Prediction of surface distress using neural networks
Hamdi, Sigit P. Hadiwardoyo, A. Gomes Correia, Paulo Pereira, and Paulo Cortez

Citation: AIP Conference Proceedings 1855, 040006 (2017); doi: 10.1063/1.4985502
View online: http://dx.doi.org/10.1063/1.4985502
View Table of Contents: http://aip.scitation.org/toc/apc/1855/1
Published by the American Institute of Physics

Articles you may be interested in
New optimization strategies of pavement maintenance: A case study for national road network in Indonesia using integrated road management system
AIP Conference Proceedings 1855, 040015 (2017); 10.1063/1.4985511
Prediction of Surface Distress Using Neural Networks

Hamdi¹,³,a, Sigit P. Hadiwardoyo¹,b, A. Gomes Correia²,c, Paulo Pereira³,d and Paulo Cortez⁴,e

¹Department of Civil Engineering, Faculty Technic, Universitas Indonesia Depok 16242, Indonesia
²Department of Civil Engineering, ISISE, School of Engineering, University of Minho, Azurém 4800-058 Guimarães, Portugal.
³Department of Civil Engineering, CTAC, School of Engineering, University of Minho, Azurém 4800-058 Guimarães, Portugal.
⁴Department of Information Systems/ALGORITMI R&D Center, University of Minho, Azurém 4800-058 Guimarães, Portugal.

a)Corresponding author: hamdi_pmu@yahoo.com
b)sigit@eng.ui.ac.id
c)agc@civil.uminho.pt
d)ppereira@civil.uminho.pt
e)pcortez@dsi.uminho.pt

Abstract. Road infrastructures contribute to a healthy economy throughout a sustainable distribution of goods and services. A road network requires appropriately programmed maintenance treatments in order to keep roads assets in good condition, providing maximum safety for road users under a cost-effective approach. Surface Distress is the key element to identify road condition and may be generated by many different factors. In this paper, a new approach is aimed to predict Surface Distress Index (SDI) values following a data-driven approach. Later this model will be accordingly applied by using data obtained from the Integrated Road Management System (IRMS) database. Artificial Neural Networks (ANNs) are used to predict SDI index using input variables related to the surface of distress, i.e., crack area and width, pothole, rutting, patching and depression. The achieved results show that ANN is able to predict SDI with high correlation factor (R² = 0.996%). Moreover, a sensitivity analysis was applied to the ANN model, revealing the influence of the most relevant input parameters for SDI prediction, namely rutting (59.8%), crack width (29.9%) and crack area (5.0%), patching (3.0%), pothole (1.7%) and depression (0.3%).

INTRODUCTION

Road infrastructures contribute to a healthy economy throughout a sustainable distribution of goods and services. Every country prioritizes highway pavement projects according to limited budgets using pavement distress rating in one form or another [1]. A road network requires appropriately programmed maintenance treatments in order to keep roads assets in good condition, providing maximum safety for road users under a cost-effective approach.

Road management would be supported by a robust strategy comprising monitoring and evaluation of pavement distress so as to maintain an adequate condition of roads [2]. The main surface distresses found in flexible pavement could include: cracking, pothole, rutting, patching, depression, distortion, disintegration, polished aggregate, blooding or flushing. These different distresses are analyzed together in order to produce performance indicators to evaluate pavement condition, essential to define type and priority of treatment.
PURPOSE OF STUDY

In a view to identify and prioritize a program maintenance treatment, road deterioration is used to evaluate and forecast pavement condition during the analysis period. This paper presents a new approach to predict surface distress of pavement based on performance indicator Surface Distress Index (SDI), following a data-driven approach. The Directorate-General of Highways (DGH) has been applying SDI comprising several variables: crack area and crack width, pothole and rutting. This paper proposes a new approach to predict SDI by using additional components, namely patching and depression.

LITERATURE REVIEW

Currently, using a single model to produce road condition models for all types of the road has been widely used by many researchers [3,4]. Although a single model may produce reasonable results, but it may not obtain the best solution due to the differences nature condition or variability of road performance [5]. This paper adopted a combined performance indicator (SDI) to analyze and evaluate the road condition. The surface distress index can be defined as the value of damage due to changing of pavement condition. SDI, as a measure of highway system condition is widely used in monitoring, planning and rehabilitation [6]. The method of collection is conducted through different approaches, including manual, infrared and laser. SDI in Pennsylvania integrates seven distresses [5]: Excess Asphalt, Block Cracking, Raveling and Weathering, Transverse and Longitudinal Cracking, Widening Drop-off, Edge Deterioration, and Rutting. Development of road deterioration model has been carried out by many researchers [7-13]. Reference [14] proposed several measurements of performance flexible pavement. ANNs are the most widely used for determination of road condition. Reference [15] applied ANN and Support Vector Machines as methods for determining the visual condition of roads on inventory and traffic data set with 16 continuous-valued and 2 nominal-valued dimensions.

Data mining (DM) is the scientific field that analyzes raw data in order to extract valuable knowledge [16,17]. DM has been widely successfully applied in several areas, including the Civil Engineering domain, used in various studies around the world. DM is a set of processes to explore additional value in the form of information that had been unknown manually from. For instance, DM has been used for the earthworks optimization by a new genetic programming model, showing the advantages in comparison with traditional methods [18]. DM has also been used to predict individual wine taste preferences based on physic-chemical analytical tests performed at the wine certification step [19]. Another application example of DM is the prediction of the risk of organ failure in intensive care units [20].

METHOD

DM is an emerging area that lies at the intersection of statistics, artificial intelligence and data management. There are several DM techniques, each one with its own purposes and advantages. In the process of executing this DM project, several computational tools were adopted, including Microsoft Excel for the storage of distress data, in term of the CSV (comma separated value) format, while the data-driven modeling was executed using the R tool (http://www.r-project.org). An example of the adopted Artificial Neural Network (ANN), used during the modeling phase, is illustrated in Fig. 1. The ANN was set with an intermediate layer composed of 3 neurons, which led to the best modeling results. All neurons of a given layer are connected to all neurons in the subsequent layer. A single output layer node is used to compute the SDI value. The data was divided into segments I and II. The adopted ANN was evaluated using the holdout 2/3 method. ANN was selected as a reference model and to get robust results, 20 runs of the holdout scheme were executed. The hold out is usually used to estimate the generalization capability of a model [19]. This method divides on two partitions of the data into training and validation test. The previous training data is used to fit the model (2/3 of the data), while the latter (with the remaining 1/3 of data) is used to compute the quality of the prediction estimates.

IRMS Data

The road network from which data supporting this research was collected is in Jambi province, based on the assumption that the corridors are considered as a primary road network system. This research has been performing data from IRMS. The data of survey from DGH were obtained for 824 road sections. The data is related with 3
corridors, i.e., Eastern, Center and Connecting Corridor in Jambi province and to train the ANN, six input variables were considered as shown in Table I. These corridors are divided into 2 Segment Data Set, namely; Segment I and II.

![Figure 1: Artificial Neural Network Model adopted](image)

**TABLE I.** Main Statistics of The Numeric Input Variables Each Corridor

<table>
<thead>
<tr>
<th>Corridor</th>
<th>SDI</th>
<th>Area Crack (%)</th>
<th>Width Crack (mm)</th>
<th>Pothole (%)</th>
<th>Rutting (mm)</th>
<th>Depression (mm)</th>
<th>Patching (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern</td>
<td>20.96</td>
<td>4.96</td>
<td>1.73</td>
<td>3.97</td>
<td>0.00</td>
<td>0.53</td>
<td>2.60</td>
</tr>
<tr>
<td>Center</td>
<td>31.62</td>
<td>12.08</td>
<td>2.24</td>
<td>2.90</td>
<td>0.51</td>
<td>0.59</td>
<td>2.28</td>
</tr>
<tr>
<td>Connecting</td>
<td>26.43</td>
<td>9.72</td>
<td>2.05</td>
<td>470</td>
<td>0.24</td>
<td>0.58</td>
<td>1.60</td>
</tr>
</tbody>
</table>

**Development of Data Mining Model**

ANN is powerful learning method with widespread applications in many fields [16]. In this study, an ANN was applied as the most standard of NN type, the multilayer perceptron. In this model, the equation is adopted from [17] as in Equation (1). ANN in this class consists of H processing units as one hidden layer, which uses a feed forward multilayer perception [21,22].

\[
\hat{y} = W_{o0} + \sum_{j=1}^{I-1} f \left( \sum_{i=1}^{I} x_i \cdot W_{ij} + W_{jo} \right) \cdot W_{oi}
\]

(1)

where \(W_{o0}\) is a zero weight represents no connection between two neurons and negative weight represents a prohibited relationship, \(W_{ij}\) is the weight of the connection from neuron \(j\) to the unit \(i\) (if \(j = 0\), threshold for node \(j\) or a bias connection), \(o\) the output unit, \(f\) is logistic function \(\frac{1}{1+e^{-x}}\), and \(I\) is the number of input neurons [11]. To obtain an accurate model in this study, the ANN hyper parameter (H) is optimized using a grid search of H (4, 8,12). For each H value, a NN is trained and its generalization estimate is measured. The database divided into Segment I and Segment II dataset in order to avoid over fitting. At Segment I data set comprises 607 random sections (74%) of total data set. After Segment I stage, the remaining 217 sections (26%) are used for Segment II.
Performance Indicator

A performance indicator can be defined in the form of technical parameters as a rule dimensional [23]. Performance indicator consists of a single performance indicator (e.g., IRI) or combined performance indicator (e.g., SDI, PSI) [2]. Performance indicator can be used as a pavement prediction model based on a scale of pavement distress. The pavement surface is considered in good condition if the SDI value is inferior or equal to 50 and in bad condition if SDI value is higher than 150. In Indonesia, the road condition is classified according to IRI and SDI rating [2] as shown in Table 2.

<table>
<thead>
<tr>
<th>Surface condition</th>
<th>IRI scale</th>
<th>SDI scale</th>
<th>Type of treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>IRI &lt;= 4</td>
<td>SDI &lt;= 50</td>
<td>Routine maintenance</td>
</tr>
<tr>
<td>Fair</td>
<td>IRI &gt; 4 &amp; IRI &lt;= 8</td>
<td>SDI &gt; 50 &amp; SDI &lt;= 100</td>
<td>Periodic maintenance</td>
</tr>
<tr>
<td>Poor</td>
<td>IRI &gt; 8 &amp; IRI &lt;= 12</td>
<td>SDI &gt; 100 &amp; SDI &lt;= 150</td>
<td>Rehabilitation</td>
</tr>
<tr>
<td>Bad</td>
<td>IRI &gt; 12</td>
<td>SDI &gt; 150</td>
<td>Reconstruction</td>
</tr>
</tbody>
</table>

Assessing the validity of ANN Model

The application of ANN model for pavement performance prediction has been used to predict transportation project [24,25]. The assessing of the interconnection between the nodes and weights is processed onward a perfect relationship to estimate the results [26,27]. Many research studies used the root mean square (RMS) as a standard statistical metric to measure model performance [17,24,27]. If the number of the sample reaches 100 or more, then, by utilizing the calculated root mean square error (RMSE), it can reconstruct the error distribution close to its accuracy and will be even more reliable. The lower values of mean absolute error (MAE) and RMSE correspond to a higher forecasting capacity, while an $R^2$ with a value closer to 1 indicates a model that accounts for the greater proportion of variance. In this study, RMSE, correlation coefficient ($R^2$) and MAE are used for assessing model quality. The corresponding equations are defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_k - \bar{y})^2}{N}}$$

(2)

$$R^2 = \left(\frac{\sum_{i=1}^{N} (y_k - \bar{y})(\bar{y_k} - \bar{y})}{\sqrt{\sum_{i=1}^{N} (y_k - \bar{y})^2 \cdot \sum_{i=1}^{N} (\bar{y_k} - \bar{y})^2}}\right)^2$$

(3)

$$MAE = \frac{\sum_{i=1}^{N} |y_k - \bar{y_k}|}{N}$$

(4)

Where, $y_k$ is desired value, $\bar{y}$ is predicted value, $\bar{y}$ is mean of variables and $N$ is number of sample.

RESULTS AND DISCUSSION

Multivariate Statistical Model

The purpose of any data analysis is to extract information from raw to produce an accurate estimation of an individual variable, single or composed. The most important and shared question is: which statistical relationship between a response variable ($y$) and explanatory variables ($x_i$) can be established. SDI data values are used to try to develop a mathematical model using program Statplus. This study objective is to define the priority type of model to be tried and found all correlation variables with $R^2 = 0.93$, this value is generated based on Regression Statistic.

Based on analysis statistic multivariate above, SDI model obtained the Equation (5) as following:
\[ SDI = \sum_{i=1}^{n} (AC_i \times 1.68) + (WC_i \times 1.13) + (P_i \times 2.91) + (R_i \times 1.06) + (PC_i \times 0.02) + (D_i \times 3.38) - 1.262 \] (5)

where SDI is Surface Distress Index, AC is Area crack, CW is crack width, P is Pothole, R is rutting, PC is Patching and D is depression.

**Data Mining Technique of Segment I**

The summary of database value input variables (minimum, maximum, mean and standard deviation) for Segment I and II are shown in Table III. In this study, algorithm ANN was used to train the dataset of Segment I, obtained from Indonesia Integrated Road Management System (IIRMS) in order to predict the SDI.

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>The Summary of Database Values Input Variables for Segments I and II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Area Crack (segments) I II</td>
</tr>
<tr>
<td>Min.</td>
<td>0.08 0.00</td>
</tr>
<tr>
<td>Sd.</td>
<td>8.31 6.46</td>
</tr>
<tr>
<td>Mean</td>
<td>9.03 6.16</td>
</tr>
<tr>
<td>Max.</td>
<td>40.00 40.00</td>
</tr>
</tbody>
</table>

Based on Fig. 2(a), it can be analyzed that the relationship between predicted versus observed by the ANN model, if the SDI value under than 75, the prediction was more accurate. Similar to Fig. 3(a), SDI is more accurate if it has value under than 75. This is due to the SDI value under 75 is the majority (75%) of the database.

Figure 2(b) shows the relative importance of each input variable, as measured by a painful analysis procedure. Based on this Fig. 2(b) the key variables in SDI prediction present the following weight: rutting with more than 50% (corridors are a primary road, with intense and congested, submitted to truck overloading), width crack 29.9%, area crack 5.0%, patching 3.3%, pothole 1.7% and depression 0.3%. These findings strengthen the empirical data that consider of rutting, width crack and area crack as the most relevant parameters to develop of surface distress index.

![FIGURE 2. (a) Observed versus predicted by ANN model for Segment I. (b) Relative importance of each parameter according to ANN model](040006-5)
The MAE, RMSE and $R^2$, 0.173, 0.012 and 0.996, respectively, confirm the quality of prediction Segment I, as shown in Table 4. Based on this study, the formulation of SDI can be achieved by using the parameters rutting, width crack, area crack, patching, pothole and depression.

**TABLE 4. Performance Measure of Models Segment I and Segment II**

<table>
<thead>
<tr>
<th>Model</th>
<th>MAE</th>
<th>RMSE</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment I</td>
<td>0.173</td>
<td>0.012</td>
<td>0.996</td>
</tr>
<tr>
<td>Segment II</td>
<td>0.492</td>
<td>0.017</td>
<td>0.954</td>
</tr>
</tbody>
</table>

**FIGURE 3.** (a) Box Plot Graphic of SDI Segment I (b) Box Plot Graphic of SDI Segment II

**Data Mining Technique of Segment II**

After training stage of Segment I, the remaining 217 sections (26%) were used for Segment II. The input variables for a Segment I was similar to Segment II, i.e., the total area of cracks, crack width, pothole, rutting, depression and patching. The input variable as statistical for Segment II are composing of a summary of the minimum, maximum, standard deviation and mean value, shown in Table 3.

Regarding the modeling of ANN for Segment II, the predicted versus observed results are shown in Fig. 4(a). Analyzing the relationship between predictive and measured value shows the case of SDI higher than 75, predictions are less accurate. Otherwise, for SDI under 75, the prediction was more accurate, as shown in Fig. 3(b). The fact of spreading values above 75 in the database is only under 25% and contributes to weaker prediction, because of SDI more than 75, based on scattered box, is out layer data. The importance of each parameter, obtained by sensitivity analysis, is shown in Fig/ 4(b), with rutting contributing with 47.9%, crack width 34.2%, followed by patching, pothole, and area crack area, 9.9%, 7.6% and 0.4%, respectively. Although this model presents a high performance, it is only valid for data found in the IRMS and for road submitted to the overloaded truck.

Table 4 shows the performance of quality of prediction Segment II obtained by MAE, RMSE and $R^2$, 0.492, 0.017 and 0.954, respectively. This indicates a high accurate prediction. The performance achieved by ANN model more accurate. Because, it has $R^2$ considerably high. When comparing ANN model of Segment I and Segment II, the Segment I is superior, as shown in Table 4. Whereas $R^2$ in the Segment I is higher than Segment II; as found 0.996 and 0.954.
CONCLUSIONS

Surface Distress is the key element to analyze and evaluate the road condition. Through the NN Data Mining approach, the predicted SDI can be obtained, providing a valuable contribution in terms of improving the assessment process of road conditions to be more accurate in the definition of rehabilitation. Surface Distress Index (SDI) value was computed for 607 sections of Segment I and 217 sections of Segment II for a complete set of data: area crack, width crack, pothole, rutting, depression and patching. The Multivariate Statistic Regression and NN Data Mining were used to try modeling SDI value.

Based on the test results on Segment I and Segment II, it was concluded that the Segment I provides the best result because NN Data Mining provided a high coefficient correlation of R², of 0.996 and 0.954 for Segment I and Segment II, respectively. When applying Data Mining to Segment I and Segment II, the more accurate prediction is obtained as SDI value below 75 (Fig. 3 (a) and (b)). The variables weight in SDI prediction for Segment I and Segment II is as follow: rutting with more than 50% and 48%, respectively, followed by width crack, 29%, and 34%, respectively. These findings strengthen the empirical data that considers rutting and width crack as the most relevant parameters to develop surface distress index for road network in Province of Jambi.

Furthermore, it is possible to model the data using multivariate linear statistics model and NN Data Mining has proved to be powerful tool to predict surface distress index and with similar models that we propose to develop the models predicting the performance service index and another indicator from database as well as data traffic and other types of pavement structure.

ACKNOWLEDGMENT

The Authors are grateful to the Lembaga Pengelola Dana Pendidikan (LPDP) for the financial support provided to this thesis through the grant financed by the Ministry of Finance Republic of Indonesia.

REFERENCES

8. W. Xiaodan, (1999), Research on the International Roughness Index (IRI) as the Pavement Roughness Index