ARTICLE

**Deep Learning Models for Time-series Traffic Volume Data**

**Jieun Lee†, *222DBG04 and* Doyeon Lee††, *222DBG02***

† The author is in the Ewha University Data warehousing lab.

†† The author is in the Ewha University AI applications lab.

**SUMMARY**

This study analyzes the impact of weather, temperature, and holidays on traffic volume using data from the Hourly Minneapolis-St Paul, MN traffic dataset. It employs deep learning models—LSTM, GRU, and CNN-LSTM—to predict traffic volume, with a focus on comparing their effectiveness. Extensive data preprocessing, including variable transformation and encoding, was conducted to enhance model accuracy. The results reveal that the CNN-LSTM model outperforms others, especially when incorporating additional variables like weather and holidays. The study concludes with the potential for these models to improve traffic predictions on general roads, beyond highway scenarios.

***Keywords:*** *Traffic Flow Prediction, Deep Learning, Long Short-Term Memory Networks(LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Network*

**1. Introduction**

Traffic volume increases or decreases significantly under the influence of certain factors, and such rapid changes in traffic volume can increase the risk of road traffic or lead to traffic accidents. In order to prevent this, it is important to predict traffic volume appropriately.

To create a traffic volume prediction model, we used not only historical traffic volume data but also the provided weather environmental data. Furthermore, an evaluative comparison of deep learning models was conducted to improve model effectiveness and performance metrics.

Based on our prior research, we employed Long Short-term Memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network-Long Short-term Memory (CNN-LSTM) models and proceeded to compare their performance. When making predictions using a deep learning model, we used three models suitable for predicting time series: LSTM, GRU model, and CNN-LSTM model. After examining all three models, we determined which type of model is more suitable for predicting traffic volume.

As a result of model analysis, the performance increased and errors were further reduced when we used the past traffic volume and other variables like holidays, and weather variables compared to when only past traffic volume data was included as input data. Additionally, it was found that the CNN-LSTM model has the best performance among LSTM, GRU, and CNN-LSTM models.

**2. Datasets**

**2.1. Dataset Description**

The dataset we used is the Hourly Minneapolis-St Paul, MN traffic volume for westbound I-94, provided by UCI Machine Learning Repository. It includes a traffic dataset of Hourly Interstate 94 westbound traffic volume for MN DoT ATR station 301, which is roughly midway between Minneapolis and St Paul, MN. As well as the traffic variable, weather, and holiday features from 2012 to 2018 are included for impacts on traffic volume. The traffic data is provided from the Minnesota Department of Transportation and the weather data is provided from Open WeatherMap.

The ‘holiday’ variable is a categorical variable representing US national holidays, regional holidays, and events such as the Minnesota State Fair. The ‘temp’ variable is a numeric variable indicating the average temperature in Kelvin during the specified hour. The ‘rain\_1h’ and ‘snow\_1h’ are numeric variables of the amount of rainfall and snowfall in millimeters that occurred in the hour. The ‘clouds\_all’ is a numeric variable representing the percentage of cloud cover during the specified hour. The ‘weather\_main’ variable is a categorical variable providing a short textual description of the current weather, indicating the main weather condition. The ‘weather\_description’ variable offers a longer textual description of the current weather. The variable ‘date\_time’ that we used is an index that means the hour of data collection in the local Central Standard Time (CST). For the last, the main variable ‘traffic\_volume’ means the hourly number of cars crossing highway I-94.

**2.2. Dataset Preprocessing**

In the data preprocessing stage, we made several modifications to the dataset, including altering variable types, removing outliers, and dropping certain variables. Additionally, a linear interpolation method was employed and a sliding window approach was utilized to process the data in 24-hour intervals. Following this, label encoding and one-hot encoding were performed, and the remaining variables underwent a scaling process.

The ‘holiday’ variable was transformed from a categorical form to a binary 0 or 1 representation, and the 'weather\_main' data underwent label encoding. The original dataset presented temperature in Kelvin, which we converted to Celsius for easier interpretation. Additionally, we reformatted the ‘date\_time’ variable to a date type and eliminated outliers in the ‘rain\_1h’ variable. The ‘weather\_description’ variable is excluded from our analysis to avoid adding complexity to the prediction.

**3. Proposed Approach**

Currently, two prevalent techniques are extensively used for predicting short-term traffic flow because of their superior capability in managing time-based sequences. These methods are LSTM and GRU [6].

**3.1 LSTM**

LSTM is created to manage long-term data dependencies. They have shown effectiveness in utilizing their unique memory cells to recognize long dependencies.

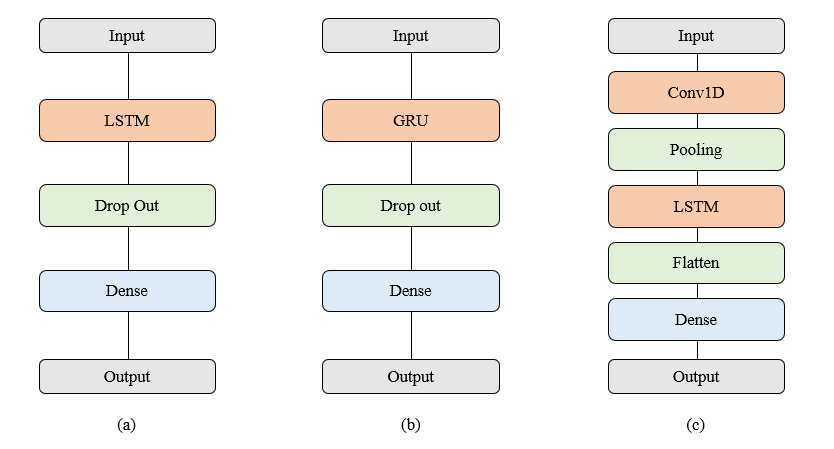
**3.2 GRU**

GRU is slightly different from LSTM, and it was introduced in 2014. Like LSTMs, GRUs handle long-term dependencies and are efficient with sequential data.

**3.3 CNN-LSTM**

To address the challenges encountered when employing LSTM and GRU for short-term traffic flow prediction, particularly in urban road networks with intricate structures and significant travel delays, we have incorporated the use of the CNN-LSTM model. The CNN-LSTM model enhances the predictive accuracy by effectively capturing these complex spatial-temporal interactions, thereby providing a more robust solution for short-term traffic prediction in challenging urban environments.

This paper evaluates the performance of LSTM, GRU, and CNN-LSTM in traffic volume prediction using a suitable regularization method. The structure of the three models proposed for use in this study can be viewed in Figure 1.



**Figure 1.** Architecture diagrams of (a) LSTM, (b) GRU, and (c) CNN-LSTM.

**4. Evaluation Metrics**

Once the training of the model is finished, it's crucial to assess its performance using the validation set data. Evaluation metrics vary across different tasks, and applying various measurement techniques for comparing the efficacy of distinct models often yields varied outcomes. To effectively examine the regression issue addressed in this paper, metrics such as the average absolute error (MAE), mean square error (MSE), mean absolute percentage error (MAPE), and (R-squared) will be utilized as key indicators of performance in the experiment.

=

**5. Experimental Analysis**

**5.1. Overall Performances**

To understand the impact of additional variables such as weather, holidays, and precipitation on prediction performance, we conducted modeling in two phases: one using solely traffic volume data and another incorporating these additional variables.

The results of the modeling demonstrated a significant reduction in error rate and a substantial improvement in prediction performance when these additional variables were utilized in conjunction with traffic volume data, as opposed to using traffic volume data alone.

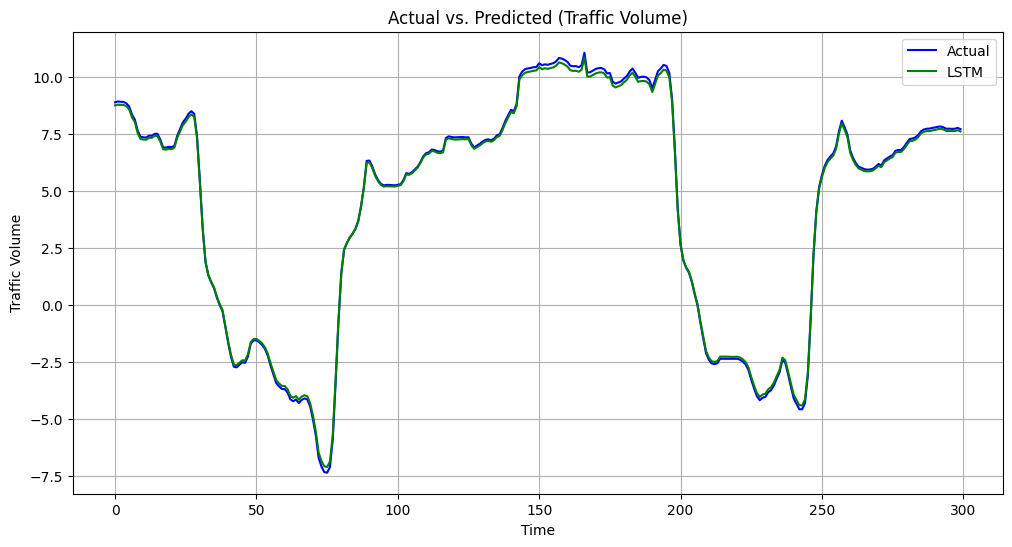
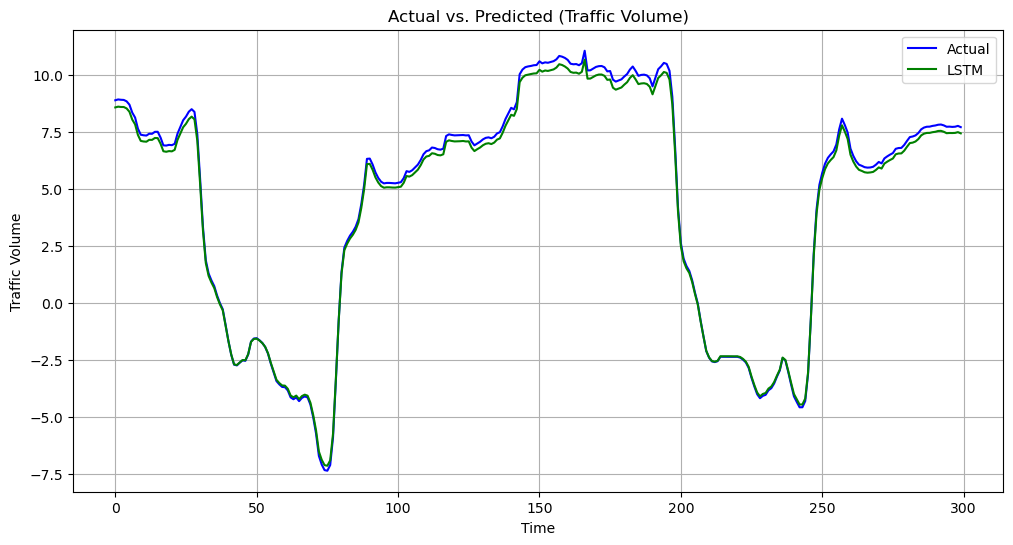
Additionally, the predictive results of the LSTM, GRU, and CNN-LSTM models all showed a high explanatory power of 99%, with the CNN-LSTM model exhibiting the highest performance among them. The specific performances of each model can be found in Table 1 and Table 2.

**Table 1.** Prediction accuracy of train model including   
only traffic volume variables

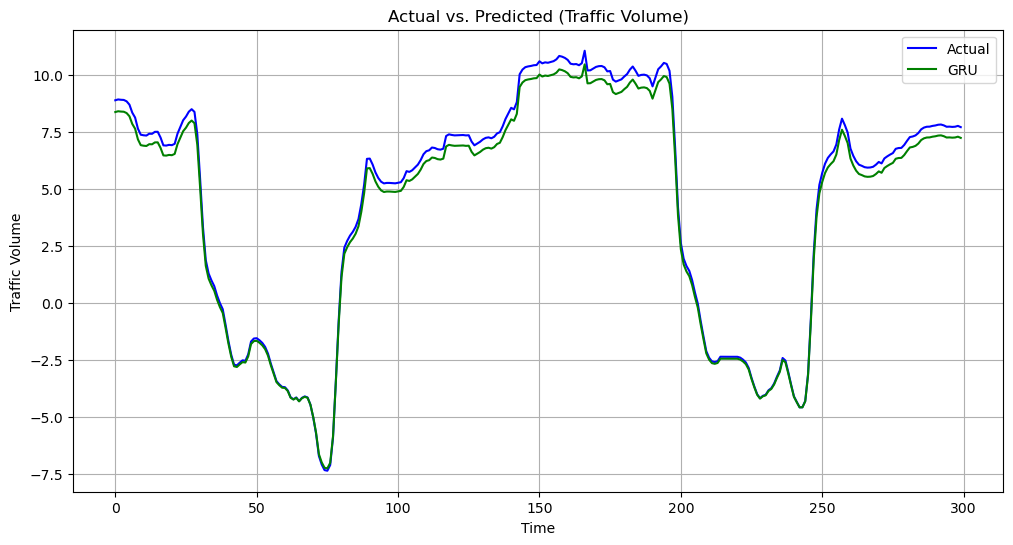
| **Models** |  | **MSE** | **MAE** | **MAPE** |
| --- | --- | --- | --- | --- |
| **GRU** | 0.9960 | 0.1488 | 0.3426 | 0.1397 |
| **LSTM** | 0.9984 | 0.0578 | 0.2132 | 0.0549 |
| **CNN + LSTM** | 0.9999 | 0.0014 | 0.0347 | 0.0248 |

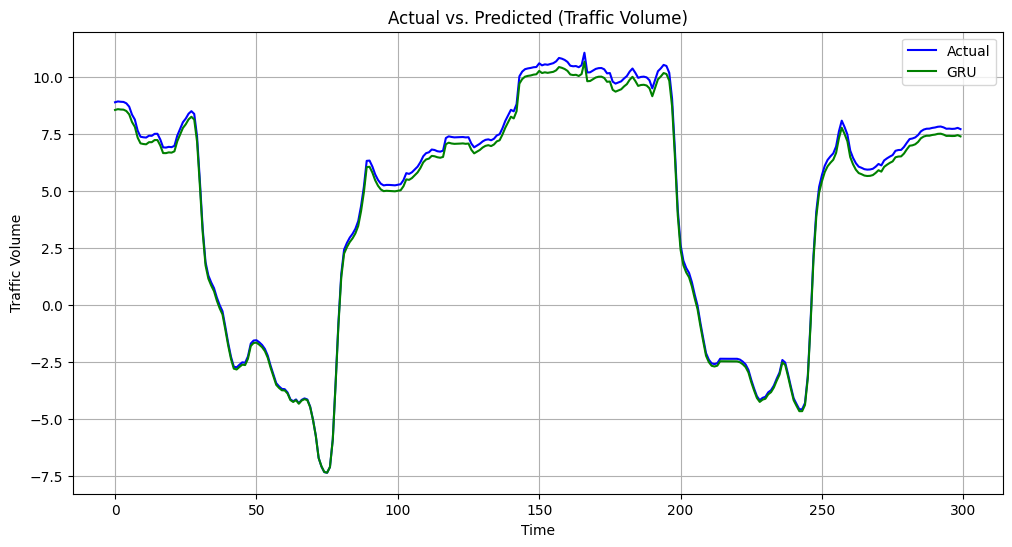
**Table 2.** Prediction accuracy of train model including   
the additional variables

| **Models** |  | **MSE** | **MAE** | **MAPE** |
| --- | --- | --- | --- | --- |
| **GRU** | 0.9986 | 0.0529 | 0.2023 | 0.0874 |
| **LSTM** | 0.9993 | 0.0232 | 0.1273 | 0.0402 |
| **CNN + LSTM** | 0.9999 | 0.0014 | 0.0312 | 0.0  209 |

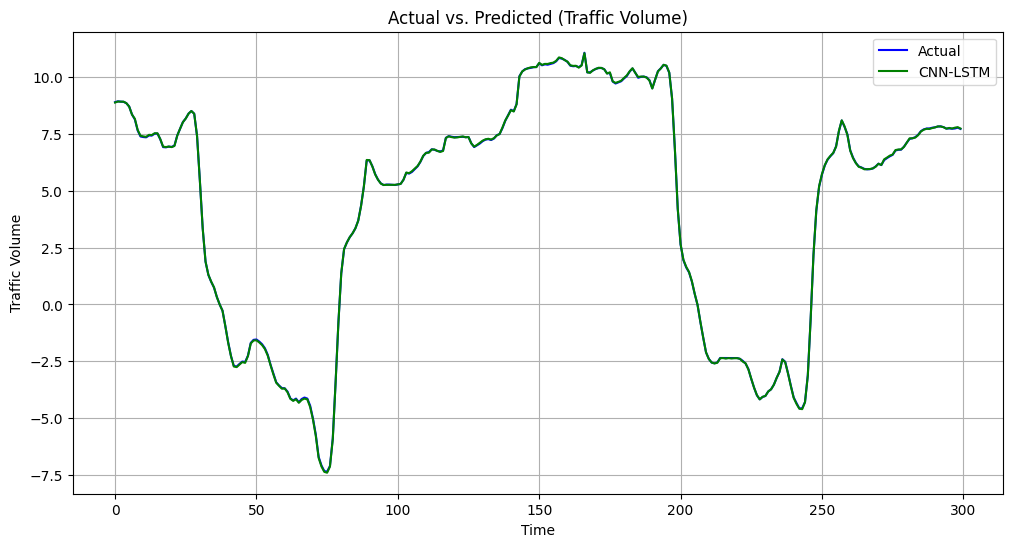
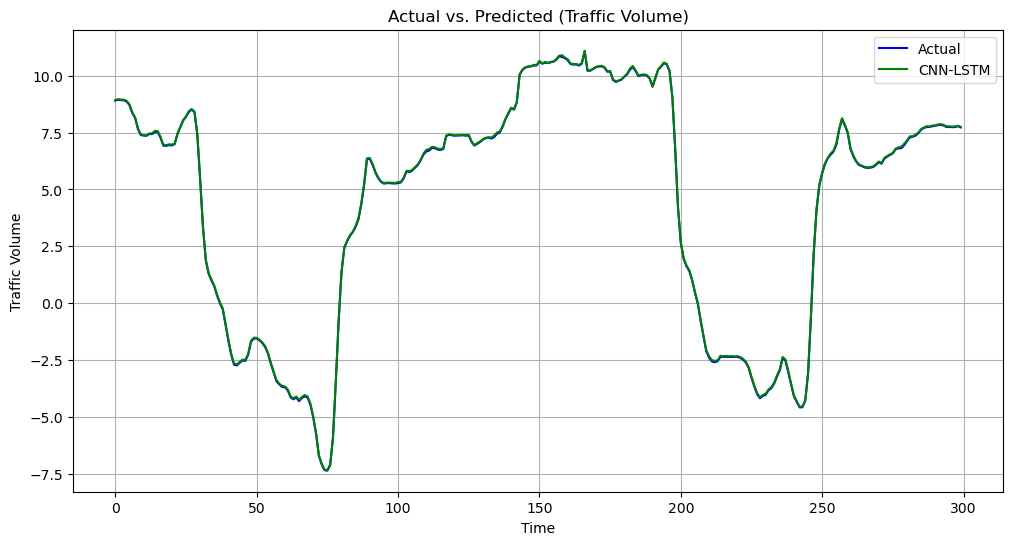


**Figure 2.** Comparison results of the LSTM model excluding and including additional variables.





**Figure 3.** Comparison results of the GRU model   
excluding and including additional variables.

**Figure 4.** Comparison results of the CNN-LSTM model excluding and including additional variables.

**6. Conclusion**

When using variables such as weather conditions, holidays, and rainfall in addition to the traffic volume variable for traffic prediction, a significant reduction in error was observed compared to predicting traffic solely based on traffic volume. The predictive results of the three models all exhibit an explanatory power of over 99%, with CNN-LSTM showing the smallest errors (MSE, MAE, MAPE) among them.

In this study, data collected from highways was utilized, and the traffic volume on highways is less influenced by factors like holidays or rainfall compared to general roads. Therefore, in future research, we aim to predict traffic volume not only on highways but also on general roads. We will explore the differences in the distribution of traffic volume between highways and general roads and seek suitable models and analyses for these scenarios.

**References**

[1]Hogue, John. (2019). Metro Interstate Traffic Volume. UCI Machine Learning Repository. https://doi.org/10.24432/C5X60B.

[2]Minnesota Department of Transportation. Minnesota Department of Transportation - MnDOT. (n.d.). https://www.dot.state.mn.us/

[3]Сurrent weather and forecast - openweathermap. (n.d.). https://openweathermap.org/

[4]P. Poonia and V. K. Jain, "Short-Term Traffic Flow Prediction: Using LSTM," 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3), Lakshmangarh, India, 2020, pp. 1-4, doi: 10.1109/ICONC345789.2020.9117329.

[5]Ke Wang, Changxi Ma, Yihuan Qiao, Xijin Lu, Weining Hao, Sheng Dong, <A hybrid deep learning model with 1DCNN-LSTM-Attention networks for short-term traffic flow prediction, Volume 583, 2021,126293, ISSN 0378-4371,https://doi.org/10.1016/j.physa.2021.126293.

[6]Y. Tao, P. Sun and A. Boukerche, "A Delay-Based Deep Learning Approach for Urban Traffic Volume Prediction," ICC 2020 - 2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020, pp. 1-6, doi: 10.1109/ICC40277.2020.9149018.