A Comparison of ARIMAX, VAR and LSTM on Multivariate Short-Term Traffic Volume Forecasting

Bhanuka Dissanayake, Osanda Hemachandra, Nuwan Lakshitha, Dilantha Haputhanthri, Adeesha Wijayasiri
University of Moratuwa
Sri Lanka
{bhanuka.16, osanda.16, nuwan.16, dilantha.15, adeeshaw}@cse.mrt.ac.lk

***Abstract*—Traffic volume forecasting is a key objective in Intelligent Transportation Systems (ITS) since its importance for both the general public and authorities in decision making, optimizing navigation strategies and avoid traffic congestions. Various research projects have been conducted for identifying the best approach to solve that issue. This paper proposes a comparison of statistical learning models, Vector Auto Regression, ARIMAX and a deep learning model, LSTM neural network, in the context of multivariate short-term (24 hours) time series forecasting using traffic volume, speed, and average waiting time, integrating weather attributes in Austin city, Texas. Models were evaluated using the rolling forecast origin method for three main feature sets generated utilizing feature selection. VAR model produced the best performance with an accuracy of 91.459% and proved to be used successfully in short term traffic forecasting in ITS applications.**

I. INTRODUCTION

The development of Intelligent Transportation Systems (ITS) is an important component in the design and implementation of smart cities. Active traffic management is one of the main objectives of the existing ITS and traffic volume forecasting is an important research area in that context [1]. Traffic volume forecasting can be utilized in optimizing navigation strategies and journey planning to identify the best routes and times avoiding traffic congestions [2]. Traffic volume forecasting can be beneficial for both the public and the authorities. Authorities can mitigate traffic congestions and accidents by identifying the anomalies of traffic volume data and optimize resource allocation to achieve greater efficiencies and profits. Also, there is a huge impact on long term policies, evidence-based planning, and procurement decisions.

Traffic is a complex structure, which is characterized by multiple variables such as volume and speed. At the same time traffic is significantly affected by other factors, especially weather. In the U.S., inclement weather like snow, ice, and fog cause delays of 544 million vehicle-hours annually, accounting for 23% of the total non-recurrent delay on highways [3]. Poor weather could have a significant impact on traffic pattern and driver behaviour, rendering these signal plans suboptimal or even unsafe for these conditions [4]. Snowfall has a significant effect on traffic safety by influencing vehicle performance, driver behaviour, and transportation infrastructure [3]. In high traffic areas, vehicle speed would be low, and the average time spent on a location would be high. It has a positive and negative effect on road traffic volume, respectively. Hence, integrating vehicle speed data, average time spent on a single location and weather data to predict traffic volume is useful. Nowadays, univariate time series forecasting models such as ARIMA and other classical statistical models are weak and inadequate for modelling non-stationary traffic flows in complex road settings [5]. Therefore, the need for multivariate analysis occurred to model non-stationary traffic flows in complex road settings.

Traffic-related data are collected as time-series, hence time series analysis and forecasting techniques such as ARIMA modelling can be employed for the traffic forecasting [6]. In that case, ARIMAX modelling which is an extended version of ARIMA enables to integrate other influencing factors as exogenous variables for multivariate traffic volume forecasting. Vector Autoregressive model (VAR) is also a multivariate time series model which can be used in scenarios where forecasting with multiple time-series that influence each other. Apart from the statistical models, deep learning techniques such as Long short-term Memory (LSTM) are also used for forecasting. LSTM uses feedback connections to keep track of previous steps of the time series utilizing one of the characteristics of time series, the dependency of the current time step to the previous steps. The main objective of this research is to compare ARIMAX, VAR and LSTM models in the context of traffic volume forecasting, integrating traffic information and weather data.

The paper is structured as follows: Section 2 explores the background of the techniques used and the existing literature, Section 3 describes the methodology, Section 4 discusses the experimental results and Section 5 concludes the paper.

II. BACKGROUND

Time series can be defined as a sequence of a metric that is recorded over regular time intervals. There are two things which Time-series make different from the regular regression problem, namely the time-dependency and the seasonality trends. In linear regression models, observations are independent but in time series data, the observations depend on time. Then, there can be seasonality trends, where variations specific to a repeated time frame. Mainly two methods are used for time series forecasting, univariate and multivariate. In univariate time-series forecasting, there are only two variables which are time and the parameter that is forecasted depending on time. Multivariate time-series forecasting contains multiple variables, which are time and multiple parameters that influence the forecasted parameter [7].

In time series analysis, there are some special characteristics needed to be considered, trend, seasonality, and stationarity [8]. The trend is the general tendency of the data to increase or decrease with time. Seasonality is a regular pattern of changes that repeats over S periods, where S defines the number of periods until the pattern repeats. And stationarity is a statistical property of a time series which are mean, variance and covariance do not change over time [9]. Several methods can be used to identify whether the time series is stationary or not. Visual test which identifies the series simply by looking at each plot, Augmented Dickey-Fuller (ADF) test which is used to determine the presence of unit root in the series and KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test are the ones commonly used [10]. In the ADF test, there is a null hypothesis which the time series is considered as non-stationary. So, if the p-value of the test is less than the significance level (0.05) then it rejects the null hypothesis and considers that the time series is stationary.

In statistical time series forecasting models, if the time series is non-stationary, it needs to be converted to stationary. One method is to use the differencing method to remove the varying mean by subtracting the previous value from the current value. The complexity of the time-series causes the number of times that the differencing is needed to remove the seasonality. Next one is seasonal differencing which is the difference between a value and a value with lag that is a multiple of seasonality factor

A. Learning models

1) ARIMAX

This section reviews the mathematical background of the compared time series techniques used and studied in this article. More specifically, the background knowledge of Autoregressive Integrated Moving Average (ARIMA) and the deep learning-based technique, Long Short-Term Memory (LSTM) is presented. Before delving into the ARIMAX model, it is necessary to have an explanation of the Auto-Regressive (AR), Moving Average (MA), ARIMA modelling techniques.

The AR model depends only on its lags. In an AR model with p lags, the value for the period t in a time series is calculated using Equation (1) where Y_t is the value of observation in period t, α is a constant, β is a numeric constant by which multiply the lagged variables and ε is residual values.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \beta_1 \quad (1)$$

MA model is one where the model depends only on the lagged forecast errors which are the errors of the AR models of the respective lags. In an MA model which uses up to q residuals, the value for the period t in a time series is calculated using Equation (2) where Y_t is the value of observation in period t, α is a constant, ϕ is numeric constants by which multiply the residuals, ε is Residual for the period t.

$$Y_t = \alpha + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + .. + \phi_q \varepsilon_{t-q} \quad (2)$$

Auto-Regressive Moving Average (ARMA) is the basic model for analysing a stationary time series. ARMA model is about merging AR and MA models. The AR model explains the momentum and mean reversion effects and the MA model captures the shock effects observed in the white noise terms. In an ARMA model which has p lags and q residuals, the value for the period t in a time series is calculated using Equation (3) where X_t is the predicted value in time t, ϕ numeric constants by which multiply the lagged variables, θ are numeric constants by which multiply the residuals.

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + \theta_1 \varepsilon_{t-1} + ... + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (3)$$

Auto-Regressive Integrated Moving Average (ARIMA) is a class of models that is based on its lags and the lagged forecast errors. Any non-seasonal time series that exhibits patterns and is not a random white noise can be modelled with ARIMA models. The ARIMA model is characterized by 3 terms. AR term where the number of lags to be used as predictors, MA term where the number of lagged forecast errors and a minimum number of times that the differencing is needed to make the series stationary. In an ARIMA model which has p lags and q residuals, the value for the period t in a time series is calculated using Equation (4) where Y_t is the value of observation in period t, α is a constant, β is a numeric constant by which multiply the lagged variables and Φ is numeric constants by which multiply the residuals.

$$Y_t = \alpha + \beta_1 Y_{t-1} + ... + \beta_p Y_{t-p} \varepsilon_t + \phi_1 \varepsilon_{t-1} + ... + \phi_q \varepsilon_{t-q} \quad (4)$$

ARIMAX is an extended version of the ARIMA model which utilizes multivariate time series forecasting using multiple time series which are provided as exogenous variables to forecast the dependent variable. Auto ARIMA function is commonly used to select the best model by automatically generating a set of optimal parameters by testing all possible combinations of (p,d,q) and returns the model with the lowest Akaike information criterion (AIC) and Bayesian Information Criterion (BIC) values [6] With the exogenous variables and the seasonality factor increased this process can be very expensive.

2) VAR

VAR is a multivariate time series model that can be used to forecast more than one variable collectively. It can be used in scenarios where multiple variables have a dependency on each other. In VAR modelling, each variable is modelled as a linear combination of past observations of itself and other variables. Therefore, it can be modelled as a system of equations, where each variable gets one equation that can be represented as vectors. Suppose we have a vector of time series data Y_t, then a VAR model with k variables and p lags can be expressed mathematically in Eq. (5) where, Y_t, β_0 and are k × 1 column vectors and $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$ are k × k matrices of coefficients.

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + \varepsilon_t \quad (5)$$

If a time series is not stationary, it is essential to differentiate the time series before training the model and invert the predicted values to get the real forecast by the number of times differentiated.

3) LSTM

LSTM is a special kind of recurrent neural network which makes use of sequential observations and learns from the prior stages to figure future patterns with additional features to memorize the sequence of information. The memorization of the prior trend of the data is done through a few gates alongside a memory line associated in an ordinary LSTM. Each LSTM is a set of cells where the data streams are captured and stored. LSTMs create a transport line that connects one module to another, carrying data from the past and keeping them for the present. Using gates in each cell, data can be disposed of, filtered, or added for the next cells. Those gates are based on a sigmoidal neural network layer which can enable the cells to optionally let data pass through or dispose of.

A sigmoid layer takes input in the range of zero and one, indicating the amount of data goes through in each cell. Estimation of zero value says nothing passes through the cell and one indicates that everything passes through the cell. There are three types of gates involved in each LSTM to control the state of each cell, Forget Gate, Memory Gate and Output Gate. Forget Gate outputs a number between 0 and 1 to say completely ignore this and completely keep all. Memory Gate chooses which new data need to be stored in the cell. Output Gate decides what will be yielded out of each cell.

B. Evaluation

1) Evaluation metrics

In time series forecasting, to evaluate the models, a comprehensive evaluation criterion is essential to measure the performance of the model. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) are commonly used metrics to reliably evaluate the performance of the models [6].

RMSE is a commonly used metric to evaluate the accuracy of predictions obtained by a model. It takes the residuals between actual and predicted values and compares prediction errors of different models for particular data. The main benefit of using RMSE is that it penalizes large errors and scales the scores in the same units as the forecast values. Let y_i is the actual value and $\hat{y}_i$ is the predicted value and for n observations, the RMSE for given n values is calculated using Equation (6).

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}} \tag{6}$$

MSE is the squared form of RMSE and is commonly used as a regression loss function [11]. MAE is used commonly in multivariate regression models [3]. Generally, MAE outperforms RMSE for measuring an average model accuracy [12]. Let y_i is predicted value and x_i is the actual value and for n data points MAE is calculated using Equation (7).

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{7}$$

MAPE is the average of absolute percentage errors which is popular in the industry since it is scale-independent and easy to interpret (Byrne, 2012). Let A_t and F_t denote the actual and forecast values at data point respectively and n is the number of data points. Then, MAPE is calculated using Equation (8).

$$M = \frac{1}{n}\sum_{t=1}^{n} \left|\frac{A_t - F_t}{A_t}\right| \tag{8}$$

2) Rolling forecast origin

The evaluation of all the time series forecasting models is based on Rolling Forecasting Origin technique [13]. This approach uses training sets, each one containing one more observation than the previous one, one-month look-ahead view of the data. There are three main variations of rolling forecast origin method namely, one-step forecasting without re-estimation, multi-step forecasting without re-estimation, and with re-estimation [14].

In one-step forecasting without re-estimation, the model estimates a single set of training data and then forecasts one-step on the remaining data sets. Multi-step forecasting is similar to one-step forecasting but forecasts multiple steps forward. Multi-step forecasting with re-estimation is an alternative approach where the model is trained at each iteration before each forecasting is performed. An intuitive and fundamental way to do the rolling forecast origin is to rebuild the model when each time a new observation is added.

C. Related work

Alghamdi, T. et al. [6] have proposed a solution to analyse and predict the traffic flow for short-term time-series observations, measured at an hourly basis, in a designated area of study in California, USA. Data were gathered as a short-term time series for three months and 13 attributes with different data types. They have considered several ARIMA models and did a comparison of models. Analysing a short time series of small-scale data requires a comprehensive checking tool to determine the suitable ARIMA parameters. Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) were used in their performance comparison. These measuring parameters are a reliable indicator of the model performance.

D. Pavlyuk et al. [15] done a systematic review of multivariate models in the context of application to short-term traffic flow forecasting. They have considered classical ARIMA specifications, which assume spatial independence of traffic flow data; multivariate ARMA models, which allow interrelationships between space points, Space-time models, which include a spatial structure as an integral model component, Error-correction models, oriented to the analysis of cointegrated time series and Relatively general multivariate space state models. Traffic information includes average speed values, aggregated by 5-minutes time frame from the data of traffic flows, publicly available from the Minnesota Department of Transportation.

S. Siami-Namini et al. [16] compared the accuracy of ARIMA and LSTM, as representative techniques when forecasting time series data. Researchers implemented and applied on a set of financial data and the results from the historical monthly financial time series from Jan 1985 to Aug 2018 from the Yahoo Finance Website. The ARIMA and LSTM algorithms are based on “Rolling Forecasting Origin” technique [13] It is important to use such a technique as there is a dependency in prior time steps. From three variations of rolling forecast origin method, they used 2 of them for the evaluation of ARIMA and LSTM. They used multi-step forecasting with re-estimation for the ARIMA model and multi-step forecasting without re-estimation for the LSTM. They showed that the LSTM-based algorithm improved the prediction by 85% on average compared to ARIMA in financial data.

Zheng Zhao et al. [17] proposed a novel traffic forecast model based on long short-term memory (LSTM) network. It was different from other forecasting models which included temporal-spatial correlation in a traffic system via a two-dimensional network which is composed of many memory units. Researchers did a comparison with other representative forecast models and their LSTM network got a better performance. Data was collected by the Beijing Traffic Management Bureau from over 500 observation stations with a frequency of 5 minutes. Dataset contained with traffic data such as vehicle volume, lane occupancy and average velocity in 25.11 million validated records and 0.81 million missing or invalid data. Mean absolute error (MAE), mean square error (MSE) and mean relative error (MRE) were used as the evaluation criteria for the traffic volume forecast. They selected three observation points where traffic volume is high, medium, and low. And they got MREs of 6.41, 6.05, and 6.21%, respectively. According to their forecast results, the proposed method is effective and reliable for traffic flow forecast in ITS.

III. METHODOLOGY

The proposed approach is designed with four main modules, namely pre-processing, feature selection, learning process, and evaluation. Fig. 1 shows the abstract view of the proposed system addressed in this paper. Three feature sets were generated after the feature selection process and then sent to the learning process where three learning algorithms were used. VAR and ARIMAX models were selected according to each feature set. ADF test and AIC based model selection was employed for the VAR model selection and Auto ARIMA was used for the ARIMAX model selection. After the learning process, the results were evaluated.

A. Materials and Pre-processing

The traffic count dataset used in the research was obtained from the open data portal of Austin, Texas. Traffic count data were collected from several GRIDSMART optical traffic detectors deployed by the city of Austin, at signalized intersections in the city. It is maintained by the arterial management division of the city of Austin transportation department. The dataset contains information regarding the traffic count and speed details in 15mins intervals, gathered from 56 optical traffic detectors. Weather dataset was purchased from openweathermap.org and it contains 15 weather parameters recorded in a one-hour interval.

For the analysis, the intersection with the most data points was selected. As mentioned before, traffic data were collected in 15 mins intervals and weather data were collected on an hourly basis. To combine traffic data with weather data, only one-hour data from the traffic dataset was selected, making the final dataset to have data of one intersection with an interval of one hour. Table I lists the attributes of the final dataset.

TABLE I. TRAFFIC AND WEATHER DATASET DETAILS

Column	Description
date	Date value formatted: Year-month-day Hour: Minute: Seconds
traffic_volume	Total traffic count of the intersection
speed_avg	The average speed of the vehicles that entered the intersection, including those that had started after the end of a red light
seconds_zone	The average amount of time that vehicles were in the measurement zone, including the time that vehicles stopped at the red lights
temp	Temperature
pressure	Atmospheric pressure (on the sea level), hPa
humidity	Humidity, %
wind_speed	Wind speed. Unit Default: meter/sec
wind_deg	Wind direction, degrees (meteorological)
rain_1h	Rain volume for the last hour, mm
rain_3h	Rain volume for the last 3 hours, mm
snow_1h	Snow volume for the last hour, mm (in liquid state)
snow_3h	Snow volume for the last 3 hours, mm (in liquid state)
clouds_all	Cloudiness, %

There are some missing values in rain_1h, rain_3h, snow_1h, and snow_3h columns due to non-occurrences of rainfall or snowfall. Therefore, those missing values were replaced by 0. There were 124 missing values in traffic data. For the imputation, interpolating by time was used for traffic_volume, speed_avg, and seconds_zone.

Deep learning models such as LSTM learn from the error on the training data via an optimization algorithm and update the randomly initialized weights. Unscaled input variables can slow down or destabilize the learning process Besides, scaling is essential when features are from different ranges. Normalizing data leads to faster convergence which speeds up the learning process. Hence, each feature is scaled to (0,1) before training the LSTM model.

B. Correlation analysis

The first step of the correlation analysis was to draw a heatmap of the correlation matrix. Correlation matrix displays the correlation coefficient which is the linear historical relationship between the variables of the dataset. With the use of correlation analysis heat map, it was found that average speed, Average amount of time that vehicles were in the measurement zone and the snow volume for the last 3 hours has the highest correlation with the traffic volume.

C. Learning models

First, all the features were categorized into traffic parameters and weather parameters. As traffic features, traffic volume, the average speed of the vehicles that entered the intersection and Average amount of time that vehicles were in the measurement zone were considered. As weather features, temperature, atmospheric pressure, humidity, wind speed, wind direction, rain volume for the last hour/3 hours, snow volume for the last hour/3 hours and cloudiness were considered. Then the study was done for three feature sets obtained from feature selection. As mentioned in the correlation analysis, snow volume for the last 3 hours was identified as the weather feature with the best correlation with the traffic volume and considered in a separate feature set with traffic features.

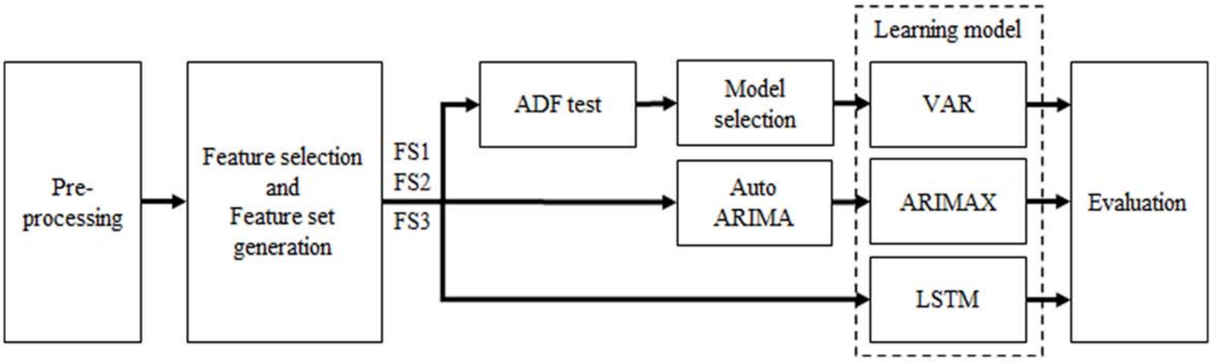

Fig. 1. Abstract view of the proposed methodology

So, the three feature sets used for the classification were, all traffic features (FS1), all traffic and weather features without feature selection (FS2), and traffic features with snowfall for the last 3 hours (FS3). All three learning models were trained and evaluated on these three feature sets.

1) ARMIAX

Auto ARIMA was used for the model selection which is provided by pmdarima library. The dataset was split with the ratio 0.2. Exogenous variables were provided for the traffic volume prediction. The model which gave the minimum AIC value was selected as the best.

For the feature set in which all traffic and weather features were used as exogenous variables, ARIMAX (4,0,3) was selected with the AIC value of 196593.141. For the feature set with only the traffic features, ARIMAX (2,0,3) was selected with an AIC value of 209674.526. ARIMAX (4,0,1) was selected with an AIC of 191865.005 for the feature set with the traffic features and snowfall for the last 3 hours. Then the models were evaluated using rolling forecast origin technique.

2) VAR

To identify whether the time series is stationary or not, the ADF test was used and a package provided by the python Statmodels called Adfuller was used in the implementation. Adfuller function has several parameters to be configured. AIC was used as the autolag, which is the parameter used for selecting the method for automatically determining the lag. The null hypothesis is the time series is not stationary. If the p-value is less than the significance level, the null hypothesis is rejected. Significance level was set to 0.05 and performed the test. Then it was found that all the time series are rejecting the null hypothesis. Hence, it was considered as all the time series are stationary.

Since all the time series were identified to be stationary, there was no need to differentiate the dataset when training the model and invert the predicted values to get the real forecast. Three VAR models were implemented for each feature set. The only hyperparameter that needed to be configured in the models was the order of the model which was selected using the AIC value where the order which produced the minimum AIC value was selected. Table II summarizes the implemented models.

TABLE II. ORDER SELECTION FOR MODELS

Feature set	Order	AIC
FS1	195	8.926026755301605
FS2	184	16.814587278934944
FS3	172	2.2530862418071367

3) LSTM

A neural network which consists of an input layer, two hidden LSTM layers and an output layer was used. A model was implemented for each feature set and the models were trained. The early stopping technique was used with a threshold of 10 steps and a validation split of 0.2. It stops the training process after 10 consecutive steps where the difference between validation error and the training error increases. This avoids the model overfitting rather than training for a fixed number of epochs.

IV. RESULTS AND EVALUATION

For the evaluation of the learning models, rolling forecast origin was used for two months of test data. ARIMAX and VAR models were evaluated using multi-step forecasting with re-estimation technique where the model is trained at each iteration before each forecasting is performed and LSTM models were evaluated using multi-step forecasting without re-estimation technique where the model estimates a single set of training data and then forecasts one-step on the remaining data sets.

In forecasting, the VAR models expect up to the lag order number of observations from the past data. The models were trained and forecasted in a rolling forecast origin method. In each iteration lag order number of data points starting from the dataset is kept for feeding as the input. Then the next 24 points were predicted and appended into the forecasted data.

The results of the VAR model for the three feature sets show that the best model is VAR (172) which used FS3. MAPE value of 0.08541 shows that it has 91.459% accuracy when forecasting next 24 hours of traffic volume while other two models that used FS2 (VAR (184)) and FS1 (VAR (196)) got accuracies of 87.478% and 87.704% respectively. VAR (172) model also produced the lowest RMSE and MAE values. Table III illustrates the summary of the evaluation of VAR models.

TABLE III. EVALUATION SUMMARY OF VAR MODELS

Metric	FS1 VAR(196)	FS2 VAR(184)	FS3 VAR(172)
MAPE	0.122959	0.12522	0.08541
MSE	3016.5147	3682.1441	2733.4961
MAE	42.63417	46.3753	41.91831
RMSE	54.9228	60.68067	52.2828

Table IV summarizes the results of the ARIMAX model for the three feature sets. Overall results show that all ARIMAX models are not performing well in traffic volume forecasting. ARIMAX (4,0,3) which used FS2 produced the best performance. Nonetheless, it also got a lower accuracy of 48.49% in forecasting for the next 24 hours, according to the MAPE value of 0.5151. All models got higher values for other error metrics compared to other learning models. ARIMAX (4, 0, 3) achieved the lowest RMSE value of 240.86979 which is not acceptable for forecasting comparing to the other models.

TABLE IV. EVALUATION SUMMARY OF ARIMAX MODELS

Metric	FS1 ARIMAX(2,0,3)	FS2 ARIMAX(4,0,3)	FS3 ARIMAX(4,0,1)
MAPE	1.0252660	0.5151	0.928552
MSE	433873.41008	58018.2558	117565.05293
MAE	583.685988	208.14596	306.40887
RMSE	658.690679	240.86979	342.877606

Table V summarizes the results of the LSTM models. According to that, the model which used FS2 produced the lowest accuracy and along with the highest errors. The model that used FS1 produced a higher accuracy compared to the FS2. Furthermore, it was observed that the performance is improved with FS3 which produced the highest accuracy of 69.97% when forecasting for 24 hours. It also produced the lowest error values for other error metrics.

TABLE V. EVALUATION SUMMARY OF LSTMS

Metric	FS1	FS2	FS3
MAPE	0.341184	1.751867	0.303467
MSE	29683.18442	243568.77422	27376.61385
MAE	136.063689	419.6359265	124.844981
RMSE	172.2881	493.526873	165.458798

V. DISCUSSION

Traffic volume forecasting is a key feature in Intelligent Transportation Systems and there are various ongoing research projects for identifying the best approach for the solution. In this paper, we have proposed an approach for forecasting traffic volume with speed, average waiting time, and weather variables using statistical models, ARIMAX and VAR and a deep learning model, LSTM. The results produced by each learning model were compared to select the best model to deploy in an ITS.

In traffic forecasting, identifying the distinguishable features between normal and abnormal traffic patterns is important. It is not practical to identify these patterns in small scale datasets with fewer features and data points [3,4]. Nonetheless, in this study, we used a dataset with a considerable number of features and data points for nearly 3 years. Besides, the learning models should be trained on large datasets to obtain acceptable performance in the context time series forecasting [5]. In this research, we have compared several statistical and deep learning models, namely ARIMAX, VAR and LSTM in multivariate time series forecasting for traffic volume data using traffic-related variables such as speed, the average time that vehicles were in the intersection, and weather variables. Before the learning process, a correlation analysis was performed to identify the correlation between the weather and traffic-related variables with the traffic volume. According to that, average speed and average time of vehicles produced the highest correlation with the traffic volume and among the weather variables, snowfall for the last 3 hours achieved the highest correlation.

In the learning algorithms used, several parameters need to be configured to select the best model for the features used. For ARIMAX, we used Auto ARIMA which is an efficient way to do a grid search by parallelizing the processes and automatically discover the optimal order for the model. For VAR, a grid search was performed to select the appropriate lag which produced the minimum AIC value. We employed the same LSTM network for all three feature sets. During the model training, the validation loss and the was monitored by early stopping call back function which halts the training when there is an increment observed in loss values. The number of epochs for the training to be continued after the first halt was set to 10.

Rolling forecast origin method is important in the evaluation of time series forecasting because of the sequential dependency between the values. In this research, we employed two variations of the rolling forecast origin method for the evaluation. Multi-step forecasting with re-estimation was used for ARIMAX and VAR models and multi-step forecasting without re-estimation was used for LSTM models.

According to the overall results, the performances of the ARIMAX models were the lowest and the VAR models performed the best. VAR and LSTM algorithms produced lower results with FS2 than with FS1. This can be caused by the high dimensionality of FS2 in which all the traffic and weather features were used. Nonetheless, since feature selection was utilized in FS3, it produced the best results for VAR and LSTM models. VAR models produced the best results and since VAR enables forecasting all the features as well as the traffic volume is an added advantage. Nonetheless, if the VAR and ARIMAX models were used in long term forecasting, it would not perform better than LSTM.

As for future work, the ARIMAX model can be extended to the Seasonal ARIMAX (SARIMAX) to improve the forecasting, utilizing the seasonality of the traffic data. At the same time, LSTM models can also be improved by hyperparameter tuning such as the number of layers, learning rate, optimizers, etc.

Fig. 2 shows the graphical representation of the comparison of models with RMSE values. VAR models got the lowest RMSE values while ARIMAX model got the highest. Out of all feature selections, FS3 of the VAR got the lowest RMSE of 52.2828 which is the selected model to deploy as the forecasting model.

VI. CONCLUSION

This paper proposes an approach for 24 hours short term traffic volume forecasting, comparing the learning algorithms, VAR, ARIMAX and LSTM. After utilizing, a correlation-based feature selection and AIC based model selection, VAR model produced the best performance of 52.282 RMSE with an accuracy of 91.46%. These promising results illustrate that VAR can be used as a successful learning model for short term traffic forecasting in ITS applications.

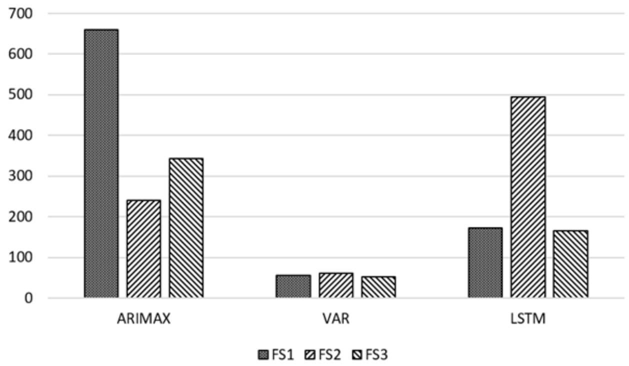

Fig. 2. Comparison RMSE of the models

ACKNOWLEDGEMENT

This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Higher Education, Sri Lanka funded by the World Bank. Also, we acknowledge the support received from the Senate Research Committee Grant, University of Moratuwa, Sri Lanka

REFERENCES

[1] K. Iqbal, M. A. Khan, S. Abbas, Z. Hasan, and A. Fatima, "Intelligent transportation system (ITS) for smart-cities using Mamdani Fuzzy Inference System," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 2, pp. 94–105, 2018.

[2] A. Koesdwiady, R. Soua, and F. Karray, "Improving Traffic Flow Prediction with Weather Information in Connected Cars: A Deep Learning Approach," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9508–9517, Dec. 2016.

[3] S.-M. Chin, O. Franzese, D. Greene, H.-L. Hwang, and R. Gibson, *Temporary Losses of Highway Capacity and Impacts on Performance: Phase 2*. 2002.

[4] M. Agarwal, T. Maze, and R. Souleyrette, "Impact of Weather on Urban Freeway Traffic Low Characteristics and Facility Capacity," 2005.

[5] E. Vlahogianni, J. Golias, and M. Karlaftis, "Short-term traffic forecasting: Overview of objectives and methods," *Transp. Rev.*, vol. 24, pp. 533–557, 2004.

[6] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting traffic congestion using ARIMA modeling," in *2019 15th International Wireless Communications and Mobile Computing Conference, IWCMC 2019*, 2019, pp. 1227–1232.

[7] R. Chuentawat and Y. Kan-ngan, "The Comparison of PM2.5 forecasting methods in the form of multivariate and univariate time series based on Support Vector Machine and Genetic Algorithm," 2018, pp. 572–575.

[8] F. Martínez, M. P. Frías, M. D. Pérez-Godoy, and A. J. Rivera, "Dealing with seasonality by narrowing the training set in time series forecasting with kNN," *Expert Syst. Appl.*, vol. 103, pp. 38–48, Aug. 2018.

[9] A. Adib, M. M. K. Kalaee, M. M. Shoushtari, and K. Khalili, "Using of gene expression programming and climatic data for forecasting flow discharge by considering trend, normality, and stationarity analysis," *Arab. J. Geosci.*, vol. 10, no. 9, pp. 1–14, May 2017.

[10] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?," *J. Econom.*, vol. 54, no. 1–3, pp. 159–178, 1992.

[11] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Berlin, Germany: Springer, 2001.

[12] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.

[13] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, 2018.

[14] R. J. Hyndman, "Variations on rolling forecasts," 15-Jul-2014. [Online]. Available: https://robjhyndman.com/hyndsight/rolling-forecasts/. [Accessed: 05-Nov-2020].

[15] D. Pavlyuk, "Short-term Traffic Forecasting Using Multivariate Autoregressive Models," in *Procedia Engineering*, 2017, vol. 178, pp. 57–66.

[16] S. Siami-Namini, N. Tavakoli, and A. S. Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," pp. 1394–1401, Jan. 2019.

[17] Z. Zhao, W. Chen, X. Wu, P. C. V. Chen, and J. Liu, "LSTM network: A deep learning approach for short-term traffic forecast," *IET Image Process.*, vol. 11, no. 1, pp. 68–75, Jan. 2017.

[18] M. Gjoreski *et al.*, "Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors," *Inf. Fusion*, vol. 62, pp. 47–62, Oct. 2020.

[19] [H. Xia, Y. Qiao, J. Jian, and Y. Chang, "Using smart phone sensors to detect transportation modes," *Sensors (Switzerland)*, vol. 14, no. 11, pp. 20843–20865, Nov. 2014.

[20] A. Jahangiri and H. A. Rakha, "Applying Machine Learning Techniques to Transportation Mode Recognition Using Mobile Phone Sensor Data," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 5, pp. 2406–2417, Oct. 2015.

[21] M. A. Shafique and E. Hato, "Travel mode detection with varying smartphone data collection frequencies," *Sensors (Switzerland)*, vol. 16, no. 5, May 2016.

[22] P. Castrogiovanni, E. Fadda, G. Perboli, and A. Rizzo, "Smartphone Data Classification Technique for Detecting the Usage of Public or Private Transportation Modes," *IEEE Access*, vol. 8, pp. 58377–58391, 2020.