EVALUATION OF ROAD CONDITION INDICES METHODS AND APPLICABILITY FOR USE IN MACHINE LEARNING

M SIMELANE[1]* and A RAMPERSAD[1]**

[1]Smart Mobility, Council for Scientific and Industrial Research, Pretoria, South Africa;
*Tel: 012 841 2924; Email: msimelane@csir.co.za
**Tel: 012 841 3831; Email: arampersad@csir.co.za

ABSTRACT

Road maintenance is a crucial process for pavement management systems. South African local roads managed by lower road authorities (municipality, etc) are in critical condition, and their management is not at optimum level which is evident from their poor condition. The aim of this paper is to provide a Machine-Learning algorithm to assist road authorities to provide optimal maintenance strategies. The objective of the study was to determine the most effective condition index for management of flexible pavements. This is achieved by conducting descriptive and inferential statistical analysis of two case studies (Low volume roads and High-volume roads). Statistical analysis indicated that the visual condition index (VCI) has inconsistencies compared to the deduct point surface condition index (CI_{SURF}) and deduct point pavement condition index (CI_{PAVE}) found in TMH 22. Four machine learning models were created which included the Gradient Boosting Classifier, Random Forest Classifier, Support Vector Machine Classifier, and Decision Tree Classifier. Of the four models explored, the model with the greatest potential for deployment was the Gradient Boosting Classifier (GBC) model. The GBC model had an accuracy of 74 %, 85 % and 93 % in relation to the VCI, CI_{SURF} & CI_{PAVE} respectively. The CI_{SURF} and CI_{PAVE} was identified as the most effective index for use in flexible pavements.

1. INTRODUCTION

A Pavement Management System (PMS) is a set of defined procedures for collecting, analysing, maintaining, and reporting pavement data, to assist the decision makers in finding optimum strategies for maintaining pavements in serviceable condition over a given period of time for the least cost. Key components of a PMS include, but are not limited to (Sabita, 2020):

- Road inventory.
- Pavement condition surveys.
- Database for information recording.
- Analysis schemes.
- Decision criteria.
- Implementation procedures.

The decision criteria and implementation procedures are dependent on the information collected as part of the road inventory and pavement condition surveys. This paper addresses the Machine Learning (ML) component of the analysis schemes within PMS. Having effective ML tools to analyse road data can advance the sector in making critical decision on the maintenance and rehabilitation (M & R).

Balaram (2022) exposed the shortcomings of manual visual assessment methods which has been used in industry which slowed production and performance of various organisations. The study further highlighted the evolution to the existing Pavement Management System through technology. The contribution technology can make creating an effective and efficient Pavement Management System whereby synergising the components of the PMS life cycle creating an integrated solution for South Africa (SA). This includes Designs, Construction, and a fixed Asset evaluation of the road network for forecasting and budget allocation at government level which in the future can dissolve the maintenance backlog that is experienced.

As the road indices contribute to decision making for M & R, they must be accurate and a true reflection of road conditions, since delayed M & R can be detrimental financially. It is important to note that M & R activities are not solely based on road visual condition indices, but also other factors such as traffic, funding, type of road class, etc. Nonetheless, visual condition indices contribute significantly to M & R. A critical evaluation of indices is necessary to determine if the visual condition indices methods are still precise as pavement engineering is in an ever-changing state due to increase vehicle loadings and climate (which can alter pavement distress needs and importance). For example, in South Africa increasing temperatures can necessitate distresses such as rutting to be more crucial than freeze thawing. Thus, such methods require continuous research and evaluation to ensure the design of resistant flexible pavements. Discrepancies in the methods affects road maintenance plans, and budgets. Improvement in the methods can result in better road infrastructure maintenance and contribute positively to the economy as roads are the most important transportation mode for any country (Rampersad et al., 2023).

In South Africa, the Visual Condition Index (VCI), Condition Index Surfacing (CI_{SURF}), and Condition Index Pavement (CI_{PAVE}) are the main indices used (Committee of Transport Officials, 2013). The main objective of the study is to investigate the most effective index for the management of flexible pavements. Rampersad et al. (2023) noted that there were discrepancies in the methods. The VCI uses arithmetic aggregating (weights and averages all present distresses), while both the CI_{SURF} and CI_{PAVE} use a deductive point method (subtracts the 6 highest distress from 100) (Committee of Transport Officials, 2013). Both methods have their advantages and disadvantages, and must be used accurately, so that pavement networks can be managed efficiently.

The VCI is a good indicator of the pavement overall condition trend. It suffers from averaging disadvantages such as not providing localised values (a portion/part of a section with poor condition can be classified as good since it can be improved by other good sections), and it uses unrelated data combinations. The deduct point method provide localised values as per the dominant distress and extent. A disadvantage occurs when one distress is visible and significant enough to qualify as a very poor condition, that it recommended that for the index must be less than 50 % for action to be taken (Van Zyl & Van Der Gryp, 2013). In 2013, Van Zyl & Van Der Gryp proved how the VCI had shortcomings and adapted the deduct point condition index method in SA. However, the VCI is still used to avoid changes in condition trends of pavements constructed prior to 2013. This highlights a gap into determining the relationship and effects of the two methods. This will be determined by quantitative statistics analysis, and the use of Machine Learning (ML) to test which method works best with Artificial Intelligence (AI) technology.

Pavement management is in an ever-changing state, with methods needing to be constantly evaluated and improved to cater for increasing traffic, changes in the environment, climate change and new materials. This study provides correlation analysis and a transformative philosophical view on the use of ML to develop a tool to be incorporated into a PMS using ML. Thus, using ML for M & R can identify trends and relationships that may be overlooked by manual operations. A model was developed which can reduce data processing and evaluation time by categorising road section according to their conditions in timeously manner.

1.1 Aim of Paper

The aim of this paper is to provide ML algorithms to assist road authorities to provide maintenance strategies. The aim can be achieved by illustrating how ML methods can be used to determine the pavement performance based on South African pavements condition indices (VCI and CI). The second aim is to determine which of the condition index methods is the true reflection of road conditions. Thus, a conclusion can be inferred from descriptive, inferential data analysis and the applicability of ML models for this purpose.

1.2 Problem Statement

The transport network system plays a crucial role in the daily life of the citizens and the economy, as it provides employment, connects places, provide government income (Tollgates) and the mobility of people and goods. The replacement of the South African road network is estimated to cost over R2 trillion which includes flexible, rigid, block and unpaved pavements (Tetley, et al., 2022). The current replacement cost of SA's paved roads is R1.1 trillion rands, with about 160 000 km paved roads (Tetley, et al., 2022). Since most South African roads are in critical conditions, it is crucial that condition rating methods are evaluated and innovated to better service delivery as roads play an integral part to the public and the economy.

Maintenance and rehabilitation must be determined by effective methods, as funds, time and specialist labour are limited. The correct M & R must be determined at the accurate location and time. Neglecting to provide M & R in the early stages of deterioration can lead to higher future cost and provide excessive M & R needs, which is unwarranted and not economical. Thus, the conducted study aims at providing solutions to the identified problem. In addition, this work is in line with the current South African National Road Agency Ltd (SANRAL) research priority projects. The South African National Roads Agency (SANRAL) (2024) has initiated a research panel for academics and researchers to address road-related problems as well as innovation in the field. Several topics were prioritized by the research panel as being critical. Due to the dynamic nature of ML and its time-dependent relevancy, it is important to relate the research to the activities of the industry.

The SANRAL priority projects are related and share the same research interest with the study, which proves its relevancy. The following is the list of SANRAL current priority projects related to this study:

- Pavement testing & performance: Artificial intelligence / Machine learning modelling to create self-learning models for pavement performance prediction, based on LTPP and APT data.
- Pavement testing & performance: LTPP/APT/Lab/Field correlation modelling using AI and ML techniques; data mining; data correlation analysis.
- TRH6: Nomenclature and methods for describing the condition of asphalt pavements. Latest tools which need to be incorporated.
- TMH9: Pavement Visual Condition Assessment Manual Update, and linkage with TRH6.
- TMH22: Road asset manual update.

2. LITERATURE REVIEW

2.1 Background

The deduct point method was developed in the early 2010s due to the VCI no longer being a true reflection of the road condition (Van Zyl & Van Der Gryp, 2013). Van Zyl and Van der Gryp localized the deduct point method in a PMS of the Western Cape provincial government, which was later adopted to be used nationally, included in TMH 22 (2016). Although the VCI results worked considerably well in the past, as time progressed it failed due to road networks deteriorating at a higher rate and coupled with insufficient maintenance (Van Zyl & Van Der Gryp, 2013). The CI_{SURF} and CI_{PAVE} were illustrated to correctly determine funding needs, prioritizing projects, reflect true current condition of road sections and road networks, and identifying specific distress for M & R (Van Zyl & Van Der Gryp, 2013).

The VCI provides a basis for the development of new indices. The VCI uses weights (Wn) of all present distresses, and factors to calculate the index, which were provided by expert panels in SA. The CI method provide a deduct point per defect in relation to the extent and degree, then the 6 dominant distresses are multiplied by their constants to get the total deduct point which is subtracted from 100.The difference between the VCI and CI_{SURF} is that the weighting of surfacing distresses in CI_{SURF} have higher weighting compared to VCI surfacing distresses. The difference of the CI_{PAVE} compared to the VCI is that surfacing drainage has a higher weighting, while in the VCI surfacing drainage is the 4th lowest weighting (3) out of 22 distresses (Committee of Transport Officials, 2013).

The average design life of roads is about 20 years for asphalt pavements, meaning that changing the rating indices will lead to inconsistency for pavement built prior to the early 2010's. This makes the VCI still relevant. According to a study by Balaram & Mostert (2014), the deduct point method was found to have a strong correlation with the International Roughness Index (IRI). This indicates that the deduct point method is consistent. Due to the study being specific for South Africa, literature on statistical analysis among VCI, CI_{SURF} and CI_{PAVE} was limited. Literature based on other pavement indices, such as Pavement Condition Index (PCI), Overall Pavement Condition Index (OPCI), and Pavement Serviceability Index (PSI), will be used as an analysis guide, source for methodology, and criteria on how to determine the most effective method (VCI/CI) (Mubaraki & Sallam, 2021).

In a study conducted by Mubaraki & Sallam (2021) to investigate the most effective index for pavement management for urban road networks in Jazan city in the Kingdom of Saudi Arabia, it was concluded that the first choice to use was the PSI, and the second choice was the Urban Distress Index (UDI). The two indices were compared with the IRI, and the Present Serviceability Rating (Mubaraki & Sallam, 2021). The PSI was reported to be a very good index (measuring 3 major distress, namely (i) cracking, (ii) rutting, and (iii) patching), while the UDI measured all the road distresses (Mubaraki & Sallam, 2021). Figure 1 illustrates the correlation of the methods. It is shown that the UDI has a higher correlation compared to the PSI. The reason to recommend the PSI however, was because the UDI was deemed to be inefficient (requires all pavement distress types), and the PSI indicates an accurate index efficiently. Multiple literature sources recommended using an index that focus mainly on structural capacity (rutting, cracking, & patching) failure as they are more detrimental to the pavement health (Suma et al., 1998; Shah et al., 2013).

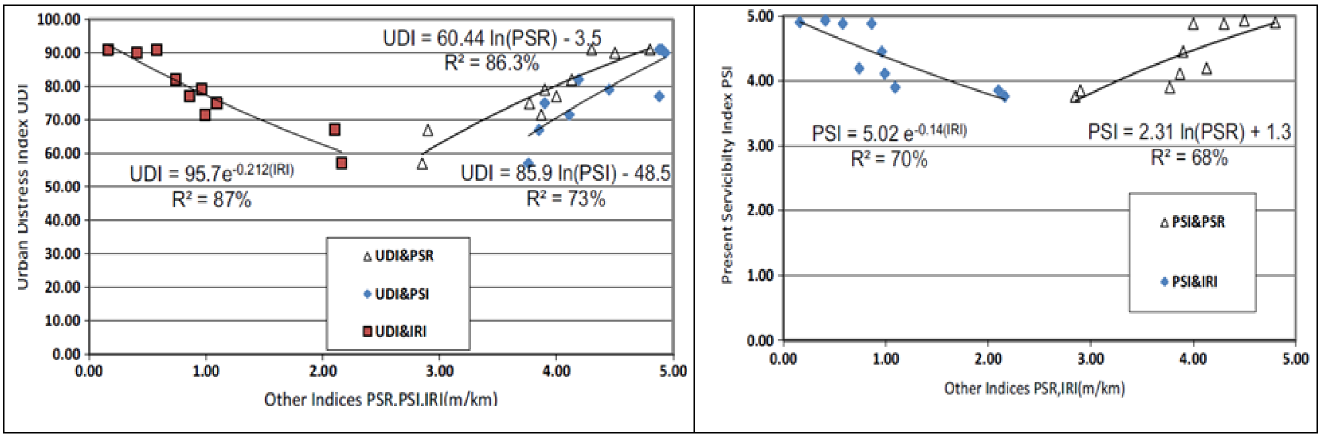

Figure 1: A comparison of pavement indices (Mubaraki & Sallam, 2021)

A study conducted to develop an overall pavement condition index for maintenance strategy, illustrated that it is crucial to analyse the index results with the type of M & R at each condition class (Kumar & Suman, 2022). When keeping other parameters constant (control variable), an index must not be conservative or overestimated. This can exponentially increase the cost when the extent of M & R is conducted unnecessarily or is not done, leading to an increase in future M & R costs. The constant parameters are factors such as available funding, traffic demand, and road class (importance). Pavement management requires the correct M & R at the right location and time since most of the time M & R funding is limited from the government or road authority (Ahmed, et al., 2017). Figure 2 illustrates how different index methods can be analysed in relation to M & R actions. Figure 2 and Table 1 illustrates how different methods can provide different M & R strategies.

Table 1: Maintenance Strategies linked with PCI/OPC

PCI/OPCI	Rating	Strategy
85-100	Excellent	Routine Maintenance (RM)
70-85	Very Good	Preventative Maintenance (PM)
55-70	Good	Minor Rehabilitation (MIR1)
40-55	Fair	Minor Rehabilitation (MIR2)
25-40	Poor	Major Rehabilitation (MAR)
10-25	Very Poor	Reconstruction (RC1)
0-10	Failure	Reconstruction (RC2)

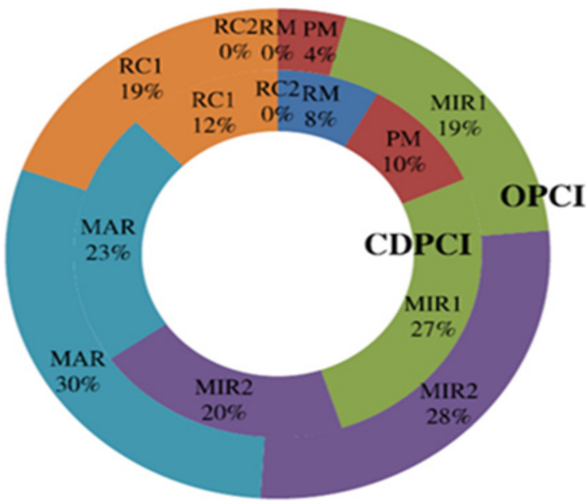

Figure 2: M & R strategies (Rajnish & Suman, 2022)

2.2 The Use of Machine for Pavement Performance Index

Machine learning (ML) is defined as a developed algorithm designed to mimic human processes/methods for analysis of certain problems by allowing the machines to automatically develop its own algorithms to solve problems (Alpaydin, 2020). Numerous studies have been published which were aimed at utilising machine learning methods to predict pavement performance. Table 2 highlights the investigated ML models. This approach has been used since the 1990s (Roberts & Attoh-Okine, 1998). Figure 3 presents the methodology used to develop models, by considering data inaccuracies, data preprocessing, and the evaluation of models. In each stage, the objectives (OBJ) are identified, and procedure/mitigation methods (M) are provided. The process is not linear, as it is dictated by the performance of the model. If the performance is low, it requires the methodology used in Stage 2 to be reevaluated and updated. When the performance still does not improve, Stage 1 must be reevaluated and updated. Incorrectly sourced and pre-processed data will result into poor model performance. This highlights the importance of Stage 1.

Table 2: ML models used for pavement condition indices prediction

Source	Type of Index (Model)	Type of Data	Results
(Issa, et al., 2022)	PCI ANN Models*	Road Distresses Extent Distresses	ML was capable of pavement performance prediction. Accuracy of 0.9975
(Hossain, et al., 2017)	IRI ANN Model	Climate Data Traffic Data IRI Data	The model had a RMSE** of 0.012 (Low error)
(Hanandeh, 2022)	SR*** PSR ANN, MLR & GA****	SR and PSR Data Pavement Age	Training Accuracy was 0.94-0.98. GA performed best

*Artificial Neuron Network (A subset of machine learning 'deep learning model')
** Root Mean Square Error
***Surface Rating- Minnesota Department of Transportation crack and surface distress index
**** MLR- Multiple Linear Regression & GA-Genetic Algorithm.

Figure 3 is an author illustration which was used to guide the process of the current study for the use machine learning. The process was grouped into 3 stages, the stages consider the objective (OBJ), Methods/Techniques (M), and Benefits (B). Stage 1 deals with data sourcing and preprocessing with the goal of obtaining high quality data. Stage 1 objectives are to identify and rectifying data inaccuracies, and data preprocessing to obtain high quality input data. The accuracy of data is crucial as the model performance is based on the data quality, hence erroneous data will output low performance and inaccurate results. Numerous engineering criteria were utilised such as the T-test method (treating the models' predictions as an assessor) sourced from THM 9-part b (2016). The T-test determine whether two sets of data are significantly different. The ML model's results in a $T_{value} < T_{95}$ the significance level rating is Non-Significant (NS). Stage 2 deals with the model development with a goal of developing models with high accuracy, this is achieved by utilising numerous different ML techniques (further discussed in section 3- Methodology). Stage 3 is model evaluation and deployment with a goal of improving methods and deploying models with the highest accuracy and performance efficiency. Stage 3 is used to refine, evaluate and identifying potential models for deployment and includes the objectives, methods, and benefits of the models in society.

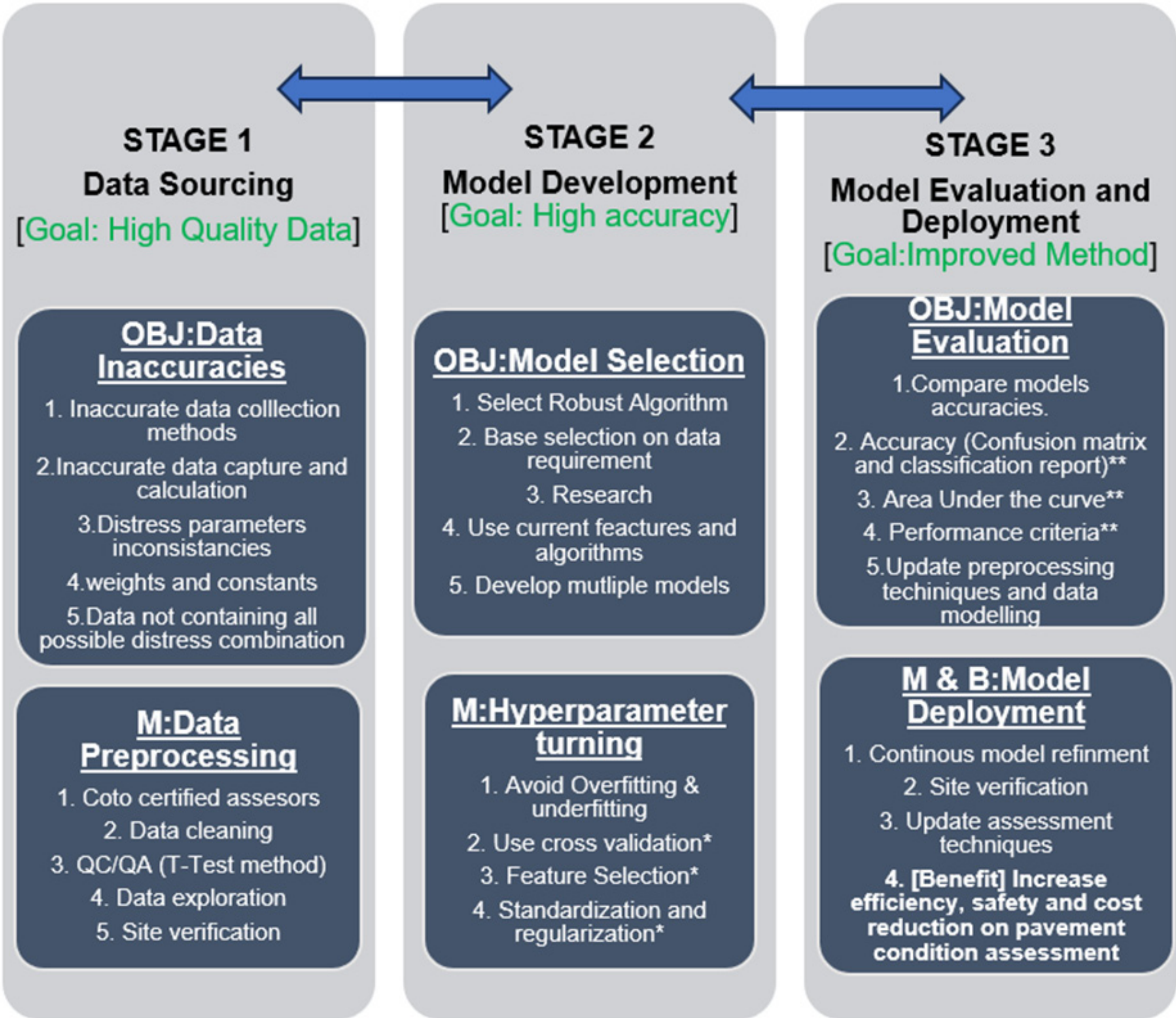

Figure 3: Author illustration on the use of ML in pavement condition prediction

*These are methods to limit the model overlearning/under-learning the data (fitting), and statistical approaches to reduce the error of the models.
* *These are different types of methods to measure the accuracy of models

3. METHODOLOGY

The data used in the study was collected from two case study projects, Case Study 1 was the visual condition data of a low traffic volume network spanning 8 km with each section ranging from 50 m to 200 m (2 way, 1 lane) (169 points) and Case Study 2 was the visual condition data from a national road (high volume roads) spanning 33 km (2 lanes, 2 way), with each point representing a 100 m section. Case study 1 comprises of 50 m flexible pavement parking lots and 500 m – 2 km major roads in a residential/business complex. Case study 2 had 1348 points (337 per lane).

The assessment was made with reference to TMH 9-part A and B for flexible pavements (Committee of Transport Officials, 2016). The case studies were sourced from past projects and the authors had no means of increasing the data to be of equal size. The simulation of data to increase the sample size was abandoned, since data simulation can lead to road distress combinations which can result in conflicting combination rating as stated in TMH 22 (2013), appendix J-4. The calculation of indices was guided by THM 22, for all the methods (VCI, CI_{SURF} and CI_{PAVE}) (Committee of Transport Officials, 2013). Table 4 highlights the data combination of the combined case studies.

As per TMH 9, road distresses are divided into engineering assessment (surfacing and structural distresses) and functional assessment (distresses for riding quality). Each distress is rated based on degree (0-5) and extent (1-5), and the product of these two aspects was used as a road distress parameter. The product is named the road defect quantity rating (Author's concept). This means that the range of the inputs is from 0 to 25. The degree rating is a measure of the severity of the distress (Committie of Transport Officials, 2016). Degree 0 is used when the distress is absent, 1-5 with 1 being slight increasing to 5 for warning-severe degrees. The extent rating is a measure of how widespread the distress is observed in relation to the total length of the assessed road section. Extent rating range from 1-5 (isolated occurrence to extensive occurrence).

Equation 1 illustrates the equation for the road defect quantity rating. This was done to reduce the complexity of the parameters which would make the model architecture more complex. See equation 1. The novelty of the study is that weights, and constant factor are not used to determine the index, as the ML algorithm will develop its own procedure to deduct the relationship between road defect rating and the index. The algorithm is only given the road defect rating (input), and the index (output), from which it finds an efficient method to predict the index by learning underlying relationships of the input and output. Pavement management is in an ever-changing state, requiring methods to be constantly evaluated and improved to cater for increasing traffic, changes in the environment, climate change and new materials. Thus, using ML algorithms for M & R can identify trends and relationships which may be overlooked by humans. With the continuous training of the models, necessary changes in weights/importance of certain distresses can be found easily and the model will adapt to future current needs.

$$Road\ defect\ quantity\ rating = Degree\ (Severity) * Extent\ (occurence) \qquad (1)$$

Table 4 present the data summary of the study. The data combination indicates a class imbalance for the very poor/reconstruction class. The very good class data distribution is 30 %, 50 %, & 54 %, and the very poor class data distribution is 17 %, 4 % and 7 % in relation to the VCI, CI_{SURF} and CI_{PAVE} respectively. This will create a problem for ML use in the CI method as the data is imbalanced, it will overlearn the very good class and can result in poor performance. For CI_{PAVE}, the imbalance would be tested by under-sampling (reducing data distribution percentage in relation to the total data percentage used) the very good class, to increase the percentage of the very poor class and test the effect of the overfitting/ underfitting when using ML. Overfitting is a scenario whereby the model overlearns the data, and not learn the principle relationships in the data. The resultant model will perform poor when encountering instance of the classes with the small data size in relation to the total data size. Underfitting occurs when a very small class in terms of total data size, will result in the model not having enough data to learn all the relationships/characteristics of the class to successfully differentiate the class from other classes in the data. Underfitting is when the model training did not reach the global minimum (state of balance between training and testing accuracy), overfitting is when there is a high mean square error meaning the model overpassed the global minimum. The CI_{PAVE} mainly measure structural distresses, thus under-sampling it will not affect the result extensively, since the CI_{SURF} is the index closely related to the VCI. The results were analysed using descriptive and inferential statistics concepts, which will be discussed in section 4.1. This was done to summarize, understand the main features of the dataset, and to draw conclusions about the population.

Table 4: VCI, CI_{SURF} & CI_{PAVE} Data Summary

Maintenance and Rehabilitation guide based on Condition Index				VCI		CI_{SURF}		CI_{PAVE}	
Type of M & R	Action	Category	Range	Points	Data %	Points	Data %	Points	Data %
Reconstruction	Design a new Road	Very Poor	0-30	255	17	64	4	104	7
Major Rehabilitation	Structural Enhancement	Poor	30-50	245	16	148	10	179	12
Minor Rehabilitation	Non-structural Enhancement	Fair	50-70	293	19	268	18	234	15
Preventative Maintenance	Retard future deterioration	Good	70-85	264	17	272	18	177	12
Routine Maintenance	Day-to-Day Activities	Very Good	85-100	459	30	764	50	822	54
TOTAL				1516	100	1516	100	1516	100

Figure 4 refers to the methodology followed while using ML. Data pre-processing aspect of ML is not yet a precise science, instead it has a rule of thumb that at least all class have a minimum of 5-10 % of the total data to avoiding underfitting or overfitting (Hanandeh, 2022).

Pre-processing techniques to solve overfitting and underfitting (Alpaydin, 2020):

- Increase data quality (having highly informative information of the minority class)
- Chosen algorithms (some algorithms are more robust to imbalance dataset, e.g. decision tress, random forests, gradient boosting)
- The use of cross-validation (helps to ensure that performance is constant)

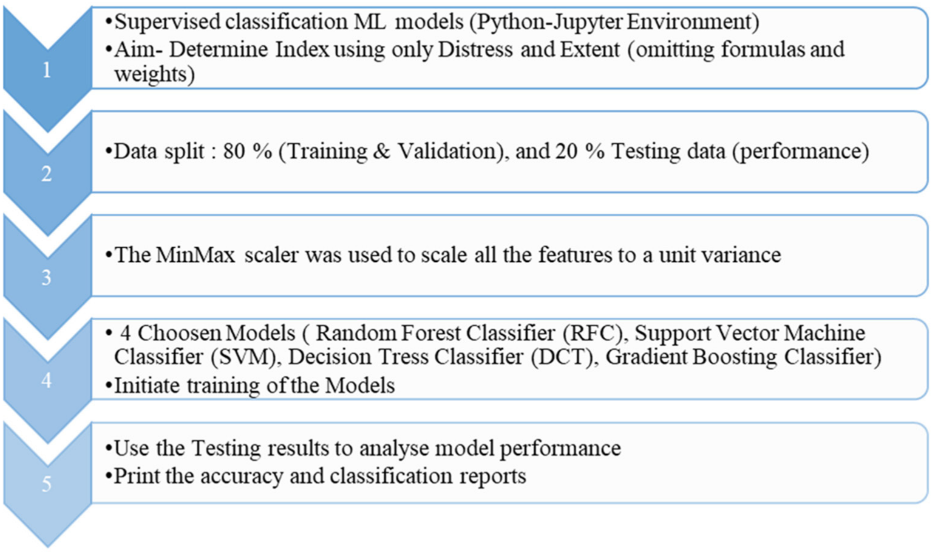

Figure 4: Author's Illustration of the Methodology for using ML

Supervised learning is a technique of ML that train the model by utilising the input data (ratings), and output (index), to learn relationships between the two and predict the output for new data. A MinMax scaler is a standardization method for the features to have a similar range (Alpaydin, 2020). The selection of the algorithm was because these types of algorithms can handle data imbalance, and a majority of literature source suggests that they perform better for pavement conditions index prediction (Ahmed, et al., 2017; Hanandeh, 2022; Issa, et al., 2022).

Table 5 present a guideline on how to use the developed model in relation to the suggested activities M & R activities. The output of the model is the type of M & R strategy, which is linked to the suggested type of actions. It is important to note that M & R of pavement depends on other factors such as traffic demands, financial budgets, etc. The type of M & R to be used depends on other factors not solely dependent on the road condition indices. It is key that the correct action be implemented at the correct location and time, thus making the accuracy of the indices methods to be important, and constantly need improvement. These guidelines were made with reference to the methodology by Rajnish & Suman (2022), localised for SA road condition classes, and pavement management principles.

Table 5: Model use guideline

Maintenance and Rehabilitation guide based on Condition Index				Suggested Action
Type of M & R	Action	Category	Range	
Reconstruction	Design a new Road	Very Poor	0-30	Full depth reconstruction, Full depth reclamation, New design
Major Rehabilitation	Structural Enhancement	Poor	30-50	Thick overlay, Mill & Overlays, Full depth patching, Premix carpet
Minor Rehabilitation	Non-structural Enhancement	Fair	50-70	Chip seal, Micro-surfacing, Thin overlay, Fog seal
Preventative Maintenance	Retard future deterioration	Good	70-85	Patching, Pothole filling, Crack sealing
Routine Maintenance	Day-2-Day Activities	Very Good	85-100	Drain Unblocking, Vegetation control, Paint, and signage replacement

4. RESULTS AND DISCUSSION

Figure 5 represents the type of M & R strategy found from the VCI and CI methods (combination of both studies). It is seen that the VCI contain more road sections in the reconstruction class, indicating that the method can lead to decisions which require more funding for M & R as reconstruction is expensive. Another class which had significant differences in the methods is the routine maintenance class. This indicates that the VCI method is conservative (gives lower ratings), with fewer sections recommended for routine maintenance, which will cost more if defects were not addressed early. The CI methods promotes the need for routine maintenance, which is more advantageous as the method promotes proactive maintenance. This can result in cost savings, as expensive M & R future needs are reduced.

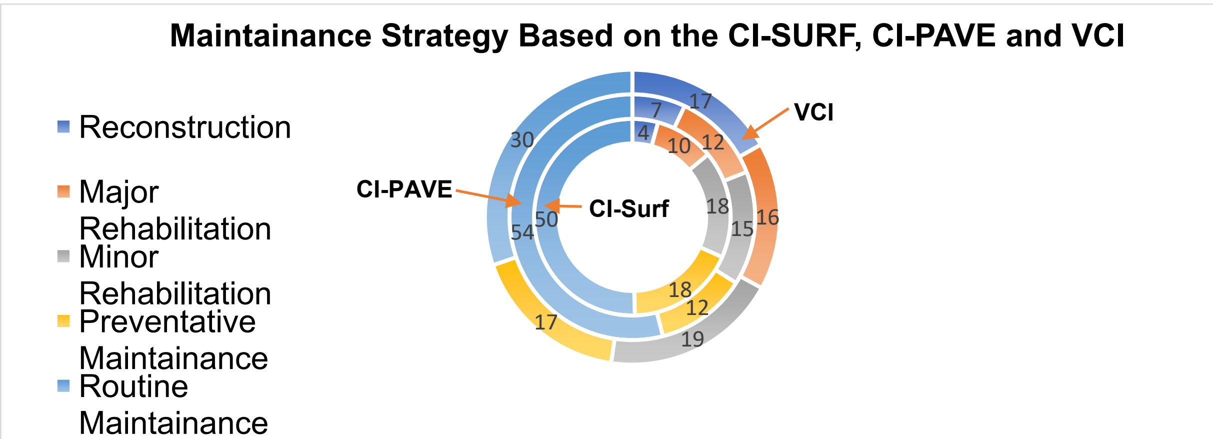

Figure 5: M & R Strategy

4.1 Statistical Analysis

Table 6 presents the descriptive data results. When considering the mean, the CI indices have higher mean values indicating that these methods produce higher average values than the VCI. The standard deviation (SD) of the VCI is higher indicating more significant variability or dispersion of the data. The lower coefficient of variation (COV) of CI_{SURF} and CI_{PAVE} indicated that the methods provide more consistent results. From this analysis the VCI performance is considered inferior.

Table 6: Descriptive Statistics of the Indices (Rampersad, et al., 2023)

	VCI	$C\text{-}_{SURF}$	$CI\text{-}_{PAVE}$	VCI	$CI\text{-}_{SURF}$	$CI\text{-}_{PAVE}$
	Case study 1			Case study 2		
Mean	38	54	56	67	81	84
SD	22	20	20	29	21	21
COV	57	37	36	42	26	25

Table 7 presents the inferential statistics results. Another measure of comparison used was the average differences (d, average absolute value difference per section) per road condition index class. The coefficient of determination (R^2) of VCI VS $CI\text{-}_{SURF}$ confirms the initial hypothesis, that CI_{SURF} is more related to the VCI, than CI_{PAVE}. The average difference between the methods is highest for the very poor class (high variation as seen in Table 7). The results indicate a higher difference in rating (d) in Case Study 2 compared to Case Study 1. This suggests that the VCI and CI methods vary more for high volume roads.

Table 7: Inferential statistics of the Indices (Rampersad, et al., 2023)

<table>
<tr><td colspan="3" rowspan="2">Index Correlation</td><td colspan="5">d (Average difference between ratings)</td></tr>
<tr><td rowspan="3">Class</td><td colspan="2">Case Study 1</td><td colspan="2">Case Study 2</td></tr>
<tr><td rowspan="3">VCI VS CI_{SURF}</td><td>Case Study 1</td><td>$R^2 = 0.75$</td><td rowspan="2">VCI VS CI-Surf</td><td rowspan="2">VCI VS CI-Pave</td><td rowspan="2">VCI VS CI-Surf</td><td rowspan="2">VCI VS CI-Pave</td></tr>
<tr><td>Case Study 2</td><td>$R^2 = 0.79$</td></tr>
<tr><td colspan="2"></td><td>0-30</td><td>22</td><td>25</td><td>28</td><td>32</td></tr>
<tr><td>VCI VS CI_{PAVE}</td><td>Case Study 1</td><td>$R^2 = 0.72$</td><td>30-50</td><td>12</td><td>13</td><td>24</td><td>29</td></tr>
<tr><td></td><td>Case Study 2</td><td>$R^2 = 0.70$</td><td>50-70</td><td>11</td><td>12</td><td>18</td><td>23</td></tr>
<tr><td colspan="3" rowspan="2"></td><td>70-85</td><td>9</td><td>14</td><td>14</td><td>18</td></tr>
<tr><td>85 – 100</td><td>5</td><td>5</td><td>7</td><td>10</td></tr>
</table>

When comparing the overall pavement conditions, the VCI tends to be more closely related to the overall pavement conditions, refer to Figure 6. The overall pavement condition is based on TMH 9-Part B (Table B.27), [very good (1) - very poor (5)]. This is due to the VCI considering all the distresses, rather than the CI_{SURF} method that considers only the 6 dominant distresses. The 6 dominant distresses are specific to the distress present on site for the road section.

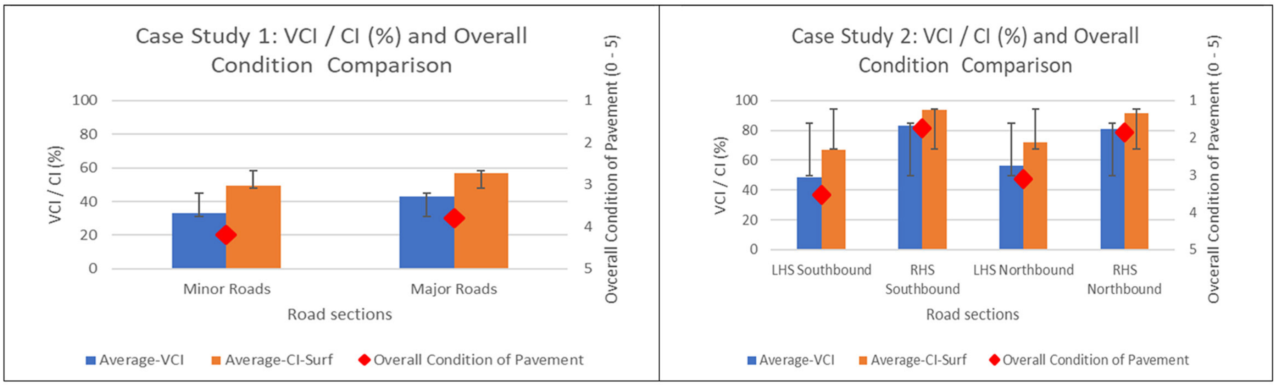

Figure 6: VCI/CI (%) VS Overall Condition (Rampersad, et al., 2023)

4.2 Use of ML

The accuracies of the models were validated by considering the predictions of the models and assuming the models as trainee assessors compared to the results obtained by approved assessors using traditional methods to calculate the indices as per TMH 9 (2016) T-test method. The model predictions indicated an overall $T_{value} < T_{95}$ for the 95% and 99% significant difference confidence interval. The data size was divided into 80 % training data and 20% testing data (performance validation). The performance of the models was tested using the unweighted and weighted accuracy averages, precision, recall and F1-Score (Refer to figure 7 for the concept definitions). The overall testing accuracy ranged from 67 - 93 %, with the VCI performing low and the CI_{PAVE} having the highest accuracy performance.

Figure 7 and 8 illustrates the classification report of the VCI and CI_{SURF}, respectively. The high accuracy in the models indicates, that ML can be used successfully in determining the indices. The model with the highest accuracy is the GBC (74 - 85%). This model can be improved and deployed by increasing the data quantity, using deep learning, and improving the data quality. The imbalance of the very poor, can be the reason the models did not have an accuracy above 95%. The minor rehabilitation class had the lowest average accuracy for the VCI, and the preventative maintenance class had the lowest accuracy for the CI_{SURF}. Figure 9 illustrates the classification report of the CI_{PAVE} index. The hypothesis test for the models trained with the under-sampled data (1138 sections), did not give a clear reason for solving the data imbalance. The model with the full dataset (1517 sections) performed on average like the under-sampled. The under-sampled data for the VCI and CI_{SURF} had low accuracy, which were immediately scrapped. This can be due to factors such as loss of information, biased representation, loss of rare patterns and dependence on random sampling. The poor performance indicates that data imbalance can be solved in this case by performing parameter turning, cross validation, sampling techniques, the choice of models such as GBC which is more robust to imbalanced datasets. However, the models trained with the CI_{PAVE} index performed better compared to the other indices with the GBC recording an accuracy of 93%. This can be due to the CI_{PAVE} mainly focusing on structural defects, making it less complex compared to an index that considers all the defects.

The aim and objectives of the study were answered by the results which highlighted that ML can be used for South African condition indices methods. The deployment proposed model can be used to successfully assist road authorities to determine M & R with a high accuracy and 95 & 99 % confidence interval. The CI methods was identified to be the most accurate method, and is compatible which machine learning methods, which can assist road authorities predict the necessary M & R to achieve satisfactory road condition.

Random Forest Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.76	0.57	0.65	51
Minor Rehabilitation	0.56	0.62	0.59	53
Preventative Maintainance	0.79	0.64	0.70	58
Reconstruction	0.68	0.94	0.79	49
Routine Maintainance	0.83	0.82	0.82	93
accuracy			0.73	304
macro avg	0.72	0.72	0.71	304
weighted avg	0.74	0.73	0.72	304

(a)

Support Vector Machine Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.72	0.57	0.64	51
Minor Rehabilitation	0.45	0.47	0.46	53
Preventative Maintainance	0.55	0.62	0.58	58
Reconstruction	0.78	0.78	0.78	49
Routine Maintainance	0.80	0.81	0.80	93
accuracy			0.67	304
macro avg	0.66	0.65	0.65	304
weighted avg	0.67	0.67	0.67	304

(b)

Decision Tree Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.66	0.57	0.61	51
Minor Rehabilitation	0.57	0.62	0.59	53
Preventative Maintainance	0.73	0.62	0.67	58
Reconstruction	0.70	0.86	0.77	49
Routine Maintainance	0.82	0.82	0.82	93
accuracy			0.71	304
macro avg	0.70	0.70	0.69	304
weighted avg	0.71	0.71	0.71	304

(c)

Gradient Boosting Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.69	0.67	0.68	51
Minor Rehabilitation	0.66	0.62	0.64	53
Preventative Maintainance	0.71	0.64	0.67	58
Reconstruction	0.73	0.90	0.81	49
Routine Maintainance	0.83	0.83	0.83	93
accuracy			0.74	304
macro avg	0.73	0.73	0.73	304
weighted avg	0.74	0.74	0.74	304

(d)

Figure 7: Classification report of the VCI, (a) RFC (b) SVM (c) DTC (d) GBC

Precision = (Positive prediction/ True positive + False positives) - How accurately the model identifies classes (models' ability to avoid false positives)
Recall = (True Positives/ True Positives + False Negatives) - How accurately the model predicts M & R for all scenarios (Mode's ability to identify all instances)
F1-Score = (2 / (1/Precision) + (1/ Recall)) - A balanced parameter to evaluate the overall performance of the classes
Macro avg = (Macro average of the Precision, Recall, F1-Score) - provides the average of all the classes as an unweighted measure for all classes
Weighted avg = (Weighted average of the Precision, Recall, F1-Score) - provides the average of all the classes according to class weights, considers class distribution (useful for data with data containing class imbalance)

Random Forest Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.68	0.66	0.67	29
Minor Rehabilitation	0.77	0.87	0.81	53
Preventative Maintainance	0.74	0.69	0.71	54
Reconstruction	1.00	0.64	0.78	14
Routine Maintainance	0.90	0.92	0.91	154
accuracy			0.83	304
macro avg	0.82	0.75	0.78	304
weighted avg	0.83	0.83	0.83	304

(a)

Support Vector Machine Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.79	0.79	0.79	29
Minor Rehabilitation	0.81	0.83	0.82	53
Preventative Maintainance	0.81	0.56	0.66	54
Reconstruction	1.00	0.93	0.96	14
Routine Maintainance	0.87	0.96	0.91	154
accuracy			0.85	304
macro avg	0.86	0.81	0.83	304
weighted avg	0.85	0.85	0.84	304

(b)

Decision Tree Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.62	0.62	0.62	29
Minor Rehabilitation	0.84	0.70	0.76	53
Preventative Maintainance	0.62	0.63	0.62	54
Reconstruction	0.61	0.79	0.69	14
Routine Maintainance	0.89	0.92	0.90	154
accuracy			0.79	304
macro avg	0.72	0.73	0.72	304
weighted avg	0.80	0.79	0.79	304

(c)

Gradient Boosting Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.81	0.86	0.83	29
Minor Rehabilitation	0.85	0.85	0.85	53
Preventative Maintainance	0.70	0.61	0.65	54
Reconstruction	0.92	0.86	0.89	14
Routine Maintainance	0.89	0.92	0.90	154
accuracy			0.85	304
macro avg	0.83	0.82	0.83	304
weighted avg	0.84	0.85	0.84	304

(d)

Figure 8: Classification report of the CI_{SURF}, (a) RFC (b) SVM (c) DTC (d) GBC

The performance of the SVM was 67, 85 & 88%, similar to the accuracies of the DTC with 71, 79 & 89% in relation to the VCI, CISURF, and CIPAVE respectively. The performance of the RFC was 73, 83 & 88%, similar to the accuracies of the GBC with 74, 85 & 93% in relation to the VCI, CI_{SURF}, and CI_{PAVE} respectively. The average accuracy of the methods was 84% for the GBC, 81% for the RFC and 80% for both the SVM and DTC.

Random Forest Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.74	0.77	0.75	30
Minor Rehabilitation	0.75	0.91	0.82	44
Preventative Maintainance	1.00	0.73	0.85	41
Reconstruction	1.00	0.75	0.86	20
Routine Maintainance	0.94	1.00	0.97	93
accuracy			0.88	228
macro avg	0.89	0.83	0.85	228
weighted avg	0.89	0.88	0.88	228

(a)

Support Vector Machine Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.85	0.93	0.89	30
Minor Rehabilitation	0.86	0.82	0.84	44
Preventative Maintainance	0.82	0.66	0.73	41
Reconstruction	1.00	0.95	0.97	20
Routine Maintainance	0.90	0.98	0.94	93
accuracy			0.88	228
macro avg	0.88	0.87	0.87	228
weighted avg	0.88	0.88	0.88	228

(b)

Decision Tree Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.88	0.77	0.82	30
Minor Rehabilitation	0.76	0.95	0.85	44
Preventative Maintainance	0.97	0.76	0.85	41
Reconstruction	0.94	0.75	0.83	20
Routine Maintainance	0.94	1.00	0.97	93
accuracy			0.89	228
macro avg	0.90	0.85	0.86	228
weighted avg	0.90	0.89	0.89	228

(c)

Gradient Boosting Classifier Report:

	precision	recall	f1-score	support
Major Rehabilitation	0.89	0.83	0.86	30
Minor Rehabilitation	0.86	0.98	0.91	44
Preventative Maintainance	1.00	0.80	0.89	41
Reconstruction	1.00	0.90	0.95	20
Routine Maintainance	0.94	1.00	0.97	93
accuracy			0.93	228
macro avg	0.94	0.90	0.92	228
weighted avg	0.93	0.93	0.93	228

(d)

Figure 9: Classification report of the CI_{PAVE}, (a) RFC (b) SVM (c) DTC (d) GBC

Figure 10 presents the Area Under the Receiver Operating Characteristic (ROC) Curve (AUC). The ROC AUC is a performance metric which compares the positive class (each class) versus other classes (false positive). It indicates how well the models identify each classification category (Alpaydin, 2020). The results can be summarized by noting that in the four developed models, the GBC has a higher accuracy for all the indices evaluated. The ROC curve also indicates that the class data imbalance does not affect the accuracy of the GBC for class identification (AUC = 97%). In addition, the GBC has a higher accuracy in identifying the critical classes that are the major and minor rehabilitation which requires immediate attention as the road would been in a critical condition and rapid mitigation is required.

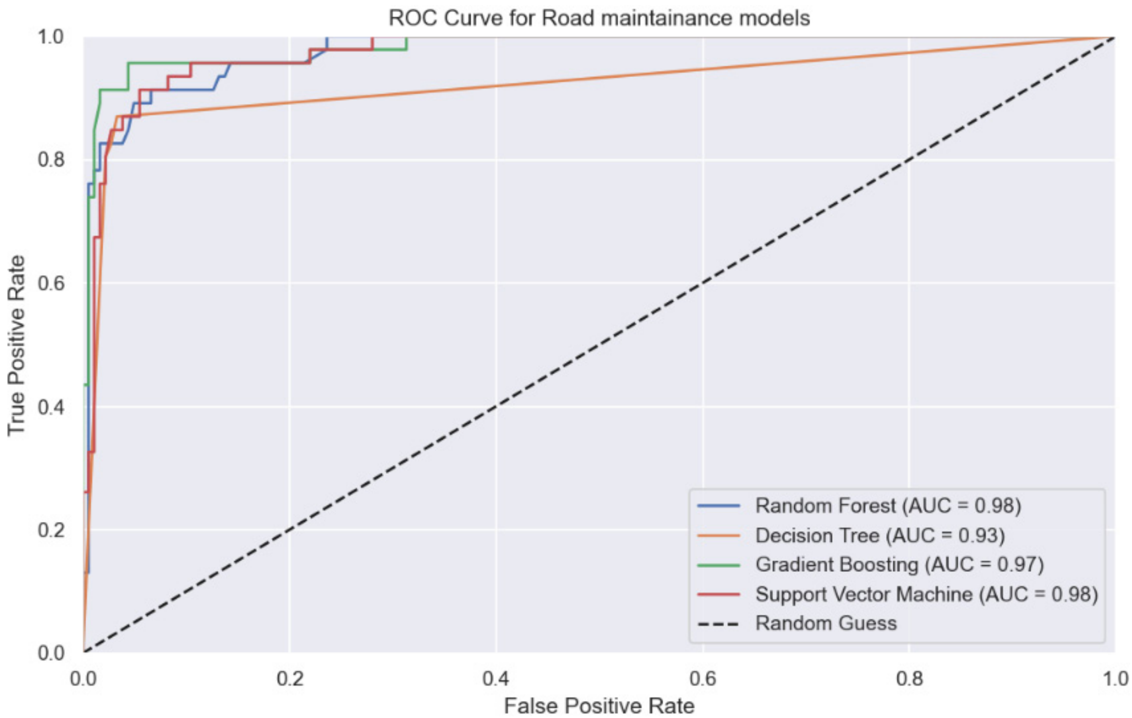

Figure 10: Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for models

5. CONCLUSIONS

It can be concluded from the results that ML can be successfully incorporated into a PMS when considering South African indices. The benefit of ML is that it can improve the pavement management methodology, reduce visual condition assessment time and cost, and can remove human error in the calculations. This ultimately will improve the economy as roads are critical for freight and traveling. The final chosen model (GBC) had an accuracy of 74, 85 and 93 %, in relation to the VCI, CI_{SURF}, and CI_{PAVE} respectively. The GBC also had the second highest AUC which was 97 %. When considering all the performance metrics, the GBC performed better, followed by RFC, SVM, and DCT respectively. The ROC curve indicates that the models have developed a systematic process and do not give random guesses.

Based on statistical analyses, it can be concluded that the VCI indicates excessive inconsistencies, which will result into higher spending on M & R. The average difference amongst the classes can be as high as32 %, and this will likely push road sections to the upper classes (e.g., meaning a road classified as very poor class using VCI can be a fair class for the CI indices). It was concluded that the most effective index is the deduct point condition indices (CI_{PAVE}) and (CI_{SURF}). Since not all distress combinations or possible combination were available, increasing the dataset to refine the models, will potentially increase the model accuracies because there was a data imbalance. The VCI M & R will result into less routine maintenance, as about 10 % more sections were suggested for major rehabilitation compared to the CI_{SURF} and $CI_{PAVE.}$ The CI method works best as it is specific for surface and pavement conditions. This will give the pavement engineer a precise mode of failure whether surfacing or pavement distress need attention, and maintenance solution can be specific. Incorporating the GBC model can help to assist road authorities to predict M & R needs.

6. RECOMMENDATIONS

The following recommendations are made:

- Further investigation is needed to improve the accuracy to at least above 98 %.
- Increasing the data availability of pavement sections so that the model can learn more relationship with the extent, distress, and the condition indices.
- More studies are needed which will utilize other types of algorithms/models to identify which model performs more accurately.
- More advanced methods should be utilized to better either overfitting/underfitting such as regularization and ensemble methods.
- There should be more collaboration with the data science industry and pavement engineering sector to increase the ideas generation for pavement management.

7. ACKNOWLEDGEMENTS

The authors wish to thank the Council for Scientific and Industrial Research (CSIR), Republic of South Africa, for funding the research into the use of Machine Learning in road management. The support of the Pavement Design and Construction (PDC) research group has been crucial to the vision of the study and offered technical support.

8. REFERENCES

Ahmed, S, Vedagiri, P & Rao, K. 2017. Prioritization of pavement maintenance sections using objective based Analytic Hierarchy Process. *International Journal of Pavement Research and Technology,* 10(2):158-170.

Alpaydin, E. 2020. *Introduction to Machine Learning.* 4 ed. MIT: MIT press.

Balaram, Y & Mostert, M. 2014. *Lessons learnt from the measurement of gravel road roughness in KwaZulu-Natal.* KwaZulu-Natal, In Proceedings of 5th SARF/IRF Regional Conference.

Balaram, V. 2022. An integrated solution for pavement management and monitoring systems in South Africa. Masters Degree. University of KwaZulu-Natal, South Africa

Committee of Transport Officials. 2013. *Draft TMH 22: Road asset management manual.* South Africa: Committee of Transport Officials (COTO).

Committe of Transport Officials. 2016. *TMH 9 manual for visual assessment of pavement roads, Part A: General.* Pretoria: South African Road Agency SOC Limited.

Committie of Transport Officials. 2016. *TMH 9 manual for visual assesstment of pavement roads, part b: flexible roads.* Pretoria: South African National Road Agency SOC limited.

Hanandeh, S. 2022. Introducing mathematical modeling to estimate pavement quality index of flexible pavements based on genetic algorithm and artificial neural networks.. *Case Studies in Construction Materials,* Volume 16, p. 00991.

Hossain, M, Gopisetti, L & Miah, M. 2917. Prediction of international roughness index of flexible pavements from climate and traffic data using artificial neural network modeling. *In Airfield and highway pavements,* pp. 256-267.

Issa, A, Samaneh, H & Ghanim, M. 2022. Predicting pavement condition index using artificial neural networks approach. *Ain Shams Engineering Journa,* 13(1):101490.

Kumar, R & Suman, S. 2022. Development of overall pavement condition index for maintenance strategy selection for Indian highways. *International Journal of System Assurance Engineering and Management,* 13(2):832-843.

Mubaraki, M & Sallam, H. 2021. The most effective index for pavement management of urban major roads at a network level. *Arabian Journal For Science And Engineering,* 46: 4615-4626.

Mubaraki, M & Sallam, H. 2021. The Most Effective Index for Pavement Management of Urban Major Roads at a Network Level. *Arabian Journal for Science and Engineering,* 46:4615-4626.

Rajnish, K & Suman, S.K. 2022. Development of overall pavement condition index for maintenance strategy selection for Indian highways. *Int J Syst Assur Eng Manag,* April, 13(2):832-843.

Roberts, C & Attoh-Okine, N. 1998. A comparative analysis of two artificial neural networks using pavement performance prediction. *Computer-Aided Civil and Infrastructure Engineering,* 13(5):339-348.

Sabita. Undated. Sabita information sheet - Pavement Management Systems. Available at: https://www.sabita.co.za/documents/sabbrochure-pms-pm.pdf

SANRAL. 2024. Sanral Research Panel. Request number RFI1002/58600/2022. Expression of Interest - Research Panel. South Africa.

Shah, Y, Jain, S, Tiwari, D & Jain, M. 2013. Development of overall pavement condition index for urban road network. *Procedia-Social and Behavioral Sciences,* Volume 104:332-341.

Suma, N, Chakraborty, N & Veeraragavan, A. 1998. *Simplified methodology for prioritisation of pavement maintainance based on pavement condition index.* s.l., 4th International Conference on Managing Pavements.

Van Zyl, G & Van Der Gryp, A. 2013. *Pavement condition description using a deduct point approach.* Pretoria, Document Transformation Technologies cc, pp. 383-393.