactions between the datasets, an ANN model was adopted. This model comprises a Multi-Layer Perceptron structure featuring 10 to 12 input neurons that adapt to the discrete characteristics of each input dataset.

In the calculation process between the input layer and hidden layers, the adaptive weight coefficient \((\omega)\) plays an important role, which ensures the transfer of information. Equation (12) describes the relationship between the inputs and weight coefficient [24]:

\[
net_j = \sum_{i=1}^{n} (w_j \times x_i + b_j)
\]

(12)

In Equation (12), \(net_j\) denotes the unit that computes the total weighted input, \(x_i\) denotes the unit of the previous layer, \(w_j\) denotes the weight of the connection between the previous layer and current layer; and \(b_j\) denotes the bias of the current layer.

Furthermore, to ensure optimal convergence and facilitate the network's ability to decipher underlying patterns, Standardization was employed to process the input features and target variable. Adam optimization algorithm [44] was also used to improve the convergence speed and adaptively adjust the learning rate in different dimensions. The regularization coefficient, maximum iterations, warm start, and random state were carefully chosen to optimize training efficiency and reproducibility [44].

3) Evaluation Criteria

To assess the comprehensive performance of regression prediction model, Mean Square Error (MSE) and the Mean Absolute Error (MAE) were used to assess the accuracy of the models in predicting the target variables. Additionally, the determination coefficient, represented as the R-squared (R2) score was used to evaluate the models' capacity to explain the variability present within the observed data.

Additionally, a multi-output ANN model was developed to predict 10 distinct visible depths of each individual crack. The obtained outcomes revealed a notable deficiency in the feasibility of accurately predicting crack depth. Therefore, no relevant outcomes or models pertaining to visible depth are demonstrated within this paper. Due to the subpar regression prediction performance observed in the SVR, RF, and KNN models concerning crack volume prediction, only the individual output Xgboost and ANN models, and the multi-output ANN damage penetration depth classification prediction model are presented in this paper.

B. Damage Penetration Depth Classification Prediction Model and Evaluation Criteria

1) Multi-output ANN model

While a standard ANN typically terminates in a single
output neuron, a Multi-output ANN is characterized by having multiple neurons in its output layer. Each of these neurons corresponds to a distinct target variable. The architectural of developed model is shown in Fig. 9.

Fig. 9. The architecture of the developed Multi-output ANN model

This architectural variation addresses scenarios where multi-targets are pertinent. This not only streamlines computational resources but also enables the network to capture interdependencies between the outputs, enhancing overall predictive accuracy. The significance of this feature becomes particularly evident in the context of crack damage penetration depth classification, where the outputs are a set of depth which includes inner relationship.

2) Evaluation Criteria

The fundamental objective of the assessment metric for the classification predictive model resides in the discernment of model precision concerning the demarcation between the dual categories of fissures. This evaluative yardstick conventionally comprises four cardinal facets, namely True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN) [34]. Within the context of this investigation, TP correspond to accurate identification of profound cracks as deep, while FP denote erroneous classification of superficial cracks as deep. FN entail the model's misclassification of deep cracks as shallow, whereas TN denote the correct identification of shallow cracks as such.

Within the framework of evaluating crack severity in flexible pavements, this study takes into account the practical significance of individual parameters for decision-making, with detailed discussions provided in Section 5.3. The evaluation of the predictive model's performance ultimately revolves around three pivotal parameters: accuracy, recall, and F1 score [36].

V. RESULTS AND DISCUSSION

A. Crack Volume Regression Prediction

The Xgboost and ANN outperformed the KNN, SVR, and RF models in terms of prediction accuracy and model interpretability. This superior performance can be attributed to the Xgboost and ANN ability to better capture the complex non-linear relationship between width, length, area, and volume of cracks. Specifically, ANN has the capability to automatically acquire feature representations from data, while Xgboost is able to automatically select the important features to further enhance prediction accuracy. Consequently, this paper only presents the prediction results associated with the Xgboost and ANN models, as detailed in Table IV.

As can be seen from Table IV, when employing width in isolation as the input feature and subsequently juxtaposing it with combinations such as width-length or width-area in the model's input, a substantial enhancement in prediction performance is observed for both the Xgboost and ANN models. In this context, the utilization of width and length as input variables yields notable improvements in prediction accuracy for the Xgboost model, with reductions of 38.3% (MSE) and 24.5% (MAE) and a concurrent enhancement in interpretability by 42.4%. Similarly, the ANN model demonstrates substantial gains in predictive performance, achieving reductions of 51.5% (MSE) and 31.6% (MAE) in prediction errors, coupled with a 36.5% improvement in interpretability. Furthermore, when width and area are employed as input features, the Xgboost model exhibits even more pronounced improvements, with reductions of 49.5% (MSE) and 36.2% (MAE) in prediction errors and an impressive 60.7% enhancement in interpretability. The ANN model also showcases substantial predictive enhancements, with reductions of 57.7% (MSE) and 32.2% (MAE) in prediction errors and a concurrent boost of 42.5% in interpretability.

The observed improvements in predictive performance can be explained through Pearson correlation analysis. Notably, a robust correlation coefficient of 0.762 between area and length signifies a substantial association, suggesting the presence of feature overlap between these two parameters. Conversely, the correlation coefficients of -0.036 between width and length, and 0.525 between width and area indicate comparatively weaker associations. This underscores the inadequacy of relying solely on width features as input, thereby rationalizing the remarkable and parallel enhancements in predictive performance for both models.

Moreover, as evidenced by the data presented in Table IV, the utilization of width and area features for volume prediction outperforms the use of width and length features. This could be attributed to the stronger correlation between area and volume (r=0.760), surpassing the correlation between length and volume (r=0.417).

When width, area, and length are concurrently employed as input variables (Table IV), both the Xgboost and ANN models exhibit a decrement in performance. This phenomenon is likely attributed to the limited size of dataset. When two highly linearly correlated features, i.e., area and
length, are simultaneously introduced, the model might inadvertently capture complex and variable combinations that lack practical significance. Consequently, the test set samples may lack corresponding features, leading to a deterioration in model performance.

Furthermore, when comparing the performance of the Xgboost and ANN models, it becomes evident that, with the exception of the combined input of length and width, the Xgboost model consistently exhibits smaller errors and relatively higher prediction accuracy compared to the ANN model. However, it is worth noting that, on the whole, the ANN model demonstrates superior capability in elucidating the variability of the target variable when compared to the Xgboost model. Therefore, this study recommends selecting between these two models based on specific analytical needs.

Eventually, to highlight the predictive abilities of ML models, the comparison between predicted values and actual values for ANN and Xgboost models with test datasets was graphically shown in Fig. 10.

![Graphs](image1.png)

**Fig. 10.** Volume predictions by ANN model (a-c) and Xgboost model (d-f) versus ground truth, using variable input feature: Width and Length (a and d), Width and Area (b and e) and Width, Length, and Area (c and f)

---

### B. Sensitivity Analysis with Shapley Additive exPlanations (SHAP)

It was evident from the obtained results that the changes in feature correlation can affect model performance. In this regard, this study used sensitivity analysis [41] to further explore the importance of different features to determine which features have a greater impact on model performance (Fig. 11 and 12).

![Graph](image2.png)

**Fig. 11.** SHAP results – crack volume prediction of the Xgboost model, colors transitioning from blue to red, indicating feature values from low to high, noting that isolated red feature values on the left side of the images, can be regarded as noise.

![Graph](image3.png)

**Fig. 12.** SHAP results – crack volume prediction of the ANN model, colors transitioning from blue to red, indicating feature values from low to high, noting that isolated red feature values on the left side of the images, can be regarded as noise.

As we can see from Fig. 11 and 12, there is a noticeable trend where elevated eigenvalues align with higher SHAP values. Notably, for Width 2 in the Xgboost model and Width numbers 7 and 3 in the ANN model. Disregarding these outliers, the general observation is that larger and more densely clustered eigenvalues are indicative of greater contributions from their respective input features to the prediction of crack volume.

By moving to the sequence of the input features to do a comprehensively comparison, it can be seen from Fig. 11 and 12 that the area is the most influential feature in the Xgboost model and second in the ANN model. Having said that length leads in the ANN model and ranks fourth in the Xgboost model. This suggests that length and area have more
substantial contributions to crack volume prediction than width. Finally, in contrast to Pearson’s linear regression, where the linear relationship between area and volume is deemed relatively robust, with average width following closely behind and length exhibiting the weakest correlation, it is noteworthy that ML results offer a different perspective. This disparity in findings could be attributed to the limitations of linear regression in capturing non-linear relationships between input and output features, as well as the intricate and multifaceted interplay among area, length, and width. Consequently, in practical engineering applications, relying solely on a single easily measurable metric to predict crack volume is an inaccurate approach, given the complexities inherent in these relationships.

C. Crack Damage Penetration Depth Classification Prediction

The training outcomes for binary classification of crack damage penetration depth utilizing the multi-output ANN model are summarized in Table V.

<table>
<thead>
<tr>
<th>TABLE V</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GRADE CLASSIFICATION PREDICTION PERFORMANCE OF ANN MODEL</strong></td>
</tr>
<tr>
<td><strong>Evaluation Criteria</strong></td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
</tr>
<tr>
<td><strong>Recall</strong></td>
</tr>
<tr>
<td><strong>F1 Score</strong></td>
</tr>
</tbody>
</table>

\(^1\)Type include transverse, longitudinal and net-shape. 
\(^2\) Easily measurement parameters include 10 sets of width, length, and area dataset.
\(^3\) Uneasily measurement parameters include volume and 10 sets of visible depth dataset.

Table V reveals a commendable level of accuracy in the classification of damage penetration depth. This observation suggests a strong possibility of nonlinear correlations between crack damage penetration depth and various 3D metrics. Furthermore, it is foreseeable that augmenting the training dataset with more extensive data could lead to an enhanced classification accuracy for the model in subsequent iterations. This finding underscores the viability of predicting crack damage penetration depth through the utilization of easily accessible (type, length, area, and width) and less accessible (visible depth and volume) crack indicators.

Moreover, it is evident that the increase of input features has a pronounced positive impact on the model’s performance. This observation underscores the intrinsic importance of each feature dimension in predicting crack damage penetration depth, especially for the comb of the length and area. This implies that in addition to width and type metrics, the dimensions of length and area hold significant importance in the accurate prediction of crack depth. Furthermore, it is of significance to observe that with the incorporation of volume and visible depth features, albeit resulting in an overall performance enhancement, the gains are modest. In particular, the recall metric appears to have attained an extremum. Therefore, this study recommends future crack inspection efforts focus on acquiring a wide range of 3D metrics, especially width, area, length, and type of cracks.

Moreover, it is worth noting that in Table V, the practical context for the decision-making based on this data, e.g., pavement maintenance, during the computation of optimal values for model evaluation indicators such as Accuracy, Recall, and F1 Score has been taken into consideration. These metrics are intricately linked to the parameters of TP, TN, FP, and FN [34], [36]. Specifically, under the condition of maximizing the sum of TP and TN, there is a greater inclination to accept the misclassification of crack damage points in the resurfaced layer as belonging to the original layer, further reduce the likelihood of misclassifying crack damage points from the original layer into the resurfaced layer. In other words, it is more acceptable for crack severity to be underestimated than overestimated, as the latter might lead to the formulation of a maintenance program that falls short of restoring the pavement to its optimal performance. This strategy aims to ensure that the maintenance plan, even if somewhat conservative, is sufficiently effective in bringing the pavement back to its peak performance [45].

D. Application of Damage Penetration Depth Classification

To quantitatively evaluate crack severity in terms of penetration propagation, further optimization of the existing crack severity evaluation system is required. Damage Index (DI), which gauges crack severity by considering damage penetration depth in flexible pavements, has been introduced in this paper. It is expected to provide more references for the development of future crack intelligence detection, evaluation, and maintenance strategies. The DI can be determined through the utilization of Equations (13).

\[
DI = 1 - \left(1 - \frac{a_n}{b}\right)^n
\]  

(13)

In which, \(n\) denotes the count of depth points of cracks penetrating into original layers of asphalt pavement, with a permissible range from zero to five. The parameter \(a\) denotes the critical fracture energy required for the propagation of an existing crack that has penetrated the resurfaced layer. While \(b\) denotes the critical fracture energy required for the propagation of an existing crack that has penetrated the original layer.

To secure parameters \(a\) and \(b\), this study suggests the acquisition of fractured asphalt samples through a dual approach: in-situ coring samples and/or dedicated laboratory preparation of Marshall specimens. Subsequently, indirect tensile tests are administered to assess the fracture energies of specimens categorized as \(a\) and \(b\). This analysis involved the systematic variation of parameters, including loading rate, loading type, sample geometry, preload, data acquisition rate, and testing environment, among others [46], [47].

Furthermore, in cases where the parameters \(a\) and \(b\) pose challenges in their quantification, this research proffers a maintenance reference strategy based on the predicted
variable \( n \) as defined in Equation (13) and presented in Table VI. The maintenance recommendations provided in Table VI are specifically designed to address various crack types and depth distributions. It should be noted that while each country might have its own standards, such as [2], [3], [4], [5], [6], [7], [8], [9], [10] for pavement maintenance strategies, these standards normally provide a range of possible strategies and hence our suggestions in Table VI could be used to select the most appropriate one(s).

<table>
<thead>
<tr>
<th>Maintenance decisions based on ( n )</th>
<th>Type</th>
<th>Maintenance options and suggestions</th>
</tr>
</thead>
</table>
| \( n = 0 \)                           | Transverse, Longitudinal, and Net-shaped crack | Crack filling (Filling cracks after debris removal)  
Crack sealing (Grooving and cleaning debris before sealing. It is recommended to cut at the interface of the resurfaced layer and the original layer, with the depth of the resurfaced layer varying based on different pavement structures)  
Crack banding |
| \( 0 < n \leq 3 \)                    | Stable original layer  
Transverse, Longitudinal, and Net-shaped crack | Crack filling (Filling cracks after debris removal. This method is suitable for dormant cracks or those exhibiting severe aging, where slot cutting is not a feasible option)  
Crack sealing (Following the pavement structure, groove the pavement down to a stable location where is impermeable to water. The depth of these grooves should not be less than the maximum depth of the crack, followed by filling the cracks.)  
Crack banding (This approach should not be employed independently. It is advisable to utilize it in combination with other strategies and subsequent to their application.)  
Unstable original layer (pumping, settlement, voids, and other distresses) (Firstly, repairing the original layer (e.g., using a large-size permeable asphalt mixture). Subsequently, filling both the original and resurfaced layers with the identical material as that of the original pavement.) |
| \( 3 < n \leq 5 \)                    | Transverse and Longitudinal crack  
Net-shaped crack | Consider maintenance strategies within the range of \( 0 < n \leq 3 \).  
The procedure for repairing net-shaped cracks can be summarized in four steps:  
Along the outer contour of the crack, excavate/mill down to the interface between the resurfaced layer and the original layer.  
Continue excavating/milling at the location of the visible crack, extending the operation to reach either the bottom of the crack or the unstable portion of the original layer.  
Repairing the original layer.  
Resurfacing the resurfaced layer |

VI. CONCLUSIONS

This study accumulated a dataset comprising 200 sets of 3D parameters for flexible pavement cracks. A comprehensive interrelationship among these parameters led to the successful development of predictive models for crack volume and damage penetration depth [29].

Initially, Pearson correlation analysis was employed, which showed a notably weak linear correlation between the most commonly used indicators of crack severity, crack width, and damage penetration depth. Even other easily measurement indicators, such as length and area, failed to establish linear correlations with damage penetration depth. This underscores the limitations of relying solely on width to assess crack severity.

Furthermore, we explored the possibility of utilizing five single-output and one multi-output ML models to predict the average visible depth, visible depth, and volume of cracks. The results indicated that these models were unsuccessful in identifying correlations with easily measurable metrics for average visible depth and visible depth. However, for volume prediction, single-output ANN model and Xgboost model achieved relatively high R2 scores of 0.832 and 0.748, respectively.

Additionally, the developed multi-output ANN model successfully classified the crack damage penetration depth with high accuracy, achieving optimal precision, recall, and F1 scores of 0.790, 0.779, and 0.761, respectively. This demonstrates the feasibility of predicting damage penetration depth in flexible pavement cracks. Based on this, we introduced the calculation method for the damage penetration depth parameter, \( D_I \), and recommended corresponding maintenance strategies based on the count of depth points of cracks penetrating into original layers of asphalt pavement.
It should be noted that the research conclusions and maintenance recommendations of this study primarily apply to flexible pavements with resurfaced layers [47], [48], where the thickness of the resurfaced layer ranges from 3–5 cm. Based on our dataset, we advise against allowing crack damage penetration depth to exceed 17 cm.

Future work can be categorized into three areas:

1) The dataset used in this study may benefit from expansion in terms of size. With more 3D parameter data on flexible pavement cracks, the accuracy of predicting crack severity using easily measurable metrics could potentially improve, leading to more reliable results.

2) While this study successfully demonstrated the predictability of crack volume using easily measurable metrics, it did not explore how volume could further quantify crack severity. This remains a potential avenue for future research.

3) The specific calculation methods for coefficients $a$ and $b$ in the crack severity prediction equation proposed in this study require further investigation and experiments.

REFERENCES


Zhe Li received the bachelor's degree in Traffic and Transportation Engineering from Qingdao University of Technology in 2018. He is currently pursuing the Ph.D. degree with Railway and Highway Engineering, Chang'an University, in corporation with University of Birmingham. During his studies, he was granted the opportunity to work with Micromechanics & Smart Infrastructure LAB, Boise State University. His research interests include pavement condition evaluation, machine learning-based 3D modeling, non-destructive testing, and pavement intelligent inspection.

Dr. Mehran Eskandari Torbaghan is a lecturer in Infrastructure Asset Management at the Department of Civil Engineering with a PhD focused on risk management and linear infrastructure systems from University of Birmingham. Mehran spent five years in the civil engineering industry, as a geotechnical engineer, before returning to academia to study a MSc in Construction Management and then the PhD both at the University of Birmingham. He then worked as a research fellow at the University of Birmingham for five years, before becoming a lecturer. He was also a visiting research fellow at the University of Nottingham. His research portfolio and interest lie in the field of smart management of infrastructure systems, investigating the interaction between roads, buried utilities, and the ground. He is a member of the Editorial Panel of the Proceedings of the Institution of Civil Engineers – Transport, a Senior Fellow of Higher Education Academia, and a member of CIHT.

Tuo Zhang received the B.S. degree in intelligent science and technology from Beijing University of Posts and Telecommunications in 2016. He is currently pursuing the Ph.D. degree in computer Science with the University of Birmingham's Computer Science Department. His research interests include the theoretical aspects of Artificial Intelligence, with a particular emphasis on the intersection of Reinforcement Learning and Game Theory.
Xia Qin received the B.S. degree in civil engineering from Heilongjiang University in 2018, and the M.S. degree in civil engineering from University of Birmingham. He is currently pursuing the Ph.D. degree in civil engineering with the University of Birmingham. His research interests include recycled materials used in reinforced concrete structure, railway and transportation systems, and zero carbon emissions.

Dr. Wenda Li currently is an assistant professor of telecommunication at the Institute of Sensors, Signals and Systems, Heriot Watt University. He was a lecturer in the Department of Biomedical Engineering, School of Science and Engineering, University of Dundee between 2022 and 2023. He obtained his PhD at the University of Bristol in 2018 with research focused on remote sensing techniques. After that, he joined the University of Birmingham as a Research Fellow and moved to the University College London in 2019 also as a Research Fellow. His research focuses on the signal processing for passive radar and high-speed digital system design for wireless sensing applications in healthcare, security, and positioning. His research in passive WiFi radar has led to a number of IEEE conference and journal publications.

Yongjian Li received the B.Eng. degree in Chongqing University of the Posts and Telecommunications, China, in 2020. He was the research associate in West China Hospital of Sichuan University from 2020 to 2021. He is currently pursing the Ph.D. degree in environmental health and risk management with the University of Birmingham, UK. His main research interests including machine learning, city sustainability, smart city, road weather information system, transportation resilience, IoT, etc.

Prof. Jiupeng Zhang received the B.S. degree in civil engineering and Ph.D. degree in transportation engineering from Southeast University, China, in 2004 and 2009 respectively. From September 2014 to September 2015, he was a visiting scholar at the University of Illinois Urbana-Champaign, Champaign, IL, USA. He is currently a Full Professor and Vice Dean of the School of Highway, Chang’an University. He is also a Vice Director of the National Key Laboratory of Civil Aviation Administration of China. His broad research areas are the infrastructure performance evolution, intelligent maintenance for infrastructure state sensing, and integrated road transport, environment, and energy development.

He is serving on the Editorial Board for the International Journal of Transportation Science and Technology, Intelligent Transportation Infrastructure, etc. He is a member of the ASCE, CHTS, and IACIP.