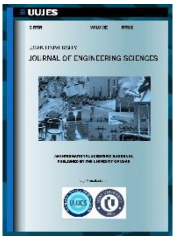

Usak University

Journal of Engineering Sciences

An international e-journal published by the University of Usak

Journal homepage: dergipark.gov.tr/uujes

Research article

OPTIMIZED SURFACE CONDITION CLASSIFICATION OF FLEXIBLE ROAD PAVEMENT USING AUTOWEKA MODEL

Paul Terkumbur Adeke, Aper E. Zava and Manasseh Tyogo*

Department of Civil Engineering, College of Engineering, Federal University of Agriculture Makurdi, Nigeria.

Received: 19 Sep 2020 Revised: 1 Dec 2020 Accepted: 8 Dec 2020 Online available: 28 Dec 2020
Handling Co-Editor: Jülide Öner

Abstract

The development of pavement management tools using intelligent algorithms requires a robust form of data mining – data classification for efficient and reliable results. The aim of this study is to investigate and optimally classify the surface condition of flexible road pavement along 60 km length of the Zaria – Kaduna Federal Highway in Northern Nigeria for maintenance decision. The study used data mining technique for the classification of pavement surface condition into good, satisfactory, fair, poor, very poor, serious or failed. A field survey was carried out to examine the surface area and length of various surface defects such as cracks, potholes, rutting and edge failure within chainages measuring 200 meters apart, which was used to compute the Pavement Condition Index (PCI) values and section classification in accordance with procedures stated in ASTM D6433. The AutoWEKA model of Waikato Environment for Knowledge Analysis (WEKA) software was used to optimally classify the surface condition of the highway. Results indicated that, 79.67% of the 300 total instances considered by the model were correctly classified while 20.33% of the instances were incorrectly classified. The optimum surface condition classification showed that worse pavement surface conditions of the sampled site were 'Poor', 'Very Poor' and 'Failed' at 77 (32.22%), 51 (21.34%) and 54 (22.59%) instances respectively of the correctly classified 239 instances out of the 300 total instances sampled. Based on its present condition, 76.15% of the road segment was bad. The rehabilitation or reconstruction of the Zaria – Kaduna Federal Highway was therefore recommended for improved condition and optimum performance.

Keywords: *Zaria – Kaduna highway; data mining; pavement surface condition; optimization; WEKA software.*

*Corresponding author: Adeke Paul Terkumbur
E-mail: adeke.pt@outlook.com (ORCID: 0000-0003-2939-8465)

DOI: 10.47137/uujes.791586

1. Introduction

In most African countries, highway transportation constitutes the major means of the transport system [1, 2]. This therefore requires conscious developmental strides for highway facilities to cater for the dire need to travel. Like other infrastructures, highway facilities especially the commonly used flexible road pavement, usually undergo continuous deterioration over time due to the impact of damaging factors such as traffic load, weather condition, material properties, age of pavement, original design of pavement structure, construction quality, road geometry and maintenance policy [2 – 5]. Flexible road pavement is the commonest type often used. It is an elastic type of pavement whose structure is made of different unbound layers of soil materials that exhibit nonlinear behaviour under the influence of damaging factors. The typical structure of a flexible pavement designed and built according to Nigeria specifications [6] is as shown in Fig. 1;

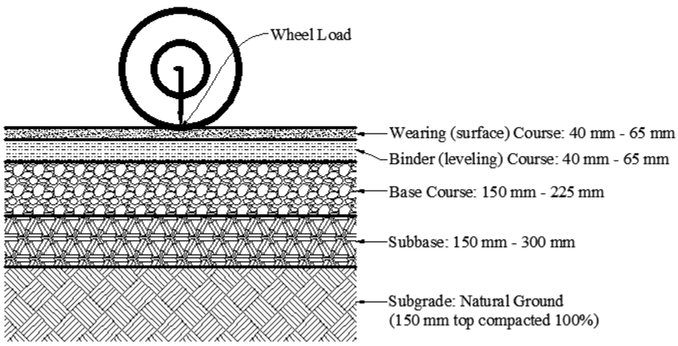

Fig. 1 Structure of a Flexible Pavement

The wearing course made of asphalt concrete is designed to carry traffic load directly, provide seal for the pavement structure against surface water and resist skidding [7]. The base and sub-base layers lay on the natural ground soil known as the subgrade. The massive use of flexible pavement for highway construction in most developing countries like Nigeria is due to its relatively cheap cost of construction and maintenance [8]. The performance of road pavement is usually measured on a classification scale of good to worse with subdivisions of; good, satisfactory, fair, poor, very poor, serious and failed states [9]. These classifications are a function of the severity (extent) and quantity (frequency) of surface defects caused by deterioration of the pavement materials. Common defects of flexible pavement include; cracks, potholes, ravelling, rutting, edge failure, shoving, swell, etc.

1.1 Pavement Management Systems

The practice of pavement management through performance monitoring and condition classification started decades ago. In 1950s, the American Association of State Highway and Transportation Officials (AASHTO) formerly known as American Association of State Highway Officials (AASHO) first carried out road test program to evaluate pavement conditions in America [10]. The approach used regression equations to examine relationship between effective variables. Several researches on pavement management were carried out thereafter. In recent times, pavement management systems employed the use of intelligent algorithms which involves data mining for condition classification for efficient and reliable maintenance decisions [11-13]. The use of intelligent algorithms in pavement surface condition classification is apt in road pavement management system, since the process is characterised with inconsistencies caused by subjective judgment and measurement of performance variables [12-14]. Hence the need for optimum classification of road pavement surface condition based on damaging effects at different stages of the life cycle for efficient and sustainable management policy. Previous studies had made concerted efforts in this regard; Yin [15] used probability theorem for integrating instrumentation data in probabilistic performance prediction of flexible pavement based on condition classifications, while Mahmood [16] studied at the network-level, maintenance decisions for flexible pavement using a soft computing-based framework which classified the road network into different regions. Studies in the past have established that initial surface condition classification is an essential aspect of pavement management system using intelligent algorithms [17-24].

There are different indices used for pavement condition classification. Setyawan et al. [25] examined the condition of road performance and remaining service life of some selected road pavements using the Pavement Condition Index (PCI) which is a function of the surface distresses per road section to predict the remaining service life using deflection data obtained from a Falling Weight Deflectometer (FWD) measurements. In Adeke *et al.* [26] assessment of road pavement condition on the Federal University of Agriculture Makurdi Campus was carried out to present classified and quantitative facts to the university management board for maintenance decision. Dabous *et al*. [27] developed a probabilistic distress-based evidential reasoning method for assessing pavement infrastructure condition and rating as the basis for maintenance decisions and budget allocations. The above mentioned studies employed the use of subjective evaluation techniques which were prone to estimation errors, hence require advanced analytical techniques for data training. Most existing pavement performance prediction models were based on pavement surface condition classification such as; Condition Rating Survey (CRS) or Pavement Condition Index (PCI) and the International Roughness Index (IRI) methodologies among others [27]. The modelling of pavement management system requires classified assessment of initial pavement distress conditions as independent variable on a section-by-section assessment [25, 28, 29]. In most developing countries like Nigeria, it is usually difficult to classify pavement condition using standard scales such as CRS, PCI and IRI methods for performance prediction over time due to cost of operation, lack of machineries and operational skills. Pavement management system generally requires large data on all the influencing factors which is usually expensive in terms of cost and time [3, 30]. Salpisoth [31] identified major challenges in road pavement management system to include; the ability to carry out routine monitoring and evaluation of pavement condition regularly with limited budget and expertise; and how to predict pavement performance using appropriate models based on incomplete data for optimum maintenance and repair decisions. Arifuzzaman et al. [32] used the predictive modelling and machine learning technique also known as the classification and regression tree (CART) method to predict the behaviour of adhesive properties of modified asphalt subjected to oxidation. The approach showed more explanatory relationship between the input variables at a nanoscale. The use of data science and machine learning techniques have become the most efficient and reliable methods for analysing data generated from systems behaviour for information and knowledge discovery [12, 33-35].

1.2 The Concept of Data Mining

Data mining is a technique of machining learning which is capable of using some algorithms to discover deep and hidden relationships between variables in a given dataset based on its elements and attributes [14, 35, 36]. It is suitable for numeric analysis of systems characterised by noise and incomplete or missing data for classification and mapping to reason and discovering relationship between variables in a dataset [14, 37 – 39, 40]. Data mining uses non-generative, black-box models or exploratory techniques to examine variables that define systems behavior [41]. It is used for system classification and performance prediction to aid categorical understanding of systems behaviour [12, 14]. Classification in data mining simply means the grouping of individual items or entries into common groups based on characteristic similarities as relates to specific groups using a known technique or classifier [42]. The use of data mining techniques has been commended by previous studies due to its successes in diverse fields of research; Araujo *et al*., [43] used data mining for the estimation of energy consumption on the tire-pavement interaction for asphalt mixtures with different surface properties. Also, Luca *et al.* [44] used the data mining approach on friction data to investigate runway pavement friction decay at the International Civil Airport of Lamezia Terme, Italy. These studies all stated the suitability and efficiency of data mining technique for describing system behaviour based on condition classification.

1.3 The AutoWEKA Model

Data classification and many other methods of data mining can be implemented in the Waikato Environment for Knowledge Analysis (WEKA) software. It is a machine learning and data mining toolkit developed at the Waikato University, New Zealand. It is an open source software written in Java programming language, used for research and project works [40, 45]. The software is capable of carrying out data mining in the form of pre-processing, data classification, visualisation, association, clustering and filtering [35, 46]. Algorithmic models used by WEKA for behavioural classification and performance predictions of systems include the Naïve Bayes, Decision Tree, Random forest, etc. [40]. When the most suitable classifer or regression model in terms of performance and parameterism among the aforementioned is required, the Automatic WEKA (AutoWEKA) command is used. Its major function is to peruse the dataset and select the optimum model and hyperparameters for system analysis [35]. Not much of AutoWEKA applications are found in previous researches especially in pavement management, its theoretical function and strength motivated this study. Methodologies based on intelligent algorithms employed by previous studies for data classification yielded reliable results [14, 47 - 49]

The aim of this study therefore, is to investigate surface condition classification of flexible road pavement using intelligent algorithms. Objective of the study is to optimally classify surface condition of the Zaria – Kaduna Federal Highway using data mining technique implemented in WEKA software.

2. Methodology

2.1 Description of Study Area

A measure of road segments along the Zaria - Kaduna Federal Highway was examined by this study. The route length considered measured 60 km from Zaria - Kaduna towns. The site is located within the North-West region of Nigeria. It is the major route connecting the North-Central and North-West regions of Nigeria running through Kaduna State. A map of Nigeria showing the proposed site by its Average Daily Traffic flow is as presented in Fig. 2;

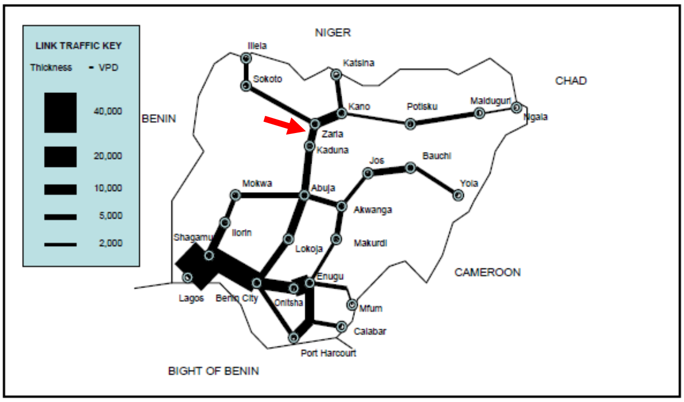

Fig. 2 Road Network and ADT in Nigeria [4]

2.2 Data Collection and Analysis

The study involved an assessment and classification of pavement condition along the Zaria – Kaduna Federal Highway using empirical dataset based on procedures stated in Road Sector Development Team [4] and ASTM D6433 [9]. The process involved measurement of geometric features of pavement defects along the 60 km length of the Federal Highway. Measuring instruments used for the field work included, steel tape, range rod, plastic ruler and safety kits. The route length was divided into equal chainages measuring 200 meters. Based on their relatively obvious occurrence, types of defects considered and methods of measuring affected surface area and level of severity were as presented below;

Pothole: this is a bowl-shaped depression on the pavement surface with sharp edges and vertical or inclined sides as a result of propagation of alligator cracking promoted by the presence of surface water and impulse of wheel load. Features of a typical pothole are as shown in Fig. 3. Plate A1 presents a pothole and method of measuring its severity levels are as shown in Plates A2 and A3. The impact of potholes to PCI is measured based on the quantity of potholes within a road segment counted according to their respective levels of severity.

Fig. 3 Pothole and Measurement of Severity Level

Categories of severity levels measured as a function of the mean depth of pothole are as presented in Table 1;

Table 1 Level of Severity for Potholes [9]

Maximum Depth of Pothole (mm)	Average Diameter (mm)		
	100 to 200	200 to 450	450 to 750
13 to ≤ 25	Low	Low	Medium
> 25 and ≤ 50	Low	Medium	High
> 50	Medium	Medium	High

Rutting: this is a surface depression along the line of traffic load (wheel path) due to permanent deformation in any pavement layer or the subgrade caused by consolidation of materials. A typical rut and its features are as shown in Fig. 4. Ruting is measured in square meters of the distressed surface area shown in Plate B1. Its levels of severity are the effective length and the mean depth of the rut as shown in Plates B2 and B3 respectively. The severity levels for Low, medium and high are measured as the mean rut depth on a scale of 6 to 13 mm, 13 to 25 mm and > 25 mm respectively.

Fig. 4 Rutting and Measurement of Severity Level

Edge Failure: also known as edge cracking are edge cracks which run parallel to, and usually measured within 0.3 to 0.5 m length of the outer edge of the road pavement. Essentails details of an edge failure defect used for this study are as shown in Fig. 5. It is usually caused by traffic loading and weak subsoil layers. The impact of edge failure as shown in Plate C1 is measured along linear distance in meters as shown in Plate C2 and the severity levels depend on the quantity of breakup, cracks and ravelling.

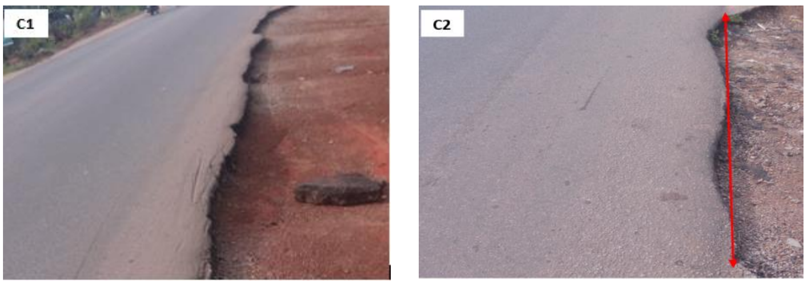

Fig. 5 Edge Failure and Measurement of Severity Level

Cracks: this refers to the alligator or fatigue crack which is a series of interconnected cracks caused by fatigue failure of the pavement surface due to repeated traffic loading as shown in Plate D1. Its effect is measured in square meters of the affected surface area and the level of severity is based on crack sizes which range from; fine and parallel longitudinal hairlines cracks with few interconnections for the low severity level, to light network of alligator cracks that are lightly spalled for the medium level and well defined network of cracks with spalled at the edges for the high level, measured within a square area as shown in Plate D2;

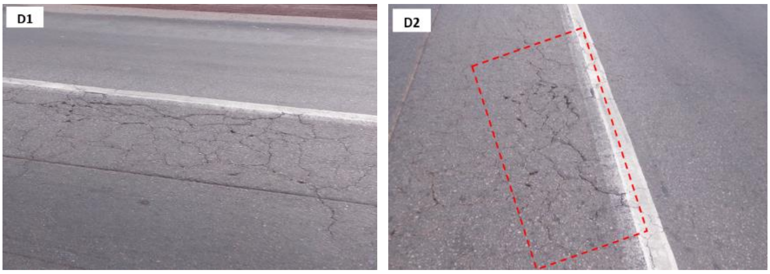

Fig. 6 Cracks and Measurement of Severity Level

Data obtained from the field survey were collated into a spreadsheet package – Microsoft Excel and procedures stated in ASTM D6433 [9] manual were followed chronologically to classify the pavement condition. The summary of the procedure used for computing Pavement Condition Index (PCI) for condition classification of the road pavement were as follows;

Step 1: Carry out a classified collation of various defects according to suitable format (square meter area or meter length) at respective levels of severity in a spreadsheet specially designed for the purpose.

Step 2: Sum up the total quantity of each distress type at each distress severity level and record as the total severity

Step 3: Determine the density of each distress type at each severity level in every road segment using Equation 1;

$$Density\ of\ Defect\ (\%) = \frac{Amount\ of\ Defect}{Amount\ of\ Road\ Segment} \times 100 \tag{1}$$

Step 4: Determine the deduct value for each distress type and severity level combination from the distress deduct value curves from the manaual.

Step 5: Determine the maximum corrected deduct value using the following procedure; if none or only one individual deduct value is greater than 2.0, use the total value in place of the maximum Corrected Deduct Value (CDV) in determining the PCI; otherwise, maximum CDV should be determined using the following procedure;

a) List the individual deduct values in descending order
b) Determine the allowable number of deducts using suitable curve from the manual or Equation 2;

$$\mathrm{m} = 1 + (9/98)(100 - HDV) \leq 10 \tag{2}$$

where, m is the allowable number of deducts including fractions (which must be less than or equal to 10.0) and HDV is the Highest Individual Deduct value.

c) Reduce the number of individual deduct values to the m largest deduct values including fractional part. If less than m deduct values are available, use all the deduct values.
d) Determine the maximum CDV iteratively from the manual
e) Determine the total deduct value by summing individual deduct values
f) Determine q as the number of deducts with a value greater than 2.0
g) Determine the CDV from total deduct value and q by looking up the appropriate correction curve from the manual for Asphalt concrete pavement

h) Reduce the smallest individual deduct value greater than 2.0 to 2.0 and repeat e), f) and g) until q = 1

i) Maximum CDV is the largest of the CDVs.

Step 6: Compute PCI value by subtracting the maximum CDV from 100.

Step 7: Classify the PCI value of each road segment using Table 2;

Table 2 Scale for Pavement Condition Classification [9]

Pavement Surface Condition	PCI
Good	86 – 100
Satisfactory	71 – 85
Fair	56 – 70
Poor	41 – 55
Very poor	26 – 40
Serious	11 – 25
Failed	0 – 10

2.3 Optimisation of Pavement Surface Condition Classification

The original file saved in column-separated values (.csv) format contained values for variables entered in the file such that; rows represent instances, columns were for defined attributes per segment such as density of cracks, potholes, rutting and edge failure with the last column representing the targeted attribute or outcome of pavement condition classification – good, satisfactory, fair, poor, very poor, serious and failed. All variables entered were defined into strings, numeric and nominal values. Using the pavement surface condition classifications scale, surface condition of the road length was classified along chainages as shown in Table 3;

Table 3(a) Classification of Pavement Surface Condition

Chainage (m)	PCI	Condition Classification
CH 000+000	-	-
CH 000+200	61	Fair
CH 000+400	57	Fair
CH 000+600	50	Poor
CH 000+800	47	Poor
CH 001+000	66	Fair
CH 001+200	58	Fair
CH 001+400	69	Fair
CH 001+600	70	Fair
CH 001+800	80	Satisfactory
CH 002+000	68	Fair
CH 002+200	81	Satisfactory
CH 002+400	45	Poor
CH 002+600	72	Satisfactory
CH 002+800	77	Satisfactory
CH 003+000	83	Satisfactory
CH 003+200	42	Poor
CH 003+400	52	Poor
CH 003+600	74	Satisfactory
CH 003+800	85	Satisfactory
CH 004+000	59	Fair
CH 004+200	73	Satisfactory
CH 004+400	75	Satisfactory
CH 004+600	66	Fair
CH 004+800	84	Satisfactory
CH 005+000	56	Fair
CH 005+200	67	Fair
CH 005+400	71	Satisfactory
CH 005+600	44	Poor
CH 005+800	65	Fair
CH 006+000	77	Satisfactory
CH 006+200	73	Satisfactory
CH 006+400	44	Poor
CH 006+600	75	Satisfactory
CH 006+800	59	Fair
CH 007+000	68	Fair
CH 007+200	76	Satisfactory
CH 007+400	79	Satisfactory
CH 007+600	88	Good
CH 007+800	92	Good
CH 008+000	80	Satisfactory
CH 008+200	77	Satisfactory
CH 008+400	82	Satisfactory
CH 008+600	80	Satisfactory
CH 008+800	72	Satisfactory
CH 009+000	76	Satisfactory
CH 009+200	71	Satisfactory
CH 009+400	69	Fair
CH 009+600	84	Satisfactory
CH 009+800	65	Fair
CH 010+000	55	Poor
CH 010+200	36	Very Poor
CH 010+400	57	Fair
CH 010+600	21	Serious
CH 010+800	40	Very Poor

Chainage (m)	PCI	Condition Classification
CH 011+000	30	Very Poor
CH 011+200	9	Failed
CH 011+400	35	Very Poor
CH 011+600	7	Failed
CH 011+800	6	Failed
CH 012+000	4	Failed
CH 012+200	9	Failed
CH 012+400	10	Failed
CH 012+600	3	Failed
CH 012+800	45	Poor
CH 013+000	10	Failed
CH 013+200	17	Serious
CH 013+400	6	Failed
CH 013+600	37	Very Poor
CH 013+800	40	Very Poor
CH 014+000	28	Very Poor
CH 014+200	41	Poor
CH 014+400	7	Failed
CH 014+600	40	Very Poor
CH 014+800	54	Poor
CH 015+000	50	Poor
CH 015+200	10	Failed
CH 015+400	6	Failed
CH 015+600	5	Failed
CH 015+800	29	Very Poor
CH 016+000	50	Poor
CH 016+200	30	Very Poor
CH 016+400	32	Very Poor
CH 016+600	41	Poor
CH 016+800	38	Very Poor
CH 017+000	67	Fair
CH 017+200	27	Very Poor
CH 017+400	37	Very Poor
CH 017+600	55	Poor
CH 017+800	54	Poor
CH 018+000	8	Failed
CH 018+200	43	Poor
CH 018+400	69	Fair
CH 018+600	58	Fair
CH 018+800	55	Poor
CH 019+000	50	Poor
CH 019+200	48	Poor
CH 019+400	49	Poor
CH 019+600	55	Poor
CH 019+800	40	Very Poor
CH 020+000	2	Failed
CH 020+200	36	Very Poor
CH 020+400	44	Poor
CH 020+600	7	Failed
CH 020+800	28	Very Poor
CH 021+000	8	Failed
CH 021+200	3	Failed
CH 021+400	6	Failed
CH 021+600	15	Serious
CH 021+800	10	Failed

Table 3(b) Classification of Pavement Surface Condition Continued

Chainage (m)	PCI	Condition Classification	Chainage (m)	PCI	Condition Classification
CH 022+000	20	Serious	CH 033+000	26	Very Poor
CH 022+200	10	Failed	CH 033+200	0	Failed
CH 022+400	7	Failed	CH 033+400	55	Poor
CH 022+600	8	Failed	CH 033+600	15	Serious
CH 022+800	5	Failed	CH 033+800	7	Failed
CH 022+000	1	Failed	CH 034+000	2	Failed
CH 023+200	9	Failed	CH 034+200	6	Failed
CH 023+400	41	Poor	CH 034+400	0	Failed
CH 023+600	28	Very Poor	CH 034+600	5	Failed
CH 023+800	20	Serious	CH 034+800	7	Failed
CH 023+000	23	Serious	CH 035+000	8	Failed
CH 024+200	44	Poor	CH 035+200	10	Failed
CH 024+400	49	Poor	CH 035+400	0	Failed
CH 024+600	50	Poor	CH 035+600	20	Serious
CH 024+800	5	Failed	CH 035+800	7	Failed
CH 024+000	8	Failed	CH 036+000	50	Poor
CH 025+200	27	Very Poor	CH 036+200	35	Very Poor
CH 025+400	45	Poor	CH 036+400	55	Poor
CH 025+600	55	Poor	CH 036+600	46	Poor
CH 025+800	60	Fair	CH 036+800	49	Poor
CH 026+000	33	Very Poor	CH 037+000	33	Very Poor
CH 026+200	30	Very Poor	CH 037+200	67	Fair
CH 026+400	1	Failed	CH 037+400	21	Serious
CH 026+600	45	Poor	CH 037+600	0	Failed
CH 026+800	55	Poor	CH 037+800	22	Serious
CH 027+000	20	Serious	CH 038+000	50	Poor
CH 027+200	23	Serious	CH 038+200	8	Failed
CH 027+400	47	Poor	CH 038+400	37	Very Poor
CH 027+600	8	Failed	CH 038+600	35	Very Poor
CH 027+800	5	Failed	CH 038+800	26	Very Poor
CH 027+000	4	Failed	CH 039+000	29	Very Poor
CH 028+200	40	Very Poor	CH 039+200	45	Poor
CH 028+400	36	Very Poor	CH 039+400	48	Poor
CH 028+600	29	Very Poor	CH 039+600	54	Poor
CH 028+800	50	Poor	CH 039+800	33	Very Poor
CH 028+000	37	Very Poor	CH 040+000	55	Poor
CH 029+200	5	Failed	CH 040+200	54	Poor
CH 029+400	7	Failed	CH 040+400	55	Poor
CH 029+600	2	Failed	CH 040+600	40	Very Poor
CH 029+800	30	Very Poor	CH 040+800	45	Poor
CH 030+000	7	Failed	CH 041+000	38	Very Poor
CH 030+200	26	Very Poor	CH 041+200	36	Very Poor
CH 030+400	10	Failed	CH 041+400	50	Poor
CH 030+600	15	Serious	CH 041+600	38	Very Poor
CH 030+800	1	Failed	CH 041+800	77	Satisfactory
CH 031+000	3	Failed	CH 042+000	80	Satisfactory
CH 031+200	5	Failed	CH 042+200	54	Poor
CH 031+400	3	Failed	CH 042+400	55	Poor
CH 031+600	29	Very Poor	CH 042+600	65	Fair
CH 031+800	30	Very Poor	CH 042+800	54	Poor
CH 032+000	35	Very Poor	CH 043+000	47	Poor
CH 032+200	30	Very Poor	CH 043+200	71	Satisfactory
CH 032+400	25	Serious	CH 043+400	59	Fair
CH 032+600	1	Failed	CH 043+600	55	Poor
CH 032+800	26	Very Poor	CH 043+800	49	Poor

Table 3(c) Classification of Pavement Surface Condition Continued

Chainage (m)	PCI	Condition Classification	Chainage (m)	PCI	Condition Classification
CH 044+000	2	Failed	CH 052+000	41	Poor
CH 044+200	5	Failed	CH 052+200	55	Poor
CH 044+400	46	Poor	CH 052+400	39	Poor
CH 044+600	66	Fair	CH 052+600	85	Satisfactory
CH 044+800	34	Very Poor	CH 052+800	70	Fair
CH 045+000	47	Poor	CH 053+000	44	Poor
CH 045+200	18	Serious	CH 053+200	53	Poor
CH 045+400	6	Failed	CH 053+400	40	Very Poor
CH 045+600	34	Very Poor	CH 053+600	55	Poor
CH 045+800	50	Poor	CH 053+800	48	Poor
CH 046+000	34	Very Poor	CH 054+000	41	Poor
CH 046+200	40	Very Poor	CH 054+200	46	Poor
CH 046+400	50	Poor	CH 054+400	55	Poor
CH 046+600	32	Very Poor	CH 054+600	50	Poor
CH 046+800	50	Poor	CH 054+800	55	Poor
CH 047+000	34	Very Poor	CH 055+000	36	Very Poor
CH 047+200	54	Poor	CH 055+200	40	Very Poor
CH 047+400	8	Failed	CH 055+400	40	Very Poor
CH 047+600	10	Failed	CH 055+600	49	Poor
CH 047+800	2	Failed	CH 055+800	37	Very Poor
CH 048+000	38	Very Poor	CH 056+000	55	Poor
CH 048+200	40	Very Poor	CH 056+200	26	Very Poor
CH 048+400	47	Poor	CH 056+400	5	Failed
CH 048+600	50	Poor	CH 056+600	25	Serious
CH 048+800	55	Poor	CH 056+800	41	Poor
CH 049+000	49	Poor	CH 057+000	10	Failed
CH 049+200	60	Fair	CH 057+200	40	Very Poor
CH 049+400	44	Poor	CH 057+400	28	Very Poor
CH 049+600	7	Failed	CH 057+600	26	Very Poor
CH 049+800	58	Fair	CH 057+800	36	Very Poor
CH 050+000	50	Poor	CH 058+000	40	Very Poor
CH 050+200	54	Poor	CH 058+200	39	Very Poor
CH 050+400	27	Very Poor	CH 058+400	55	Poor
CH 050+600	43	Poor	CH 058+600	11	Serious
CH 050+800	27	Very Poor	CH 058+800	2	Failed
CH 051+000	41	Poor	CH 059+000	27	Very Poor
CH 051+200	56	Fair	CH 059+200	10	Failed
CH 051+400	48	Poor	CH 059+400	40	Very Poor
CH 051+600	70	Fair	CH 059+600	36	Very Poor
CH 051+800	37	Very Poor	CH 059+800	25	Serious

Results from Table 3 (a) to (c) were collated into an excel spreadsheet file. The file was then imported into WEKA Explorer in the form of an Attribute-Relation File Format (ARFF) for analysis. The AutoWEKA classifier model was used for optimally training the dataset. İt is an automatic model used for selection and hyper-parameter optimization in the WEKA software. Classifications obtained from the analysis were optimally implemented using the WEKA software to derive an optimum distribution of flexible road pavement surface condition classification in the investigation.

3. Results and Discussion

The summary of results obtained from simulation of the AutoWEKA model using pavement surface condition classification data is as presented in Fig. 7;

```
=== Summary ===

Correctly Classified Instances         239               79.6667 %
Incorrectly Classified Instances        61               20.3333 %
Kappa statistic                          0.742
Mean absolute error                      0.0561
Root mean squared error                  0.1675
Relative absolute error                 28.1792 %
Root relative squared error             53.1521 %
Total Number of Instances              300

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     -
                 0.900    0.059    0.628      0.900    0.740      0.720    0.979     0.806     Fair
                 0.906    0.144    0.713      0.906    0.798      0.715    0.963     0.885     Poor
                 0.536    0.011    0.833      0.536    0.652      0.643    0.976     0.793     Satisfactory
                 0.500    0.000    1.000      0.500    0.667      0.706    0.999     0.833     Good
                 0.729    0.039    0.850      0.729    0.785      0.729    0.974     0.903     Very Poor
                 0.722    0.004    0.929      0.722    0.813      0.809    0.995     0.919     Serious
                 0.818    0.004    0.982      0.818    0.893      0.871    0.994     0.972     Failed
Weighted Avg.    0.797    0.058    0.823      0.797    0.797      0.753    0.978     0.894
```

Fig. 7 Model Performance

Fig. 7 presented results of an optimised model of the algorithm. It revealed that 76.67% of the 300 total instances or entries considered by the model were correctly classified into good, satisfactory, fair, poor, very poor, serious and failed; while 20.33% of the instances were incorrectly classified. This implied that majority of the instances were fittingly classified into the defined classifications of attributes. The Kappa statistics also defined as the coefficient of correlation of the optimized model was 0.742. In spite of the significant Relative Absolute Error (RAE) of 28.1792% and Root Relative Square Error (RRSE) of 53.1521% caused by the incorrectly classified pavement conditions, the Kappa statistics value is approaching unity which explains a relatively strong relationship among the independent and the dependent variables of the optimised classification model. The relatively low value of the mean absolute error 0.0561 indicates high accuracy of the prediction outcome. Also, the significant values of True Positive (TP) rates against the False Positive (FP) predictions for the various instances indicate high level of prediction accuracy. The high degree of precision, recall and F-score (F-measure) values further justify the agreement that classification by this AutoWEKA model is within a tolerable error.

Another important results that further explained the characteristics of the surface condition classification model for flexible road pavement is as presented in Fig. 8;

```
=== Confusion Matrix ===

  a  b  c  d  e  f  g  h   <-- classified as
  1  0  0  0  0  0  0  0 |  a = -
  0 27  3  0  0  0  0  0 |  b = Fair
  0  7 77  1  0  0  0  0 |  c = Poor
  0  2 11 15  0  0  0  0 |  d = Satisfactory
  0  1  0  0  1  0  0  0 |  e = Good
  0  4 12  2  0 51  0  1 |  f = Very Poor
  0  1  2  0  0  2 13  0 |  g = Serious
  0  1  3  0  0  7  1 54 |  h = Failed
```

Fig. 8 Confusion Matrix

Fig. 8 presented the confusion matrix. The diagonal of this matrix explains the share of all correctly classified attributes used for the classification of pavement defects on the site, while other entries on the matrix represent the unexplained or incorrectly classified outputs. It can also be deduced from Fig. 8 that, the overall or optimum classification of pavement condition on the site was 'Poor' as indicated by its relatively high density value of 77 (32.22%) instances. The 'failed and very poor' pavement conditions of segments also recorded relatively significant values of 54 (22.59%) and 51 (21.34%) instances respectively. This indicated that a significant portion of the highway segments (76.15%) is undergoing rapid and continuous deterioration which actually tends to total failure as anticipated, hence calls for quick maintenance intervention by rehabilitation or reconstruction.

4. Conclusion

Due to the relevance of initial surface condition classification of road pavement in performance management system, this study attempted to investigate and optimally classify pavement surface condition along the Zaria – Kaduna Federal Highway in Northern Nigeria using data mining technique. Procedures stated in ASTM D6433 manual were used for data collection, while the AutoWEKA model of WEKA software was used for data analysis. Results indicated that the optimum classification of the flexible road pavement was 'Poor' at 77 (32.22%) instances. Other significant classifications were 'Very Poor' and 'Failed' conditions at 51 (21.34%) and 54 (22.59%) instances respectively of the correctly classified 239 instances out of the 300 total instances considered. With this high proportion of poor road segments, the rehabilitation or reconstruction of the Zaria – Kaduna Federal Highway was recommended to improve its present condition.

Acknowledgements

Special thanks to Richard Tyokyaa and Raymond Tume for assisting us for the field work.

References

1. African Development Bank Group. Rail infrastructure in Africa – Financing Policy Options. International d'Abidjan, Abidjan, Côte d'Ivoire, 2015.
2. Adeke PT, Atoo AA and Joel E. A Policy framework for efficient and sustainable road transport system to boost synergy between urban and rural settlements in developing countries: A case of Nigeria, 1st International Civil Engineering Conference (ICEC 2018), Department of Civil Engineering, Federal University of Technology, Minna, Nigeria, 2018a.
3. Abiola OS, Owolabi AO., Odunfa SO and Olusola A. Investigation into causes of premature failure of highway pavements in nigeria and remedies. In proceedings of the Nigeria Institution of Civil Engineers (NICE) Conference, Abuja, 2010
4. Road Sector Development Team. Configuration and Calibration of HDM-4 to Nigeria Conditions, Government of the Federal Republic of Nigeria. Nigeria, 2014. p. 33.
5. Dong S, Zhong J Hao, P Zhang W, Chen J, Lei Y and Schneider A. Mining multiple association rules in LTPP database: An analysis of asphalt pavement thermal cracking distress, Construction and Building Materials, 2018;191:837-852.
6. Government of the Federal Republic of Nigeria. General Specifications (Roads and Bridges), Volume II, Revised, Abuja, Nigeria, 2016.
7. Garber NJ and Hoel LA. Traffic and Highway Engineering, 4th Edition, Cengage Leaning, Canada, 2009.

8. Taylor MAP and Philip ML. Investigating the impact of maintenance regimes on the design life of road pavements in a changing climate and the implications for transport policy. Transport Policy, 2015;41;117-135.
9. ASTM D6433-07 Standard Practice for Road and Parking Lots Pavement Condition Index Survey, American Standard for Testing and Materials, USA, Philadelphia; 2007.
10. American Association of State Highway and Transportation Officials (AASHTO). AASHTO Guide for the Design of Pavement Structures. Washington, D.C., 1993.
11. Lina JD, Huangb WH, Hungc CT, Chend CT and Leee JC. Using decision tree for data mining of pavement maintenance and management. Applied Mechanics and Materials, 2013;330;1015-1019.
12. Inkoom S, Sobanjo J, Barbu A. and Niu X. Prediction of the crack condition of highway pavements using machine learning models. Structure and Infrastructure Engineering, 2019, DOI: 10.1080/15732479.2019.1581230.
13. Li Z, Cheng C, Kwan MP, Tong X and Tian S. Identifying asphalt pavement distress using UAV LiDAR point cloud data and random forest classification. International Journal of Geo-information, 2019;8(39);1-26. DOI:10.3390/ijgi8010039.
14. Miradi M. Knowledge discovery and pavement performance: Intelligent data mining, PhD Thesis, Road and Railway Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, The Netherlands, 2009.
15. Yin H. Integrating instrumentation data in probabilistic performance prediction of flexible pavement, PhD Thesis, Civil and Environmental Engineering, Graduate School, The Pennsylvania State University, 2007.
16. Mahmood MS. Network-Level maintenance decisions for flexible pavement using a soft computing-based framework. PhD Thesis, Nottingham Trent University, United Kingdom, 2015.
17. Fwa TF and Shanmugam R. Fuzzy logic technique for pavement condition rating and maintenance-needs assessment. 4th International Conference on Managing Pavements; 1998 May 17-21; Durban, South Africa. Washington DC. USA: TRB; 1998. p. 465-476.
18. Mahmood M, Rahman M, Nolle L, and Mathavan S. A fuzzy logic approach for pavement section classification. International Journal of Pavement Research and Technology, Chinese Society of Pavement Engineering, 2013;6(5):620 – 626. DOI: 10.6135/ijprt.org.tw/2013.6(5).620.
19. Cheu RL, Wang Y and Fwa TF. Genetic algorithm-simulation methodology for pavement maintenance scheduling. Computer –Aided Civil and Infrastructural Engineering, 2004;19;446-455.
20. Chassiakos AP. A fuzzy-based system for maintenance planning or road pavements. In: Bojkovic ZS. World Scientific and Engineering Academy and Society (WSEAS): 10th International Conference on Computers; 2006 July 10 - 15; Vouliagmeni, Athens, Greece. Wisconsin, USA: Association for Computing Machinery; 2006. p. 535 – 540.
21. Liu Y and Sun M. Fuzzy optimization BP neural network model for pavement performance assessment. IEEE International Conference on Grey Systems and Intelligent Services; Grey Systems and Intelligent Services (GSIS 2007); 2007 Nov. 18 – 20; Nanjing, China. China: IEEE; 2007. p. 1031-1034.
22. Golroo A and Tighe SL. Fuzzy set approach to condition assessments of novel sustainable pavements in the canadian climate. Canadian Journal of Civil Engineering, 2009;36:754-764.

23. Bianchini A and Bandini P. Prediction of pavement performance through neuro-fuzzy reasoning. Computer-aided civil and infrastructure engineering, 2010;25; 39-54.
24. Thube DT. Artificial Neural Network (ANN) based pavement deterioration models for low volume roads in India. International Journal of Pavement Research and Technology, 2011;5(2);115 -120.
25. Setyawan A, Nainggolan J and Budiarto A. Predicting the remaining service life of road using pavement condition index, Procedia Engineering 125; 5th International Conference of Euro Asia Civil Engineering Forum (EACEF-5 2015), 2015 Sept. 15 – 18; Surabaya, East Java, Indonesia. Elsevier; 2015. p. 417 – 423.
26. Adeke PT, Atoo AA and Orga SG. Assessment of pavement condition index: A case of flexible road pavements on the University of Agriculture Makurdi Campus. Nigerian Journal of Technology, 2018b;38(1):15-21.
27. Dabous SA, Zeiada W, Al-Ruzouq R, Hamad K and Al-Khayyat G. Distress-Based evidential reasoning method for pavement infrastructure condition assessment and rating. International Journal of Pavement Engineering, 2019;1–12. DOI: 10.1080/10298436.2019.1622012.
28. Hamed RI and Kakarash ZA. Evaluate the asphalt pavement performance of rut depth based on intelligent method. International Journal of Engineering and Computer Science, 2016;5(1):15474-15481.
29. Premkumar L and Vavrik WR. Enhancing pavement performance prediction models for the illinois tollway system. International Journal of Pavement Research and Technology, 2016;9:14-19.
DOI: http://dx.doi.org/10.1016/j.ijprt.2015.12.002
30. Surendrakuma K, Prashant N and Mayuresh P. Application of Markovian probabilistic process to develop a decision support system for Pavement maintenance management. International Journal of Scientific and Technology Research, 2013;2(8):295-303.
31. Salpisoth H. Simple evaluation methods for road pavement management in developing countries. PhD Thesis, Graduate School of Engineering, Kyoto University, Japan, 2014.
32. Arifuzzaman M, Gazder U, Alam MS and Sirin O. Modelling of asphalt's adhesive behaviour using classification and regression tree (CART) analysis. Computational Intelligence and Neuroscience, ID 3183050, Hindawi. https://doi.org/10.1155/2019/3183050, 2019.
33. Huang YH. Pavement Analysis and Design, 2nd edition. Upper Saddle River, New Jersey, USA; Pearson Prentice Hall, Inc.; 2004.
34. Saltan M, Terzi S and Kucuksille EU. Backcalculation of pavement layer moduli and Poisson's ratio using data mining. Expert Systems with Applications, 2011; 38:2600-2608.
35. Hutter F, Kotthoff L and Vanschoren J. Automated machine learning – Methods, systems, challengers. The Springer Series on Challenges in Machine Learning. Gewerbestrasse, Cham, Switzerland; Springer Nature; 2019.
36. Cox E. Fuzzy modelling and genetic algorithms for data mining and exploration. San Francisco, California, USA: Elsevier; 2005.
37. Pawlak Z. Rough Sets. In: Lin TY and Cercone N. Rough sets and data mining - Analysis of imprecise data. Dordrecht, Netherlands: Springer Nature; 1997. p. 3 – 7.
38. Pawlak Z. Rough sets and intelligent data analysis. Information Sciences 147; Natural Environment Management and Applied Systems Analysis; 2002 Sept. 6 – 9; Laxemburg, Austria. Austria: Elsevier; 2002. p. 1 – 12.

39. Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982;11:341-356.
40. Witten IH, Frank E Hall, MA and Pal CJ. The WEKA workbench – Data mining practical machine learning tools and techniques, 4th edition, Burlington, Massachusetts, USA: Elsevier; 2016.
41. Fox C. Data Science for transport; self-study guide with computer exercises. Gewerbestrasse, Cham, Switzerland: Springer Nature; 2018.
42. Mohd A, Yuk Y, Wei-Chang Y, Noorhaniza W and Ahmad M. Classification technique using modified particle swarm optimization. Modern Applied Science, 2011;5(5):150–164.
43. Araujo J P C, Palha CAO, Martins FF and Silva HMRD. Estimation of energy consumption on the tire-pavement interaction for asphalt mixtures with different surface properties using data mining techniques. Transport Research Part D, 2019;67:421-432.
44. Luca MD, Abbondati F, Pirozzi M and Zilioniene D. Preliminary study on runway pavement friction decay using data mining. In: Rafalski L and Zofka A. Transport Research Procedia 14: 6th Transport Research Arena (TRA2016); 2016 April 18 – 21; Warsaw, Poland. Austria: Elsevier; 2016. p. 3751 – 3760.
45. Sharma TC and Jain M. WEKA approach for comparative study of classification algorithm. International Journal of Advanced Research in Computer and Communication Engineering, 2013;2(4);1925-1931.
46. WEKA. Waikato Environment for Knowledge Analysis: User's Manual, The University of Waikato, New Zealand, 2018.
47. Marianingsih S and Utaminingrum F. Comparison of support vector machine classifier and Naïve Bayes classifier on road surface type classification. International Conference on Sustainable Information Engineering and Technology (SIET); 2018 Nov. 10 – 12; Malang, Indonesia. USA: IEEE; 2018. p. 35-48.
48. Marianingsih S, Utaminingrum F and Bachtiar FA. Road surface types classification using combined of k-nearest neighbor and Naïve Bayes based on GLCM. International Journal of Advances in Soft Computing and its Applications, 2019;11(2):15-27.
49. Gong H, Sun Y, Shu X. and Huang B. Use of random forest regression for predicting IRI of asphalt pavements. Construction and Building Materials, 2018;189:890- 897.