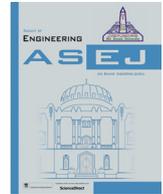




Contents lists available at ScienceDirect

Ain Shams Engineering Journal

journal homepage: www.sciencedirect.com

Civil Engineering

Predicting pavement condition index using artificial neural networks approach

Amjad Issa ^{a,*}, Haya Samaneh ^b, Mohammad Ghanim ^c^a Civil Engineering Department, An-Najah National University Nablus, P.O. Box: 7, Palestine^b Computer Engineering Department, An-Najah National University Nablus, P.O. Box: 7, Palestine^c Transport Engineering Expert, Jacobs Engineering Group, Qatar

ARTICLE INFO

Article history:

Received 3 February 2021

Revised 7 April 2021

Accepted 29 April 2021

Available online xxxxx

Keywords:

Pavement

Pavement Condition Index

Artificial Intelligence

Artificial Neural Networks

Prediction

ABSTRACT

Pavement Condition Index (PCI) is a numerical assessment of pavement conditions based on existing distresses. The PCI values are used for pavement management and rehabilitation programs. Calculating the PCIs using conventional method relies on collecting relevant field data (such as distresses types and severity) by visual inspection method. The collected data are processed to estimate the PCI values, which is a lengthy process that requires technical experience. This research aims to model the relationship between distresses type and severity and PCIs via straightforward and adaptive model. Therefore, Artificial Neural Networks (ANN) capabilities are employed to predict the PCI values of the different sections, thus reducing the required efforts and technical experiences to estimate PCI values. Moreover, the use of ANN enables the possibility of introducing new localized variables, such as the presence of man-holes in pavement sections. The total of 348 directional sections from 10 different roads located in the City of Nablus, Palestine were examined to collect the distresses-related data and to estimate the corresponding PCI values using ASTM 6433 07 method. The results revealed low correlation between distresses and PCI, where the highest absolute correlation between PCI and any distress type and severity did not exceed 0.38. The results indicated that the ANN model is capable to predicting the PCI with high level of reliability, with an R^2 value of 0.9971, 0.9964 and 0.9975 for training, validation and testing datasets, respectively. The regression slope between observed and predicted PCIs ranges between 0.9964 and 0.9974.

© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pavements are considered the major asset of highway infrastructure in Palestine. Pavement performance is often measured using a set of indicators, such as Present Serviceability Rating (PSR), Pavement Condition Index (PCI), and International Roughness Index (RI). They have been widely employed to develop pavement maintenance strategies [1]. Evaluation of Pavement conditions which includes evaluation of friction, surface roughness, pavement structure, and existing distresses are considered

as one of the main components of pavement design and rehabilitation in any Pavement Management System (PMS). The majority of the cost-effective Maintenance and Rehabilitation (M&R) strategies which were developed using the PMS have resulted in accurate pavement evaluation [2].

Due to the lack of financial resources and allocated budgets targeting pavement maintenance and rehabilitation in Palestine, the evaluation of existing pavement distresses in terms of PCI is considered one of the main components of PMS, which is used to identify the deterioration in the pavement sections and identify a proper maintenance strategy.

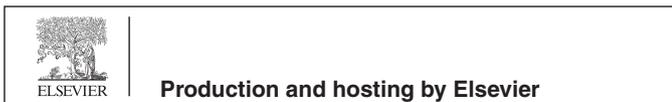
Two different categories of pavement distresses are usually classified into functional and structural failures. In most cases there is usually one type of failure that can be observed. While there are other cases where one type of failure would potentially lead to the development of the other failure type [3].

The functional failure is often expressed by the degree of surface roughness. The second type of failure on the other hand (i.e., the structural failure) is often characterized by the existence of

* Corresponding author.

E-mail address: amjadissa@najah.edu (A. Issa).

Peer review under responsibility of Ain Shams University.



<https://doi.org/10.1016/j.asej.2021.04.033>

2090-4479/© 2021 THE AUTHORS. Published by Elsevier BV on behalf of Faculty of Engineering, Ain Shams University.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

fatigues (or what is called alligator cracks), shear-developing or consolidation which is presented in one or more layer of the pavement structure [3,4]. On the other hand, and due to the limited pavement rehabilitation allocated funds, there is usually immediate needs to prioritize any allocated funds. This prioritization is accomplished by developing systematic procedures for scheduling M&R activities to maximize the anticipated benefits of the road users and reduce the associated M&R cost. Thus, PMS would allow local agencies and engineers to allocate the required budgets and funds, personnel and resources in an effective manner [5].

PCI can be defined as a numerical index with a value between 0.0 and 100. The PCI is widely acceptable to describe the overall pavement surface condition of a roadway section. The perfect score (i.e., a score of 100) indicates the best possible pavement condition, while the score of 0.0 is representing the worst possible pavement condition.

The rating of a roadway in terms of the pavement condition index is usually based on examining the existing surface distresses. Accordingly, PCI does not directly measure skid resistance, the capacity of pavement structure or surface roughness. [6]. Adopting the PCI system supports the process of identifying any immediate and necessary M&R of roads [7], which is expected to identify appropriate preventive maintenance strategies, allocate budgets and evaluate pavement design methods and its construction materials.

The ASTM standards for roads and parking lots pavements are used in the PCI survey procedures and calculation methods [8]. The ASTM standards identified terms related to PCI through development calculation sheets for PCI which can be filled automatically.

The PCI score of a road segment is initiated by identifying the pavement section, which is defined as an adjoining pavement area with a unified maintenance, design, structure, climatic conditions, same traffic volume demand, usage history, structural and geometric characteristics, this section is later visually inspected for any pavement distresses [33].

Several studies discussed the pavement performance in terms of common distresses such as rutting and fatigue and illustrated the methods to predict pavement performance. For example, Mousa, et al predicted the performance of constructed pavement with base layer consisting of reclaimed asphalt pavement (RAP)/virgin aggregate blends using Multi-Layer Elastic Analysis software (KENLAYER) considering horizontal tensile strain at the bottom of AC layer and the vertical resilient strain at critical locations within the pavement system [9]. They determined the total pavement rutting and fatigue cracking using the critical strains computed by the Multi-Layer Elastic Analysis along with the Mechanistic Empirical Pavement Design Guide performance models and transfer functions. Based on the results, the RAP blends showed superior/comparable performance compared to natural aggregates for the application in base/subbase layers for the Egyptian conditions. They concluded that the effect of the rate of loading and climate conditions was significant on both asphalt concrete layer fatigue cracking and rutting.

Arab et al investigated the combined impact of compaction level and initial matric suction of construction and demolition (C&D) waste material used as a based layer on simulated pavement performance using the AASHTO are Pavement ME Design software [10]. The results showed that the compaction level has more significant effect on the base and subgrade rutting and bottom-up fatigue cracking than the initial matric suction. Moreover, both initial matric suction and SWCC input level were found to have a significant effect on all simulated distresses. Furthermore, the interaction between the compaction and initial matric suction was found to be significant.

Azam et al investigated the rheological properties of a conventional asphalt binder (typically used in Egypt) modified with different polymer products and wax [11]. They estimated the dynamic modulus (E^*) of the asphalt mixes considering the popular predictive equations for the prediction of pavement performance using the quality-related specifications software (QRSS) at three climatic conditions and two different rates of loading (traffic speed) in Egypt. The results showed better mechanical properties and better resistance to moisture damage for the modified mixtures. Finally, the findings indicated that the use of polymer products or wax as modifiers is helpful in improving the pavement performance.

Tarbay et al presented the use of waste materials (marble and granite) and by-product material (steel slag) as alternative to the mineral conventional filler [12]. Moreover, they used Quality-Related Specifications Software which is a simplification of the Mechanistic-Empirical Pavement Design Guide software, in order to predict the field performance. They found that all materials enhanced rutting resistance compared to the traditional limestone filler.

The use of artificial intelligence application (AI), including Artificial Neural Networks (ANNs) has wide application in civil and infrastructure engineering, and the use is rapidly increasing. The application of ANNs includes transportation, pavement engineering, structure, and environment [13–15,35]. Inspired from how the human brain works, ANNs uses nonlinear statistical algorithms to model complex relationship between a set of inputs and outputs. ANN has gained popularity through its ability of solving prediction and recognition problems [13].

2. Problem Statement and Study Objectives

In Palestine, long and exhausted procedure is used for calculating the pavement condition index based on detailed visual inspections through collection the targeted 100 m sections in terms of extent (area of the distress divided by the area of the section) and severity (low, medium, and high). However, in this paper, the modeling of PCI is implemented through using ANN model for estimating and predicting PCI. Accordingly, a proper maintenance and repair strategies and corresponding costs can be easily defined and prioritized. Furthermore, this research represents the first study to consider this innovative approach in estimating the PCI in Palestine.

Another key factor that is considered in this research is the significant number of manholes that is found roadway segments. The manholes tend to contribute toward reducing the pavement conditions due to the improper installation and poor maintenance. Their negative impact is further induced by their relatively high density, where they were spaced every 10 m in certain sections. The manhole densities per section was not considered when the PCI is estimated, despite its role in reducing the pavement quality. Therefore, the proposed ANN model in this research has considered the manholes when the PCI is predicted.

Accordingly, the main objective of the paper is to calculate and predict the PCI utilizing the ANN and depending on several indicators such as pavement distresses density and severity and the number of manholes in each section.

In order to have a better idea about the pavement performance of the studied segments, the PCI scores for those segments were estimated [16]. The PCI value ranges between 0 and 100, where high values are associated with better the pavement conditions and quality.

The main expected advantages of the proposed models are:

- Providing the Palestinian municipalities and the local agencies with a tool of evaluation pavement conditions.

- Performing what-if analysis and consequently prioritizing pavement rehabilitation programs and budget allocation.
- The proposed ANN model can be integrated within an image processing platform that would automatically detect and identify pavement defects and predicted the PCI value. Ultimately, this integration is expected to reduce the human efforts and errors when it comes to quantify the pavement conditions or defects.

The contribution of this paper is presented through developing an Artificial Neural Networks model. The developed model is used to estimate PCI as a function of several inputs that characterizes the pavement sections' distresses and conditions, in addition to minimizing the number of inputs in the model by considering the main category of each distress. The developed model can be further utilized by different municipal engineers to provide an easy and reliable assessment of the PCI score for the various roadway sections.

3. Literature Review

The new model uses the artificial neural network (ANN) approach. This approach is considered innovative and is applied to Palestine pavement management for the first time. The following studies summarize relevant regional and international works.

Different ANN models related to flexible pavement management and maintenance were developed. For instance, Eldin, Neil N et al. presented an overview of management tool that is used for flexible pavement maintenance [17]. The tool was designed neural networks methods based on Oregon Department of Transportation. The authors used a set of 744 training and 1736 testing records to develop the artificial neural networks model. Yang, Jinglin et al. used the ANN methodology to assess pavement condition based on a set of three performance indices by developing three ANN models [18]. The three indices are used to rate cracks, rut and ride. The results indicate the ANN models capabilities of forecasting a five-year pavement conditions index. The ANN models were incorporated within a computer-based software to facilitate the predictions.

The optimization techniques (genetic programming) incorporated with ANN were developed for predicting PCI by Shahnazari, Habib, et al. developed an alternative approach using ANN and genetic programming optimization techniques for predicting the PCI [19]. A soft computing method was proposed by the authors to estimate the PCI and accordingly to be used in the PMS. The results showed that both models are compatible with the measured data in the field, although the first model has resulted in more accurate results when compared to the second model [19]. Furthermore, Amin et al. developed pavement performance optimization method using linear programming [20]. The case study was Montreal road network city. They considering simulated traffic for 50 years period. Moreover, they took into account modeling the uncertainty of pavement performance. Furthermore, they applied the generalized delta rule learning algorithm with the backpropagation neural network (BPN) where uncertainties were ignored. The estimated PCI values were determined based on a set of inputs, including Annual Average Daily Traffic (AADT), equivalent single axle loads (ESAL), and pavement condition index. In addition to, M. Jalal et al. used an optimized ANN model to forecast the PCI. The authors performed an experimental study that took place at Texas A&M University Campus. They carried out a multi-step in order to find the optimized ANN model to predict the PCI obtained from the experimental data. The PCI prediction results show that for, the accuracy of optimal ANN utilizing this approach is well above the normal ANN models [21].

Moreover, several studies including developing models that predict the PCI from International Roughness Index (IRI) such as

Arhin, Stephen A., et al. [22]. The authors developed models that predict PCI from International Roughness Index (IRI) by using two years of IRI-PCI data sets considering pavement type and functional classification in Columbia District [22]. The results based on the mean PCI and IRI values showed that highway and local roads have first and final ranking in terms of smooth riding. For pavement type, the ranking from rough to smooth was: Concrete Pavement, Asphalt Pavements, Composite Pavements. Furthermore, Elhadidy et al. evaluated the pavement surface condition based on IRI and PCI values [23]. The authors applied regression analysis model on a group of data that was extracted from the LTPP database which correlates IRI to the PCI. The case study estimated PCI through includes pavement distresses from 1,448 LTPP representing 12,744 data points. The underlying relationship between IRI and PCI was best modeled with a sigmoidal function. The coefficient of determination (R^2) of this model was relatively high, with a value of 0.995. Moreover, Mazari et al. utilized the Long-Term Pavement Performance (LTPP) database to develop a new approach to estimate the IRI [24]. In this research, the authors extracted two datasets and applied a hybrid approach that combines two AI methods, namely Gene Expression Programming (GEP) and ANN. IRI was predicted through developing a reasonable algorithm using structural properties of pavement and traffic parameters. Moreover, estimation of present IRI utilizing another set of LTPP data historical data was assessed. A reasonable performance power was produced by the second prediction model.

Majidifard, Hamed et al. studied nine categories of distress classes manually for approximately 7,237 google street view images [25]. Then, they implemented and trained the deep learning framework using the labeled dataset. The authors also developed U-net based model that was used to quantify distresses' severity. A hybrid model was created to integrate the YOLO and U-net model that can classify the detected distresses according to their type and severity in one step. Furthermore, Vyas, Vidhi et al. tried to develop reliable correlations using a number of environmental structural, subgrade soil attributes and functional as input in order to estimate two different deflection basin parameters [26]. About 124 km long pavement network were studied by the authors. They trained different ANN models with different architects that consists of one or more hidden layers. They compared the ANN results against those obtained by applying multiple linear regression models and other non-intelligent approaches. The ANN outperforms the other models.

The gap that the researchers want to cover is illustrated in providing the developing countries such as Palestinian municipalities and agencies with a tool for evaluating existing pavement conditions. Furthermore, based on that what-if analysis and scenarios for pavement maintenance prioritization can be applied in the future. Moreover, for future perspective, the proposed ANN model can be integrated within image processing platform that would automatically detect and identify pavement defects and predicted the PCI value. Ultimately, this integration is expected to reduce the human efforts and errors when it comes to quantify the pavement conditions or defects

4. Research Methodology

The pavement condition deterioration and degradation are expressed by an external indicator called pavement resulting from different causes such as environmental factors, loading, or both of them. The common distresses in flexible pavement are rutting cracks, and weathering which are usually appeared on the surface of pavement. Each distress is classified into three levels of severity according to its effect on functional and structural pavement performance and riding quality: Low (L), Moderate (M), and High

(H). Federal Highway Administration (FHWA) published in 2014 comprehensive and completed Distress Identification Manual [16].

In general, and as summarized in Table 1, there are 19 different types of distresses that can be caused by different factors, including example pavement mixture design, traffic loads, quality of the materials and more. It is worth mentioning that out of the 19 different types distresses in total, the “Rail-Road Crossing” distress does not exist in the City of Nablus, simply because there are no railroads within the city.

The following paragraphs illustrate the ASTM procedure used in calculating the PCI:

Distress density is defined as the ratio of observed or detected distress area and the section area. It is expressed in percentage. When calculating the distress density, each level of distress severity is often treated independently from the other levels, even if they all relate to the same distress type.

A series of adjustments are then applied to calculate the PCI value. For each distress level of distress, a statistical weight number called Deduct Value (DV) is applied toward calculating the PCI value. The DV is retrieved from a set of curves. More details can be found in ASTM 6433–07 [8].

The summation of the all the DVs is called the Total Deduct Value (TDV). Once estimated, further adjustments are made through applying Corrected Deduct Value (CDV). The CDV is used to normalize the TDV so that its range is between 0.0 and 100. PCI is calculated based on the maximum of CDV (maxCDV) as illustrated in Equation 1 [8].

$$PCI = 100 - \max CDV \quad (1)$$

Different approaches were found in the literature to model or compute the PCI as a function of pavement distress type. For instance, PCI rating is conducted as per the literature [8,27–29]. These researches modeled the PCI as a function of the main pavement distresses in Palestine such as alligator cracks, longitudinal and transverse cracks, fatigue, slippage, shoving, lane-shoulder drop off, patching, polishing, potholes, bleeding, rutting, raveling, etc. These distresses are the main common flexible pavement defects in Palestinian roads network.

The first step in meeting the objectives of this research is to assess different pavement sections based on their physical characteristics distresses conditions., therefore, 10 different roads were selected and inspected in each direction. Each street was divided into multiple directional sections with 100 m section length. The widths of those sections were also measured as they varied. The type, severity and area of each distress type were collected for each section. All the collected data were stored in a database structure for further analysis. In total, 348 sections were individually inspected, and their data were gathered.

Secondly, the database used in this research was collected from the targeted roads in the study area. As stated before, all existing pavement distresses were collected using the ASTM 6433–07 [8]. In summary, the collected parameters are distress type and level of severity, section width, number of exiting manholes. All these

collected data are then used to estimate the existing Pavement Condition Index (PCI) for each section. Once the PCI for each section is estimated, further statistical analyses were performed to see the correlation between the PCI and the inputs.

The third step in this research was to develop an Artificial Intelligence approach (namely ANN model) that can be used to predict the PCI value for pavement sections using the inputs described earlier. This approach is used because the use of many input variables that are categorical ones (distress type and severity), which makes the development of classical statistical models infeasible. Several ANN architectures will be evaluated to identify the near-optimum architecture that is generalized and accurate enough to predict the PCI performance based on the pavement sections' characteristics (i.e., physical inputs associated with the pavement sections). Fig. 1 summarizes the proposed methodology. Meanwhile, Section 5 provides more insights on the data collection process. Section 6 provides more details regarding the development of the ANN model.

5. Data collection and Processing

5.1. Field data collection

The data were collected from ten different roads located in the City of Nablus, Palestine. Each road was divided into several directional segments, where each segment is 100 m in length. The directional pavement sections from these different roads were visually inspected and evaluated. Those roads vary in terms of width and length. Table 2 summarizes the characteristics of these sites. In total, 34,800 m (348 segments of 100 m each) were investigated, covering an overall area of approximately 248,000 m-squared. Fig. 2 illustrates a map of Nablus City, showing the locations of the studied segments.

For the total 348 investigated segments, the segment width, number of manholes, and distress area per type and severity were collected. Out of the possible 19 type of distress listed in Table 1, only one distress (Railroad Crossing) was not presented in the investigated segments. This is due to the nonexistence of railroads in the City of Nablus. Each distress type was ranked with low, medium and high severity. Table 3 summarizes the collected data for each segment. Those fields are then used as inputs to train and develop the ANN-based prediction model. Finally, the PCI is calculated for each paved road section as per the ASTM 6433–07 method.

5.2. Descriptive analysis of collected data

The collected data were qualitatively analyzed in order to identify any potential trends. The first step was to identify the distresses likelihood, by determining distresses that have the highest occurrence (or probability), which would help later in identifying proper preventive measures and budget allocation. It was found that approximately 22.5% of the overall pavement sur-

Table 1
List of all considered distresses.

The Different Types of Distresses				
Cracking	(B) Patching and Potholes	(C) Surface Deformation	(D) Surface Defects	Group (E) Miscellaneous Distresses
Longitudinal and Transverse Cracks	Patching	Depressions	Polished Aggregate	Lane Shoulder Drop
Alligator Cracking	Landing	Corrugation	Bleeding	Rail-Road Crossing*
Edge Cracking	Potholes	Rutting	Raveling	
Reflection Cracks		Shoving		
Block Cracking		Swelling		
Slippage Cracks				

* This distress was not observed in this study, due to the lack of railroads in The City of Nablus.

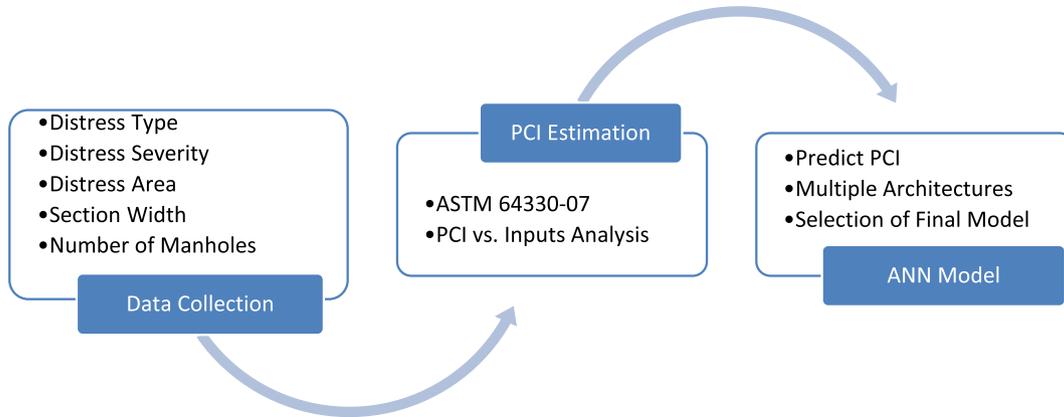


Fig. 1. Proposed research methodology steps.

Table 2
Sample road geometric configurations.

Street Name (St.)	Direction 1		Direction 2	
	Length	Width Range	Length	Width Range
Al-Makhfiya St.	1,000	4.3–5.5	1,000	4.3–5.5
Amman St.	1,200	7.0–8.5	1,200	7.0–8.5
Askar St.	1,700	8.5–11.5	1,700	8.5–13.0
Fadwa Tuqan St.	1,300	2.4–5.5	1,300	2.4–5.5
Haifa St.	1,400	10.0	1,400	10.0
An-Najah Hospital St.	2,000	3.6	2,000	3.6
Titi St.	1,000	3.5	1,000	3.5
Tunis St.	1,200	7.0–9.0	1,200	6.0–9.0
Yafa St.	1,100	6.0	1,100	6.0
Yaser Arafat St.	5,500	4.0–12.0	5,500	4.0–11.5
Total Length (m)	17,400	–	17,400	–

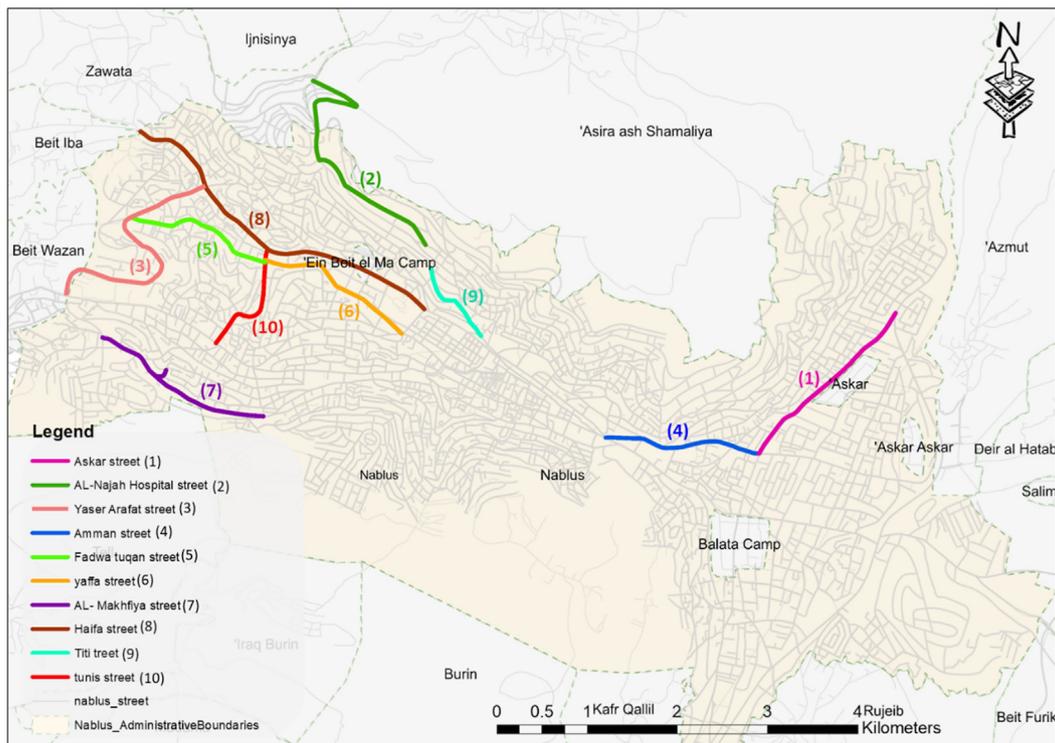


Fig. 2. Study area showing the study roadway segments (Nablus City).

Table 3
Summary of collected data (ann input and output variables).

Variable	Variable range	Variable type
Input Variables		
Distress Type	(Refer to Table 1 for Distress Types)	Nominal
Distress Severity	Low, Medium, High	Nominal
Distress Area	0–850	Numeric
Section Width	2.4–13	Numeric
Number of Manholes	0–10	Numeric
Output Variable		
Pavement Condition Index (PCI)	0–100	Numeric

face area is subject to a given type of distress. Fig. 3 shows the overall distress area per distress type, regardless of the severity level. The figure shows that Patching, Longitudinal and Transverse Cracks, Polished Aggregate, Alligator Cracking and Raveling distresses are responsible for 87.3% of the overall observed distresses. Patching distress was found to be the associated with the highest occurrence (with 24%). On the other hand, Raveling distress is found to be lowest of the five common distresses, with a percentage of 10%. Note that the existence of the other type of distresses is less common compared to the top five distresses. For instance, Rutting is ranked sixth, but was only covering a percentage of 3.7%.

Fig. 4 on the other hand classifies the distresses according to the severity level for the distresses with the highest occurrences, (namely Patching, Longitudinal and Transverse Cracks, Polished Aggregate, Alligator Cracking and Raveling). In general, distresses with high severity do not seem to describe the majority of distressed areas, with the exception of Patching. For the other four distresses, the medium level tends to be the dominant.

Fig. 5 on the other hand provides an idea about the number of manholes per segment once normalized over the width to account for width differences. As can be seen in the figure, most of the investigated segments do not have any manhole. While some other segments have five manholes or more. Since the manholes are often constructed after the construction of the roads, they are often associated with reducing the pavement quality, especially when improper construction procedures are followed with very minimal quality control. Thus, low PCI scores is expected. The relationship between number of manholes and PCI values are shown in Fig. 6. The figure shows a general trend of lower PCI values to be associated with more manholes per section.

Fig. 7 distribution of segments according to their PCI score. The figure shows that approximately 50% of the investigated segments are associated with a PCI value of 50 or high, with a median PCI score value of 54.

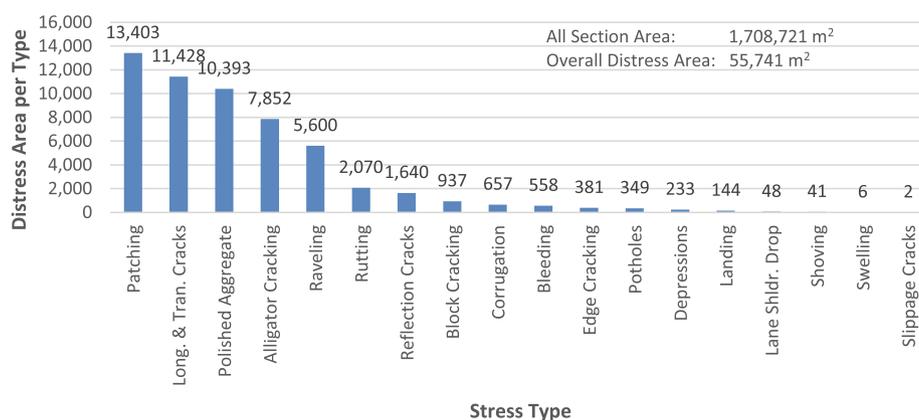


Fig. 3. The total distressed area per distress type.

5.3. Correlation between distresses and PCI values

A correlation matrix is calculated between the different sets of input variables and the outputs (PCI values). Due to the relatively high number of inputs and their nature, the correlation matrix is visually shown in Fig. 8, in the form of a heat map. The figure is coded with color scheme, where the dark red color indicates the highest possible positive correlation value (+1.0), while the dark blue indicates the lowest observed negative correlation value of approximately -0.40 , with a gradual coloring scheme in between. The green color indicates a correlation value of zero.

Fig. 8 shows that in general, none of the inputs is highly correlated with the PCI. For instance, it was noted that the highest positive correlation between any of the inputs and the PCI was the width of the section, with a correlation coefficient of 0.0742. Nonetheless, it is an indication of weak correlation between the section width and the PCI. On the other hand, the highest negative correlation (i.e., lowest correlation) was between High Severity Alligator Cracks, with a negative correlation value of -0.3758 . It can also be concluded that most of the inputs are not correlated with each other, the only exception is Block Cracking with Mediums Severity and Raveling with High Severity. The lack of high correlation between the inputs indicate their independencies from each other, despite the fact that all of them are used to indicate pavement performance.

6. Artificial Neural Networks (ANN)

6.1. ANN modeling

A feedforward backpropagation ANN typically consists of three layers [15,30,31,35]. The first layer is called the Input Layer, which is a vector that represents all the input variables. Usually, the inputs are normalized before being fed into the input layer. This normalization is used to make sure that the ANN is unbiased, where all inputs will have the same range once normalized. The second layer is called the hidden layer, where the neurons belong to this layer collect the transformed signals from the input layer and transfer them to the output layer. Finally, the third layer is the output layer, and it collects all the transformed signals from the hidden layer and then processes those signals into the output vector. In each iteration or epoch, the weights and biases of each neuron is adjusted based on several parameters, such as the learning algorithm and learning rate. The objective is to minimize the cost function.

As indicated earlier, the total number of records was 348 records. Those records were randomly assigned to one of three datasets, as follow:

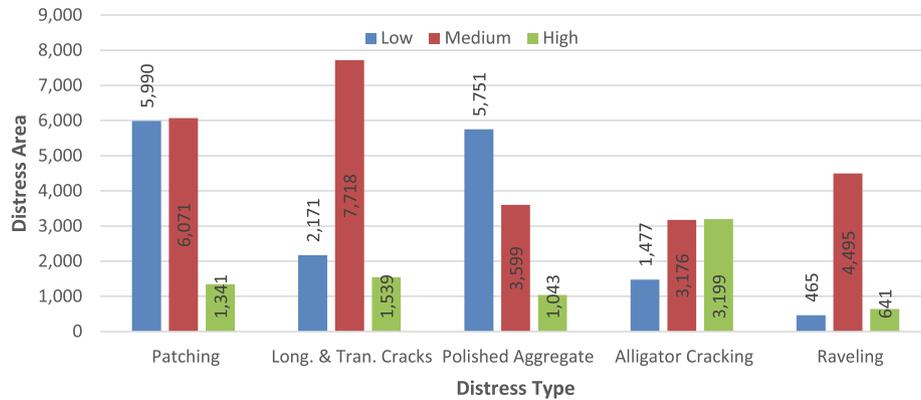


Fig. 4. Severity distribution of the top five distresses.

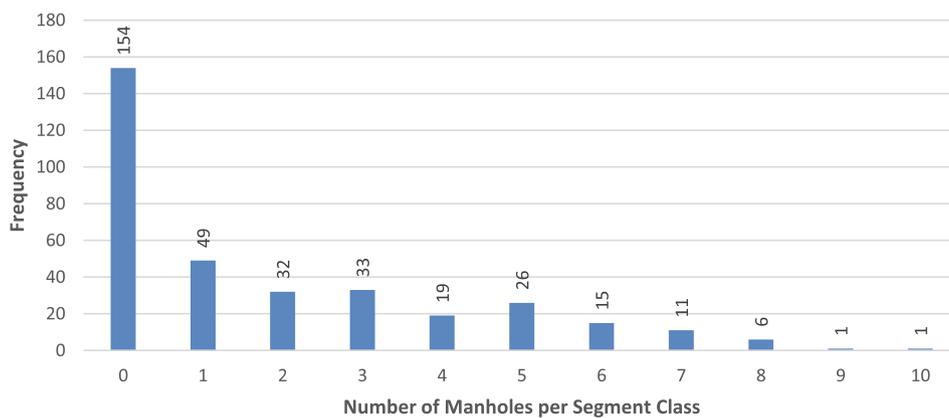


Fig. 5. Number of segments against number of manholes per segment.

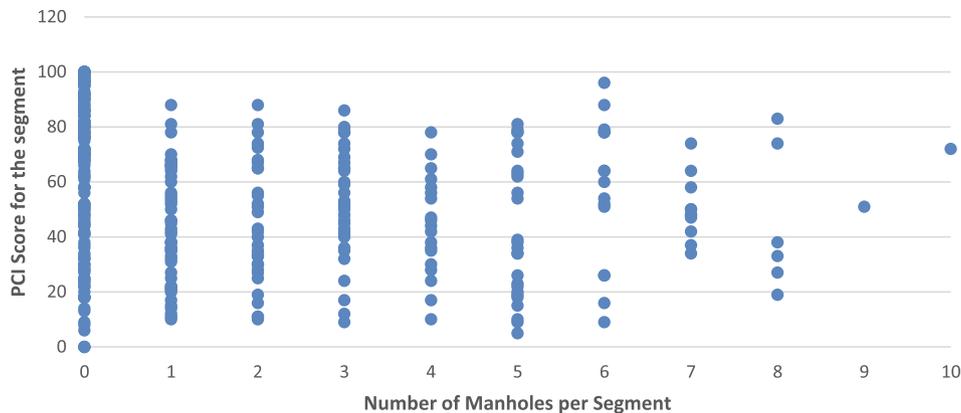


Fig. 6. Relationship between number of manholes and PCI score.

Training Dataset: This dataset is used for the ANN learning (or training) process, by adjusting the weight and bias vectors to minimize the differences between the outputs and the targets. The size of this group is 70% of the collected data, which contains 244 records.

Validation Dataset: This dataset is used to monitor the convergence of ANN learning process, and it is often used to avoid overfitting so that the ANN model is applicable to new inputs beyond the ones used in training or validating the ANN model. This dataset used 15% of the collected data, with 52 records.

Testing Datasets: This dataset is used to check the performance of the trained ANN once completed. It is completely independent

from the first two datasets (training and validation). The remaining records of the collected data (that were not used in training or validation datasets) were assigned to this group, with a total of 52 records that represents 15% of the collected data.

6.2. ANN architecture and parameters

The ANN architecture is generally characterized by its structure and its associated parameters. In terms of the layers, a typical ANN consists of one input layer, one hidden layer and one output layer. The number of neurons in each of the input and the output layer depends on the number of input and output signals. The perfor-

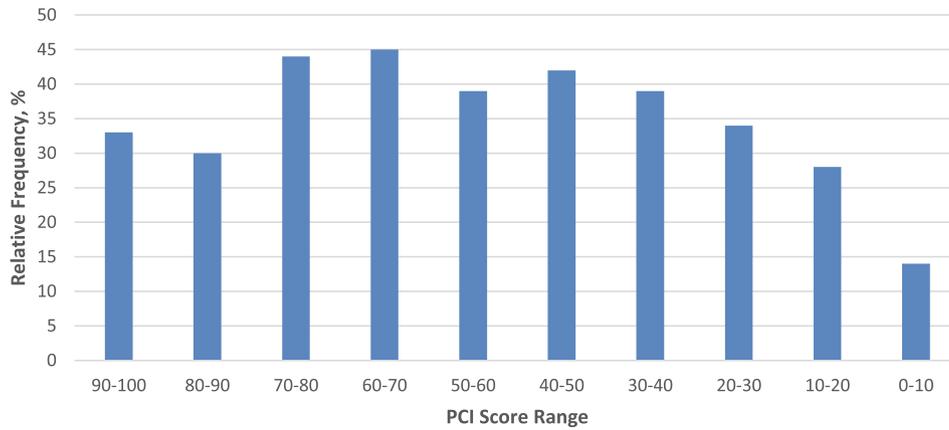


Fig. 7. Relative frequency of PCI Distribution.

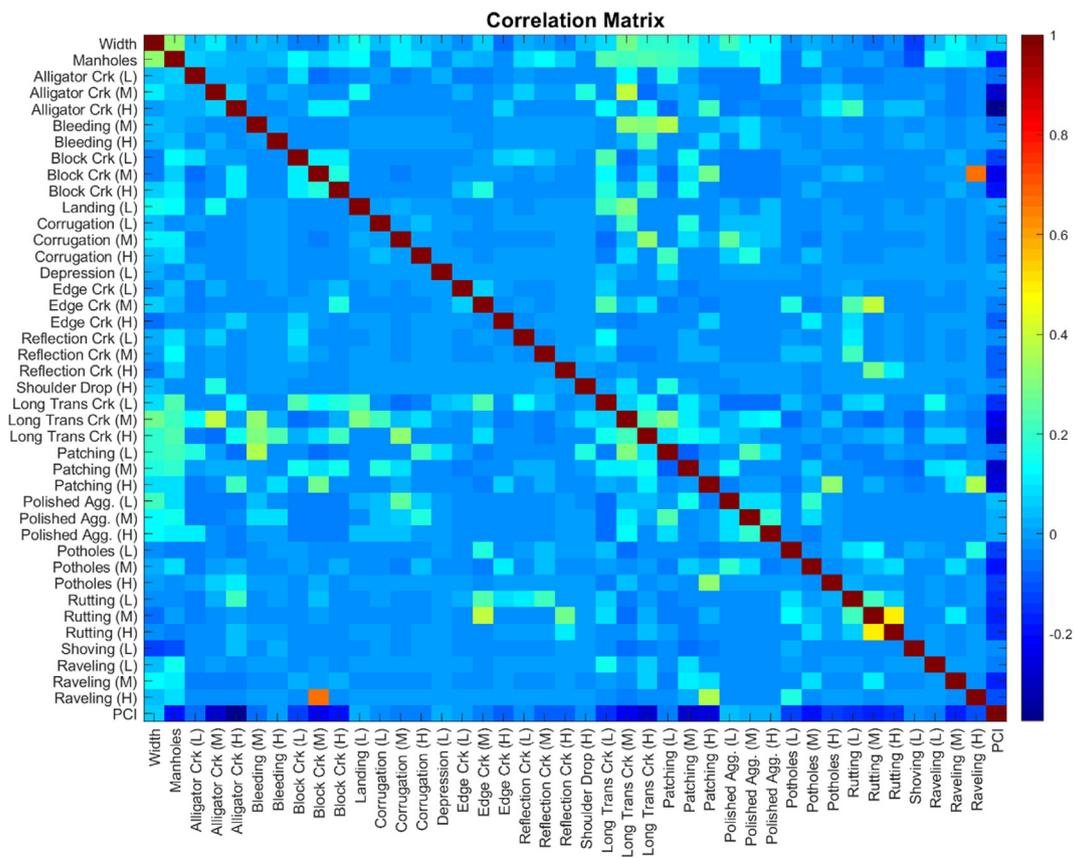


Fig. 8. Correlation matrix visualization between inputs and output variables.

mance of ANN is highly influenced by several attributes, including the number of neurons in the hidden layer, transfer functions and learning algorithms. Determining the characteristics of these attributes, however, depend on several factors, such as the sample size, the number and the type of inputs and outputs. For instance, one categorical variable can be presented with a number of dummy variables that are binary coding. If one categorical variable has four possible outcomes, then this one variable must be coded using four dummy variables (or three dummy variables in certain conditions). Each one of those dummy variables has a value of either 0 or 1, to represent the categorical value. Therefore, the four input variables (i.e., 19 stress type, 3 stress severity, section width and number of manholes per section) were coded with 41 neurons in the input layer after getting rid of any redundant variables.

The use of high number of neurons in the hidden layer does not necessarily mean that the ANN will perform better than those networks with lower number. Such an increase might potentially lead to overfitting (i.e., under-characterized performance) or creating the computational needs to optimize more parameters than what can be constrained by the input vectors. While there is no specific rule to follow in order to determine the number of neurons in the hidden layer, the common practice suggests that the size of the hidden layer (i.e., number of neurons in the hidden layer) is guided by the following guidelines:

- The number of neurons in the input layer.
- Two-third the number of neurons in the input layer and the number of neurons in the output layer.

- The ratio of the sample size to the number of neurons in the input and output layers.

With respect to the transfer functions, they are used to transfer the normalized inputs (i.e., signals) from one layer to another. A signal is transferred from the input to the hidden layers, and from the hidden to the output layers using those function, so that all the signals are normalized and unbiased. The three common transfer functions are: tan-sigmoidal (*tansig*), log-sigmoidal (*logsig*) and linear (*purelin*). Equation 2 to Equation 4 show the formulas for each of the transfer functions.

$$tansig(x) = 2/(1 + e^{-2x}) - 1 \tag{2}$$

$$logsig(x) = 1/(1 + e^{-x}) \tag{3}$$

$$purelin(x) = x \tag{4}$$

Another important attribute is the training algorithm. These algorithms control how the weights of the signals are adjusted as the learning process is going over the training datasets. In this research, several learning algorithms were examined. However, Levenberg-Marquardt backpropagation (*trainlm*) and Bayesian regularization backpropagation (*trainbr*) were found to provide realistic results. More details about those training algorithms can be found in [32], as comparing the fundamentals of these training algorithms is beyond the scope of this research.

Several ANN architectures and learning algorithms were evaluated in this research, and the result for each ANN model is shown in Table 4, which shows that the performance of the ANN model is highly influenced by the suggested architecture (i.e., the number of neurons in the hidden layer, the learning algorithm, and the transfer function).

Fig. 9 is showing a schematic sketch representing the architecture of the selected ANN model that outperforms the other models. It has one input layer that has 41 inputs, one hidden layer with 10 neurons and one output layer as presented in the figure. The training function used in the selected model is the Levenberg-Marquardt backpropagation. The *tansig* transfer function is used to transfer the signal x from the input layer to the hidden layer. On the other hand, the *purelin* transfer function transfers the signal x from hidden layer to the output layer.

6.3. ANN final model results

The training process was terminated after reaching the maximum number of epochs, which was set to 500 epochs. The results for all datasets and the three groups (i.e., training, validation and testing datasets) are shown in Fig. 10. The results show a linear trend, with a slope that is very close to 1.0 for all the three groups, which is an indicator that the model is capable of predicting the PCI value with a high level of accuracy. The results also show that there were only two instances where the ANN failed to predict the PCI correctly. One instance occurred through the training process and another instance was within the testing dataset. Further investigation to those two instances revealed that their inputs were not well presented in the collected data.

To further assess the performance of the developed ANN model, a set of performance measures were identified based on the errors in predicting the pavement PCI. These performance measures are mean-squared error (*MSE*), standard error (*SE*), root mean-squared percentage error (*RMSPE*), and coefficient of determination (R^2). Equations 4 to Equations 7 shows the mathematical formula for each of the performance measures based on the observed values or the target vector (t) and the predicted values vector, (y). The results of the selected performance measures are summarized in Table 5. The results indicate that the ANN was able to predict the PCI, with a standard error of approximately 25% across all groups. The results are promising despite the fact that there were several qualitative (or nominal) variables to consider.

$$MSE = \frac{\sum (y - t)^2}{N} \tag{4}$$

$$SE = \sqrt{\frac{\sum (y - \bar{t})^2}{N}} \tag{5}$$

$$RMSPE, \% = 100 \sqrt{\frac{\sum (\frac{y-t}{t})^2}{N}} \tag{6}$$

$$R^2 = 1 - \frac{\sum (t - \bar{t})^2}{\sum (y - \bar{t})^2} \tag{7}$$

Table 4
List of performance results of different ANN models.

ANN Architecture	Input-to-Hidden Transfer	Hidden-to-Output Transfer	ANN Network Performance	
			Training Algorithm	
			<i>trainlm</i>	<i>trainbr</i>
[41 -3 -1]	<i>tansig</i>	<i>purelin</i>	216.21	282.71
[41 -4 -1]	<i>tansig</i>	<i>purelin</i>	98.17	110.80
[41 -5 -1]	<i>tansig</i>	<i>purelin</i>	64.07	78.55
[41 -6 -1]	<i>tansig</i>	<i>purelin</i>	160.40	67.78
[41 -7 -1]	<i>tansig</i>	<i>purelin</i>	115.35	58.10
[41 -8 -1]	<i>tansig</i>	<i>purelin</i>	41.45	41.38
[41 -9 -1]	<i>tansig</i>	<i>purelin</i>	46.43	41.21
[41 -10 -1]	<i>tansig</i>	<i>purelin</i>	25.61	38.36
[41 -11 -1]	<i>tansig</i>	<i>purelin</i>	28.88	36.73
[41 -12 -1]	<i>tansig</i>	<i>purelin</i>	70.65	134.26
Same Hidden Layer Size with Different Transfer Functions				
[41 -10 -1]	<i>tansig</i>	<i>purelin</i>	25.61	38.36
[41 -10 -1]	<i>tansig</i>	<i>logsig</i>	681.82	390.84
[41 -10 -1]	<i>tansig</i>	<i>tansig</i>	39.43	42.01
[41 -10 -1]	<i>logsig</i>	<i>purelin</i>	52.27	40.38
[41 -10 -1]	<i>logsig</i>	<i>logsig</i>	681.82	470.39
[41 -10 -1]	<i>logsig</i>	<i>tansig</i>	126.08	82.78
[41 -10 -1]	<i>logsig</i>	<i>purelin</i>	292.65	292.53
[41 -10 -1]	<i>purelin</i>	<i>logsig</i>	681.82	681.82
[41 -10 -1]	<i>purelin</i>	<i>tansig</i>	279.23	261.41

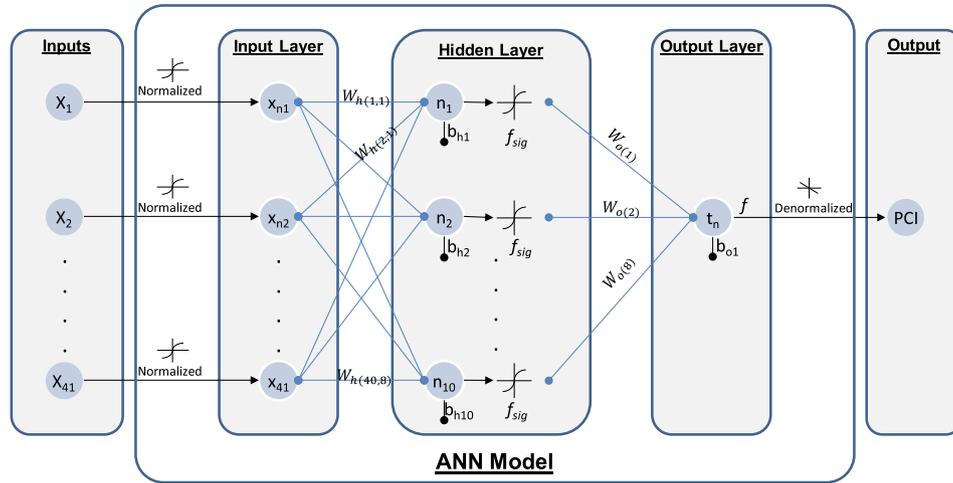


Fig. 9. Schematic architecture of the developed ANN model.

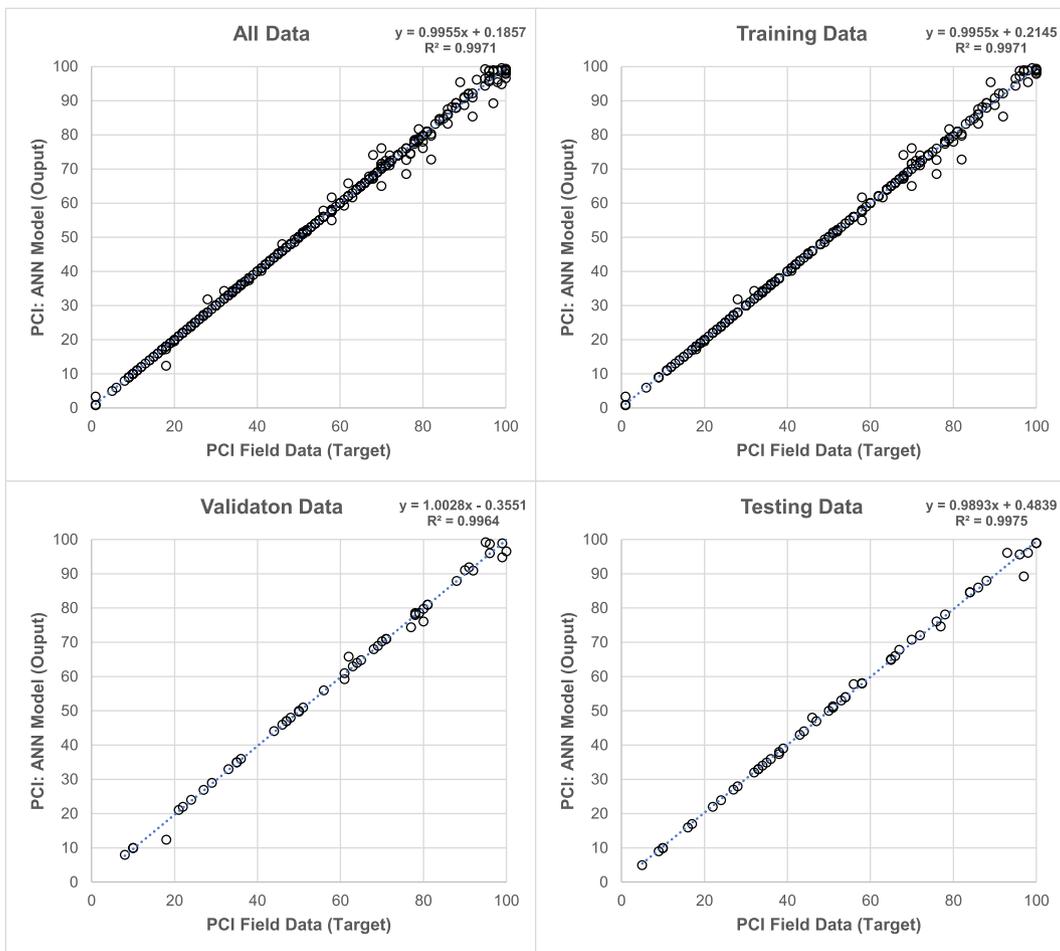


Fig. 10. Artificial neural networks model end results.

Table 5

Summary of the MOEs for the different ANN groups.

Data Source	N	MSE	SE	RMSPE	R ²
All Data	348	2.16	26.04	12.9%	0.9971
Training Dataset	244	2.2	25.8	15.3%	0.9971
Validation Dataset	52	3.6	26.3	4.6%	0.9964
Testing Dataset	52	2.7	26.1	1.6%	0.9975

7. Conclusions and future Perspectives

The PCI is used to assess and evaluate the existing pavement conditions based on distress density and severity as they observed in the field. The calculation required high level of technical knowledge and experience which might not be available in the developing countries. The calculation of the PCI using the conventional method relies mainly on visual inspection method, which identifies the existing pavement distresses and quantify the type, severity and extend of distresses. Moreover, the lack of a simplified approach to estimate PCI values is always a challenge due to complex relationship between inputs and PCI.

This paper illustrates the development of an artificial intelligence approach to predict the Pavement Condition Index (PCI). Unlike the conventional method, an AI approach is proposed to model the complex relationship between the PCI and other measured variables including distress types, severity and areas, along with the number of manholes within the section.

A database was created by visually collecting and recording distresses characteristics of 34.8 km of road segments located within the municipal boundary of the City of Nablus, Palestine. This database is then used to develop an Artificial Neural Networks (ANN) model that is capable of predicting the PCI values based on a set of inputs that characterize the roadway segment. The tested model shows a linear relationship between observed and predicted PCI values with a slope that is closed to 1.0, which indicates an accurate and reliable prediction model. The coefficient of determination for this relationship was 0.9971.

Although the results indicate that the proposed ANN model can predict the PCI values with high level of accuracy, the authors recommend the following:

- The use of the proposed model would ultimately provide municipalities and local agencies with a tool that is capable of evaluate pavement conditions.
- This tool can also be used to perform what-if analysis and then prioritize pavement rehabilitation programs and budget allocation.
- The ANN model can be integrated within an image processing platform that would automatically detect and identify pavement defects and predicted the PCI value. Ultimately, this integration is expected to reduce the human efforts and errors when it comes to quantify the pavement conditions or defects.
- The ability to use other variables (such as traffic demand, trucks presence and weather data) as part of predicting the PCI values, [34].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors would like to thank An-Najah National University for funding the research. The reference number is ANNU-1920-Sc003.

References

- [1] Shah YU et al. Development of overall pavement condition index for urban road network. *Procedia-Soc Behav Sci* 2013;104(2):332–41.
- [2] Huang YH. *Pavement Analysis and Design*. Englewood Cliffs, NJ: Prentice-Hall; 1993
- [3] Karim FM, Rubasi KAH, Saleh AA. The Road Pavement Condition Index (PCI) evaluation and maintenance: a case study of Yemen. *Org Technol Manag Construction Int J* 2016;8(1):1446–55.
- [4] Yoder E, Witczak M. *Principles of Pavement Design*. New York: Wiley; 1975.
- [5] Hall JW., et al. Performance Indications from Army Airfield Pavement Management Program, in *Pavement Management Implementation*. 1991, ASTM International.
- [6] Hajj EY et al. Optimum Time for Application of Slurry Seal to Asphalt Concrete Pavements. *Transp Res Rec* 2011;2235(1):66–81.
- [7] Galehouse L, Moulthrop JS, Hicks RG. Principles of pavement preservation: definitions, benefits, issues, and barriers. *TR News* 2003;228.
- [8] American Society for Testing and Materials, ASTM-E-647-15-e1: Standard Test Method for Measurement of Fatigue Crack Growth Rates, in *Annual book of ASTM standards*. 2011. p. 591–630.
- [9] Mousa E, El-Badawy S, Azam A. Effect of reclaimed asphalt pavement in granular base layers on predicted pavement performance in Egypt. *Innovative Inf Solut* 2020;5:1–18.
- [10] Arab MG et al. Combined effect of compaction level and matric suction conditions on flexible pavement performance using construction and demolition waste. *Constr Build Mater* 2020;261:119792.
- [11] Azam AM, El-Badawy SM, Alabasse RM. Evaluation of asphalt mixtures modified with polymer and wax. *Innovative Inf Solut* 2019;4(1):1–12.
- [12] Tarbay EW, Azam AM, El-Badawy SM. Waste materials and by-products as mineral fillers in asphalt mixtures. *Innovative Inf Solut* 2019;4(1):1–13.
- [13] Ceylan H, Bayrak MB, Gopalakrishnan K. Neural networks applications in pavement engineering: a recent survey. *Int J Pavement Res Technol* 2014;7(6).
- [14] Zeiada W et al. Machine Learning for Pavement Performance Modelling in Warm Climate Regions. *Arab J Sci Eng* 2020:1–19.
- [15] Ghanim MS, Abu-Lebdeh G. Real-time dynamic genetic signal priority optimization for coordinated traffic networks using genetic algorithms and artificial neural networks. *J Intell Transport Syst* 2015;19(5):327–38.
- [16] Miller JS, Bellinger WY. *Distress Identification Manual for the Long-Term Pavement Performance Program (Fifth Revised Edition)*. May 2014, Federal Highway Administration. Office of Infrastructure Research and Development: United States.
- [17] Eldin NN, Senouci AB. A pavement condition-rating model using backpropagation neural networks. *Comput-Aided Civ Infrastruct Eng* 1995;10(6):433–41.
- [18] Yang J, Lu JJ, Gunaratne M. Application of neural network models for forecasting of pavement crack index and pavement condition rating. 2003.
- [19] Shahnazari H et al. Application of soft computing for prediction of pavement condition index. *J Transp Eng* 2012;138(12):1495–506.
- [20] Amin MSR, Amador-Jiménez LE. Pavement management with dynamic traffic and artificial neural network: a case study of Montreal. *Can J Civ Eng* 2016;43(3):241–51.
- [21] Jalal M, Floris I, Quadrifoglio L. Computer-aided prediction of pavement condition index (PCI) using ANN. in *Proceedings of the international conference on computers and industrial engineering*, 2017.
- [22] Arhin SA et al. Predicting pavement condition index using international roughness index in a dense urban area. *J Civil Eng Res* 2015;5(1):10–7.
- [23] Elhadidy AA, El-Badawy SM, Elbeltagi EE. A simplified pavement condition index regression model for pavement evaluation. *Int J Pavement Eng* 2019:1–10.
- [24] Mazari M, Rodriguez DD. Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. *J Traffic Transport Eng (English Edition)* 2016;3(5):448–55.
- [25] Majidifard H, Adu-Gyamfi Y, Buttler WG. Deep Machine Learning Approach to Develop A New Asphalt Pavement Condition Index. *Constr Build Mater* 2020;247:118513.
- [26] Vyas V, Singh AP, Srivastava A. Prediction of asphalt pavement condition using FWD deflection basin parameters and artificial neural networks. *Road Mater Pavement Des* 2020:1–19.
- [27] Shahin MY. *Paver Asphalt Distress Manual*. 1997, CONSTRUCTION ENGINEERING RESEARCH LAB (ARMY) CHAMPAIGN IL.
- [28] Shahin, MY, Walther, JA. *Pavement Maintenance Management for Roads and Streets Using the PAVER System*. 1990, CONSTRUCTION ENGINEERING RESEARCH LAB (ARMY) CHAMPAIGN IL.
- [29] Shahin MY, Kohn SD. Overview of the PAVER Pavement Management System and Economic Analysis of Field Implementing the PAVER Pavement Management System. 1982, CONSTRUCTION ENGINEERING RESEARCH LAB (ARMY) CHAMPAIGN IL.
- [30] Ghanim MS, Shaaban K. Estimating Turning Movements at Signalized Intersections Using Artificial Neural Networks. *IEEE Trans Intell Transp Syst* 2018;20(5):1828–36.
- [31] Ghanim M, Abu-Lebdeh G. Projected statewide traffic forecast parameters using artificial neural networks. *IET Intel Transport Syst* 2018.
- [32] Haykin S. *Network neural: a comprehensive foundation*. neural networks. Vol. 2. 2004: Macmillan.
- [33] ISSA A. Development of Pavement Performance Model for Proper Rehabilitation and Maintenance using First Order Markov Chain. In: *The Tenth International Structural Engineering and Construction Conference* May 20–25. Chicago, Illinois, United States: International Structural Engineering Society; 2019.
- [34] Cunagin, W., R.L. Reel, M. Ghanim, et al. Generating Site-Specific Axle Load Factors for Florida's MEPDG Implementation. in 92nd Annual Meeting of the Transportation Research Board. 2013. Washington DC.
- [35] Ghanim, M.S. Florida Statewide Design-Hour Volume Prediction Model. in Transportation Research Board 90th Annual Meeting. 2011.