2022 8th International Conference on Control, Decision and Information Technologies
CoDIT'22 | Istanbul, Turkey / May 17-20, 2022

Deep Learning-Based Prediction Models for Freeway Traffic Flow under Non-Recurrent Events

Fahad Aljuaydi
EECMS, Curtin University
Perth WA, Australia
f.aljuaydi@postgrad.curtin.edu.au

Benchawan Wiwatanapataphee
EECMS, Curtin University
Perth WA, Australia
b.wiwatanapataphee@curtin.edu.au

Yong Hong Wu
EECMS, Curtin University
Perth WA, Australia
Y.Wu@curtin.edu.au

2022 8th International Conference on Control, Decision and Information Technologies (CoDIT) | 978-1-6654-9607-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/CODIT55151.2022.9803892

***Abstract*—This paper concerns predictions of freeway traffic flow under non-recurrent events using multivariate machine learning models, including the multilayer perceptron network and the one-dimensional CNN long short-term memory network. The machine learning architectures and loss functions for training neural networks are presented. The study region is a portion of the Kwinana Freeway northbound in Perth, Western Australia. The study dataset, obtained by matching the timestamp of all available data, has various features, including traffic volume (flow rate), speed, density and road incident. Using the root mean squared error and mean absolute error, results from the two learning models are compared to the baseline model to determine the suitable model for traffic prediction under non-recurrent events.**

***Index Terms*—Traffic prediction, Non-recurrent events, MultiLayer Perceptron, Convolutional Neural Network, Long short-term memory.**

I. Introduction

Predicting future traffic flow conditions is essential in managing traffic operations and providing advanced traveller information. Uncertainties in traffic conditions result in high travel time variability, traffic congestion, delays, and greenhouse gas emissions. Therefore, traffic flow prediction has been one of the critical research areas in transport engineering over the last two decades.

In literature, prediction models comprise statistical models and artificial neural networks (ANNs) using two types of data whose order matters, including sequential and time-series data. There are two types of statistical models, parametric and non-parametric. Various parametric models based on the Auto-regressive Integrated Moving Average model (ARIMA) have been proposed for short-term traffic prediction [1]–[6]. Machine learning models may be classified into four classes: supervised, semi-supervised, unsupervised, and reinforcement. The supervised learning algorithm involves classification, regression and forecasting. It analyses available data to determine the correlations and relationships, concludes a known dataset, and identify data patterns. Examples of popular supervised learning algorithms are Naïve Bayes Classifier Algorithm, Support Vector Machine Algorithm, Linear Regression, Logistic Regression, Random Forests, and Nearest Neighbours. For reinforcement learning, different options and possibilities are explored using a given set of actions, parameters and end values (hyperparameters) to determine the optimal one. Similarly, the semi-supervised learning model combines the data with essential information and other data lacking. Unlike the supervised and the semi-supervised algorithms, the unsupervised learning algorithm does not determine the data structure but identifies the data pattern using clustering and dimension reduction techniques. The standard and popular unsupervised learning algorithm is the K Means Clustering Algorithm.

For time-series data, a recurrent neural network (RNN), a type of artificial neural network (ANN), is recommended [7]. Most RNN learning algorithms have been refined to improve their accuracy for real-world problems. In general, various RNN models have been designed based on machine learning approaches, including the Multilayer Perceptron (MLP), the Convolutional Neural Networks (CNN) and the Long Short-term Memory (LSTM) networks. The MLP is a simple case of the feed-forward ANN, comprising input, hidden, and output layers. The method allows for more than one hidden layer, referred to as the depth of a neural network. Like the MLP, CNN also comprises an input and output layer between which lies a series of hidden layers that can be convoluted, pooled or fully connected for feature identification. These convolution layers use filters to record characteristics of multi-dimensional data. A pooling layer reduces dimension in the data and can thus provide an abstract representation of the data. A Long Short-term Memory (LSTM) network is a particular case of RNN, which utilizes hidden components such as memory cells. LSTM is beneficial in studying long time-series data and making forecasts considering autocorrelations [8].

Many machine learning models have been proposed for time-series forecasting problems such as stock markets [9], wind speed [10], solar radiation [11], and more recently in forecasting competitions [12]. A deep reinforcement learning model has been proposed for traffic prediction [13]. To deal with variability and non-linearity of traffic flow, many researchers have combined machine learning models with other data mining methods such as the k-nearest neighbour algorithm [14], [15] and Bayesian model averaging algorithms [16]–[18] for traffic flow prediction. The critical problems in machine learning models are overfitting of the training set and extended training time. For preventing overfitting model, training with more data, using early stopping to halt the training of the network at the right time and using a suitable number of

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on March 02,2023 at 03:00:44 UTC from IEEE Xplore. Restrictions apply.

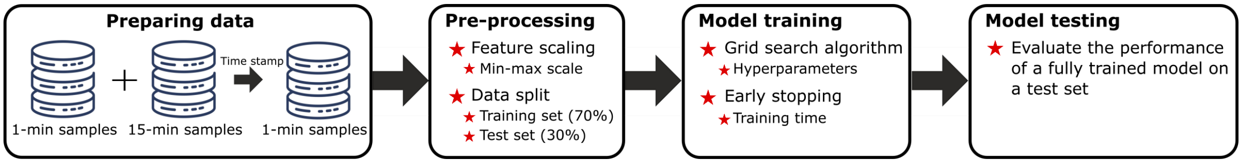

Fig. 1. Machine learning workflow.

epochs are required. Recently, research has been focused on predicting multiple outputs with multiple inputs using the data set with various classes (features). As unexpected accidents on the freeway significantly affect road capacity, road incident detection is required for traffic operation. Recently, research has focused on the development of a deep learning algorithm to detect vehicle movement and traffic accidents in tunnels [19] from time series of images from CCTV. A little attempt has been made to predict traffic flow under non-recurrent events using a multivariate learning model with multiple inputs. Therefore, the deployment of machine learning models is a challenging task.

This paper predicts freeway traffic flow under non-recurrent events using multivariate learning models based on the multilayer perceptron and the one-dimensional CNN long short-term memory network. Using the root mean squared error, results obtained from these models are compared to the baseline model, which is the average traffic variables from the training set (Fig. 10) to find the suitable model for the freeway traffic predictions under non-recurrent events. A 1-min data set is generated from the matching timestamp of the 1-min traffic flow data set and the 15-min incident data set. As shown in Fig. 2, the study region, the red curve, is a road segment of the Kwinana Smart Freeway from Cranford Avenue to Manning Road, which experiences severe congestion and stop-start conditions due to traffic incidents.

The work is structured as follows. Section II concerns Methodology. Section III describes the study area and available data sets. Results and discussion are given in section IV. The conclusion is given in section V.

II. Methodology

This section concerns building multivariate machine learning (ML) models based on either the MLP or the one-dimensional CNN long short-term memory neural networks to predict traffic flow rate, speed and density under non-recurrent events. As shown in Fig. 1, the machine learning workflow comprises preparing the data, pre-processing, model training, and model testing.

A. Preparing data and pre-processing

Constructing a predictive learning model is a complex process involving getting the correct data, cleaning it and creating useful data features, testing different learning algorithms, and validating the model. Failure in preparing the relevant data will result in predictive failure analysis.

In this study, the time series data has n observations (samples) and four features (classes) denoted by

$$\mathbf{X} = (X^1, X^2, X^3, X^4)_{i=1}^n,$$

where X^1, X^2 and X^3 denote traffic flow rate, speed and density, and X^4 is a traffic incident feature. As time-series traffic flow data and incident data have different scales, feature scaling is required. Here, we apply the min-max normalization linearly transforms x_i^c to ξ_i^c, $0 \le \xi \le 1$:

$$\xi_i^c = \frac{x_i^c - \min_c}{\max_c - \min_c}, \quad c = 1, ..., 4. \tag{1}$$

where $\min_c$ and $\max_c$ are the minimum and maximum values in X^c, where X^c is the set of observed values of x_i^c.

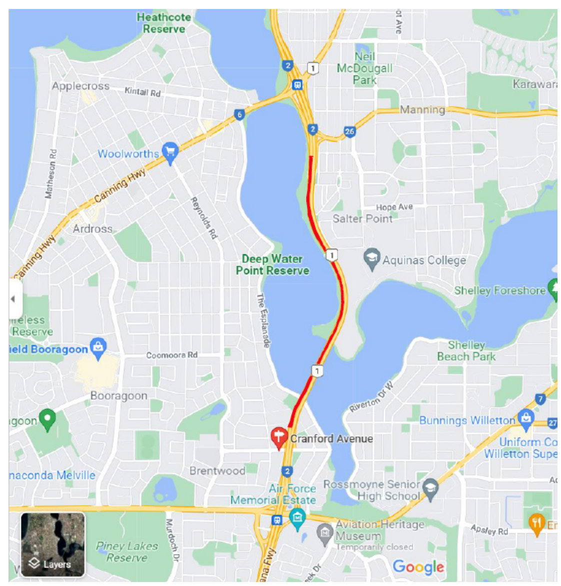

Fig. 2. Study region (red curve) between the Cranford Avenue on-ramp and the Canning Highway northbound off-ramp.

The study data with 429,120 samples are split into the training set (70%) and the test set (30%).

B. Model training and testing

Grid searching is used to find the optimal set of hyperparameters for each learning model: the MLP model (learning rate, drop out rate and dense units) and the 1-D CNN LSTM model (filters and kernel size, learning rate. and drop out rate). Early stopping is also applied for the right time to train the model. The models were fully trained using the Adam optimizer [20], where the number of hidden layers was chosen empirically to improve the overall performance.

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on March 02,2023 at 03:00:44 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RMSE AND MAE FOR EACH CASE OF THE TRAINING AND THE TEST SETS.

Parameter	Baseline				MLP				CNN-LSTM			
	RMSE		MAE		RMSE		MAE		RMSE		MAE	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Speed	10.2598	9.8137	5.4993	6.1823	5.2311	5.7965	2.753	2.6917	5.891	6.7006	2.0547	2.0801
Flow Rate	11.0281	11.7503	6.5284	6.2812	3.5864	3.5861	4.1855	4.639	2.8134	2.8403	4.5135	5.2321
Density	24.1516	24.419	11.144	10.956	7.3215	6.9089	5.5157	5.4881	6.8161	6.3472	3.6161	3.6738

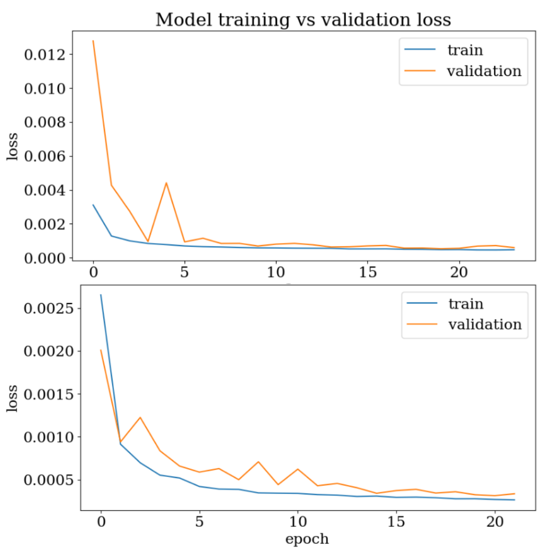

Fig. 3. Squared error loss, $L_2 = (\mathbf{y} - \hat{\mathbf{y}})^2$, for the training and the test datasets in each ML model: the MLP model (top) and the 1-D CNN LSTM model (bottom).

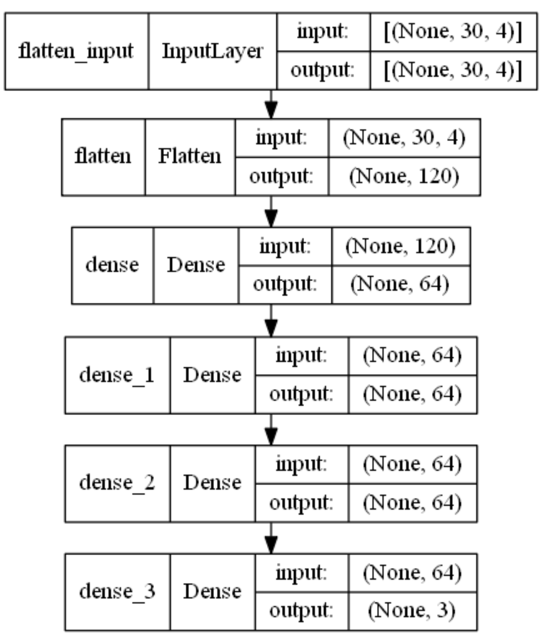

Fig. 4. MLP architecture.

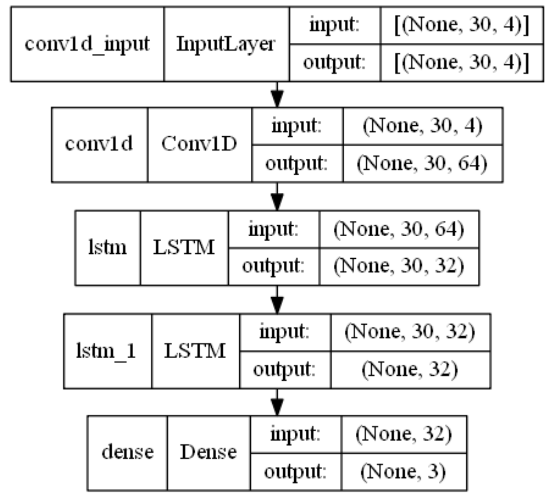

Fig. 5. 1D-CNN LSTM architecture.

For model testing, the goodness of each learning model is assessed in terms of accuracy and efficiency by the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE). The RMSE and MAE, determining the deviation around the mean value predicted by the model, are calculated by

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

and

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{3}$$

respectively. Lower values of the RMSE and the MAE signify a better model fit as the level of deviation is low. As shown in Table. I, the root means squared errors (RMSEs) and the mean absolute errors (MAEs) of the training dataset are larger than those of the test dataset, which are reasonable as the ML model is better at predicting the known dataset that it has learned than the unknown (the test set).

The MLP and 1-D CNN LSTM models with a batch size of 30 are fit over 20 epochs, as shown in Fig. 3. The model architectures of the MLP and the 1-D CNN LSTM are thus obtained and are shown in Fig. 4 and Fig. 5, respectively.

III. STUDY AREA AND STUDY DATASET

The study area is a major arterial, link 9, of the Kwinana Freeway northbound in Perth, Western Australia. It is between

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on March 02,2023 at 03:00:44 UTC from IEEE Xplore. Restrictions apply.

the Cranford Avenue on-ramp and the Canning Highway northbound off-ramp as shown in Fig. 2. The 1-min study data is generated by matching timestamp between 1 January and 25 October 2018 of the 15-min road incident data with that of the 1-min traffic flow data of flow rate, speed and density.

A. *Non-recurrent event datasets*

The data of road incidents affecting traffic capacity during the study period come from the following two sources

- MRWA
  - Weather flood/fog hazard
  - Heavy/moderate/stand still Jam
  - Major and minor accidents
  - Road closure
- Web emergency operation centre (WebEOC)
  - Break down/tow away
  - Road crash
  - Debris/trees/lost loads
  - Vehicle fire
  - Animal/livestock
  - Pothole/road surface damage

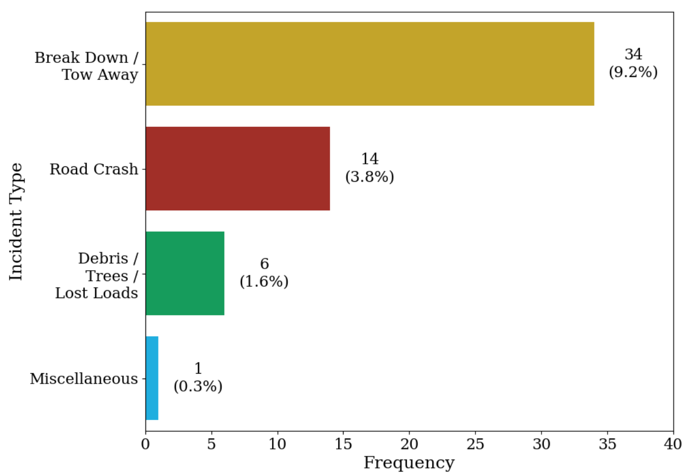

Fig. 6. Frequency of road incidents on the study region.

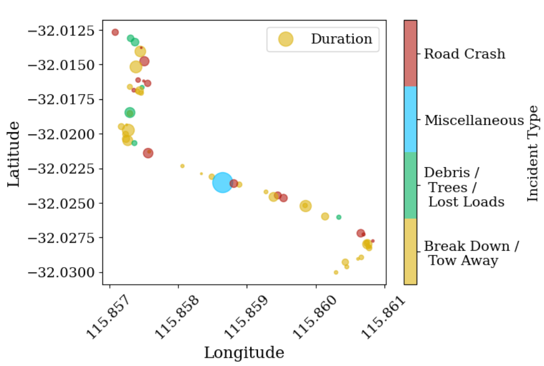

Fig. 7. Locations in latitude and longitude coordinates of traffic incidents and incident duration on the study region.

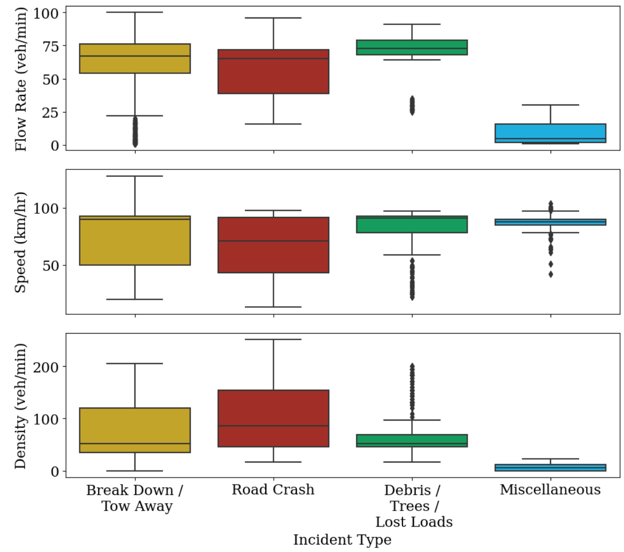

Fig. 8. Box plot showing distribution and skewness of traffic variables associated with road incidents on the study region.

Breakdown, road crashes, and road debris are common incident types in this link. Fig. 6 shows the study region's frequency of road traffic incidents. Locations in latitude and longitude coordinates of traffic incidents are shown in Fig. 7, in which a bigger circle size indicates a longer incident duration. As shown in Fig. 8, the box plot shows the distribution and skewness of traffic data associated with road incident types. It is noted that traffic capacity is associated with road crashes and breakdowns.

B. *Traffic data*

Traffic data of flow rate (volume), speed and density is provided by Main Roads Western Australia (MRWA) between 1 January and 25 October 2018. It is 1-min dataset with 429,120 observations and 3 features including traffic flow rate (veh/min), speed (km/hr) and density (veh/km).

Fig. 9 shows fundamental traffic flow diagrams on the study road, i.e., the relationship of the speed with the flow rate, the relationship of the flow rate with the density, and the relationship of the speed with the density, respectively.

C. *The study dataset*

Road incident message data were first converted into Boolean type data, being one if a road incident exists and zero otherwise. The study data is then obtained by matching the timestamp of the traffic data and the boolean incident data. Fig. 11 shows the observed traffic flow under road crash on 18 July 2018 between 12:51 and 14:24. It indicates that road crash reduces traffic capacity.

IV. Experimental Results

Traffic variables under a long-period road crash on Wednesday 18 July 2018, between 12:51 - 14:24, are predicted using the proposed learning models, and are compared with the

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on March 02,2023 at 03:00:44 UTC from IEEE Xplore. Restrictions apply.

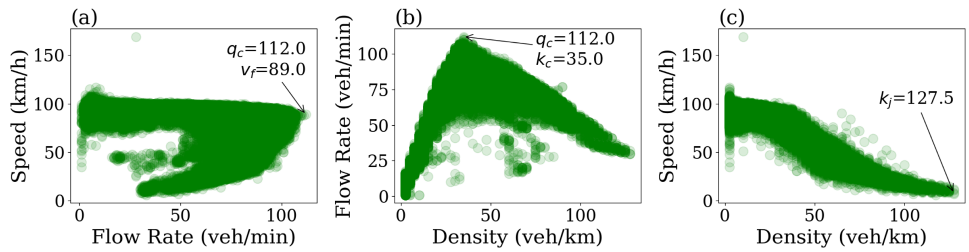

Fig. 9. Fundamental diagrams showing relationship of traffic variables: (a) speed (km/hr) and volume (veh/min); (b) flow rate (veh/min) and density (veh/km); (c) speed (km/hr) and density (veh/km).

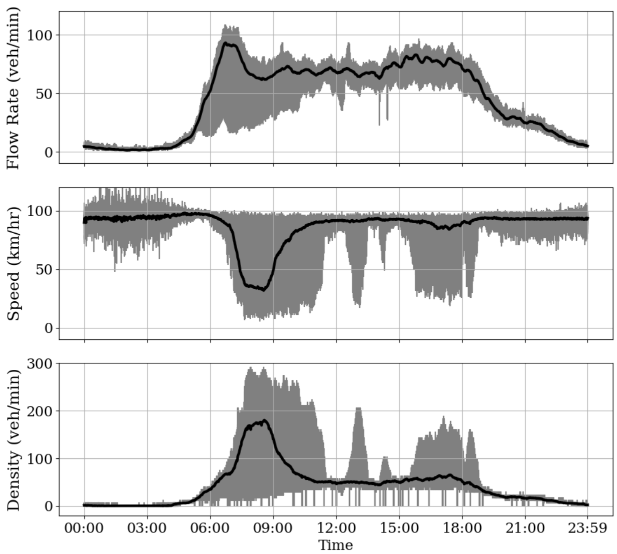

Fig. 10. Average traffic profiles of flow rate, speed and density during the study period.

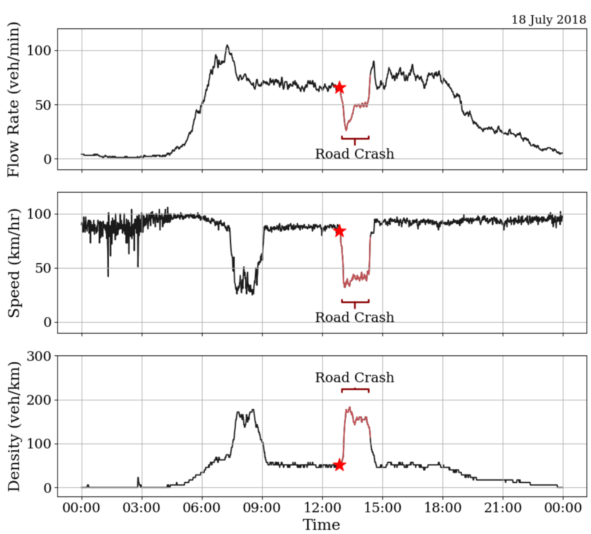

Fig. 11. Observed traffic flow rate (top), speed (middle) and density (bottom) with road crash during the prediction period.

results of the baseline model which is the average traffic variables from the training set in Fig. 12.

Fig. 13 and Fig. 14 present predictions of flow rate, speed and density obtained from the MLP and the 1D-CNN LSTM networks, respectively. It is noted that

- Predictions of traffic flow rate, speed and density obtained from both models are reasonable. Compared with the observed data, the proposed machine learning models give the general pattern and trend of the traffic under non-recurrent events. However, predicted traffic flow rates are not exactly equal to the observed data, especially around the incident.
- RMSEs in estimating the flow rate, speed and density in each case of the training and the test sets obtained from the 1-D CNN LSTM model are much smaller than those obtained from the baseline and the MLP models.

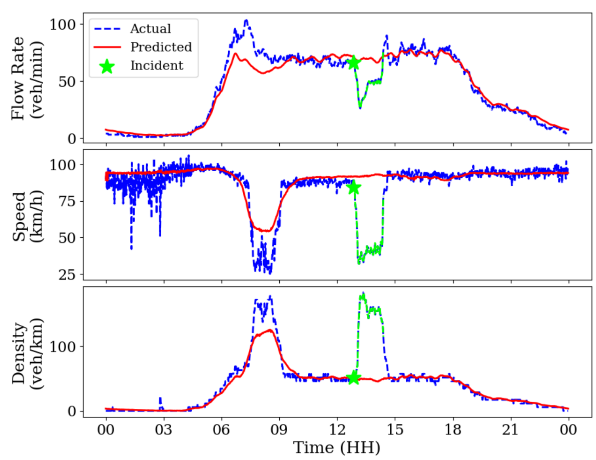

Fig. 12. Baseline predictions of traffic flow rate (veh/min/lane), speed (km/hr) and density (veh/km/lane) with a road incident ($*$) and its impact on traffic parameters (a solid green line).

V. Conclusions

Two multivariate ML models were designed based on the multilayer perceptron and the one-dimensional CNN long short-term memory networks. Traffic flow predictions under non-recurrent events using the proposed ML models with four inputs and three outputs have been made. Both ML models can accurately predict traffic flow under non-recurrent circumstances compared with the baseline model. The RMSE

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on March 02,2023 at 03:00:44 UTC from IEEE Xplore. Restrictions apply.

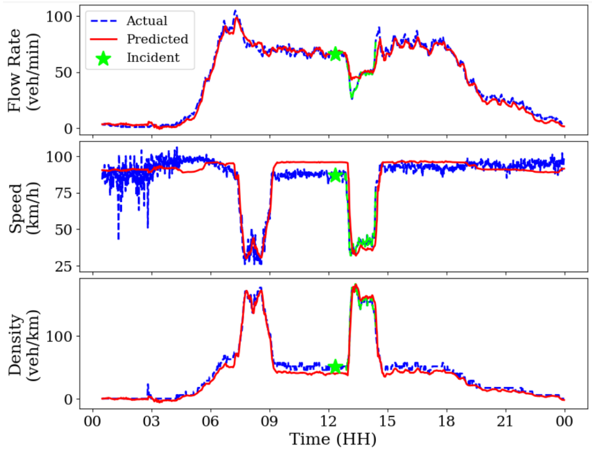

Fig. 13. MLP predictions of traffic flow rate (veh/min/lane), speed (km/hr) and density (veh/km/lane) with a road incident (∗) and its impact on traffic parameters (a solid green line).

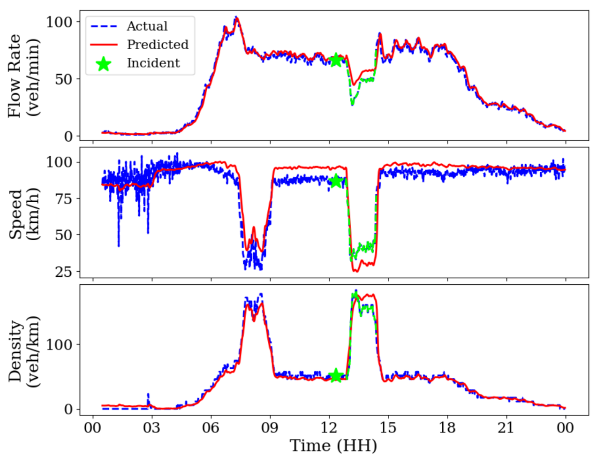

Fig. 14. 1D-CNN LSTMs predictions of traffic flow rate (veh/min/lane), speed (km/hr) and density (veh/km/lane) with a road incident (∗) and its impact on traffic parameters (a solid green line).

and MAE of the proposed ML models are much lower than those obtained from the baseline model. Compared to the MLP model, the 1-D CNN LSTM model gives better predictions of all traffic parameters.

As traffic flow is disrupted and delayed around on-ramp and off-ramp areas, future research will consider the non-recurrent event and congestion due to lane changes to develop a deep time-delay neural network for predicting traffic flow characteristics. A deep traffic congestion model will also be developed for predicting traffic congestion. This congestion model will provide information on congestion propagation on the study road to improve traffic prediction.

VI. Acknowledgments

This research is supported by the Australian Research Council through the linkage project funding scheme (project LP170100341). The authors would like to thank the Main Roads Western Australia (MRWA) for providing traffic data for this research, and also would like to acknowledge the Road and Maritime Service NSW and the Australia Sustainable Environment National Research Centre for their support of the project.

References

[1] L. Li, X. Su, Y. Zhang, Y. Lin, and Z. Li, “Trend modeling for traffic time series analysis: An integrated study,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 3430–3439, 2015.

[2] M. M. Hamed, H. R. Al-Masaeid, and Z. M. B. Said, “Short-term prediction of traffic volume in urban arterials,” *Journal of Transportation Engineering*, vol. 121, no. 3, pp. 249–254, 1995.

[3] 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.

[4] 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.

[5] 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.

[6] B. Ghosh, B. Basu, and M. O'Mahony, “Multivariate short-term traffic flow forecasting using time-series analysis,” *IEEE transactions on intelligent transportation systems*, vol. 10, no. 2, pp. 246–254, 2009.

[7] J. L. Elman, “Finding structure in time,” *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.

[8] Z. Liu, Z. Li, K. Wu, and M. Li, “Urban traffic prediction from mobility data using deep learning,” *Ieee network*, vol. 32, no. 4, pp. 40–46, 2018.

[9] H. Y. Kim and C. H. Won, “Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models,” *Expert Systems with Applications*, vol. 103, pp. 25–37, 2018.

[10] C. Yu, Y. Li, Y. Bao, H. Tang, and G. Zhai, “A novel framework for wind speed prediction based on recurrent neural networks and support vector machine,” *Energy Conversion and Management*, vol. 178, pp. 137–145, 2018.

[11] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, “Lasso and lstm integrated temporal model for short-term solar intensity forecasting,” *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2933–2944, 2018.

[12] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, “The m4 competition: Results, findings, conclusion and way forward,” *International Journal of Forecasting*, vol. 34, no. 4, pp. 802–808, 2018.

[13] W. Hu, Y. Liu, L. Li, and S. Xin, “The short-term traffic flow prediction based on neural network,” in *2010 2nd International Conference on Future Computer and Communication*, vol. 1, pp. V1–293, IEEE, 2010.

[14] H. Chang, Y. Lee, B. Yoon, and S. Baek, “Dynamic near-term traffic flow prediction: system-oriented approach based on past experiences,” *IET intelligent transport systems*, vol. 6, no. 3, pp. 292–305, 2012.

[15] L. Lin, Y. Li, and A. Sadek, “A k nearest neighbor based local linear wavelet neural network model for on-line short-term traffic volume prediction,” *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 2066–2077, 2013.

[16] W. Zheng, D.-H. Lee, and Q. Shi, “Short-term freeway traffic flow prediction: Bayesian combined neural network approach,” *Journal of transportation engineering*, vol. 132, no. 2, pp. 114–121, 2006.

[17] S. Qiu, J. R., and N. N., “Multivariate bayesian structural time series model,” *Journal of Machine Learning Research*, vol. 19(1), p. 2744–2776, 2018.

[18] J. R., S. Qiu, and N. N., “Predicting a stock portfolio with the multivariate bayesian structural time series model: Do news or emotions matter?,” *International Journal of Artificial Intelligence*, vol. 17(2), pp. 81–104, 2019.

[19] K. B. Lee and H. S. Shin, “An application of a deep learning algorithm for automatic detection of unexpected accidents under bad cctv monitoring conditions in tunnels,” *IEE Xplore: 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)*, pp. 7–11, 2019.

[20] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

Authorized licensed use limited to: CURTIN UNIVERSITY. Downloaded on March 02,2023 at 03:00:44 UTC from IEEE Xplore. Restrictions apply.