Assessment of pavement degradation through statistical analysis model: A case study of the Department of Transportation (DOT) of Iowa, USA

S. Rezvani* and Marta Castilho Gomes **

* Instituto Superior Técnico, Universidade de Lisboa, Portugal email address: seyedi.rezvani@tecnico.ulisboa.pt

** Instituto Superior Técnico, Universidade de Lisboa, Portugal email address: marta.gomes@tecnico.ulisboa.pt

Abstract. Transportation network, especially highways, is considered a national or international asset, and by proper maintenance system, public and private organizations can prioritize the budgeting of repair and reconstruction. The problem is to have a reliable and practical model creating a solid understanding of the pavement degradation condition by inexpensive measurable parameters for municipalities. This study focuses on the road pavement condition, particularly the statistical evaluation of the processes of degradation involved in various road sections. Quantitative statistical analysis of a sample taken from the Iowa Department of Transportation (DOT) in the United States provides a better understanding of the needs in pavement maintenance processes. In addition, it can identify the critical factors of pavement maintenance. Through a case study, it is shown that organizations can develop a solid based statistical decision-making model using basic and low-priced parameters. The model has two approaches, with and without pavement type (used by creating several dummy variables to include each pavement type as independent variables). This study will positively enhance the pavement degradation prediction through a statistical analysis model and a case study of the Department of Transportation (DOT) of Iowa, USA. These details include explanatory models, bivariate correlation, principal component analysis, hierarchical and non-hierarchical clustering, creating dummy variables, and developing multivariate regression models.

1. Introduction

In recent years, roads and highways are playing a vital role, and preserving previously constructed highways is essential. Municipalities need to preserve them by continuous maintenance programs (Mascio, 2019) to reduce degradation (Sadeghi et al., 2017), to make a higher national growth rate; they are of crucial importance. Furthermore, access to work, social, healthcare, and schooling systems make a transportation system vital for reducing poverty. Roads are connecting several regions and boost economic and social growth. Robust transport infrastructure, such as the highway network, strengthens the transport network, decreasing transportation costs, which certainly reduces the cost (Al-Suleiman et al., 2020) of production while increasing enterprises' efficiency.

The highway network also allows for more efficient and timely transport schedules and deliveries and often contributes to social development. Families are more often able to visit their loved ones, promoting great relationships with relatives and friends. Highways will reduce travel time to villages, cities, and towns while enabling individuals to travel for business and trade. Highways make journeys shorter, more convenient, and smoother than conventional roads while reducing the carbon footprint at the same time by reducing fossil fuel consumption.

Researchers developed different variables and parameters of pavement conditions that are practical tools to understand pavement conditions better. In this context, the parameters and variables that are going to use are diverse; all of them are not crucial in the prediction model; the variables are divided into eight main groups (Iowa DOT, 2020) used in this paper.

This paper aims to reduce the assessment cost of pavement degradation by using cheaper parameters in terms of evaluation to overcome time-consuming and costly pavement assessment that is not feasible to perform for many organizations and municipalities.

This paper starts with the introduction **Error! Reference source not found.** and follows by the literature review in **Error! Reference source not found.** and follows by methodology in **Error! Reference source not found.** **Error! Reference source not found.** and 0, respectively are discussing correlation coefficients matrix and non-hierarchical clustering (k-means) that are the pre-requirement of the multivariate linear regression in 0 and the result and discussion has been discussed in **Error! Reference source not found.** follows by conclusions in **Error! Reference source not found.** In this article, there are several terms that are listed and explained in Appendix A. (Iowa DOT, 2020).

2. Literature review

The most exposed part of the road deteriorating by both usage and weather condition is the pavement. So, having comprehensive pavement condition monitoring would be a critical component for the maintenance decision-makers. For example, asphalt will contract and shrink at shallow temperatures in winter, and this contraction and shrinkage create tension within the asphalt, meaning the asphalt is attempting to draw away from itself. If this strain gets very high inside the asphalt, cracks may form on the surface to release some stress (Gökçekuş et al., 2019).

Temperatures that drop quickly allow the asphalt to compress faster and create stress faster, associated with an increased risk (Alberti & Fiori, 2019) of cracking, and in spring, the snow and ice in winter start to melt in the spring. It saturates the soil under the surface as the ice below the pavement melts. The spring also brings more rain, which can puddle due to frost heave in deformed places and penetrate cracks and seep under the pavement. Asphalt deteriorates (very slowly), water is the most significant concern to asphalt pavements in the spring. The more asphalt is in contact with water, the more it is likely to damage. It means puddling water on top of asphalt, running water through asphalt pavements, water flowing into asphalt paving holes, and water under asphalt pavements. Therefore, asphalt roads and parking lots are sloped to allow the water to drain rapidly.

To avoiding water penetration, the asphalt pavements should be preserved by crack sealing as soon as cracks occur, then repairing the cracked pavement or excessively cracked areas. In summer, moisture and more sunlight during the summer cause the asphalt pavements to be gradually affected by oxidation caused by UV radiation penetration. It results from mixing light oils in asphalt to form heavier oils, and the pavement becomes more fragile and more prone to crack with more heavy oils. Finally, fall means an increase in precipitation. Moisture again becomes the most prominent threat. Fall is among the year's highest pavement construction times in terms of temporary maintenance schemes. Rainfall is the greatest obstacle to delaying the project's completion while freezing weather also triggers the freezing of water collected under the pavement. It expands when the water freezes and allows the surface to lift (David Farris, 2017). This phenomenon is known as frost heave, and if frequent maintenance or proper drainage is not constructed, it will occur below the pavement.

In the modern economy, there is always more pressure on the budget. Optimizing the way to expend budget is a vital element of transportation departments and municipalities, forcing them to analyze the pavement condition more sophisticatedly. Furthermore, in developing countries, inflation plays a crucial role in fluctuating machinery prices, workforce, materials, and equipment, resulting in the deviation of the project's design and closing costs (Musarat et al., 2020). The cost of road construction is difficult to predict (Soncim et al., 2018), and other costs (Alqadhi et al., 2018) are continually rapidly increasing, leading economic activity to become more difficult to predict. Consequently, it makes it challenging for construction companies to bid in developing countries.

The International Roughness Index (IRI) (Carvalho, n.d.) and Pavement Condition Index (PCI) (Loprencipe & Pantuso, 2017), which are indexing variables and indicating the present condition of the asphalt-based on the other variables, and historical variables show the layers past years detail. The Project Management Information System (PMIS) (Amin, 2015) and Department of Transportation (DOT) need to observe and analyze to let municipalities know when to do maintenance. However, pavement degradation (Zhao et al., 2018) variables do not differ much from other pavement with the same condition. They do not differ much between the same road and region. The database that comes out of the pavement in one region is most likely repeated in the same region. The measured items are based on the distance between two stations and, it could be used to make a new variable by weighing other parameters based on this. Based on the fact that the percentages in a sample of each pavement index are comparable due to this ratio factors to one hundred. Each pavement section contributes to the model predictions that the road pavement is repaired every time, and specific technicians and equipment must measure all the pavement parameters.

The length of the observations is 11127.64 miles (17904 km); there are geometric lengths and lane length, which is calculated by multiplying the number of the lane by a geometric length in the length variables.

Some missed data (Abdelaty et al., 2018) in the database is not because of the dataset problem's errors (Piryonesi & El-Diraby, 2020). However, it is more because of the nature of the data marked and labeled as missing data in SPSS. The outliers also need to be considered and find why they happen that could be a statistical way to interpret the odds. Time depending analysis illustrates how the time will affect a road's durability based on the frequency of maintenances.

3. Methodology

In this article, first of all, the descriptive statistics performed to analyze the effect of the variables and parameters alone and on each other (correlation coefficient); and continued by analyzing the clustering based on variables and entities to observe the clusters by performing hierarchical and non-hierarchical clustering approach. Consequently, by knowing the nature of the data and each variable, we can draw a broader picture of the interrelation, that is leads to create the multivariate regression model with and without dummy variables. Dummy variables in this study consist of the pavement types.

To this extent, the frequency and normal distribution of key variables have been calculated to indicate their characteristics and insight into what is interpreted by that variable. For example, there are two different road pavement construction year and resurfacing year variables then subtract, and the maximum of them using from a fixed date, obtaining age that is highly correlated to layer1 (top layer) year or similarly can be transferred to age. There has been a strong correlation between the pavement degradation factors and the age or the top layer construction year.

Also, this study uses a statistical approach to separate/allocate overlapping distributions to each pavement degradation as there is no revolving program available in the database. In the following analysis, the effect of each variable on each other will be calculated. For example, the traffic speed and the pavement thickness will change the pressure on the road surface, and rutting depth gets deeper if the lane is one and the vehicle cannot shift between lanes, it will affect the depth to tire in the asphalt. The study aims to find a regress to predict the high-cost indexes by the more practical ones and introduce the whole state's best timeframe for maintenance.

Furthermore, a considerable analysis of the variable is not related to finding out the true effect on each other. This research begins with indexing and simplifying, and combining factors to calculate the true variables that give the trustable results. The pavement variables were distinguished by main descriptive statistical indicators like mean and dispersion ratio, discussing their variance.

The initial phases primarily describe the variation of the surface combination, thicknesses, and geometric variables, because of the correlated bivariate correlations between these variables' combinations. The second main part describes the traffic variables and their variances, that pair of variables indicate a strong correlation. Indeed, although the variation in pavement thicknesses because of the region and design standards tends to be related to degradation factors due to the volume of vehicles traveling on the paved surface and previous maintenance activity, a low pavement thickness may be durable.

4. Correlation coefficients matrix

Correlation is to evaluate whether there is a statically linear relationship significance between variables. It is to find the direction and strength (the dispersion of this association's cases). These are commonly used to determine a relation between continuous variables and having the best picture of variables. With the correlation's help, we can narrow down the variables' size, contributing to predicting the independent variable. The linear regression aims to find out a relation between an independent variable and some other dependent variables. A strong group of points alongside a straight line indicates a linear correlation. A curved set of points can imply a nonlinear relationship, or a random pattern may show no relationship (Hair et al., 2019).

A correlation matrix would be a useful tool to understand the relation between variables. However, there are various ways of doing correlation (Gideon, 2007). The most common ones are Pearson's correlation (linear) and the Spearman correlation (non-parametric); in this research, we used both to better overview the variables while considering it when we do not have the normal distribution in a variable. Both matrices indicate that the indexes rows and columns have more density of color, both negative and positive. These colors are adjusted and made in excel to have a better and clearer view of bidirectional correlation.

After checking the bivariate correlation for both linear and non-parametric tests, we can identify there is not much difference between them, which indicates the dispersion of residuals are normal, and variables have the normally distributed variance around the mean in the bivariate correlation matrix. The next step is to look deeper into those that have the highlighted spot in the matrix. Moreover, narrow down the variable

size to finally check the scatter plot of each. Suppose we check the high correlated ones in one matrix. In that case, we can obtain the contributor in the linear regression model and find out the nonlinear ones and decide where to use dummy variables in the nominal and ordinal variables. Based on ASTM 6433 (ASTM D 6433, 2011). we can check those variables, namely, indicate the pavement condition index, which is a visual indicator of the pavement and their contribution to the PCI prediction.

However, those variables calculated have the highest Pearson correlation for pavement performance (Mohammadi et al., 2019) index. Based on (Hair et al., 2019), when we want to have more than 50 percent of the variable's explanatory power, we need to have a 0.707 correlation. We need to check out which linear bivariate correlation is high enough to give the model enough explanatory power (see FIGURE 1).

FIGURE 1. Heat map of the Pearson correlation matrix of the most effective variable for predicting the pavement performance index.

PCI is classified as class (PCI range), good (85-100), satisfactory (70-85), fair (55-70), poor (40-55), very poor (25-40), serious (10-25), and failed (0-10) based on (ASTM D 5340, 2018). We allocated a binned variable to visualize the pavement quality in FIGURE 2, which shows the highly correlated variables bivariate scatter plot and their distribution in the graph's diagonal line. This graph indicates the binned PCI in different colors: cyan and red are good and satisfactory, and the other colors mean an acceptable range. Based on road importance. it would need repair and maintenance services.

FIGURE 2. Scatter plot of highly correlated variables

5. Non-hierarchical clustering (k-means)

The procedure starts by adding selected variables into the k-mean clustering method for creating three groups or clusters for all entities, checking the result of the initial and final clustering of non-hierarchical of entities, and looking at the differences in the clusters; all 4020 entities are set into clusters which means indicating the deviation internally. This result shows the same size clusters as they are going to minimize the group mean and maximize their distances; the center of each cluster will play a crucial role in the model, and we can discriminate and get the result on our dataset and allocate the cluster to each entity (a record).

The three clusters are showing the pavement condition. It is related to the quality of the pavement, the best quality, which is the cluster number 2 center, shows the highest IRI index by 86%, structural average by 7.43, the latest the construction year 2012, and the minimum rut by just 0.1 foot as well as minimum fault average by 0.0111, highest speed among by 59mph and pavement thickness of 14.9 inches that is nested in the highest PCI binned group, number 7. This result shows the 1st and second cluster which is built constructed averagely after 2002 and has rutting better than average 0.13 foot with IRI index higher than 66% would be considered as satisfactory and good based on PCI binned; while in the 3rd cluster which the top layer age more than about 25 and have rutting of 0.15 foot and low than 50% should be added to the resurfacing schedule of DOT. That is not to say if all these conditions are met, the pavement needs maintenance, while the IRI index, which is the most practical, would be the initiator. By checking the other parameters, there would be a tool to order the urgency of the project. See

FIGURE 3.

FIGURE 3. PCI and IRI scatter plot colored by clusters (k-mean)

6. Multivariate linear regression

Automatic Linear Modelling is a way to understand the general regression model between all variables and the target variable. We consider having PCI as the most complex variable into the target and the cheap and practical obtaining variable by the predictors, resulting from the automatic linear modeling. The result shows more than 77% accuracy and the IRI (Hossain et al., 2020) as an essential predictor with 55.8% contribution in the result, and an acceptable significance level of 0.000; as the layer1_transformed with 12.8% have acceptable importance among them while it shows speed does not have a considerable contribution in the prediction result. Then, we run another multi regression with more variables and result in which some variables have a high contribution. However, we need to remove them because they cost to obtain, among others. For example, Crack_indx predicts well the PCI (Nguyen, 2017), which is the most important index in

pavement maintenance because they have already constructed PCI. So, we continue to remove it and keep removing it until we reach only three simple variables, IRI, RUT, and LAYR1, which can predict with 76.9% accuracy.

Linear regression is the main objective of this study. We obtain the regression with the variable we obtain in the automatic linear model, which helps us have coefficient and intercept. We can also check the multicollinearity and variance inflation factor (VIF). R is equal to 0.933, and R squared of 0.871, so the adjusted R square is 0.871, which means predictors predict 87.1% id PCI variance. While the model's significance is less than 0.05, the model is significant, and the F (3,3663) =8254.447, p=0.000, in ANOVA also the VIF is less than the critical value of 5. Therefore, it can be considered that there is no multicollinearity.

The model will be for the unstandardized variable is (1) and for the standardized variable is (2), are shown in the coefficients table with low standard errors.

$$PCI=0.532 * IRI_INDX-58.632 * RUT+0.129 * LAYR1-213.398 \quad (1)$$

$$PCIs=0.717 * IRI_INDXs-0.313 * RUTs+0.106 * LAYR1 \quad (2)$$

While there is also an SI metric unit with the exact age of the highest pavement layer, two new variables are computed and a new regression by this new approach.

The model will be for the unstandardized variable is (3) and for the standardized variable is (4), which is shown in the coefficients table with low standard errors.

$$PCI=0.532 * IRI_INDX-1.924 * RUT_CM+0.129 * AGE+47.873 \quad (3)$$

$$PCIs=0.717 * IRI_INDXs-0.313 * RUT_CM s+0.106 * AGE \quad (4)$$

There is also the same significance in the model as the previous one. The reason for obtaining this model is to use the SI unit system and age instead of the construction year. So it can be appliable for all the different road with the same weather and usage condition.

Using dummy variables for different surface and base types would be a chance to narrow down the result and increase the quality of the model. After generalizing the previous section regarding the variable to transfer year to age and rut from feet to cm, we decide to include dummy variables that are shown the type of the surface generally and in the subbase. So, the transportation of the variable performed by SPSS from Transform > Creates dummy variables. Creating dummy variables taken from the top surface variables like SURTYP1 and BASETYP1 is summarised and selected in TABLE 1. So, there are BT1_AAC, BT1_BAC, BT1_GSB, BT1_HMA, BT1_MSB, ST1_AAC, ST1_HMA, ST1_MMS, and ST1_PCC set of dummy variables.

The illustration of nine dummy variables get the value of 0 and 1 for each entity, so we are making new 4020x9= 36180 inputs to use these variables regarding the type of pavement.

Variable Creation	
Var Name	Label
BT1_1	BASTYP1=
BT1_2	BASTYP1=AAC
BT1_3	BASTYP1=ATB
BT1_4	BASTYP1=BAC
BT1_5	BASTYP1=BTB
BT1_6	BASTYP1=CIP
BT1_7	BASTYP1=CTB
BT1_8	BASTYP1=ECB
BT1_9	BASTYP1=GBF
BT1_10	BASTYP1=GSB
BT1_11	BASTYP1=HMA
BT1_12	BASTYP1=MSB
BT1_13	BASTYP1=MSS
BT1_14	BASTYP1=PCB

BT1_15	BASTYP1=PCC
BT1_16	BASTYP1=RAC
BT1_17	BASTYP1=RPC
BT1_18	BASTYP1=RSB
BT1_19	BASTYP1=SBF
BT1_20	BASTYP1=TBB
ST1_1	SURTYP1=
ST1_2	SURTYP1=AAC
ST1_3	SURTYP1=ARC
ST1_4	SURTYP1=BAC
ST1_5	SURTYP1=BRK
ST1_6	SURTYP1=BSC
ST1_7	SURTYP1=GND
ST1_8	SURTYP1=HMA
ST1_9	SURTYP1=MSS
ST1_10	SURTYP1=PC7
ST1_11	SURTYP1=PC8
ST1_12	SURTYP1=PCC
ST1_13	SURTYP1=RAC
ST1_14	SURTYP1=SS

TABLE 1. dummy variable list with their indicators

The selected variables for this method are all the created and selected dummy variables and the remaining variables that have been used in the previous regression models. We choose the probability of F to enter $\leq .050$ and the probability of F to remove ≥ 0.100 , which creates a general condition for all of the variables to enter and remove. TABLE 2 shows the entered and removed variables in each step for the dependent variable PCI.

As discussed before, the model summary includes the R, R square, adjusted R square Standard error of the estimate for the whole process of eleven steps. The purpose of the model is to minimize the sum of square error. In this process, the adjusted R square increases while the estimated standard error will decrease to optimize the dependent variable explanatory power. The standard error of the estimated from 7.285 surged down to 5.331, and the adjusted R square rises from 0.772 to 0.878 in model number 11.

Analysis of Variance (ANOVA) test the overall significance level of a regression model in each step as TABLE 2 indicates the sum of the square of the regression through each step increases from 657664.001 to 748235.823, and the sum of the square of the residuals is decreasing from 194504.937 in the first step to 103933.115. the significant level for all the models is the same and equal to 0.000, while the F is different for then that when the model evolves through the process, it decreases from 12392.172 to 2925.277.

TABLE 2 illustrates stepwise linear regression including dummy variables: Coefficients (Left) Excluded variables (Right) which is visible in step six when a significance level of Surface type 1 PCC is as high as 0.174 then it removed.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.878 ^a	.772	.772	7.285
11	.937 ^k	.878	.878	5.331

a. Predictors: (Constant), IRI_INDEX

k. Predictors: (Constant), IRI_INDEX, RUT, LAYR1, SURTYP1=AAC, BASTYP1=GSB, BASTYP1=MSB, SURTYP1=MSS, BASTYP1=BAC, BASTYP1=AAC

ANOVA

Model	Sum of Squares	df	Mean Square	F	Sig.
11 Regression	748235.823	9	83137.314	2925.277	.000 ^l
Residual	103933.115	3657	28.420		
Total	852168.939	3666			

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
11	(Constant)	-211.042	17.369		-12.151	.000
	IRI_INDX	.530	.005	.714	100.095	0.000
	RUT	-51.950	1.269	-.277	-40.942	.000
	LAYR1	.128	.009	.105	14.630	.000
	SURTP1=AAC	-.657	.379	-.019	-1.735	.083
	BASTYP1=GSB	2.121	.267	.048	7.949	.000
	BASTYP1=MSB	2.969	.642	.027	4.626	.000
	SURTP1=MSS	-1.489	.429	-.021	-3.474	.001
	BASTYP1=BAC	-2.209	.478	-.035	-4.620	.000
	BASTYP1=AAC	-1.344	.421	-.029	-3.195	.001

TABLE 2. Stepwise linear regression model summary, ANOVA, and coefficients

7. Result and discussion

For the optimization maintenance monitoring systems, the significant component of the research introduced particularly crucial details, which are required to be considered in the repair job. The pattern in the degradation appears to be directly correlated with the crack ratio. Also, while there is a strong correlation between the resurfacing year (layer 1) and other indexing variables, there is a weak correlation between the construction year and no resurfacing step. So, there is a chance to choose the most effective variables and use them in the calculation, which is highly the same as the latest layer construction year.

To synthesizing the data, a PCA analysis was conducted in suitable groups of variables. Then, the pavement findings are extracted from the sample with 4020-pavement sections that DOT records used for conducting a clustering analysis to promote interpretation and analysis.

Different clustering methods are adopted so that the sample is better aggregated, and various parameters and strategies are used to achieve consistencies. However, the final choice is often the form of clustering, which offers the most significant insight into the pavement properties, contributing to various pavement changes. Clustering methods are classified into hierarchical (agglomeration) and non-hierarchical (k-means) processes. Hierarchical approaches and metrics use different connection approaches and cluster processes.

The findings are compared via the coefficient of cophenetic correlations and are more robust according to this criterion with the rotated principal components. The distance between the cluster level and the ideal number of clusters is three, based on the closest neighborhood connection and Chebyshev distance, in line with the hierarchical method. By running several k-means analyses with different parameters as heuristic techniques and grouping processes, from previously established two clusters to three clusters (number of

hierarchy clusters selected). Finally, three clusters are chosen: pavement entities divide into the priority for repair, second priority, and acceptable condition.

The k-mean, non-hierarchical clustering has resulted in three key clusters that indicate the pavement's overall quality. In comparison, the hierarchical classification clustered the variable due to highly variable. For this sample of road plots, the trend is to group the type plots in a large group and then add the pavement to a larger group. Besides, the pavement's PCI and IRI are grouped into another group. Eventually, the other pavement observations, which do not have identical characteristics as most pavement observations of these five last groups, are grouped in the remaining groups separately, most often alone in one group.

Based on pavement observations are analyzed, such as the pavement condition index version 2.3 of the pavements is the highest correlated factor compare to other factors and followed by international roughness index rating and crack index. The clustering study helps us infer that research findings appear to add to other pavements of the same kind in earlier stages, based on the clustered data. This inference indicates that the 'outliers' of this subsample, in a more extensive study, would make a more similar observation of the pavement and add them of the same kind, indicating three cluster classes in larger samples.

Finally, a model was developed using multiple linear regression (Mahmood et al., 2019) to deprecate the pavement condition due to multiple variables, which are the most significant in the PMIS for the life cycle (Pérez-Acebo et al., 2018). The characteristics with a more significant impact on the pavement condition, shown in the result, are IRI, RUT, layer 1, thickness, fault, speed, and structural average.

After generalizing the previous section regarding the variable to transfer year to age and rut from feet to cm, we decide to include dummy variables that are shown the type of the surface generally and in the subbase. So, creating dummy variables from the top surface variables like SURTYP1 and BASETYP1 is summarized and selected. So, there are BT1_AAC, BT1_BAC, BT1_GSB, BT1_HMA, BT1_MSB, ST1_AAC, ST1_HMA, ST1_MMS, and ST1_PCC set of dummy variables.

The new stepwise model gives adjusted R square 0.878 with the less the standard error of the estimated equal to 5.331. the type of the material (Surface type and base type as dummy variables) is important to our model, and we can make it more accurate by doing this.

This study also shows that the pavement maintenance responsible organizations can consider the pavement type and the age and region condition as vital for calculating their road surface maintenance frequency. Furthermore, this research shows how variables and parameters affect the pavement condition and the analysis for the most efficient time to repair and create a model to predict the PCI from other practical variables obtained easily from the pavement.

8. Conclusions

To conclude, there is an obvious need to have a more significant look at road maintenance as vital for society and the economy. The responsible organization would need tools to evaluate the pavement condition with a metric that makes them decide when and with what priority they need to spend the budget. However, tons of metrics are available to illustrate the exact situation of the roads and highways. However, there would be by the roads' nature that assessing all of them for organizations would not be possible due to the cost, time, and matter of human resources. So, we distinguish among all different variables to find the cheapest and most effective one, which can have an excellent contribution to illustrate the pavement condition and predict the most challenging and complex obtaining index (PCI) for most of the maintenance group. So, after checking many methods, narrowing down the path, and reaching the most important variables: IRI, RUT, LAYR1, and introducing a linear regression model that predicts the PCI with an accuracy of 87.1%. This model is also reinforced with the various automatic linear regression model. Also, there is a chance to use some more tests inside the model to enhance the prediction. However, it would be optional and based on the need of the maintenance organization. The model can be improved by adding the weather condition and crack photography pictures with a deep learning model in future studies. Following by a stepwise regression which is obtained with the variables are extracted from the dummy variable section. The new model shows that we can have adjusted R square 0.878 with less the standard error of the estimated equal to 5.331. the type of material is important to our model, making it more accurate by doing this. As future work, there could be adding some other variables that are able to measure based on the municipality's capacities and including other information, i.e., local weather conditions. Additionally, another possible opportunity to increase the usability of the model to introduce a stochastic approach based on the result of this study besides developing a web application to make it available for maintenance in charge body to use.

REFERENCES

- Al-Suleiman, T. I., Bazlamit, S. M., & ... (2020). Pavement Deterioration Rate and Maintenance Cost for Low-Volume Roads. *MATEC Web of ...*. https://www.matec-conferences.org/articles/mateconf/abs/2020/08/mateconf_eppm2018_06002/mateconf_eppm2018_06002.html
- Alberti, S., & Fiori, F. (2019). Integrating Risk Assessment into Pavement Management Systems. *Journal of Infrastructure Systems*. [https://ascelibrary.org/doi/abs/10.1061/\(ASCE\)IS.1943-555X.0000472](https://ascelibrary.org/doi/abs/10.1061/(ASCE)IS.1943-555X.0000472)
- Alqadhi, S., Ghahari, S. A., Chen, S., & ... (2018). Costs and benefits of highway resurfacing: a case study of Interstate 465 in Indiana, USA. *Infrastructure Asset ...*. <https://www.icevirtuallibrary.com/doi/abs/10.1680/jinam.17.00036>
- Amin, M. S. R. (2015). *Pavement Management Systems: Integration of Transportation Modeling, Land Use, Economy and Indicators of Development*. spectrum.library.concordia.ca. <https://spectrum.library.concordia.ca/980669/>
- ASTM D 5340. (2018). Standard Test Method for Airport Pavement Condition Index Surveys. *American Society for Testing and Materials*, 12(Reapproved 2018).
- ASTM D 6433. (2011). Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys. *ASTM International*, D6433(11), 49. <https://www.astm.org/Standards/D6433>
- Carvalho, A. F. C. (n.d.). *Maintenance of airport pavements: the use of visual inspection and IRI in the definition of degradation trends*. 8.
- David Farris, P. E. (2017). *How Does Weather Effect Asphalt Pavement*. Bituminous Roadways Blog. <https://bitroads.com/About-Us/Blog/entryid/6/how-does-weather-effect-asphalt-pavement>
- Gideon, R. A. (2007). The correlation coefficients. *Journal of Modern Applied Statistical Methods*, 6(2), 517–529. <https://doi.org/10.22237/jmasm/1193890500>
- Gökçekuş, H., Kassem, Y., & Musa, M. K. (2019). Behavior of asphalt pavement structure under influenced of alteration climatic condition. *International Journal of Innovative Technology and Exploring Engineering*, 8(8), 3199–3206.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate Data Analysis. In *TORUS 1 – Toward an Open Resource Using Services*.
- Hossain, M., Gopiseti, L. S. P., & Miah, M. S. (2020). Artificial neural network modelling to predict international roughness index of rigid pavements. *International Journal of Pavement ...*. <https://link.springer.com/content/pdf/10.1007/s42947-020-0178-x.pdf>
- Iowa DOT. (2020). *Pavement - Iowa DOT Open Data*. IowaDOT_GIS. https://data.iowadot.gov/datasets/6e82c3f3e73b4aa2b91cc20e0848458f_0
- Loprencipe, G., & Pantuso, A. (2017). A specified procedure for distress identification and assessment for urban road surfaces based on PCI. *Coatings*. <https://www.mdpi.com/2079-6412/7/5/65>
- Mahmood, M., Rahman, M., & ... (2019). A multi-input deterioration-prediction model for asphalt road networks. *Proceedings of the ...*. <https://www.icevirtuallibrary.com/doi/abs/10.1680/jtran.16.00170>
- Mascio, P. Di. (2019). Implementation of a pavement management system for maintenance and rehabilitation of airport surfaces. *Case Studies in Construction Materials*, 11.
- Mohammadi, A., Amador-Jimenez, L., & ... (2019). Simplified pavement performance modeling with only two-time series observations: a case study of Montreal Island. ... , *Part B: Pavements*. <https://ascelibrary.org/doi/abs/10.1061/JPEODX.0000138>
- Musarat, M. A., Alaloul, W. S., & Liew, M. S. (2020). Impact of inflation rate on construction projects budget: A review. In *Ain Shams Engineering Journal*. Ain Shams University. <https://doi.org/10.1016/j.asej.2020.04.009>

- Nguyen, L. H. (2017). Research on the correlation between international roughness index (IRI) and present serviceability index (PSI), recommendations on evaluation rates in *International Journal of Engineering* <https://pdfs.semanticscholar.org/96b7/5cd259a6de320d7ebb59d46d0f3f3755e816.pdf>
- Pérez-Acebo, H., Bejan, S., & Gonzalo-Orden, H. (2018). Transition Probability Matrices for Flexible Pavement Deterioration Models with Half-Year Cycle Time. *International Journal of Civil Engineering*, 16(9), 1045–1056. <https://doi.org/10.1007/s40999-017-0254-z>
- Piryonesi, S. M., & El-Diraby, T. E. (2020). Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. *Journal of Transportation Engineering, Part B: Pavements*, 146(2), 4020022. <https://doi.org/10.1061/jpeodx.0000175>
- Sadeghi, J., Najafabadi, E. R., & ... (2017). Development of degradation model for urban asphalt pavement. ... *Journal of Pavement* <https://www.tandfonline.com/doi/abs/10.1080/10298436.2015.1095912>
- Soncim, S. P., de Oliveira, I. C. S., Santos, F. B., & Oliveira, C. A. de S. (2018). Development of probabilistic models for predicting roughness in asphalt pavement. *Road Materials and Pavement Design*, 19(6), 1448–1457. <https://doi.org/10.1080/14680629.2017.1304233>
- Zhao, B., Silva, E., & Soga, K. (2018). Pavement degradation: a city-scale model for San Francisco, USA. *Proceedings of the Institution of Civil* <https://www.icevirtuallibrary.com/doi/abs/10.1680/jsmic.18.00001>

Appendix A.

- NULL means that it can leave empty for entry of value.

Name	Units	Description	Alias	Valid Values
Section Control Info				
ORIGKEY		Original smart key	Primary Key	String Value {Route, System, Direction, Begin Post, End Post, County}
PMISYR	year	PMIS year	Pavement Management Year	1981 to 9999
ROUTE_ID		Route identification number	Route ID	String Value
FROM_MEASURE	mile	LRS begin milepost	From Measure	Numeric Value
TO_MEASURE	mile	LRS end milepost	To Measure	Numeric Value
SYSTEM		Highway system	Highway System	1 =Interstate 2=US 3 = Iowa
ROUTE		Route	Route	1 to 998
DIR	direction	Direction	Direction	1 = non-divided or divided, NB/EB 2 = divided only, SB/WB
BPOST		Beginning milepost	Beginning Milepost	String Value
EPOST		Ending milepost	Ending Milepost	String Value

PAVTYP		Pavement type	Pavement Type	1 = Concrete Pavement 2A = Continuous Reinforced Concrete with Asphalt Treated Base 2B = Continuous Reinforced Concrete with Granular Subbase or Cement Treated Base 3 = Composite Pavement 3A = Composite with Jointed Concrete Pavement 3B = Composite with Continuous Reinforced Concrete Pavement 4 = Asphalt
DESCRIPT		Pavement management section description	Description	String Value NULL
CONYR	year	Year of construction or reconstruction	Construction Year	1900 to 9999
RESYR	year	Year of last resurfacing	Resurfacing Year	1900 to 9999 NULL
AGE	year	Number of years since last resurface/construction	Age	Numeric Value NULL
PMIS_LENGTH	mile	Length of roadway calculated from geometry	Geometry Length	Numeric Value NULL
LANE_MILES	mile	Calculated lane miles	Lane Miles	Numeric Value NULL
MDIST		Maintenance district area	Maintenance Region	1 = Central 2 = North East 3 = North West 4 = South West 5 = South East 6 = East Central
COUNTY		County number	County Number	1 to 99
CITY		City number	City Number	15 to 8637 NULL
Condition Data				
PCI_2	index	Pavement condition index version 2.3	Pavement Condition Index Version 2.3	0 to 100

PCI_2DEF		Flags important changes to PCI_2	Pavement Condition Index Version 2.3 Default Flag	M = Manipulated Segment R = Reset Values F = Manual Fixes to data X = Old or unknown distress data NULL
RUT_INDX	index	Rutting index rating	Rutting Index	0 to 100 NULL
IRI_INDX	index	International roughness index rating	International Roughness Index Rating	0 to 100 NULL
FAULT_INDX	index	Faulting index rating	Faulting Index	0 to 100 NULL
CRACK_INDX	index	All cracking combined index rating	Cracking Index	0 to 100 NULL
IRI	inch / mile	International roughness index	International Roughness Index	Numeric Value NULL
FRICT		Ribbed tire friction value	Friction Value	0 to 75 NULL
FRIDATE	date	Friction test date	Friction Test Date	DD/MM/YYYY NULL
FAULT	inch	Average faulting only on faulted joints in a segment	Max Faulting	Numeric Value NULL
FAULTAV	inch	Average faulting on all joints in a segment	Average Faulting	Numeric Value NULL
RUT	foot	Rut depth	Rut Depth	Numeric Value NULL
CRACK_RATIO	percent	Percent of cracking for segment	Crack Ratio	0 to 100
T_INDX	index	Transverse cracking index rating	Transverse Cracking Index	0 to 100 NULL
LLW_INDX	index	Longitudinal and longitudinal Wheel path Cracking Index Rating	Longitudinal + Longitudinal Wheel path Crack Index	0 to 100 NULL
L_INDX	index	Longitudinal Cracking Index Rating	Longitudinal Cracking Index	0 to 100 NULL
LW_INDX	index	Longitudinal Wheel path Cracking Index Rating	Longitudinal Wheel path Cracking Index	0 to 100 NULL
A_INDX	index	Alligator Cracking Index Rating	Alligator Cracking Index	0 to 100 NULL

Distress Data				
ACRACKH	square foot / mile	High Severity Alligator Cracking	Alligator Cracking - High	Numeric Value NULL
ACRACKM	square foot / mile	Moderate Severity Alligator Cracking	Alligator Cracking - Medium	Numeric Value NULL
ACRACKL	square foot / mile	Low Severity Alligator Cracking	Alligator Cracking - Low	Numeric Value NULL
TCRACKH	count / mile	High Severity Transverse Cracks / mi	Transverse Cracks High	Numeric Value NULL
TCRACKM	count / mile	Moderate Severity Transverse Cracks / mi	Transverse Cracks Medium	Numeric Value NULL
TCRACKL	count / mile	Low Severity Transverse Cracks / mi	Transverse Cracks Low	Numeric Value NULL
LCRACKH	foot / mile	High Severity Longitudinal Cracks (ft/mi)	Longitudinal Cracks High	Numeric Value NULL
LCRACKM	foot / mile	Moderate Severity Longitudinal Cracks (ft/mi)	Longitudinal Cracks Medium	Numeric Value NULL
LCRACKL	foot / mile	Low Severity Longitudinal Cracks (ft/mi)	Longitudinal Cracks Low	Numeric Value NULL
LCRACKW H	foot / mile	High Severity Long Cracks - wheel path (ft/mi)	Longitudinal Wheel path Cracks High	Numeric Value NULL
LCRACKW M	foot / mile	Moderate Severity Long Cracks - wheel path (ft/mi)	Longitudinal Wheel path Cracks Medium	Numeric Value NULL
LCRACKW L	foot / mile	Low Severity Long Cracks - wheel path (ft/mi)	Longitudinal Wheel path Cracks Low	Numeric Value NULL
DCRACKH	count / mile	Joints w/High Severity DCRACK / mi	D Cracking - High	Numeric Value NULL
DCRACKM	count / mile	Joints w/moderate Severity DCRACK / mi	D Cracking - Medium	Numeric Value NULL
DCRACKL	count / mile	Joints w/Low Severity DCRACK / mi	D Cracking - Low	Numeric Value NULL
JTSPALLH	count / mile	Joints w/High Severity Spalling / mi	Joints Spalling High	Numeric Value NULL

JTSPALLM	count / mile	Joints w/Moderate Severity Spalling / mi	Joints Spalling Medium	Numeric Value NULL
JTSPALL_LONG	count / mile	Longitudinal joints with spalling / mi	Longitudinal Joint Spalling	Numeric Value NULL
JTSPALL_TRANS	count / mile	Transverse joints with spalling / mi	Transverse Joint Spalling	Numeric Value NULL
PATCHAB	square foot / mile	Moderate Severity Patch-Bad Condition (ft^2/mi)	Bad Condition Patching	Numeric Value NULL
PATCHAG	square foot / mile	Moderate Severity Patch-Good Condition (ft^2/mi)	Good Condition Patching	Numeric Value NULL
PATCHES	count / mile	Number of patches / mi	Number of Patches	Numeric Value NULL
COVERAG E		Portion of section covered by Condition Data (%)	Condition Data Percent	0 to 100 NULL
CAPDAT	date	Crack & Patch collected by vender Test Year	Distress Testing Year	DD/MM/YYYY NULL
Structural Data				
STRUC80		80% Structural Rating	80% Structural Rating	0.00 to 15.00 NULL
STRUCT_N EED80		80th percentile Structural Need	80% Structural Need	Numeric Value Null
STRUCAV		Average Structural Rating	Average Structural Rating	0.50 to 15.00 NULL
STRUCJTR		Ratio between joint and mid-panel average structural rating values for PCC	Structural Joint Ratio	0 and 200 NULL
AVEK	PSI / inch	Average "K" Rating	Average K Rating	Numeric Value NULL
FWD_DATA E	date	Falling weight deflectometer date	Falling weight deflectometer date	DD/MM/YYYY NULL
Traffic Data				
TRYR	year	Traffic Year	Traffic Year	1981 to 9999 NULL
AADT	count / day	Average Annual Daily Traffic by	Average Daily Traffic	Numeric Value NULL

		direction		
TRUCKS	count	ADT - Trucks	Truck Count	Numeric Value NULL
Detail Data				
NHS		Boolean Indicator of National Highway System	National Highway System	Y = Yes N = No C = Connection
SYS_NHS		System break by NHS and Interstate	National Highway System Breakout	Interstate NHS Non-NHS NULL
URBAN		Urban area code	Urban Area Code	0 to 999 NULL
PCLASS		Planning Classification	Planning Classification	1 to 5
FCLASS		Federal Functional Class	Federal Functional Class	1 to 7
MEDIAN		Boolean Indicator of Median	Boolean Indicator of Median	Y = Yes N = No
MEDIAN_PCT	percent		Percent Median	
LANES	count	Number of Lanes	Lane Number	2 to 9
CONRTE1		1st concurrent route	1st Concurrent Route	0 to 998 NULL
CONRTE2		2nd concurrent route	2nd Concurrent Route	0 to 998 NULL
AREA		Maintenance area	Maintenance Area	1 to 9 NULL
SPECIAL		Boolean indicator of special section	Special Section Indicator	E = Experimental M = Multipavement S = SHRP Site NULL = Normal Section
SPEED	mile / hour	Speed limit	Speed Limit	0 to 100 NULL

SURFACE_ TYPE		Surface Type	Surface Type	0 = Unknown 3 = Grade and drained earth 20 = Gravel or stone 30 = Bituminous 31 = Bituminous over gravel or stone 60 = Asphalt 65 = asphalt over Portland cement concrete 69 = asphalt over asphalt 70 = Portland cement concrete 74 = Portland cement concrete not reinforced 76 = Portland cement concrete fully reinforced 77 = Portland cement over Portland cement 79 = Portland cement concrete over asphalt 81 = Brick 92 = Combination Surface
PAVTHICK	inch	Pavement thickness	Pavement Thickness	Numeric Value NULL
TPCCDEPTH	inch	Total PCC depth	Total Concrete Depth	Numeric Value NULL
TACCDEPTH	inch	Total asphalt depth	Total Asphalt Depth	Numeric Value NULL
BASEDEPTH	inch	Total construction base depth	Total Base Depth	Numeric Value NULL
MLEVEL		Maintenance service level	Maintenance Service Level	A to D
WIDTH	foot	Pavement Width	Pavement Width	Numeric Value NULL
WDRIVEN		Boolean Indicator of a Widened Driving Lane	Widened Driving Lane Indicator	Y = Yes NULL
RCURB		Boolean indicator of a Curb or Gutter on the right side	Right Curb Indicator	Y = Yes N = No NULL
LCURB		Boolean indicator of a Curb or Gutter on the left side	Left Curb Indicator	Y = Yes N = No NULL
ISHLDTIE		Boolean Indicator of Inside Shoulder Tied	Inside Shoulder Tied	Y = Yes N = No

ISHLDTYP		Inside Shoulder Type	Inside Shoulder Type	E = Earth G = Granular N = No shoulder P = Paved PE = Paved / Earth PG = Paved / Granular PP = Paved / Paved Paved? = Unknown
ISHLDWID	foot	Inside Shoulder Width	Inside Shoulder Width	Numeric Value NULL
OSHLDTIE		Boolean indicator of Outside shoulder Tied	Outside shoulder Tied	Y = Yes N = No
OSHLDTYP		Outside Shoulder Type	Outside Shoulder Type	N = No shoulder E = Earth G = Granular P = Paved PE = Paved / Earth PG = Paved / Granular PP = Paved / Paved ? = Unknown NULL
OSHLDWID	foot	Outside Shoulder Width	Outside Shoulder Width	Numeric Value NULL
Construction history				
COMPLEX		Boolean Indicator of Multiple Surface Types	Multiple Surface Types	Y = Yes N = No NULL
SUBDRAIN		Subdrain Project Number	Subdrain Project Number	String Value NULL
SUBDPCT		Subdrain--Percent of Section	Subdrain Percent	0 to 100 % NULL
SUBDMULT		Boolean Indicator of Multiple Subdrain Projects	Multiple Subdrain Indicator	Y = Yes N = No NULL
TREATMENT		Surface Treatment (Friction Surface Type)	Surface Treatment Type	AC = Asphalt MS = Micro-surfacing PC = Concrete SC = Seal Coat SS = Slurry Seal NULL
LAYR (1-8)		Layer Year # (1-8)	Layer Year	1900 to 9999 NULL
PROJECT (1-8)		Project Number # (1-8)		String Value NULL

PROJTYP (1-8)		Project Type # (1-8)		W = widening project L = left (inside) lane(s) only R = right (outside) lane(s) only V = various locations NULL = entire pavement width
SURTYP (1-8)		Surface Type # (1-8)		AAC = Type A Asphalt Cement Concrete ARC = Asphalt Rubber Cement Concrete ASC = Asphalt-Sand Surface Course ATB = Asphalt Treated Base BAC = Type B Asphalt Cement Concrete BRK = Brick BSC = Bituminous Seal Coat BTB = Bituminous Treated Base CTB = Cement Treated Base GND = Diamond Grinding HMA = Hot Mix Asphalt MSS = Micro Surfacing PC7 = 10"-7"-10" PCC Concrete Slab PC8 = 10"-8"-10" PCC Concrete Slab PCC = Portland Cement Concrete RAC = Recycled Asphalt Cement Concrete RPC = Recycled PCC Pavement RSB = Rolled Stone Base SAS = Soil-Aggregate Subbase SLS = Soil-Lime Subbase SS = Slurry Seal TBB = Type B Asphalt Base NULL
SURTHK (1-8)		Surface Thickness # (1-8)		Numeric Value NULL

BASTYP (1-8)		Base Type # (1-8)		AAC = Type A Asphalt Cement Concrete ATB = Asphalt Treated Base BAC = Type B Asphalt Cement Concrete BRK = Brick BTB = Bituminous Treated Base CIP = Cold in Place Recycled Asphalt CTB = Cement Treated Base ECB = Econocrete Base FOA = Foamed Asphalt GBF = Granular Backfill GSB = Granular Subbase HMA = Hot Mix Asphalt MSB = Modified Subbase PCB = Portland Cement Base PCC = Portland Cement Concrete RAC = Recycled Asphalt Cement Concrete RPC = Recycled PCC Pavement RSB = Rolled Stone Base SAS = Soil-Aggregate Subbase SBF = Special Backfill SCS = Soil-Cement Subbase TBB = Type B Asphalt Base ? = Unknown NULL
BASTHK (1-8)		Base Thickness # (1-8)		Numeric Value NULL

SUBTYP (1-8)		Subbase Type # (1-8)		AAC = Type A Asphalt Cement Concrete ATB = Asphalt Treated Base BAC = Type B Asphalt Cement Concrete BTB = Bituminous Treated Base CIP = Cold in Place Recycled Asphalt CTB = Cement Treated Base FOA = Foamed Asphalt GSB = Granular Subbase HMA = Hot Mix Asphalt MSB = Modified Subbase PCB = Portland Cement Base PCC = Portland Cement Concrete RAC = Recycled Asphalt Cement Concrete RPC = Recycled PCC Pavement RSB = Rolled Stone Base SAS = Soil-Aggregate Subbase SBF = Special Backfill SCS = Soil-Cement Subbase SGT = Subgrade Treatment SLS = Soil-Lime Subbase TBB = Type B Asphalt Base ? = Unknown NULL
SUBTHK (1-8)		Subbase Thickness # (1-8)		Numeric Value NULL
RMVTYP (1-8)		Removal Type # (1-8)		GND = Grinding MIL = Milling SCR = Heater Scarification NULL
RMVTHK (1-8)		Removal Thickness # (1-8)		Numeric Value NULL
AGGCLAS (1-8)		Aggregate Class # (1-8)		1 = Class 1 Durability (<10 yrs) 2 = Class 2 Durability (10- 20 y) 3 = Class 3 Durability (>20 yrs) I = Class 3i Durability (20-30) NULL

AGGTYP (1-8)		Aggregate Type # (1-8)		C.LST. = Crushed Limestone GRAVEL = Gravel NULL
AGGSRC (1-8)		Aggregate Source # (1-8)		String Value NULL
REMARKS (1-8)		Remarks # (1- 8)		String Value NULL
Hyperlinks				
Road Analyzer (Chrome)		Hyperlink to Road Analyzer for users with default Chrome Browser		
Road Analyzer (I.E.)		Hyperlink to Road Analyzer for users with default Internet Explorer Browser		
Path web (I.E.)		Hyperlink to Path web		