International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:1, 2016

The Design of a Vehicle Traffic Flow Prediction Model for a Gauteng Freeway Based on an Ensemble of Multi-Layer Perceptron

Tebogo Emma Makaba, Barnabas Ndlovu Gatsheni

***Abstract*—**The cities of Johannesburg and Pretoria both located in the Gauteng province are separated by a distance of 58 km. The traffic queues on the Ben Schoeman freeway which connects these two cities can stretch for almost 1.5 km. Vehicle traffic congestion impacts negatively on the business and the commuter's quality of life. The goal of this paper is to identify variables that influence the flow of traffic and to design a vehicle traffic prediction model, which will predict the traffic flow pattern in advance. The model will unable motorist to be able to make appropriate travel decisions ahead of time. The data used was collected by Mikro's Traffic Monitoring (MTM). Multi-Layer perceptron (MLP) was used individually to construct the model and the MLP was also combined with Bagging ensemble method to training the data. The cross—validation method was used for evaluating the models. The results obtained from the techniques were compared using predictive and prediction costs. The cost was computed using combination of the loss matrix and the confusion matrix. The predicted models designed shows that the status of the traffic flow on the freeway can be predicted using the following parameters travel time, average speed, traffic volume and day of month. The implications of this work is that commuters will be able to spend less time travelling on the route and spend time with their families. The logistics industry will save more than twice what they are currently spending.

***Keywords*—**Bagging ensemble methods, confusion matrix, multi-layer perceptron, vehicle traffic flow.

I. Introduction

THE Gauteng province of South Africa (SA) is currently experiencing traffic congestion, especially during peak hours of (06:00hrs-09:00hrs) and (15:00hrs-18:00hrs). The cities of Johannesburg and Pretoria are both located in the Gauteng province. The traffic queues on the Ben Schoeman freeway, which connects these two cities, can stretch for almost 1.5 km. The traffic on this freeway has been growing at a rate of approximately 7% per year for the past ten years [1]. Traffic congestion on this freeway is currently estimated to cost the Gauteng economy over 30 million US dollars per year due to time lost, higher transport costs and higher delivery costs of goods among other factors. Traffic congestion also has a negative impact on air quality due to emissions from vehicles, as well as on the quality of life [1]. When there are vehicle accidents the total time for the journey increases by over 150% of the free flowing traffic travel time on average. Various economic studies indicated the negative effect of traffic congestion on the Gauteng economy and standard of living [2].

The main benefits of this research are free flowing vehicle traffic and the reduction in travel time. Other benefits of the study are the improvement in environmental conditions resulting from the reduction in exhaust gases (which is not addressed in this paper).

In 2005 to 2010, the department of road and transport implemented the Gauteng Freeway Improvements Project (GFIP).This involved upgrading and expanding the provinces freeway network with the goal of reducing vehicle traffic congestion. The expansion of freeways is not a sustainable solution. The South African National Roads Agency Limited (SANRAL) has installed electronic tollgates (e-tolls) in all the freeways in Gauteng. This is a billing system which bills road users as they use the freeway. The system allows users to register to qualify for a discount. In 2010, a rapid commuter train link called the Gautrain was introduced in Gauteng. This rail spans 80 km, and it links the cities of Johannesburg, Pretoria and Johannesburg is OR Tambo International Airport. This rapid railway covers the distance between Johannesburg and Pretoria in 30 minutes compared to the 1hour or 2hours when using a car. In 2009, the bus rapid transit (BRT) system also called Rea Vaya was implemented in Johannesburg and in Pretoria BRT is known as Are Yeng was rolled out in 2015. These public transport interventions allow the public to travel quickly around the cities since there are dedicated bus lanes. In this paper the traffic congestion problem has been solved using a model derived from historical vehicle traffic data. Variables that are known to cause traffic congestion include road intersections, traffic volumes, pedestrian traffic signals, connecting roads [3] and road infrastructure. Thianniwet et al. [4] used the Decision Tree algorithm (J48) and sliding windows to predict traffic congestion. Data was collected from Thailand cities using a GPS device, webcam and opinion survey from the road users. Main parameters used were time, speed, volume, service level and the cycles of traffic signal, that motorist had to wait on the traffic queue. The study focused on the vehicle speed covering the greater traffic ranges. The evaluations revealed that J48 model achieved an overall accuracy of 91.29%. The approach in this case is weak as vehicle owners might deny permission for data to be collected from their devices. In our case the data was collected in Gauteng (GP) by the company called Mikro's Traffic Monitoring (MTM).

Tebogo Makaba is with the Department of Applied and Information System, University of Johannesburg, Auckland Park, Gauteng Johannesburg (e-mail: te.makaba@gmail.com/bgatsheni@uj.ac.za).

He et al. [5] used the ensemble-based methods to predict traffic jam. This method was used to fuse information from several base predictors in order to come up with better-combined predictors. The attributes used were speed, number of cars and the time when the traffic jam first occurred. Inputs of the simulation were given 5 road segments with roadwork as a sequence of major roads where the first jams occurred during the initial 20 minutes of the simulation. The goal of the study was to predict a sequence of major road where the next jam will occur in the next 40 minutes. To evaluate the model cross-validation was used and performance on the final test set proved the effectiveness of the methods used. This study did not include evening peak hours data and they did not do cost calculation to be able to determine the model that best performed.

Mao et al. [6] used multilayer perceptron (MLP) optimised using the genetic algorithm to predict dynamic vehicle traffic flow. None of these computed the cost of prediction. Yang et al. [7] used Kalman Filtering and an Estimation Technique to predict traffic volume using Global Positioning System (GPS) test vehicle technique. Traffic volume was predicted using collected real time and historical data. The traffic volume was predicted using a graduation ceremony event as a case study to carry the study. The Mean Absolute Relative Error (MARE) was used to calculate different values for the real time data and the historical data, collected using GPS. The approach is weak as their data was collected using only one graduation ceremony event, which may not provide accurate traffic data that is required to produce good results.

Chen et al. [8] used an ensemble learning method namely, the bagging of radial basis function (RBF) for short-term traffic flow prediction. They used data that collected from freeway in Beijing. The data was collected from August 1 2003 to August 7 2003. The data was collected every 2 minutes from 00:00 am to 11:59 pm, which lead to 720 patterns every day and 5040 patterns in 7 days. Data traffic variables used was traffic flow, occupancy, and speed. Results showed that the bagging of RBF has a better prediction performance that one single RBF predictor. Thus, the ensemble learning method demonstrates great potential in improving the capability of unstable procedures like RBF. Their use of traffic data for 1-month period is weak as there may be other months, which may have high experience of traffic congestion beside the chosen month. Their approach did not show any calculation of the cost and cross- validation was not used to evaluate the model.

Pongpaibool et al. [9] used fuzzy logic and adaptive neuro-fuzzy to predict traffic congestion, vehicle detection (traffic videos) and tracking software to collect traffic information and evaluate levels of road traffic congestion. The model predicted the traffic congestion, depending on how the rules were defined, time-of-day and day-of-week variations and determined the accuracy and Average Deviation. MATLAB tools 'Fuzzy Logic Control' and 'ANFIS' were used to implement the fuzzy logic and the adaptive neuro-fuzzy. Results showed that manually tuned fuzzy logic achieved 88.79% accuracy, while the adaptive neuro-fuzzy technique achieved only 75.43% accuracy. In this approach, they may have limited the number of parameters that affects traffic congestion.

A model that will give commuters the ability to know the state of the traffic condition ahead of time has been constructed. This prediction model will assist in reducing vehicle traffic congestion and it will be beneficial to commuters, the economy of Gauteng, and that of South Africa traffic volume, accidents, average speed, and time of the day [10].

II. Methods

A. Multi-Layer Perceptron (MLP)

The MLP is an information-processing paradigm that is inspired by the way biological nervous system processes information. The MLP is composed of a large number of highly interconnected processing elements (neurons) organised in layers. MLP is applicable to non-linearly separable data [11].

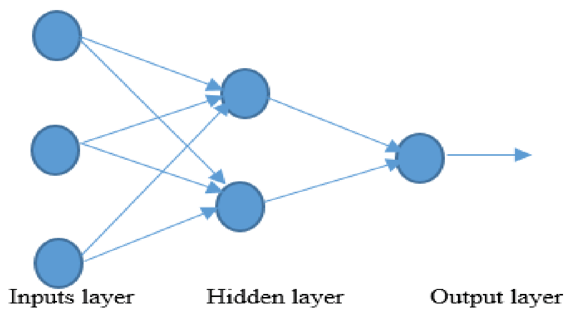

Fig. 1 A Multi-layer perceptron with one hidden layer

The layer in the MLP consists of the input layer, the hidden layer (s) and the output layer as shown in Fig. 1. The hidden layer does intermediate computation before directing the input to the output layer. The MLP takes a different approach to solving problems than that of traditional computers. The latter use the algorithm approach, meaning that a computer follows a set of instructions in order to accomplish a task.

The MLP approach does the following [12]:

- Self-organization: The MLP can create its own organization or representation of the information it receives during the learning time.
- Linear and Non-linear relationship: The power and advantage of the MLP lies in its ability to represent both linear and non-linear relationships and in its ability to learn these relationships directly.
- Evidential response: In the context of pattern classification, neural network can be designed to provide information not only about which particular pattern to select but also about the confidence in the decision made.

The advantages of the MLP are as follows:

- Adaptive learning: An ability to learn how to do tasks based on the data given for training which can also be called the initial experience.

- It is one of the preferred techniques for gesture recognition.
- MLP does not make any assumption regarding the underlying probability density functions or other probabilistic information about the pattern classes under consideration [13].
- It yields the required decision function directly via training.
- A two-layer back propagation network with sufficient hidden nodes has been proven a universal approximate [14].
- The disadvantage of Multi-layer Perceptron is as follows:
- The MLP network finds out how to solve the problem by itself, therefore its operation can be unpredictable.

MLP has been applied in areas such as Text to Phoneme Mapping [15], breast cancer cell analysis and market analysis [16], Speech recognition [17] and manufacturing process control.

B. Bagging Ensemble Method

The idea of ensemble learning methods is to select a whole collection (ensemble) of hypotheses (weak or based learners) from the hypotheses space and combine or aggregate their predictions [18] into a single learning model. Weak or base learners combined through a voting or averaging process. The ensembles have been shown to be accurate in many cases than the individual predictors [19], but it is not always meaningful to combine models. The ensemble methods competes with data fusion [20] which combines data from multiple predictors, and related information from associated databases, in order to achieve improved accuracy and to make better inferences than could be achieved by the use of a single model or data set alone. In this paper, the bagging ensemble method is used together with Multi-layer perceptron (MLP). Bagging, which stands for bootstrap aggregating, is a method for generating diverse ensembles for model combination [19]. Bootstrap aggregating is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical prediction and regression. It also reduces variance and helps to avoid over fitting. Bagging is an intuitive algorithm, with a good performance [21]. The method works by reducing variance by voting and averaging.

Bagging has the following advantages:

- Improved accuracy.
- Solve problem of unstable base classifiers/predictors by reducing the errors associated with random fluctuations in the training data.
- It reduces the error due to variance of the base classifier/predictors.
- It is noise-tolerant, but not so accurate [22].

Bagging has the following disadvantage:

- Is not a simple to interpret.

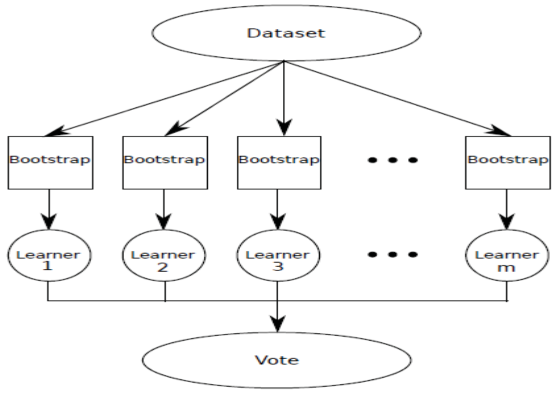

Fig. 2 A Bagging ensemble method diagram

C. Cross-Validation (Cv)

Cross-validation is a model validation technique for assessing how a model will generalize to an independent data set. One can decide on a fixed number (n) of folds for the data. The data is split into n folds. In each turn, one split fold is used as test and the remaining folds for training. In this study 10-folds cross-validation were used. Thus 9-folds of the data were used for training 1-fold of the data was used for testing (re-evaluating models). An average test error rate is computed from ten trials and this is the estimated error rate for the model.

The procedure for 10 folds Cross-validation works [19]:

- Step 1: Break data into 10 sets of size n/10.
- Step 2: Train on 9 datasets and test on 1.
- Step 3: On the next iterations ensure each of the 10 folds have been used as a test set.
- Step 4: Repeat 10 times and take a mean accuracy.

This approach it is recommended when there is insufficient data.

D. Root Mean Square Error (RMSE)

RMSE is derived by squaring the differences between known (locations) and unknown (interpolated location) points, adding those together, dividing that by the number of test points, and then taking the square root of the results as shown in (1) [24]. The RMSE of a prediction model with respect to the estimated variable X_{model} is defined as the square root of the mean squared error as shown in (1):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_{obs,i} - X_{model,i})^2}{n}} \quad (1)$$

where X_{obs} are observed values and X_{model} are modelled values at time or place i.

The RMSE values can be used to distinguish model performance in a calibration period with that of a validation period as well as to compare the individual model performance to that of other predictive models. A high RMSE indicates a poor result and a low RMSE indicates a good result.

III. Data Collection

The data used for the study was collected from Gauteng freeways by Mikro's Traffic Monitoring (MTM). This vehicle traffic flow data was for the freeway that links Johannesburg with Pretoria (M1 North extending to the N1 North) also called Ben-Schoeman freeway. MTM is contracted by the Department of Transport (DOT) to collect vehicle traffic flow data.

MTM has data loggers installed at different location sites on the freeway. The data loggers count the number of cars, the average speed, and the volume of vehicles on the freeway. The data was then obtained from MTM in this form. The data loggers are connected wirelessly to the MTM server where data is stored. TelDialer is a device that is mounted on the data loggers on site which can communicate with the MTM server to store data in real-time. TelWin is a software program that is used by the technical team of MTM on site to check if the data loggers are functioning properly. This device can record and play back incidents on the freeway. It is used by both the technical team and the data analyst. The main function of the program is to communicate with the logger. Traffic is also viewed in real-time from this program. MTM uses MonCam cameras for quality assurance of the data collection process. They provide frame-grab images for all recorded vehicles and a synchronized video stream with the recorded traffic data for data analysis and processing. TrafBase application software was used. The TrafBase product is designed to validate, store and manage large amounts of traffic data collected by the traffic logging equipment. Traffic data is made available to a user either in its original or in summarized form through data files, spreadsheet files, and physical reports. The data was received in a Microsoft Excel Spreadsheet format from MTM.

A. Data Pre-Processing

The data used for the study was collected from Mikro's Traffic Monitoring (MTM). The data, when collected, was in numeric format as shown in Table I, which shows only 5 of the 5533 instances of the vehicle traffic flow data. The number of instances used to build models was 5533. The data was received in an excel spreadsheet and it contained numeric values which represented values of attributes that influence traffic flow such as TravelTime, AverageSpeed and TrafficVolume as shown in Table I. In addition to the attributes that came with the data, another new one called "day of month" was created. For the data, public holidays and weekends were left out in this paper since the traffic pattern during this time is unstable on the freeways.

TABLE I
ONLY 5 INSTANCES OF THE 5533 DATA BEFORE PRE-PROCESSING

Instance	Date	Time	Total Road	Average Speed (Dir1)
1	2013-01-28	14:00:00	4169	100
2	2013-01-28	15:00:00	6719	100
3	2013-01-28	16:00:00	7660	100
4	2013-01-28	17:00:00	8682	93
5	2013-01-28	18:00:00	7793	97

The data in Table II is the data from Table I that has been converted to nominal values. The vehicle traffic flow data was categorized into three targets namely. Freeflow, meaning that vehicles are travelling at the required speed that is greater than 100km, Flowing Congestion meaning that vehicles enter the congestion state at the speed between (90 km– 100 km) and Congested meaning that vehicles are travelling at a speed less than or equal to 90 km as shown in Table I.

TABLE II
5 INSTANCES FROM TABLE I THAT HAVE BEEN CONVERTED TO NOMINAL VALUES

Instances	Attributes				Predictions
	Day of Month	Travel Time	Traffic Volume	Average Speed	Target Concept
1	DM-Jan	Off-Peak	Average-Traffic	Average-Speed	Free flow
2	DM-Jan	Peak	Heavy-Traffic	Average-Speed	Flowing Congestion
3	DM-Jan	Peak	Heavy-Traffic	Average-Speed	Flowing Congestion
4	DM-Jan	Peak	Heavy-Traffic	Average-Speed	Flowing Congestion
5	DM-Jan	Peak	Average-Traffic	Average-Speed	Flowing Congestion

TABLE III
THE DISTRIBUTION OF INSTANCES FOR THE TESTING AND TRAINING DATA CATEGORIZED AND THEIR TARGET CONCEPT

Instances	Target Concepts			Total
	Congested	Flowing Congestion	Free flow	
Training dataset	210	1033	2446	3689
Testing dataset	108	482	1254	1844
Total Instances				**5533**

Table III contains summary of the training and testing data organised with target concept. The testing dataset will be used to re-evaluate all designed models during experiments.

IV. Experimentation And Results

In this part of the section experiments and results of this paper will be provided.

A. Experiments

The methods used in this section are supervised learning methods. What makes it supervised is that the algorithm is given training examples where the values of the target variable are provided. This algorithm may learn which values of target variables are associated with which values of the predictor variable. The used data was divided into two-thirds (2/3) which was used for training and one-third (1/3) of the data was used to test the model. Open source WEKA software was used to carry out the experiments. The algorithms used for this paper are Multi-layer perceptron (MLP) individually and MLP used together with the Bagging ensemble method. The confusion matrix and the loss matrix were used to calculate the cost of prediction for each model. The model with the lowest cost was chosen as the best model. The rmse was used to determine good results.

Experiment 1: Model Designed Using the Multi-Layer Perceptron Algorithm

The procedure for constructing a model based on the MLP is as follows:

- Define the input parameter.
- Select the number of hidden layers and nodes within the hidden layer.
- Determine the learning rate and momentum.
- Determine the number of Epochs.
- Set the validation threshold.

The strategy used for constructing the prediction model using MLP: 5533 Instances were used and the evaluation method used was cross-validation.

Data was presented to the MLP for training.

- If results were good the model was saved.
- The results were analysed looking at the RMSE, Attributes involved and trained success rate.

Experiment 2: Model Designed Using Bagging With The MLP Algorithm Together.

The procedure for constructing the bagging method is as follows [14]:

- Phase1: Predictor generation

Step 1. Create t- (iteration) data sets from a database by applying the sampling with replacement strategy. This will form the training dataset.

Step 2. Apply a learning algorithm to each sample training data set to create models.

- Phase 2: Prediction

Step 3. For an object with unknown decision, make predictions with each of the t-models predictors.

Step 4. Select the most frequently predicted decision.

The strategy used for constructing the prediction model using Bagging with MLP:

- 5533 instances were used and the evaluation methods used was cross-validation.
- Bagging was used as a 'meta' data and the MLP was set as the predictor under Bagging properties box.
- The data was presented to Bagging as 'meta' and the MLP set as a predictor for training.
- If the results were good the model was saved.
- The results were analysed looking at the RMSE.

Experiment 1 and 2 procedures were used when constructing the vehicle traffic model using WEKA. The model was used to build the traffic prediction models. The model were saved and used to predict/classify the test data set. The results of the training and test data are shown in Table IV.

TABLE IV
THE RESULTS FOR THE MLP WHEN USED INDIVIDUALLY AND WHEN THE MLP IS USED TOGETHER WITH BAGGING ENSEMBLE METHOD

#	Learning Algorithm	Attributes	Correct Predicted instances training (%)	RMSE Training	Predicted Instance Testing (%)	RMSE Testing
1	**MLP**	TT, TV, DOM and AS	99.973	0.0137	100	0.0012
2	**Bagging with MLP**	TT, TV, DOM and AS	99.973	0.0136	100	0.002

B. Results

The results in Table IV when using MLP individually and when combining the MLP with Bagging, the results are still the same as 99.973%. This shows that MLP performed best when used individually and that bagging ensemble method does not improve the results. The attributes that were used were TravelTime (TT), TrafficVolume (TV), AverageSpeed (AS) and DayOfMonth (DOM).

C. Post Processing

The models developed in the previous section, were used for predicting the state of vehicle traffic flow on the freeway. The desire was to minimise the probability of getting a wrong prediction. There might be serious consequences if the model predicts traffic as flowing freely yet there is in fact traffic congestion. Predictions that are costly to the commuter had to be penalised. The MLP model and the Bagging (MLP) were evaluated using a loss matrix with elements L_{kj} which specify the penalty (quantifying risks) associated with predicting a novel instance to be class C_j when in fact it is class C_k. Thus for all instances $\mathbf{x}$ which belong to C_k, the expected loss for those instances is given by (3) [23].

$$R_k = \sum_{j=1}^{c} L_{kj} \int_{R_j} P(\mathbf{x} \mid C_k) d\mathbf{x} \tag{2}$$

The overall expected loss or risk for patterns from all classes is given by (2).

$$R = \sum_{k=1}^{c} L_{kj} P(\mathbf{x} \mid C_k) \tag{3}$$

The risk is minimised if the integrand is minimised at each point **x, i.e.** if the regions R_j are chosen such that $\mathbf{x} \in R_j$

In assigning the loss matrix in Table V a loss of 1 or more was assigned if an instance was placed in a wrong class and a loss of zero if the instance was placed in a correct class. The values of the coefficients of L_{kj} (values for cells in Table V) were chosen by hand based on the views of experienced MTM staff.

Since there was a 3 x 3 confusion matrix from three classes, a 3 x 3 loss matrix had to be designed. This loss matrix in combination with a confusion matrix was used for selecting the best vehicle traffic congestion prediction model.

The procedure for computing the cost of prediction is as:

- Compute the confusion matrix.
- Get a suitable loss matrix of the same dimension as the confusion matrix.
- Compute the cost by multiplying values in corresponding cells of these matrices and then perform a linear combination of the results as shown in (4).
- If the result has a high RMSE value and a high cost and reject the predictor model, otherwise accept it.

TABLE V
A LOSS MATRIX FOR COMPUTING THE COST OF VEHICLE TRAFFIC CONGESTION PREDICTION, A = FREE FLOW, B = FLOWING CONGESTION AND C = CONGESTED

		Predicted		
		A	**B**	**C**
Actual	**A**	0	2	3
	B	4	0	1
	C	4	1	0

The cost for prediction for Table V is as:

$$C_{\cos t\_MLP} = \sum Conf_{kj} \mathbf{x} L_{kj} \tag{4}$$

The costs for Tables VI and VII were computed using the procedure for cost prediction. Thus, their cost is only shown with the confusion matrices and in the summary of results in Table VIII.

TABLE VI
THE CONFUSION MATRIX FOR THE VEHICLE TRAFFIC CONGESTION BASED ON THE MLP

		Predicted		
		a	**b**	**c**
Actual	**a**	2445	0	1
	b	0	1033	0
	c	0	0	210

Total results are obtained by multiplying the values of the Loss Matrix in Table V with the corresponding cell values of the Confusion Matrix in Table VI, Total Cost = 3

TABLE VII
THE CONFUSION MATRIX OF BAGGING METHOD FOR VEHICLE TRAFFIC CONGESTION PREDICTION MODEL USING MLP

		Predicted		
		a	**b**	**c**
Actual	**a**	2445	0	1
	b	0	1033	0
	c	0	0	210

Total results are obtained by multiplying the values of the Loss Matrix in Table V with the corresponding cell values of the Confusion Matrix in Table VII.

Total Cost = 3

The model with the lowest cost is the best performing model. Looking at Tables VI and VII where the MLP was used individually, the results show that the MLP has the lowest cost of 3. Bagging when used with MLP was also the best performing model with the lowest cost of 3. This can be concluded that MLP is the best performing model.

TABLE VIII
THE SUMMARY OF PREDICTION PERFORMANCE SUCCESS, RMSE AND THE TOTAL COST FOR MLP AND BAGGING WHEN USED TOGETHER WITH THE MLP ALGORITHM

	Attributes	**Prediction Performance (%)**	**RMSE**	**Total Cost**
MLP	TT,TV,AS and DOM	99.973	0.0137	3
Bagging (MLP as a classifier)	TT,TV,AS and DOM	99.973	0.0136	3

Because bagging has a low RMSE value, it becomes the best model.

V. DISCUSSION

The results in Table IV show that the Multi-Layer perceptron (MLP) predicts vehicle traffic flow with the prediction performance of 99.973%. The MLP algorithm when combined with bagging ensemble has the best performance as it has the lowest RMSE value.

The results are consistent with what is experienced on the freeways. Thianniwet et al. [4] used the Decision Tree algorithm (J48) and sliding windows to predict traffic congestion. Data was collected from Thailand cities using a GPS device, a webcam, and an opinion survey from the road users. Main parameters used were time, speed, volume, service level and the cycles of traffic signals. The study focused on the vehicle speed. The J48 model achieved an overall accuracy 91.29%. The study used by Thianniwet et al, used parameters that are more similar as the current paper.

These results also compares well with those by [6] that used the MLP optimised using the genetic algorithm to predict dynamic vehicle traffic flow. Yang et al. [7] used Kalman Filtering and an Estimation Technique to predict traffic volume. The use of the Global Positioning System (GPS) test vehicle technique. Traffic volume was predicted using collected real-time and historical data. The traffic volume predictors were tested using a graduation ceremony event as a case study. The Mean Absolute Relative Error (MARE) was used to calculate different values for the real-time data and the historical data, which were collected using GPS. The results are better compared to ours as both historical and real-time data was used for this study. None of this computed the cost of prediction.

The implication of this MLP traffic prediction model is that commuters will be able to make travel decisions ahead of time. Data for weekends and public holiday was excluded for this work. In addition, commuters will be able to choose alternative routes to avoid getting stuck in traffic. Businesses will also see improvement in productivity due to on-time delivery of goods and services. Goods are likely become cheaper because of the decrease in delivery costs and thus improving the economic competitiveness of Gauteng.

VI. Conclusion

An intelligent vehicle traffic flow prediction model for Gauteng's freeway based on Ensemble methods has been designed. The results show that Multi-Layer perceptron model when combined with bagging ensemble method MLP used individually in predicting traffic flow on the freeways.

The Gauteng Department of Transport (DoT) and other agencies working with the department can look into using this prediction models to assist motorists to avoid being caught in a traffic jam.

References

[1] Gauteng Rail Link between Johannesburg, Pretoria and Johannesburg International Airport, Nov 01, 2014, pp. 1-12.

[2] A. Van Niekerk, Gauteng Freeway Improvement Project (GFIP) and Open Road Tolling. (PowerPoint Presentation). SANRAL (Mar 2013), pp.9 Accessed at: http://www.dbsa.org/en/prodserv/PACBP/MPIM%20Documents/Gauteng%20Freeway%20Improvement%20Project%20and%20Open%20Road%20Tolling%20by%20Alex%20van%20Niekerk%20-%20SANRAL.pdf.

[3] Y.L Dahee Hong, Development of Traffic Accident Prediction Models. Proceedings of the Eastern Asia Society for Transportation Studies, 5, 2011, pp.2046-51.

[4] T. Thianniwet and S. Phosaard, Classification of Road Traffic Congestion Levels from GPS Data using a decision Tree Algorithm and Sliding Windows, IEEE Conference on intelligent trans. Sys., Proceedings of the 12th International , 2009 , pp. 978-988.

[5] J. He, Q. He, G. Swirszcz, Y. Kamarianakis, R. Lawrence, W. Shen, and L. Wynter,"Ensemble-based Method for Task 2:Predicting Traffic Jam", Int. IEEE., Conf. on Data mining workshop New York 10598, 2010, pp.1363-1365.

[6] S. Mao, Y. Chen, and X. Xiao, Short-term Traffic Flow Prediction Based on GM (1, 1, Exp) model. International Conference on grey systems and intelligent services. (Sept. 2011), pp. 271-275.

[7] J. Yang , Travel Time Prediction Using the GPS Test Vehicle and Kalman Filtering Techniques., American control conference Portland, OR, USA, 8-10.Jun.2005, pp. 2128-2133.

[8] L. Chen and C.L. Phillip Chen, Ensemble Learning Approach for Freeway Short-Term Traffic Flow Prediction, IEEE, Department of Electrical and Computer engineering, The University of Texas at San Antonio, 2007, pp. 1-6.

[9] P. Pongpaibool, P. Tangamchit P. and Noodwong K., "Evaluation of traffic congestion using fuzzy techniques." IEEE Trans., University of Technology Thouburi, Thailand, 2007, pp. 1-4.

[10] Jungme Park, Z.C.L.K.M.L.K.M.A.M., 2009. Intelligent Vehicle Power Control Based on Machine Learning of Optimal Control Parameters and Prediction of Road Type and Traffic Congestion. Vehicular Technology, IEEE Transactions on (Volume: 58, Issue: 9), 2009,pp 4741 -- 4756

[11] P. Le and W. Zuidema, 2014, Perceptron and Multi-layer Perceptron, International Conference on Measuring Technology and Mechatronics Automation, 2014, pp. 1-8.

[12] S. Haykin, neural network: A comprehensive foundation. New Jersey: Macmillan, 1994, pp.4-5.

[13] G. Cybenko, "Approximation by super positions of a sigmoidal function." Math Contr. Signals Syst., vol. 2, 2003, pp. 303-314.

[14] M.C. Su, W.F. Jean, and H.T. Chang, " A Static Hand Gesture Recognition System Using a Composite Neural Network," in Fifth IEEE Int. Conf. on Fuzzy Systems, 1996, pp. 786- 792, New Orleans, U.S.A. (NSC85-2213-E-032-009).

[15] S. Beale and T. Jackson, Text to Phoneme Alignment and Mapping for Speech Technology: A Neural Networks Approach. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN 2011), Bullinaria, 2011, 625-632. IEEE.

[16] H.H. Makwana, D.J. Shah, and P.P. Ghandhi, FPGA Implementation of Artificial Neural Network. *International Journal of Emerging Technology and Advanced Engineering,* January 2013, 3(1):672-679.

[17] J. Park, F. Diehl, M.J.F. Gales and M. Tomalin, 2009. Training and adapting MLP features for Arabic speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing,* 19-24 April 2009:4461-4464.

[18] S. Russell and P. Norvig, 2003, Artificial Intelligence: A modern approach. 2nd ed, 2003.

[19] P. Tan, M. Steinbach and V. Kumar, Introduction to data mining: Pearson International Edition. 2006, pp 25-35.

[20] D.L. Hall, and J. Llinnas, "An introduction to multi-sensor data fusion," in Proc. Of IEEE, vol. 85, no. 1, 1997, pp. 6-23.

[21] I. Syarif, E. Zaluska, A. Prugel-Bennett and G. Wills, 2012, Application of Bagging, Boosting and Stacking to Intrusion Detection, 2012.

[22] R.P. Sweety, Ensemble of classifier using artificial neural networks as base classifier, Int Journal of Comp Sci and Mob App, vol.1 issue. 4 Oct -2013, pp. 7-16

[23] C.M. Bishop, 1995.Neural Networks for Pattern Recognition. Oxford University Press (Month 1995).

[24] T. Chai and R.R. Draxler, 2014, Root Mean square error (RMSE) or Mean absolute error (MAE), "Arguments against avoiding RMSE in literature", 2014, pp. 1-4.