DEEP LEARNING NEURAL NETWORK APPROACHES TO LAND USE-DEMOGRAPHIC- TEMPORAL BASED TRAFFIC PREDICTION

Abul Azad and Xin Wang

Department of Geomatics Engineering, University of Calgary,
Calgary, Alberta, Canada

ABSTRACT

Land use and transportation planning are inter-dependent, as well as being important factors in forecasting urban development. In recent years, predicting traffic based on land use, along with several other variables, has become a worthwhile area of study. In this paper, it is proposed that Deep Neural Network Regression (DNN-Regression) and Recurrent Neural Network (DNN-RNN) methods could be used to predict traffic. These methods used three key variables: land use, demographic and temporal data. The proposed methods were evaluated with other methods, using datasets collected from the City of Calgary, Canada. The proposed DNN-Regression focused on demographic and land use variables for traffic prediction. The study also predicted traffic temporally in the same geographical area by using DNN-RNN. The DNN-RNN used long short-term memory to predict traffic. Comparative experiments revealed that the proposed DNN-Regression and DNN-RNN models outperformed other methods.

KEYWORDS

Deep Neural Network (DNN), Deep Neural Network Regression (DNN-Regression), Deep Neural Network Recurrent Neural Network (DNN-RNN), traffic prediction

1. INTRODUCTION

Cities are multifaceted organizations consisting of complex, interconnected sub-systems that are influenced by a number of variables, such as how the land is developed, zoning regulations, population growth, infrastructure, and market fluctuations. The two most influential factors that impact how existing and future development will occur is centred around land use and transportation systems [1]. Predictive analysis of traffic is of fundamental importance for transportation management and planning. Knowledge of the traffic flow counts contribute to a transportation planning system that can result in optimal resource allocation.

Land use, demographic, and socioeconomic factors can impact the transportation planning processes [2]. The relationship between demographics, land use and transportation are the key components of planning. Several spatial and temporal approaches have been developed to predict traffic [3] [4]. There have been a number of traffic prediction methods that have attempted to understand traffic patterns, have significantly contributed to transportation planning. These traffic prediction methods allowed researchers to study the past and predict future trends in traffic by examining land use and traffic-related datasets.

The traffic prediction models found in the literature can be generally divided into two categories: parametric and nonparametric models. For parametric modelling, linear relationships between variables have been widely applied. Due to the stochastic and nonlinear characteristics of traffic

flow, it is difficult to overcome the limitations of parametric models. Limitations involved with traffic prediction approaches include difficulty discovering nonlinear relationships between the variables of the datasets and redundancy of the datasets.

Nonparametric machine learning methods have become increasingly popular. Artificial neural networks (ANN) have been commonly utilized for this problem [5], which can be regarded as the general pattern of machine learning application in traffic planning.

ANN is a computational model, considered to be a nonlinear process that analyzes the multifaceted interactions between inputs and outputs. ANN's focus is on the structure and function of biological neural networks. These are considered to be nonlinear data modelling processes, where the complex relationships between inputs and outputs are modelled and patterns are found. ANN-based modelling has recently become a successful field with many realistic applications and active research topics. However, shallow ANNs have very few hidden layers, whereas Deep Neural Networks (DNN) have at least three or more hidden layers. The purpose of having more layers is to extract finer features of the input. Generally speaking, increasing the depth of the model increases its power but also creates greater difficulties with complexity. There are few studies [5] [6] that focus on the relationship between traffic volume and land use changes using Artificial Neural Networks (ANNs). Typical shallow neural network methods face challenges in predicting traffic when the data have over-fitting issues or redundancy-dependency. DNNs have been employed to recognize common characteristics from datasets. These were accomplished by allowing a DNN to learn collective features of datasets and pulled these features into an integrated feature. DNNs use a more refined computational modelling approach than other similar approaches to processing data and they have been found to perform at a higher level of satisfaction than other neural networks due to the ability to overcome the weaknesses of typical ANN methods.

In this paper, a systematic procedure to predict traffic using the deep learning method was presented. The main contribution of this work was to investigate whether the deep learning models could outperform typical methods. Deep Neural Network (DNN)-based traffic prediction model was developed using the City of Calgary's datasets. Specifically, the focus of this paper was as follows:

For bulleted lists predicting traffic with a Deep Neural Network using two types of Deep Neural Network Models-DNN-Regression (employing the regression method in the Deep Neural Network process) and DNN-RNN (using the temporal data to predict future traffic)

Comparing the model's performance to other existing approaches, this demonstrated the effectiveness of the proposed models.

This study focused on land use and demographics as inputs in Dissemination Areas (DAs–a few blocks of each neighbourhood) in Calgary for traffic prediction while using only DNN-Regression. DNN-RNN was used for temporal traffic data from 2008 to 2018 to predict overall traffic temporally in the same area.

This paper is organized as follows: related works of literature are covered in Section 2, while Section 3 outlines the study area and the methodology. In Section 4, the experiments and analyses are presented, and the conclusions are covered in Section 5.

2. LITERATURE REVIEW

Predicting traffic volume is of great importance when seeking to optimize and manage traffic in an urban area. It is important to meet the goals of enhanced routing, scheduling, and detecting inconsistencies, as well as saving energy. Having a more refined process to effectively predict traffic would have an overall positive impact in all areas of transportation planning and management.

Methods used to solve traffic prediction have usually been traditionally a parametric approach. As the deep learning approach has become more popular, attention has turned to deep learning models but the continuation of exploration using traditional methods has not stopped. Researchers continue to implement parametric and nonparametric approaches such as using

regression and neural networks. The following sections introduce the evolution of these two approaches for traffic prediction.

Many successful models for traffic-related prediction have been developed. Traditional algorithms such as Linear Regression and Kalman Filtering have been found to predict traffic. Hou and Ma [7] selected the Grey Linear Regression approach to predict railway ridership in Henan Province, China. Yo and Cho [8] presented variables that could be used to infer traffic conditions, relationships between variables and traffic jams, that led them to construct a Bayesian Network model to represent the relationship among the variables to infer traffic jams. They also identified the variables (unexpected incidents) that could be used to infer traffic conditions - accidents, weather conditions (rain/snow), road works (construction), road blocks, and speed of vehicles to infer traffic conditions.

Considerable work has focused on increasing the ability to interpret the dynamics of traffic prediction, but time-series analytic approaches have mainly been used. However, most often methods such as Auto-Regressive Integrated Moving Average (ARIMA) and its sub-categories (i.e., ARIMAX, Mixed ARIMA, and SARIMA) were found to be limited to capturing information that exposed temporal traffic prediction [9]. Hybrid models combined the benefits of two or more types of methods in order to change the focus of how researchers approached predictions. A hybrid method of combined maps and ARIMA model was developed to forecast traffic at different time intervals and its performance was found to be superior to ARIMA [10]. In addition, many other modified ARIMA models have been examined, such as subset ARIMA [11], ARIMA with explanatory variables [12], and seasonal ARIMA [13]. Another widely used method to predict traffic volume is the filtering approach, such as the Kalman Filter model. Okutani and Stephanedes introduced Kalman filtering theory into this field and results indicated improved performances [14] [8]. Using these approaches tended to provide robust results from the fluctuating data, leading to issues in predicting traffic by skewing results with an over reproduction of observed values [15].

However, all the above-mentioned models or approaches had issues with non-linearity. Traffic prediction has a number of characteristics that have to be considered. These include the notion they are nonlinear, random, and unstable and the fact that they are a time-series. It was determined that nonparametric Neural Networks (ANN) could address these issues.

Nikravesh et al. [16] investigated the accuracy of Multi-Layer Perceptron (MLP), which was a typical architecture of ANN, for predicting the future behaviour of network traffic. Nikravesh et al. [16] also compared the model to Support Vector Machine (SVM). The results showed MLP was more accurate than SVM. Wang et al. used Local Stacked Auto-Encoders (LSAEs) and

Global Stacked Auto Encoder (GSAE) to extract local traffic and then utilized Long Short-Term Memory (LSTM) to predict traffic [17]. Wang et al. [17] found that ANN could efficiently approach a continuous function with an expectation of being nearly accurate. ANNs were able to demonstrate their effectiveness in solving nonlinear problems, as long as they contained a sufficient number of factors. However, if the number of ANN variables became extremely large, the amount of computation became too heavy to be applicable. Using Deep Neural Network (DNN), where it had more layers, provided more satisfactory results. DNN had at least three or more hidden layers compared to ANN [18]. Having more layers in DNN could provide better results from inputs. A DNN-based prediction model generally performed more satisfactorily than other models [18].

Liu et al. [19] used a hybrid LSTM model to predict the vehicle flow of each road section and intersection in an actual traffic network. They successfully used the maximum relative error to compare the difference between the actual and predictive vehicle flows. Ateeq et al. [20] used DNN to predict delays in traffic. Hua et al. [21] reinvestigated the field of traffic prediction with deep learning and proposed a Random Connectivity LSTM (RCLSTM) based on the conventional LSTM.

As mentioned above, there were recent popular deep learning-based traffic prediction studies available. Zheng et al. [22] examined a deep learning traffic prediction model based on a convolution neural network (CNN) framework. Zhou et al. [23] proposed a multiscale and high-precision LSTM model to predict the short-term traffic flow data. Zheng et al. [22] proposed a deep learning based model which used hybrid and multiple-layer architectures to extract traffic flow data. Experimental results were presented to show that the proposed model achieved better prediction performance when compared with other existing approaches. Wu et al. [3] examined LSTM to capture the temporal correlations in traffic time series data. Narmadha and Vijaykumar [24] showed multivariate traffic prediction by using LSTM in their paper. Results were compared with LSTM univariate and convolutional neural network (CNN) models and it was found that the deep learning LSTM could be used in several learning and prediction scenarios. Deep Neural Network-based models have contributed considerably to prediction in many studies. However, existing works did not take advantage of using two deep learning approaches in the same city and did not determine relationships between different variables to predict traffic. The proposed approach was designed to use two Deep Neural Network approaches to predict traffic in Calgary.

3. STUDY AREA AND METHODOLOGY

3.1. Study Area

Calgary, the city with the largest population (1.2 million) in Alberta Province, Canada owes its growth to being the center of Canada's oil industry. It has an extensive transportation network that combines road, rail, air, public transit, and public infrastructure. Calgary (Figure 1) was chosen as a representation of the challenges seen in urban growth. Nevertheless, there are always influential surrounding factors that have an effect should be acknowledged such as estimates of intra-area traffic flow proportion.

The data that was provided for the purposes of this study contained land use (2011), demographics (2015) (population, age, gender, and employment), as well as traffic count data obtained in a geographic information system (GIS) format at a Dissemination Area (DA) (2011) level. All datasets were collected from the Spatial and Numeric Data Services from the University of Calgary Library [2].
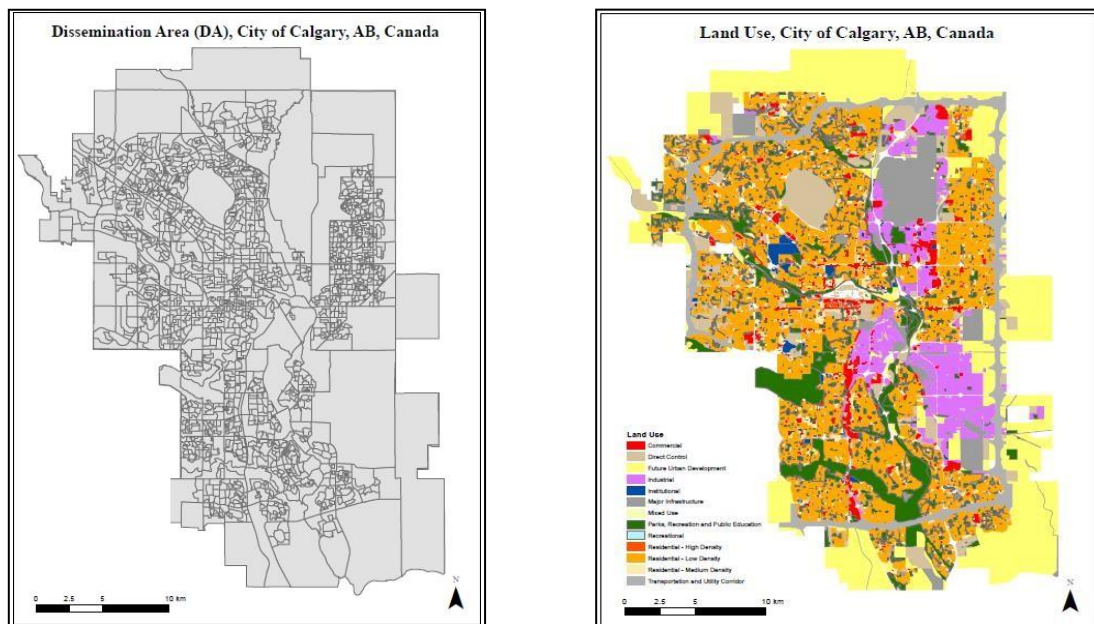
4

In the obtained database, different types of land usage were identified. Categories of land use were labelled as follows: unclassified, commercial, direct control, future development, industrial, institutional, major infrastructure, mixed-use, recreation, transportation utility corridor, as well as parks and recreation. Additionally, there were three types of residential land uses that were classified based on the city of Calgary's data sources—residential high, medium, and low density [2] (Table 1).

Table 1. Land Use Data, Transportation-Related Data and Demographic Data. Table 1. Description statistics of the data. Table 1 provides an example of the composition of the data (Traffic, land use and demographic data), such as the specific values under the different attributes

	N	DESCRIPTION	MINIMUM	MAXIMUM	MEAN	STD. DEVIATION
TOTALM (DEPENDENT VARIABLE, 2010)	1590	Total number of vehicles in each DA dissemination area (DA)	0.00	3545.00	329.84	260.98
INDEPENDENT VARIABLE						
AREA	1590	DA Area	0.01	450.36	5.22	23.80
LAND USE VARIABLES						
UNCLASS	1590	Unclassified	0.00	72 592.07	434.96	4203.16
COMMERCA	1590	Commercial	0.00	99 375.86	5085.13	12 790.80
DIRECTCA	1590	Direct Control	0.00	99 168.91	4749.42	13 617.06
FUTUREA	1590	Future Development	0.00	92 043.22	1456.85	7981.12
INDUSTRA	1590	Industrial	0.00	83 711.75	354.00	4206.49
INSTITUTA	1590	Institutional	0.00	96 700.00	1957.27	8265.11
MAJINFRA	1590	Major Infrastructure	0.00	98 117.36	3033.97	11 785.12
MIXEDUSEA	1590	Mixed Use	0.00	65 702.47	90.31	2026.29
PARKRECA	1590	Parks and Recreation	0.00	99 695.69	16329.8	23 592.38
RECREATA	1590	Recreation	0.00	18 508.36	13.23	468.42
TRANUTIA	1590	Transportation and Utility Corridor	0.00	92 233.65	955.78	7921.20
RESIDHDA	1590	Residential High Density	0.00	97 268.68	857.69	5884.98
RESIDMDA	1590	Residential Medium Density	0.00	99 994.88	14080.40	21 332.28
RESIDLDA	1590	Residential Low Density	0.00	98 983.94	22087.72	34 248.61
DEMOGRAPHIC VARIABLES						
TOTALP	1590	Total Population	75.00	6785.00	690.31	511.86
MALE	1590	Gender Male	40.00	3450.00	344.51	255.31
FEMALE	1590	Gender Female	35.00	3340.00	345.88	258.49
EMPLOYED	1590	Total Number of Employed	0.00	3934.00	388.04	291.89
UNEMPLOYED	1590	Total Number of Unemployed	0.00	270.00	20.73	25.08
NOTINLF	1590	Not in Labor Force	0.00	1295.00	144.19	97.48

Traffic count refers to the total number of vehicles in each DA prepared by the City of Calgary. A Dissemination Area is a small, relatively stable geographic unit composed of one or more adjacent road blocks. It is the smallest standard geographic area for which all census data are distributed (Figure 1).

The dependent variable in this study is the traffic count. The list of independent variables assessed in this study is presented in Table 1. Land use and transport provisions are closely related issues, each being reactive to the other. Therefore, this study explores the predictive ability of land use on traffic count. Land use, such as residential density, allows a closer look at land uses where people live and provides a link to measures of people that contribute to traffic on the road. Population density measures the number of people per square kilometre; residential density measures the number of living units per square kilometre. A proportional increase in residential density may correspond with a proportional increase in population density. Since residential density will have an impact on activity, it is apparent that it should have a strong correlation to traffic count prediction.

(a) Study Area: Dissemination Areas   (b) Study Area: Land Uses

Figure 1. Study area and land uses City of Calgary, AB, Canada. Figure 1. Study Area, City of Calgary, AB, Canada. Number of Dissemination Area (DA): 1590

Demographic data were collected from the City of Calgary that covered population, gender, and employment rates. This information was also included in the model as additional input. Further demographic data (population, gender, employment) were also obtained from Spatial and Numeric Data Services from the University of Calgary Library. These data were made available as an attribute of the polygons at the Dissemination Area (DA) level and were part of the same area as land uses [2]. Table 1 illustrates all independent demographic variables explored in this study.

Land use and demographic data (Figure 1 and Table 1) were used as inputs to predict traffic. It is important to acknowledge traffic prediction models, which provided good approximations for traffic loads using population and employment. However, other important variables such as road network variables (capacity, maximal speed) were not considered here because of unavailability

of data during the time of this study. Generally, in addition to land use and demographic variables, other factors such as weather and road status are also important considerations for predicting traffic [25] [26] [27].
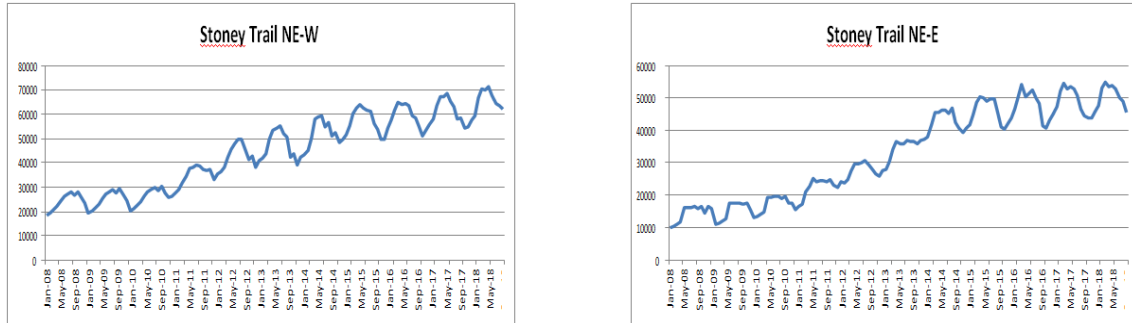

Figure 2. 24-hour period on an average weekday, monthly data from the year 2008 to 2018 at two monitoring stations, northeast Calgary. Source: City of Calgary.

The temporal traffic data collected for this case study contained ten years (2008 to 2018) of monthly traffic obtained from two monitoring stations located at intersection of two major highways Stoney Trail - Deerfoot Trail in the city of Calgary, Canada. Calgary operated these two traffic monitoring stations which collected traffic volumes 365 days of the year. The average annual monthly traffic data provided the numbers for vehicles passing a traffic monitoring station over a 24-hour (daily) period, on an average weekday. Figure 2 demonstrates two stations' 24-hour periods on an average weekday, monthly data from the year 2008 to 2018 in Calgary. In Figure 2, the horizontal axis is the time-in-month unit and the vertical axis is the daily (24-hour) traffic count. The patterns of daily traffic count can be observed from this figure.

3.2. Methodology

For traffic prediction, this study proposed two deep learning methods deep neural network regression (DNN-Regression) and Recurrent Neural Network (DNN-RNN). In traffic prediction processes using DNN, three tasks were addressed: 1) exploring the datasets, 2) traffic prediction using the DNN-Regression and DNN-RNN models, and 3) comparison of the models with other data mining models. As shown in Figure 3, two types of deep neural network models were proposed (DNN-Regression and DNN-RNN models) to predict traffic.
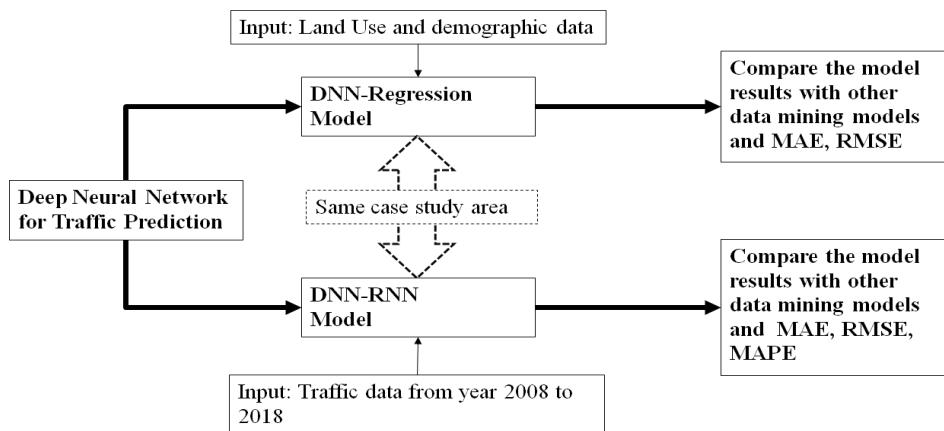

Figure 3. Overall methodology.

Deep Neural Networks (DNN-Regression) use sophisticated mathematical modelling to process data in complex ways. It was an extension to the FFNN (Feed Forward Neural Network) model but it was a more efficient method in that it contained more than one hidden layer. This was accomplished by incorporating several levels of non-linearities [28], resulting in a much improved non-linear regression function.

The proposed prediction model was composed of three layers (see Figures 4). The first input layer was compatible with the land use and demographic input variables. The next layer was hidden and was used to capture non-linear relationships between variables. The values in the hidden layer were standardized in order to avoid variation of large values. It prevented the learning deletion caused by large values. The third output layer provided the expected values of traffic prediction.
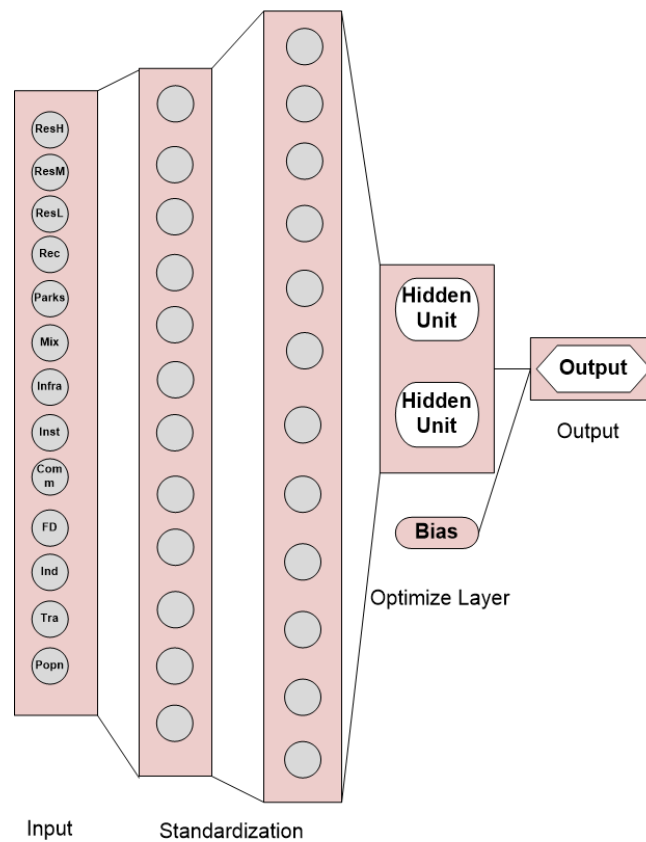

Figure 4. Proposed DNN-Regression Model (Input, Standardization, Optimize, Output).

Preprocessing stage: This stage consisted of different steps, including features extraction, transforming nominal to numeric and correlation of data if there were any multicollinearity between variables. The correlation between variables was explored to minimize any possible bias that might arise. Statistical tests were conducted by computing the Pearson correlation coefficient to examine the correlation between the independent variables. Two variables were considered to be strongly correlated to each other if the computed Pearson correlation coefficient was less than -0.5 or greater than $+0.5$ (significance value less than 0.01 for the considered data). The ranges of this value were available in different studies [29]. Showing the predictor variables that exhibited a strong correlation with other predictor/independent variables were omitted from the modelling process to minimize multicollinearity [29]. In this study, various land uses and each DA's total population were considered for the DNN-Regression Model. The remaining

independent variables were excluded from the modelling process to minimize the multicollinearity.

Prediction stage: This stage involved building a new prediction model which could forecast traffic in each Dissemination Area. This was achieved by building a regression model for traffic prediction, using the deep learning concept of neural networks. This model depended on solving the problem completely from the input to predicting traffic. It consisted of three sequential layers – a generation layer, a standardization regression layer, and a regression predictor layer. The topology of the suggested deep neural network is shown in Figure 5.

Layer 1: Indicator generation layer. The aim of this layer was to enhance the accuracy of prediction by taking the standard features of land use and demographics as inputs for the network. The output from this layer went on to the next step. The major characteristic of this layer was the output neuron (indicator) only connected with input neurons that ultimately contributed to finding the indicator.

Layer 2: Standardization regression layer: The values in all features were standardized. The inputs of this layer merge the inputs and outputs of layer 1.

Layer 3: Regression predictor layer: This layer developed a prediction model through Optimize Multilayer Perceptron (OMLP) ANNs with one hidden layer by minimizing the given loss function (squared error).

A sigmoid (logistic) (see following) activation function was used in the hidden layer. Also, this activation function was used in the output layer. Two neurons were determined in the hidden layer in addition to bias. The value of tolerance was used and weights were optimized by reducing the error.

$$\int (x) = \frac{1}{1 + e^{-x}}$$

DNN with fully connected layers could be trained to obtain more accurate results in a time series prediction. The architecture of RNN was naturally suitable for modelling sequence data because the recurrent connections within each neural cell enabled the cells to map the input of sequential data and output from each of them. Recurrent Neural Networks (DNN-RNN) had a fundamental component called Long Short-Term Memory (LSTM), which was an addition of the Feed Forward Neural Networks. By using the power of RNN's Long Short-Term Memory (LSTM) units, a more in-depth and predictive process could be obtained. Solving traffic prediction problems could be addressed by using these deep neural network structures [23].

RNNs consider two sources of input (i.e., the present and the recent past), which were combined to determine how the network should respond to. DNN-RNN used cells that were a function of inputs from previous time steps, also known as memory cells. DNN-RNN were found to be flexible in their inputs and outputs for both sequences and single vector values.

RNNs for time series presented their own gradient challenges. A possible solution was to shorten the time steps used for prediction, but this made the model worse at predicting longer trends. Another issue RNN faced was that, after a while, the network began to 'forget' the initial inputs as information was lost at each step. The LSTM (Long Short-Term Memory) cell was created to help address these RNN issues. LSTMs, as a specific subgroup of RNNS, were introduced in Hasim et al.[30] and were specifically created to bypass or eliminate any long reliance on

dependency issues. This resulted in the exclusion of any vanishing-gradient issues, which were often seen in normal RNNs. The structure of the LSTM (Figure 5) was able to learn long-term dependencies due to its structure. It was also useful in integrating and regulating the learning process.
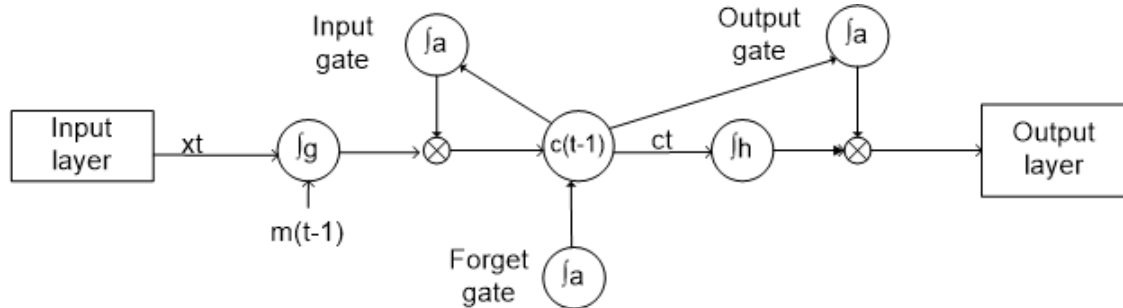

Figure 5. Proposed architecture for the DNN-RNN LSTM traffic prediction. Input and output data: temporal traffic volume data.

Monthly data collected over a 10-year period (2008-2018), from two traffic monitoring stations found in Calgary, were used for the purposes of this study. To evaluate the performance of the proposed architecture, this traffic data was used, as described in Section 3.1.

The following steps illustrate the development of the DNN-RNN model and parameters:

Step 1. Use traffic data to prepare training and test data set.

Step 2. Use architecture optimization data set to decide optimal architecture parameters, including input dimension, layer size, hidden units size per layer, and epochs.

Step 3. Use training data set to train the LSTM based on optimal architectures.

Step 4. Use test data set to examine prediction performance. Compare the results with benchmark models Naive, ANN and ARIMA.

3.3. Fitting and Comparing Models

In order to evaluate the performance of the models, a comparative study with popular prediction methods – Neural Network, K-Nearest Neighbour (K-NN) and ARIMA - was proposed in this study. Additionally, an evaluation error metric - Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE), were applied.

A simple Naive model was used to serve as a benchmark approach. It can be assumed that any complex model such DNN-RNN would outperform a Naive model. Naive forecasts, adapted from Smith et al. [31], were calculated in a way that took into account both the current conditions and the historical pattern. Average flow rates were calculated for each month using the data for the same locations. For the test data, Naive forecasts were calculated using these historical average flow rates by the equation:

$$T_{pred} = (T(t))/(T_h(t)) * T_h (t+1)$$

Where T_{pred} the predictor forecast for the next traffic, where $T_{(t)}$ is the traffic rated at the current time and Th the historical average of the traffic associated with time interval t.

Neural Network (NN or ANN): ANN considered being a nonlinear process that analyzed the interactions between inputs and outputs. It focused on the structure and function of biological neural networks [32]. The activation function of ANN applied in this study to have a non-linear transformation of input.

K-Nearest-Neighbour (KNN): KNN was another type of learning model that predicted the entity based on the K nearest training instances in the feature space. K nearest neighbors was a simple algorithm that stored all available cases and predicted the numerical target based on a similarity measure (e.g., distance functions).

ARIMA (Auto Regressive Integrated Moving Average): ARIMA was a popular and widely used statistical method for time series forecasting. This model captured a suite of different standard temporal structures in time series data.

To evaluate the performance of the predictor, three performance measurements were selected, which were the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the root mean square error (RMSE), as demonstrated below. The mean absolute error (MAE) is an indication of the average deviation of the predicted values from the corresponding observed values. MAE can present information on the long-term performance of a model, whereby lower MAEs indicated improved long-term model prediction. Root Mean Squared Error (RMSE) is the aggregate squared error that produced results relative to what the error would have been if the prediction had been the average of the absolute value. Lower RMSE values were an indication of greater accuracy and were a better prediction model. The expressions of all measures are provided below. The units of MAE and RMSE were both vehicles per day (veh/day).

$$MAE = \frac{1}{n} \sum_{i=0}^{n} |\overline{y}_i - y_i|$$

$$RMSE = \frac{\sum_{i=1}^{n}(\overline{y}_i - y_i)^2}{\sum_{i=1}^{n}(mean(y) - y_i)^2}$$

$$MAPE = \frac{1}{\overline{n}} \sum_{i=0}^{n} |\overline{y}_i - y_i| / \overline{y}_i$$

N (n) denoted the total number of points, while and were the predicted values and their corresponding observations. This same metric was used to compare the accuracy of the proposed architecture with that obtained using other predictive algorithms adapted from Xu et al.[33]. The implementation of the traffic prediction algorithm was done in Python, using Tensorflow as backend.

4. EXPERIMENTAL RESULTS

In this section, experimental results of using two deep learning approaches DNN-Regression and DNN-RNN to predict traffic are described. As demonstrated in Section 3, research has mainly focused on administrating multiple linear regression to investigate interactions between land use,

demographic, and transportation-related variables in order to predict traffic. Recently, machine learning-based models for predicting traffic have been much more successful than other statistical methods. Due to lack of consensus amongst researchers as to which method for accurately predict traffic works best, several well-known models have been incorporated as a baseline for comparative studies. For the purposes of this study, the selected models were as follows: Artificial Neural Network (ANN), K-Nearest Neighbour (KNN) and ARIMA. Experiment group 1 (presented in Section 3) evaluated the DNN-Regression model in comparison with three typical models in traffic prediction. Experiment group 2 (presented in Section 3) evaluated the DNN-RNN model in comparison with two baseline models by using time series variables. The effectiveness and outcomes of using DNN-Regression and DNN- RNN models were also demonstrated in Section 3.

All data preparation and processes of this research study were completed using the programming language Python. Pandas library in Python was used to explore the data and the two proposed models were implemented with Tensorflow [34].

4.1. Deep Neural Network Regression (DNN-Regression) Results

DNN-Regression was used to predict traffic using variables: land uses and demographics. The DNN-Regression model is an extension to the FFNN (Feed Forward Neural Network) model. Before the model could be implemented, the data needed to be split into two parts - one for training and the other for testing. The type of training depended on the amount of data provided, as well as the training design (see pseudo algorithm below).

Pseudo Code for DNN-Regression

Data: Trained samples input X, output traffic prediction Y, and hidden layers h

Result: Training the model

1. w = weight of sparsity;
2. MinMax fit scaler;
3. Initialize weights;
4. Input X;
5. Creating index for xtrain data
6. Foreach layer in h do
7. Calculate n' = nlayer(input)
8. Calculate loss (n', Y);
9. Create the estimator model. Use a DNN Regressor.
10. Minimize loss and repeat from n' (until reach to convergence);
11. Let input be layer
12. Train the prediction layer with calculate errors until convergence
13. Calculate the RMSE and MAE, comparing between y test and final prediction

The Pseudo code illustrated the DNN-Regression model used in this study. The output of the traffic prediction was the Y values; while land uses and demographics were X values. Data was split by using 'sklearn' in Python, 70% for training and 30% for testing. Initially, only the training data was used for the purpose of studying this model, whereas the testing data was excluded in order to avoid biased results. The DNN-Regression model used 13 variables as inputs. The model was trained for 1,000-25,000 steps, which created a prediction input function.

The maximum number of iterations in our experiments was set to 100,000. The number of hidden units in each layer varied from 1 to 13, while the learning rate was set to 0.01.

In this study, ANN used 13 variables as input. The activation function of ANN was applied to have a non-linear transformation. The same logic applied for all input as forward propagation. Next the derivative of cost function (sigmoid) was used to generate a list of derivatives called backward propagation. Finally gradient descent was applied. Then a new round iteration started with updated parameters. The algorithm was not stopped until it converged.

KNN had been used in this study for traffic prediction as a non-parametric technique. The implementation of the KNN regression calculated the average of the numerical target of the K nearest neighbors. Choosing the optimal value for K was done by first inspecting the data. In general, a large K value was more precise as it reduced the overall noise. The optimal K value of 201 was used in this study.

Next, it used the prediction method from the estimator model to create a list of predictions on the test data. The output was created as a continuous traffic count as the numeric column. Then, it calculated the RMSE and MAE by using sklearn comparing y test and final prediction.

The DNN-Regression was proposed as a new model which involved an automatic and integrated process, resulting in an optimal mathematical model that solved the convergence problem in the Neural Networks. As mentioned earlier, land use and demographic data were used to predict traffic in Calgary. The DNN-Regression, ANN and KNN results were compared for checking the difference from the prediction. The results for the proposed model indicated better results when compared with other models (ANN and KNN) Two evaluation methods, RMSE and MAE, were incorporated to evaluate the proposed regression predictor model. The results of RMSE and MAE were applied to the Calgary's datasets and have been summarized in Table 2. Table 2 presents the performance evaluation results of the prediction model and the ANN and KNN models. The DNN-Regression model's RMSE and MAE values were lower compared to the other two models. This indicated DNN-Regression's higher efficiency compared to the other two models, as this was a more appropriate method of predicting traffic than ANN and KNN.

Table 2. DNN-Regression, KNN and ANN models results comparisons.

MODEL	RMSE	MAE
DNN-REGRESSION (PROPOSED METHOD)	120	190
NEURAL NETWORK (ANN)	135	214
KNN	225	308

Table 2 also depicts the predictive performance of the proposed DNN-Regression model and two other common methods of traffic prediction. It was observed that DNN-Regression achieved the lowest results among all approaches. Table 2 shows that the DNN-Regression model produced a RMSE error of 120, while other models had RMSE error of 135 for ANN model and 225 for KNN models. The MAE error of the proposed DNN regression model was 190, while the MAE of the ANN and KNN were 214 and 308 respectively. DNN-Regression showed improvements over the best baseline models in terms of MAE and RMSE, respectively. Table 2 indicates an RMSE of 120 in DNN-Regression model; this meant errors on average, was 120 times larger than the variance of traffic counts. A possible reason to have a small variance could be that these areas are located on busy Deerfoot Trail (highway) and other nearby highways. Another possible reason for this is that they relied entirely on historical traffic values for prediction without considering other critical environmental variables.

The DNN-Regression model predicted traffic based on land use, as well as demographic data used in all Dissemination Areas. There was, however, another method that could predict temporal traffic in the same area. DNN-RNN could predict traffic temporarily. DNN-Regression took the ability to predict traffic using different but was not as successful for temporal data.

4.2. Deep Neural Network Recurrent Neural Network (DNN-RNN) Results

As mentioned in the Study Area section, this study used traffic data collected from the City of Calgary as the dataset. The time period selected was from January 2008 to December 2018. In order to make the evaluation more consistent, effective, and efficient, only temporal data from the same area in Calgary (used for DNN-Regression) was used in the experiments. The DNN- RNN models were found to produce positive prediction results when used with sequential dates. All traffic data had similar types of sequences, i.e., each level was dependent upon the previous one. The goal of exploring this model was to find a way to mine time-series information from traffic data.

The architecture of the DNN-RNN predictor depended on several principal parameters, including the number of previous time intervals utilized for prediction, the hidden layer size, the units in each hidden layer, and the number of epochs for training a DNN-RNN. To decide the model architecture, the previous time intervals for prediction were set from 12 to 36, and the hidden layer size is from 1 to 10. The units in each DNN-RNN layer were chosen from 100 to 1000 with gap value 100, and the epochs are set from 10 to 50 with gap value of 10.

In this paper the DNN-RNN predictor was utilized to predict the next 12, 24, and 36 months daily traffic flow. The traffic count data collected from three monitoring stations aggregated as one data point for this study. For each of the three prediction horizons, grid searches had been implemented to decide the most effective architecture for predicting based on the training set. Regarding the units in each hidden layer, the number could be neither too small nor too large. Furthermore, the epoch time ranges were from 10 to 40. Using a grid search, the optimal LSTM architecture was determined for arterial traffic flow prediction in northeast Calgary.

The difference between MAE and RMSE's observed and data are presented in Table 3 and Figure 6 below. For all the predictors, performance outcomes showed better results when use for a 12 month horizon versus a 24 or 36 month horizon. This finding was expected and consistent with other studies wherein larger errors were observed for longer prediction horizon [35].

Table 3. Model results comparisons.

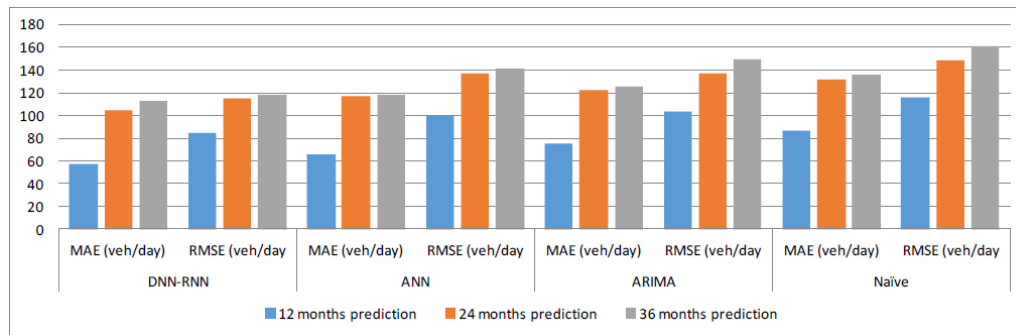| MODEL		12 MONTHS PREDICTION	24 MONTHS PREDICTION	36 MONTHS PREDICTION
DNN-RNN	MAE	5696	10431	11259
(LSTM)	RMSE	8446	11459	11790
ANN (FFNN)	MAE	6623	11656	11855
	RMSE	10016	13677	14074
ARIMA	MAE	7551	12221	12585
	RMSE	10366	13677	14904
NAIVE	MAE	8652	13212	13586
	RMSE	11577	14845	16106

Figure 6. Model results comparison. RMSE and MAE values were divided by 100 for graphical presentation purposes.

In comparison with the predictors, DNN-RNN performed better for all time horizons, indicating that the accuracy of deep learning could be improved with LSTM. This emphasized the significance of using the DNN-RNN model for traffic flow prediction. Furthermore, the LSTM performed better than ARIMA and ANN, which resulted in providing the advantage of LSTM to capture the time series characteristics of traffic data.

When comparing with the other existing methods, ANN and ARIMA were chosen to represent a shallow machine learning model and a typical parametric model, respectively. For ARIMA, the optimum structure was found an ARIMA (4,1,2) model. This set the lag value to 4 for autoregression, used a difference order of 1 to make the time series stationary, and used a moving average window of 2.

This study also used the Friedman Test [36] to better understand the performance of each of the models. A comparative study of the methodology was adapted from Xu et al. [33] who found the lower value of absolute percentage errors (MAPE) provided better results. As seen in Table 4, the MAPE from each of the modelling forecasts were analyzed. MAPE results are shown in the table where the null hypothesis of the Friedman test could be rejected at the alpha = .05 significance level. Table 4 demonstrates how the related sample tests for this case study supported the DNN-RNN model as the better performer, followed by the other three models.

From Table 4, it was noted that these three shallow models were less effective than the LSTM. Furthermore, the ANN had better accuracy than ARIMA. Another interesting outcome was that the LSTM worked better than regular ANN, while the ARIMA was not as effective as the regular ANN. This phenomenon indicated that the ability to conduct ARIMA was limited for the time series model in this particular case.

Table 4. Statistical significant results by using Friedman Test.

MODEL	MEAN RANK	MAPE
DNN-RNN (LSTM)	4.82	10.27
ANN (FFNN)	6.37	13.41
ARIMA	6.37	13.77
NAIVE	6.82	14.52

The DNN-Regression model predicted traffic using demographics and land uses and the DNN-RNN had the ability to temporarily predict traffic. Both deep learning models demonstrated advantages over other modelling techniques.

15

5. CONCLUSIONS

Multiple researchers have explored a number of approaches to predicting traffic, with varied success. In this study, a Deep Neural Network-based modelling framework for predicting traffic was used. The DNN-Regression model incorporated different variables such as land uses and demographics. The results of using this model indicated that it outperformed the corresponding values produced by other traditional methods (e.g., KNN). This study also proposed the implementation of a time-related DNN-Recurrent Neural Network (DNN-RNN) model. By taking advantage of the unique properties of RNN, temporal traffic prediction was explored. Completing an in-depth literature review led to the belief that using two deep learning methods

for predicting traffic had never been attempted. Speculative DNN-RNN experiments were completed on a dataset which was generated from reported traffic counts in the northeast part of Calgary from 2008 to 2018. These experiments validated the suitability and appropriateness of the DNN-RNN model, compared with other models (e.g., Naive, ARIMA). Notably, the study incorporated information that was directly pulled from actual three traffic monitoring stations. The model was implemented to predict traffic for forthcoming multiple steps. This was considered to be of great importance in terms of making predictions using fewer data procedures, as well as using a smaller pool of observations.

This study considered only the temporal pattern in DNN-RNN into account. In future work, the spatial pattern could be considered, and the model could be expanded to have the ability to learn spatial-temporal dependence. In addition, the model predicted traffic well when the time interval was short, but it performed poorly when the time interval was a little longer. Future work could extend the model to a more generalized version so that it could perform better when the time interval is both short and long.

Some potential issues that remain to be explored are the consideration of variables that impact the accuracy of predicting traffic such as weather and road conditions. Adaptive approaches should explore and address these variables and their correlation in order to better understand the non-linear behaviour inherent to traffic prediction. Since weather conditions impact traffic, making the network have the ability to learn the correlation between weather conditions and traffic should be considered in future studies. Ongoing integrated and innovative approaches for making accurate predictions may lead to a more effective means of augmenting transportation planning and management.

ACKNOWLEDGEMENTS

REFERENCES

[1] T. Straatemeier and L. Bertolini, "How can planning for accessibility lead to more integrated transport and land-use strategies? Two examples from the Netherlands," European Planning Studies, vol. 28, no. 9, pp. 1713–1734, 2020, doi: 10.1080/09654313.2019.1612326.
[2] A. Azad and X. Wang, "Prediction of Traffic Counts Using Statistical and Neural Network Models," Geomatica, vol. 69, no. 3, pp. 297–311, 2015.
[3] Q. Wu, Q. Fu, and M. Nie, "Graph Wavelet Long Short-Term Memory Neural Network: A Novel Spatial-Temporal Network for Traffic Prediction," J. Phys.: Conf. Ser., vol. 1549, p. 042070, 2020, doi: 10.1088/1742-6596/1549/4/042070.

[4]   L. Kang, G. Hu, H. Huang, W. Lu, and L. Liu, "Urban Traffic Travel Time Short-Term Prediction Model Based on Spatio-Temporal Feature Extraction," Journal of Advanced Transportation, 2020. https://www.hindawi.com/journals/jat/2020/3247847/.

[5]   A. Boukerche, Y. Tao, and P. Sun, "Artificial intelligence-based vehicular traffic flow prediction methods for supporting intelligent transportation systems," Computer Networks, vol. 182, 2020, doi: 10.1016/j.comnet.2020.107484.

[6]   B. C. Pijanowski, D. G. Brown, B. A. Shellito, and G. A. Manik, "Using neural networks and GIS to forecast land use changes: A Land Transformation Model," Computers Environment and Urban Systems, vol. 26, no. 6, pp. 553–575, 2002, doi: 10.1016/S0198-9715(01)00015-1.

[7]   L. Hou and G. Ma, "Forecast of Railway Passenger Traffic Based on a Grey Linear Regression Combined Model," Computer Simulation, vol. 7, pp. 1–10, 2011.

[8]   Y. J. Yu and M. Cho, "A Short-Term Prediction Model for Forecasting Traffic Information Using Bayesian Network," in 2008 Third International Conference on Convergence and Hybrid Information Technology, 2008, vol. 1, pp. 242–247, doi: 10.1109/ICCIT.2008.355.

[9]   Y. Jia, P. He, S. Liu, and L. Cao, "A combined forecasting model for passenger flow based on GM and ARMA," International Journal of Hybrid Information Technology, vol. 9, no. 2, pp. 215–226, 2016.

[10]   M. Van Der Voort, M. Dougherty, and S. Watson, "Combining kohonen maps with arima time series models to forecast traffic flow," Transportation Research Part C: Emerging Technologies, vol. 4, no. 5, pp. 307–318, 1996, doi: 10.1016/S0968-090X(97)82903-8.

[11]   S. Lee and D. B. Fambro, "Application of Subset Autoregressive Integrated Moving Average Model for Short-Term Freeway Traffic Volume Forecasting," Transportation Research Record, vol. 1678, no. 1, pp. 179–188, 1999, doi: 10.3141/1678-22.

[12]   B. M. Williams, "Multivariate Vehicular Traffic Flow Prediction: Evaluation of ARIMAX Modeling," Transportation Research Record, vol. 1776, no. 1, pp. 194–200, 2001, doi: 10.3141/1776-25.

[13]   B. M. Williams and L. A. Hoel, "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results," Journal of Transportation Engineering, vol. 129, no. 6, pp. 664–672, 2003, doi: 10.1061/(ASCE)0733- 947X(2003)129:6(664).

[14]   I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," Transportation Research Part B: Methodological, vol. 18, no. 1, pp. 1–11, Feb. 1984, doi: 10.1016/0191-2615(84)90002-X.

[15]   H. D. Trinh, L. Giupponi, and P. Dini, "Mobile Traffic Prediction from Raw Data Using LSTM Networks," 2018, doi: 10.1109/PIMRC.2018.8581000.

[16]   A. Y. Nikravesh and S. A. Ajila, "An experimental investigation of mobile network traffic prediction accuracy," Services Transactions on Big Data, vol. 3, no. 1, 2016.

[17]   J. Wang et al., "Spatiotemporal modeling and prediction in cellular networks: A big data enabled deep learning approach," in INFOCOM 2017-IEEE Conference on Computer Communications, 2017, pp. 1–9.

[18]   "Neural networks and deep learning [Book]." https://www.oreilly.com/library/view/neural- networks-and/9781492037354/ (accessed May 29, 2019).

[19]   Y. Liu, Y. Wang, X. Yang, and L. Zhang, "Short-term travel time prediction by deep learning: A comparison of different LSTM-DNN models," in 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Oct. 2017, pp. 1–8, doi: 10.1109/ITSC.2017.8317886.

[20]   M. Ateeq, F. Ishmanov, M. K. Afzal, and M. Naeem, "Predicting Delay in IoT Using Deep Learning: A Multiparametric Approach," IEEE Access, vol. 7, pp. 62022–62031, 2019, doi: 10.1109/ACCESS.2019.2915958.

[21]   Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Traffic Prediction Based on Random Connectivity in Deep Learning with Long Short-Term Memory," arXiv:1711.02833 [cs], Nov. 2017, Accessed: Nov. 30, 2018. [Online]. Available: http://arxiv.org/abs/1711.02833.

[22]   H. Zheng, F. Lin, X. Feng, and Y. Chen, "A Hybrid Deep Learning Model With Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction," IEEE Transactions on Intelligent Transportation Systems, pp. 1–11, 2020, doi: 10.1109/TITS.2020.2997352.

[23]   J. Zhou, H. Chang, X. Cheng, and X. Zhao, "A Multiscale and High-Precision LSTM-GASVR Short-Term Traffic Flow Prediction Model," Complexity, 2020, Accessed: Sep. 23, 2020. [Online]. Available: https://www.hindawi.com/journals/complexity/2020/1434080/.

[24] S. Narmadha and D. V. Vijayakumar, "Multivariate Time Serious Traffic Prediction Using Long Short Term Memory Network," vol. 9, no. 04, pp. 1–6, 2020.

[25] C. M. Queen and C. J. Albers, "Intervention and Causality: Forecasting Traffic Flows Using a Dynamic Bayesian Network," Journal of the American Statistical Association, vol. 104, no. 486, pp. 669–681, Jun. 2009, doi: 10.1198/jasa.2009.0042.

[26] M. Wachs, "Fighting Traffic Congestion with Information Technology. (cover story)," Issues in Science & Technology, vol. 19, no. 1, p. 43, 2002.

[27] Guozhen Tan, Zhipeng Liu, and Yaodong Wang, "The determination and analysis of traffic congestion evacuation priority," in 2010 Second IITA International Conference on Geoscience and Remote Sensing, Qingdao, China, 2010, pp. 484–487, doi: 10.1109/IITA-GRS.2010.5603003.

[28] "Traffic sign recognition with multi-scale Convolutional Networks - IEEE Conference Publication." https://ieeexplore.ieee.org/abstract/document/6033589 (accessed Nov. 29, 2018).

[29] R. Taylor, "Interpretation of the Correlation Coefficient: A Basic Review," Journal of Diagnostic Medical Sonography, vol. 6, no. 1, pp. 35–39, 1990, doi: 10.1177/875647939000600106.

[30] H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition," 2014. http://arxiv.org/abs/1402.1128 (accessed Nov. 30, 2018).

[31] B. L. Smith, Williams B. M., and Oswald R. J., "Comparison of parametric and nonparametric models for traffic flow forecasting," Transportation Research Part C: Emerging Technologies, vol. 10, no. 4, pp. 303–321, 2002.

[32] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research," J Pharm Biomed Anal, vol. 22, no. 5, pp. 717–727, 2000.

[33] W. Xu, Q. Wang, and R. Chen, "Spatio-temporal prediction of crop disease severity for agricultural emergency management based on recurrent neural networks," GeoInformatica, vol. 22, no. 2, pp. 363–381, 2018, doi: 10.1007/s10707-017-0314-1.

[34] G. Zaccone, M. R. Karim, and A. Menshawy, Deep Learning with TensorFlow. Packt Publishing Ltd, 2017.

[35] Z. Zhao, W. Chen, X. Wu, P. C. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," IET Intelligent Transport Systems, vol. 11, no. 2, pp. 68–75, 2017.

[36] "Friedman Test - an overview | ScienceDirect Topics." https://www.sciencedirect.com/topics/medicine-and-dentistry/friedman-test.

[37] Y. Xu and D. Li, "Incorporating Graph Attention and Recurrent Architectures for City-Wide Taxi Demand Prediction," ISPRS International Journal of Geo-Information, vol. 8, no. 9, p. 414, 2019, doi: 10.3390/ijgi8090414.

AUTHORS

Abul Azad is currently specializing in GIS-based transportation and land use planning as a PhD researcher at the Department of Geomatics Engineering, University of Calgary. He received his MSc in Civil Engineering (Specialized in Transportation) from the University of Calgary. He also received his graduation in Urban Planning from University of Twente (ITC), Netherlands, and Bangladesh University of Engineering and Technology, Bangladesh. His research interest includes GIS-based land use planning, transportation planning, risk analysis and traffic safety.

Xin Wang is a Professor at the Department of Geomatics Engineering, University of Calgary. She holds a BSc in Computer Science, MEng in Software Engineering from Northwest University, China, and a PhD in Computer Science from the University of Regina. Her current research interests are spatial databases and spatial data mining, data mining for engineering applications, ontology and knowledge engineering in GIS, web GIS and location-based social networks.