Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

# Proposal on Implementing Machine Learning with Highway Datasets

Steve Efe[1], Mehdi Shokouhian[2]
[1] Department of Civil Engineering,
Morgan State University, Baltimore Maryland, U.S

***Abstract*:- Every year State Highway Agencies (SHA) invests millions of dollars into testing materials to optimize engineering designs, and challenges are bound to be encountered in the areas of projects cost, rehabilitation process, delays in construction activities, rate of materials, maintenance costs, risk analysis etc. which are highly complicated in nature. Many Departments of Transportation regularly evaluate the condition of infrastructure through visual inspections, nondestructive evaluations, image recognition models and learning algorithms. State Highway Agencies (SHAs) across the United States are now able to collect a large amount of pavement condition information because of these advanced technological data collection methods. SHA and many other agencies collect pavement performance data, which encompass measurements of the international roughness index (IRI) and rut depth using electronic sensing devices that utilize laser, acoustic, and infrared technologies. These agencies also use imaging technologies and automated image processing techniques to estimate the levels of severity and extent of surface distresses. These effort by SHA result in high-density data that is used to support a variety of decision-making. The challenge is how these historical data can be managed and analyzed while extracting these data to improve practices and decision-making processes. With the advent of big data, from large-scale databases to data mining applications, there has been a tremendous progress in machine learning intelligence for understanding and optimizing engineering processes. This provides an opportunity for SHA to implement machine learning (ML) for large datasets in materials and testing including pavement data, construction history, slope stability, and geologic risk. There can be significant cost savings in many SHA transportation and infrastructural projects geared from supervised learning which leverages on historic data sets. Artificial Intelligence (AI) techniques like fuzzy logic, case-based reasoning, probabilistic methods for uncertain reasoning, classifiers and learning methods, Artificial Neural Networks (ANN), Genetic Algorithms and hybrid techniques have been widely used in the many applications in the engineering field. Thus, a dynamic pavement condition algorithm that allows agencies to detect pavement segments in need of rehabilitation would be beneficial for a variety of decision making including pavement maintenance performance evaluation, pavement deterioration model development, and budgeting. This project develops a pavement test condition- prediction algorithm using ANN (Artificial Neural Network) to dynamically detect untested road segments for SHA programming. This project uses an artificial neural network simulator to suggest locations for pavement testing program from historical datasets.**

## INTRODUCTION

Every year State Highway Agencies (SHA) and many other agencies collect pavement performance data, which encompass measurements of the international roughness index (IRI) and rut depth using electronic sensing devices that utilize laser, acoustic, and infrared technologies. These agencies also use imaging technologies and automated image processing techniques to estimate the levels of severity and extent of surface distresses. These effort by SHA's result in high-density data that is used to support a variety of decision-making. The challenge is how these historical data can be managed and analyzed while extracting these data to improve practices and decision-making processes. With the advent of big data, from large-scale databases to data mining applications, there has been a tremendous progress in machine learning intelligence for understanding and optimizing engineering processes. This provides an opportunity for SHA to implement machine learning (ML) for large datasets in materials and testing including pavement data, construction history, slope stability, and geologic risk.

This project seeks to develop and implement an automatic procedure for screening and recommending roadway sections for SHA pavement rehabilitation program. It uses the artificial neural network to reduce the level of effort required to identify candidate sections for the pavement in need of test and repair.

To accomplish this, the objectives of this project are:

1. Extraction and cleaning of pavement datasets such as cracking, rutting, faulting, pavement condition index and IRI.
2. Development of an artificial neural network model for screening and selecting untested roadway sections in order to support support pavement management decisions.
3. Validation of prediction model and integration into existing work processes.

## LITERATURE REVIEW

***Artificial Neural Network for Condition Assessment:***

Change in pavement surface roughness over time is one of the most important performance indicators of rehabilitation needs, because it affects vehicle ride quality and dynamic loads. One of the main objectives of every road authority is to provide a comfortable ride for users, and pavement roughness is a good indicator of whether this criterion will be fulfilled. Therefore, International Roughness Index (IRI) has been used in PMS as the major indicator of pavement functional performance [3]. Various studies have been conducted regarding pavement deterioration trends and factors affecting performance. Most of these studies have

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

limitations such as the correlation of input variables, and difficulty of variables data collection, among others [7]. Due to the large number of variables and the complex ways in which they affect one another, use of simple statistical approaches such as linear regression is not a viable means to develop pavement roughness prediction models. In addition, the shape of pavement performance curve is not known beforehand and multiple arrangements have to be tested in order to develop a model using nonlinear regression [9]. Hence, studies have attempted to use computational intelligence techniques such as artificial neural networks (ANN) to develop more accurate models [1]. Positive results on ANN application for evaluation of present pavement performance have encouraged researchers in the application of neural networks for prediction tasks. A large number of researchers deal with developing pavement distress prediction models. These models are aimed at predicting progression of single distress (e.g. cracks, roughness, rutting) or combination of various distresses through pavement performance indicator.

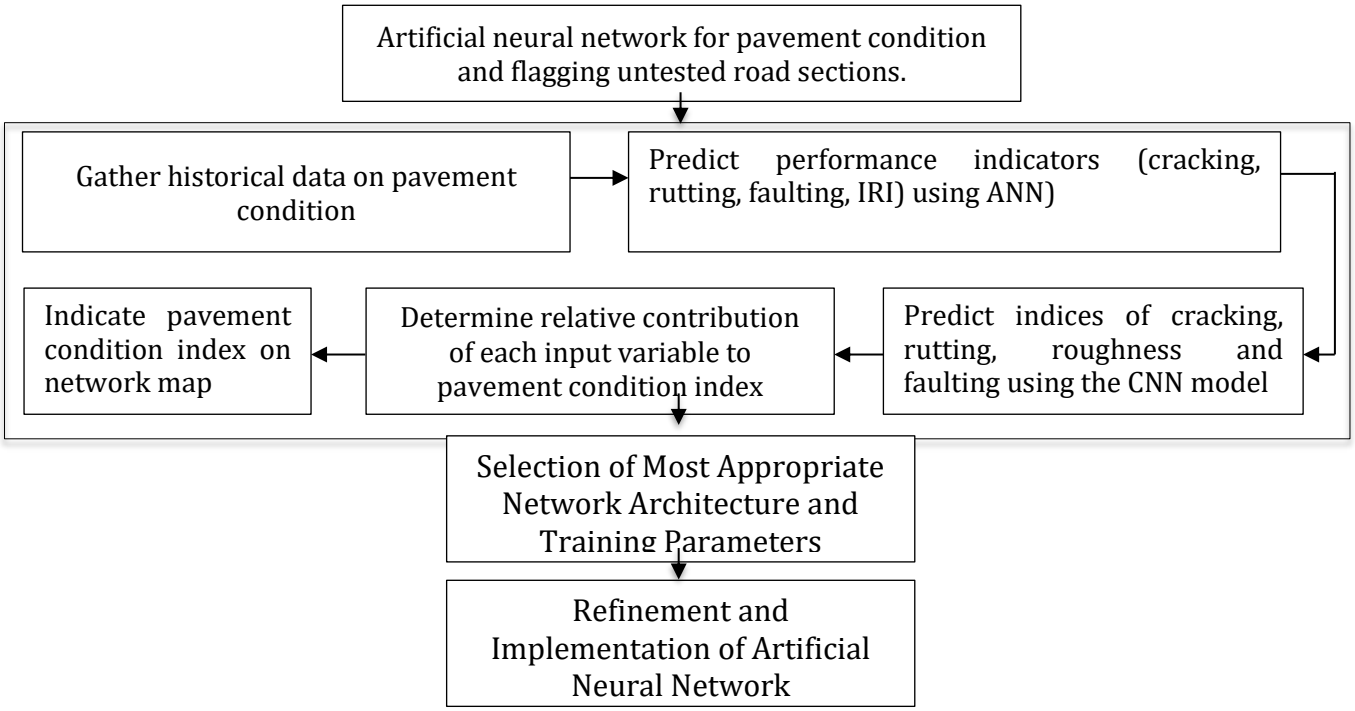


Figure 1: Research Approach Summary

It was suggested that four unified ANN models are needed to predict the progression of different pavement distresses (cracking, raveling, rut depth and roughness) on low volume roads [8]. Obtained results indicate high correlation between observed and ANN predicted distresses. The results showed that suggested ANN models would be useful in the accurate prediction of investigated distresses. Three individual ANN models for prediction of three key indices, crack rating, ride rating, and rut rating used by Florida Department of Transportation for pavement evaluation purposes were developed [10]. Results of the combination of the individual models suggest that the developed ANN models have the capability to satisfactorily forecast the overall pavement condition index up to a future period of five years. Roughness level probability prediction using multiple linear regression and two ANN was studied [4]. Obtained results indicate that ANNs have a superior ability to predict the probability of roughness distress level compared with multiple regression methods. A back-propagation ANN was constructed for prediction of International Roughness Index (IRI) based on distress rating results obtained from pavement video images [5]. Results of conducted study showed high correlation between IRI and the distress variables, which lead to a conclusion that IRI may completely reflect pavement distress conditions. Possibility of using ANN as a tool for screening and condition rating of pavement was studied. [2]. They developed neural network system for the determination of flexible pavements condition rating based on cracking and rutting indices. Conducted study shows that neural networks are capable of accurately determining pavement condition rating in a systematic and objective manner. Correlating the pavement roughness to other performance measures was studied [6].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

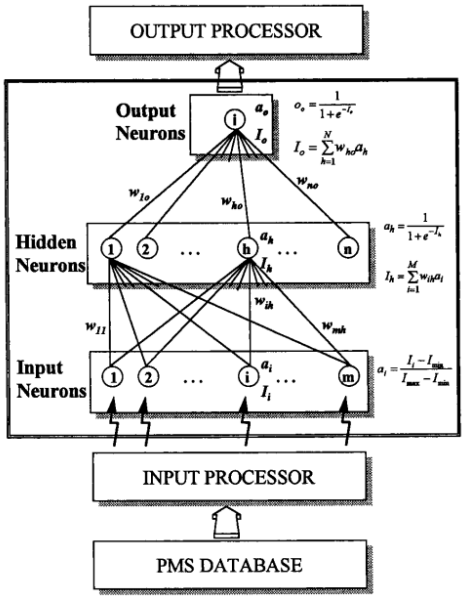


Fig. 1. Developed artificial neural network [10]

They proposed methodology which included the application of a hybrid technique which combines the gene expression programming (GEP) and artificial neural network (ANN). The developed algorithm showed good results for prediction of IRI using traffic parameters and structural properties of pavement.

## RESEARCH METHODOLOGY

This research assembles and evaluates information regarding present practices used for assessing crack and making repairs. Primary emphasis is placed on crack assessment methods of PCBCP's by DOTs and various federal and state agencies involved in pile substructure restoration projects. The severity of crack and damage states including field testing of repair methods to ascertain the effectiveness of applicability and aid guidelines for inspection, crack assessment, and selection of repair methods will be developed and presented.

To accomplish these, this study implemented the following tasks (see Figure 1):

**1. Explore historical data and identify possible data quality problems: Cleaning, filtering, and extracting dataset:** Many transportation agencies rely on robust models to evaluate pavement performance and improve pavement asset management. Accurate pavement performance models that include adequate pavement data are needed as the basis for future pavement maintenance and rehabilitation strategies. Data are the main building blocks in performance modeling, so obtaining good quality data is essential to getting accurate results. Pavement distress data including section identification, construction history, pavement type, maintenance history, traffic loading, structure parameters, and pavement distress will be obtained from SHA or DOT Pavement Management Information System (PMIS). While these pavement datasets can facilitate training more sophisticated ML models, systematic data errors can make model training unreliable and lead to incorrect decisions. In this project, data cleaning will be performed in two stages: error detection and error repairing. Error detection would identify common errors including duplicates, missing values, integrity constraints violations, typos, mixed formats, and replicated entries. All inconsistencies, redundant and irrelevant data will be removed while noisy data will be smoothed out. To understand correlations and patterns in existing data, the data will be cleaned and trained.

**2. Develop the machine learning algorithms:** The architecture of an ANN model is developed and the datasets randomly divided into training and validation datasets to train the neural networks. The training data set will be used to develop the model, while the validation data will be used to assess the accuracy of the developed ANN model and avoid overfitting in the model. A supervised learning process utilizing historical data both for network inputs and desired outputs will be implemented. Factors affecting pavement surface roughness will be used as input variables to develop a model to predict the output variable, roughness. The predicted output is compared to measured values to calculate an error used to adjust the connection weights between the model inputs and outputs. Back-propagation algorithm is then used during the training process to minimize error until the model produces a lower error value. Several preprocessing steps will be taken for smoothing, outlier detection, normalization (mean removal), and de-correlation. As the pavement deteriorates, roughness typically increases with time. Data records where the IRI measurement is less than its previous IRI reading (while no testing or rehabilitation treatment had happened) will be considered outliers and will be ignored in the database created. This allows the creation of network map of pavement areas that needs testing and rehabilitation.

**3. Preparation of the artificial neural network database:** Before the application of the developed ANN model, data will be collected and sorted on the network to conduct learning and training. The database for entry in the neural

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

network will comprise of a set of input data and output data. The input data are the measured values of the technical parameters (IRI, rut depth, texture depth, surface cracks and patches) shown as their mean value for an individual pavement segment. The results calculated global performance index (GPI) and the selected maintenance strategy represent the output results of this database. The intention of this database is the preparation and sorting of data in a format suitable for entering it into the neural network.

**4. Validation of prediction models and and integration into existing work processes:** The process of developing the neural network of a validation dataset is similar to the training process, except no weight matrices will be produced from the validation process. The model will be run with the architecture in Fig. 2 ensuring similar error values for training and validation datasets are obtained. The architecture simulation will be repeated for each uniform pavement section in the database to prepare the recommended list of sections for the testing and rehabilitation program. The project recommendation procedure will be incorporated into existing work processes of SHA to assist in the rehabilitation program processes.

**Implementation Plan for the Design and Implementation of Artificial Neural Network**

**1. Selection of Artificial Neural Network Software**

The first step was to review available artificial neural network software to select an artificial neural network simulator. A backpropagation simulator, Brainmaker, which works in both DOS and Windows environments and has very good utilities for data preparation and manipulation will be selected.

**2. Data Preparation and Selection of Input and Output Parameters:** Historical data from MD-Datasets from SHA pavement rehabilitation programs will be used to prepare the examples for training and testing the neural network using the following steps:

Step 1: Create a database with all of the historical data of projects between 2010 and 2018. Once the database is created, the average pavement condition and traffic at the time that the road sections are programmed will be compiled for each road section in the database. In addition, cracking and roughness roughness 1, 2, and 3 years before the programming year (the year when the sections were selected) and the average maintenance cost for the last 3 years will be computed.

Step 2: Selection of outlying road sections not included in the rehabilitation program. This is important because the artificial neural network needs to learn to differentiate sections that require rehabilitation from those that did not. All of the information needed will be obtained for all mileposts that do not have a rehabilitation programmed or constructed for the 3 years following the programming year. Homogeneous sections will be delimited by using past and programmed project information, and the averages for all of the parameters considered will be computed and recorded for all sections longer than 5 miles. The neural network will then be trained ensuring the number of nonprogrammed sections are similar to the number of programmed sections.

Step 3: Ensuring good performance of network. Seventy-five percent of these datasets or examples will be used for training and 25 percent will be reserved for testing. A higher percentage of test cases is expected in this project which would provide a good estimate of the capability of the network to generalize. This is also important for ensuring a good performance of the network in the future.

**3. Selection of Most Appropriate Network Architecture and Training Parameters.** The design of the artificial neural network architecture, training, and testing will be conducted concurrently. A fractional factorial experiment will be designed for this purpose. The objectives of the experiment is to identify the neural network design factors that significantly affect the network performance and the levels at which these factors should be used. The following factors will be selected for the experiment: Learning rate, Training tolerance, Stratification of the input data, and number of neurons in the hidden layer. Each experimental run would consist of a separate training section that will start with a complete randomized set of a connection's weights. The learning rate is a factor used to scale the corrections done to the weights while the network is learning. Since the learning rate is necessary to improve the speed of convergence of the network, it could affect network performance. The training tolerance is the maximum deviation from the actual output accepted to consider a network output correct. Since most neural networks respond better to ranges in inputs and outputs than to precise numeric value, two separate sets of training and testing examples will be prepared. In the first set the input variables are the numeric values of each field as they are computed (e.g., 5 percent cracking). In the second set the input are coded or stratified by classifying each input item in up to a maximum of five ranges or categories (low, medium or average, high, very high, and extremely high). These categories were represented by an integer from 21 to 13. All missing values were assigned a 0. The number of categories used for each item are indicated in Table 1. For example, a pavement section of an Interstate highway with cracking lower than 1 percent was assigned a -1, one with cracking of between 1 and 5 percent was assigned a 0, one with cracking between 5 and 12 percent was assigned a +1, one with cracking of between 12 and 25 percent was assigned a +2, and one with more than 25 percent cracking was assigned a +3. A Yes in Table 2 indicates that the input was stratified, and a No indicates that the actual values (actual percentage of cracking) were used for each item.

The number of hidden neurons or neurons in the hidden layer (only one hidden layer was used) normally affects the performance of a neural network. Brainmaker user's guide recommendation will be followed in the selection of neurons [14].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

Table 1: Selected Levels for Designed Fractional Experiment

| Training Session | Learning Rate | Training Tolerance | Stratified Input | Hidden Neurons |
|---|---|---|---|---|
| 1 | 0.2 | 0.1 | Yes | 12 |
| 2 | 1 | 0.4 | Yes | 12 |
| 3 | 0.2 | 0.4 | No | 12 |
| 4 | 1 | 0.1 | No | 12 |
| 5 | 0.2 | 0.4 | Yes | 48 |
| 6 | 1 | 0.1 | Yes | 48 |
| 7 | 0.2 | 0.1 | No | 48 |
| 8 | 1 | 0.4 | No | 48 |

## 4. REFINEMENT AND IMPLEMENTATION OF ARTIFICIAL NEURAL NETWORK

Several trial runs will be made to determine the number of hidden neurons that uses a stratified input with a slightly modified stratification scheme, training tolerance and learning rate that would achieve the best performance. The simulation of the artificial neural network would be implemented through a computer program that basically follows the scheme presented in Fig. 1. For each pavement section the input processor would computes the input values $I_i$ for each field and passes them to the artificial neural network simulator. The simulator is expedted to use the defined neural network architecture and computes a prog number for each section in the network. Fig. 2 provides the scheme for the artificial network simulator. The simulation starts computing the activation level of each input neuron i. As shown there is one input neuron for each pavement characteristic to be considered. The corresponding

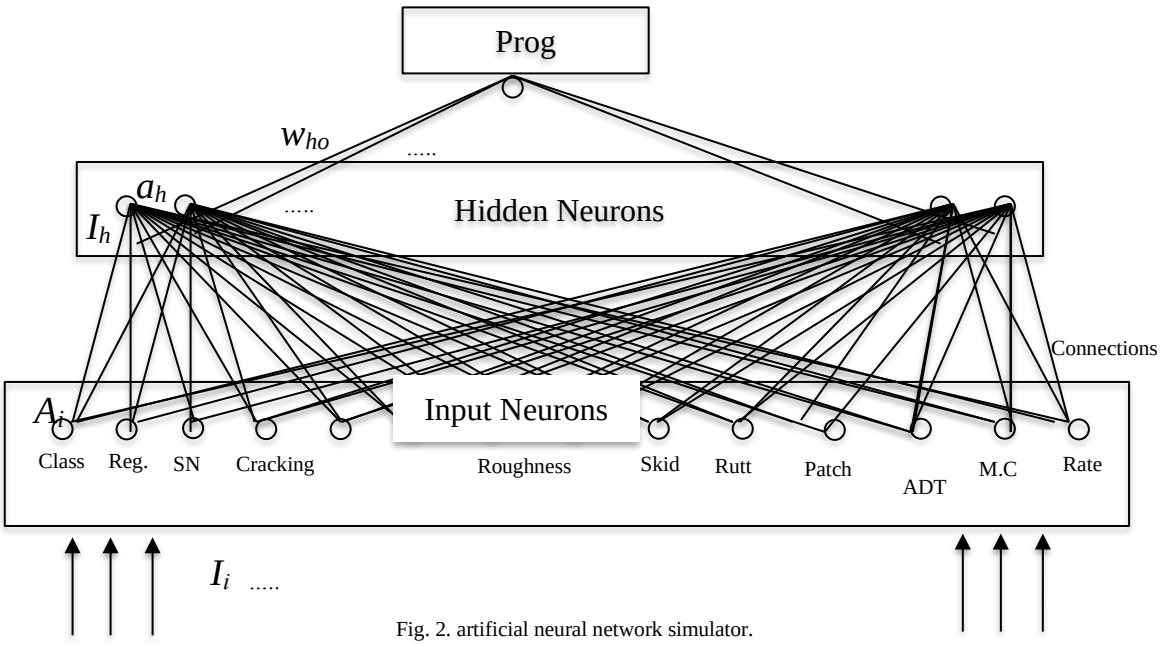


Fig. 2. artificial neural network simulator.

input value $I_i$ will be processed to a scale of 0 to 1 by using the minimum and maximum values of all pavement sections for that field. Finally, the simulator would compute the input of the output neuron ($I_o$) and its level of activation $O_O$. That is, the prog value that is transmitted to the output processor. The simulation is repeated for each uniform pavement section in the database to prepare the recommended list of untested sections for rehabilitation program. This recommended procedure of the developed ANN will then be applied to existing work processes of SHA.

## REFERENCES

[1] Attoh-Okine, N. O. (1994). "Predicting Roughness Progression in Flexible Pavements Using Artificial Neural Networks," *Proceedings of the 3rd International Conference On Managing Pavements*, Vol. 1, San Antonio, Texas, pp. 55-62.

[2] Eldin, N. N. & Senouci, A. B. (1995). A Pavement Condition-Rating Model Using Backpropagation Neural Networks. *Microcomputers in Civil Engineering*, 10, 433-441. https://doi.org/10.1111/j.1467-8667.1995.tb00303.

[3] Haas, R., Hudson, W. R., and Zaniewski, J. (1994). *Modern Pavement Management*, Krieger Publishing Company, Malabar, Florida.

[4] Huang, Y. & Moore, R. K. (1997). Roughness level probability prediction using artificial neural networks. *Transport Research Record 1592, Paper No. 970419*, 89-97. https://doi.org/10.3141/1592-11.

[5] Lin, J. D., Yau, J. T., & Hsiao, L. H. (2003). Correlation analysis between international roughness index (IRI) and pavement distress by neural network. *82nd Annual Meeting of the Transportation Research Board*, Washington, DC, USA, Retrieved from https://www.researchgate.net/ profile/Jyh-Dong_Lin/publication/228848218_Correlation_ nalysis_between_international_roughness_index_IRI_and_pavement_distress_by_neural_network/links/02e7e52f385af7c205000000/Correlation-analysis-between-international-roughness-index-IRI-and-pavement-distress-by-neural-network.pdf.

[6] Mazari, M. & Rodriguez, D. D. (2016). Prediction of pavement roughness using a hybrid gene expression programming-neural network technique. *Journal of Traffic and Transportation Engineering*, *3*(5), 448-455. https://doi.org/10.1016/j.jtte.2016.09.007.

[7] Perera, R. W. and Kohn, S. D. (2001). *LTPP Data Analysis: Factors Affecting Pavement Smoothness.* NCHRP Web Document

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 9 Issue 05, May-2020

40. National Cooperative Highway Research Program, Transportation Research Board.

[8] Thube, D. T. (2011). Artificial Neural Network (ANN) Based Pavement Deterioration Models for Low Volume Roads in India. *International Journal of Pavement Research and Technology*, *5*(2), 115-120. Retrieved from http://www.ijprt.org.tw/reader/pdf.php?id=223.

[9] Von Quintus, H. L., Eltahan, A., and Yau, A. (2001). "Smoothness Models for Hot-Mix Asphalt-Surfaced Pavements; Developed from Long-Term Pavement Performance Data," *Transportation Research Record*, No. 1764, pp. 139-156.

[10] Yang, J., Lu, J. J., Gunaratne, M., & Xiang, Q. (2003). Forecasting overall pavement condition with neural networks – an application on Florida Highway Network. *Transportation Research Record 1853, Paper No. 03-3441*, 3-12. https://doi.org/10.3141/1853-01.

[11] Brainmaker, Neural Network Simulation Software, User's Guide and Reference Manual. California Scientific Software, Nevada City, 1993.