Chair of Transportation Systems Engineering
TUM School of Engineering and Design
Technical University of Munich

Factors influencing transfer learning of traffic forecasting models

Master's Thesis

A thesis presented in part fulfilment of the requirements of the Degree of Master of Science in Environmental Engineering at the TUM School of Engineering and Design, Technical University of Munich.

Supervisor M.Sc. Vishal Mahajan
M.Sc. Cheng Lyu
Chair of Transportation Systems Engineering

Submitted by Abu Sayed

Submitted on München, 18.01.2023

Abstract

Traffic flow forecasting is an important strategy for reducing congestion in Intelligent Transport Systems (ITS). According to many kinds of literature, there are several methods of traffic forecasting ranging from statistical time series, and machine learning to combined deep learning. Data insufficiency is a crucial problem for many city areas, and transfer learning (TL) can solve this issue. Also, another advantage of this TL method is that it can reduce the computational burden of the re-training model. A pre-trained model can help to forecast traffic flow on a new limited amount of traffic dataset. This study is proposed for experimenting with the data insufficiency issue, where prestigious Caltrans (PeMS) traffic datasets from California State, USA are used. Among several external factors, weather factor datasets were collected and integrated with traffic datasets, and their influences were checked for different locations. Forecasting models were built by three different algorithms, univariate ARIMA, GRU, and LSTM. Model performances were checked by three different accuracy metrics MAE, MAPE, and RMSE. Six different scenarios were made for transfer learning models in LSTM models to check the performances of three different roads' traffic flow forecasting in California. In transfer learning models, there are three different strategies applied in traffic flow forecasting. Model accuracies were checked for each strategy in each scenario by MAE, MAPE, and RMSE. Although direct forecasting of the target baseline limited dataset model B did not give a good prediction, transfer learning techniques give a good prediction of traffic flow forecasting on new target datasets with a limited amount of data for all scenarios. Overall, strategies 2 and 3 give better forecasting than the strategy-1 transfer learning model in almost all scenarios of transfer-learned traffic flow forecasting. With five-months ahead datasets, for example, with September 2019 datasets, transfer-learned models give better accuracy than target baseline model B. In this case, strategies 2 and 3 also give better forecasting accuracy than strategy-1 traffic flow forecasting. TL method can help with data scarcity issues for forecasting. It is a crucial technique in ITS, which can help in the field of advanced traveler information systems (ATIS), traffic management, smart-signal design, route guidance, etc.

Acknowledgements

I am grateful to my supervisors Vishal Mahajan and Cheng Lyu for their unconditional support of my thesis. I would especially be grateful to Vishal Mahajan for the nice thesis idea and for growing my motivation for deep learning in transportation. I would like to give my gratitude to both supervisors for their time-to-time support and feedback on my work. Finally, I would like to thank my friends and family during the challenging time of this thesis. I would like to dedicate this thesis to my parents and my wife Chumki for their unconditional love and support.

Table of Contents

1 Introduction 1
1.1 Background and Motivation 1
1.2 Objective and Research Questions 2
1.3 Contributions 3
1.4 Thesis Outline 3
2 Literature Review 4
2.1 Traffic Flow Theory 4
2.2 Traffic Forecasting 5
2.3 Traffic State Estimation 8
2.4 Traffic Forecasting Methods 9
2.5 Transferability and Transfer Learning 17
2.6 Research Gaps Related to Literature Review 19
3 Methodology 21
3.1 Data Collection 21
3.2 Data Preprocessing 24
3.3 Model Building 25
3.4 Model Evaluation Metrics 29
3.5 Transfer Learning with Different Strategies 29
4 Data Analysis 32
4.1 District 3 Traffic Data 32
4.2 District 6 Traffic Data 33
4.3 Weather Data 35
4.4 Data Visualization 35
4.5 Train-test Split 43
4.6 Transfer Learning Scenarios 43
5 Result and Discussion 45
5.1 Effects of training data size on model performance 45
5.2 Effects of weather data with traffic dataset 53
5.3 Transfer Learning Results 57
5.4 Transfer learning with September target dataset 69
5.5 Summary of Results 74
6 Conclusion 75
6.1 Summary 75
6.2 Limitations and Future Works 76
References 77
Declaration 85

List of Figures

Figure 1.1: Causes of congestion on highway and roadway (Source: Olayode et al., 2022) 1

Figure 2.1: Greenshield's fundamental diagram of traffic flow vs. density (Source: Monteil, 2014) 4
Figure 2.2: Greenshield's original linear relation between the traffic density and the mean speed (Source: Maerivoet & de Moor, 2005) 5
Figure 2.3: Fundamental diagram regarding traffic flow (q) vs. space mean speed (vs) (Source: Maerivoet & de Moor, 2005) 5
Figure 2.4: The repeating module structure of LSTM (Source: Boukerche & Wang, 2020) 13
Figure 2.5: The cell structure of Gated Recurrent Unit (GRU) (Source: Mateus et al., 2021) 15
Figure 2.6: Representation of (a) Typical traditional machine learning model, (b) Typical transfer learning model (Source: Sagheer et al., 2021) 19

Figure 3.1: Methodological flow chart 21
Figure 3.2: Directional distance covered by each district (Source: PeMS, 2022) 22
Figure 3.3: Number of stations covered by each district (Source: PeMS, 2022) 23
Figure 3.4: Number of detectors covered by each district (Source: PeMS, 2022) 23
Figure 3.5: California district 3 traffic approximate map (Source: Google Maps) 24
Figure 3.6: California district 6 traffic approximate map (Source: Google Maps) 24
Figure 3.7: PACF plot for 4 lane road (station id 319351) data 26
Figure 3.8: ACF plot for 4 lane road (station id 319351) data 26
Figure 3.9: Stacked LSTM layers architecture (source network) 28
Figure 3.10: Sliding window forecasting system for time series data (Source: Hota et al., 2017) 29
Figure 3.11: The architecture of transfer learning strategies adapted from J. Li et al. (2021) 30

Figure 4.1: Elk Grove Blvd Road in Sacramento, California (Station: 313190) (Source: OpenStreetMap) 32
Figure 4.2: Sunrise Blvd Road in Sacramento, California (Station: 314042) (Source: OpenStreetMap) 32
Figure 4.3: SB Riverside Avenue, Placer County, California (Station: 319351) (Source: OpenStreetMap) 33
Figure 4.4: Herndon Avenue in Fresno, California (Station: 601722) (Source: OpenStreetMap) 33
Figure 4.5: Barstow Avenue in Fresno, California (Station: 602538) (Source: OpenStreetMap) 34
Figure 4.6: Cedar Avenue in Fresno, California (Station: 601348) (Source: OpenStreetMap) 34
Figure 4.7: Methodological flow chart of combining weather data 35
Figure 4.8: Traffic flow, occupancy, speed vs day (2-lane road data for station 313190) 35
Figure 4.9: Traffic flow, occupancy, speed vs day (3-lane road data for station 314042) 36
Figure 4.10: Traffic flow, occupancy, speed vs day (4-lane road data for station 319351) 37
Figure 4.11: 2 lane road's district 3 vs district 6 data comparison 38
Figure 4.12: 3 lane road's district 3 vs district 6 data comparison 39
Figure 4.13: 4 lane road's district 3 vs district 6 data comparison 39
Figure 4.14: Correlation plot of 2 lane road in district 3 (Station id 313190) 40
Figure 4.15: Correlation plot of 3 lane road in district 3(Station id 314042) 41
Figure 4.16: Correlation plot of 4 lane road in district 3 (Station id 319351) 41
Figure 4.17: Fundamental diagram (Flow vs. speed) 42
Figure 4.18: Fundamental diagram (Flow vs. density) 42
Figure 4.19: Fundamental diagram (Speed vs. density) 43

Figure 5.1: Changes of MAE based on different data length for 2 lane road (station id 313190) 46
Figure 5.2: Changes of MAPE based on different data length for 2 lane road (station id 313190) 46
Figure 5.3: Changes of RMSE based on different data length for 2 lane road (station id 313190) 47
Figure 5.4: Forecasting plots of 2 lane road (station id 313190) for 4 months of datasets 47
Figure 5.5: Changes of MAE based on different data length for 3 lane road (station id 314042) 48
Figure 5.6: Changes of MAPE based on different data length for 3 lane road (station id 314042) 49
Figure 5.7: Changes of RMSE based on different data length for 3 lane road (station id 314042) 49
Figure 5.8: Forecasting plots of 3 lane road (station id 314042) for 4 months of datasets 50
Figure 5.9: Changes of MAE based on different data length for 4 lane road (station id 319351) 51
Figure 5.10: Changes of MAPE based on different data length for 4 lane road (station id 319351) 51
Figure 5.11: Changes of RMSE based on different data length for 4 lane road (station id 319351) 52
Figure 5.12: Forecasting plots of 4 lane road (station id 319351) for 4 months of datasets 53
Figure 5.13: MAE, MAPE and RMSE plots for 2 lane road for 5 runs 54
Figure 5.14: MAE, MAPE and RMSE for 3 lane road for 5 runs 55
Figure 5.15: MAE, MAPE and RMSE for 4 lane road for 5 runs 56
Figure 5.16: Scenario 1 MAE, MAPE and RMSE 57
Figure 5.17: Scenario 1 prediction plots for 2 lane to 2 lane transfer learning 58
Figure 5.18: Scenario 1 MAE, MAPE and RMSE with 1 month source data 59
Figure 5.19: Scenario 2 MAE, MAPE and RMSE 60
Figure 5.20: Scenario 2 prediction plots for 3 lane to 3 lane transfer learning 60
Figure 5.21: Scenario 2 MAE, MAPE and RMSE with 1 month source data 61
Figure 5.22: Scenario 3 MAE, MAPE and RMSE 62
Figure 5.23: Scenario 3 prediction plots for 4 lane to 4 lane transfer learning 63
Figure 5.24: Scenario 3 MAE, MAPE and RMSE with 1 month source data 64
Figure 5.25: MAE, MAPE and RMSE for 2 lane transfer learning 65
Figure 5.26: Scenario 4 prediction plots for 3 lane to 2 lane transfer learning 65
Figure 5.27: MAE, MAPE and RMSE for 4 lane to 3 lane road transfer learning 66
Figure 5.28: Scenario 5 prediction plots for 4 lane to 3 lane transfer learning 67
Figure 5.29: Scenario 6 MAE, MAPE and RMSE 68
Figure 5.30: Scenario 6 prediction plots for 4 lane to 2 lane transfer learning 69
Figure 5.31: MAE, MAPE and RMSE with target dataset for September (2 lane) 70
Figure 5.32: Prediction plot for 2 lane road transfer learning with September data 70
Figure 5.33: MAE, MAPE and RMSE with target dataset for September (3 lane) 71
Figure 5.34: Prediction plot for 3 lane road transfer learning with September data 72
Figure 5.35: MAE, MAPE and RMSE with target dataset for September (4 lane) 73
Figure 5.36: Prediction plot for 4 lane road transfer learning with September data 73

List of Tables

Table 3.1: Caltrans PeMS spatial description (Source: PeMS, 2022) 22
Table 3.2: Model settings for LSTM and GRU multivariate models 27

Table 4.1: Different scenario for transfer learning for different link type 44

Table 5.1: MAE, MAPE, RMSE value for 1 to 6 months of dataset by three algorithms for 2 lane road 45
Table 5.2: MAE, MAPE, RMSE value for 1 to 6 months of dataset by three algorithms for 3 lane road 48
Table 5.3: MAE, MAPE, RMSE value for 1 to 6 months of dataset by three algorithms for 4 lane road 50
Table 5.4: 2 lane road MAE, MAPE for with and without weather models 53
Table 5.5: 2 lane road RMSE for with and without weather models 54
Table 5.6: 3 lane road MAE, MAPE for with and without weather models 55
Table 5.7: 3 lane road RMSE for with and without weather models 55
Table 5.8: 4 lane road MAE, MAPE for with and without weather models 56
Table 5.9: 4 lane road RMSE for with and without weather models 56
Table 5.10: MAE, MAPE and RMSE for scenario 1 transfer learning 57
Table 5.11: MAE, MAPE and RMSE for scenario 1 transfer learning with 1 month source data 58
Table 5.12: MAE, MAPE and RMSE for scenario 2 transfer learning 59
Table 5.13: MAE, MAPE and RMSE for scenario 2 transfer learning with 1 month source data 61
Table 5.14: MAE, MAPE and RMSE for scenario 3 transfer learning 61
Table 5.15: MAE, MAPE and RMSE for scenario 3 transfer learning with 1 month source data 63
Table 5.16: MAE, MAPE and RMSE for scenario 4 transfer learning 64
Table 5.17: MAE, MAPE and RMSE for scenario 5 transfer learning 66
Table 5.18: MAE, MAPE and RMSE for scenario 6 transfer learning 67
Table 5.19: Traffic forecasting accuracy with target dataset of September (2 lane) 69
Table 5.20: Traffic forecasting accuracy with target dataset of September (3 lane) 71
Table 5.21: Traffic forecasting accuracy with target dataset of September (4 lane) 72

1 Introduction

The first chapter of this thesis describes the introduction of this thesis. First, background and motivation are described in this chapter. Then the contribution and thesis outline are mentioned in this chapter.

1.1 Background and Motivation

Traffic forecasting is an important strategy for predicting future traffic based on historical traffic data on both freeways and urban roadways. It is a technique for predicting traffic information for a specific region by historical traffic data. Stationary sensors (loop detectors, radars, magnetometers) and moving sensors (floating car data, GPS technology) are the instruments that collect real-time traffic data from urban roadways and freeways. Some important traffic variables are flow, speed, occupancy, Vehicle Miles Travel (VMT), etc. Traffic forecasting can be divided into two categories: short-term forecasting and long-term forecasting. Short-term traffic forecasting is either with less than or equal to one-hour granular traffic data (Lana et al., 2018). In our study, first, we will do short-term traffic-flow forecasting, and then transfer learning using pre-trained models.

Every day people lose time and fuel due to congestion, and it increases frustration to people. Congestions also increase environmental pollution and hamper the transportation of goods and passengers on time. The accident is an example of a road incident, and it can happen several times a day and has a negative impact on traffic flow. Accidents can create congestion, and secondary incidents can happen, which causes delays. Environmental effects, for example, bad weather conditions, another cause of delay reduce speed on highway and urban roadways. Also, an increasing number of vehicles is another issue for increasing congestion. The higher number of vehicles and lower capacity of roadways create congestion. As urbanization advances and vehicles become more popular, transportation issues become more and more challenging (Boukerche & Wang, 2020). Some problems happened due to the higher amount of traffic on the highway and urban roadways. For example, congestion is a common issue. Another crucial problem is traffic incidents. An accident can create secondary accidents and traffic congestion in the upstream direction of traffic flow due to no information on the previous incident. Another cause of traffic congestion is higher traffic density.

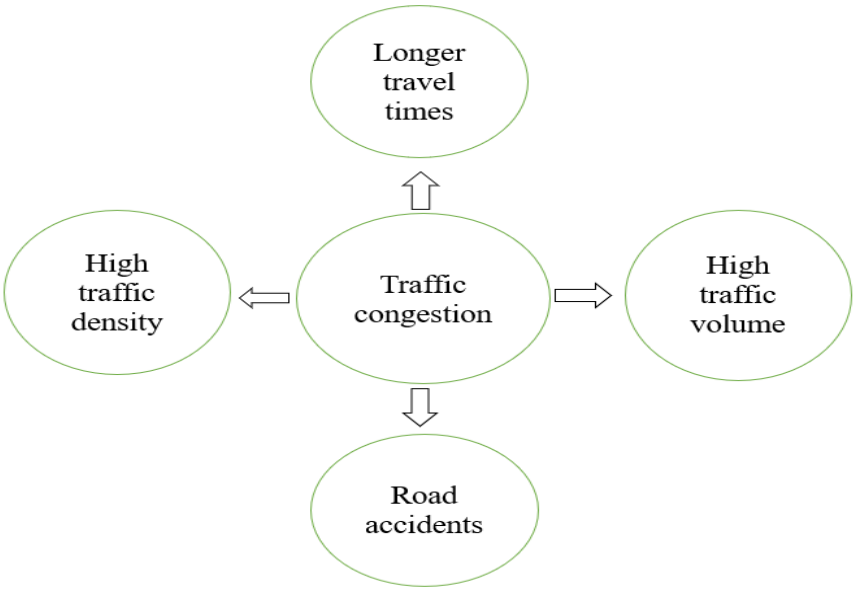


Figure 1.1: Causes of congestion on highway and roadway (Source: Olayode et al., 2022)

Building additional roadways and adding more lanes on the roads is the first answer that comes to most people's minds when trying to fix these problems (Boukerche & Wang, 2020). However, increasing the

road capacity would result in worsening traffic conditions, according to research by Dechenaux et al. (2014). One efficient solution to these problems could be better transportation systems with upgraded transportation resources and modern traffic control systems. One most well-known systems of them are the Intelligent Transportation System (ITS). As an essential component of ITS, an accurate and effective road traffic prediction system can deliver consistent and exact information on the state of the road based on previous road conditions (Alipour et al., 2019; Boukerche & Wang, 2020; Domingues et al., 2019). Traffic forecasting can give an advanced idea about the traffic situation on roadways.

Data scarcity is another problem in traffic forecasting because traditional traffic data collection needs a lot of sensors, which costs a lot. Data scarcity or insufficiency is the problem, where fewer amount of data is available, and which do not give good predictions but rather do overfitting. For the transferability of traffic forecasting from one region to another, missing data is another problem. Implementing a previous region's pre-trained model to a new limited dataset is a recent smart strategy that can give better traffic forecasting. For the issue of data scarcity, this technique of model transferability helps a lot. There is a crucial problem due to the complex spatial and temporal dependencies for route guidance and traffic control (Y. Li et al., 2017; Y. Li & Shahabi, 2018). As traffic time series data shows strong temporal dynamics, incidents such as rush hours or accidents can create non-stationary time series, which will be a challenge in traffic forecasting for better accuracy (Y. Li & Shahabi, 2018).

Urban congestion reduction strategies include a variety of techniques, such as traffic management, route guidance, congestion fees, and infrastructure improvements for transportation (J. Ma et al., 2020; J. Sun et al., 2020; Zheng & Huang, 2020). However, if a city's transportation infrastructure is fully developed, it is challenging to reduce traffic congestion by expanding road facilities (Zheng & Huang, 2020). By improving traffic management in this situation digitally and intelligently, that is, by creating an intelligent traffic system (ITS), traffic congestion might be successfully decreased (Nallaperuma et al., 2019; Zheng & Huang, 2020). For reducing congestion and incident, traffic forecasting is a strategy for predicting future flow and speed in ITS, advanced traveler information systems (ATIS), and route guidance. Long-term traffic forecasting helps with the design and building of a sensible road network and will support the timely scheduling of construction progress and operation of road maintenance (Yu et al., 2015). As traffic flow changes are unusual and stochastic, it is challenging to get exact forecasting. Recent technological development, for example, the improvement of computers helps a lot in analyzing high volumes of data. By using these data, different statistical analyses and machine learning techniques help to forecast the future state of any traffic events.

Traffic forecasting techniques can help to speed up the traffic flow and traffic speed, reduce congestion, save money, and increase road safety. The short-term traffic forecasting model is a modern technology for real-time traffic control and management (Habtemichael & Cetin, 2016). Such technologies are crucial for deploying emergency management systems, enhancing traffic signals, and giving travelers accurate travel time information (Habtemichael & Cetin, 2016). In transportation sectors, traffic data is a core thing for future demand estimation of traffic and infrastructure development. Historical traffic data are applied for getting the accurate forecasting of traffic parameters. In this study, machine learning and deep learning algorithms are used. Transfer learning is another technique of using a pre-trained forecasting model for predicting traffic of a new experimental limited dataset.

1.2 Objective and Research Questions

This thesis is motivated to analyze traffic prediction in the state of California, USA. We will assess and forecast traffic flow with the PeMS traffic data of different cities in California. The main objectives of this study are:

- To develop a methodology of transfer learning (TL) of traffic forecasting.

- To select source datasets and target datasets using two different district data of California, USA.
- To select several machine learning and deep learning algorithms for forecasting and transfer learning.
- To create several scenarios of TL with several link-type datasets.
- To investigate the data scarcity issues in traffic forecasting.
- To compare the accuracy of the target baseline model and TL strategies.

Following research questions raised for this study of traffic forecasting:

(1) What is the best amount of input data in traffic forecasting models, which will be used as a source dataset for transfer learning?
(2) How does weather as a factor affect the transferability of traffic forecasting models from one link to another?

1.3 Contributions

This thesis uses different types of machine learning (ML) models, and several data attributes to forecast traffic flow based on different types of detectors (e.g., loop detectors, radars and other types of detectors, and meteorological sensors) information. Machine learning has become more advanced in recent decades. Due to the potentiality of ML for artificial intelligence, it is now frequently employed in the transportation industry during the fourth industrial revolution (Olayode et al., 2022). In the context of the detector's historical data, the thesis defines appropriate tools and methodologies of data analytics. The thesis contributes to the research of ITS based on the ML approach of traffic forecasting by real-world data of different districts of California State, USA.

ML models, for example, Autoregressive Moving Average (ARIMA), Gated Recurrent Unit (GRU), and Long Short-Term Memory (LSTM) are used to train models to predict traffic flow in two different districts of California state. The transfer learning (TL) strategy is used to forecast the traffic flow of a new District by a pre-trained model from another District. Weather data are collected and integrated with the model to evaluate the influences. Rainfall, wind speeds, visibility, and other features are available in the weather dataset. Techniques that extract dependence between historical data using different ML methods and estimate the traffic state is called the data-driven approach (Seo et al., 2017). Based on all the historical data, using machine learning algorithms for future traffic flow prediction is a data-driven technique, and we use this technique in this thesis.

1.4 Thesis Outline

The thesis consists of six chapters.

- **Chapter 1** introduces the thesis with background and motivation, objectives, and research questions.
- **Chapter 2** describes the literature review that focuses on traffic flow models, traffic forecasting, traffic state estimation, factors on traffic flow, time series models, different methods of traffic forecasting models, and transfer learning.
- **Chapter 3** describes the methodology where data collection, data cleaning, data filtering, model building with different algorithms, and transfer learning architecture are available.
- **Chapter 4** includes data analysis of selected road datasets, and weather datasets, where district 3 traffic data, district 6 traffic data, weather data, data visualization, macroscopic-fundamental diagrams, and train-test split are available.
- **Chapter 5** includes results and discussions, where answers to research questions are made, and transfer learning several scenarios results and comparisons are made.
- **Chapter 6** describes conclusion of this thesis and future research direction.

2 Literature Review

In this chapter, we will present the literature review of the different contexts of traffic forecasting. We split this chapter into the following sub-chapters, namely traffic flow theory, traffic forecasting, traffic state estimation, traffic forecasting methods, time series prediction, transferability of traffic flow, and transfer learning.

2.1 Traffic Flow Theory

John Glen Wardrop pioneered the developing field of traffic flow theory by utilizing mathematical and statistical concepts to describe traffic flows (Maerivoet & de Moor, 2005; WARDROP, 1952). Understanding the precise spatiotemporal dynamics of traffic flow is the goal of traffic flow theory (Rempe, 2018). The study of traffic flow modeling had a renewed interest towards the start of the 1990s (Maerivoet & de Moor, 2005). On the one hand, the LWR model's appealing simplicity renewed academics' attention, but on the other hand, one of the biggest driving factors came from the field of statistical physics for the traffic flow model (Maerivoet & de Moor, 2005). Traffic flow modeling has become one of the simplest strategies to reduce the congestion of city and highway traffic. Traffic flow theory advancements allow for improved design, operation, and development of Intelligent Transportation Systems (ITS) (Rempe, 2018). The traffic macroscopic fundamental diagram was first proposed by Greenshields (1934). According to Monteil (2014), the first model proposed by Greenshields that have the following relation:

$$q_{eq}(\rho) = 4q_{max}\rho\left(1 - \frac{\rho}{\rho_{max}}\right)1_{0\leq\rho\leq\rho_{max}} \tag{1.1}$$

Where: q_{max} is the maximum traffic flow and ρ_{max} is the maximum traffic density. The Greenshield's fundamental diagram shows in figure 2.1, where ρ_c is the critical traffic density, q_s is the arbitrary flow and r_1, r_2 are the solutions to equation $q_{eq}(\rho) = q_s$, which is the flux function for hyperbolic systems of conservation laws.

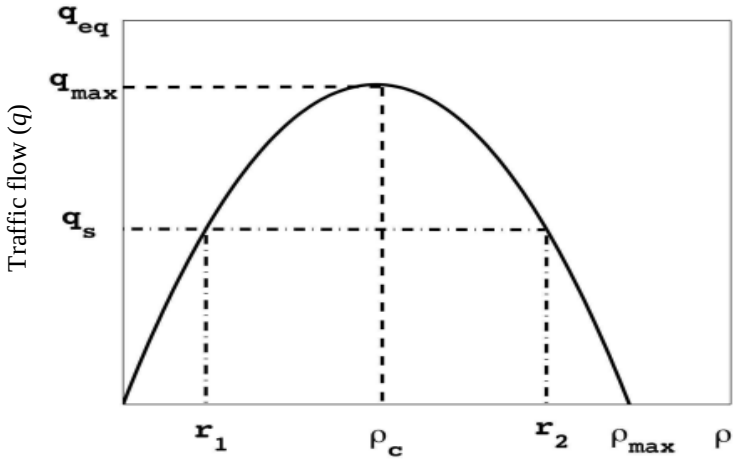


Figure 2.1: Greenshield's fundamental diagram of traffic flow vs. density (Source: Monteil, 2014)

Figure 2.1 represents Greenshield's fundamental diagram of traffic flow vs. traffic density, the x-axis shows traffic density, and the y-axis shows traffic flow. Initially, when traffic density increases, traffic flow also increases. At critical density ρ_c , traffic flow is maximum, and then due to increasing density,

traffic flow decreases. At maximum density ρ_{max}, traffic flow is almost zero. The upward parabolic region is called the free-flow region, and the downward parabolic region is called the congestions region.

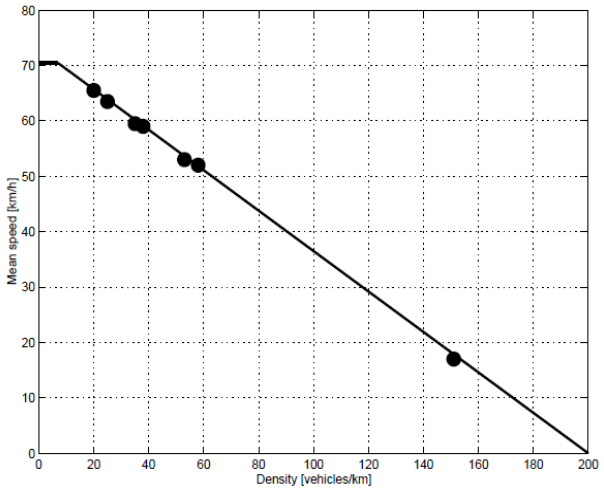


Figure 2.2: Greenshield's original linear relation between the traffic density and the mean speed (Source: Maerivoet & de Moor, 2005)

Figure 2.2 describes Greenshield's original relationship between the density and the mean speed. The x-axis shows the density (vehicles/km), and the y-axis shows the mean speeds of vehicles. Based on this relationship, the mean speeds of vehicles are initially higher with lower density. The maximum speed is between 0 and approximately 5 vehicles/km. Then after this point, speed linearly decreased due to increasing the density. The speed is completely zero at the highest density, which is 200 vehicles/km.

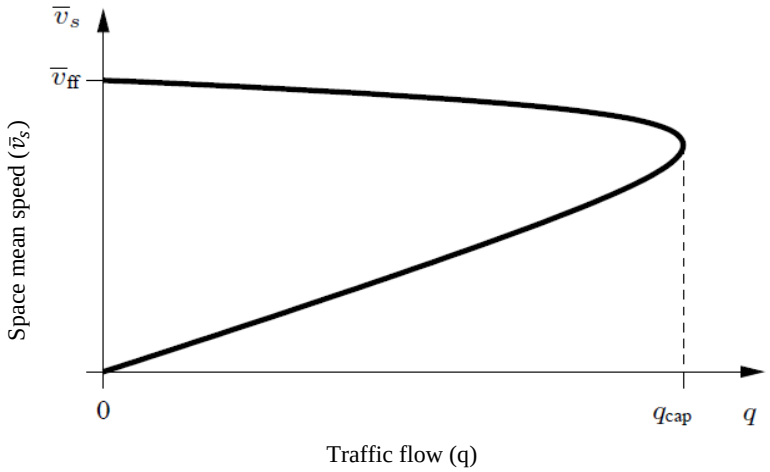


Figure 2.3: Fundamental diagram regarding traffic flow (q) vs. space mean speed ($\bar{v}_s$) (Source: Maerivoet & de Moor, 2005)

Figure 2.3 depicts the fundamental diagram relating to the traffic flow vs traffic space mean speed. Capacity flow q is the maximum traffic average flow. According to Maerivoet & de Moor (2005), there are two regimes in the fundamental diagram. The upper branch is called the free-flow regime, and another one is the lower branch which is the congested regime.

2.2 Traffic Forecasting

Traffic forecasting is a crucial step in planning road infrastructure, from the investment feasibility assessment to the creation of functioning documentation (Dombalyan et al., 2017). Due to rapid technological development in transportation industries, traffic flow forecasting gained importance in

intelligent transportation systems (ITS) for modern traffic management. It is a process of highway and urban traffic flow, or volume prediction based on historical traffic data. For improving highway operations and preventing traffic congestion, a reliable prediction model might be helpful (Abduljabbar et al., 2021). A new generation of intelligent transportation systems has come, and with it, new approaches to the problems associated with traffic forecasting (Gao et al., 2022). These approaches are divided into two different types: classical statistical approaches and data-driven approaches (Gao et al., 2022; Lippi et al., 2013; Vlahogianni et al., 2014). The data-driven approach is when decisions are made based on the analysis and interpretation of numerical data rather than on observation. Dai et al. (2019) proposed a study where they suggested a model for predicting short-term traffic flow that coupled spatiotemporal analysis with a Gated Recurrent Unit (GRU). In this study, the traffic flow was first submitted to time correlation analysis and spatial correlation analysis; the optimal input time and spatial data volume were then determined using the spatiotemporal feature selection approach. Finally, the prediction result of their proposed method (GRU + STFSA) was compared with Convolutional Neural Network (CNN) and (GRU). Their proposed method outperforms GRU and CNN models in accuracy and stability.

Ren et al. (2021) proposed a study on traffic flow prediction by combining deep learning models in two Chinese road traffic flow datasets. They combined CNN, Long-Short Term Memory (LSTM), and GRU with a Dynamic Optimal Weighted Coefficient Algorithm (DOWCA), namely a CDLP model in a sequential combination structure. They experimented with different neuron numbers in the network and found the accuracy of each scenario. They compared the CDLP model with other individual and combined models and got a better performance by the CDLP model. According to MAPE, MAE and RMSE CDLP model outperforms all other models.

Another study was done by Z. Sun et al. (2022) for forecasting short-term traffic flow based on different algorithms on the PeMS CalTran dataset. They compared the performance of each model, and finally, they found that combination model K-means-GRU performs better in comparison with the GRU network, stacked autoencoders (SAEs), Random Forest (RF), and Support Vector Machine Regression (SVR). They also found that there are significant differences in traffic flow distribution patterns.

S. Sun et al. (2020) proposed a study of city-wide traffic flow forecasting using a deep convolutional neural network (CNN). They chose the Chinese state Wuhan city dataset. This dataset also included some external factors like holidays and weather conditions. They used one hot encoder for adding the external factors. Finally, the accuracy of the model was evaluated by accuracy metrics. Tran et al. (2022) proposed a study on short-term traffic speed forecasting for a multi-lane road using GPS-monitored data using different deep learning algorithms. They used a Vietnamese city road for the experiment on AR, ARMA, ARIMA, SARIMAX, SES, HWES, PROPHET, MLP, and LSTM models to find out the performance differences in these models. Finally, they found that the LSTM model performs better in comparison with other algorithms.

Z. Yu et al. (2015) proposed a research article on traffic flow forecasting based on combinational algorithms. They used a combination of Grey Model (GM), ARIMA, and Generalized Regression Neural Network (GRNN) algorithms for forecasting traffic flow. Lee et al. (2021) did two experiments on traffic data prediction. They used STGCN, DCRNN, ASTGCN, St-MetaNet, Graph-WaveNet, STG2Seq, GMAN, etc. models on METR-LA, the PeMS speed, and the traffic flow dataset for prediction. According to their experiments, Graph-WaveNet and GMAN showed better performance. Qu et al. (2020) did research on intersection traffic flow forecasting. A two-layer stacked network in a combination of K-Nearest Neighbor (KNN) and Elman Neural Network (ENN) was used for an experimental intersection in China. Non-parametric and machine learning models were also used in this research to compare their proposed combined model for the signalized intersection. Their proposed

model outperformed other models in this study. Shao & Soong (2016) proposed a study on short-term traffic flow prediction on Caltrans PeMS historical data at mainline Irvine Center Drive by the LSTM model. They experimented with the LSTM model has the capability of capturing long-term dependency traffic data. They compared this LSTM model with other models. Moreover, LSTM outperforms other models.

2.2.1 Factors on Traffic flow

Many factors affect traffic flow patterns on highways and roadways. For example, these factors are adverse weather conditions, construction sites, or incidents. Bad weather conditions such as snowfall, rainfall, and high wind speed reduce the speed of vehicles and traffic flow (Cools et al., 2010). There will be a comparable drop in flows in a region when temporary traffic control is implemented due to road construction (Xie et al., 2019). Road construction and maintenance works can create delays and increase longer travel time. Extreme weather, such as thunder, lightning, and pollution, can influence urban traffic movements (Xie et al., 2019). When it's cloudy outside, it makes it more likely that fewer people will go out, however when it's bright outside, more people may go out (Xie et al., 2019). Additionally, there are cross-regional and increasing flows during holidays like National Day and Spring Festival, which occur on a cyclical basis each year (Xie et al., 2019). The day before the holiday, also has some effects on the urban flows in terms of time, increasing the fluctuation in crowd flows (Xie et al., 2019).

Zhang & Kabuka (2017) proposed a study on combining weather conditions for the prediction of traffic flow in deep learning. They make two different scenarios for traffic flow prediction. One scenario is weather as a factor, and another is without weather as a factor. They used the gated recurrent unit (GRU) based recurrent neural network for the prediction, calculated the accuracy of both models, and found better accuracy for the model which is made with weather data. In future research, they will take a much larger dataset for this similar type of experiment, will build a deep neural network, and will make comparisons with other classical machine learning frameworks. Braz et al. (2022) proposed research on traffic flow forecasting using historical traffic data with weather data using three different algorithms. These algorithms are LSTM, autoregressive LSTM, and CNN in terms of MAE and time intervals. They used two beaches of Portugal's historical traffic and weather data. Overall, CNN performs better in terms of MAE with 10 minutes intervals of input variables compared to the two algorithms. Another study is proposed by Rahman (2020) on forecasting traffic flow with weather data using different machine learning algorithms like K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Artificial Neural Network (ANN). They used Dublin Airport route data for short-term traffic flow forecasting and set several parameters based on different weather conditions and different day types. Then they compared the results and found better results for the parameter set 1, where they used weather parameters as sunny + not rain, sunny + rain, not sunny + not rain, and not sunny + rain. In this study, they didn't consider other parameters like traffic jams, accidents, etc.

2.2.2 Application of Traffic Forecasting

According to Navarro-Espinoza et al. (2022), traffic sensors that monitor the number of vehicles passing through a lane once per specified amount of time will be used to feed traffic forecasting models in smart traffic signal controllers. With these readings, a database like the one utilized in this study can be generated (Navarro-Espinoza et al., 2022). The ML model will be trained for each intersection when the databases are created (Navarro-Espinoza et al., 2022). Then, using a certain number of prior data, the traffic flow for the subsequent time will be forecasted (Navarro-Espinoza et al., 2022). Once the forecast did, it will be possible to better program for the timings of each state either manually by a user or automatically using an algorithm to determine the best times for the states of the traffic light. Wirelessly connecting the traffic light to a central station or the controller can complete the entire operation

(Navarro-Espinoza et al., 2022). This strategy can help to speed up the traffic flow and speed at intersections. This smart technology can reduce traffic congestion in busy intersections.

2.3 Traffic State Estimation

Traffic state estimation is the first part of traffic forecasting. Traffic flow, speed, and density are three major parameters that explain the characteristics of traffic on road segments (Grumert & Tapani, 2018). The flow/ volume is defined as the number of vehicles passing a point or location over a specified time interval in minutes, hours, etc. Traffic flow can be either short-term traffic flow or long-term traffic flow. According to Grumert & Tapani (2018), speed is defined by time mean speed and space mean speed. Time mean speed is the average speed of all vehicles crossing a particular location in a predetermined time. Space mean speed is the average speed of all vehicles going over a particular road segment at a specified time. The traffic density is the total number of vehicles occupying a unit length of the lane of the roadway at any time. These three traffic parameters are commonly known as traffic state.

Stationary detectors like loop detectors and radar detectors are the conventional method for estimating traffic state (Coifman, 2003; Grumert & Tapani, 2018; Kurkjian et al., 1980; Singh & Li, 2012). To make these sensors more efficient, it is crucial to think about how to collect historical traffic data from them (Hong & Fukuda, 2012). The position of these stationary sensors is a factor for collecting traffic data, and effective location intervals give a better estimation of traffic flow and speed. Hong & Fukuda (2012) did a case study on the effects of sensor position that can help to collect better traffic data. In practice, detectors and cameras are some common examples of stationary sensing equipment that observe flow (e.g., PeMS) (van Erp et al., 2018). CalTrans PeMS covers the whole of California State by stationary detectors like magnetic loop detectors, radars, and magnetometers for collecting traffic data. According to Bouyahia et al. (2021), these conventional technologies needed higher costs to install and maintain, which doesn't give benefits for large-scale deployment. Also, some other recent technologies which estimate traffic state by dynamic detectors like floating car data, drone videography, or satellite imagery (Janecek et al., 2015; Mahajan et al., 2022), cellular phones (Janecek et al., 2015) etc. S. M. Khan et al. (2017) proposed research, that applied connected vehicle technologies (CVT) for collecting traffic data. They found that with a 10% or more CV penetration with PeMS data, higher accuracy can be achieved using the CVT-AI algorithm than PeMS density algorithm. CVT is a recent potential method for collecting continuous real-time traffic data (Bhavsar et al., 2007; S. M. Khan et al., 2017; Y. Ma et al., 2010; Yongchang Ma et al., 2009).

Fundamental diagram proposed by Greenshields (1934) gives an idea about traffic flow and speed patterns through the different diagrams. If the underlying relationship holds in fundamental diagrams, it is possible to predict traffic flows using easily accessible traffic flow, and speed data along with additional factors (Mahajan et al., 2022). Agalliadis et al. (2020) proposed a study on traffic state estimation for highway networks with potentially incomplete, limited data from different sensors. They applied an Unscented Kalman Filter (UKF) which will help in the data fusion process in the approach of the algorithm. Another second-order traffic model named METANET was combined with UKF, which was modified to consider the spatiotemporal interrelations between the corresponding error terms. Hawes et al. (2016) proposed a study on traffic state estimation within road segments using particle filters. They initiated two different approaches for handling missing measurements, one is compressive sensing, and another is Bayesian compressive sensing. Simulated and real data was tested by two algorithms. They got better improvement in terms of missing data.

2.3.1 Data Definition

The datasets collected by point detectors are the most often utilized in the literature in terms of data sources (Lana et al., 2018). Data from traffic speed sensors gathered by the Caltrans Performance

Measurement System (PeMS) and the Los Angeles Metropolitan Transportation Authority Service (METR-LA) were applied in a lot of research (Zhao et al., 2017; Lv et al., 2014; Z. Wu et al., 2019; B. Yu et al., 2017). These datasets are widely used for predicting traffic flow, incident detection, and different factors in the traffic flow of California and Los Angeles. Every day, a large amount of data are generated by traffic in several locations in these states. The traffic detectors are placed in multiple locations on urban roadways and highways in California State. These detectors can read real-time traffic movement, flow, speed, and other parameters. Although the datasets are critical in new traffic forecasting techniques, they are associated with technology that has substantial installation and maintenance costs (Buroni, 2021). The detectors are prone to malfunction and mistakes (daily in California, 30 percent out of the sensors installed do not work according to the PeMS system) (Herrera et al., 2010).

On the other hand, just a few studies in the literature employ datasets collected by moving sensors (Buroni, 2021; Lana et al., 2018). Moving sensor data, for example, floating car data, and GPS tracking systems give quite accurate data regarding traffic movement and location. These data help a lot in traffic management systems for both urban roadways and highways. Real-world internet traffic forecasting is a bit challenging for three reasons: i) data heterogeneity, ii) anomalous data, and iii) data scarcity (Saha et al., 2022). In the temporal and spatial dimensions, internet traffic shows heterogeneity, because of different traffic patterns at different time-lapses (Saha et al., 2022). Internet traffic in the real world is subject to many internal and external influences, which results in nonstationary complicated traffic patterns additionally, such events contain data points that fall outside the parameters of the data distribution, which would limit the model's capacity to generalize during deployment (Saha et al., 2022). For better performance in traffic prediction, it is needed a high volume of historical data, but it is tough to get all regions data for improve traffic prediction, data scarcity is another major issue in traffic forecasting (Saha et al., 2022).

2.4 Traffic Forecasting Methods

This section describes different methods of traffic forecasting. According to several kinds of literature, these methods are categorized based on underlying assumptions about different traffic states(C. van Hinsbergen et al., 2007; Lana et al., 2018; Vlahogianni et al., 2004, 2014). Four methods are (i) the naïve methods, (ii) the parametric methods, (iii) the non-parametric methods, and (iv) the adaptive strategies. These methods are used widely in traffic flow, speed, and other parameters forecasting.

2.4.1 Naïve Method

The Naïve method is considered the simplest traffic forecasting method, which makes the forecast using fundamental statistical assumptions (J. J. Q. Yu, 2021). Among them, the historical average (HA) technique and the instantaneous travel times (ITT) approach are the two most used in real-world applications (Smith et al., 2002; J. J. Q. Yu, 2021). The naïve method considers what happened previously and predicts what happens again. Because of their simplicity and cheap computing effort, these approaches are commonly applied in practice. As this method has lower accuracy, we will not use this method for forecasting traffic flow.

2.4.2 Parametric Methods

A parametric technique involves using a formula to calculate a margin of error and a sample mean to estimate the population means. The parameters are learned from the dataset of the parametric technique, although its structure is fixed (Tran et al., 2022; C. P. IJ. van Hinsbergen et al., 2009). Parametric models are based on traffic information, which assists in understanding traffic processes, especially in traffic simulations. The benefit of such parametric models is that they can catch previously undiscovered occurrences and situations. Another benefit is the parametric method needs fewer data. Due to the

assumptions made while parameterizing the models, many of these systems are demonstrated to perform poorly in unstable traffic conditions and complex road configurations (Tran et al., 2022; Williams, 2001). Parametric methods consist of i) a traffic simulation model, and ii) a parametric time-series model.

- **Traffic simulation model:** As early as 1956, the foundation of this modeling is laid (D. Boyce, 2007; D. E. Boyce & Williams, 2005). Using the concept of Network Equilibrium, traffic is assigned from an origin-destination (OD) matrix ('Wardrop's First Principle') (Buroni, 2021). Utilizing car-following theories, traffic simulation uses mathematical models to show how vehicles might move in real-time (Alghamdi et al., 2022). According to Alghamdi et al. (2022), Traffic prediction differs from traffic simulation models in a way that it uses historical data to forecast traffic value (volume, speed, density, etc.) at a future time point under predefined conditions. On the other way, traffic simulation models estimate the value of the traffic (volume, speed, density, etc.) under certain conditions. Examples of traffic simulation models are the car following model, lane changing model, and gap acceptance model. VISSIM, SUMO, AIMSUN, MATSim, and Paramics are some traffic simulation tools. We will not implement this simulation technique in our study, but we will implement a traffic prediction model with historical traffic data.
- **Parametric time-series model:** The prediction in these approaches involves modeling a traffic variable as a function of its previous observations and error factors, which necessitates that the process remains stable (Buroni, 2021). However, because traffic processes are rarely stationary, these models frequently account for seasonality (Buroni, 2021). Abduljabbar et al. (2021), When compared to non-parametric techniques, these models suffer from a lack of capturing the dynamic traffic patterns. Autoregressive integrated moving average model (ARIMA), seasonal autoregressive integrated moving average model (SARIMA), vector autoregressive moving average (VARMA), etc. are some parametric approaches for traffic flow forecasting.

2.4.3 ARIMA Model

The ARIMA model was first introduced by Box and Jenkins in 1970 (F. M. Khan & Gupta, 2020; Y. Wang et al., 2022). It is a single product autoregressive moving average process in which a random process with *d* unit roots is initially assumed, and after *d* times difference, the random process can be changed into a stable autoregressive moving average process (Y. Wang et al., 2022; W. Wu et al., 2015). To put it another way, the time series in the autoregressive (AR) model, the moving average (MA) model, and the autoregressive moving average ARMA model are all stationary series (Y. Wang et al., 2022). Autoregressive (AR) process and Moving Average (MA) processes together build a combined model of time series, which is ARIMA model (Siami-Namini et al., 2018). There is no seasonality in these types of models, these are called stationary not moving. *ARIMA (p,d,q)*, the main elements of this model:

- *AR*: Autoregression. A regression model that takes use of the relationships between a single observation and several lagged observations (*p*) (Siami-Namini et al., 2018).
- *I*: Integrated. For making the time series stationary, comparing the differences between observations made at various times (*d*) (Siami-Namini et al., 2018).
- *MA*: Moving Average. A method that considers the dependency between observations and the residual error terms for getting lagged observations (*q*), a moving average model is used (Siami-Namini et al., 2018). According to Siami-Namini et al. (2018), a simple version of an AR model of p order can be written as a linear process:

$$x_t = c + \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t \quad (2.1)$$

Where:

x_t = stationary variable,

c = constant,

ϕ_i = autocorrelation coefficients at lags 1, 2, …, p, and

ϵ_t= residuals, are the Guassian white noise series with mean zero and variance σ_ϵ^2.

An MA model of order q can be written as:

$$x_t = \mu + \sum_{i=0}^{q} \theta_i \, \epsilon_{t-i} \tag{2.2}$$

Where:

μ = expectation of x_t (normally assumed as zero),

θ_i = weights applied to the current and prior values of a stochastic term in the time series,

θ_0 = 1,

ϵ_t= residuals, are the Guassian white noise series with mean zero and variance σ_ϵ^2.

Although time series acts as *d*-order unit root processes in several cases, it is needed to carry out differential processing on the data to convert them into a stationary time series before developing the model (Y. Wang et al., 2022). According to Y. Wang et al. (2022), the ARIMA (*p, d, q*) model can be written by following equation:

$$\left(1 - \sum_{i=1}^{p} \alpha_i L^i\right)(1 - L)^d y_t = \alpha_0 + (1 + \sum_{i=1}^{q} \beta_i L^i)\varepsilon_t \tag{2.3}$$

Where:

p = non-seasonal autoregressive order,

d = non-seasonal difference order,

q = non-seasonal moving average order,

α_i= the autoregressive term coefficient,

β_i= the moving average term coefficient,

L = the lag operator,

ε_t= the white noise series,

α_0= the constant term coefficients,

i= the coefficient number, and

y_t= the time series value at time t.

2.4.4 SARIMA Model

The seasonal autoregressive integrated moving average model (SARIMA) always deals with seasonality. Seasonality means a pattern of increases and decreases in a series' mean level that stays throughout time (Jebb et al., 2015). For the non-stationary pattern of the datatype, the SARIMA model is meaningful and gives a better result. According to Y. Wang et al. (2022), considering the seasonal factors, the SARIMA $(p, d, q)(P, D, Q)_m$ model can be represented as follows:

$$\left(1-\sum_{i=1}^{p}\emptyset_i L^i\right)\left(1-\sum_{i=1}^{P}\Phi_i L^{mi}\right)(1-L)^d(1-L^m)^D y_t = \alpha_0 + \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\left(1+\sum_{i=1}^{Q}\Theta_i L^{mi}\right)\varepsilon_t \quad (2.4)$$

Where:

P = seasonal autoregressive order,

D = seasonal difference order,

Q = seasonal moving average order,

m = number of periods (monthly data *m* = 12, quarterly data *m* = 4, hourly data *m* = 24),

$\emptyset_i$= non-seasonal autoregressive coefficient,

Φ_i= seasonal autoregressive coefficient,

θ_i= non-seasonal moving average coefficient,

Θ_i= seasonal moving average coefficient.

If, for example, the variance increases with time shown on the data's time plot, we should apply variance-stabilizing transformations and differences (Siami-Namini et al., 2018). To measure the amount of linear dependence between observations by using the autocorrelation function (ACF) in a time series are divided by a lag *p,* and the partial autocorrelation function (PACF) to determine how many autoregressive terms *q* are needed and inverse autocorrelation functions (IACF) for detecting over differencing, the preliminary values of autoregressive order *p*, the order of differencing *d*, the moving average order *q* and their corresponding seasonal parameters *P, D,* and *Q* can be identified (Siami-Namini et al., 2018). *m* is denoted by the lag or period of repeating seasonal pattern.

2.4.5 Non-parametric Methods

The non-parametric approach is a statistical method that makes no assumptions about the sample's characteristics (its parameter) or whether the observed data is quantitative or qualitative. These approaches use data to identify the structure and parameters. According to Habtemichael & Cetin (2016), when compared to parametric models, non-parametric models perform better because they learn more from complicated data and adapt to its pattern. The non-parametric patterns recognition-based algorithms, which are a subset of these, appear to be more suitable since they are good at finding similar traffic conditions needed to provide predictions (Habtemichael & Cetin, 2016). These models often demand a large amount of data than their parametric equivalents. This category involves two different methods:

(i) Non-parametric, and (ii) Neural networks (NN) method

- **Non-parametric Methods:** The approaches are standard ML methods for traffic forecasting that depend on pattern recognition and chaotic systems concepts (Smith et al., 2002). The non-parametric method aims to characterize data with behavior like the current traffic state for the predicting interval (Buroni, 2021). The nonparametric technique has several advantages, including a simple formulation, the lack of assumptions about traffic state transitions from one period to the next, and ultimately, the ease of modeling multivariate environments (Buroni, 2021). Deep learning neural network models have been used to forecast future traffic speeds, travel time, and traffic flows in several research articles (Duan et al., 2016; X. Ma et al., 2015; Zhao et al., 2017), is one example of non-parametric approaches that can manage the stochastic pattern and the noise in traffic input data (Abduljabbar et al., 2021).

- **Neural Networks (NN) Methods:** In the domains of AI, machine learning, and deep learning, neural networks mimic the activity of the human brain, allowing computer programs to spot patterns and solve common problems (IBM Cloud Education, 2020). Artificial neural networks (ANNs) and simulated neural networks (SNNs) are types of neural networks that are used in deep learning techniques (IBM Cloud Education, 2020). According to Dougherty (1995), NNs have already been used in a wide range of transportation applications, including driver behavior, autonomous vehicles, parameter estimation, pavement maintenance, vehicle detection classification, traffic pattern analysis, freight operations, traffic forecasting, transportation policy and economics, traffic control, and so on. Recurrent Neural Networks (RNN), Long-Short Term Memory (LSTM), and Gated Recurrent Units (GRU) are some mostly used neural networks in traffic forecasting. Neural networks (NN) are based on the human brain.

2.4.6 LSTM Neural Network

Traditional neural networks presupposed that each input vector observation is independent of the others (Sagheer et al., 2021). As a result, the sequential information present in time series data cannot be used by the conventional neural network (Sagheer et al., 2021). Recurrent neural network (RNN) procedures are used to produce a series of data so that each observation is believed to be reliant on the prior ones, in contrast to traditional neural network approaches (Das et al., 2021). As Recurrent Neural Networks (RNNs) have a long-term dependence problem, that's why Long Short-Term Memory (LSTM) networks are created to solve this problem (due to the vanishing gradient problem). LSTM network is used to classify and prediction of time series data (Saha et al., 2022). LSTMs have feedback connections to previous data that help to process time series data by retaining previous important information from previous data. Due to the model's ability to learn extensive sequences of data, LSTM has evolved into a trending method for time-series forecasting (Tran et al., 2022). One benefit of LSTM is that it can capture both long- and short-term seasonality, such as weekly and yearly trends (Tran et al., 2022). According to Kashyap et al. (2022); Tran et al. (2022), LSTM is still the top choice for the use of deep learning algorithms to address the traffic speed prediction problem. Several 'gates' are used in LSTMs to control how data in a sequence enters, is stored in, and departs the network.

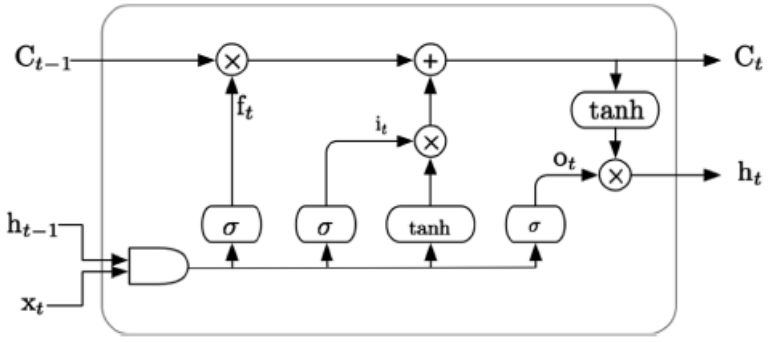


Figure 2.4: The repeating module structure of LSTM (Source: Boukerche & Wang, 2020)

LSTM contains two gates inside of one initially, the input gate and the output gate, compared to standard RNN (Boukerche & Wang, 2020). When the input time series lengthened, the two-gates LSTM was restricted (Boukerche & Wang, 2020; Gers et al., 2000). The two-gate LSTM models a linear process; hence the non-linear feature of the original dataset is ignored (Boukerche & Wang, 2020). To get the solution to this problem, Gers et al. (2000) introduced a third gate to the two-gate LSTM, termed the forget gate. Different gates of LSTM do different kinds of work. The input gate decides whether to write data to the cell state, the output gate decides what data to pass as the output hidden state, the forget gate decides whether to remove data from the cell state, and the forget gate decides whether to write data at

all (Sagheer et al., 2021). These three gates structure of LSTM have been shown in Figure 2.4. The following equations illustrate how the LSTM cell works.

$$i_t = \sigma(W_{xi}X_t + W_{hi}h_{t-1} + b_i) \quad (2.5)$$

$$o_t = \sigma(W_{x0}X_t + W_{h0}h_{t-1} + b_0) \quad (2.6)$$

$$f_t = \sigma(W_{xf}X_t + W_{hf}h_{t-1} + b_f) \quad (2.7)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}X_t + W_{hc}h_{t-1} + b_c) \quad (2.8)$$

$$h_t = o_t \odot \tanh(c_t) \quad (2.9)$$

Where:

i_t = LSTM gating for the cell state to input information

o_t = LSTM gating for the cell state to output information

f_t = LSTM gating for the cell state to forget information

c_t = Cell memory state vector

h_t = Hidden state vector

σ = Sigmoid function

tanh = Activation function (we can use also RELU, Leaky RELU activation function for better performance)

X_t= Input vector

$W_{xi}, W_{hi}, W_{x0}, W_{h0}, W_{xf}, W_{hf}, W_{xc}, W_{hc}$= Linear transformation matrices whose parameters are needed to be learned for each single gate and cell memory

b_i, b_0, b_f, b_c = Corresponding bias vector

We will utilize the gate's output to determine how much information we want to save, which will be a value between 0 and 1 (Boukerche & Wang, 2020). According to Sagheer et al. (2021), the simple expression of LSTM could be shown in one form as follows:

$$h_t, c_t = LSTM(x_t, h_{t-1}, c_{t-1}) \quad (2.10)$$

To integrate the benefits of each LSTM, the given model consists of a stack of LSTM blocks (or layers) arranged sequentially and coupled in a deep architecture (Pascanu et al., 2013; Sagheer et al., 2021). In the stacked LSTM model, several layers of LSTM are stacked, where one layer gives information to the next layer, and the final LSTM layer will give the output. LSTM is a widely used model for traffic flow and speed prediction. Many studies used this algorithm either in single or combined form for the prediction of traffic flow and speed.

2.4.7 Gated Recurrent Unit (GRU)

Gated Recurrent Units (GRUs) were first proposed by Chung et al. (2014). It is considered a simpler version of the LSTM cell and provides better network performance while requiring less training time (ArunKumar et al., 2022; Chung et al., 2014). The GRU internal unit and LSTM internal unit are almost the same (Chung et al., 2014; Mateus et al., 2021), the only exception is a single update port is created

by the GRU by combining the incoming port and the forgetting port in the LSTM model (Mateus et al., 2021). The hidden and cell states are combined into one state by GRU (ArunKumar et al., 2022). As a result, GRU is a popular and shortened version of the LSTM cell and has half as many gates as LSTM does overall (ArunKumar et al., 2022). There are two gates in the GRU unit the update gate and the reset gate. The reset gate regulates whether the current state should be mixed with the previous information, while the update gate controls how much of the state data from the previous period is maintained in the present state (Cho et al., 2014; Mateus et al., 2021).

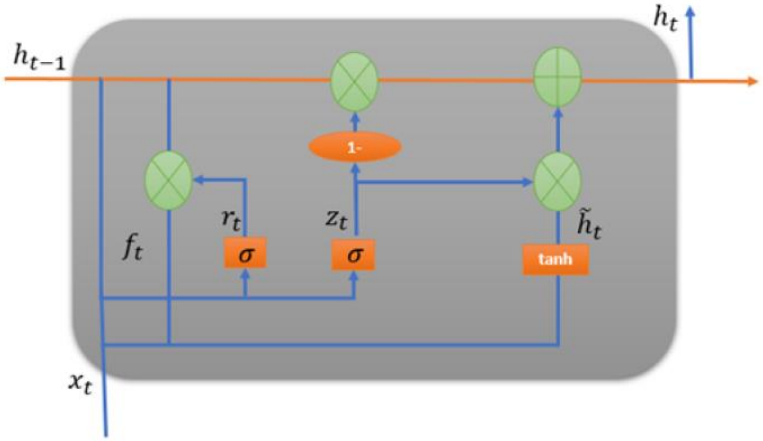


Figure 2.5: The cell structure of Gated Recurrent Unit (GRU) (Source: Mateus et al., 2021)

According to Mateus et al. (2021), figure 2.5 is showing the cell structure of GRU neural network. Following equations will show different gate's control mechanism:

$$z_t = \sigma(x_t W^z + h_{t-1} U^z + b_z) \tag{2.11}$$

$$r_t = \sigma(x_t W^r + h_{t-1} U^r + b_r) \tag{2.12}$$

$$\tilde{h}_t = \tanh(r_t \times h_{t-1} \mathrm{U} + x_t W + b) \tag{2.13}$$

$$h_t = (1 - z_t) \times \tilde{h}_t + z_t \times h_{t-1} \tag{2.14}$$

Where:

W^z, W^r, W = The weight matrices for the corresponding connected input vector,

U^z, U^r, U = The weight matrices of the previous time step,

b_z, b_r, b = Corresponding bias vector,

σ = The logistic sigmoid function,

r_t = The reset gate,

z_t = The update gate, and

$\tilde{h}_t$ = The candidate hidden layer.

2.4.8 Time-series Prediction

Time series (definition): A time series $T = [x_1, \dots, x_n]$ is a regular recording of data points at a fixed interval of time Δt following the preceding point, where each data point is either a real value or a vector of real values (Weber et al., 2021). Univariate time series (definition): A time series where, $x_i \in \mathbb{R}$ is called univariate time series, and in univariate time series there is a single time-dependent variable, e.g.,

periodic sensor readings from one specific sensor (Weber et al., 2021). Multivariate time series (definition): A time series where each x_i is a *d-dimensional* vector of real values $(x_i^1, \ldots, x_i^d), x_i^j \in \mathbb{R}$ is called multivariate time series, and multivariate time series combine multiple time-dependent variables e.g., sensor modalities, sensor channels of different sensor placed in several locations (Weber et al., 2021).

A time series prediction system consists of predicting the system's performance in the future based on data from the present and past status of the system (Sagheer et al., 2021). It is important to consider both the past and present status of the system for time-series prediction. Time series forecasting (definition): $T = [x_1, \ldots, x_n]$, time series forecasting denotes the problem of predicting the next value of the sequence x_{n+1} or the next m values $x_{n+1}, \ldots, x_{n+m}$ in univariate time series forecasting (Weber et al., 2021). It denotes the problem of forecasting the next value(s) in at least one of the *d* dimensions in multivariate time series forecasting (Weber et al., 2021). For traffic and weather prediction, time-series prediction techniques are used. Time-series works by considering a sequence of observations *(x)* measured in sequential order of a variable of interest at time *(t)* (Sagheer et al., 2021). According to Sagheer et al. (2021), a sequence of time series data can be illustrated as:

$$X = (x_t; t = 1, \ldots, N) \tag{2.15}$$

Where, *X* is the time series sequence or order, *t* denotes as the observation's time for *N* observations. According to Sagheer et al. (2021), there are three types of time series forecasting techniques. 1) Single Step ahead Forecast (SSaF), 2) Multi-Step ahead Forecast (MSaF), and 3) Hierarchical time series forecasting.

- **Single Step ahead Forecast (SSaF):** It is an easy method in time series prediction as it predicts only the next future value. According to Masum et al. (2018); Sagheer et al. (2021), at a time t + 1 is executed by passing the current and the past observations t, t – 1, t – 2, …., t – N to a selected model,

$$F(t + 1) = M(o(t), o(t - 1), o(t - 2), \ldots, o(t - N)) \tag{2.16}$$

Where, *F (t + 1)* is the prediction at time *(t + 1)*, *M* is the selected model, and *o(t)* is an observation at time *t*.

- **Multi Step ahead Forecast (MSaF):** Multi-step ahead forecast (MSaF) requires forecasting the next *H* sequence of future time series values $[y_1, y_{2,\ldots,} y_N]$ with *N* observations and with *H* > 1 being the prediction horizon (Sagheer et al., 2021). There are three main techniques in MSaF, where one is the recursive technique, another one is the direct technique and the last one is the multi-input multi-output technique (MIMO) (ben Taieb et al., 2012; Sagheer et al., 2021).

1. The recursive technique is the oldest one used to treat MSaF where a single model *f* is trained to perform only one step ahead forecast (Sagheer et al., 2021). Then, the value just forecasted is used as part of the following input variable to predict the next step with the same step-ahead model (Sagheer et al., 2021). According to Sagheer et al. (2021), considering the trained one-step ahead model $\hat{f}$ and the following are the given forecasts:

$$\hat{y}_{N+h} = \begin{cases} \hat{f}(y_N, \ldots, y_{N-d+1}), & if\ h = 1 \\ \hat{f}(\hat{y}_{N+h-1}, \ldots, \hat{y}_{N+1}, y_N, \ldots, y_{N-d+1}), & if\ h \in 2, \ldots, d \\ \hat{f}(\hat{y}_{N+h-1}, \ldots, \hat{y}_{N-d+h}) & if\ h \in d+1, \ldots, H \end{cases} \tag{2.17}$$

Where, d represents the number of past values used to predict the future value. These predications are used as an input for making further predictions (ben Taieb et al., 2012).

2. The direct technique also called an independent technique which depends on forecasting each horizon independently from the others (Hamzaçebi et al., 2009; Sagheer et al., 2021; Sorjamaa et al., 2007). According to Sagheer et al. (2021), the forecasts equation is:

$$\hat{y}_{N+h} = \hat{f}_h(y_N, \dots, \dots, y_{N-d+1}) \quad (2.18)$$

The predictions are protected from the error accumulation created by the recursive method since the direct strategy does not compute the forecasts using any approximations (Sagheer et al., 2021).

3. Multi-input multi-output (MIMO) technique directly derives from data, the dependence between previous observations and future values (Sagheer et al., 2021), instead of being scalar values in this instance, the future values are a vector of values from the time series itself (Bontempi, 2008; Chen & Wang, 2010; Sagheer et al., 2021). The advantages of the MIMO technique are its ability to conserve the temporal stochastic dependency included in the forecasted future values as it uses only one multi-output model (Sagheer et al., 2021). It requires that all horizons can be predicted using the same learning method and unified structure, and that's why the predictions are gained in one step at once by using the following equation:

$$[\hat{y}_{t+H}, \dots, \hat{y}_{t+1}] = \widehat{F}(y_N, \dots, y_{N-d+1}) \quad (2.19)$$

Where, $\hat{F}$ is a multi-output model.

2.5 Transferability and Transfer Learning

In the past, short-term traffic forecasting was done by statistical models and machine learning techniques. Advanced machine learning approaches have been widely applied in this study area due to their capacity to handle missing data and their computational efficiency, dependency, and resilience (J. Li et al., 2021). The frequency of inadequate data over an entire network is a difficult challenge, that hinders the deployment of these technologies in the real world (J. Li et al., 2021; Vlahogianni et al., 2014). It is unusual to have access to all the city's connections' historical training data, which is essential for model training (Abadi et al., 2014; J. Li et al., 2021). Statistical approaches have been used extensively in the previous several decades to investigate the transferability of various transportation models across time periods and several locations, such as travel demand models and route choice models (Bekhor & Prato, 2009; J. Li et al., 2021). For short-term travel time prediction, Luan et al. (2018) evaluated the link-to-link transferability of three distinct machine learning models. Many machine learning and data mining techniques assume that the training and testing data must be in the same feature space with the same distribution (J. Li et al., 2021; Pan & Yang, 2010), and that model calibration requires a large amount of prior training data (J. Li et al., 2021).

Transfer learning is a potential method for dealing with data scarcity, where a model is trained for one task and is reused and/or altered for another (Mallick et al., 2020). In the text domain, transfer learning is commonly used for image classification, sentiment analysis, and document classification (Mallick et al., 2020; Tan et al., 2018; Zhuang et al., 2019). When labeled training data is sparse, Transfer Learning solves the machine learning challenge (Pan & Yang, 2010; L. Wang et al., 2019). To undertake chain store site suggestions, Guo et al. (2018) proposed a cross-city transfer learning system using collaborative filtering and AutoEncoder. In the field of computer vision and natural language processing, the study of transfer learning is vast (Pan & Yang, 2010; B. Wang et al., 2019). In practice, replacing or expanding traffic-related systems necessitates new sensors, which lack historical data for model training (B. Wang et al., 2019). In this case, using the fine-tuning approach, gained knowledge from previously used sensors might aid in predicting accuracy for freshly placed sensors, but for short-term traffic prediction details research on model fine-tuning is still unknown (B. Wang et al., 2019).

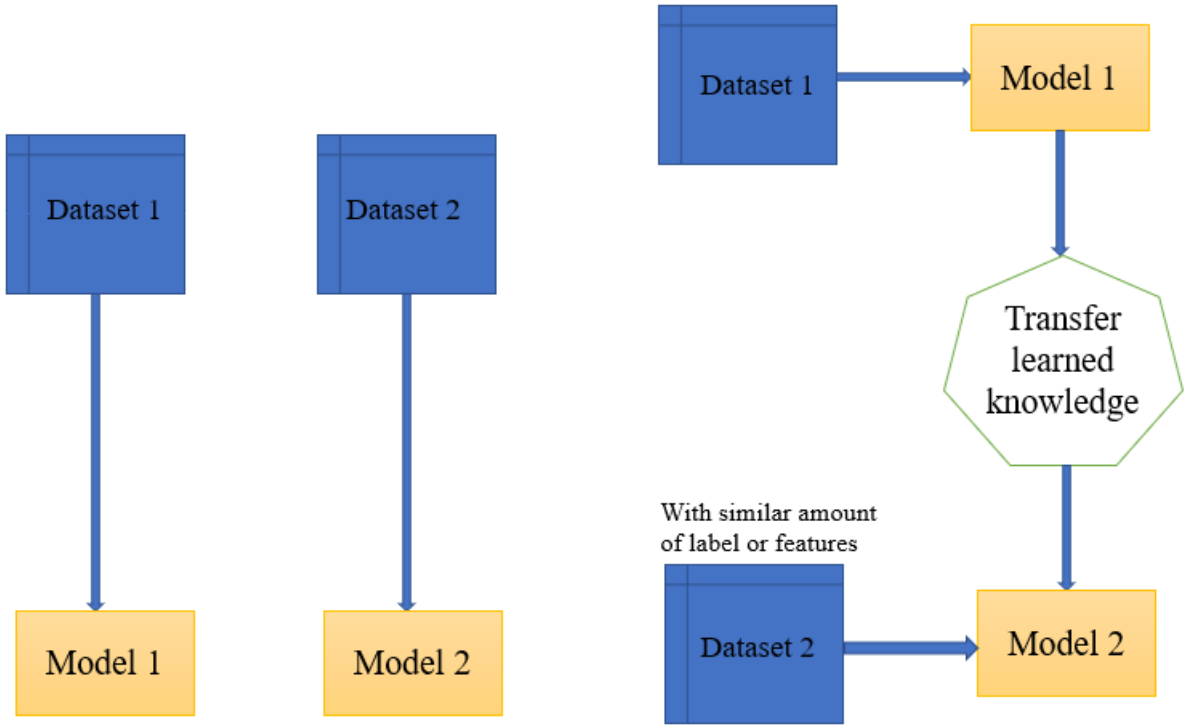
Transfer learning techniques' greater generalization has the potential to overcome insufficient training data issues by enhancing the transferability of a traffic prediction model, and the faster training process

can greatly reduce the associated computing cost (J. Li et al., 2021). According to Saha et al. (2022), three method makes up the transfer learning approach: parameter transfer, domain adaptation, and multi-task learning. The parameter transfer approach applies the learned parameters from the source data and builds a model for the target data using this information (Saha et al., 2022). This method can be useful in two ways: modeling smaller datasets and reusing knowledge for rapid learning (Saha et al., 2022). According to J. Li et al. (2021), A feature space *X* and a marginal probability distribution *P(X)* over the feature space comprise a domain $\mathcal{D}$, where *X* is the learning sample. The relationship between feature space *X* and learning sample *X* can be described as $X = \{x_1, \dots, x_n\} \in X$, where x_i is the ith vector in the existing learning sample *X* with different features and *X* has all the features (J. Li et al., 2021). Given a domain $\mathcal{D} = \{X, P(X)\}$, a task *T* is made up of a label space *Y* containing all labels y_i corresponding to learning vectors x_i and a conditional probability distribution *P(Y|X)* (i.e., the objective prediction function *f(·)*, which is generally learned from the training pairings $x_i \in X$ and $y_i \in Y$. According to Pan & Yang (2010); J. Li et al. (2021), Transfer learning tries to improve the learning of the target conditional probability distribution $P(Y_T|Y_T)$ in T_T with the knowledge learned from $\mathcal{D}_S$ and T_S, where $\mathcal{D}_S \neq \mathcal{D}_T$ or $T_S \neq T_T$, given a source domain $\mathcal{D}_S$, a matching source task T_S, as well as a target domain $\mathcal{D}_T$ and a target task T_T.

According to J., Li et al. (2021), the concept of $\{\mathcal{D}_S, T_S\}$ and $\{\mathcal{D}_T, T_T\}$ which are included in the neural network-based transfer learning technique concept, are defined as follows:

- A source dataset A: expressed by $\mathcal{D}_S = \{X_S, P(X_S)\}$, where the learning samples in X_S comprise almost all of the features in feature space *X* and the distribution *P(X)* throughout feature space.
- A source task A: expressed by $T_S = \{Y_S, f_S(\cdot)\}$, where a sufficient training pairs of $x_i \in X_S$ and $y_i \in Y_S$ to accurately estimate the objective prediction function $f_S(\cdot)$.
- A source network A: expressed by $f_S(\cdot)$, where each network parameter is precisely estimated to allow $f_S(\cdot)$ to make accurate predictions.
- A target dataset B: expressed by $\mathcal{D}_T = \{X_T, P(X_T)\}$, where the learning samples in X_T are limited that it prevent $\{X_T, P(X_T)\}$ from accurately representing $\{X, P(X)\}$.
- A target task B: expressed by $T_T = \{Y_T, f_T(\cdot)\}$, where the insufficient training pairs of $x_i \in X_S$ and $y_i \in Y_S$ cannot accurately estimate the objective prediction function $f_S(\cdot)$.
- A target network B: expressed by $f_T(\cdot)$, where the majority of the network's parameters are not accurately estimated, it is uncertain to ensure the correctness of the predictions.

Given the above descriptions, J. Li et al. (2021) used three transfer learning strategies in the deep neural network. One is transfer-without fine-tuning, another is fine-tuning with freezing transferred layers, and the last is fine-tuning without freezing transferred layers. In conventional learning systems, the individual dataset will create an individual model. But in transfer learning, one dataset will learn the knowledge and send this pre-learned knowledge to another dataset, and that pre-learned knowledge will create another model. This technique will reduce computational burden and can be used for a small dataset of the same features. The following figure describes this typical transfer learning technique.



(a) (b)

Figure 2.6: Representation of (a) Typical traditional machine learning model, (b) Typical transfer learning model (Source: Sagheer et al., 2021)

Depending on the availability of labels in the target and source dataset, there are different problem settings in transfer learning (Weber et al., 2021). Transfer learning is used when the target domain has a little quantity of labeled data, which is frequently insufficient for independently training an appropriate target model (Weber et al., 2021). According to Pan & Yang (2010); Weber et al. (2021), transfer learning solution approaches divided into four categories: instance-based, feature-based, parameter-based, and relational-knowledge-based. The selection or reweighting of samples from the source domain occurs during instance-based transfer (Weber et al., 2021). Based on the assumption that examples from the source domain are comparable to the target domain examples, this approach is more or less helpful for training models (Weber et al., 2021). By transferring feature representations, data can be transformed into a common feature space, allowing learning to be done on features that reflect characteristics from both domains (Weber et al., 2021). The more generic term model-based transfer also refers to parameter transfer, which reuses parameters from a model that has already been trained in the source domain for the target model-building process (Weber et al., 2021). Finally, time series cannot use relational knowledge transfer since such is reserved for relational domains (Weber et al., 2021).

2.6 Research Gaps Related to Literature Review

It is clear from the literature review that traffic parameters like flow, speed, and VMT forecasting are core tasks for reducing traffic congestion and real-time traffic information. Most of the research did traffic flow and other parameters forecasting and their accuracy using different statistical and machine learning models. They analyzed traffic forecasting models, and transfer learning models in some literature, but there are some research gaps in this field of research.

- Spatial regional attributes (shopping malls, governmental organizations, stadiums, etc.) are not considered with temporal features for the traffic forecasting (Cheng et al., 2012; Du et al., 2020; Yao et al., 2019).

- Factors (weather conditions, incidents, signal settings, etc.) are not considered during traffic forecasting by transfer learning (J. Li et al., 2021).

According to the above research gaps, in this thesis, we will use weather factor data with different features for finding the influences on traffic flow forecasting. We will integrate weather data with traffic data for traffic flow forecasting by transfer learning.

3 Methodology

This chapter discusses data collection, preprocessing of data, and model building using univariate ARIMA, multivariate GRU, LSTM, and transfer-learning strategies. The initial part of this chapter describes how datasets are collected, the next step describes how data are preprocessed, how these data are cleaned and filtered, and next it describes model buildings. The next chapter will be data analysis of several roads in California State, USA. The further chapter is about results and discussions, and the next chapter is about conclusions and future research in this field.

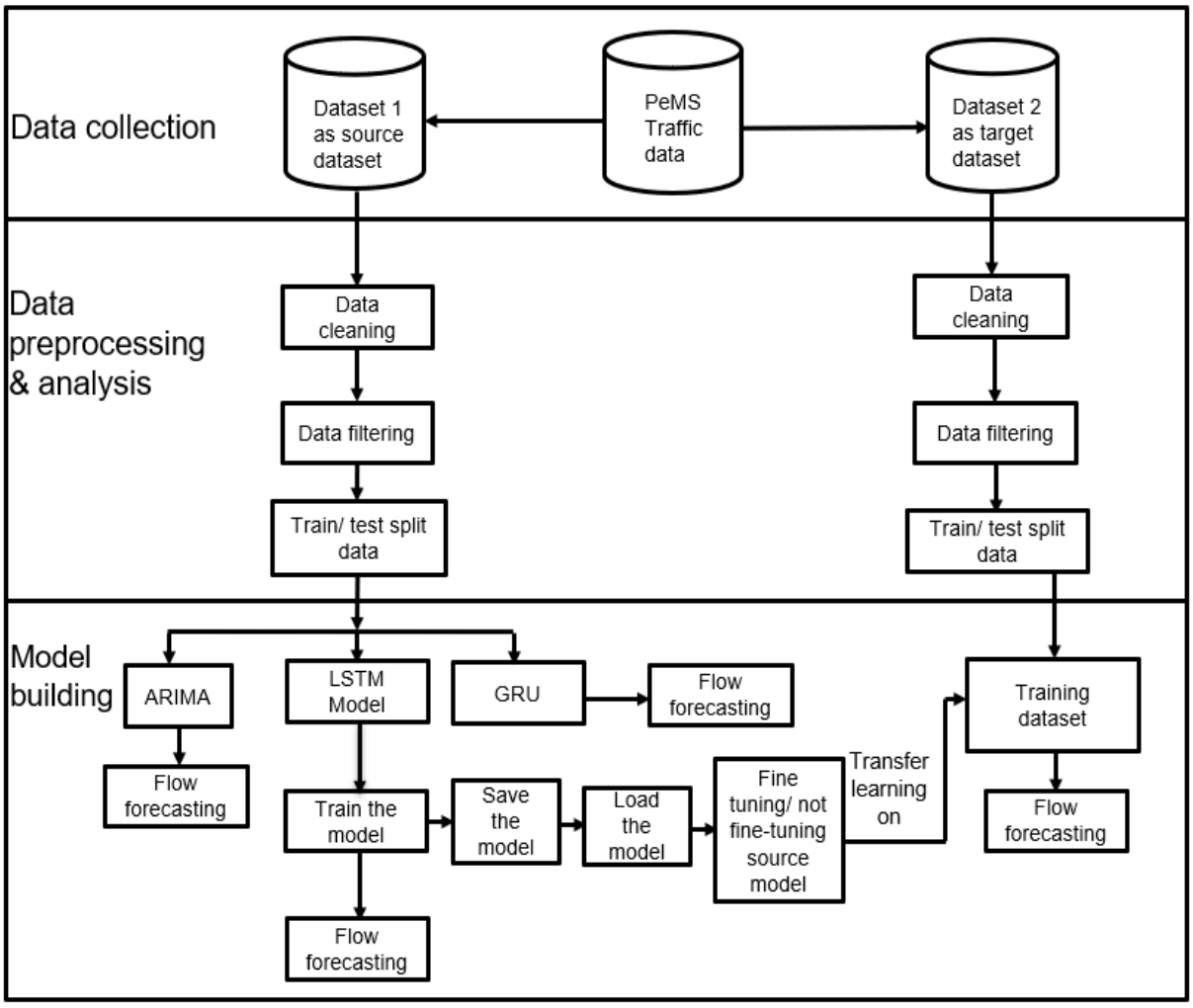


Figure 3.1: Methodological flow chart

3.1 Data Collection

Datasets are collected from the prestigious California Transportation Performance Measurement System (PeMS, 2022) database. In this database, all traffic data are well organized. Traffic data are collected by many stations situated in different parts of districts of California State. California state is separated into 12 different districts for traffic data collection. For collecting traffic data, they used loop detectors, radars, magnetometers, and many other detectors for collecting traffic count data, traffic flow data, traffic speed data, occupancy data, VMT data, and many more. Currently, a total of 46,335 detectors are available and working in all California districts where there are in total 41,236-mile total distances covered for the traffic data. A total of 18,986 stations and a total of 18,348 traffic census stations are available for collecting traffic data. PeMS also check the detectors' health and always tries to maintain the quality of detectors in all California districts. They also analyze the data on where the detector should be changed or repaired. The granularity of PeMS traffic data is 5 minutes, 1 hour, and 1 day.

3.1.1 Caltrans PEMS Data Description

Table 3.1: Caltrans PeMS spatial description (Source: PeMS, 2022)

District	Name	Directional distance (mile)	Number of stations	Number of detectors
1	Northwest	1889.3	0	0
2	Northeast	3463.2	0	0
3	North central	3004.5	1544	3177
4	Bay area	2869.0	4027	11603
5	Central coast	2333.2	561	1033
6	South central	11579.3	690	1790
7	LA/Ventura	2318.2	4895	11241
8	San Bernardino /Riverside	3890.4	2100	5289
9	Eastern Sierra	4598.2	0	0
10	Central	2658.9	1253	2385
11	San Diego/Imperial	2053.9	1440	4153
12	Orange County	577.9	2511	5728

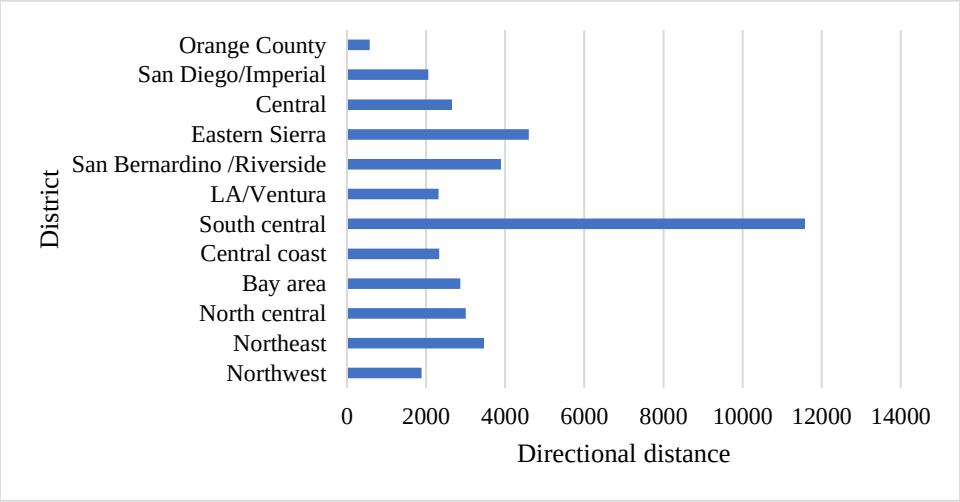


Figure 3.2: Directional distance covered by each district (Source: PeMS, 2022)

Figure 3.2 illustrates the directional distance covered by each of the 12 districts. South Central covers the maximum directional distance, and the lowest number of directional distances are covered by Orange County. Eastern Sierra covers the second-highest distance, and all other districts cover distances from 2000 miles to 3000 miles range.

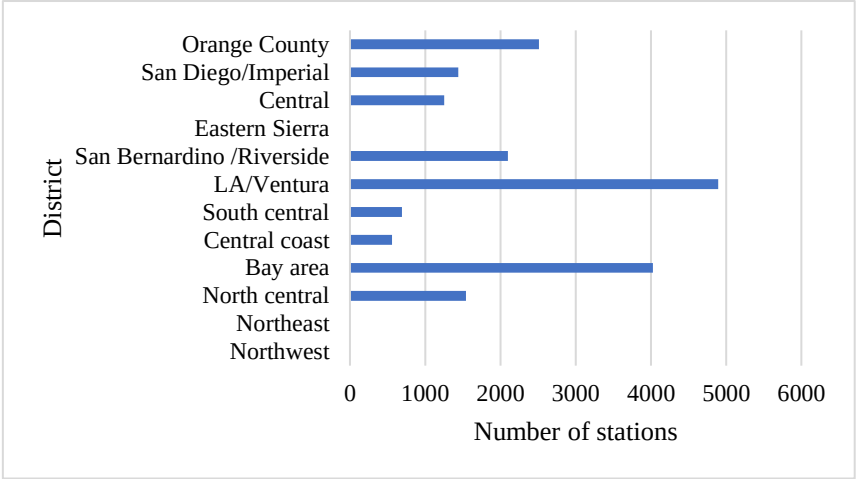


Figure 3.3: Number of stations covered by each district (Source: PeMS, 2022)

Figure 3.3 illustrates the number of stations covered by every 12 districts of California state. Here, there is no station available for three districts, which are Northwest, Northeast, and Eastern Sierra. LA/Ventura has the most stations, followed by Bay Area with the second most. The third highest number of stations are covered by district 12 or Orange County and then San Bernardino/ Riverside. District 3 North Central is the 5th largest district in terms of stations cover.

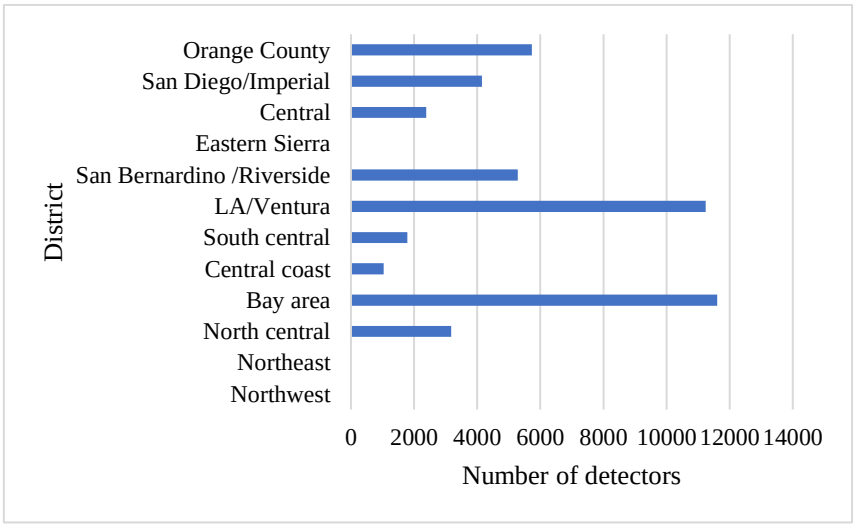


Figure 3.4: Number of detectors covered by each district (Source: PeMS, 2022)

Figure 3.4 illustrates the number of detectors covered by every 12 districts of California state. It shows that there is no detector available for district 1 which is Northwest, district 2- Northeast, and district 9- Eastern Sierra. Bay Area and LA/ Ventura cover the highest number of detectors. District 12- Orange County covers the third highest number of detectors. District 3- North Central covers approximately 3000 detectors and district 6- South Central covers approximately, 1800 detectors. These detectors collect traffic data directly from the road.

3.1.2 California District 3 Data

California district 3 is called the North-Central part of California state. This district consists of Sacramento, Roseville, Colusa, Yuba City, Willows, Chico, Oroville, Placerville, etc. city of California. This district covers a total of 3,004.5 miles of roadway, where a total of 603 controllers, 1,540 stations, 3,174 detectors, and 1,728 traffic census stations are available for collecting different traffic data.

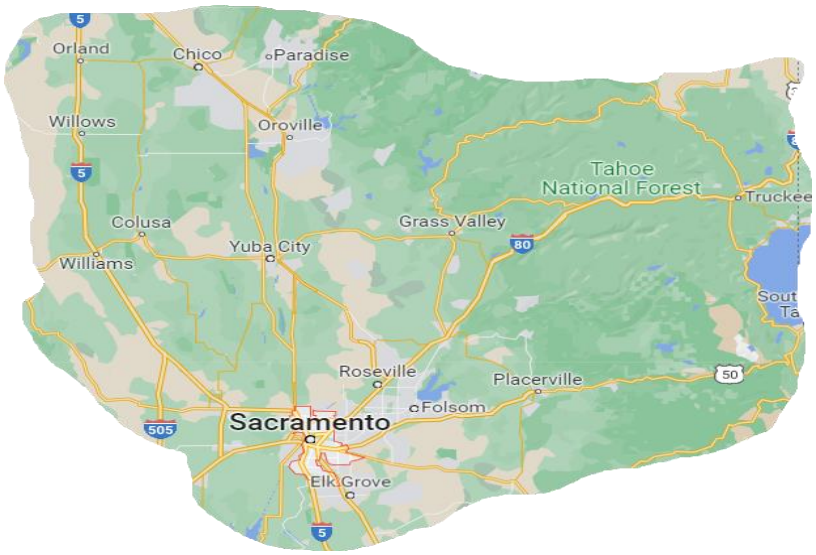


Figure 3.5: California district 3 traffic approximate map (Source: Google Maps)

We collected a total of 6 months of data on California district 3 from April 1, 2019, to September 30, 2019, for forecasting traffic flow. The granularity of these data is 1 hour. These data cover Mainline (ML), High Occupancy Vehicle Lane (HOV), On-Ramp (OR), Off-Ramp (Off-R), Freeway-Freeway (Fwy-Fwy) connector road.

3.1.3 California District 6 Data

California district 6 is called the South-Central part of California state. This district consists of Fresno, Visalia, Bakersfield, Madera, etc. city of California. Currently, this district covers 11,579.3 miles of directional distance, where 342 controllers, 690 stations, 1,790 detectors, and 1,582 traffic census stations are available for collecting traffic data. We collected 1-month data for transfer learning and performance of traffic flow forecasting.

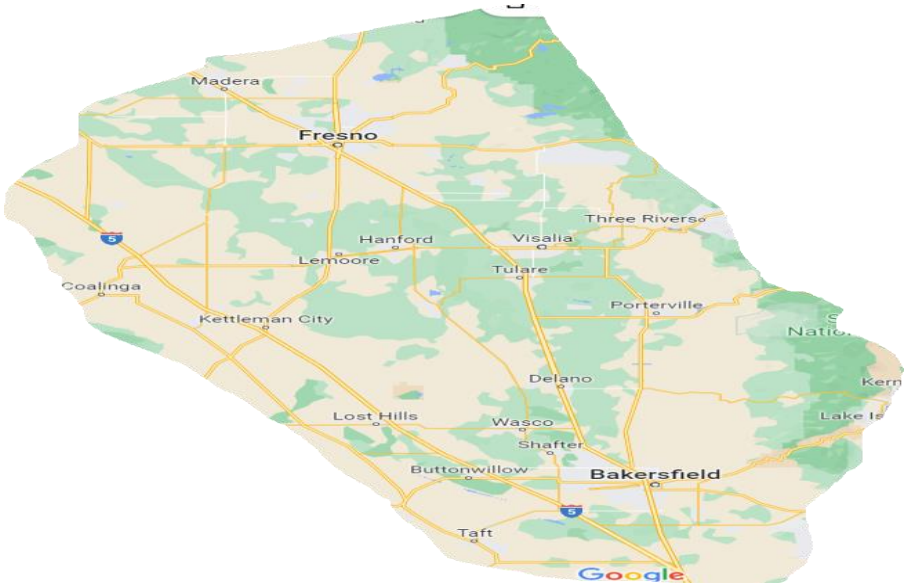


Figure 3.6: California district 6 traffic approximate map (Source: Google Maps)

3.2 Data Preprocessing

Data preprocessing include data cleaning and data filtering. Each of these sub-chapters is described below.

3.2.1 Data Cleaning

In both California district 3 and district 6 datasets, there are some missing values in the total flow data. This happened due to faulty detectors or there are other reasons. Speed and occupancy data are completely missing on the on-ramp and off-ramp roadways. According to PeMS (2022), either they did not collect the speed and occupancy data, or they did not calculate these data for on-ramp and off-ramp. On freeway-freeway connector roads, sometimes occupancy data are available, but speed data are completely missing. Approximately 35% of data are missing in flow, occupancy, and speed data, where the majority of data are on-ramp, and off-ramp link speed and occupancy data. A total of 65% of data are completely in good shape for traffic flow, speed, and occupancy data on mainline (ML), and high occupancy vehicle (HOV) lane data in district 3 data. We just removed on-ramp, off-ramp, and freeway-freeway connector road data for univariate and multivariate time series analysis and model building.

For district 6 data, there are approximately 85% of the data are in good shape, rest 15% data has missing data for the on-ramp, off-ramp, and freeway-freeway connector roads. For the mainline and HOV lane, almost all the data of total flow, average speed, average occupancy, and each lane parameters are available. In that case, we will consider only mainline (ML) and high occupancy vehicle (HOV) link-type data for our model building. We will select individual station data-specific spatial and temporal data of district 3 data as the source model dataset and district 6 data as the target model dataset in transfer learning models.

3.2.2 Data Filtering

District 3 datasets are filtered for six months duration from 2019. We took data from April to September 2019 data for the model building. Data from District 6 are filtered to 1-month, and the pre-trained model will be used for transfer learning based on district 3 datasets. We have selected 2-lane, 3-lane, and 4-lane link data in the dataset. Total flow, average occupancy, and average speed of all lanes for the individual link are given, also individual lane flow, average speed, and average occupancy are given. We filtered data to hourly data, then we took one-month data of district 3 for individual link type data for analysis and performance. We made different datasets for each of the 3 different link types. There are many features available in traffic datasets. From the raw dataset, we have created another feature named ‘lane type’. In this lane type column, relevant rows of each lane are mentioned by this lane splitting. Then we can find out how many link types are available in the datasets. From 50 different features, we just selected three traffic features that will be used for the multivariate time series modeling. Total flow, average occupancy, and average speed for each link type are taken for the model building.

3.3 Model Building

3.3.1 Univariate ARIMA

For the ARIMA model structure, we take total flow as an endogenous variable, and average occupancy and average speed are exogenous variables in our multivariate time-series forecasting. When we do traffic flow forecasting with the multivariate ARIMA model, it does not perform well. In that case, we will not use the multivariate ARIMA model. Hence, we will check the univariate ARIMA model only with our target variable ‘Total Flow’.

A Dicky-Fuller-Test (ADF) test is done for checking if the data time series is stationary or not. In the ADF test, the *p*-value will evaluate the stationarity of the time series data. If the *p*-value is less than 0.05, then the series is stationary. Otherwise, the series is non-stationary. In our studies, all three roads’ data in all shapes of datasets, *p*-values are less than 0.05. Our selected 3 road datasets (Station id: 313190, 314042, and 319351) from District 3, California State, are stationary. There will be no differencing needed before model building. We can directly use the training dataset for model building and evaluations. In this case, differencing value, $d = 0$. Seasonality and trend-line are also checked.

From (*p*, *d*, *q*) order, we will get *p* and *q* value from partial autocorrelation function (PACF) and autocorrelation function (ACF) plots.

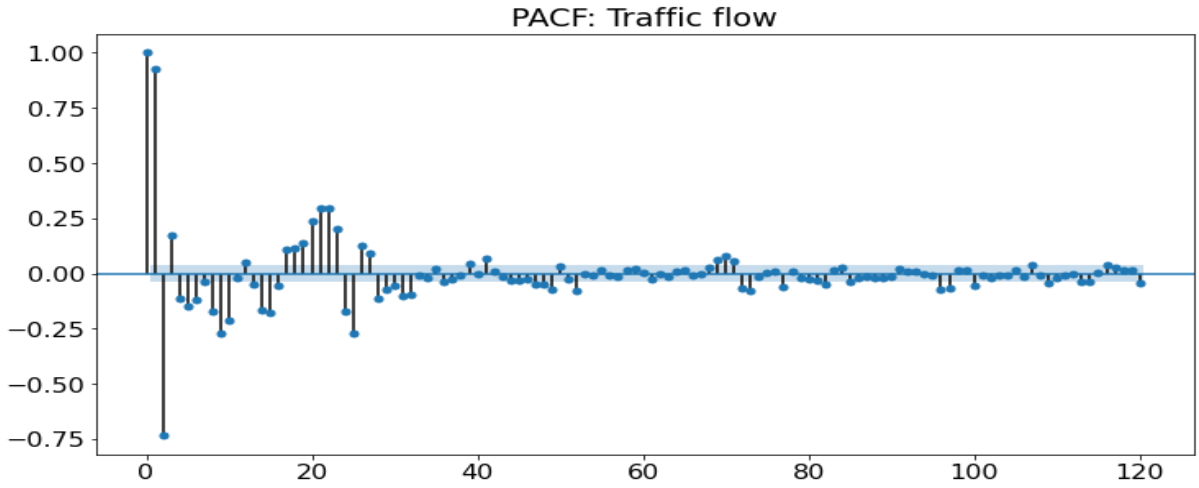


Figure 3.7: PACF plot for 4 lane road (station id 319351) data

As per the PACF plot, the *p* value will be found. Here in the figure, lags 2 is significant as there is an exponential decay available here. This time-series p-value will be 2.

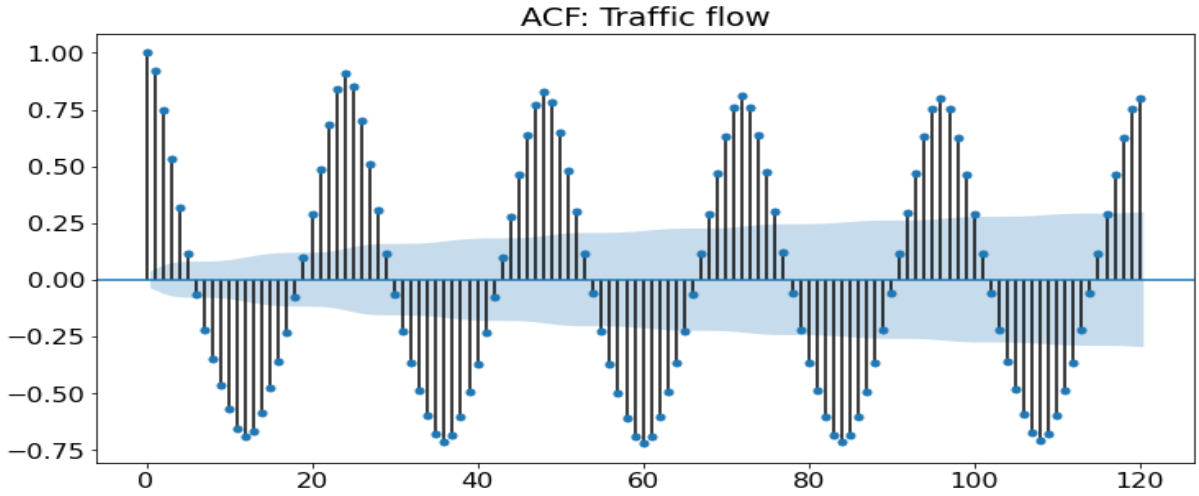


Figure 3.8: ACF plot for 4 lane road (station id 319351) data

By ACF plot we can get the *q*-order value of the ARIMA model. Here, the figure shows the ACF plot for the data of Placer County, district 3 station 319351. It is evident that here after 2 orders that means at order 2 the series is significant as per the plot as $q = 2$ order lies in a significant level. We will select the *q* value as 2 for this data type. From the Statsmodels library of Python programming language, the ARIMA model will be run. Also, the (*p*, *d*, *q*) value as (2, 0, 2) gives a lower Akaike information criterion (AIC) value. Seasonal ARIMA models with (2, 0, 2) and $(1, 1, 0)_{24}$ are taken for model building. These same orders of ARIMA models will be used in the other two links (station id: 313190 and 314042) traffic flow forecasting because the ACF and PACF plots are almost the same as the 4-lane link (Station id 319351) and the ADF tests' *p*-value is less than 0.05.

3.3.2 LSTM and GRU Neural Network Model

For LSTM and GRU models, we will use the Python Keras library (Attri et al., 2020)for model building, and Python Keras library (Chollet, 2020) for transfer learning and fine tuning. The dataset is transformed by the MinMaxScaler function. Then the dataset is split into the train, validation, and test dataset.

The dataset is transformed in the Min-Max scaler. Using the Min-Max scaler, input features or variables will be transformed to a scale of 0 to 1. The mathematical formulation of the MinMaxScaler function is shown by:

$$x_{scaled} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{3.1}$$

Where: x_{scaled} is the Min-Max scaled value; x is the original value; $\min(x)$ is the minimum value of that feature; and $\max(x)$ is the maximum value of that feature or variable.

Multiple variables will be used in LSTM and GRU models where the target variable is total flow (vehicles/h), and other traffic variables are average occupancy (%) and average speed (mph). This type of model will be called the without-weather multivariate model. For the with-weather multivariate model, additional weather variables will be used for checking research question 2. Additional weather variables are hourly precipitation (in), hourly dry bulb temperature (°F), hourly wind speed (mph), and hourly visibility (miles).

For LSTM and GRU models, the following model settings are used. Both models are used as stacked LSTM and stacked GRU models. Which means, multiple hidden layers will be used.

Table 3.2: Model settings for LSTM and GRU multivariate models

		LSTM neural network	**GRU neural network**
Window length		24	24
Batch size		36	36
Number of neurons	**1st hidden layer**	1024	1024
	2nd hidden layer	512	512
	3rd hidden layer	264	264
Activation function		LeakyReLU	LeakyReLU
Patience		5	5
Optimizer		Adam	Adam
Number of epochs		100	100
Learning rate		Default (0.001)	Default (0.001)
Random state		123	123

3.3.3 Stacked-LSTM Model Architecture

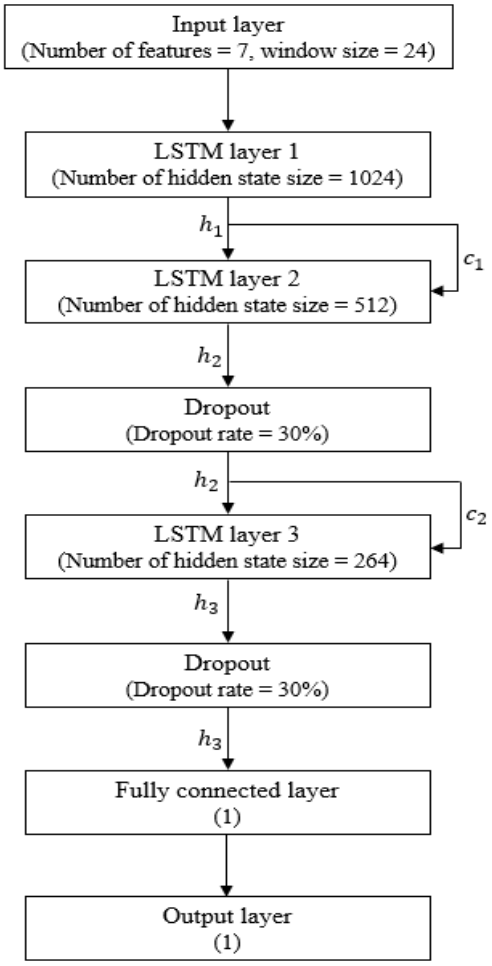


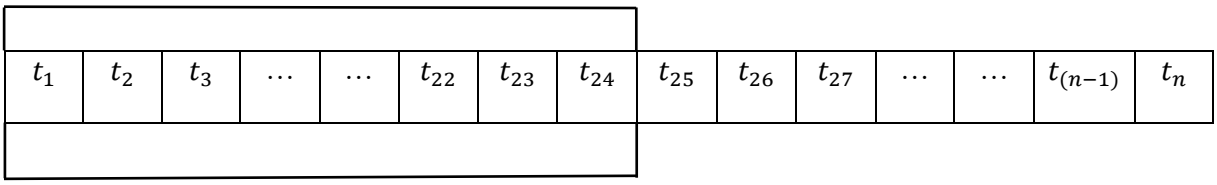
Figure 3.9: Stacked LSTM layers architecture (source network)

Figure 5.3 describes the architecture of the stacked-LSTM network in our source model; this model will be used as the pre-trained model for transfer learning. The first layer is the input layer, where the total number of features is 7 (3 variables as traffic variables and 4 variables as weather variables) and window size is taken as 24 as data is hourly data. The second layer is the first LSTM layer with 1024 hidden states. h_1 is the first hidden state vector, and c_1 is the first cell memory state vector. In the second LSTM layer, the number of hidden states is taken as 512 and it will release h_2 as a second hidden state vector and c_2 as a cell state vector. On the second LSTM hidden neurons, a dropout layer is used where the dropout size is 30%. Srivastava et al. (2014) propose the dropout technique to deal with overfitting in deep learning. By demonstrating how dropout enhances neural network performance on supervised learning tasks in vision, speech recognition, document classification, and computational biology, they achieve cutting-edge results on several benchmark data sets (Srivastava et al., 2014). During training time, dropout can ignore the randomly selected neurons. In our tasks as we selected a 30% dropout rate which means it can drop out 30% of randomly selected neurons in each cycle of weight update. The third LSTM layer contains the number of hidden neurons of 264 number with a dropout layer with the same 30% dropout rate, and finally, it will retrieve h_3 hidden state vector. After that, we use a fully connected layer, and the last layer is the output layer.

3.3.4 Sliding Window

According to Hota et al. (2017), sliding window prediction system is given below.

(a) Initial window:



(b) Window sliding:

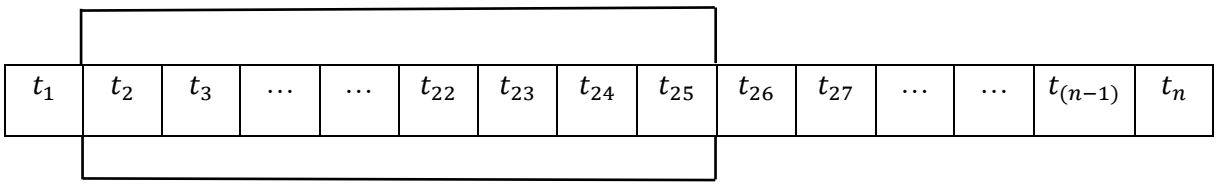


Figure 3.10: Sliding window forecasting system for time series data (Source: Hota et al., 2017)

As our traffic datasets are 1-hour granular datasets, in each 24-time step, the patterns of traffic flow are almost similar. We take the window size for traffic flow forecasting as 24. Figure 3.10 shows the initial time series where t_1 to t_{24} is the first window length that will give a prediction of the next time step which is t_{25} due to using the window. In the window sliding system, the next step will be the same window and will slide 1 step right. That means this window will cover t_2 to t_{25} time steps and will give a prediction for the next step or t_{26} time step. It will go further until t_{n} and will give the final prediction of the dataset.

3.4 Model Evaluation Metrics

To evaluate traffic flow forecasting model in terms of different approaches we selected Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) metrics for model evolution. According to Ren et al. (2021), the formulas of evaluation metrics are given below:

Mean absolute error, MAE = $\frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y}_i\right|$

Root mean squared error, RMSE = $\left(\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \widehat{Y}_i\right)^2\right)^{\frac{1}{2}}$ (3.2)

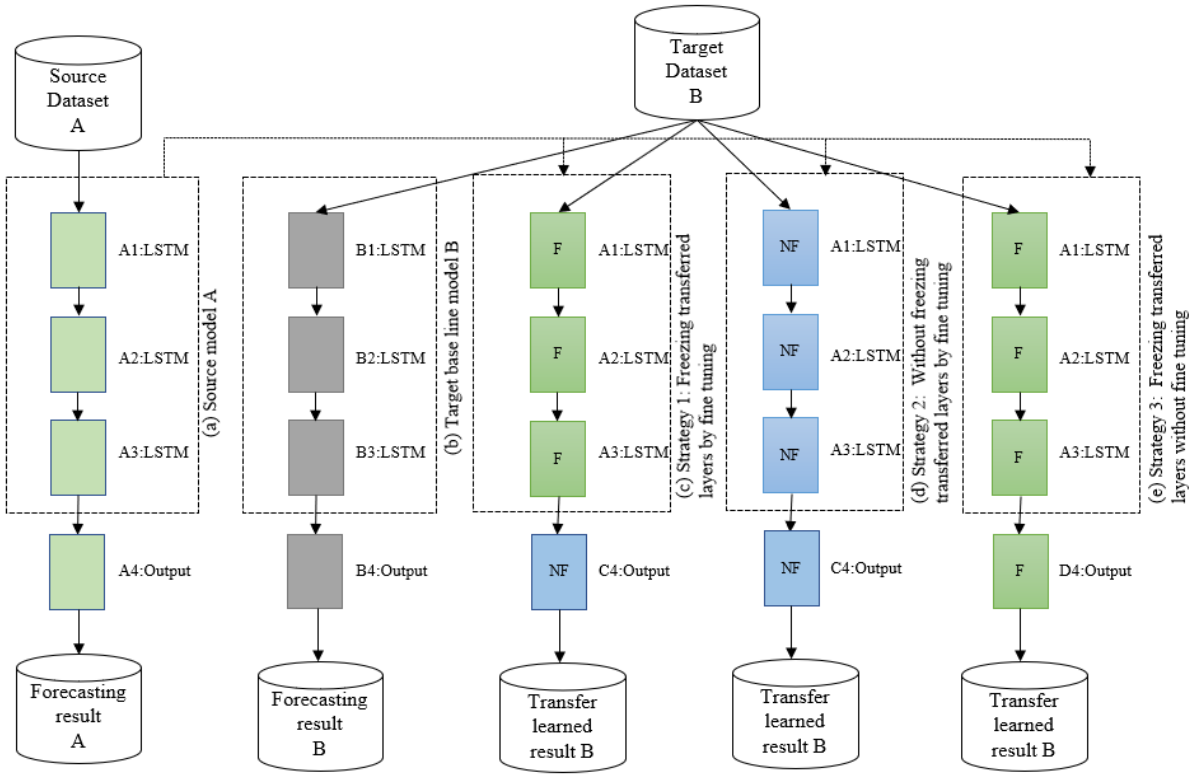
Mean absolute percentage error, MAPE = $\frac{1}{n}\sum_{i=1}^{n}\left|\frac{Y_i - \widehat{Y}_i}{Y_i}\right|$

Where: Y_i is the actual traffic flow, $\widehat{Y}_i$ is the predicted traffic flow, and n is the number of observations.

3.5 Transfer Learning with Different Strategies

Pre-trained models from district 3 datasets will be implemented here for transfer learning. There are three different strategies here that will be applied to transfer learning.

- Strategy 1: Parameters in the first n layers will be kept frozen, and parameters in the last layer will be kept unfrozen and will be trained toward the target dataset, which means the district 6 dataset of California.
- Strategy 2: Fine-tuning will be done without freezing transferred layers. All parameters within the entire neural network will be re-trained towards target dataset which means district 6 dataset of California state.
- Strategy 3: It is called without fine-tuned model, where source model's all layers will be kept frozen and pre-trained features of source model will be used on new datasets of district 6.



NF = Not frozen; F = Frozen

Figure 3.11: The architecture of transfer learning strategies adapted from J. Li et al. (2021)

Figure 3.11 demonstrates the architecture of transfer learning strategies. Figure (a) is Source model A, which is taken from district 3 datasets. This source model is transferred through three different strategies of transfer learning. Figure (b) shows the target baseline model B, which is taken from district 6 datasets with the one-month dataset, and the prediction of this model is compared with test datasets. In each source model and target model, 3 hidden LSTM layers will be used, which is called the stacked-LSTM model. Figure (c) shows strategy-1 transfer learning, where all layers except the last layer of the pre-trained model are kept frozen. This last layer will be kept unfrozen and ready for fitting on the target dataset. Figure (d) describes strategy-2 transfer learning, where all layers will be kept unfrozen and fine-tuned for transfer learning on target datasets. Figure (e) illustrates the strategy-3 transfer learning model where all layers of the pre-trained source model are kept frozen and learned on target datasets. This is called without fine-tuned typical transfer learning model.

- **Learning rate:**
  Many types of optimizers can be used in deep learning models. The learning rate is an important part of an optimizer. The amount of parameter updating during training, known as the learning rate, includes a trade-off between convergence accuracy and convergence speed (J. Li et al., 2021). Optimization methods can be divided into two categories: non-adaptive learning rate methods and adaptive learning rate methods, depending on how the learning rates are calculated at each stage, such as non-adaptive learning rate methods are stochastic gradient descent (SGD), SDG with momentum (J. Li et al., 2021). Adaptive learning rate methods examples are AdaGrad (Duchi et al., 2011), RMSProp (Tieleman & Hinton, 2012), Adam (Kingma & Ba, 2014). Due to quick training times and effective results, adaptive learning rate approaches are becoming more and more common and widely used (Karpathy, 2017; J. Li et al., 2021). In this thesis, we chose Adam optimizer as their running time is faster. We take the default learning rate value as 0.001 as this learning rate gives better performance.

- **Batch Size:**
  Batch size, a crucial hyper-parameter that affects the dynamics of the learning algorithm, refers to the number of samples from the training dataset used in the estimation of the error gradient (J. Li et al., 2021). According to J. Li et al. (2021), the number of features and data size have a direct impact on the batch size selection. That's why when there are many training instances, mini-batch training is typically utilized since a smaller batch size would lead to more frequent gradient changes due to lower computing costs (J. Li et al., 2021). In our case, we will utilize mini-batch size as our models are multivariate time series models. Smaller batch size will help to reduce computation burden as a single batch size for a large dataset takes a lot of time to gradient update. In our case, the batch size is taken as 36 because it will help to estimate a faster gradient error.

- **Number of epochs:**
  According to J. Li et al. (2021), how many times the complete dataset is fed through a neural network is shown by the number of training epochs. The size and diversity of the training dataset heavily influence the number of training epochs, and there is no one ideal amount of training epochs for training the neural network (J. Li et al., 2021). Early stopping is also used in our models, where it can control the overfitting issues.

4 Data Analysis

We selected a total of 6 different road types from district 3 and district 6. 2-lane, 3-lane, and 4-lane road datasets are selected from both districts 3 and 6. These are described in the following sub-chapters.

4.1 District 3 Traffic Data

4.1.1 Selected 2 Lane Road

The 2-lane road is selected from District 3, which has station number 313190 from Sacramento County, California. Traffic data collected by this station's detector is the loop detector. This road is selected for the traffic flow forecasting model and source dataset of transfer learning.

Figure 4.1: Elk Grove Blvd Road in Sacramento, California (Station: 313190) (Source: OpenStreetMap)

4.1.2 Selected 3 Lane Road

The 3-lane road is selected from district 3 which has station number 314042 from West Sacramento, California. Traffic data collected by this station's detector is the loop detector. This road is selected for the traffic flow forecasting model.

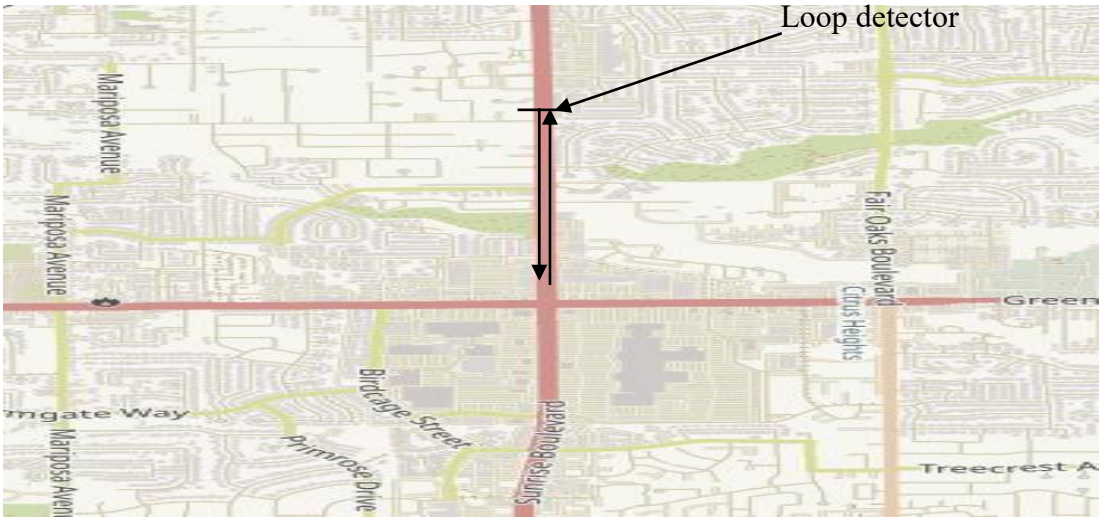


Figure 4.2: Sunrise Blvd Road in Sacramento, California (Station: 314042) (Source: OpenStreetMap)

4.1.3 Selected 4 Lane Road

We selected SB Riverside Avenue from Placer County of district 3 as 4 lane road for our study. The station number for this 4-lane road is 319351. This road collects data by loop detectors and then combines all lanes data to this station number. This road dataset is selected for the model building as a source model dataset.

Figure 4.3: SB Riverside Avenue, Placer County, California (Station: 319351) (Source: OpenStreetMap)

4.2 District 6 Traffic Data

We will use three road datasets for district 6 as the target model datasets for transfer learning. Three different road types are 2-lane, 3-lane, and 4-lane.

4.2.1 Selected 2 Lane Road

We selected a 2-lane road at Herndon Avenue in Fresno County, District 6 in California. At station 601722, traffic data is collected using loop detectors and aggregated and recalculated for total traffic flow, average traffic occupancy, and average traffic speed. 1-month of traffic data from April 2019 and 1-month of traffic data from September 2019 are taken separately as the target datasets for the transfer-learning model.

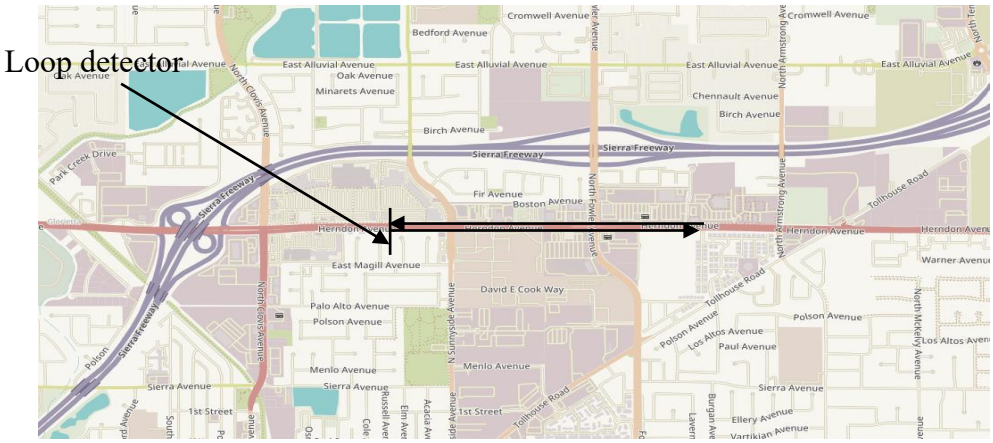


Figure 4.4: Herndon Avenue in Fresno, California (Station: 601722) (Source: OpenStreetMap)

4.2.2 Selected 3 Lane Road

A 3-lane road Barstow Avenue which is situated in Fresno County in district 6 in California State was selected. Loop detectors collect traffic data, which are accumulated and recalculated for total flow, average occupancy, and average speed at station 602538. 1-month of traffic data from April 2019 and 1-month of traffic data from September 2019 are taken separately as the target datasets for the transfer-learning model.

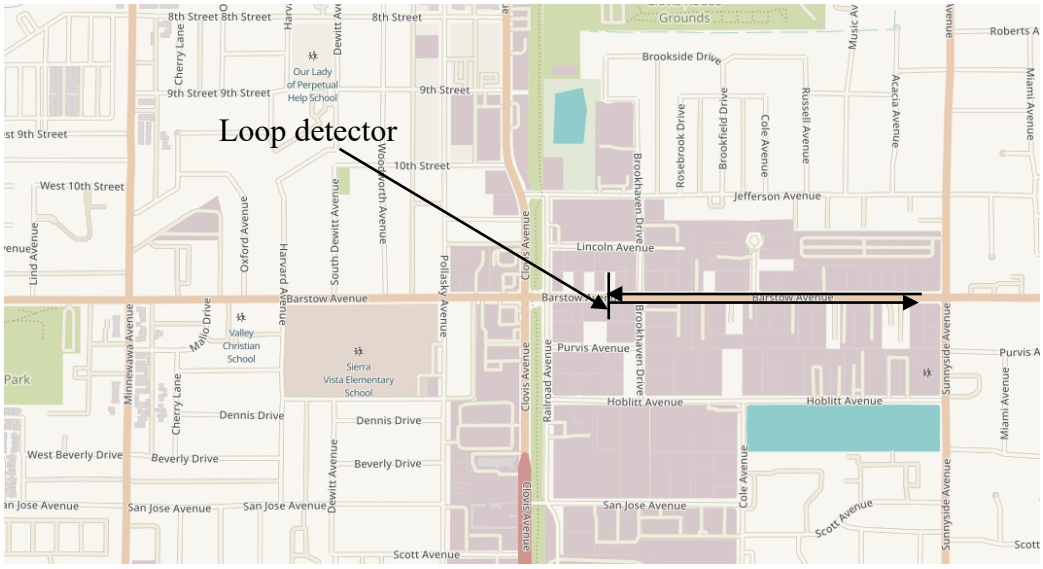


Figure 4.5: Barstow Avenue in Fresno, California (Station: 602538) (Source: OpenStreetMap)

4.2.3 Selected 4 Lane Road

Another 4-lane road Cedar Avenue which is situated in Fresno County in district 6 in California State was selected. Loop detectors collect traffic data, which are accumulated and recalculated for total flow, average occupancy, and average speed at station 601348. 1-month of traffic data from April 2019 and 1-month of traffic data from September 2019 are taken separately as the target datasets for the transfer-learning model.

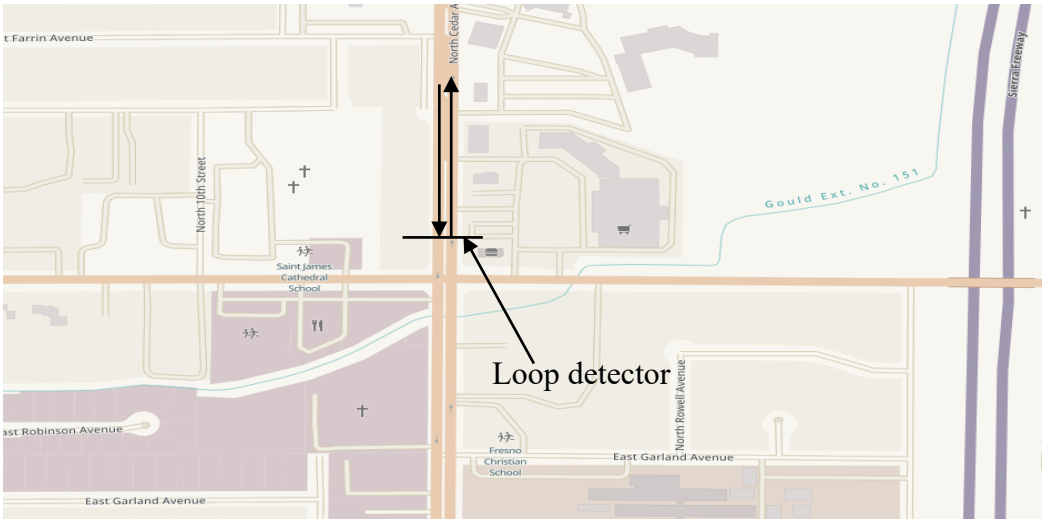


Figure 4.6: Cedar Avenue in Fresno, California (Station: 601348) (Source: OpenStreetMap)

4.3 Weather Data

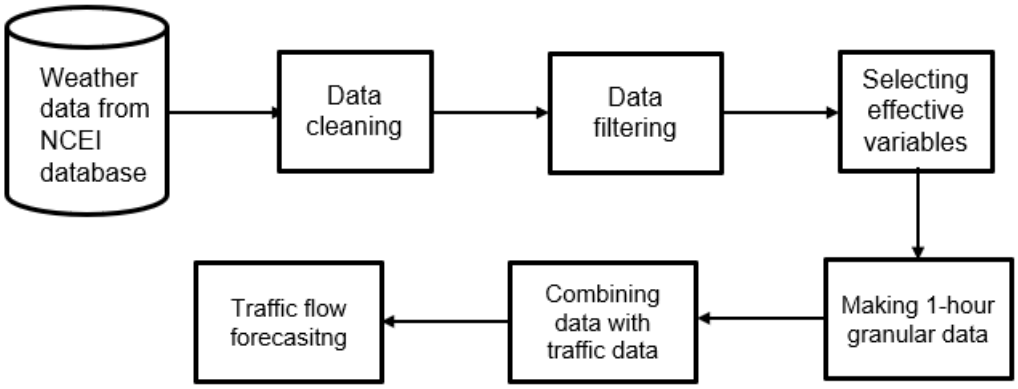


Figure 4.7: Methodological flow chart of combining weather data

There are a lot of external factors which influence traffic flow forecasting. These might be incidents, weather conditions, spatial conditions, football games, concerts, or special occasions according to the literature. As all these factors' data are not available, we will evaluate our model by external factor weather. Which will answer research question 2. We collected weather data for California state from the National Centers for Environmental Information (NCEI) (*National Centers for Environmental Information*, 2022) weather database, where almost more than 70 years of environmental data are available for all US states. We collected for the relevant County of district 3 and district 6 for checking the model and how it performs with weather data. We collected weather data with a granularity of 20 minutes for each experimental location from April 2019 to September 2019. Then from this 20-minute granular data, we got the average value for 1-hour granular data. There are many variables available in this weather data. From these variables, we take hourly precipitation, hourly dry bulb temperature, hourly visibility, and hourly wind speed as the weather factors. Then, these weather data were combined with the traffic data.

4.4 Data Visualization

4.4.1 2 Lane Road in District 3

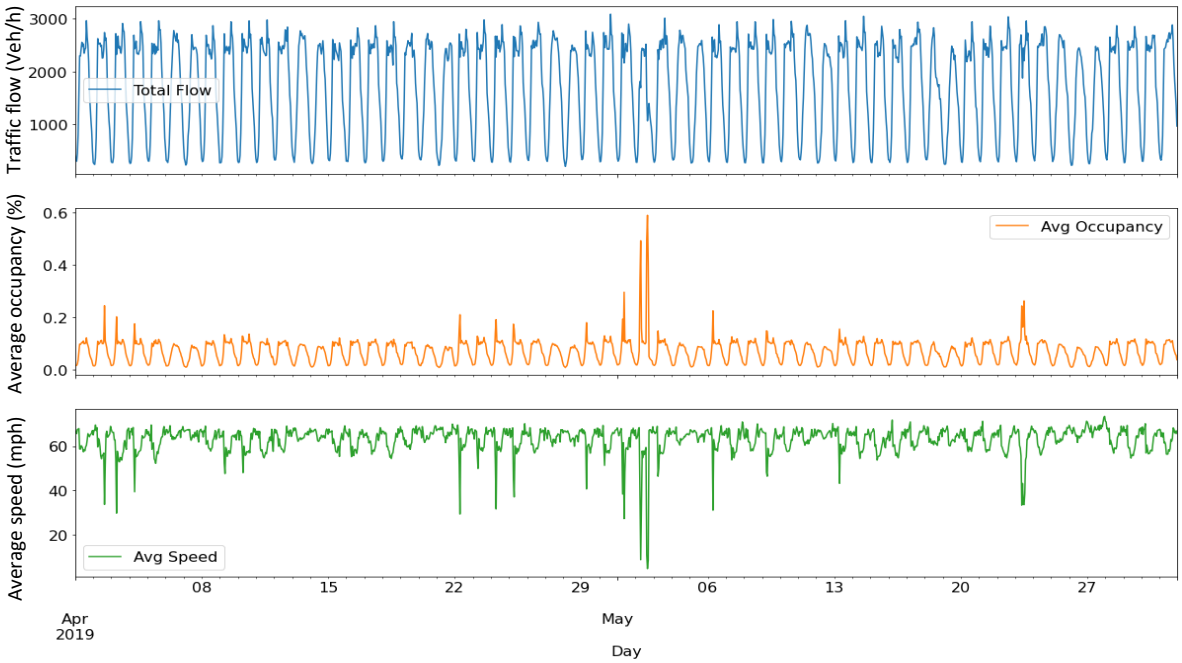


Figure 4.8: Traffic flow, occupancy, speed vs day (2-lane road data for station 313190)

Figure 4.8 shows traffic flow, average occupancy, and average speed vs. day plot where two months of data from April 2019 to May 2019 are available. In the top plot, in the blue color line graph, total flow vs. day is shown. Here, the total flow for this road is higher during peak hours in the morning time and in the evening time because in these time limits people go outside to the workplace or school or other activities in the morning and come back home in the evening. In off-peak hours flow decreased due to less demand. Especially at mid-day and night-time the traffic flow decreases a lot. This pattern follows all weekdays, which means weekday traffic flows are almost similar. During weekends, traffic flows are low. At day flow are higher than at night-time but not like on weekdays.

The middle figure in the orange-colored line shows two months' average occupancy vs. day for station 313190, which is Elk Grove Blvd of district 3 data. It illustrates that patterns of occupancy are almost similar to traffic flow. If the flow is higher, occupancy is also higher, and if the flow is lesser, occupancy is also lesser. During weekdays, occupancy values are higher like flow values, and during weekends, occupancy values are lower like flow values. At one point on 3rd May, it shows a sudden increase in occupancy value. The reason behind this is here either some traffic jam or an incident occurred.

The bottom figure in the green color line represents two months' average speed vs. day for station 313190, which is Elk Grove Blvd of district 3 data. It describes that the pattern of speed data is opposite to traffic flow and occupancy data. If traffic flow and occupancy are higher speed of traffic will be lower, and if traffic flow and occupancy are lower speed of traffic will be higher. On 3rd May it is showing that the speed of the traffic suddenly reduced in the peak hours, where either a traffic congestion or any incident occurred.

4.4.2 3 Lane Road in District 3

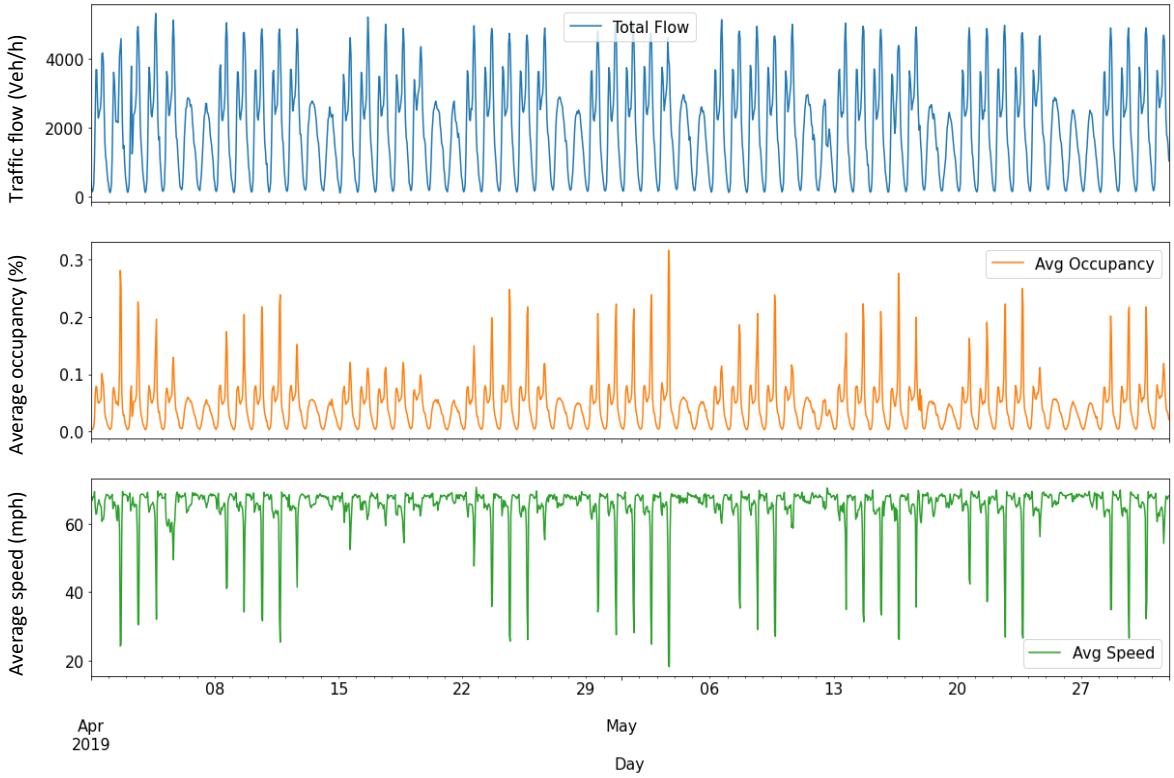


Figure 4.9: Traffic flow, occupancy, speed vs day (3-lane road data for station 314042)

Figure 4.9 shows the traffic flow, average occupancy, and average speed vs. day plot for the 3-lane road named Sunrise Blvd at Sacramento County in district 3 of California, where two months of data from April 2019 to May 2019 are available. The top plot in the blue color line illustrates total flow vs. day.

Here, the total flow for this road is higher during peak hours in the morning-time and evening-time because people go outside to the workplace or school or do other activities in the morning and come back home in the evening. In off-peak hours flow decreased due to less demand. Especially, at mid-day and night-time, flow decreases a lot. This pattern follows all weekdays, which means weekday traffic flows are almost similar. During weekends, traffic flows are low. At day-time flow are higher than at night-time but not like on weekdays.

The middle figure in the orange-colored line shows two months' average occupancy vs. day for station 314042, which is Sunrise Blvd of district 3 data. It describes that pattern of occupancy are almost similar to traffic flow. If the traffic flow is higher, occupancy is also higher, and if the traffic flow is lesser, occupancy is also lesser. During weekdays, occupancy values are higher as traffic flow, and during weekends, occupancy values are lower as traffic flow.

At the bottom figure in green color line is showing two months' average speed vs. day for the station 314042, which is Sunrise Blvd of district 3 data. It is stated that speed data patterns are opposite to traffic flow and occupancy data. If traffic flow and occupancy are higher speed of traffic will be lower, and if traffic flow and occupancy are lower speed of traffic will higher.

4.4.3 4 Lane Road in District 3

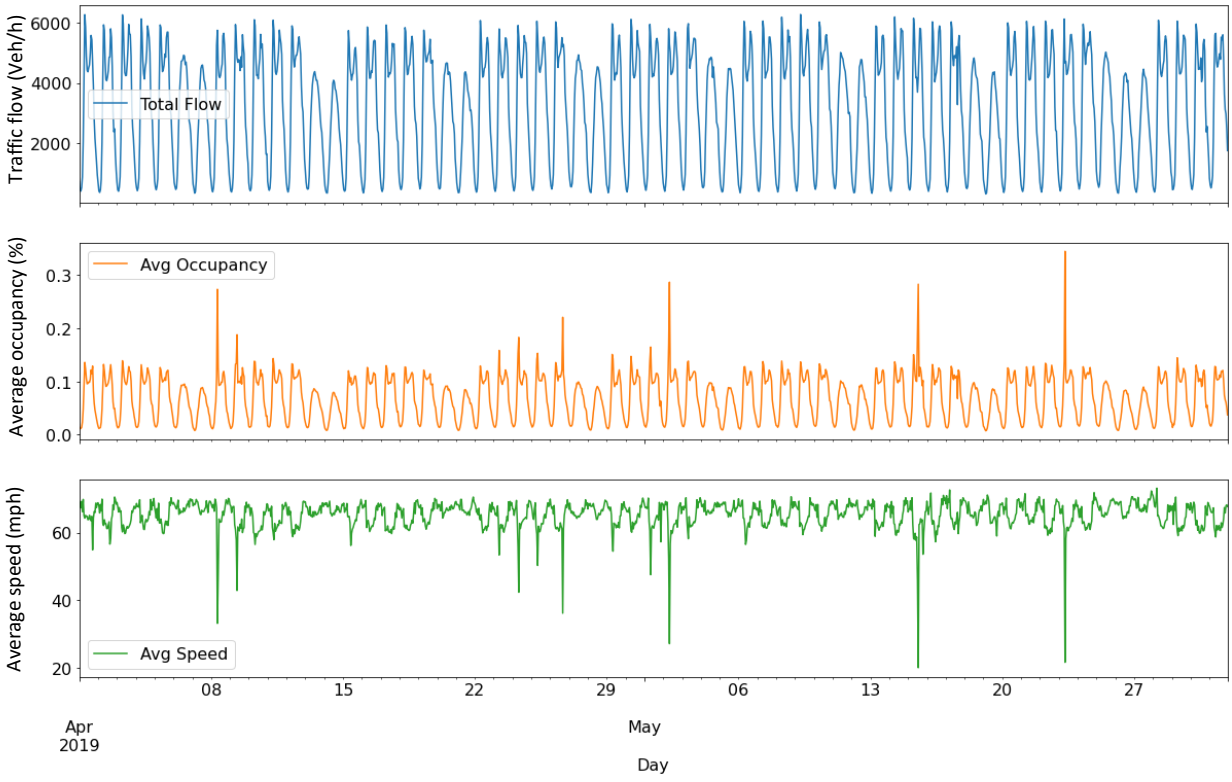


Figure 4.10: Traffic flow, occupancy, speed vs day (4-lane road data for station 319351)

Figure 4.10 shows traffic total flow, average occupancy, and average speed vs. day plot for the 4-lane road named SB Riverside Avenue at Placer County in district 3 of California, where two months of data from April 2019 to May 2019 are available. In the top plot in the blue color line, total flow vs. day shows where total flow for this road is higher during peak hours in the morning-time and evening-time, people go outside to the workplace or school or other activities in the morning and come back home in the evening. In off-peak hours flow decreased due to less demand. Especially at mid-day and night-time flow decreases a lot. This pattern follows all weekdays, which means weekday traffic flows are almost

similar. During weekends, traffic flows are low. At day flow are higher than at night-time but not like on weekdays.

The middle figure in the orange-colored line shows two months' average occupancy vs. day for station 319351, which is SB Riverside Avenue of district 3 data. It illustrates that patterns of occupancy are almost similar to traffic flow. If the flow is higher, occupancy is also higher, and if the flow is lesser, occupancy is also lesser. During weekdays, occupancy values are higher like flow values, and during weekends, occupancy values are lower like flow values. In some points, on 8th April, 1st May 15th May, and 24th May, the occupancy values are higher means here speeds of the vehicles are lower for the traffic congestion or other incident.

The bottom figure in the green colored line shows two months' average speed vs. day for station 319351, which is SB Riverside Avenue of district 3 data. It describes that speed data patterns are opposite to traffic flow and occupancy data. If traffic flow and occupancy are higher speed of traffic will be lower, and if traffic flow and occupancy are lower speed of traffic will be higher. Similarly, as average occupancy graph, on 8th April, 1st May, 15th May, and 24th May, speed values are suddenly decreasing, which means here speeds of vehicles are lower for a traffic jam or other incident.

4.4.4 District 3 and District 6 Data Comparison (2 Lane)

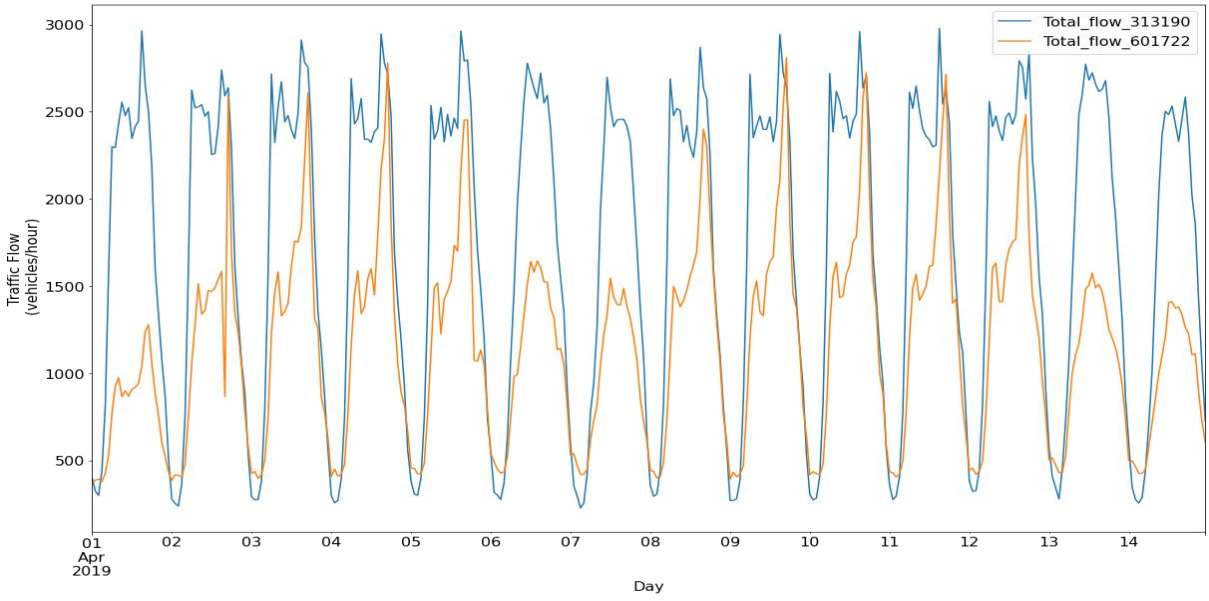


Figure 4.11: 2 lane road's district 3 vs district 6 data comparison

Figure 4.11 shows 15 days of traffic data from April 1, 2019, to April 14, 2019, for 2-lane roads in both District 3 and District 6. The district 3 2-lane road (station id 313190) is Elk Grove Blvd in Sacramento County, which was mentioned in previous plots. The district 6, 2-lane road (station id 601722) is Herndon Avenue in Fresno County. In this plot, the blue line shows the data pattern of station 313190, and the orange line shows the data pattern of station 601722. In this comparison plot, flow patterns are completely different, where overall, the total flow of 313190 is higher than the total flow of 601722. Although the flow of 60177 is low during weekends, the flow of 313190 is high during weekends, and the flow is almost the same every day on 313190 roads. The station 313190 dataset will be used as a source dataset for the pre-trained model, which will be used for transfer learning to the district 6 target dataset (601722) with a few days of data.

4.4.5 District 3 and District 6 Data Comparison (3 Lane)

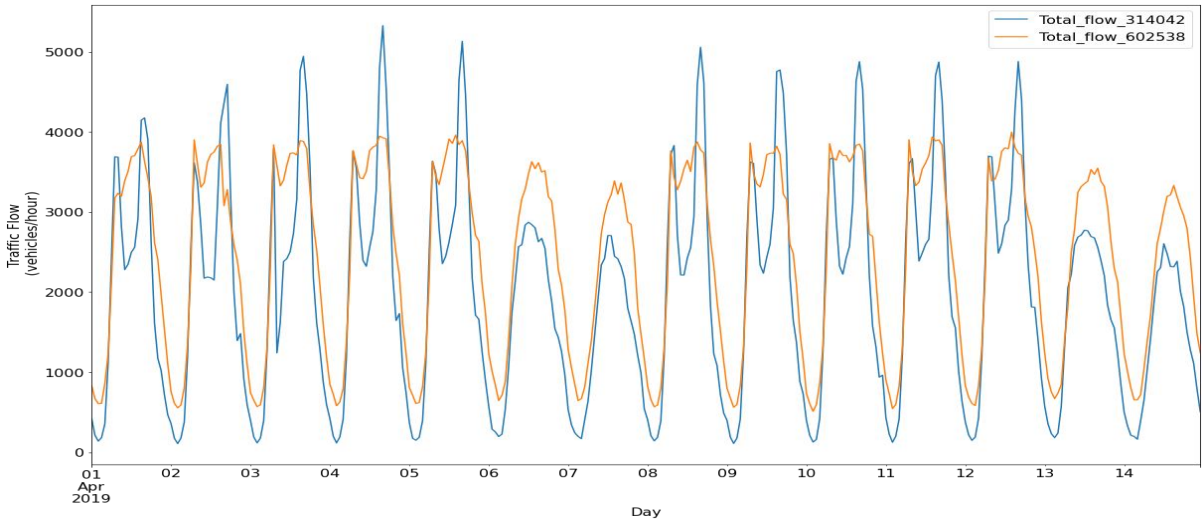


Figure 4.12: 3 lane road's district 3 vs district 6 data comparison

In figure 4.12, the plot illustrates a data pattern comparison of district 3 and district 6 for 3-lane roads. Here the blue line represents the actual total flow of district 3, which is Sunrise Blvd Road (Station id 314042) in Sacramento County. The orange line shows the actual total flow of district 6, which is Barstow Avenue (Station id 602538) in Fresno County. This graph shows 14 days of data plots from April 1 to April 14 are showing for 3-lane roads. Overall, the actual total flow of 314042 is higher than 602538 for weekdays and weekends. For 602538, traffic flow is a little bit less on weekends than on weekdays traffic flow. 314042 has a greater weekday actual flow than district 6 dataset 602538, but a smaller weekend actual flow. The 314042 dataset will be used as a source dataset for the pre-trained model, and the 602538 dataset will be used as the target dataset in transfer learning.

4.4.6 District 3 and District 6 Data Comparison (4 Lane)

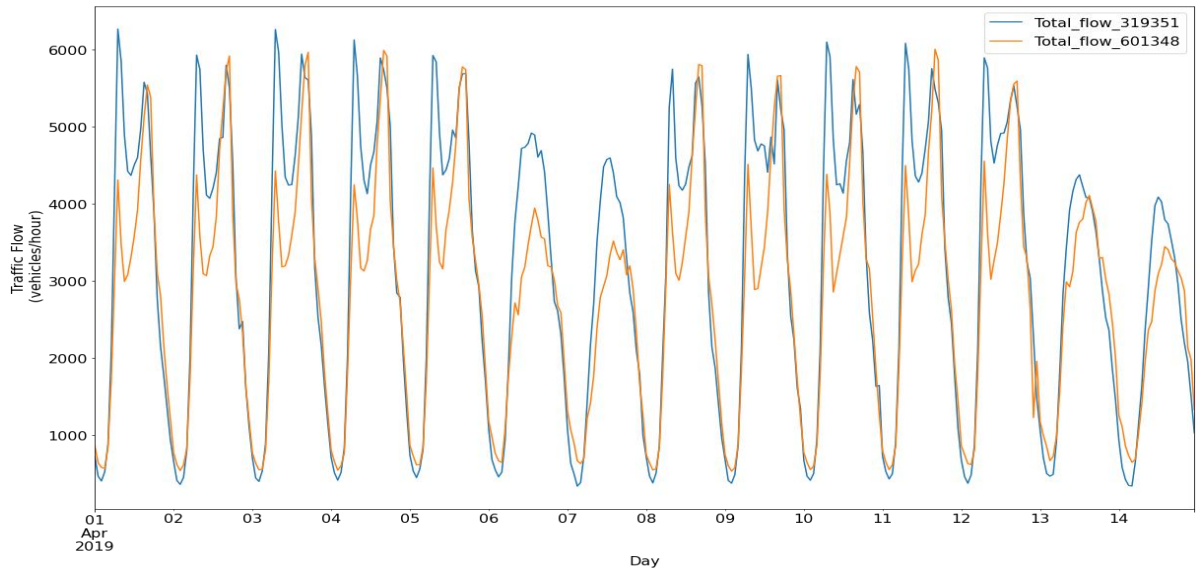


Figure 4.13: 4 lane road's district 3 vs district 6 data comparison

Figure 4.13 illustrates the actual traffic flow pattern comparison for both district 3 and district 6 4 lane road where from April 1, 2019, to April 14, 2019, 14 days of traffic flows are used. The blue plot shows

the actual total flow of SB Riverside Avenue 4 lane road (Station id 319351) from Placer County, District 3. The orange line describes the actual total flow of Cedar Avenue (Station id 601348) from Fresno County of District 6. If we see at the plots, flow patterns are different in both districts' flow plots. Although the maximum traffic flow on weekdays in both districts is almost the same actual traffic flow of 319351 is higher during weekends. The 319351 dataset will be used as a source dataset for the pre-trained model, and the 601348 dataset will be used as a target dataset in transfer learning.

4.4.7 District 3 Data Correlation Plots

4.4.7.1 2 lane

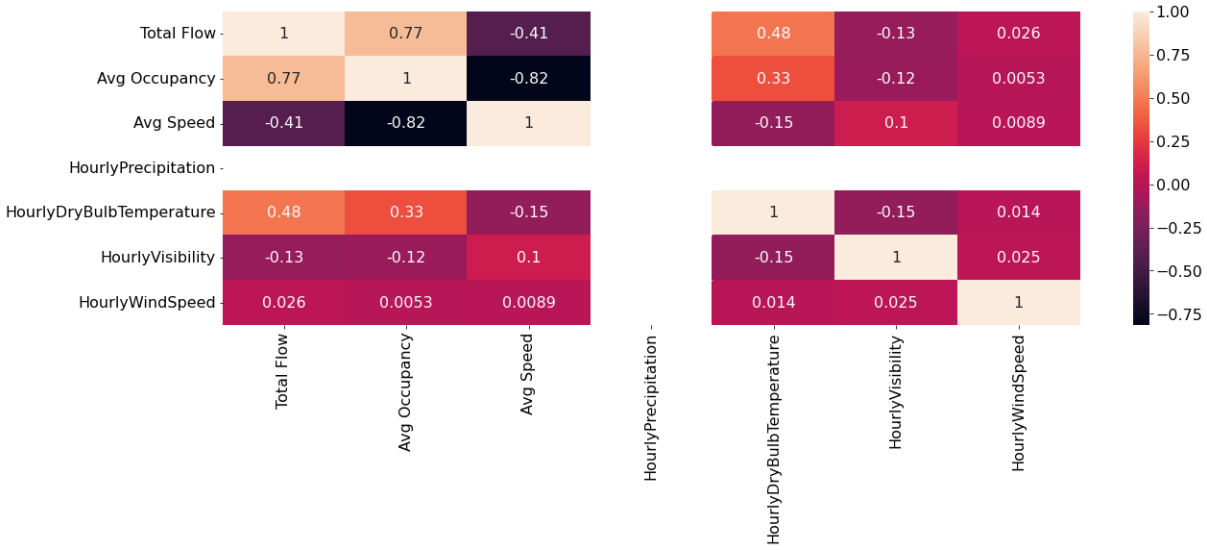


Figure 4.14: Correlation plot of 2 lane road in district 3 (Station id 313190)

As city traffic behaves stochastic nature correlation of each variable (total flow, average occupancy, and average speed) behave differently at different time. According to figure 4.14, the total flow and average occupancy are 77% positively correlated. The correlation varies from case to case and from spatial condition to spatial condition, but it is always positive. Concerning average speed, it shows that for higher traffic flow average speed is lower, which means it is negatively correlated with an average speed. It shows that total flow -41% correlated with average speed. It also varies by different spatial and temporal conditions. For the relation of average occupancy with average speed, the scenario is the same as before, that occupancy is negatively correlated with traffic average speed. For other factors like hourly precipitation, hourly dry bulb temperature, hourly visibility, and hourly wind speed main variables (total flow, average occupancy, and average speed) behave differently.

4.4.7.2 3 lane

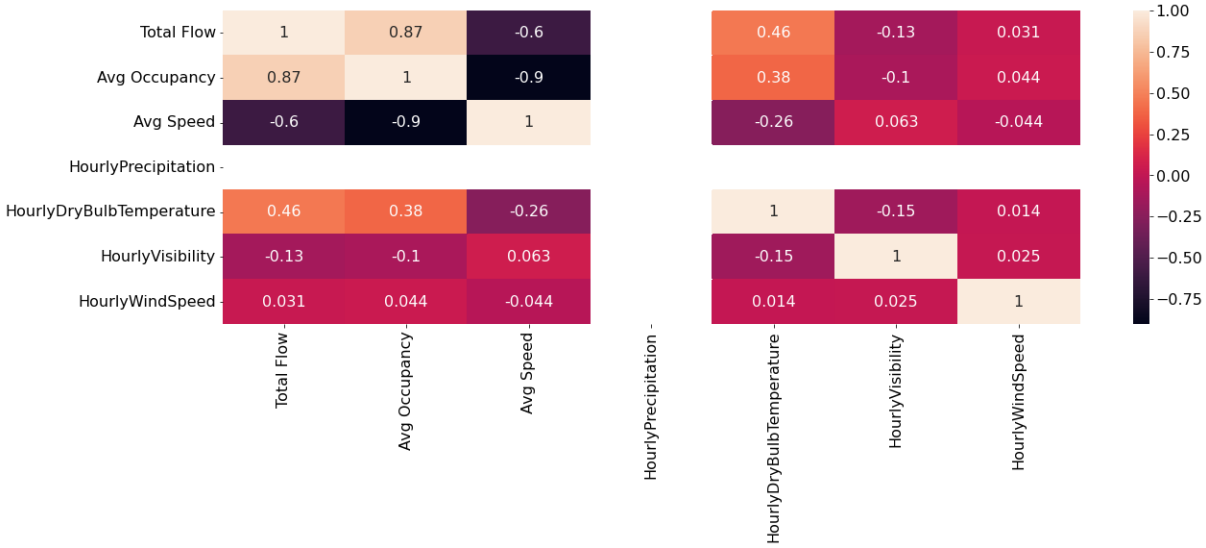


Figure 4.15: Correlation plot of 3 lane road in district 3(Station id 314042)

Figure 4.15 illustrates the correlation plot of 3 lane road from district 3 where average occupancy is 87% positively correlated with the total flow and average speed is 60% and 90% negatively correlated with both total flow and average occupancy, respectively. Hourly dry bulb temperature and hourly wind speed are positively correlated to total flow and average occupancy, but hourly wind speed has less effect. On the contrary, these two weather variables are negatively correlated to average speed. Another weather variable, hourly visibility is negatively correlated to total traffic flow and average occupancy and positively correlated to average traffic speed, but the effects are small.

4.4.7.3 4 lane

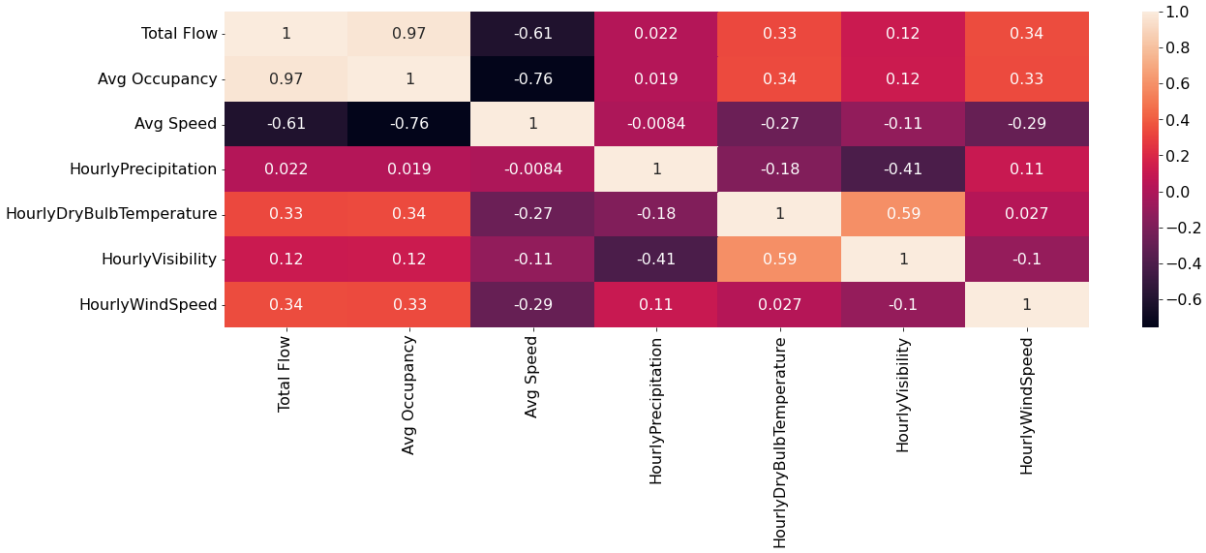


Figure 4.16: Correlation plot of 4 lane road in district 3 (Station id 319351)

Figure 4.16 shows the correlation plot of 4 lane road from district 3, where total traffic flow is 97% positively correlated with average occupancy and 61% negatively correlated with average traffic speed.

Also, average traffic occupancy is 76% negatively correlated with average speed. If we compare the correlation between weather variables with traffic variables, hourly precipitation, hourly dry bulb temperature, hourly visibility, and hourly wind speed all are positively correlated with traffic variables total flow, and average occupancy, but hourly precipitation has a small influence on both traffic variables. On another side, all weather variables are negatively correlated with average traffic speed, and hourly precipitation has less amount of correlation with average traffic speed.

4.4.8 Fundamental Diagrams

For fundamental diagrams, we selected all stations dataset from District 3. April 1 to April 30, 2019, which means 30 days of the dataset of district 3 whole region macroscopic traffic data were taken for making the fundamental diagram. Total flow means all lanes' total traffic flow in vehicles per hour data, and total density means all lanes' total vehicles per mile.

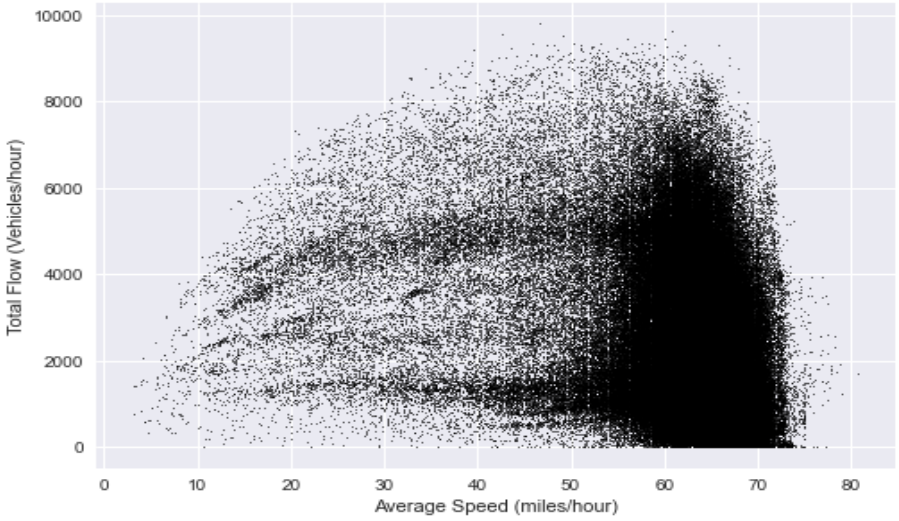


Figure 4.17: Fundamental diagram (Flow vs. speed)

Figure 4.17 represents the fundamental diagrams of different traffic parameters. The graph shows the total flow vs. average speed relation. Here, the total flows are not dense in the range of 0 miles/hour to 50 miles/hour of average speed. Most of the traffic flows are between 50 vehicles/hour to approximately 75 vehicles/hour or free flow conditions. According to Greenshield's fundamental diagram, this criterion is fulfilled for the flow vs. speed diagram.

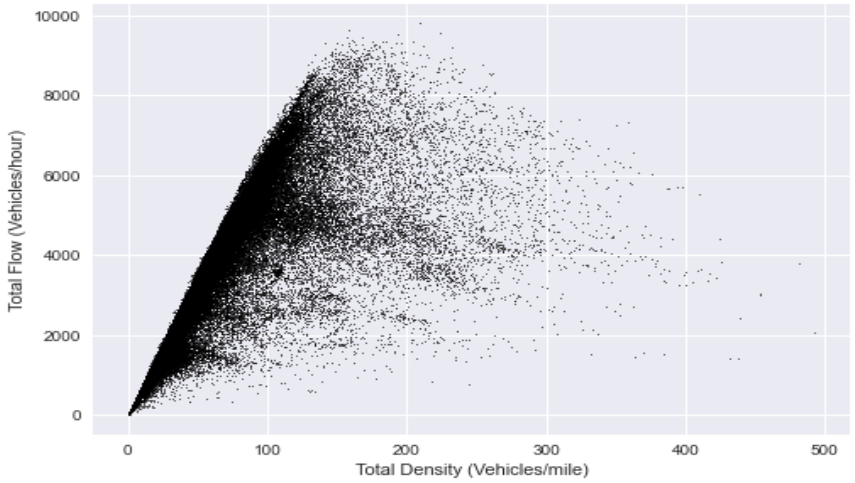


Figure 4.18: Fundamental diagram (Flow vs. density)

Figure 4.18 illustrates the relation between total flow vs. total density, where from 0 vehicles/mile density to approximately 120 vehicles/mile density, traffic total flow increases, and most of the traffic flows are in this range. Then, the total flow decreases. The maximum value of total flow here is approximately, 8500 vehicles/hour.

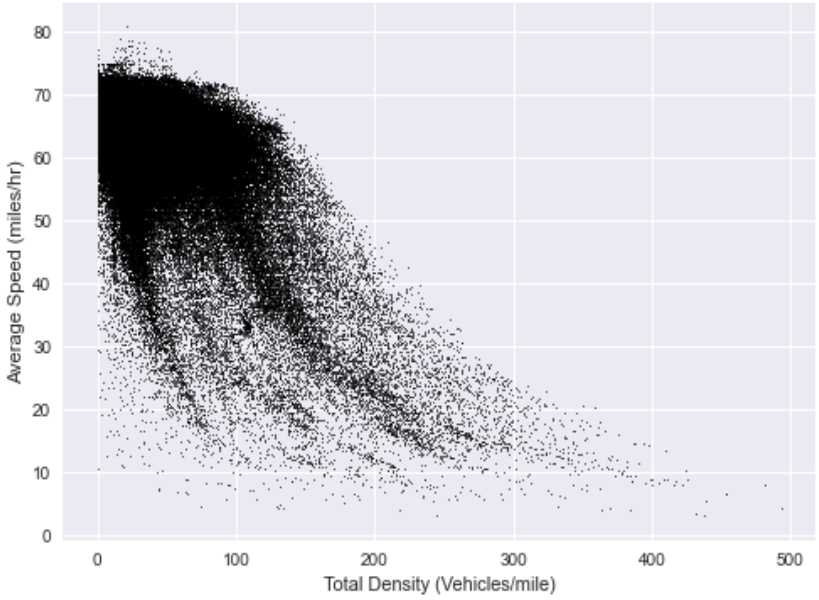


Figure 4.19: Fundamental diagram (Speed vs. density)

Figure 4.19 shows the fundamental diagram of average speed vs. total density, wherein the x-axis shows total density, and the y-axis shows average speed. From 0 to 100 (vehicles/mile) density, the average speed of the vehicle is maximum. Then the speed gradually decreases for the increased amount of density. This boundary line is approximately a linear line with a negative slope. For higher density, the speed of the vehicles is minimum in that case.

4.5 Train-test Split

For the univariate ARIMA model, the dataset is split into a train-test dataset. No shuffling is used when the dataset is divided. The first 95% of the dataset is picked as the training dataset, and the next 5% dataset is picked as the test dataset. In the GRU and LSTM models, the first 90% of the data is selected for the training dataset. The next 5% of data is selected for the validation dataset, and the last 5% is selected for the test dataset. In the train-test split model, the training dataset will be used for the prediction in the time series forecasting model. The test dataset will be used to determine the accuracy of the prediction. Similarly, the dataset for transfer learning in the LSTM model is divided into three parts. The first 90% dataset is taken as the training dataset, the next 5% dataset is taken as the validation dataset, and the last 5% dataset is selected as the test dataset in source model datasets. In target model datasets, we take 60% data as a training dataset, the next 20% data as a validation dataset, and the last 20% data as a test dataset.

4.6 Transfer Learning Scenarios

The following table is showing the different transfer learning scenarios for different link-type LSTM multivariate models. Source models are chosen as station IDs 313190, 314042, and 319351. Target models are chosen as station IDs 601722, 602538, and 601348.

Table 4.1: Different scenario for transfer learning for different link type

Timeline	**April 2019 to September 2019**		**April 2019**	
Scenarios	**Source link**		**Target link**	
	Station ID	**Number of lanes**	**Station ID**	**Number of lanes**
Scenario 1	313190	2	601722	2
Scenario 2	314042	3	602538	3
Scenario 3	319351	4	601348	4
Scenario 4	314042	3	601722	2
Scenario 5	319351	4	602538	3
Scenario 6	319351	4	601722	2

For experimenting with all six scenarios, we will check how transfer learning models will perform for April 2019 target datasets. Another three experiments will be done for 2-lane, 3-lane, and 4-lane roads of district 6 with five months ahead data (September datasets) as target datasets.

5 Result and Discussion

In this thesis, we will implement all three categories of the time series model for performance analysis. ARIMA model is the statistical classical model, gradient boosting, random forest regression, and recent XG-boost are some machine learning time-series forecasting models. A variety of deep learning models, including artificial neural networks (ANN), convolutional neural networks (CNN), recurrent neural networks (RNN), long-short-term memory (LSTM), and graph neural networks (GNN), etc. are widely used for time-series forecasting. RNN follows backpropagation, which means it considers previous values for predicting future values and retains previous important information for prediction. LSTM is a modified version of RNN. In our thesis, we will use the ARIMA model, the gated recurrent unit (GRU) model, and the LSTM model. The LSTM model will be used for transfer learning of traffic flow forecasting.

For research question one, we will get the answer by making several models with 1-month data, 2-month data, 3-month data 4- months data, 5-month data, and 6-month dataset for each univariate ARIMA, multivariate GRU, and LSTM model. We will do another experiment in transfer learning models with scenarios 1, 2, and 3 for 1 month source data and 6-month source data.

For research question two, we will collect additional weather data for each district. We will consider different models integrating traffic data with weather data and without weather data for both district 3 and district 6, and we will compare factors like weather influence how much in the traffic forecasting models for multivariate GRU and LSTM. We will compare the result of each model with weather data and without weather data.

This chapter contains the findings of forecasting models and transfer-learning models. We used the univariate ARIMA model with only one variable total flow (vehicles/h), which is the target variable. ARIMA model with exogenous variables (average occupancy, average speed, weather variables) was experimented with, but this model does not perform well. Traditional univariate ARIMA model and two different neural networks LSTM and GRU are used for research question 1. For research question 2, LSTM and GRU multivariate models are used for the performance and effects of weather variables as a factor.

5.1 Effects of training data size on model performance

5.1.1 2 Lane Road (Station id 313190)

Table 5.1: MAE, MAPE, RMSE value for 1 to 6 months of dataset by three algorithms for 2 lane road

Month of dataset	**MAE (vehicles/h)**			**MAPE (%)**			**RMSE (vehicles/h)**		
	LSTM	GRU	Univariate ARIMA	LSTM	GRU	Univariate ARIMA	LSTM	GRU	Univariate ARIMA
1	101.80	117.14	118.89	7.78	9.12	7.14	130.15	147.38	134.87
2	112.35	99.75	76.45	10.32	7.93	4.49	137.80	131.57	103.82
3	104.86	116.69	138.74	7.89	8.68	11.02	139.71	153.77	186.10
4	116.23	109.46	122.10	8.80	9.21	10.12	151.50	141.34	174.05
5	105.04	111.41	157.17	7.90	8.62	10.02	147.61	153.16	236.45
6	98.65	99.52	160.26	7.11	7.10	13.11	127.50	134.85	229.42

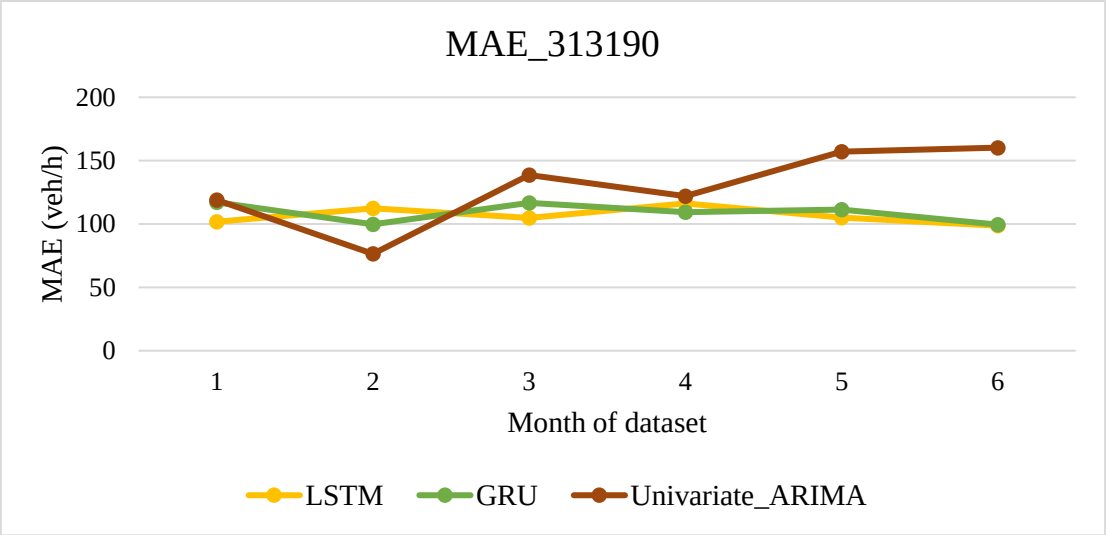


Figure 5.1: Changes of MAE based on different data length for 2 lane road (station id 313190)

In figure 5.1, the graph illustrates the changes in MAE according to different data lengths from 1 month to 6 months from District 3. For 1-month of data (April 2019), MAE is almost the same for all three algorithms. In the 3-month dataset (April 2019 to June 2019), GRU and LSTM are performing almost the same, but the ARIMA model performs not well here. Again, with the 4-months of data from April to July 2019, all three algorithms perform almost the same. Again, for the 5-month of the dataset from April to August 2019, LSTM and GRU perform better, and ARIMA is not performing well.

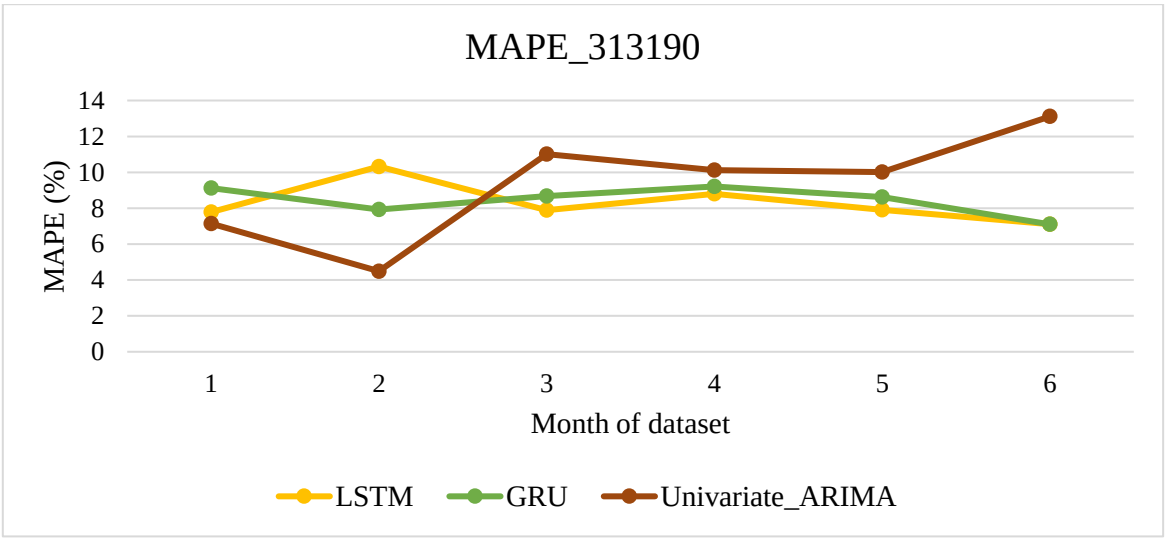


Figure 5.2: Changes of MAPE based on different data length for 2 lane road (station id 313190)

In terms of MAPE, the patterns of all algorithms are almost the same as MAE where LSTM and GRU models perform almost the same, but for the 6-month of the dataset (April 2019 to September 2019), we get the lowest value of MAPE for both LSTM and GRU model. For the univariate ARIMA model, model performance is good for 1 month of data or 2 months of data but after increasing data that means at 3 and 5 months of the dataset, the model does not perform well. Only at the 4-month of the dataset, the univariate model performs well. At 6-month dataset, ARIMA gives the highest amount of MAPE value.

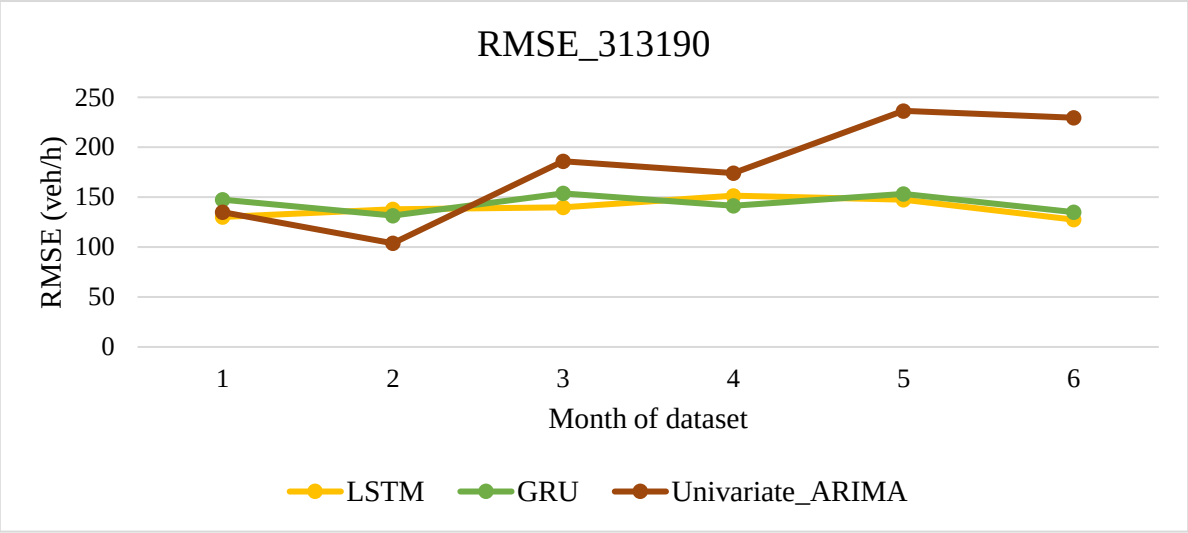


Figure 5.3: Changes of RMSE based on different data length for 2 lane road (station id 313190)

According to RMSE, model performance for all algorithms is almost the same as MAE and MAPE. Figure 5.3 shows almost the same results for both LSTM and GRU models. For the 6-month dataset (April 2019 to September 2019), model performance is well for both LSTM and GRU. At the 1-month or 2-month dataset, the univariate ARIMA model performs well. After increasing data, that means at the 3-month dataset and 5-month dataset, the univariate ARIMA model does not perform well. In the 4-month dataset from April to July, the ARIMA model performs well according to RMSE. At the 4-month and 6-month dataset the ARIMA gives the highest amount of RMSE.

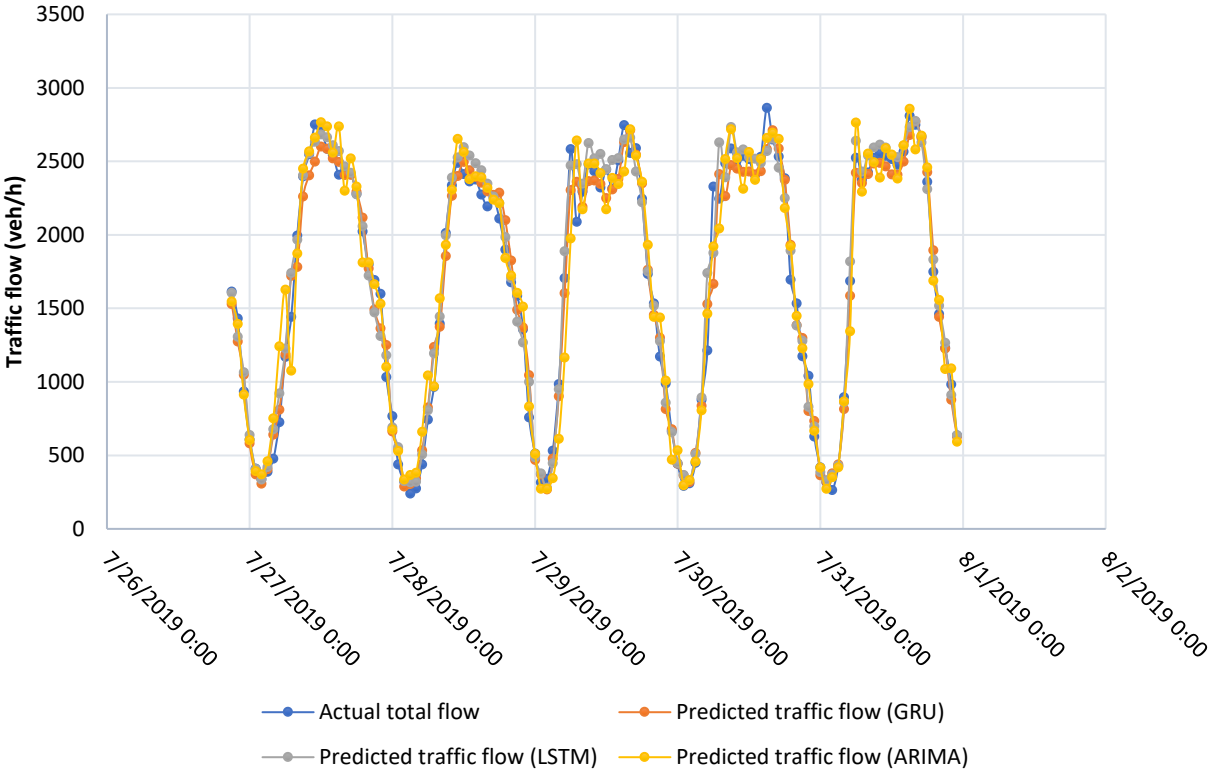


Figure 5.4: Forecasting plots of 2 lane road (station id 313190) for 4 months of datasets

Figure 5.4 illustrates forecasting plots for the 2-lane road from Sacramento. The actual total traffic flow is shown in the blue colored-line. The predicted traffic flow for the multivariate GRU model is shown

in the orange-colored line. The predicted traffic flow for the multivariate LSTM model is shown in the grey-colored line, and the predicted traffic flow for the univariate ARIMA model is shown in the yellow-colored line. Overall, LSTM and GRU models give better fits to actual traffic flow.

5.1.2 3 Lane Road (Station id 314042)

Table 5.2: MAE, MAPE, RMSE value for 1 to 6 months of dataset by three algorithms for 3 lane road

Month of dataset	MAE (vehicles/h)			MAPE (%)			RMSE (vehicles/h)		
	LSTM	GRU	Univariate ARIMA	LSTM	GRU	Univariate ARIMA	LSTM	GRU	Univariate ARIMA
1	195.72	181.05	169.20	12.22	11.83	9.16	297.86	251.78	296.64
2	134.74	185.31	86.82	13.57	14.33	7.34	166.90	305.18	119.73
3	128.50	168.86	185.67	10.45	13.47	13.75	174.17	214.32	257.71
4	125.69	111.45	162.34	10.30	9.74	11.02	168.30	146.43	251.61
5	99.17	126.12	220.07	8.75	11.07	13.31	147.37	177.88	404.57
6	104.19	150.00	293.50	9.30	11.45	16.57	137.24	205.09	446.18

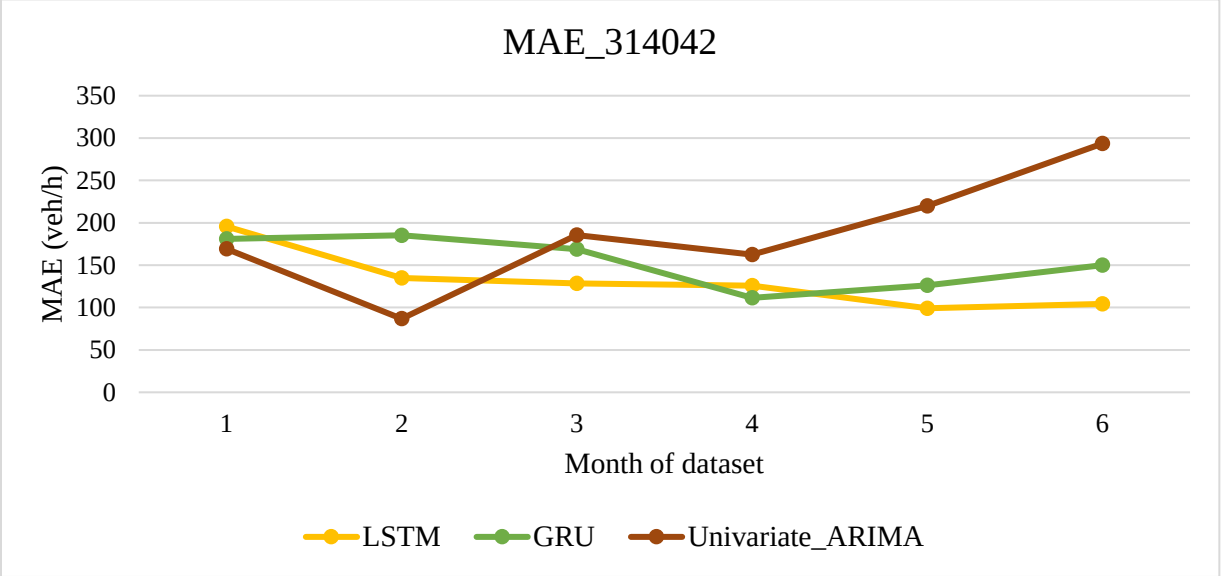


Figure 5.5: Changes of MAE based on different data length for 3 lane road (station id 314042)

Figure 5.5 represents 3-lane road forecasting model performances by MAE for different lengths of dataset patterns. Here, initially at univariate ARIMA model, the MAE value is lower for the 1-month dataset, and then it gradually decreases at 2-month dataset, for adding data after reaching a higher value at the 3-month dataset, it will decrease at the 4-month dataset and again sharply increases to peak at the 6-month dataset. For LSTM and GRU models, the MAE value is higher for the 1-month dataset, then both algorithms' MAE gradually decreases for both algorithms. In the 6-month dataset, LSTM performance is good, but after the 4-month dataset, the GRU model's MAE gradually increases again.

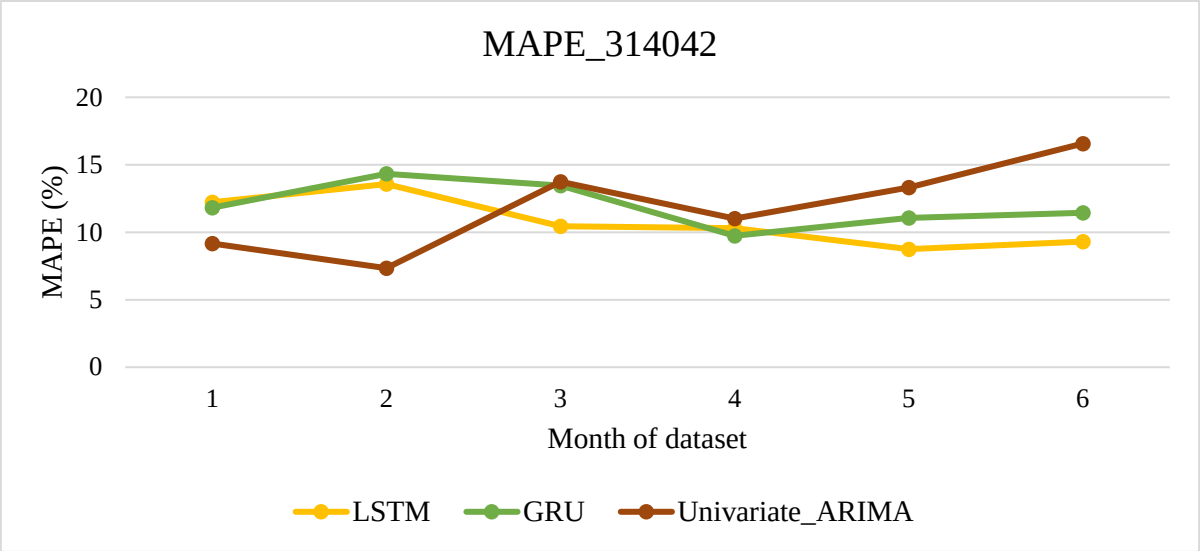


Figure 5.6: Changes of MAPE based on different data length for 3 lane road (station id 314042)

Figure 5.6 illustrates the 3-lane road traffic flow forecasting model performance for 3 different algorithms by MAPE, where the univariate ARIMA model performs well for both the 1-month dataset and the 2-month dataset. At the 3-month dataset model, the MAPE value increases and then again decreases to the 4-month dataset model and then again gradually increases and finally peaks at the 6-month dataset. For LSTM and GRU models, performance is almost the same for both algorithms where for the 1-month dataset, MAPE value is approximately 12%, then MAPE value slightly increases for the 2-month dataset and then again decreases to a minimum value at the 4-month dataset for the GRU model but at the 5-month dataset, LSTM model gives the lowest MAPE value.

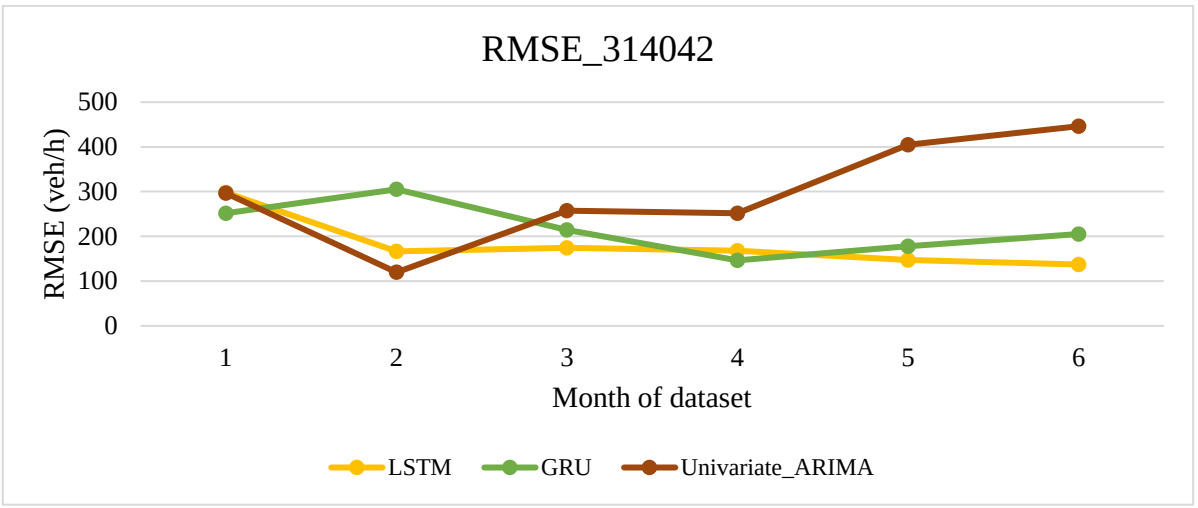


Figure 5.7: Changes of RMSE based on different data length for 3 lane road (station id 314042)

Figure 5.7 represents the traffic flow forecasting performance for the 3-lane road for 3 different algorithms by RMSE. Here, the univariate ARIMA model's RMSE value is low for the 1-month dataset and the value is approximately 300 vehicles/h, then the value decreased for the 2-month dataset and again gradually increased to the 3-month dataset and at 4 months of the dataset, the RMSE value is the approximately 250 vehicles/h and RMSE value is in peak for the 6-month dataset. For LSTM and GRU models, initially, for the 1-month dataset, the LSTM model's RMSE is higher, and the GRU model's RMSE value is lower than the 2-month dataset, LSTM model's RMSE slightly decreased, and the GRU model's RMSE a little bit increased. At the 6-month dataset, the LSTM model's performance is good,

the RMSE value is the lowest here. For the GRU model, at the 4-month dataset, the RMSE value is the lowest, and it gradually increased after adding more data.

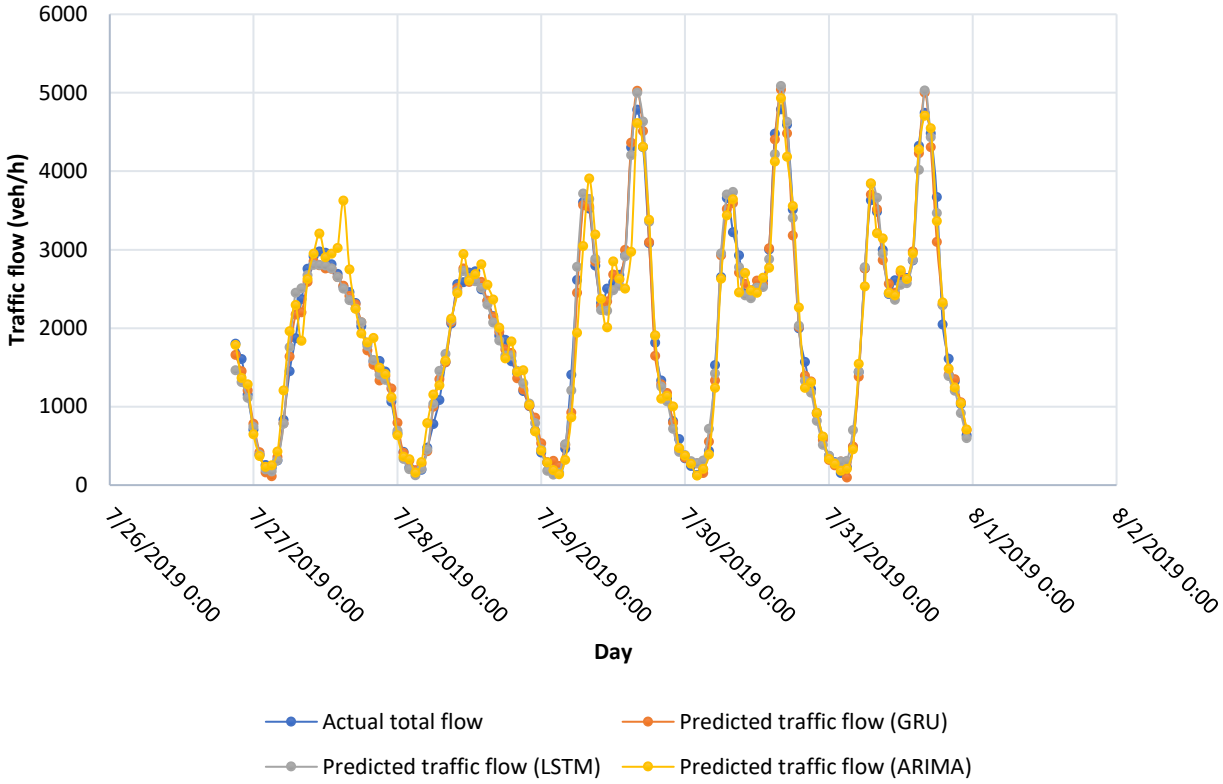


Figure 5.8: Forecasting plots of 3 lane road (station id 314042) for 4 months of datasets

Figure 5.8 illustrates the traffic forecasting plot for the 3-lane road for Sacramento County, where the 4-month dataset from April 2019 to July 2019 dataset taken as input data. Here the blue line in the graph shows the actual traffic flow, the orange line shows the multivariate GRU model's predicted traffic flow, the grey line shows the multivariate LSTM model's predicted traffic flow, and the yellow line shows the univariate ARIMA model's predicted traffic flow. The ARIMA model does not fit well, but LSTM and GRU models fit well to actual traffic flow.

5.1.3 4 Lane Road (Station id 319351)

Table 5.3: MAE, MAPE, RMSE value for 1 to 6 months of dataset by three algorithms for 4 lane road

Month of dataset	**MAE (vehicles/h)**			**MAPE (%)**			**RMSE (vehicles/h)**		
	LSTM	GRU	Univariate ARIMA	LSTM	GRU	Univariate ARIMA	LSTM	GRU	Univariate ARIMA
1	307.21	215.39	261.74	13.31	9.44	10.47	433.46	293.23	371.42
2	218.20	276.60	229.45	9.83	11.48	9.15	290.69	355.47	316.48
3	172.55	206.46	248.76	8.93	9.44	10.32	236.57	282.23	367.87
4	195.70	213.12	221.43	8.41	9.54	9.45	278.46	295.35	313.56
5	191.32	158.89	213.17	8.82	6.75	8.10	244.86	213.46	311.40
6	159.78	146.00	300.88	6.64	6.56	14.32	210.75	188.71	423.30

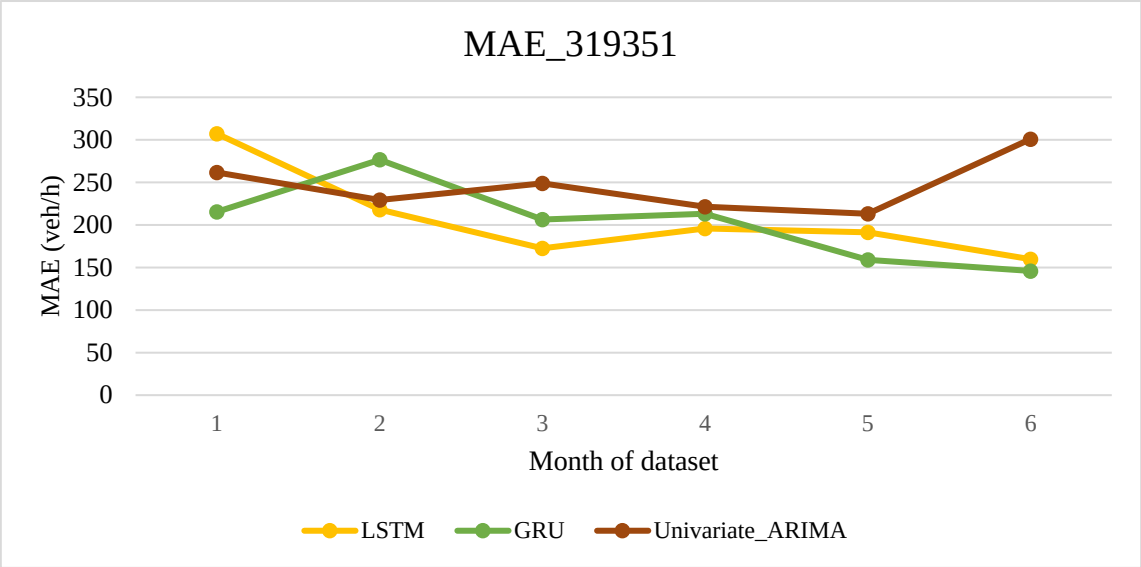


Figure 5.9: Changes of MAE based on different data length for 4 lane road (station id 319351)

Figure 5.9 shows the changes in MAE for forecasting traffic flow for the 4-lane road using three different algorithms (univariate ARIMA, LSTM, and GRU). Here for the univariate ARIMA model, the MAE value is lower in the 1-month dataset. In the 2-month dataset, the MAE value decreases and then slightly increased in the 3-month dataset, and the minimum value of MAE is in the 5-month dataset from April to August 2019 for the univariate ARIMA model. The value is in peak for the 6-month dataset in ARIMA. For the LSTM model, the MAE value is higher for the 1-month dataset, after it sharply decreased to the 3-month dataset. After that MAE value increased a small amount for the 4- month dataset, and for the 6-month dataset MAE value is the minimum. For the GRU model, the MAE value is low in the 1-month dataset, for the 2-month dataset, MAE increased to a peak, and again decreased gradually, and at the 6-month dataset, the MAE value is the minimum.

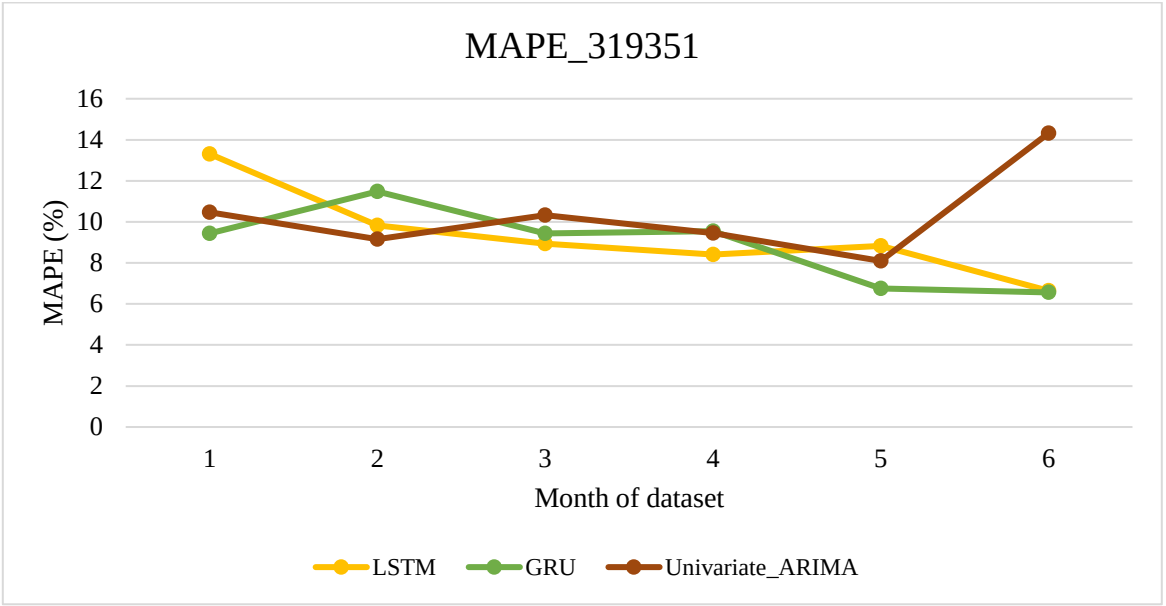


Figure 5.10: Changes of MAPE based on different data length for 4 lane road (station id 319351)

Figure 5.10 represents the changes of MAPE for forecasting traffic flow of 4 lane road by three different algorithms (univariate ARIMA, LSTM, and GRU). Here, univariate ARIMA is performing well for the 1-month traffic data, then the MAPE value decreases a bit until the 2-month dataset, at the 5-month

dataset it gives the lowest value of MAPE and at 6-month dataset the value is maximum. For the GRU model, the MAPE value is almost the same as the univariate ARIMA model performance, then the MAPE value gradually increased at the 2-month dataset, and again MAPE value decreased gradually, and the lowest value is at the 6-month dataset. For the LSTM model, at the 1-month dataset, the MAPE value is a bit higher, then for the 2-month dataset, the MAPE value reduces, and at the 6-month dataset, it shows the lowest value.

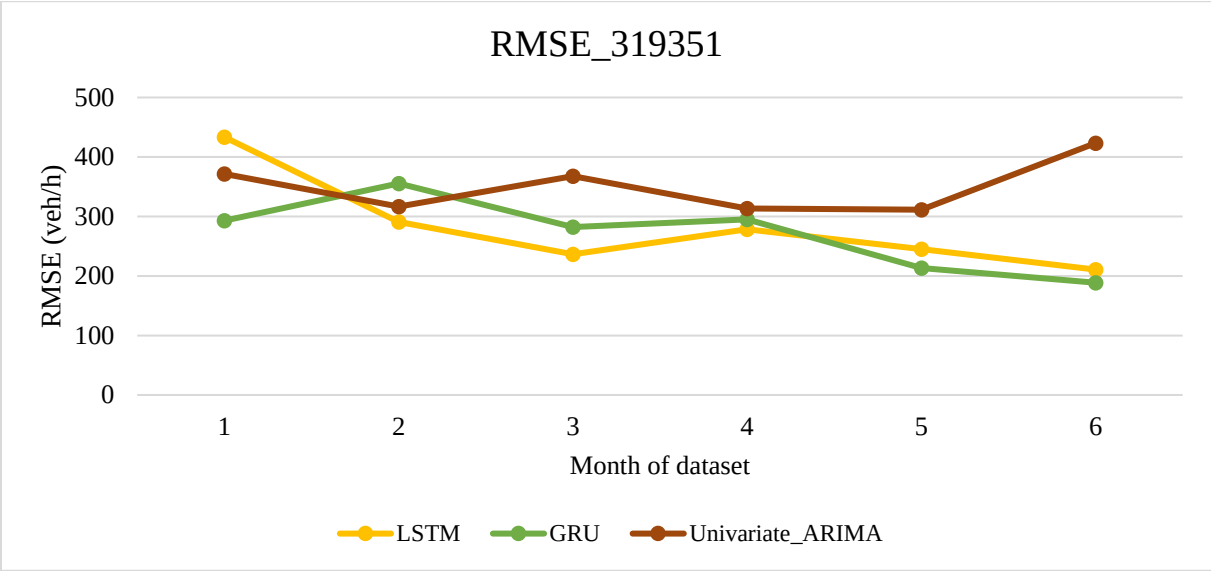


Figure 5.11: Changes of RMSE based on different data length for 4 lane road (station id 319351)

Figure 5.11 illustrates the changes of RMSE for 1 to 6 months of the datasets for the 4-lane road in 3 different algorithms (univariate ARIMA, LSTM, and GRU). Here, in univariate ARIMA, the RMSE is less for the 1-month dataset, then decreased a bit for the 2-month dataset. After fluctuating the RMSE value at the 5-month dataset, where the RMSE value is the lowest for the univariate ARIMA model and maximum at the 6-month dataset. For the GRU model, the RMSE value is lower in the 1-month dataset, and its value increased. After that, for adding data, from 2 months of the dataset to 6 months of the dataset, the RMSE value decreased. The minimum value of RMSE is at the 6-month dataset and then again increased for the 6-month dataset. For the LSTM model, at the 1-month dataset, the RMSE value is higher, then its value gradually decreases until the 6-month dataset. The RMSE value is the minimum for the 6-month dataset.

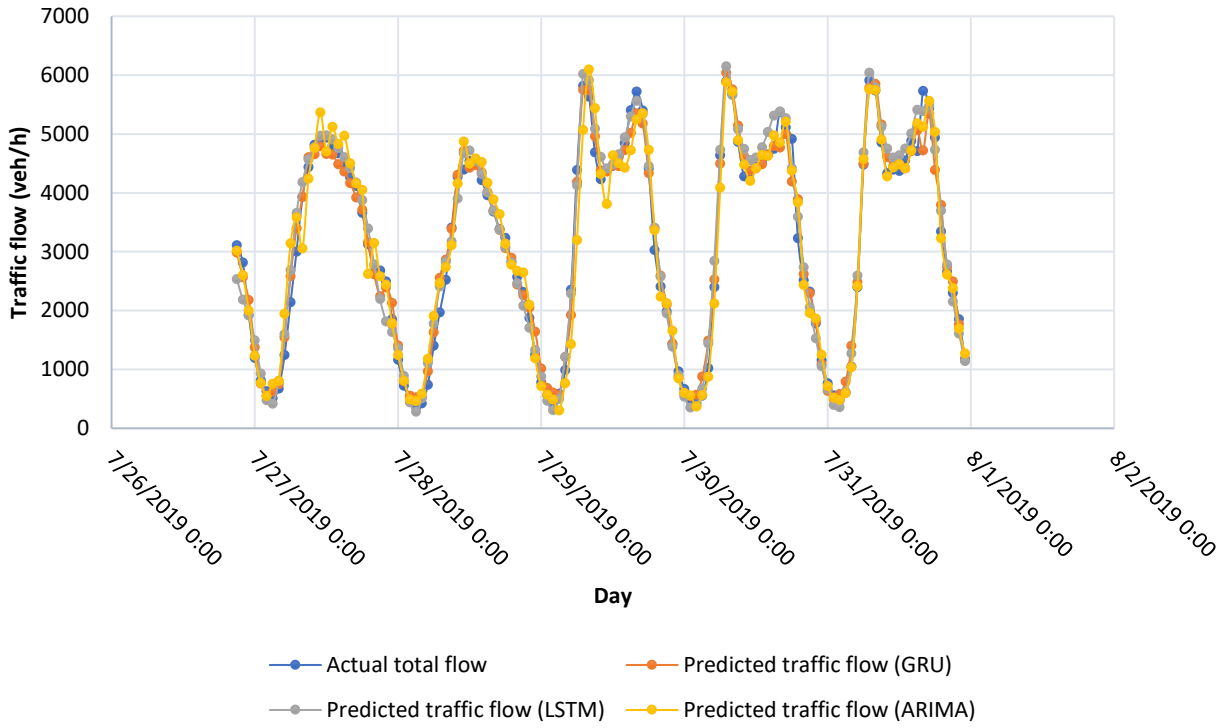


Figure 5.12: Forecasting plots of 4 lane road (station id 319351) for 4 months of datasets

Figure 5.12 represents the forecasting plots for the 4-lane road from Placer County. Here, the blue line shows the actual traffic flow. The orange line shows the predicted traffic flow by the multivariate GRU model. The grey line represents the predicted traffic flow by the multivariate LSTM model, and the yellow line represents the predicted traffic flow (vehicles/h) by the univariate ARIMA model. Overall, LSTM and GRU models give better fits, but the univariate ARIMA model did not fit well.

According to all three selected road-type data, overall, we can get better performance for the 6-month datasets in terms of MAE, MAPE, and RMSE in LSTM and GRU multivariate models in this experiment. We will use the 6-month datasets from District 3 for further experiments. Also, we will do another experiment in transfer learning models for checking how transfer learning models perform with both 1-month and 6-month source datasets. Then, we can get the complete answer to research question 1.

5.2 Effects of weather data with traffic dataset

We will get the answer to research question 2 by following experiments where additional weather data were added and made separate datasets as traffic data with weather data and traffic data without weather data. For the experiment, six-month traffic datasets (April 2019 to September 2019) are taken as at 6-month datasets, we get minimum values of MAE, MAPE, and RMSE in terms of all 3 roads from District 3 in the previous experiment. Three different roads were selected from District 3. The traffic-predicting results for each road were obtained using the multivariate LSTM and GRU models. A total 5 number of runs in each model were done, and results were presented in the following tables and boxplots for MAE, MAPE, and RMSE.

5.2.1 2 Lane Road (Station id 313190)

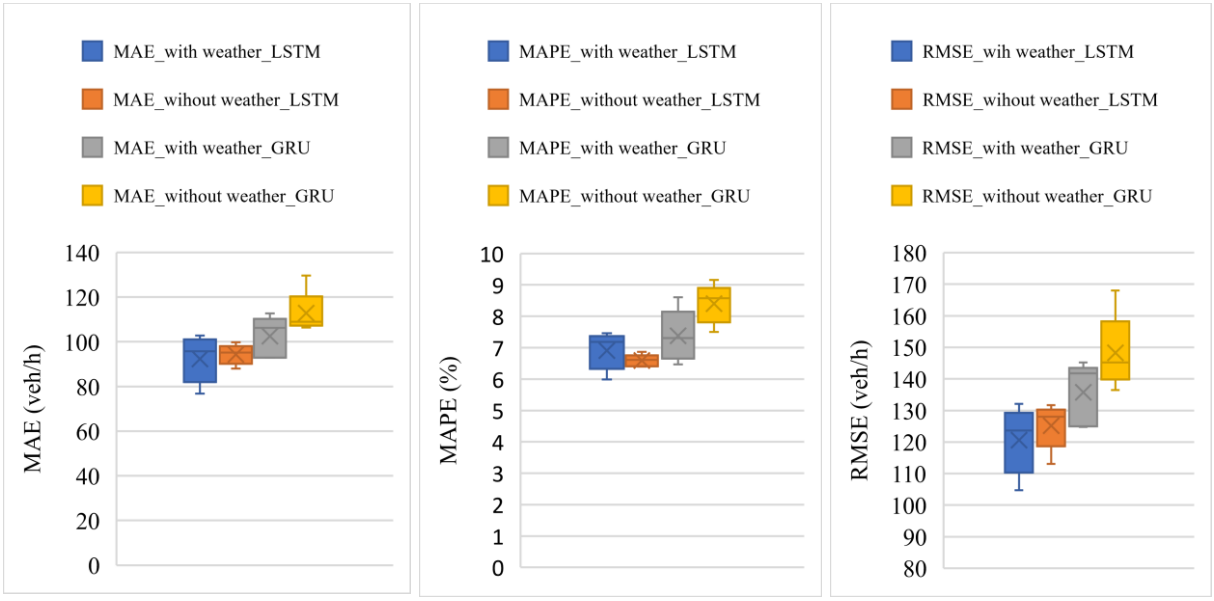
Table 5.4: 2 lane road MAE, MAPE for with and without weather models

Number of Runs	LSTM_MAE (vehicles/h)	GRU_MAE (vehicles/h)	LSTM_MAPE (%)	GRU_MAPE (%)

	With weather	Without weather	With weather	Without weather	With weather	Without weather	With weather	Without weather
1	102.76	96.29	106.29	111.06	7.46	6.64	7.31	8.11
2	76.80	99.74	107.66	108.12	5.99	6.87	7.69	7.50
3	87.00	92.32	112.77	109.05	6.65	6.40	8.61	8.58
4	95.73	87.97	92.93	129.70	7.18	6.61	6.83	9.16
5	99.50	95.14	92.85	106.41	7.27	6.42	6.47	8.65

Table 5.5: 2 lane road RMSE for with and without weather models

Number of Runs	**LSTM_RMSE (vehicles/h)**	**LSTM_RMSE (vehicles/h)**	**GRU_RMSE (vehicles/h)**	**GRU_RMSE (vehicles/h)**
	With weather	Without weather	With weather	Without weather
1	132.08	127.97	141.78	148.43
2	104.71	131.68	141.80	145.26
3	116.01	124.22	145.18	143.21
4	123.64	113.07	125.19	167.95
5	126.46	128.85	124.78	136.47



(a) MAE (b) MAPE (c) RMSE

Figure 5.13: MAE, MAPE and RMSE plots for 2 lane road for 5 runs

Figure 5.13 (a) shows the box plot of MAE for 2-lane road 313190 with weather data and without weather data for both LSTM and GRU models. The mean value of with weather data model for both models performs better than without weather data. Overall LSTM model is performing better than the GRU model in terms of MAE.

In figure 5.13 (b), in terms of MAPE, the with-weather model does not perform well as the without-weather model in LSTM. For the GRU model, the with-weather model is performing better than the without-weather model. Overall, the LSTM model is performing better than the GRU model.

Figure 5.13 (c) shows the boxplots for the 2-lane road in Sacramento, where traffic flow forecasting performance by RMSE for both multivariate models. Overall, the with-weather model for both

algorithms performs better than the without-weather data models. Also, the LSTM model is performing better than the GRU model.

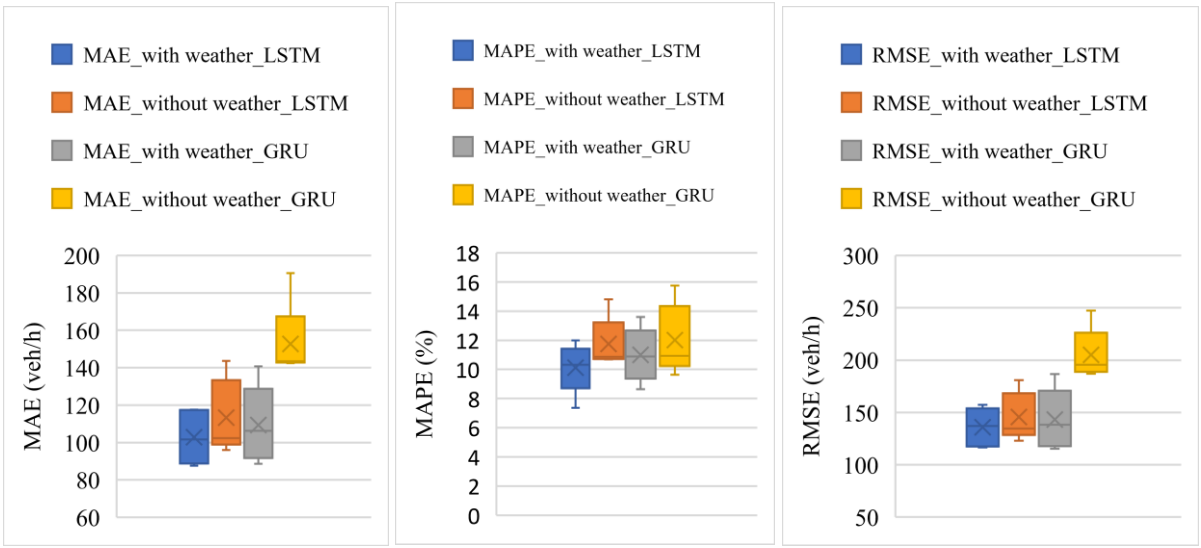
5.2.2 3 Lane Road (Station id 314042)

Table 5.6: 3 lane road MAE, MAPE for with and without weather models

Number of Runs	LSTM_MAE (vehicles/h)		GRU_MAE (vehicles/h)		LSTM_MAPE (%)		GRU_MAPE (%)	
	With weather	Without weather	With weather	Without weather	With weather	Without weather	With weather	Without weather
1	87.59	102.34	106.23	190.52	7.37	10.72	10.89	12.92
2	90.03	96.07	88.70	143.05	10.85	10.69	8.64	10.80
3	117.65	101.78	140.67	144.48	10.06	10.87	11.76	9.64
4	117.30	143.63	94.84	143.42	12.00	14.81	10.12	10.93
5	101.67	123.03	116.76	142.49	10.32	11.63	13.60	15.75

Table 5.7: 3 lane road RMSE for with and without weather models

Number of Runs	LSTM_RMSE (vehicles/h)	LSTM_RMSE (vehicles/h)	GRU_RMSE (vehicles/h)	GRU_RMSE (vehicles/h)
	With weather	Without weather	With weather	Without weather
1	118.56	134.70	138.15	247.45
2	116.44	122.98	115.27	205.00
3	157.28	134.06	186.76	190.55
4	150.38	180.97	120.43	195.72
5	137.06	155.69	154.63	186.87



(a) MAE (b) MAPE (c) RMSE

Figure 5.14: MAE, MAPE and RMSE for 3 lane road for 5 runs

Figure 5.14 (a) illustrates the boxplots of the 3-lane road in Sacramento, where traffic flow forecasting performance according to MAE is available. The with-weather data models are performing better than the without-weather data models. Overall, the LSTM model is performing better than the GRU model. In figure 5.14 (b), the MAPE value for 5 runs is shown in the boxplots where the LSTM model performs

better than the GRU model. In both algorithms, the with-weather models perform better than the without-weather models.

Figure 5.14 (c) represents the boxplots of RMSE values for the 3-lane road with-weather and without-weather datasets for both LSTM and GRU algorithms. Overall, the with-weather model's RMSE value is lower compared to the without-weather data models. Also, overall, the LSTM model's performance is better than the GRU model's performance.

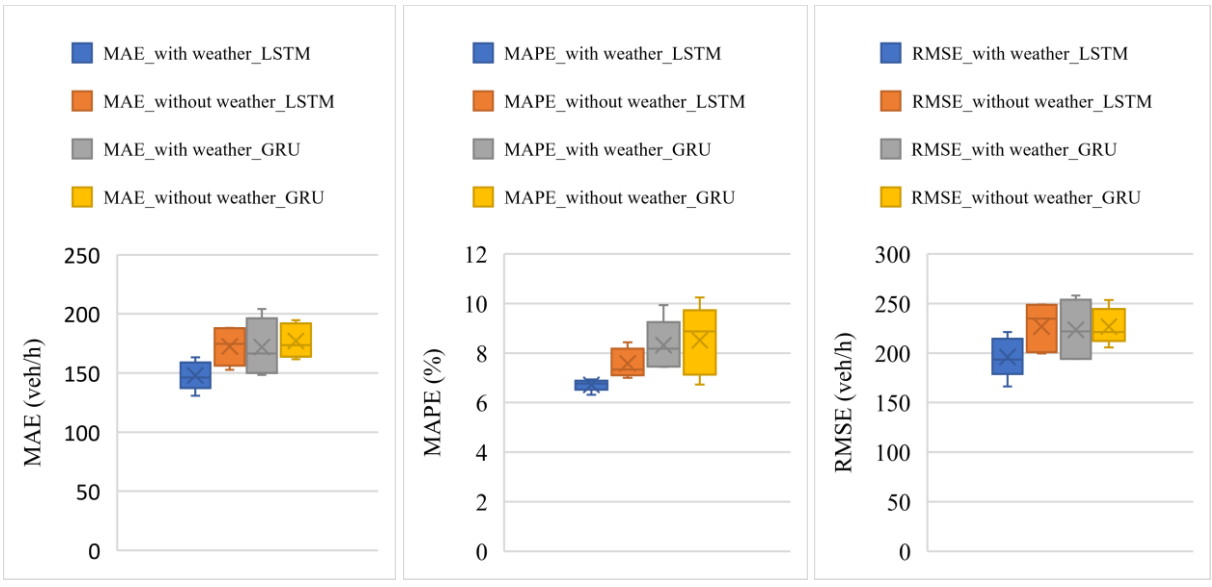
5.2.3 4 Lane Road (Station id 319351)

Table 5.8: 4 lane road MAE, MAPE for with and without weather models

Number of Runs	LSTM_MAE (vehicles/h)		GRU_MAE (vehicles/h)		LSTM_MAPE (%)		GRU_MAPE (%)	
	With weather	Without weather	With weather	Without weather	With weather	Without weather	With weather	Without weather
1	130.87	159.86	151.56	189.10	6.72	7.34	8.17	10.24
2	146.33	152.81	166.61	161.53	6.32	6.99	7.446	7.52
3	143.99	174.62	204.15	194.71	6.93	7.21	9.93	8.88
4	154.64	188.26	188.41	165.98	6.77	8.43	8.57	6.73
5	163.39	187.54	148.48	173.61	6.83	7.918	7.47	9.22

Table 5.9: 4 lane road RMSE for with and without weather models

Number of Runs	LSTM_RMSE (vehicles/h)	LSTM_RMSE (vehicles/h)	GRU_RMSE (vehicles/h)	GRU_RMSE (vehicles/h)
	With weather	Without weather	With weather	Without weather
1	166.26	201.74	194.32	235.39
2	193.40	199.69	221.85	205.80
3	191.76	234.67	258.16	253.44
4	207.37	249.00	249.59	221.10
5	221.31	248.11	193.96	218.41



(a) MAE (b) MAPE (c) RMSE

Figure 5.15: MAE, MAPE and RMSE for 4 lane road for 5 runs

Figure 5.15 (a) illustrates the MAE values for the with-weather and the without-weather models for both LSTM and GRU models for the 3-lane road, where overall with-weather models are performing better than without-weather models. Overall, the LSTM model gives better performance than the GRU model. Figure 5.15 (b) shows the boxplot of MAPE for the with-weather and the without-weather datasets for the 3-lane road in Sacramento by both deep learning algorithms, where both LSTM and GRU with weather models are performing better than without-weather data models. Also, LSTM models are performing better than GRU models. Figure 5.15 (c) represents the boxplot of the model performance by RMSE for the 3-lane road in Sacramento by LSTM and GRU model. The with-weather models are performing better than the without-weather data models.

According to all 3 road types, overall, with-weather traffic data models perform well in terms of MAE, MAPE, and RMSE in both LSTM and GRU models. Moreover, the LSTM model has better performance than the GRU model in this experiment. We will use the LSTM model in transfer learning models for traffic forecasting with weather traffic data models.

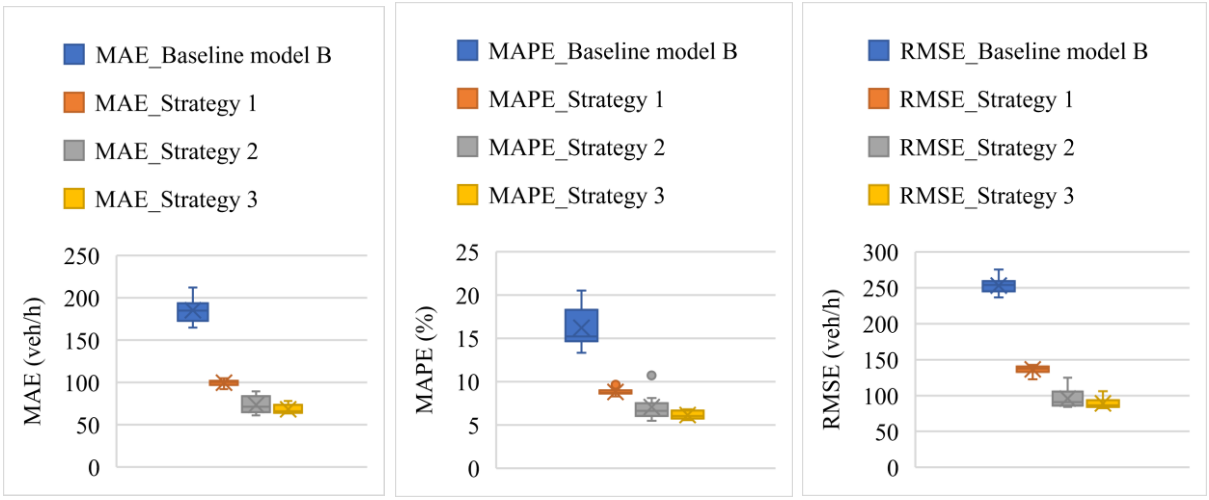
5.3 Transfer Learning Results

5.3.1 Scenario 1 Transfer Learning (LSTM Model)

Table 5.10: MAE, MAPE and RMSE for scenario 1 transfer learning

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	188.83	100.84	72.96	64.04	17.79	8.74	7.34	5.54	236.57	137.19	91.52	86.61
2	212.20	95.88	82.04	63.91	20.53	8.61	8.10	5.85	256.61	133.10	107.20	82.33
3	177.59	101.83	74.23	66.99	15.31	9.00	6.75	6.20	262.96	140.27	96.64	84.68
4	174.66	97.72	69.97	64.99	15.07	8.86	6.39	5.88	243.89	132.01	90.50	83.51
5	207.88	101.72	65.20	64.24	19.89	8.68	6.04	5.74	254.8	137.54	87.67	88.20
6	165.02	102.44	63.84	68.76	13.32	9.01	5.99	6.55	245.42	141.88	84.09	86.34
7	185.54	92.18	68.62	75.27	15.13	8.28	6.54	6.67	275.48	122.63	85.96	96.70
8	167.85	98.21	89.48	63.38	13.82	8.65	7.30	5.74	246.15	139.92	124.63	84.21
9	188.09	104.87	89.52	78.13	16.62	9.64	10.73	6.80	253.11	137.09	104.88	106.18
10	184.61	100.07	61.04	73.06	14.93	8.67	5.47	6.65	258.02	142.53	85.86	92.48

Scenario 1 accuracy plots for 2 lane road:



(a) MAE (b) MAPE (c) RMSE

Figure 5.16: Scenario 1 MAE, MAPE and RMSE

Figure 5.16 shows the accuracy of transfer learning scenario-1 with three strategies and the forecasting performance of target baseline model B. Here, the 2-lane road (313190) from district 3 is selected for the source model, and the 2-lane road (601722) from district 6 is selected for the target dataset. According to MAE, MAPE, and RMSE, baseline target forecasting model B does not give good predictions. Strategy 1 transfer learning performance is slightly higher than strategies 2 and 3 transfer learning. Overall, strategy-3 transfer learning gives the best accuracy compared to other strategies.

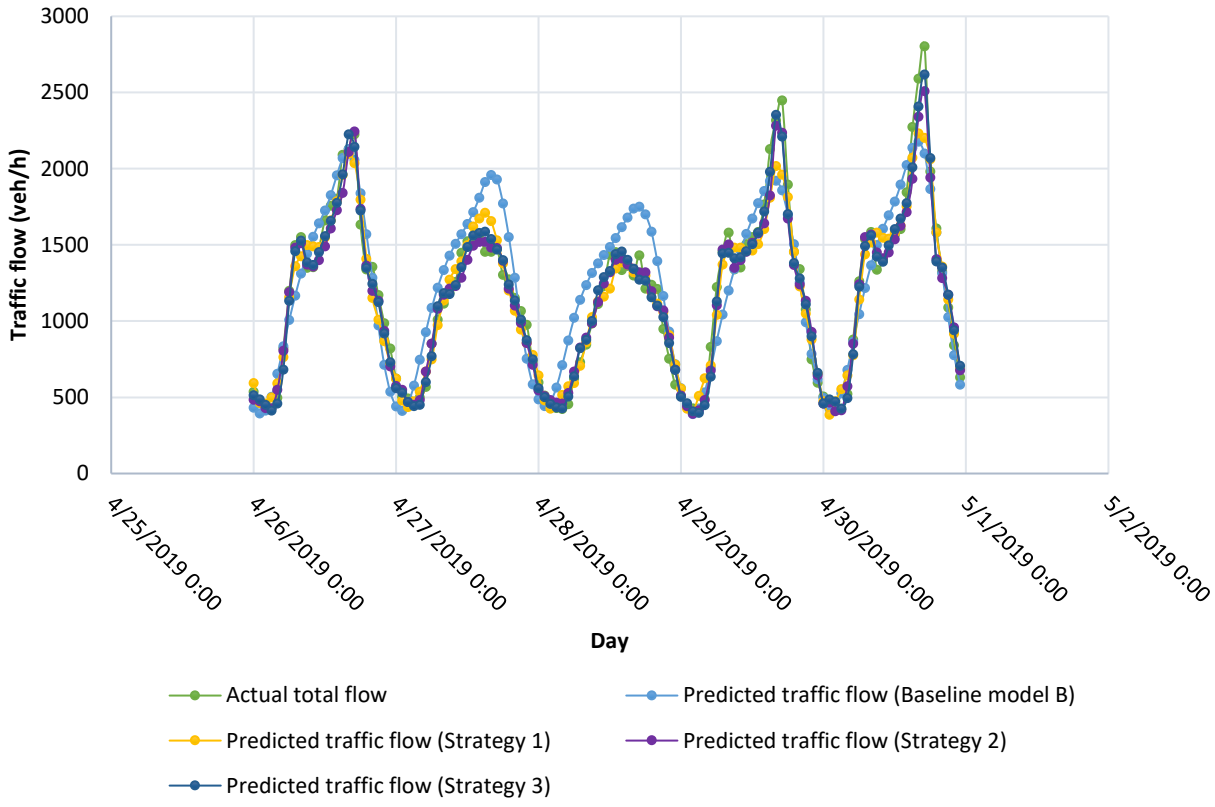


Figure 5.17: Scenario 1 prediction plots for 2 lane to 2 lane transfer learning

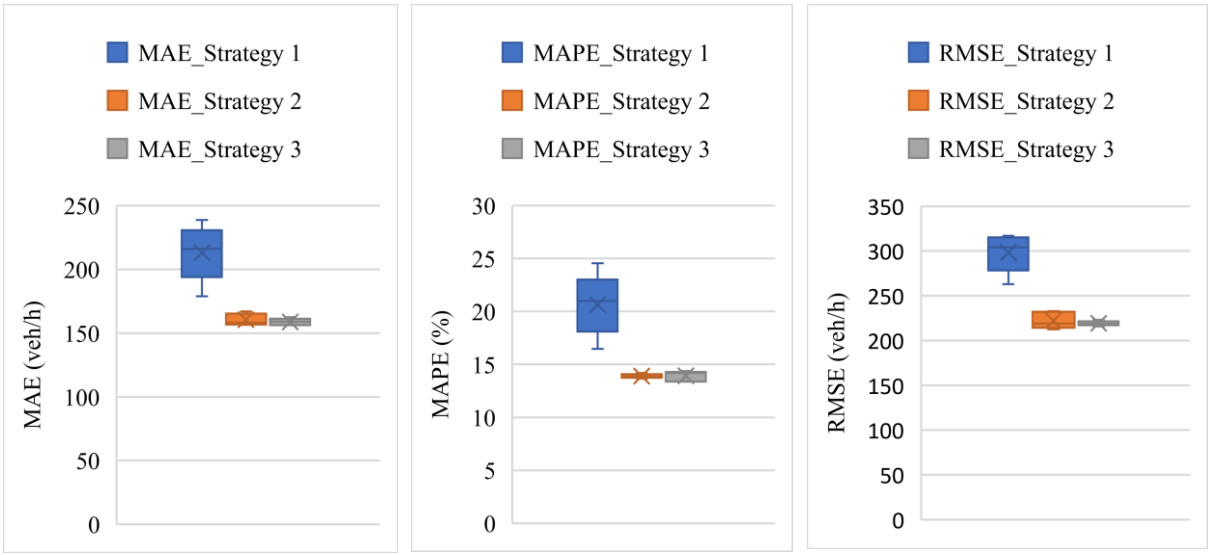
Figure 5.17 illustrates the forecasting graphs for the forecasting of target baseline model B and the forecasting of scenario-1 transfer learning. On this 2-lane road, the light-green colored line represents actual traffic flow. The light-blue line indicates the forecasted plot of target baseline model B, which did not provide a good forecast. The yellow-colored line shows forecasted traffic flow by strategy-1 transfer learning (fine-tuned), at some point, it gives under or overfits to actual traffic flow. The purple line represents the predicted traffic flow using strategy-2 transfer learning (fine-tuned), while the deep blue line represents the forecast using strategy 3 (without fine-tuning). Overall, both strategies 2 and 3 give a better fit.

Scenario 1 TL with 1-month source dataset: For research question 1, we did another experiment with 1-month source dataset. April 2019 dataset from district 3 for 2 lane road is taken. Following result will show the transfer learning for 1-month source data.

Table 5.11: MAE, MAPE and RMSE for scenario 1 transfer learning with 1 month source data

Number of runs	MAE (vehicles/h)			MAPE (%)			RMSE (vehicles/h)		
	Strat. 1	Strat. 2	Strat. 3	Strat. 1	Strat. 2	Strat. 3	Strat. 1	Strat. 2	Strat. 3
1	209.09	156.49	156.39	19.77	13.82	13.37	293.75	212.46	215.56
2	178.74	157.12	156.28	16.46	13.64	13.39	262.84	216.03	218.41
3	216.17	163.4	159.03	20.98	14.20	14.20	304.08	231.14	219.62

4	222.45	158.30	159.67	21.49	13.98	14.21	313.62	218.76	218.61
5	238.69	166.97	162.60	24.54	13.87	14.39	316.96	232.91	223.28



(a) MAE (b) MAPE (c) RMSE

Figure 5.18: Scenario 1 MAE, MAPE and RMSE with 1 month source data

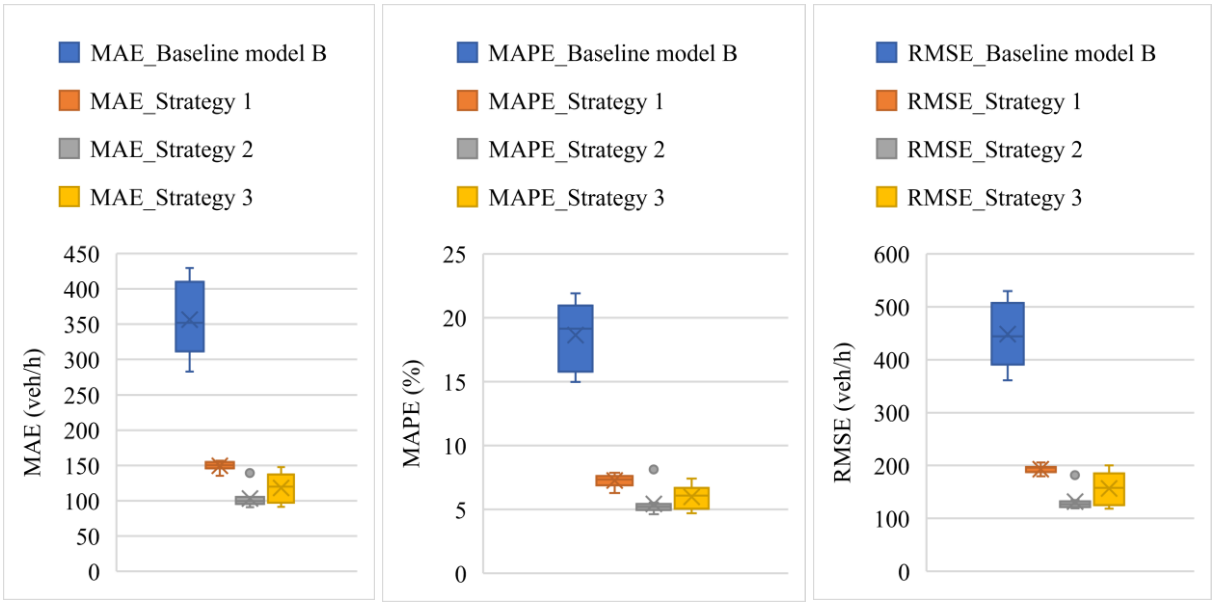
According to the above experiments in Figures 5.16 and 5.18, for both the 6-month source dataset and 1-month source dataset models, the 6-month source-dataset transfer learning model gives better accuracy of traffic flow forecasting than the 1-month source dataset model. We can summarize that more source data model gives better traffic flow forecasting, which gives the answer to research question 1.

5.3.2 Scenario 2 Transfer Learning (LSTM Model)

Table 5.12: MAE, MAPE and RMSE for scenario 2 transfer learning

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	328.20	147.77	99.16	119.29	17.53	6.95	5.36	5.90	425.99	190.39	123.38	157.62
2	409.24	148.87	94.80	98.90	21.92	7.43	4.98	5.09	520.76	195.75	124.01	127.01
3	306.63	148.69	104.20	101.62	15.24	7.29	5.54	5.25	391.12	197.31	131.77	133.59
4	283.17	156.54	105.36	135.99	14.98	7.87	5.40	6.61	360.83	192.93	130.33	183.29
5	313.41	155.76	95.51	147.67	15.95	7.37	4.92	7.42	390.38	205.96	121.28	200.23
6	378.13	135.31	99.52	91.37	19.17	6.30	5.01	4.71	477.03	179.89	129.92	120.48
7	429.64	154.69	105.18	121.02	21.74	7.80	5.37	6.30	529.95	196.04	134.92	158.59
8	336.46	153.80	139.05	141.85	20.11	7.56	8.14	6.91	433.34	198.03	181.90	190.52
9	412.44	140.11	90.90	130.39	20.70	6.68	4.63	6.58	502.82	180.08	119.18	175.39
10	367.88	151.89	96.37	92.72	19.13	7.34	5.07	5.00	454.50	196.02	120.87	118.28

Scenario 2 accuracy plots for 3 lane road:



(a) MAE (b) MAPE (c) RMSE

Figure 5.19: Scenario 2 MAE, MAPE and RMSE

Figure 5.19 represents the performance of traffic forecasting of target baseline model B and scenario-2 transfer learning forecasting in three strategies. For the source model, the 3-lane road (314042) is selected as the source dataset, and the 3-lane road (602538) dataset is selected as the target dataset. According to MAE, MAPE, and RMSE, the forecasting performance of target baseline model B is not good compared to transfer-learned models. Here, overall, strategy 2 performs better in all three strategies of transfer learning models. Mean MAE, MAPE, and RMSE values are the lowest for strategy-2 transfer learning which is the fine-tuned model with the pre-trained model's all layers kept unfrozen.

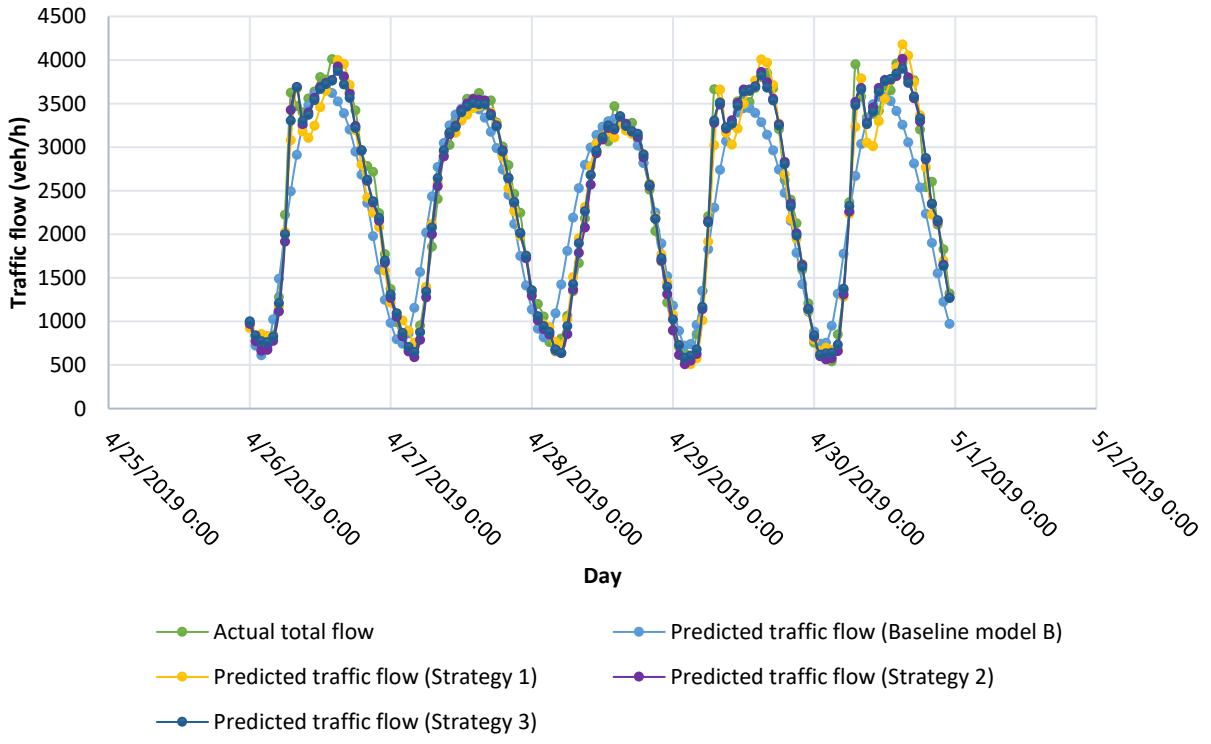


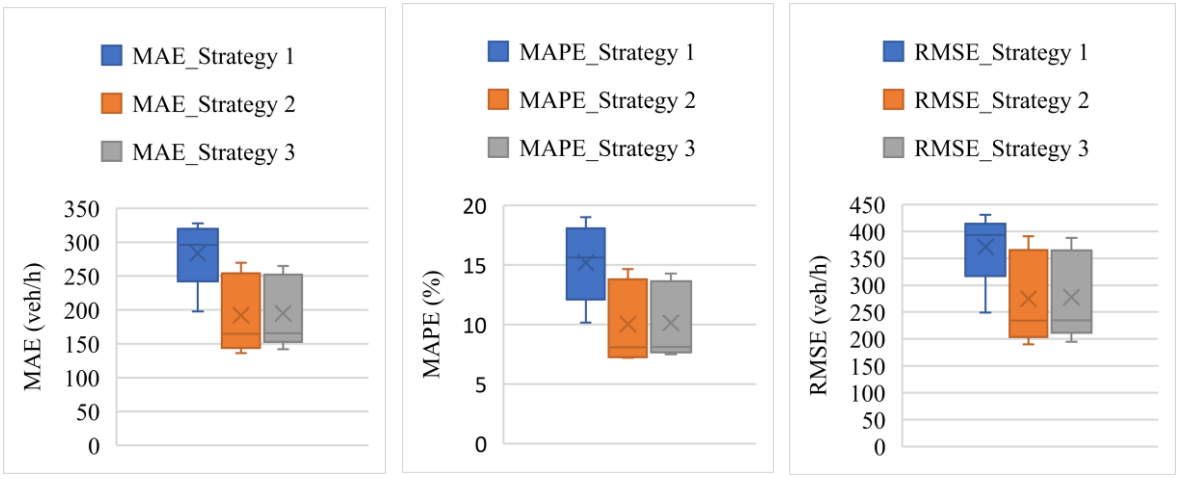
Figure 5.20: Scenario 2 prediction plots for 3 lane to 3 lane transfer learning

Figure 5.20 shows the forecasting graphs for the target baseline model B and scenario-2 transfer learning. On this graph, the light-green colored line represents actual traffic flow. The light-blue line indicates the forecasted plot of target baseline model B, which did not provide a better forecast. The yellow-colored line shows forecasted traffic flow by the strategy-1 transfer learning (fine-tuned) at some point, giving under or overfits to actual traffic flow. The purple line represents the predicted traffic flow using strategy-2 transfer learning (fine-tuned), while the deep blue line represents the forecasting plot using strategy-3 (without fine-tuning). Overall, both strategies 2 and 3 give better fits.

Scenario 2 TL with 1-month source dataset: For research question 1, we did another experiment with 1-month source dataset. April 2019 dataset from district 3 for 3 lane road is taken. Following result will show the transfer learning for 1-month source data.

Table 5.13: MAE, MAPE and RMSE for scenario 2 transfer learning with 1 month source data

Number of runs	MAE (vehicles/h)			MAPE (%)			RMSE (vehicles/h)		
	Strat. 1	Strat. 2	Strat. 3	Strat. 1	Strat. 2	Strat. 3	Strat. 1	Strat. 2	Strat. 3
1	311.21	164.73	165.75	17.10	8.07	8.10	397.21	233.89	234.90
2	295.93	151.37	163.49	14.00	7.32	7.83	385.56	216.86	228.89
3	197.64	136.01	141.79	10.15	7.18	7.48	248.79	190.18	194.53
4	327.97	269.68	264.97	19.01	14.63	14.27	430.83	390.72	387.87
5	286.96	238.35	239.50	15.60	12.91	12.95	393.29	340.50	341.36



(a) MAE (b) MAPE (c) RMSE

Figure 5.21: Scenario 2 MAE, MAPE and RMSE with 1 month source data

According to the above experiments in Figures 5.19 and 5.21, for both the 6-month source dataset and 1-month source dataset models, the 6-month source-dataset transfer learning model gives better accuracy of traffic flow forecasting than the 1-month source dataset model for 3 lane road. We can summarize that more source data give better traffic flow forecasting, which gives the answer to research question 1.

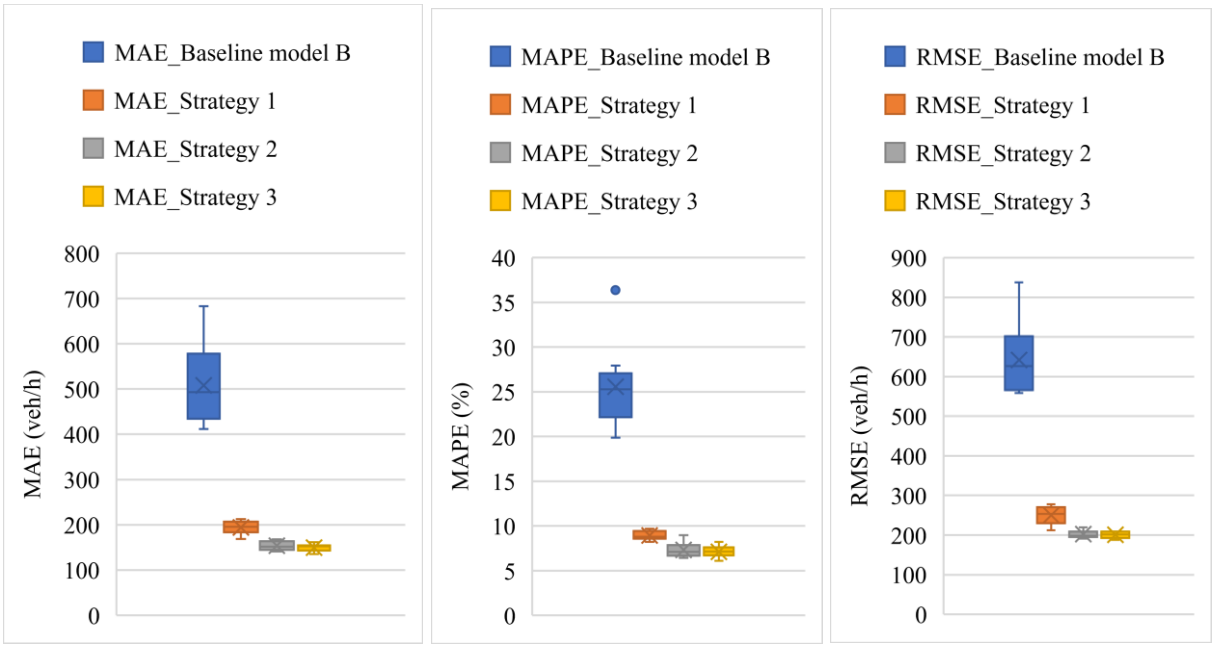
5.3.3 Scenario 3 Transfer Learning (LSTM Model)

Table 5.14: MAE, MAPE and RMSE for scenario 3 transfer learning

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3

1	484.03	197.25	143.30	145.40	24.00	9.36	6.42	7.05	624.59	255.62	194.90	193.06
2	437.40	212.34	168.08	155.99	22.56	9.25	8.08	7.57	563.87	272.75	219.36	209.04
3	465.31	171.94	149.05	135.88	23.76	8.57	7.31	6.16	595.16	211.95	202.16	192.91
4	506.68	189.76	158.33	161.31	26.58	8.20	7.80	8.21	638.67	250.52	195.34	209.04
5	594.00	186.90	147.06	151.71	26.77	9.70	6.72	6.89	710.32	230.57	195.60	209.17
6	682.95	206.76	140.77	153.04	36.37	9.62	6.55	7.69	837.92	256.67	195.44	199.09
7	411.33	168.84	153.57	148.96	19.86	8.55	6.93	7.22	566.01	227.04	205.75	200.14
8	502.79	206.65	165.96	152.96	26.79	8.72	7.50	7.02	628.55	277.65	219.74	202.71
9	425.50	205.06	162.53	154.09	20.97	8.64	8.96	7.27	558.83	269.08	201.71	208.08
10	572.98	193.46	144.83	135.32	27.91	8.86	6.91	6.11	698.49	247.49	191.03	188.19

Scenario 3 accuracy plots for 4 lane road:



(b) MAE (b) MAPE (c) RMSE

Figure 5.22: Scenario 3 MAE, MAPE and RMSE

Figure 5.22 illustrates the performance of traffic forecasting for target baseline model B and scenario-3 transfer-learned forecasting. Here the 4-lane road (319351) from district 3 is selected as the source dataset, and the 4-lane road (601348) from district 6 is selected as the target dataset. According to MAE, MAPE, and RMSE, target baseline forecasting model B does not give good predictions. Overall, strategies 2 and 3 transfer learned forecasting give better accuracy compared to strategy 1. Also, transfer-learned models perform better than baseline model forecasting.

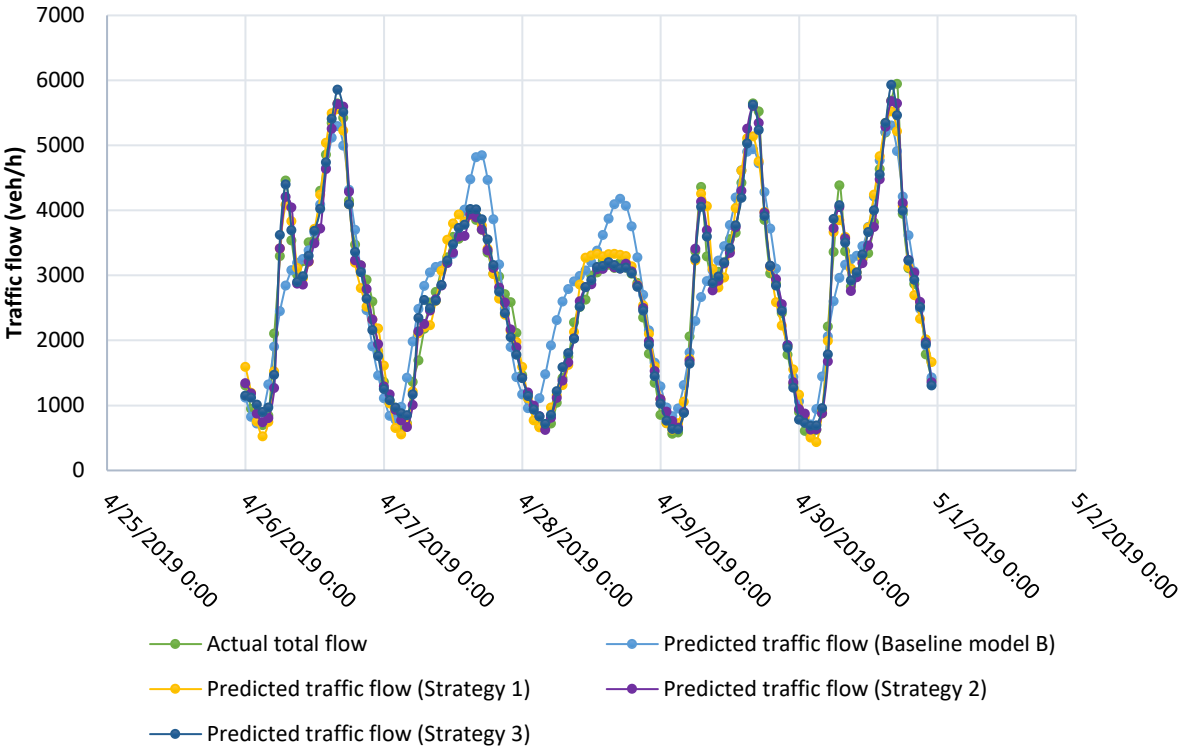


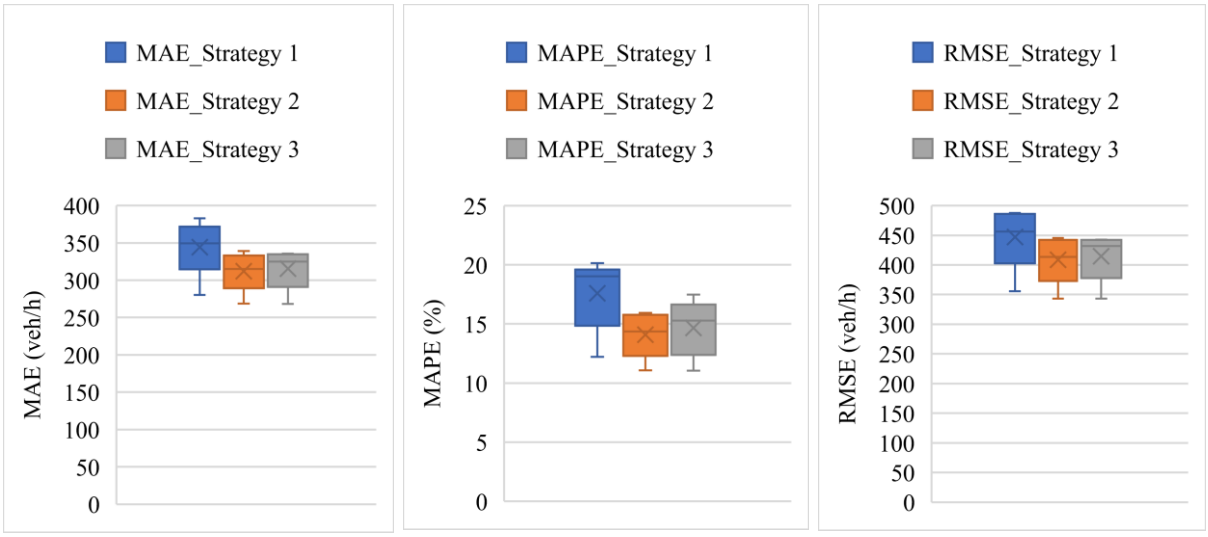
Figure 5.23: Scenario 3 prediction plots for 4 lane to 4 lane transfer learning

Figure 5.23 shows the forecasting graphs for target baseline model B and scenario-3 transfer learning. In this graph, the light-green colored line represents actual traffic flow. The light-blue line indicates the forecasted plot of target baseline model B, which did not provide a better forecast. The yellow-colored line shows forecasted traffic flow by the strategy-1 transfer learning (fine-tuned) at some point gives under or overfits to actual traffic flow. The purple line represents the predicted traffic flow using strategy-2 transfer learning (fine-tuned), while the deep blue line represents the forecast using strategy-3 (without fine-tuned). Overall, both strategies 2 and 3 give a better fit.

Scenario 3 TL with 1-month source dataset: For research question 1, we did another experiment with 1-month source dataset. April 2019 dataset from district 3 for 4 lane road is taken. Following result will show the transfer learning for 1-month source data.

Table 5.15: MAE, MAPE and RMSE for scenario 3 transfer learning with 1 month source data

Number of runs	MAE (vehicles/h)			MAPE (%)			RMSE (vehicles/h)		
	Strat. 1	Strat. 2	Strat. 3	Strat. 1	Strat. 2	Strat. 3	Strat. 1	Strat. 2	Strat. 3
1	382.90	339.21	334.44	20.13	15.60	15.83	487.43	445.26	441.70
2	349.44	314.86	313.92	17.49	13.53	13.68	449.87	413.32	412.21
3	280.35	268.45	268.09	12.21	11.07	11.05	355.65	343.36	343.12
4	348.76	310.02	335.45	19.03	14.37	17.47	456.08	403.03	431.99
5	360.63	327.28	324.84	19.09	15.92	15.29	484.85	438.58	442.69



(a) MAE (b) MAPE (c) RMSE

Figure 5.24: Scenario 3 MAE, MAPE and RMSE with 1 month source data

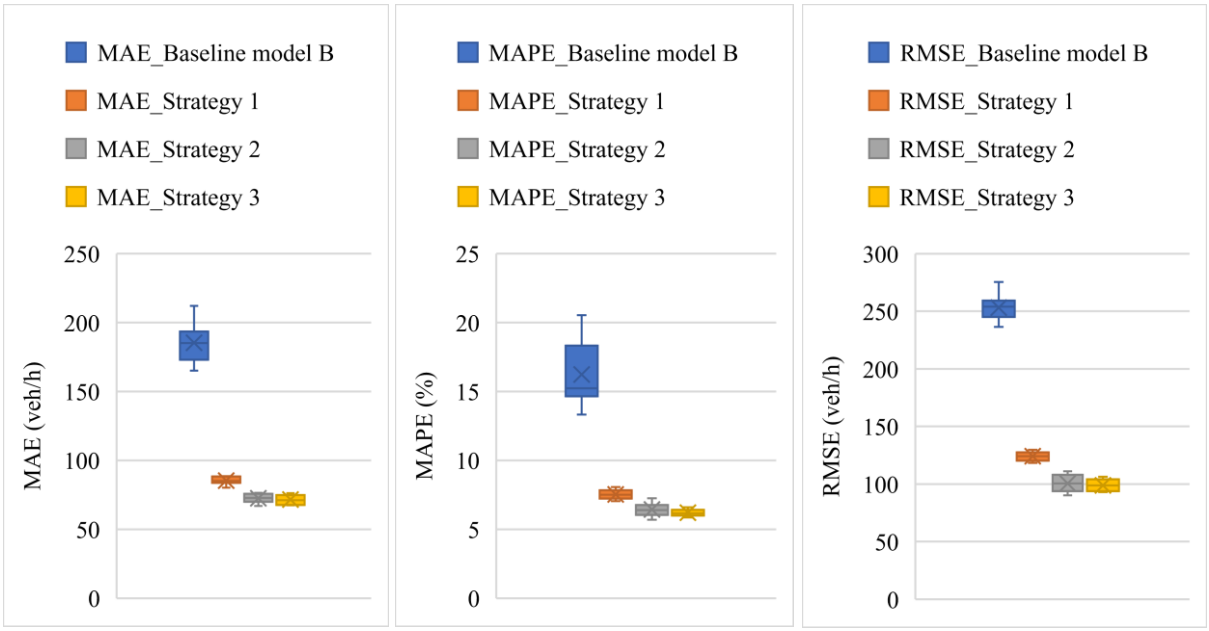
According to the above experiments in Figures 5.22 and 5.24, for both the 6-month source dataset and 1-month source dataset models, the 6-month source-dataset transfer learning model gives better accuracy than the 1-month source dataset model for the 4-lane road. We can summarize that more source data give better traffic flow forecasting. That gives the answer to research question 1.

5.3.4 Scenario 4 Transfer Learning (LSTM Model)

Table 5.16: MAE, MAPE and RMSE for scenario 4 transfer learning

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	188.83	83.27	69.97	67.45	17.79	7.23	6.65	6.00	236.57	118.40	90.26	92.99
2	212.20	84.56	70.60	67.38	20.53	7.60	6.44	5.86	256.61	120.87	97.00	94.00
3	177.59	88.25	69.94	67.54	15.31	7.82	5.93	6.01	262.96	128.83	97.47	93.14
4	174.66	80.26	72.16	71.06	15.07	7.05	7.24	5.95	243.89	118.42	90.83	99.50
5	207.88	85.28	76.10	70.95	19.89	7.43	7.05	6.10	254.80	123.57	104.53	98.07
6	165.02	88.56	74.38	76.06	13.32	8.07	6.29	6.60	245.42	126.77	107.92	103.98
7	185.54	84.94	75.57	76.14	15.13	7.58	6.09	6.60	275.48	124.66	111.03	104.03
8	167.85	85.60	66.86	73.08	13.82	7.38	5.68	6.34	246.15	127.13	94.91	100.34
9	188.09	88.40	73.29	70.42	16.62	7.82	6.35	6.21	253.11	129.79	103.41	95.92
10	184.61	83.91	76.34	74.54	14.93	7.21	6.64	6.28	258.02	122.91	108.29	106.39

Scenario 4 accuracy plots for 3 to 2 lane road:



(a) MAE (b) MAPE (c) RMSE

Figure 5.25: MAE, MAPE and RMSE for 2 lane transfer learning

Figure 5.25 illustrates the forecasting performances of target baseline model B and scenario-4 forecasting with transfer learning three strategies. Here, the 3-lane road (314042) dataset is taken for the source model, and the 2-lane road (601722) dataset is taken for the target dataset. According to MAE, MAPE, and RMSE, target baseline forecasting model B does not give good predictions. Overall, strategy-3 transfer-learned forecasting gives the best accuracy compared to the other two strategies. Overall, transfer-learned forecasting is better than baseline model forecasting.

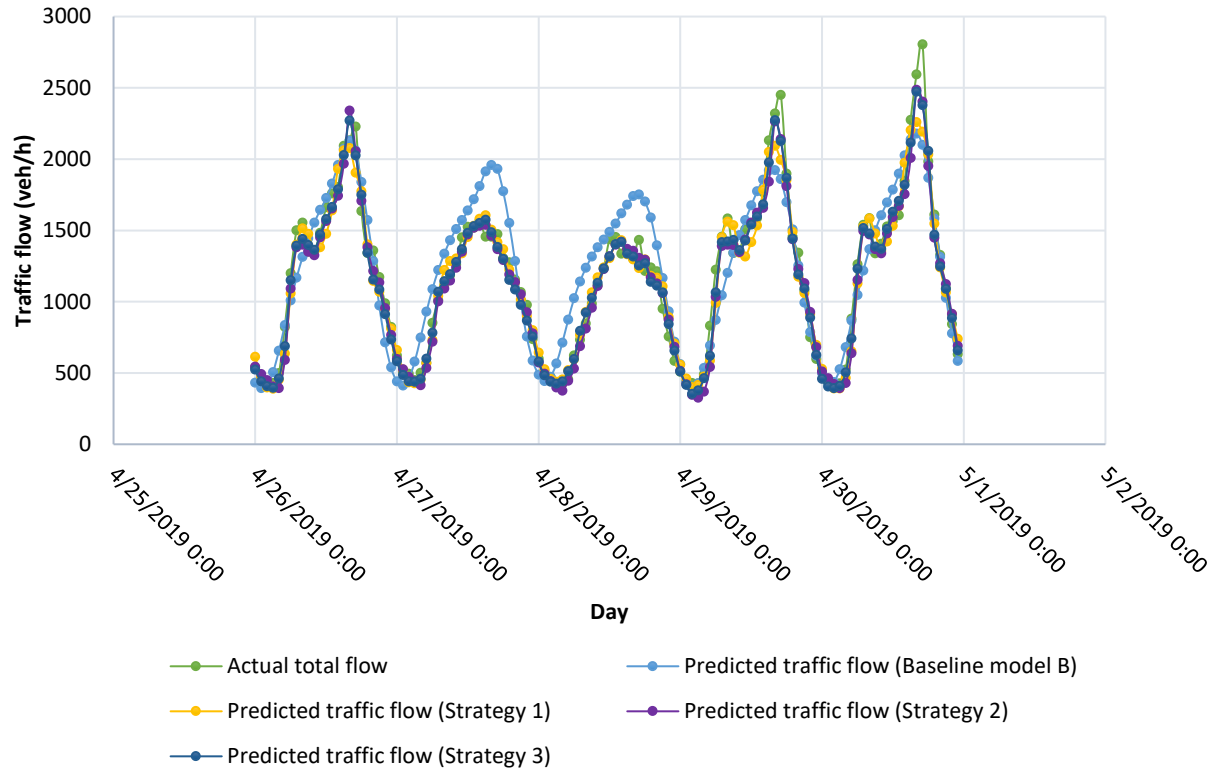


Figure 5.26: Scenario 4 prediction plots for 3 lane to 2 lane transfer learning

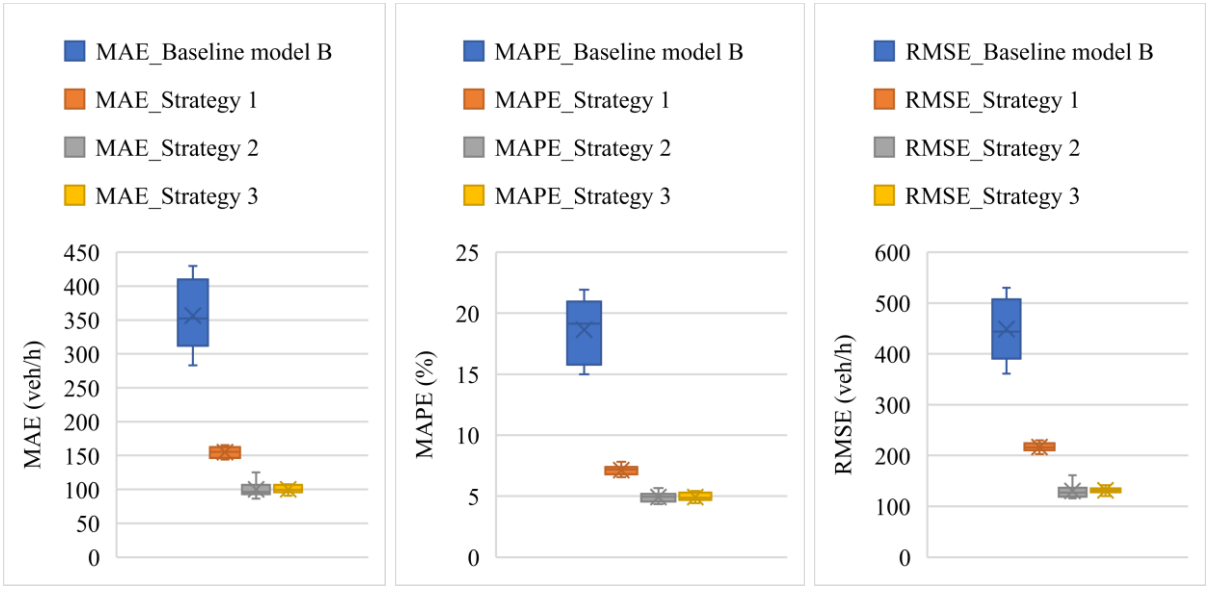
Figure 5.26 illustrates the traffic flow forecasting graphs for the target baseline model B and scenario-4 transfer-learned forecasting. Here, the 3-lane road (station id 314042) dataset is taken as the source model dataset, and the 2-lane road (station id 601722) dataset is taken as the target dataset. The light-blue line indicates the forecasted plot of target baseline model B, which did not provide a better forecast. The yellow-colored line shows forecasted traffic flow by the strategy-1 transfer learning (fine-tuned) at some point does not give good fits to actual traffic flow. The purple line represents the predicted traffic flow using strategy-2 transfer learning (fine-tuned), while the deep blue line represents the forecast using strategy 3 (without fine-tuned). Overall, both strategies 2 and 3 give better fits. Here also, transfer-learned traffic forecasting is better than baseline model forecasting.

5.3.5 Scenario 5 Transfer Learning (LSTM Model)

Table 5.17: MAE, MAPE and RMSE for scenario 5 transfer learning

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	328.20	151.58	125.17	98.63	17.53	6.85	5.67	4.75	425.99	212.05	161.03	132.02
2	409.24	143.93	86.26	98.69	21.92	6.67	4.34	4.77	520.76	205.51	115.84	131.87
3	306.63	165.28	100.58	95.76	15.24	7.83	5.15	4.69	391.12	229.95	128.85	131.23
4	283.17	157.62	88.55	95.72	14.98	7.38	4.35	4.69	360.83	218.16	117.13	131.18
5	313.41	162.15	96.48	107.63	15.95	7.25	4.77	5.34	390.38	222.81	126.90	141.93
6	378.13	146.73	104.12	106.55	19.17	6.57	5.10	5.29	477.03	213.53	133.94	138.06
7	429.64	145.18	95.25	100.17	21.74	6.86	4.64	5.00	529.95	202.79	127.45	127.45
8	336.46	163.62	95.97	100.6	20.11	7.49	5.05	5.04	433.34	228.48	124.64	127.98
9	412.44	162.09	94.60	90.85	20.70	7.27	4.86	4.44	502.82	223.14	119.99	120.45
10	367.88	153.66	115.03	107.9	19.13	7.14	5.37	5.43	454.50	213.39	145.79	134.68

Scenario 5 accuracy plots for 4 lanes to 3 lane road:



(a) MAE (b) MAPE (c) RMSE

Figure 5.27: MAE, MAPE and RMSE for 4 lane to 3 lane road transfer learning

Figure 5.27 represents the accuracy metrics of scenario-5 transfer learning three strategies and target baseline model B's forecasting performances. Here, the 4-lane road (319351) from district 3 is used as

the source dataset, and another 3-lane road (602538) from district 6 is used as the target dataset. According to MAE, MAPE, and RMSE, target baseline model B does not give good predictions. Overall, strategies 2 and 3 transfer-learned forecasting give better accuracy compared to strategy 1. Moreover, transfer-learned models give better forecasting than baseline model forecasting with a limited dataset.

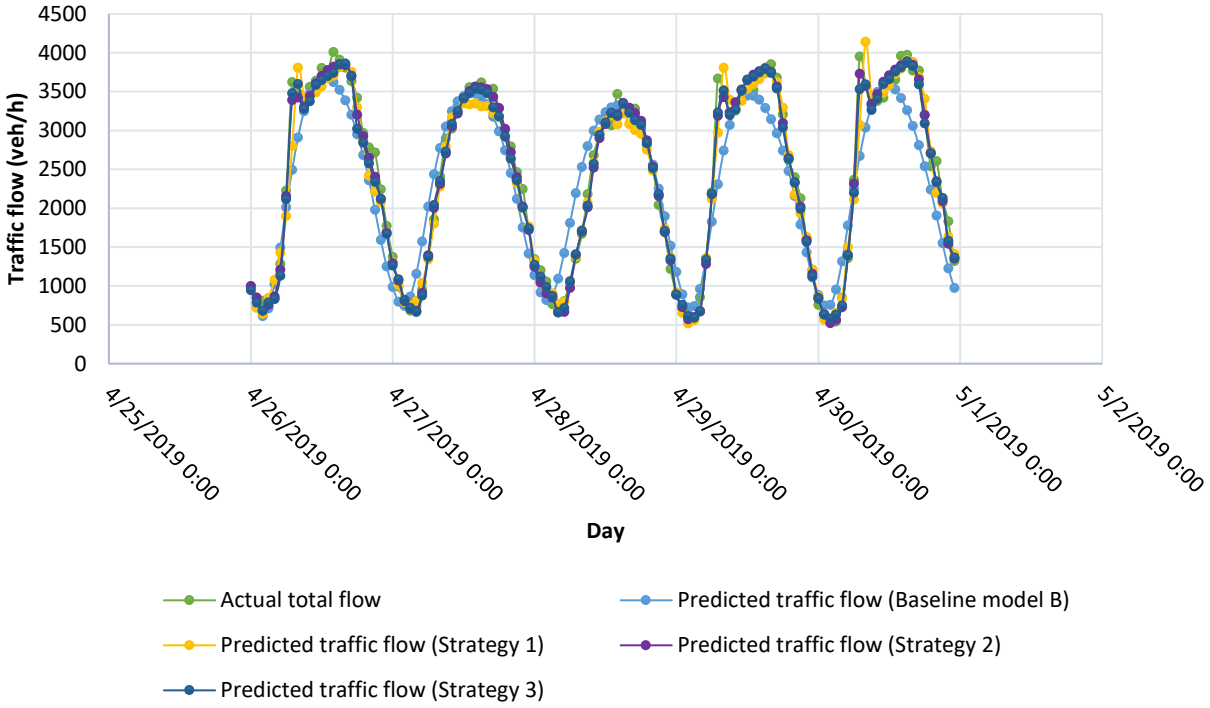


Figure 5.28: Scenario 5 prediction plots for 4 lane to 3 lane transfer learning

Figure 5.28 represents the forecasting plots of target baseline model B, and scenario-5 transfer learned forecasting graphs. Here, the 4-lane road (station id 319351) dataset from district 3 is taken as the source dataset, and the 3-lane road (station id 602538) dataset from district 6 is taken as the target dataset. The light-blue line indicates the forecasted plot of target baseline model B, which did not provide a better forecast. The yellow-colored line shows forecasted traffic flow by strategy-1 transfer learning (fine-tuned) at some point gives under or overfits to actual traffic flow. The purple line represents the predicted traffic flow using strategy-2 transfer learning (fine-tuned), while the deep blue line represents the forecast using strategy 3 (without fine-tuned). Overall, both strategies 2 and 3 give better fits with actual traffic flow. Moreover, in this scenario, transfer-learned traffic forecasting performances are better than target baseline model B forecasting.

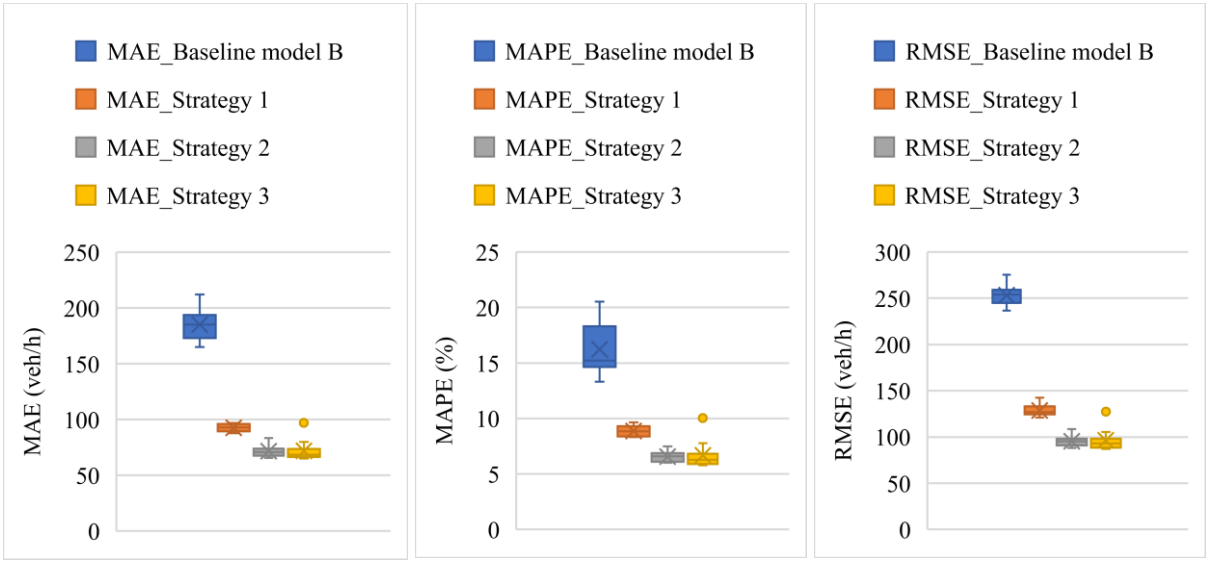
5.3.6 Scenario 6 Transfer Learning (LSTM Model)

Table 5.18: MAE, MAPE and RMSE for scenario 6 transfer learning

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	188.83	94.08	69.38	68.36	17.79	9.65	6.65	5.97	236.57	126.86	91.22	92.33
2	212.20	91.56	71.27	67.3	20.53	8.87	6.12	5.92	256.61	121.11	98.47	90.91
3	177.59	90.04	74.40	67.05	15.31	8.62	6.82	6.30	262.96	125.70	95.93	88.73
4	174.66	95.75	73.45	97.02	15.07	9.05	7.01	10.04	243.89	131.86	97.02	127.44
5	207.88	91.97	70.14	71.45	19.89	8.79	6.52	6.49	254.80	125.17	89.89	95.60
6	165.02	94.05	66.16	71.26	13.32	9.30	6.32	6.48	245.42	128.14	88.04	95.30

7	185.54	96.63	65.63	68.37	15.13	8.38	6.01	6.24	275.48	142.53	92.35	93.40
8	167.85	96.97	83.48	64.98	13.82	9.32	7.47	5.78	246.15	137.33	108.66	87.27
9	188.09	87.64	68.00	64.90	16.62	8.38	6.06	5.77	253.11	121.64	95.02	87.14
10	184.61	88.18	73.91	80.02	14.93	8.31	6.67	7.78	258.02	127.14	98.04	105.51

Scenario 6 accuracy plots for 4 lanes to 2 lanes road:



(a) MAE (b) MAPE (c) RMSE

Figure 5.29: Scenario 6 MAE, MAPE and RMSE

Figure 5.29 represents the accuracy metrics of scenario-6 transfer learning three strategies and target baseline model B's forecasting performances. Here, the 4-lane road (319351) from district 3 is used as the source dataset, and the 2-lane road (601722) from district 6 is used as the target dataset. According to MAE, MAPE, and RMSE, target baseline model B does not give good predictions. Transfer-learned forecasting performances are better than target baseline model performance with limited datasets. Overall, strategies 2 and 3 transfer learning give better accuracy compared to strategy-1 transfer-learned forecasting.

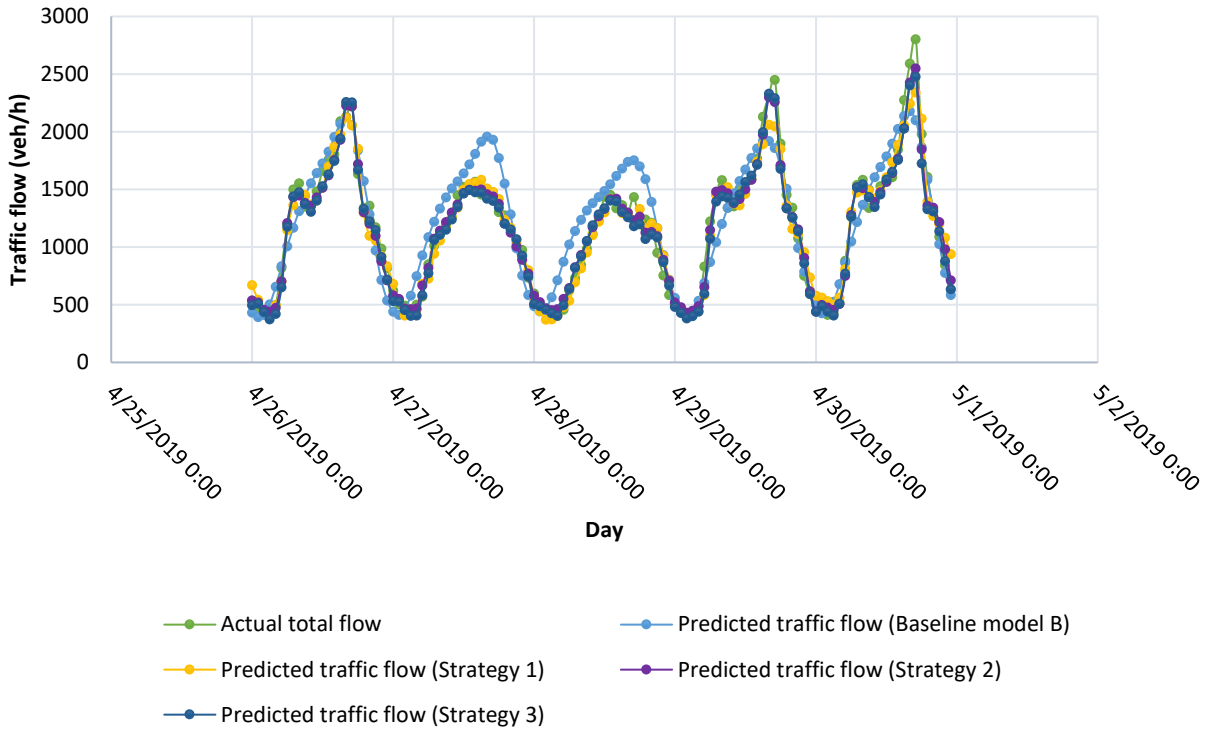


Figure 5.30: Scenario 6 prediction plots for 4 lane to 2 lane transfer learning

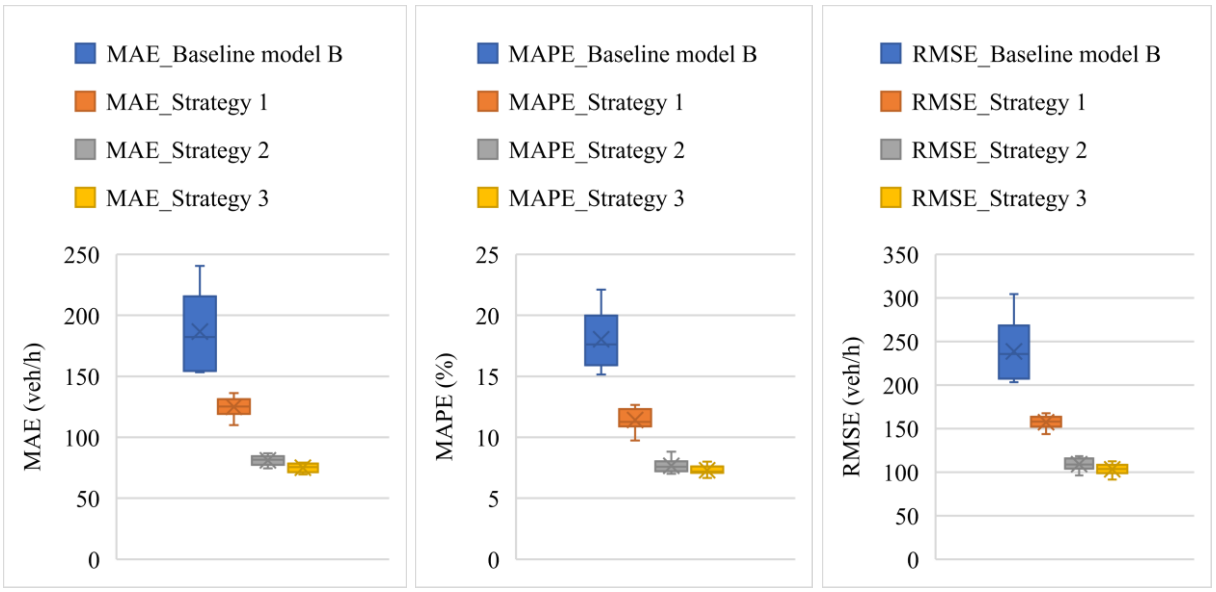
Figure 5.30 illustrates the forecasting graphs of target baseline model B and scenario-6 transfer-learned forecasting. Here, the 2-lane road (station id 601722) from district 6 is taken as the target dataset, and the 4-lane road (station id 319351) from district 3 is used as the source dataset. The light-blue line indicates the forecasted plot of target baseline model B, which did not provide better forecasting. The yellow-colored line shows forecasted traffic flow by strategy-1 transfer learning (fine-tuned) at some point gives under or overfits to actual traffic flow. The purple line represents the predicted traffic flow using strategy-2 transfer learning (fine-tuned), while the deep blue line represents the forecasting using strategy-3 (without fine-tuned). Overall, the both strategies 2 and 3 give a better fit with actual traffic flow than strategy-1 transfer-learned forecasting.

5.4 Transfer learning with September target dataset

5.4.1 Target Dataset 2 Lane

Table 5.19: Traffic forecasting accuracy with target dataset of September (2 lane)

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	168.67	127.82	74.63	70.11	16.04	11.54	7.31	7.19	236.45	159.48	96.25	91.41
2	240.67	109.93	82.43	79.35	19.71	9.73	7.24	7.31	304.52	143.61	118.02	112.29
3	197.54	120.02	78.21	74.92	22.11	10.94	7.01	7.01	241.65	153.40	115.03	110.97
4	214.55	122.17	77.64	78.45	18.58	10.98	7.13	7.10	272.81	154.16	105.06	106.97
5	154.57	131.92	84.60	77.28	15.75	12.34	7.74	7.63	203.15	160.69	113.69	102.98
6	168.31	131.19	84.68	75.36	15.98	12.66	8.81	7.62	211.88	162.65	110.26	103.88
7	195.89	116.52	80.94	71.89	20.81	10.77	7.49	7.14	234.87	147.78	106.94	100.01
8	153.21	136.20	76.46	69.43	15.15	12.30	7.66	6.66	208.66	166.90	100.24	94.83
9	218.83	122.64	84.57	78.81	19.70	10.97	8.44	8.00	266.76	156.92	106.77	100.26
10	154.02	130.63	86.84	75.58	16.67	11.84	7.88	7.21	203.39	167.56	118.22	107.75



(a) MAE (b) MAPE (c) RMSE

Figure 5.31: MAE, MAPE and RMSE with target dataset for September (2 lane)

Figure 5.31 illustrates forecasting accuracy by target baseline model B and transfer learning by MAE, MAPE, and RMSE. Here, the source model is used with the 2-lane road (313190) dataset from District 3, and the target dataset is taken as the 2-lane road (601722) dataset of September 2019 from District 6. According to the mean value of MAE, MAPE, and RMSE, the transfer-learned model's performance is better than the target baseline model B. Overall, strategy-3 transfer learned forecasting performance is better than the other two strategies. The forecasting performance of strategies 2 and 3 are close to each other.

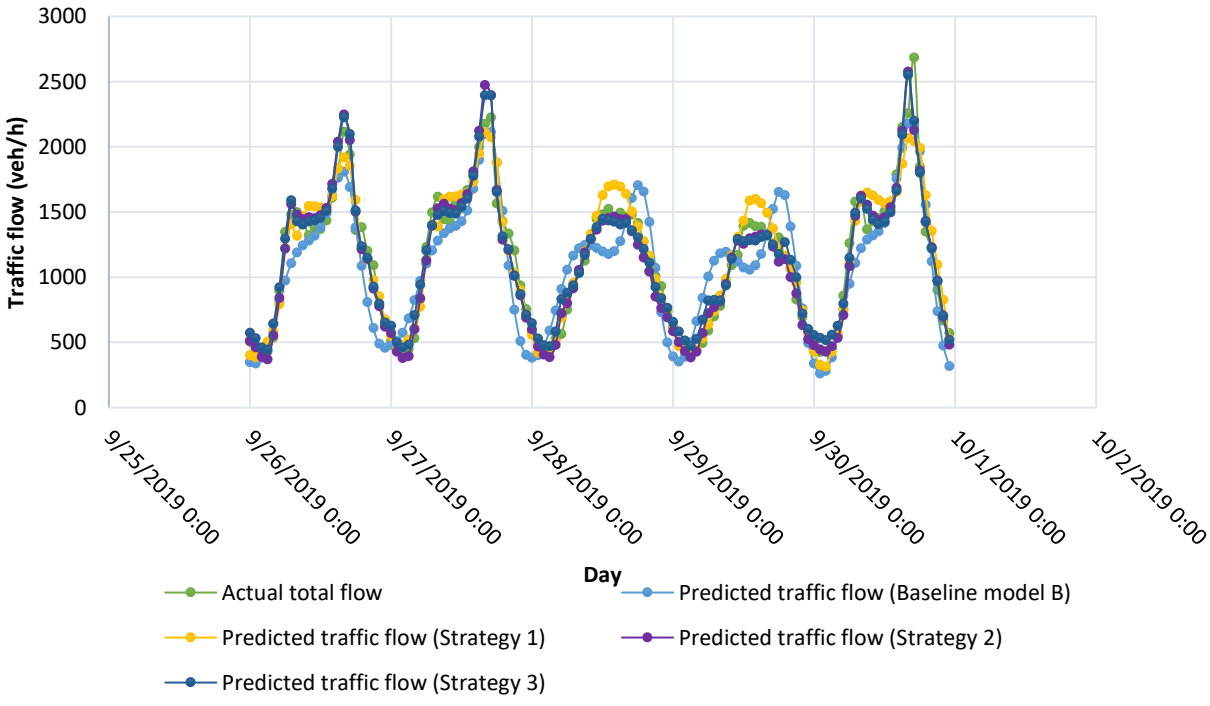


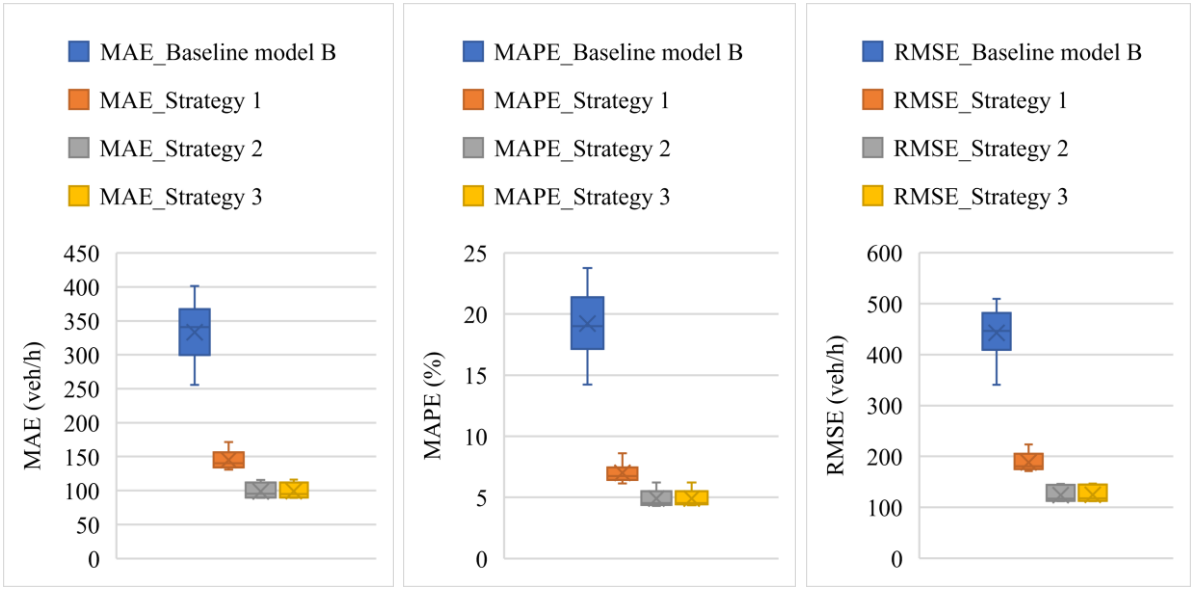
Figure 5.32: Prediction plot for 2 lane road transfer learning with September data

For checking the transfer learning model performance with another month's dataset, we took the September 2019 dataset instead of the April 2019 dataset as a target dataset. After that, we got forecasting plots for the target baseline model B and 2-lane road (601722) traffic flow for three different transfer learning strategies. The source dataset is also the 2-lane road from district 3. Here, in figure 5.32, the green line is the actual traffic flow of the 2-lane road. The light blue line is the forecasted traffic flow of target baseline model B. The yellow line represents the strategy-1 (fine-tuned) transfer learned traffic forecasting. The purple line is the forecasted traffic flow of strategy 2 (fine-tuned) transfer learning, and the deep blue line is the forecasted traffic flow of strategy 3 (without fine-tuned) transfer learning. Overall, strategies 2 and 3 give better fits for traffic forecasting than strategy-1 transfer learning and target baseline model B.

5.4.2 Target Dataset 3 Lane

Table 5.20: Traffic forecasting accuracy with target dataset of September (3 lane)

Num. of runs	MAE (vehicles/h)				MAPE (%)				RMSE (vehicles/h)			
	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3	Base line model B	Strat. 1	Strat. 2	Strat. 3
1	305.82	144.80	95.65	95.48	17.90	6.86	4.96	4.91	411.21	191.93	117.59	118.60
2	280.41	154.14	113.24	114.30	16.39	7.22	5.83	5.84	403.81	206.55	144.22	144.99
3	313.77	140.52	89.34	89.71	20.09	6.81	4.35	4.39	445.93	180.99	112.22	112.33
4	359.64	171.44	115.93	116.35	23.77	8.61	6.22	6.21	488.10	223.37	145.96	146.72
5	330.68	134.24	95.84	97.11	17.92	6.13	4.55	4.54	425.57	176.86	119.02	120.93
6	368.12	162.85	111.19	111.44	20.55	8.20	5.40	5.40	474.75	204.87	145.00	145.29
7	255.82	136.94	89.75	89.80	14.22	6.46	4.54	4.51	340.75	172.93	112.01	112.16
8	401.38	130.62	89.53	91.42	23.09	6.68	4.31	4.46	509.30	170.92	113.21	116.17
9	350.99	134.05	95.02	94.60	17.41	6.38	4.44	4.44	446.96	175.37	117.16	116.44
10	367.29	140.41	90.50	89.16	20.81	6.63	4.38	4.35	479.26	180.66	114.13	112.63



(a) MAE (b) MAPE (c) RMSE

Figure 5.33: MAE, MAPE and RMSE with target dataset for September (3 lane)

Figure 5.33 represents the accuracy boxplots of 3 lane road (602538) with the September 2019 target dataset. The source model's dataset here is taken 3-lane road (314042) from district 3. According to the

mean value of MAE, MAPE, and RMSE, the transfer-learned model's performance is better than the target baseline model B. Strategies 2 and 3 give better forecasting performance than strategy-1 transfer-learned forecasting. Here, the transfer learning with the September 2019 dataset also performs well.

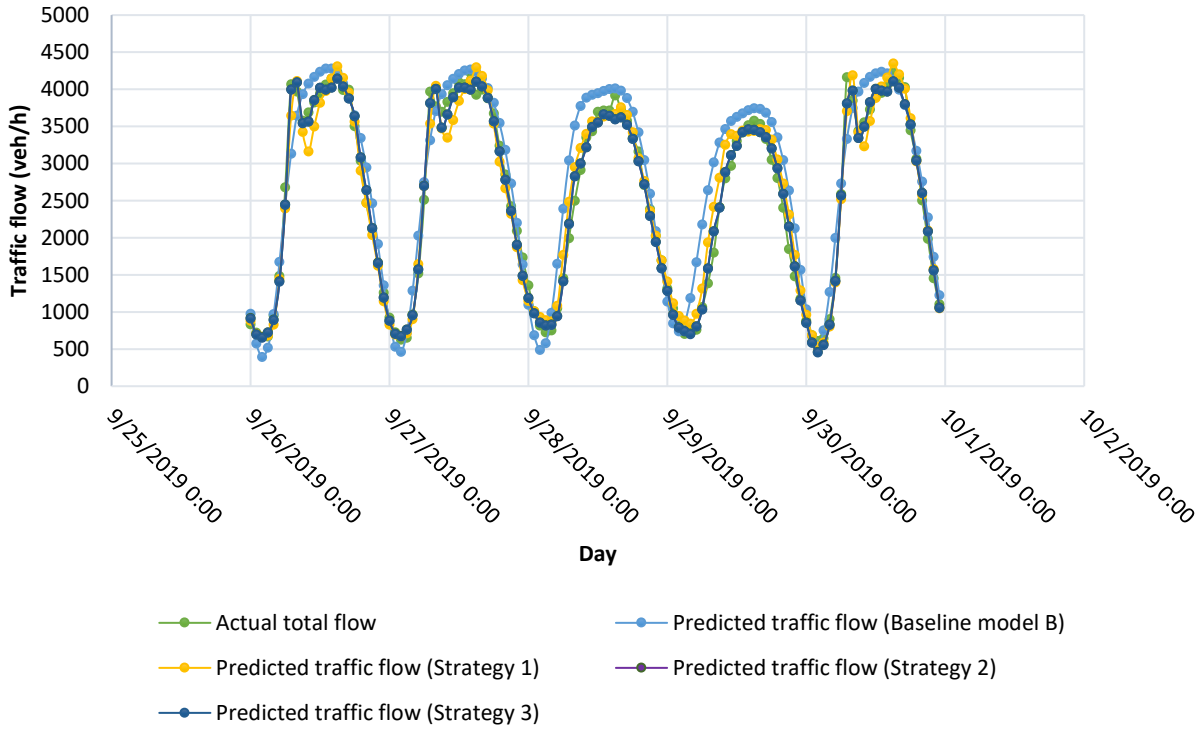


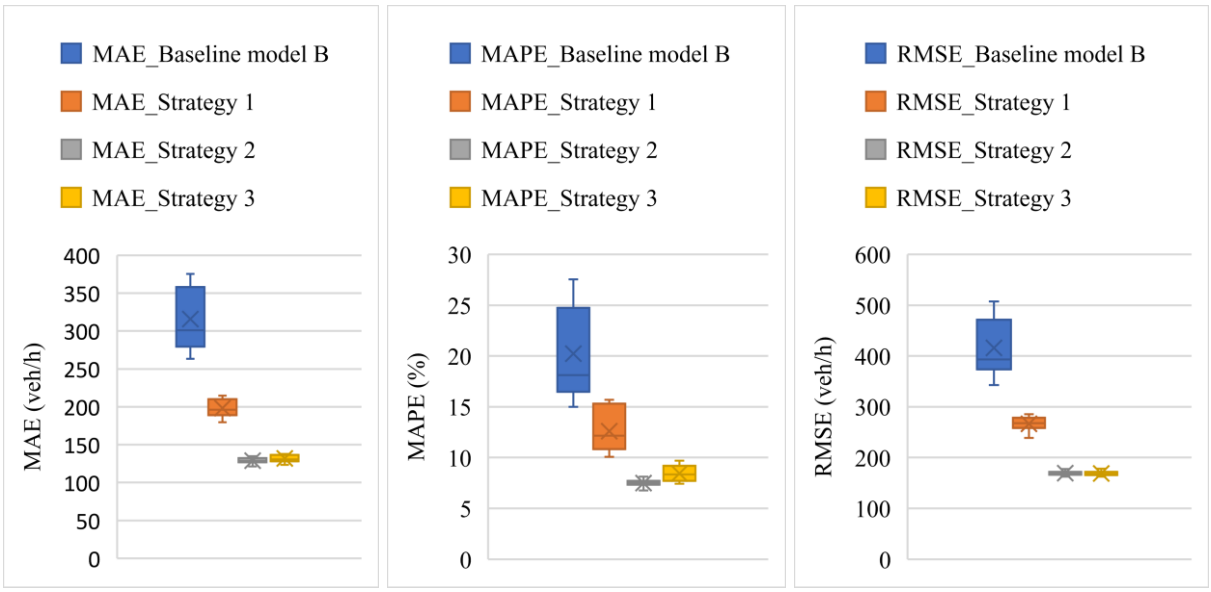
Figure 5.34: Prediction plot for 3 lane road transfer learning with September data

Figure 5.34 shows the forecasting plots of target baseline model B and transfer-learning with three strategies forecasting plots. Here, the green line represents the real-time traffic flow. The light-blue line illustrates the predicted traffic flow of baseline target baseline model B. The yellow line represents strategy-1 traffic flow forecasting, whereas the purple line presents strategy-2 traffic flow forecasting, and the deep-blue line represents the predicted traffic flow for strategy-3 transfer learning. All strategies of transfer learning perform better than target baseline model B. Strategies 2 and 3 give better fits to actual traffic flow.

5.4.3 Target Dataset 4 Lane

Table 5.21: Traffic forecasting accuracy with target dataset of September (4 lane)

Number of runs	**MAE (vehicles/h)**				**MAPE (%)**				**RMSE (vehicles/h)**			
	Baseline model B	Strat. 1	Strat. 2	Strat. 3	Baseline model B	Strat. 1	Strat. 2	Strat. 3	Baseline model B	Strat. 1	Strat. 2	Strat. 3
1	283.10	179.51	127.64	128.04	18.03	10.08	7.65	7.90	360.50	238.72	167.14	165.75
2	268.11	195.46	133.91	137.15	17.68	12.38	7.77	8.70	381.57	261.54	177.54	178.06
3	375.32	189.77	129.38	128.98	24.08	10.07	7.45	7.75	507.20	272.99	169.70	167.83
4	368.51	209.68	129.35	138.02	27.54	15.29	7.50	9.71	486.44	269.47	171.47	174.92
5	302.13	196.55	121.19	123.45	16.64	11.71	6.77	7.42	397.71	265.73	162.25	162.79
6	292.01	211.14	134.80	135.71	18.22	15.35	8.11	9.16	388.34	278.42	174.46	171.15
7	263.45	186.38	131.85	136.62	14.99	12.20	7.39	9.34	342.89	253.73	170.91	169.79
8	299.72	209.46	126.07	127.91	15.94	12.09	7.15	7.65	378.28	285.59	167.10	166.94
9	348.97	190.51	127.89	129.16	22.42	11.09	7.69	7.99	449.60	259.49	166.05	165.59
10	354.67	214.91	127.33	131.96	26.70	15.69	7.54	8.79	466.35	279.00	165.83	167.46



(a) MAE (b) MAPE (c) RMSE

Figure 5.35: MAE, MAPE and RMSE with target dataset for September (4 lane)

Figure 5.35 shows the performance of traffic flow forecasting for the target baseline model B and transfer learning three strategies by boxplots. Here September 2019 target dataset is used, and the 3-lane road dataset is taken for the source model. In terms of MAE, MAPE, and RMSE, transfer-learned model performances are overall better than target base-model B's forecasting. Both strategies 2 and 3 perform better than strategy-1 forecasting.

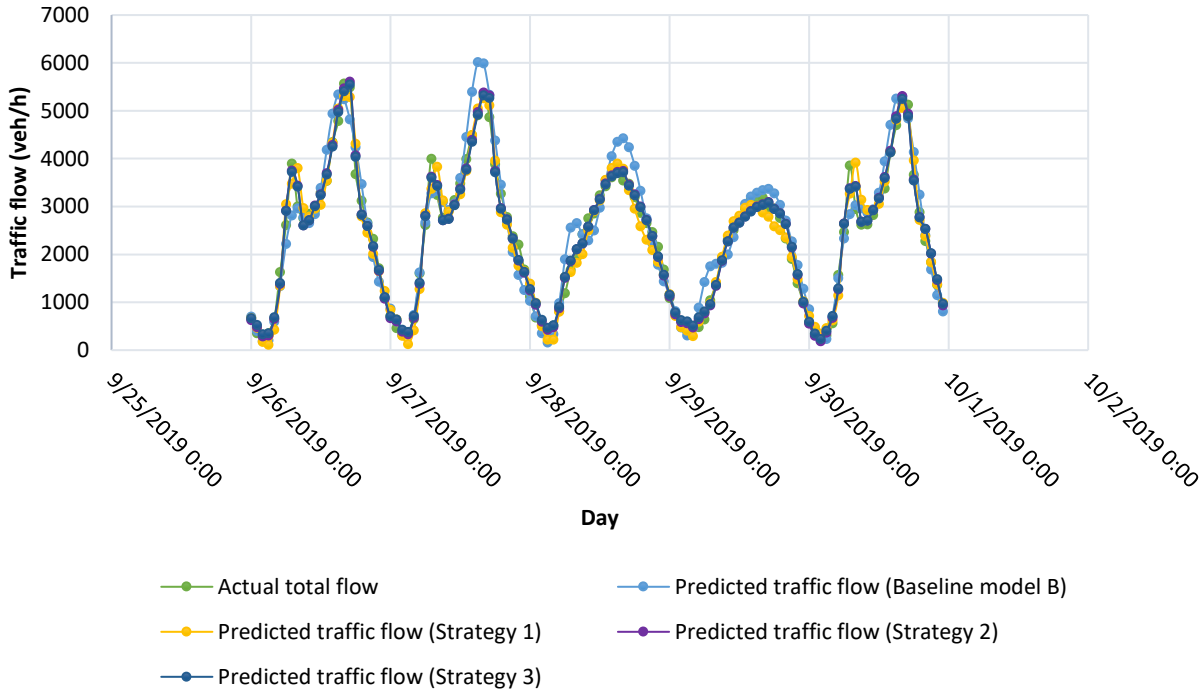


Figure 5.36: Prediction plot for 4 lane road transfer learning with September data

Figure 5.36 represents the forecasting plots of target baseline model B and transfer-learning with three strategies forecasting plots. Here, the green line represents the real-time traffic flow. The light-blue line illustrates the predicted traffic flow of target baseline model B. The yellow line represents strategy-1 traffic flow forecasting, whereas the purple line presents strategy-2 traffic flow forecasting, and the deep-blue line represents the predicted traffic flow for strategy-3 transfer learning. All strategies of transfer learning perform better than target baseline model B. Strategies 2 and 3 give better fits than strategy 1 to the actual traffic flow.

5.5 Summary of Results

Several experiments are performed to verify the accuracy of the dataset for research question 1, with lengths ranging from 1 month to 6 months from District 3, California State, USA. We used three different algorithms (univariate ARIMA, GRU, and LSTM). Overall, in 6 months of the dataset, both GRU and LSTM models give better accuracy of traffic forecasting in this experiment. Univariate ARIMA initially, for lower-length datasets, gives better accuracy, but for adding more data, it does not give a good prediction of traffic flow. We selected six months of data for further analysis.

Weather datasets from relevant counties are selected for checking their influence on traffic flow forecasting for research question 2, we integrate weather data with traffic data in two multivariate models (GRU and LSTM). Overall, with-weather datasets give better accuracy than without-weather traffic datasets. In terms of accuracy, the LSTM model gives better prediction than GRU. For the next step, we select the LSTM model with weather traffic datasets for transfer learning.

Next, six different scenarios with six different types of roads are applied in transfer learning. Datasets for three different road types are selected from District 3 for the source dataset, spanning six months (April 2019 to September 2019). 1-month of data (April 2019) and 1-month of data (September 2019) are selected from District 6 for target datasets. Overall, transfer-learned traffic flow forecasting is better than target baseline B forecasting in terms of MAE, MAPE, and RMSE. Overall, Strategy 2 and 3 transfer learning give better accuracy than strategy-1 transfer learning for traffic forecasting. We get the answer to research question 1 by experimenting with a 1-month source dataset from District 3 in scenarios 1, 2, and 3. It shows that more source traffic dataset transfer learning models give better results than 1-month source dataset models. For checking the transfer learning performance with five months ahead target datasets, we select September 2019 datasets. In this case, we got the same results of forecasting like previous experiments.

6 Conclusion

6.1 Summary

Traffic congestion is a crucial problem these days in all big cities. This is the most alarming problem for most city roads and highways to cope up. People are losing lots of time and money due to this growing problem. This research aimed to get a solution to the problem of congestions in ITS by traffic forecasting for better traffic management. To these congestion problems, traffic forecasting is a great strategy. Data scarcity is another problem in traffic forecasting, where a limited dataset cannot give a good prediction. To improve traffic forecasting accuracy with limited datasets, the transfer learning strategy can help well in these situations. To experiment with these issues and to get the solutions to these problems, this study was done for California state's two different districts.

In response to research gaps in different dimensions of this research field, two research questions were framed, and an effort was made to answer these two research questions through several experiments. The first research question is about the best amount of input data to use in traffic flow forecasting models, which can be utilized as a source dataset in transfer learning models for the selected roads. We did several experiments on univariate ARIMA, GRU, and LSTM models with 1-month, 2-month, 3-month, 4-month, 5-month, and 6-month of datasets for district 3. Overall, 6-months dataset for GRU and LSTM models gave well forecasting accuracy. Although univariate ARIMA gave a better prediction for smaller range datasets it did not perform well for higher ranges datasets. We selected 6 months of datasets for 2-lane, 3-lane, and 4-lane roads from different Counties of district 3 of California state. Also, in comparison to algorithm performances, the LSTM model performed significantly better than the other two algorithms. We also did three scenarios with 1-month source data and 6-month source data. Overall, in terms of MAE, MAPE, and RMSE, 6-month source dataset transfer learning models give better accuracy than 1-month source data models. In answer to research question 2, that is how much influence is happened when we use weather as an external factor in traffic forecasting. In this concern, the additional weather datasets for the relevant locations were collected from the NCEI weather database. Models' performances were checked by combining the weather data and without weather data in both GRU and LSTM multivariate models. In all cases, with-weather models performed well for both GRU and LSTM models. Overall, LSTM models performed better than the GRU model in terms of accuracy metrics MAE, MAPE, and RMSE.

Finally, six different scenarios were made for transfer learning with several source datasets from District 3 and target datasets from District 6 in the stacked LSTM model. In all the scenarios, April 2019 to September 2019 (6 months of datasets) were selected for all 2-lane, 3-lane, and 4-lane roads. April 2019 (1-month) datasets were selected as target datasets for all 2-lane, 3-lane, and 4-lane roads. Three different strategies for the transfer learning model were formed, where strategies 1 and 2 are fine-tuned models, and strategy 3 is not the fine-tuned model. Overall, strategy 2 and 3 transfer-learned models performed better than the strategy-1 transfer-learned model in almost all scenarios. Target baseline model, B forecasting was done with April 2019 dataset from district 6 to check the performances of target models. In these cases, all 2-lane, 3-lane, and 4-lane roads did not give well forecasting in terms of performance. Moreover, transfer learned models gave better performance than the target baseline model B. Limited datasets cannot give better forecasting independently. For data scarcity problems in traffic forecasting, transfer learning strategies give better solutions.

Last but not the least, we checked transfer learning performances with 5 months ahead target dataset. In this case, we used September 2019 dataset from district 6 for every link type transfer learning with the same source datasets from District 3. In all 3-road type forecasting of traffic in transfer learning, we used the same 3-different strategies, where strategies 2 and 3 gave better performances than fine-tuned model strategy-1 in terms of MAE, MAPE, and RMSE.

The contribution of this research on the California case study with PeMS enriched traffic datasets helps a lot for experimenting with different scenarios during the pre-Covid pandemic time. A novel deep transfer learning technique is developed for traffic forecasting in California state. The weather dataset gives information on how much influence happens in traffic forecasting. Components of different transfer learning strategies are described and analyzed. These hybrid transfer learned models give a smart solution for traffic forecasting in ITS with a limited amount of traffic data.

This research can help a lot for other city traffic forecasting with transfer learning by pre-trained models. Traffic forecasting with transfer learning will play a vital role in proper traffic management, route guidance, smart intersection techniques, etc. For reducing traffic congestion on roadways and highways in big cities, this technology will help a lot in ITS.

6.2 Limitations and Future Works

There are some limitations found during model building with the ARIMA model. When tried with exogenous variables in the multivariate ARIMA model, it did not perform well for one step ahead of traffic forecasting. We used a univariate ARIMA model for comparison with the other two neural network forecasting. The inconsistency is another limitation. A consistent dataset gives better traffic forecasting. In this research, we used link-type forecasting. In the future, more source links should be taken considered, and a larger network-wide traffic forecasting should be experimented with. Future research in this field should take consider other factors like incidents and other spatial characteristics like stadiums, supermarkets, educational institutes, etc.

References

Abadi, A., Rajabioun, T., & Ioannou, P. A. (2014). Traffic flow prediction for road transportation networks with limited traffic data. *IEEE Transactions on Intelligent Transportation Systems*, 1–10. https://doi.org/10.1109/TITS.2014.2337238

Abduljabbar, R. L., Dia, H., & Tsai, P.-W. (2021). Development and evaluation of bidirectional LSTM freeway traffic forecasting models using simulation data. *Scientific Reports*, *11*(1), 23899. https://doi.org/10.1038/s41598-021-03282-z

Agalliadis, I., Makridis, M., & Kouvelas, A. (2020). Traffic estimation by fusing static and moving observations in highway networks. *20th Swiss Transport Research Conference*, 1–15. https://doi.org/10.3929/ethz-b-000419736

Alghamdi, T., Mostafi, S., Abdelkader, G., & Elgazzar, K. (2022). A comparative study on traffic modeling techniques for predicting and simulating traffic behavior. *Future Internet*, *14*(10), 294. https://doi.org/10.3390/fi14100294

Alipour, B., Tonetto, L., Ketabi, R., Yi Ding, A., Ott, J., & Helmy, A. (2019). Where are you going next?: A practical multi-dimensional look at mobility prediction. *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems - MSWIM '19*, 5–12. https://doi.org/10.1145/3345768.3355923

ArunKumar, K. E., Kalaga, D. v., Mohan Sai Kumar, Ch., Kawaji, M., & Brenza, T. M. (2022). Comparative analysis of Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) cells, autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average (SARIMA) for forecasting COVID-19 trends. *Alexandria Engineering Journal*, *61*(10), 7585–7603. https://doi.org/10.1016/j.aej.2022.01.011

Attri, P., Sharma, Y., Takach, K., & Shah, F. (2020, June 23). *Keras_Timeseries forecasting for weather prediction*. https://keras.io/examples/timeseries/timeseries_weather_forecasting/.

Bekhor, S., & Prato, C. G. (2009). Methodological transferability in route choice modeling. *Transportation Research Part B: Methodological*, *43*(4), 422–437. https://doi.org/10.1016/j.trb.2008.08.003

ben Taieb, S., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, *39*(8), 7067–7083. https://doi.org/10.1016/j.eswa.2012.01.039

Bhavsar, P., Chowdhury, M., Sadek, A., Sarasua, W., & Ogle, J. (2007). Decision support system for predicting traffic diversion impact across transportation networks using Support Vector Regression. *Transportation Research Record: Journal of the Transportation Research Board*, *2024*(1), 100–106. https://doi.org/10.3141/2024-12

Bontempi, G. (2008). Long term time series prediction with multi-input multi-output local learning. *Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP08*, 145–154. https://www.academia.edu/40201923/Long_term_time_series_prediction_with_multi_input_multi_output_local_learning

Boukerche, A., & Wang, J. (2020). Machine Learning-based traffic prediction models for Intelligent Transportation Systems. *Computer Networks*, *181*, 107530. https://doi.org/10.1016/j.comnet.2020.107530

Bouyahia, Z., Haddad, H., Derrode, S., & Pieczynski, W. (2021). Toward a cost-effective motorway traffic state estimation from sparse speed and GPS data. *IEEE Access*, *9*, 44631–44646. https://doi.org/10.1109/ACCESS.2021.3066422

Boyce, D. (2007). Forecasting travel on congested urban transportation networks: Review and prospects for network equilibrium models. *Networks and Spatial Economics*, *7*(2), 99–128. https://doi.org/10.1007/s11067-006-9009-0

Boyce, D. E., & Williams, H. C. W. L. (2005). Urban travel forecasting in the USA and UK. In *Methods and Models in Transport and Telecommunications* (pp. 25–44). Springer-Verlag. https://doi.org/10.1007/3-540-28550-4_3

Braz, F. J., Ferreira, J., Gonçalves, F., Weege, K., Almeida, J., Baldo, F., & Gonçalves, P. (2022). Road traffic forecast based on meteorological information through deep learning methods. *Sensors*, *22*(12), 4485. https://doi.org/10.3390/s22124485

Buroni, G. (2021). *On-board-unit big data analytics: From data architecture to traffic forecasting* [PhD]. Universite Libre De Bruxelles.

Chen, J., & Wang, X. (2010). *Multi-innovation generalized extended stochastic gradient algorithm for multi-input multi-output nonlinear Box-Jenkins systems based on the auxiliary model* (pp. 136–146). https://doi.org/10.1007/978-3-642-15621-2_17

Cheng, T., Haworth, J., & Wang, J. (2012). Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems*, *14*(4), 389–413. https://doi.org/10.1007/s10109-011-0149-5

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). *On the properties of neural machine translation: Encoder-decoder approaches*.

Chollet, F. (2020, April 15). *Keras_Transfer learning & fine-tuning*. https://keras.io/guides/transfer_learning/.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). *Empirical evaluation of Gated Recurrent Neural Networks on sequence modeling*.

Coifman, B. (2003). Estimating density and lane inflow on a freeway segment. *Transportation Research Part A: Policy and Practice*, *37*(8), 689–701. https://doi.org/10.1016/S0965-8564(03)00025-9

Cools, M., Moons, E., & Wets, G. (2010). Assessing the impact of weather on traffic intensity. *Weather, Climate, and Society 2*, *2*(1), 60–68. http://www.jstor.org/stable/24907338

Dai, G., Ma, C., & Xu, X. (2019). Short-term traffic flow prediction method for urban road sections based on space–time analysis and GRU. *IEEE Access*, *7*, 143025–143035. https://doi.org/10.1109/ACCESS.2019.2941280

Das, A., Patra, G. R., & Mohanty, M. N. (2021). *A comparison study of recurrent neural networks in recognition of handwritten Odia numerals* (pp. 251–260). https://doi.org/10.1007/978-981-15-8752-8_26

Dechenaux, E., Mago, S. D., & Razzolini, L. (2014). Traffic congestion: an experimental study of the Downs-Thomson paradox. *Experimental Economics*, *17*(3), 461–487. https://doi.org/10.1007/s10683-013-9378-4

Dombalyan, A., Kocherga, V., Semchugova, E., & Negrov, N. (2017). Traffic forecasting model for a road section. *Transportation Research Procedia*, *20*, 159–165. https://doi.org/10.1016/j.trpro.2017.01.040

Domingues, A. C. S. A., Silva, F. A., & Loureiro, A. A. F. (2019). On the analysis of users' behavior based on mobile phone apps. *Proceedings of the 17th ACM International Symposium on Mobility Management and Wireless Access - MobiWac '19*, 25–32. https://doi.org/10.1145/3345770.3356739

Dougherty, M. (1995). A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies*, *3*(4), 247–260.

Du, S., Li, T., Gong, X., & Horng, S.-J. (2020). A hybrid method for traffic flow forecasting using multimodal deep learning. *International Journal of Computational Intelligence Systems*, *13*(1), 85. https://doi.org/10.2991/ijcis.d.200120.001

Duan, Y., L.V., Y., & Wang, F.-Y. (2016). Travel time prediction with LSTM neural network. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, 1053–1058. https://doi.org/10.1109/ITSC.2016.7795686

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*, 2121–2159.

Gao, Y., Zhou, C., Rong, J., Wang, Y., & Liu, S. (2022). Short-term traffic speed forecasting using a deep learning method based on multitemporal traffic flow volume. *IEEE Access*, *10*, 82384–82395. https://doi.org/10.1109/ACCESS.2022.3195353

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. https://doi.org/10.1162/089976600300015015

Google Maps. (2022). https://www.google.com/maps/@48.1457899,11.5653114,15z

Greenshields, B. D. (1934). *The photographic method of studying traffic behavior*.

Grumert, E. F., & Tapani, A. (2018). Traffic state estimation using connected vehicles and stationary detectors. *Journal of Advanced Transportation*, *2018*, 1–14. https://doi.org/10.1155/2018/4106086

Guo, B., Li, J., Zheng, V. W., Wang, Z., & Yu, Z. (2018). City transfer: Transferring inter- and intra-city knowledge for chain store site recommendation based on multi-source urban data. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *1*(4), 1–23. https://doi.org/10.1145/3161411

Habtemichael, F. G., & Cetin, M. (2016). Short-term traffic flow rate forecasting based on identifying similar traffic patterns. *Transportation Research Part C: Emerging Technologies*, *66*, 61–78. https://doi.org/10.1016/j.trc.2015.08.017

Hamzaçebi, C., Akay, D., & Kutay, F. (2009). Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. *Expert Systems with Applications*, *36*(2), 3839–3844. https://doi.org/10.1016/j.eswa.2008.02.042

Hawes, M., Amer, H. M., & Mihaylova, L. (2016). Traffic state estimation via a particle filter with compressive sensing and historical traffic data. *19th International Conference on Information Fusion (FUSION)*, 735–742.

Herrera, J. C., Work, D. B., Herring, R., Jeff Ban, X., Jacobson, Q., & Bayen, A. M. (2010). Evaluation of traffic data obtained via gps-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C: Emerging Technologies*, *18*(4), 568–583.

Hinsbergen, C. van, Lint, J. van, & Sanders, F. (2007). Short term traffic prediction models. *In Proceedings of the 14th World Congress on Intelligent Transport Systems (ITS), Held Beijing*.

Hong, Z., & Fukuda, D. (2012). Effects of traffic sensor location on traffic state estimation. *Procedia - Social and Behavioral Sciences*, *54*, 1186–1196. https://doi.org/10.1016/j.sbspro.2012.09.833

Hota, H. S., Handa, R., & Shrivas, A. K. (2017). Time series data prediction using sliding window based RBF neural network. In *International Journal of Computational Intelligence Research* (Vol. 13, Issue 5). http://www.ripublication.com

IBM Cloud Education. (2020, August 17). *Neural Networks*. https://www.ibm.com/cloud/learn/neural-networks

Janecek, A., Valerio, D., Hummel, K. A., Ricciato, F., & Hlavacs, H. (2015). The cellular network as a sensor: from mobile phone data to real-time road traffic monitoring. *IEEE Transactions on Intelligent Transportation Systems*, *16*(5), 2551–2572. https://doi.org/10.1109/TITS.2015.2413215

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: examining and forecasting change. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00727

Karpathy, A. (2017). *A peek at trends in machine learning*. https://medium.com/@karpathy/a-peek-at-trends-in-machine-learning-ab8a1085a106 .

Kashyap, A. A., Raviraj, S., Devarakonda, A., Nayak K, S. R., K V, S., & Bhat, S. J. (2022). Traffic flow prediction models – A review of deep learning techniques. *Cogent Engineering*, *9*(1). https://doi.org/10.1080/23311916.2021.2010510

Khan, F. M., & Gupta, R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, *1*(1), 12–18. https://doi.org/10.1016/j.jnlssr.2020.06.007

Khan, S. M., Dey, K. C., & Chowdhury, M. (2017). Real-time traffic state estimation with connected vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *18*(7), 1687–1699. https://doi.org/10.1109/TITS.2017.2658664

Kingma, D. P., & Ba, J. (2014). *Adam: A method for stochastic optimization*.

Kurkjian, A., Gershwin, S. B., Houpt, P. K., Willsky, A. S., Chow, E. Y., & Greene, C. S. (1980). Estimation of roadway traffic density on freeways using presence detector data. *Transportation Science*, *14*(3), 232–261.

Lana, I., Ser, J. del, Velez, M., & Vlahogianni, E. I. (2018). Road traffic forecasting: Recent advances and new challenges. *IEEE Intelligent Transportation Systems Magazine*, *10*(2), 93–109.

Lee, H., Park, C., Jin, S., Chu, H., Choo, J., & Ko, S. (2021). *An empirical experiment on deep learning models for predicting traffic data*. https://doi.org/10.1109/ICDE51399.2021.00160

Li, J. , Guo, F., Sivakumar, A., Dong, Y., & Krishnan, R. (2021). Transferability improvement in short-term traffic prediction using stacked LSTM network. *Transportation Research Part C*, *124*.

Li, Y., & Shahabi, C. (2018). A brief overview of machine learning methods for short-term traffic forecasting and future directions. *SIGSPATIAL Special*, *10*(1), 3–9. https://doi.org/10.1145/3231541.3231544

Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2017). *Diffusion convolutional recurrent neural network: Data-driven traffic forecasting*.

Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, *14*(2), 871–882. https://doi.org/10.1109/TITS.2013.2247040

Luan, J., Guo, F., Polak, J. W., Hoose, N., & Krishnan, R. (2018). Investigating the transferability of machine learning methods in short-term travel time prediction. *Proceedings of the 97th Annual Meeting of Transportation Research Board, Washington DC, USA*.

Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F.-Y. (2014). Traffic flow prediction with big data: a deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, *16*(2), 865–873.

Ma, J., Xu, M., Meng, Q., & Cheng, L. (2020). Ridesharing user equilibrium problem under OD-based surge pricing strategy. *Transportation Research Part B: Methodological*, *134*, 1–24. https://doi.org/10.1016/j.trb.2020.02.001

Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, *54*, 187–197. https://doi.org/10.1016/j.trc.2015.03.014

Ma, Y., Chowdhury, M., Jeihani, M., & Fries, R. (2010). Accelerated incident detection across transportation networks using vehicle kinetics and support vector machine in cooperation with infrastructure agents. *IET Intelligent Transport Systems*, *4*(4), 328–337. https://doi.org/10.1049/iet-its.2010.0035

Maerivoet, S., & de Moor, B. (2005). Traffic flow theory. *Physics*.

Mahajan, V., Cantelmo, G., Rothfeld, R., & Antoniou, C. (2022). Predicting network flows from speeds using open data and transfer learning. *IET Intelligent Transport Systems*. https://doi.org/10.1049/itr2.12305

Mallick, T., Balaprakash, P., Rask, E., & Macfarlane, J. (2020). *Transfer learning with graph neural networks for short-term highway traffic forecasting*. 1–17.

Masum, S., Liu, Y., & Chiverton, J. (2018). *Multi-step time series forecasting of electric load using machine learning models* (pp. 148–159). https://doi.org/10.1007/978-3-319-91253-0_15

Mateus, B. C., Mendes, M., Farinha, J. T., Assis, R., & Cardoso, A. M. (2021). Comparing LSTM and GRU models to predict the condition of a pulp paper press. *Energies*, *14*(21), 6958. https://doi.org/10.3390/en14216958

Monteil, J. G. (2014). *Investigating the effects of cooperative vehicles on highway traffic flow homogenization: analytical and simulation studies*. Université de Lyon.

Nallaperuma, D., Nawaratne, R., Bandaragoda, T., Adikari, A., Nguyen, S., Kempitiya, T., de Silva, D., Alahakoon, D., & Pothuhera, D. (2019). Online incremental machine learning platform for big data-driven smart traffic management. *IEEE Transactions on Intelligent Transportation Systems*, *20*(12), 4679–4690. https://doi.org/10.1109/TITS.2019.2924883

National Centers for Environmental Information. (2022). https://www.ncei.noaa.gov/

Navarro-Espinoza, A., López-Bonilla, O. R., García-Guerrero, E. E., Tlelo-Cuautle, E., López-Mancilla, D., Hernández-Mejía, C., & Inzunza-González, E. (2022). Traffic flow prediction for smart traffic lights using machine learning algorithms. *Technologies*, *10*(1), 5. https://doi.org/10.3390/technologies10010005

Olayode, I. O., Severino, A., Campisi, T., & Tartibu, L. K. (2022). Prediction of vehicular traffic flow using Levenberg-Marquardt artificial neural network model: Italy road transportation system.

Communications - Scientific Letters of the University of Zilina, *24*(2), E74–E86. https://doi.org/10.26552/com.C.2022.2.E74-E86

OpenStreetMap. (2022). https://www.openstreetmap.org/#map=6/51.330/10.453

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191

Pascanu, R., Gulcehre, C., Cho, K., & Bengio, Y. (2013, December 20). How to construct deep recurrent neural networks. *In Proceedings of the Second International Conference on Learning Representations ICLR, Banff, AB, Canada, 14–16 April 2014*.

PeMS. (n.d.). *California Department of Transportation (Caltrans)*. Retrieved August 10, 2022, from https://pems.dot.ca.gov/?dnode=apply

Qu, W., Li, J., Yang, L., Li, D., Liu, S., Zhao, Q., & Qi, Y. (2020). Short-term intersection traffic flow forecasting. *Sustainability*, *12*(19), 8158–8170. https://doi.org/10.3390/su12198158

Rahman, F. I. (2020). Short term traffic flow prediction using machine learning - KNN, SVM and ANN with weather information. *International Journal for Traffic and Transport Engineering, 10*(3), 371–389. https://doi.org/10.7708/ijtte.2020.10(3).08

Rempe, F. (2018). *Traffic speed estimation and prediction using floating car data*.

Ren, C., Chai, C., Yin, C., Ji, H., Cheng, X., Gao, G., & Zhang, H. (2021). Short-term traffic flow prediction: A method of combined deep learnings. *Journal of Advanced Transportation, 2021*, 1–15. https://doi.org/10.1155/2021/9928073

Sagheer, A., Hamdoun, H., & Youness, H. (2021). Deep LSTM-based transfer learning approach for coherent forecasts in hierarchical time series. *Sensors*, *21*(13), 4379. https://doi.org/10.3390/s21134379

Saha, S., Haque, A., & Sidebottom, G. (2022). *Transfer learning based efficient traffic prediction with limited training data*.

Seo, T., Bayen, A. M., Kusakabe, T., & Asakura, Y. (2017). Traffic state estimation on highway: A comprehensive survey. *Annual Reviews in Control*, *43*, 128–151. https://doi.org/10.1016/j.arcontrol.2017.03.005

Shao, H., & Soong, B.-H. (2016). Traffic flow prediction with Long Short-Term Memory networks (LSTMs). *2016 IEEE Region 10 Conference (TENCON)*, 2986–2989. https://doi.org/10.1109/TENCON.2016.7848593

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A comparison of ARIMA and LSTM in forecasting time series. *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

Singh, K., & Li, B. (2012). Estimation of traffic densities for multilane roadways using a Markov Model approach. *IEEE Transactions on Industrial Electronics*, *59*(11), 4369–4376. https://doi.org/10.1109/TIE.2011.2180271

Smith, B. L., Williams, B. M., & Keith Oswald, R. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, *10*(4), 303–321. https://doi.org/10.1016/S0968-090X(02)00009-8

Sorjamaa, A., Hao, J., Reyhani, N., Ji, Y., & Lendasse, A. (2007). Methodology for long-term prediction of time series. *Neurocomputing*, *70*(16–18), 2861–2869. https://doi.org/10.1016/j.neucom.2006.06.015

Srivastava, N., Hinton, G., Krizhevsky, A., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. In *Journal of Machine Learning Research* (Vol. 15).

Sun, J., Wu, J., Xiao, F., Tian, Y., & Xu, X. (2020). Managing bottleneck congestion with incentives. *Transportation Research Part B: Methodological*, *134*, 143–166. https://doi.org/10.1016/j.trb.2020.01.010

Sun, S., Wu, H., & Xiang, L. (2020). City-wide traffic flow forecasting using a deep convolutional neural network. *Sensors*, *20*(2), 421. https://doi.org/10.3390/s20020421

Sun, Z., Hu, Y., Li, W., Feng, S., & Pei, L. (2022). Prediction model for short-term traffic flow based on a K-means-gated recurrent unit combination. *IET Intelligent Transport Systems*, *16*(5), 675–690. https://doi.org/10.1049/itr2.12165

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). *A survey on deep transfer learning*.

Tieleman, T., & Hinton, G. (2012). Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. In *COURSERA: Neural networks for machine learning* (Vol. 4, pp. 26–31).

Tran, Q. H., Fang, Y.-M., Chou, T.-Y., Hoang, T.-V., Wang, C.-T., Vu, V. T., Ho, T. L. H., Le, Q., & Chen, M.-H. (2022). Short-term traffic speed forecasting model for a parallel multi-lane arterial road using GPS-monitored data based on deep learning approach. *Sustainability*, *14*(10), 6351. https://doi.org/10.3390/su14106351

van Erp, P. B. C., Knoop, V. L., & Hoogendoorn, S. P. (2018). Macroscopic traffic state estimation using relative flows from stationary and moving observers. *Transportation Research Part B: Methodological*, *114*, 281–299. https://doi.org/10.1016/j.trb.2018.06.005

van Hinsbergen, C. P. IJ., van Lint, J. W. C., & van Zuylen, H. J. (2009). Bayesian committee of neural networks to predict travel times with confidence intervals. *Transportation Research Part C: Emerging Technologies*, *17*(5), 498–509. https://doi.org/10.1016/j.trc.2009.04.007

Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, *24*(5), 533–557.

Vlahogianni, E. I., Karlaftis, M. G., & Golias, J. C. (2014). Short-term traffic forecasting: Where we are and where we're going. *Transportation Research Part C: Emerging Technologies*, *43*, 3–19.

Wang, B., Kim, I., & Vu, H., L. (2019). Short-term traffic prediction using a spatial-temporal CNN model with transfer learning. *Transportation Research Record*, 1–16.

Wang, L., Geng, X., Ma, X., Liu, F., & Yang, Q. (2019). Cross-city transfer learning for deep spatio-temporal prediction. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 1893–1899. https://doi.org/10.24963/ijcai.2019/262

Wang, Y., Jia, R., Dai, F., & Ye, Y. (2022). Traffic flow prediction method based on seasonal characteristics and SARIMA-NAR model. *Applied Sciences*, *12*(4), 2190. https://doi.org/10.3390/app12042190

WARDROP, J. G. (1952). Road paper: Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, *1*(3), 325–362. https://doi.org/10.1680/ipeds.1952.11259

Weber, M., Auch, M., Doblander, C., Mandl, P., & Jacobsen, H.-A. (2021). Transfer learning with time series data: A systematic mapping study. *IEEE Access*, *9*, 165409–165432. https://doi.org/10.1109/ACCESS.2021.3134628

Williams, B. M. (2001). Multivariate vehicular traffic flow prediction: Evaluation of ARIMAX modeling. *Transportation Research Record: Journal of the Transportation Research Board*, *1776*(1), 194–200. https://doi.org/10.3141/1776-25

Wu, W., An, S., Guo, J., Guan, P., Ren, Y., Xia, L., & Zhou, B. (2015). Application of nonlinear autoregressive neural network in predicting incidence tendency of hemorrhagic fever with renal syndrome. *Zhonghua Liu Xing Bing Xue Za Zhi = Zhonghua Liuxingbingxue Zazhi*, *36*(12), 1394–1396.

Wu, Z., Pan, S., Long, G., Jiang, J., & Zhang, C. (2019). *Graph WaveNet for deep spatial-temporal graph modeling*.

Xie, P., Li, T., Liu, J., Du, S., Yang, X., & Zhang, J. (2019). *Urban flows prediction from spatial-temporal data using machine learning: A survey*.

Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, 5668–5675.

Yongchang Ma, Chowdhury, M., Sadek, A., & Jeihani, M. (2009). Real-time highway traffic condition assessment framework using Vehicle–Infrastructure Integration (VII) with Artificial Intelligence (AI). *IEEE Transactions on Intelligent Transportation Systems*, *10*(4), 615–627. https://doi.org/10.1109/TITS.2009.2026673

Yu, B., Yin, H., & Zhu, Z. (2017). *Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting*. https://doi.org/10.24963/ijcai.2018/505

Yu, J. J. Q. (2021). Citywide traffic speed prediction: A geometric deep learning approach. *Knowledge-Based Systems*, *212*, 106592. https://doi.org/10.1016/j.knosys.2020.106592

Yu, Z., Sun, T., Sun, H., & Yang, F. (2015). Research on combinational forecast models for the traffic flow. *Mathematical Problems in Engineering*, *2015*, 1–10. https://doi.org/10.1155/2015/201686

Zhang, D., & Kabuka, M. R. (2017). Combining weather condition data to predict traffic flow: A GRU based deep learning approach. *2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech)*, 1216–1219. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2017.194

Zhao, Z., Chen, W., Wu, X., Chen, P. C., & Liu, J. (2017). Lstm network: a deep learning approach for short-term traffic forecast. *IET Intelligent Transport Systems*, *11*(2), 68–75.

Zheng, J., & Huang, M. (2020). Traffic flow forecast through time series analysis based on deep learning. *IEEE Access*, *8*, 82562–82570. https://doi.org/10.1109/ACCESS.2020.2990738

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2019). *A comprehensive survey on transfer learning*.

Declaration

I hereby confirm that the presented thesis work has been done independently and using only the sources and resources as are listed. This thesis has not previously been submitted elsewhere for purposes of assessment.

Munich, 18.01.2023, Abu Sayed

Place, Date, Signature