URBAN TRAFFIC FLOW PREDICTION, A SPATIAL-TEMPORAL APPROACH

Dissertation submitted in partial fulfilment of the requirements for the Degree of Master of Science in Geospatial Technologies.

Dissertation supervised by
Prof. Dr. Edzer Pebesma
Institute of Geoinformatics
University of Munster

Dissertation co-supervised by
Joel Dinis Baptista Ferreira da Silva,
PhD Instituto Superior de Estatística e Gestão de Informação,
Universidade Nova de Lisboa Lisbon, Portugal

Dissertation co-supervised by
Pro Jorge Mateu Mahiques
Department of Mathematics
Universitat Jaume I Castellón de la Plana, Spain

February, 2021

ACKNOWLEDGMENTS

ABSTRACT

Current advances in computational technologies such as machine learning combined with traffic data availability are inspiring the development and growth of intelligent transport Systems (ITS). As urban authorities strive for efficient traffic systems, traffic forecasting is a vital element for effective control and management of traffic networks. Traffic forecasting methods have progressed from traditional statistical techniques to optimized data driven methods eulogised with artificial intelligence. Today, most techniques in traffic forecasting are mainly timeseries methods that ignore the spatial impact of traffic networks in traffic flow modelling. The consideration of both spatial and temporal dimensions in traffic forecasting efforts is key to achieving inclusive traffic forecasts. This research paper presents approaches to analyse spatial temporal patterns existing in networks and goes on to use a machine learning model that integrates both spatial and temporal dependency in traffic flow prediction. The application of the model to a traffic dataset for the city of Singapore shows that we can accurately predict traffic flow up to 15 minutes in advance and also accuracy results obtained outperform other classical traffic prediction methods.

KEYWORDS

Traffic Flow

Prediction

Spatial Temporal Modelling

Machine Learning

Urbanisation

Convolutional Long Short-Term Memory

ACRONYMS

ITS	Intelligent Transport Systems
GPS	Global Positioning Systems
GIS	Geographical Information Systems
QGIS	Quantum Geographical Information Systems
ML	Machine Learning
API	Application Programming Interface
AI	Artificial Intelligence
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
RMSE	Root Mean Square Error
MSE	Mean Square error
LSTM	Long Short-Term Memory
Conv_LSTM	Convolutional Long Short-Term Memory
CNN	Convolutional Neural Networks
RF	Random Forest
LR	Linear Regression
XGB	Extreme Gradient Boosting
GPU	Graphics Processing Unit
TPU	Tensor Processing Unit

TABLE OF CONTENTS

\

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Motivation and Rationale

Forecasting urban traffic flows is purposeful for overall traffic management, public health, and land use [1]. For major urban players, traffic forecasting guides the identification of future density areas that proceeds appropriate measures to mitigate traffic challenges [2]. For business domains, traffic forecasting provides indication for investment locations or rather movement of logistics and freight, and also, the public can better schedule their travel plans [3].

Today, urban population escalation has subsequently posed great pressure on prevailing transport networks in the bid to meet mobility demand [4], The multi-modal transportation systems are characterised by traffic congestion with both solid social and environmental impact [5] . As a strategy to calming traffic congestion, various urban metropolitan expanses have implemented measures including laning, configuring traffic signals, delineating vehicle, and pedestrian paths etc., [6]   but there has been a challenge of striking a balance between the ever-growing car ownerships and traffic infrastructure [2],[7]. As urban growth speculations rise, experts and stakeholders in the transportation research domain are today continuously challenged by the ever-changing traffic mobility patterns [5]. The understanding of urban mobility involves the conceptualization of routes that greatly link the origins and destinations of travellers [8], in order to achieve effective mobility systems. Efficient urban mobility systems are crucial driving elements for achieving sustainable economic development for future growth cities [9].

Modern cities are essentially adopting intelligent transportation systems (ITS) while others are considering ITS as future growth areas as a strategy to assessing traffic mobility dynamics so as to provide solutions to traffic challenges [10]. ITS provide possibilities of capturing enormous traffic data [11], for example through; city traffic management systems that gather network traffic state data that comprises elements such as traffic volume, flow rate and speed. These traffic elements are collected using varied means including loop detectors, GPS sensors and probe vehicles trajectories[12]. Vehicle trajectories are today considered the most reliable traffic data collections for transportation research [9], [13], this is because of the leverage and inclusive supplication of user-centric movement behaviours such as; (vehicle speeds and travel times) and spatial-temporal trip accumulation [14], [15]. The growth of ICT has posed a great opportunity for major cities to strategize for Intelligent Transportation Systems (ITS), through the massive generation of traffic-related spatial-temporal data. This provides a good dimension of big data that can be analysed for a wider intuition of travel and mobility behaviour in urban transportation networks [14], [16].

Transportation studies focused on developing traffic information and management systems are taking advantage of this data to conduct analysis that importantly derive accurate short-term traffic conditions such as traffic density , speed flow, travel time and flux [8]. Tracing back to the early 1970s [17], transportation researchers have been finding ways of exploiting historical traffic data to forecast short-term traffic flow by employing different prediction methods. These prediction methods have evolved from traditional statistical methods to modern statistical modelling approaches and machine learning methods that are mainly data driven [18]. As traffic related data hugely accumulates, it poses great opportunities and yet challenges for implementing more comprehensive and focused analyses for design, planning and subsequent management of transportation systems [19], [20]. Major challenges of big data volumes are linked to the complexity of examining both the spatial and temporal perspectives [2], [21].

Most traffic flow prediction methods are designed to empirically fit the variational behaviour comprised in traffic data by putting emphasis on function concepts based on historic traffic data without considering mobility mechanisms in road networks [22]. Traffic flow mechanisms are as a result of confluences in road links that directly affect traffic flows in neighbouring roads [23]. The modelling of these confluences requires significant incorporation of all traffic characteristics (spatial and temporal) in order to achieve more meaningful traffic flow analyses and forecasts [21]. The modelling of traffic mechanisms and behaviour has been viewed as a complex task in transportation research, thus we see less research in this field [24].

Data-Driven traffic studies have recently gained focus in the effort to achieving information-based decision making for traffic monitoring and planning [8], [25]. Present day traffic modelling approaches that are statistical, or machine learning based mainly consider the temporal dimension of data [26]. Recently, studies are being undertaken to develop modelling approaches that exploit both the spatial and temporal characteristics embedded in traffic data so as to produces more accurate traffic forecasts [27]. Development in this field is still in its early stages [28] and therefore this paper undertakes a study to perform urban traffic flow prediction with a focus on the inclusion of spatial and temporal dimensions of traffic data. Using a machine learning approach, a Convolution Long Short-Term Memory (Conv_LSTM) model is proposed and implemented to achieve short-term traffic flow predictions starting from 15minutes and next intervals.

1.2 Knowledge Gap Identification

Different approaches ranging from clustering methods, statistical methods to classic machine learning techniques have been applied in transportation research [3],[18]. In [29], authors designed Logistic Regression and Spatial-Temporal Autoregressive Moving Average

statistic models that learn trajectory data to measure traffic conditions in networks. Machine learning methods such as Graph Neural Networks (GNNs), Artificial Neural Networks [30], have been applied to considerable traffic studies. It can be appreciated that the aforementioned techniques have provided ways of extracting meaningful traffic patterns, however, there lies a gap in modelling both spatial and temporal characteristics in traffic related studies [27], [31]. Existing transport research highlights that transportation networks exhibit significant spatial and temporal characteristics [32],[33] that must be incorporated and simultaneously modelled in traffic forecasting. However, most traffic studies mainly implement time series methods only considering the temporal dimension without modelling the spatial aspect as most researchers consider it to be a complex task [32], [34], [35]. The impact and importance of spatial-temporal dependencies in arterial networks is clear but it is often overlooked [36], and currently, it's still a demanding area in need of further research and development in the traffic forecasting field [37]. More recently attempts have been made to integrate both spatial and temporal dependencies in the modelling process, [38] while integrating other exogenous variables like historical weather and geo-tagged tweets to improve forecasting accuracy. As traffic data such as vehicle trajectories accumulate, it presents challenges for exploring, analysing, and visualising embedded patterns in the traffic data [39]. This calls for subsequent solutions such as exploitation of big data computational tools that provide potentials of mining and modelling spatial-temporal traffic flow patterns that can aimfully accelerate meaningful decision making for traffic systems [40].

1.3 Research Aim

The main aim of this research is to implement machine learning techniques while harnessing vehicle GPS data, to proactively incorporate both spatial and temporal characteristics in performing traffic flow prediction. The following questions will guide the achievement of the research aim;

1. How can we quantify spatial temporal patterns embedded in trajectory traffic data?

2. How can we develop traffic flow prediction models by considering both spatial and temporal dimensions of traffic data?

3. How can we assess the significance of integrating traffic spatial features on model performance?

1.4 Thesis Structure

The research is organized as follows:

• Chapter 2 Presents literature review on traffic flow prediction modelling as well as techniques being implemented.

- Chapter 3 Highlights the case study area, datasets used in implementing this research.

- Chapter 4 Demonstrates the implementation of our methodology approach and techniques used.

- Chapter 5 Presents the analysis results from methodology implementation.

- Chapter 6 Presents a discussion on research findings and also highlights some of the limitations to this research.

- Chapter 7 Presents conclusion remarks and insights about possible future research works.

2. LITERATURE REVIEW

Traffic related problems in urban areas, including accidents, congestion and, air pollution create a major impact on the livelihood quality of citizens and also directly affect sustainable development efforts of cities [39]. In the bid to alleviate traffic challenges, urban metropolitans are currently relying on traffic data detection technologies, big data analysis, mining, computational technologies, optimization concepts and designing of Intelligent Transportation Systems (ITS) as the key tools to providing solutions that help relieve traffic challenges [12]. The exploration and analysis of patterns in traffic flow as well as predicting its dynamic traits are fundamental actions to achieving applications of ITS [41], such as Advanced Traffic Signal Control Systems(ATSCS) and Advanced Traffic flow Management and Control Systems (ATMCS) [40]. This chapter gives an overview on modelling approaches related to this research in the field of traffic flow forecasting. It gives light on both the history and advances in traffic prediction modelling with a major focus on progression in methods that are incorporating both spatial and temporal features in traffic prediction.

2.1 Background on Traffic Modelling Approaches

Traffic flow modelling approaches can be divided into two major parts [42]. The first involves the specific modelling of the general traffic system itself by undertaking simulations that take into consideration of traffic parameters such as traffic signal control, vehicle counts and driver's behaviour [43]. This approach is largely built on the foundations of traffic flow theory in which traffic models are developed from mathematical equations and scientific theories to derive relationships between variables such as signalling, congestion, traffic density and flow [18]. The advantage with these methods is that they consider traffic control measures in the prediction process and try as much to quantify traffic conditions in road networks [43]. However, this approach has several drawbacks especially to do with computational complexity of model prediction parameters, calibration of mathematical equations and lack of flexibility [44]. The prediction quality of these models largely lies on the quality of inputs that are manually determined; they cannot inherently learn latent traffic patterns from traffic datasets [36]. This research paper focuses on the second major approach for traffic prediction that is characterised by data driven methods such as statistical methods and machine learning. Machine Learning approaches perform to determine functional approximations between input features and output features and learn patent relations in traffic datasets to forecast variables such as traffic speed, volume, flow etc[3], [45].

In recent years, advancements in traffic flow predictions have been majorly accelerated by improved traffic data collection in parallel with both progress in information technology and big data computation potentials [46]. Despite this, existing approaches have presented

considerable variations in traffic forecasts that still call for continued research and development in the traffic related studies as highlighted in [47]. One such area is the inclusion of both spatial and temporal relationships embedded in traffic data [21]. The following sections will give highlight on relevant literature related to traffic forecasting while giving emphasis to spatial temporal approaches for improving traffic forecasting.

2.2 Traffic Prediction Modelling

Traffic flow prediction approaches are majorly categorised as either parametric or nonparametric approaches [48]. Recent studies show that parametric models deliver powerful statistical results, however, nonparametric approaches that are mainly data driven have offered great advantages regarding handling of complexity and this has motivated continued investigations focused on improving nonparametric prediction models [49],[50].

2.3 Parametric Traffic Forecasting Approaches

Parametric models are algorithmic and mainly mathematical based methods [41], whose structure is determined in advance by estimating a set of parameter values that are usually derived from empirical data [51]. Parametric models employ conventional statistical notions to capture temporal patterns and trends in timeseries data to be able to perform traffic flow predictions[48]. Some examples of such models include the Autoregressive Moving Average (ARMA) model, and the Kalman filter that are largely considered to be classical time series models [51]. As traffic prediction is considered a time series task [51],[52] , models such as the Autoregressive Integrated Moving Average (ARIMA) [53], are today applied to short-term traffic flow prediction tasks. In [54] they implemented an Autoregressive Integrated Moving Average (ARMA) model by specifying ARMA's key parameters to predict bottleneck traffic congestion, they assert that ARMA has powerful mathematical and statistical basis to derive predictive intervals. There are varied versions of ARIMA [51], [55], in [10] they implemented a Seasonal ARIMA model to predict traffic in which the prediction parameters were found by a maximum likelihood method. Furthermore, we see more advanced statistical modelling methods applied in traffic flow prediction; In [56] they developed a hybrid prediction model that incorporates space time features by integrating vector autoregression (VAR), and ARIMA for freeway traffic speed prediction. In [57] , they designed Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models to forecast road travel time by enhancing capabilities to learn and extract trend and seasonal characteristics in the decomposition modelling processes. Results show that these models are efficient in modelling uncertainties in traffic, however, due to their modest structure and applied theoretical assumptions, they are weak in capturing complex variations inherent in traffic flow patterns especially in composite

non-linear situations, and are highly affected by the curse of dimensionality [58]. Other parametric models implemented in traffic forecasting include Historical averages (HA), Historical Mean (HM), and linear regression (LR), [59], [60].

Parametric methods provide simplicity in the modelling process, with low model complexity and are easy to understand as well as implement [50], however, road traffic is complex and is always dynamically changing, which parametric approaches cannot learn as they don't model uncertainty well [61]. Therefore, the performance of the parametric methods in traffic prediction is considerably influenced by underlaying external factors in traffic systems [62],[48].

2.4 Non-Parametric Traffic Forecasting Models

To overcome the weakness of classical statistical modelling approaches, researchers have currently put their efforts on harnessing non-parametric approaches that are majorly machine learning techniques [45],[46],[62]. Non-parametric methods are more intelligent techniques that largely depend on training big data to build the model structure[19]. They perform to mine historic conditions that are similar to the conditions at prediction stages [44]. Most studies show that non-parametric methods have produced satisfactory results in traffic prediction and so there lies a trade-off between the non-complex parametric models and the complex but effective non-parametric techniques [59],[25]. Due to their great learning capabilities and flexible structure, machine learning models are today widely applied to traffic predictions and studies show that better results have been achieved compared to the classical parametric models [46].

Renown nonparametric models applied in traffic prediction include Neural networks such as (Artificial Neural Networks (ANN) [22], Convolutional Neural Networks (CNN)[63], Recurrent Neural Networks (RNN)[64]), K-Nearest Neighbour (KNN) [65], Support Vector Regression (SVR)[66], and Bayesian network models. In [67] they implemented a Bayesian classifier and support vector regression to model traffic flow, first they define relations in the road network and finally estimate the traffic speeds for the next time step using multiple linear regression and SVR. In [5] ANN is implemented to ascertain average vehicle' speeds on arterial roads as a way of quantifying traffic congestion.

There are variants of Recurrent Neural Networks, and all have the potential to extract sequential patterns embedded in input features [68]. Since traffic speed in arterial roads has temporal traits, Recurrent Neural Network models such as Long Short-Term Memory (LSTM) [69] and Gated Recurrent Units (GRU) [70] can be considered in forecasting traffic flow. In [41], an LSTM model which implements attention mechanisms to encode long-term dense traffic flow is proposed. Nevertheless, road traffic systems have both temporal and spatial

aspects, but almost all parametric models and RNN models majorly focus on only temporal characteristics [64]. Convolutional Neural networks (CNNs) have shown possibilities of translating traffic flow into space-time matrices as images, the model is able to capture features from the image as well as learn the temporal-spatial in traffic flow [35].

2.5 Spatial Temporal Traffic Forecasting

Most recent studies have put more effort to studying how weather information can be incorporated in traffic studies and not how the spatial topology of roads can be incorporated in prediction modelling [21] . Several pioneering techniques in traffic studies have focused on traffic observations of the target location which has been mainly characterised by time series modelling without considering neighbouring locations and road topology [13]. In effort to addressing this traffic influencing aspect, studies have recently started finding ways of incorporating both spatial and temporal dimensions in traffic prediction [21], [71]. Developments have been in finding ways of enriching timeseries modelling with spatial adapting models [46], [13]. Authors in [72] presented the first attempt in considering traffic conditions of neighbouring roads in their research based applying the Kalman Filter algorithm to predict traffic flow. Basing on the first law of geography, that enlightens about how nearby things share relationships [73], studies have based on this to explore possibilities of integrating road network inherent spatial dependencies in traffic forecasting [35]. Non-parametric models especially deep learning models present possibilities of capturing important non-linear spatiotemporal associations in traffic forecasting [21],[74]. Studies such as in [75] performed time series analysis with geometric correlation techniques that resulted in producing 3D heatmap images to illustrate existing relationships of traffic states in nearby roads. In [35], authors developed a prediction framework that integrates Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) to simulate the extraction of spatial-temporal traffic traits in road networks. Patent models are being improved to learn both spatial and temporal patterns, in [67] they developed a spatial temporal Bayesian network for traffic forecasting in which they also employed a gaussian mixture model to define statistical relationships between input features and output features. Other spatial temporal traffic forecasting models Include Spatial Temporal Autoregressive Integrated Moving Average. In [76], they developed a deep learning architecture that combines Convolution and LSTM (Conv-LSTM) to forecast short term traffic, the structure was developed to fully fuse both spatial and temporal features of neighbouring links and regions of the target location. They then added another trait of LSTM to extract periodic traits in the traffic datasets. The combination of CNN and LSTM in traffic studies has considerably accelerated the capturing of sequential spatial and temporal dependencies [77]. In this research, we use a variant model

of LSTM called Convolutional LSTM (Conv-LSTM) that uses some that uses some CNN capabilities to handle both spatial and temporal relations in road networks to predict traffic flow.

2.6 Convolutional – Long Short-Term Memory (Conv_LSTM)

Conv_LSTM is a variant of the LSTM neural network introduced by authors in [78] as they were modelling the nowcasting prediction task. It was introduced due to other variants of LSTM having redundancy for spatial features. Conv-LSTM is unique in a way that it incorporates convolutional operators in every LSTM cell (Figure 1) The convolutional operation helps to extract and learn spatial characteristics embedded in input data. The application of convolution operators replaced internal matrix products in LSTM cells, to stimulate the model in reading of two dimensional (2D) spatial data inform of rows while observing their time dynamics and dependencies. The Conv_LSTM has convolutional structures (Conv_LSTM) in the input-to-state as well as the state-to-state consecutive transitions. The model is achieved by stacking several Conv-LSTM layers thus creating an encoding- spatiotemporal sequence forecasting structure [79]. Defining components of the Conv_LSTM is that all the inputs features $x_1 \ldots \ldots x_t$, cell output features $c_1 \ldots \ldots c_t$, hidden state features $H_i \ldots \ldots \ldots H_t$, and its gates i_t, f_t, o_t are 3D tensors having their last two dimensions as spatial dimensions in the form of rows and columns. The input features and state features are better understood as vectors embedded on a spatial grid [80].

The Conv_LSTM components perform to establish the future state of a particular cell in the spatial grid by incorporating both inputs and historical states of its neighbouring local cells. This operation is achieved by applying a convolution operator during the input-to-state as well as the state-to-state transitions as seen in the main equations of a Conv_LSTM shown in (eq. 1) below, where the convolution operator is denoted by '$*$' and, the Hadamard product denoted by '$\circ$':

$$i_t = \sigma(W_{xi}*X_t + W_{hi}*H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \tag{1.a}$$

$$f_t = \sigma(W_{xf}*X_t + W_{hf}*H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \tag{1.b}$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc}*X_t + W_{hc}*H_{t-1} + b_c) \tag{1.c}$$

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ c_{t-1} + b_o) \tag{1.d}$$

$$H_t = o_t \circ \tanh(C_t) \tag{1.e}$$
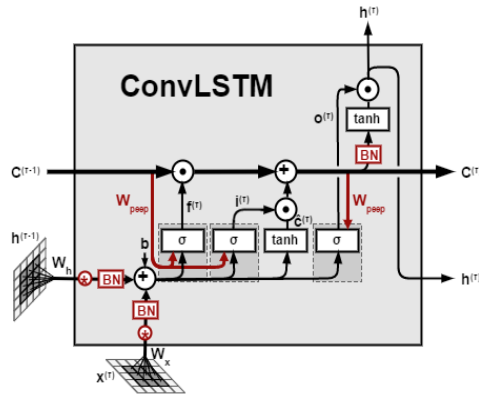
Figure 1: Internal Structure of Conv-LSTM

Initial features fed as inputs to a Conv-LSTM block, X , form a tensor whose shape is defined by the number of timesteps under consideration and the number of channels(T) present in the initial input data (N_c) as well as the height and width of the blocks inform of rows and columns defined by (N_H and N_W). X_t defines the input at the tth time-step. This gives the shape of the tensor to be (T, N_H, N_W, N_c). The output dimension shape of a Conv-LSTM block is similar to the input, though the number of channels is defined by the number in the convolution layers). Figure 2 shows the structure of a Conv_LSTM subsequent. In general, Conv-LSTM lattices apply convolution operations to the outputs of previous convolutions and by doing this, subsequent Conv-LSTM blocks successively extract and identify more hidden and complex relationships in data compared to previous Conv_LSTM blocks .
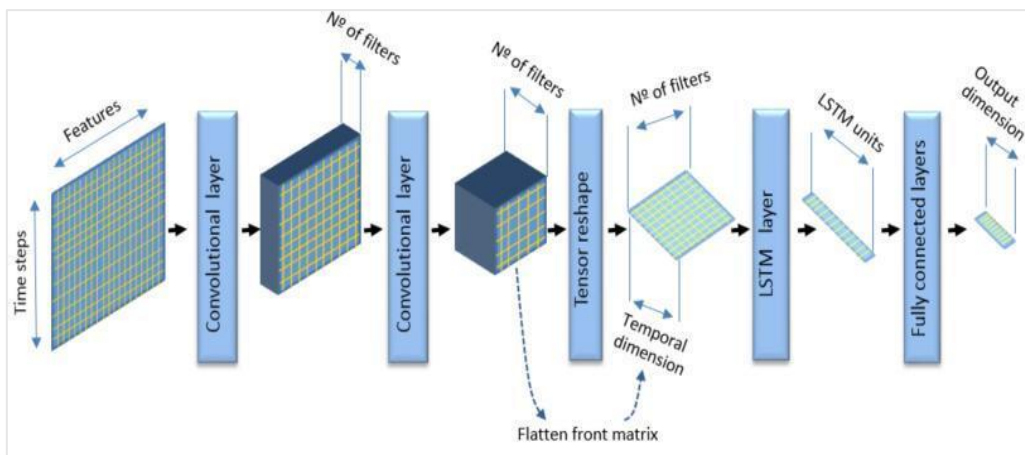
Figure 2: Operation Model structure of a Conv-LSTM model

Authors In [76] Implement modelling for traffic flow using Conv_LSTM by structuring a number Conv_LSTM units to learn traffic spatial temporal features. The first Conv LSTM unit handles input traffic flow features while creating a hidden state at every time step and continues to feed the features into spatial attention convolution layer. Furthermore, a second Conv_LSTM

is placed to learn inherent spatial- temporal dynamics passed from the attention convolution unit. In [81], they developed a traffic demand forecasting structure basing on Conv_LSTM. Their structure treats historical travel demand data as a video series stream, and then employs a Conv_LSTM model to forecast demand. They efficiently used Conv_LSTM considering both spatial and temporal features during prediction and the process was effective. In [82], [83] authors present modifications to Conv_LSTM that would improve capturing both global and local dependencies in traffic data, thus supporting the long-range prediction of traffic conditions, they implemented this approach on traffic sensor data and their structural model proved effective. Only a few studies have studied Conv_LSTM model tor traffic related studies with the most application to traffic demand [27]. Very little research has been done in the application of Conv_LSTM to traffic flow prediction. This research sets out to organise vehicle trajectory data and implement a Conv_LSTM model for traffic flow prediction for the city of Singapore.

3. STUDY AREA AND DATASET

3.1 Study Area

We chose the city of Singapore as our case study area; it is located southern of Malaysia which is the most southern part of Asia. In particular, we undertook this research for the capital city of Singapore. Being a small city, it is densely populated with its population density estimated to be about 7800 inhabitants per square kilometre [84]. The continued economic growth in Singapore has triggered a lot of economic activity characterised by increase in movement and population all of which point to increase in travel demand [85]. Traffic authorities in Singapore identified that privately owned cars are the major contributors to traffic congestion events on city arterial roads during day peak hours. It is evident that slow traffic flow in most cities is experienced during peak movement and thus we take Singapore to predict traffic flow with approaches that can be implemented to other cities in the bid to manage urban traffic.
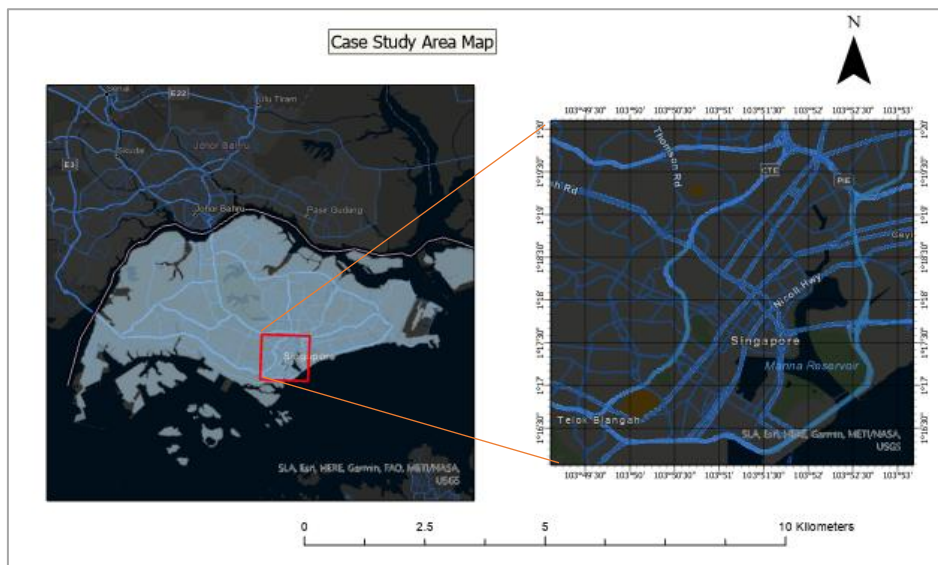
Figure 3: Study Area-Singapore city

3.2 Dataset

We used the Grab-Posisi taxi trajectory dataset containing about 80 million observation records spanning over 1miilion km of travel length covered [86]. Grab Posisi is a taxi company operating in Singapore. The GPS trajectories are collected from Grab taxi drivers' gadgets while in operation transit. Each trajectory is labelled with a trajectory ID, and for every ID attributes like coordinates, speed, timestamps are captured. The temporal range of the data is two weeks starting form 8th April 2019 to 22nd April 2019 and includes a total of 84000 trajectories.

Figure 4: Dataset extent

For this research, we chose a (6kmX6Km) bounding box centrally to the capital of Singapore to carry out our analysis for prediction. For all the dataset temporal range, we extracted 20000 trajectory events that took place within the bounding box in figure (6).

4. METHODOLOGY

This chapter describes the methodology we implemented following a three-phase approach shown in figure (5) to stimulate the exploration of traffic flow patterns from exploiting pre-processed large-scale vehicle trajectories and subsequently undertaking modelling for traffic flow prediction. In phase one, we describe the activities of data processing and transformation that improve the data quality for both traffic pattern analysis and traffic prediction. The second step includes activities for traffic pattern analysis and the final step involves activities done to predict traffic flow.
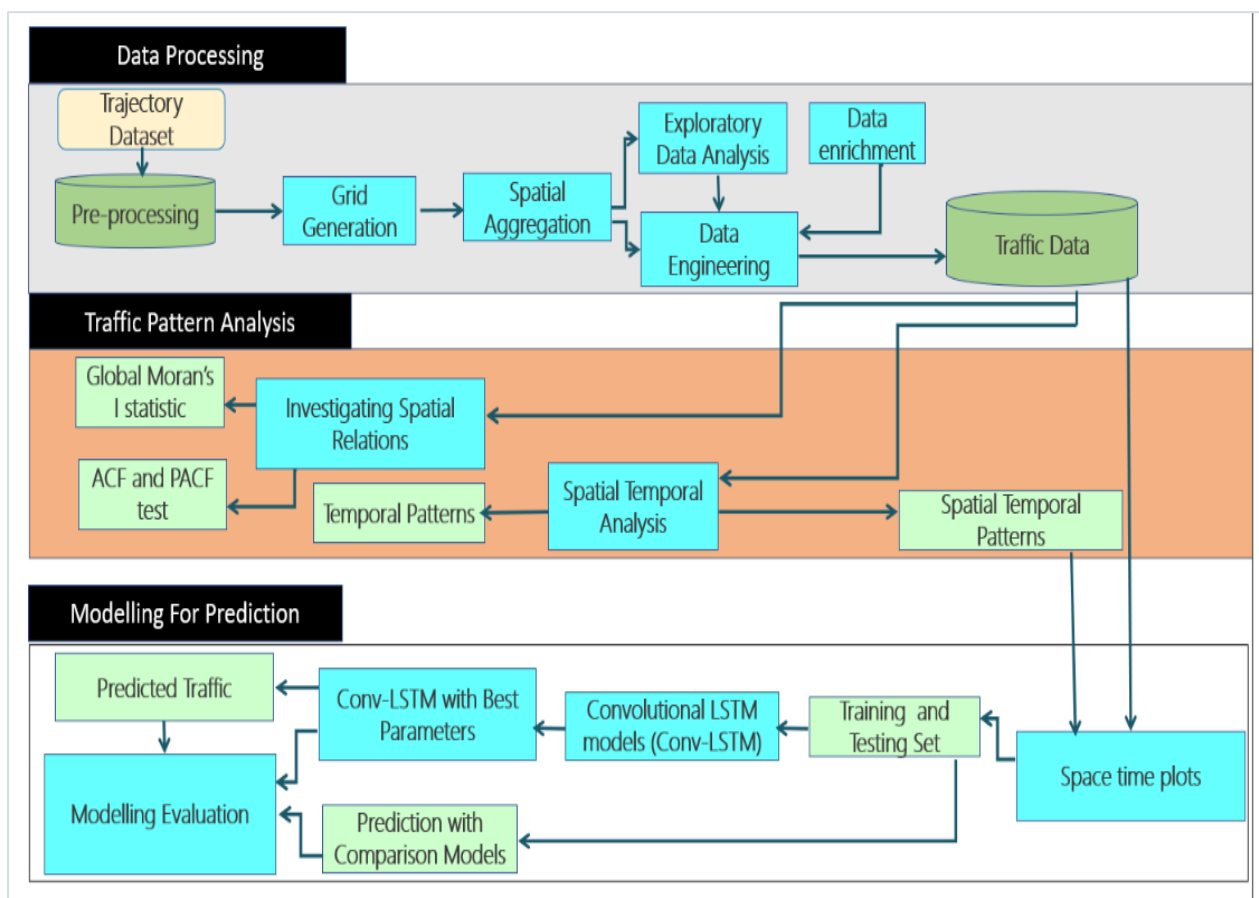
Figure 5: Methodology workflow

4.1 Phase 1 - Data Processing.

Python software and Quantum GIS (QGIS) were used to perform preliminary data pre-processing on the trajectory dataset. Major functions done in the python environment were to facilitate improving of data quality by analysing primary variables and datatypes in the dataset as well as identifying and removing missing redundant values. Quantum GIS facilitated study area exploration, selecting, and clipping to study area.

4.1.1　Grid Generation and Spatial Aggregation:

For the bounding box defining our study area, we created a 100X100m grid over the study area to facilitate the spatial joining of the road network and the GPS trajectories (Figure 7). This was done to infer the trajectory data to their relative road position, the result is that the trajectory data was joined to the grid cells intersecting the road network. This helped to overcome the tiresome map matching process for over 12 million data records. Thus, grid cells were made to contain GPS points if any trajectory event happened in that grid-cell. The points were all related to centre the grid cell whose centroid coordinate was stored as a geohash code that we later decode or encode at given points of data processing.

Using the QGIS "join attributes by location" tool, we performed a spatial join between the grid layer and the GPS points basing on the intersection rule. The GPS points were assigned the centroid coordinate of the gird to which they belong, and this facilitated proceeding spatial aggregation of the traffic data. Given that every GPS point has a timestamp, we performed spatial aggregation for every cell to aggregate and resample the vehicle trajectories into time intervals of 15 mins, 30mins and 60 minutes for every date that a trajectory event happened in the grid cell. The result formed our primary traffic data in specific timesteps with their average speeds.
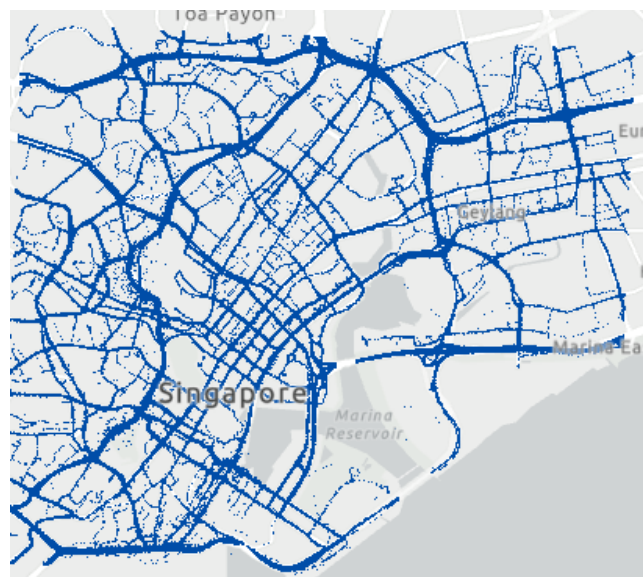
Figure 6: Raw Trajectory Data
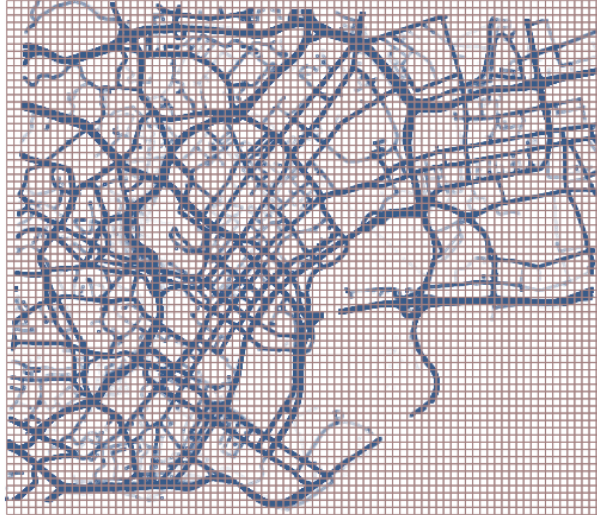
Figure 7: Spatial grid creation over trajectory data

4.1.2 Exploratory Data Analysis

Exploratory data analysis was carried out to gain more insights about the data, especially finding outliers. As we can see in figured 8, we encountered a high frequency of zero speed values, the zero values are because of vehicles stopping at traffic lights or at the end of journey. we investigated to eliminate all zero speed values observed at the end of each trip. Also, as we can see that there were several negative speed values, all these were eliminated from the dataset as we could not identify the cause of the negative values. Overall, the speed values in the dataset seem to form a normal distribution meaning that not so many values are varying from the mean speed values. We also analysed correlation between variables in the dataset (see Annex 9.2) and not much correlation was observed between the variables.
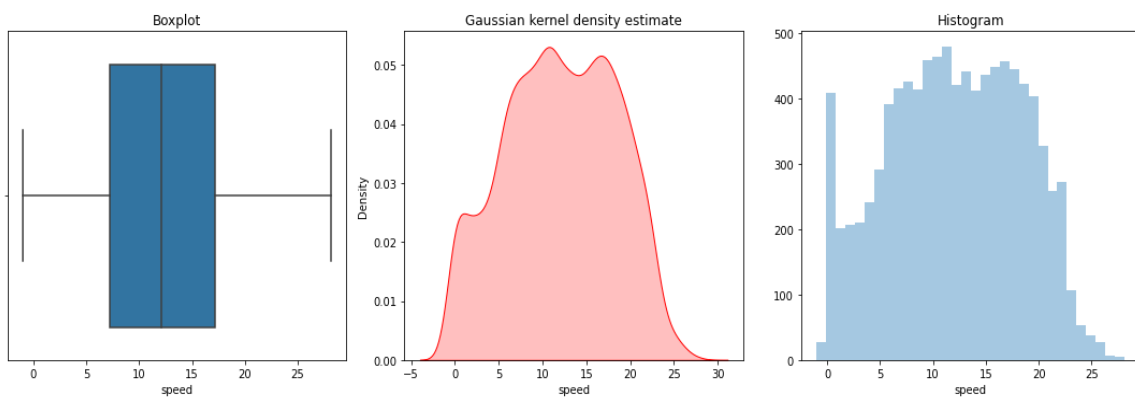
Figure 8: Distribution of speed values

4.1.3 Traffic Data Engineering:

This phase involved data scaling and feature creation. Data scaling was done to normalise speed values to lie between 0 to 1 using the formular in Equation 2. The timestamp variable

was further exploited to extract more features such as hour, day of week, weekends and holidays so as aid more focused analysis. Furthermore, we also transformed the time, latitudes and longitudes into normalised forms thus creating variables that suit the modelling process. We finally transformed the data into a matrix, and this formed our traffic dataset for the traffic pattern analysis and modelling stage.

$$Z_i = \frac{x_i - (\min{(x)}}{\max(x) - \min{(x)}} \qquad (2)$$

Where, Z_i is the new normalised value, x_i is the original variable value, $\min{(x)}$ and $\max(x)$ are the minimum and maximum values of the entire variable under consideration.

4.2 Phase 2 - Traffic Pattern Analysis

The aim of this analysis is to identify spatial-temporal dynamics embedded in the dataset. Using python software, we set out to investigate traffic patterns at varied time scales to analyse how traffic patterns evolve over time in different locations of the study area. Using focused data manipulation and computations, we produced visualisations showing traffic behaviour at different times of the day in different locations. We also investigated to see if traffic patterns are the same across the days of the week. Using heat map analysis, we were able to visualize which areas experience traffic slowness at any time of the day. One of our goal was to investigate spatial autocorrelation and temporal autocorrelation in the traffic data, this allowed us to ascertain whether traffic patterns on a given section share dependencies with patterns on neighbouring routes. The spatial autocorrelation was computed using the Global Moran's I spatial autocorrelation tool in ArcGIS software. The tool concurrently considers both the location and feature values to measure spatial autocorrelation, see equation (3). Given a set of variable elements and related attributes, the tool investigates to ascertain if the existing pattern in the features is random, clustered, or dispersed. Important insight is gained from analysing the output values of the Moran's I Index, the Z-score value and the p-value which tell much about the significance the statistic. We analysed spatial autocorrelation for the thirty-minute and one-hour average speed variables.

$$l = \frac{N}{S_0} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w(i,j)(x_i - \overline{x})(x_j - \overline{x})}{\sum_{i=1}^{N}(x_i - \overline{x})^2}, j \neq i \qquad (3)$$

where $\overline{x}$ is the mean of x_i and $w(i,j)$ is the connectivity spatial weight between *i* and *j.*

Since traffic prediction is a time series problem, we took on to assess if average speeds in the datasets are correlated with previous values. We examined the statistical significance of temporal autocorrelations contained in the dataset. The standard Autocorrelation Function

(ACF) was applied to derive graphs depicting the level of autocorrelations between traffic flow speeds with lagged values at previous time-steps. The Autocorrelation Function computes and plots the mean correlation for varied lag length between time series dataset points and previous point values [87]. Given a time series analysis; *y1, y2, y3, y4, y5.......yt* at a point under consideration, the autocorrelation of lag k is derived by equation (4) below.

$$\rho(k) = \frac{E[(Y_t - \bar{x})(Y_{t+k} - \bar{x})]}{\sigma Y_t \sigma Y_{t-k}} \quad k = 0, \pm 1, \pm 2, \dots \dots \tag{4}$$

Partial autocorrelation (PACF) was also examined, its similar to the standard autocorrelation though varies in such a way that it considers the influence of correlation between immediate observations in a time series. Given Y_t and Y_{t-1} observations in a time series, the Partial Autocorrelation Function is given by equation (5):

$$\rho Y(K) = \frac{Cov([Y_t | Y_{t-1}, Y_{t-1}, \dots, Y_{t-k+1}], [Y_{t-k} | Y_{t-1}, Y_{t-1}, \dots, Y_{t-k+1}])}{\sigma[Y_t | Y_{t-1}, Y_{t-1}, \dots, Y_{t-k+1}] \sigma[Y_{t-k} | Y_{t-1}, Y_{t-1}, \dots, Y_{t-k+1}]} \tag{5}$$

4.3 Phase 3 - Modelling for Traffic Prediction

The prediction process was implemented using Keras API and tensorflow in python software. Due to the high computation involved with modelling process, we conducted the modelling and training processes in google collaboration cloud platform which is a platform that provides free computational hardware such as high performance GPUS and TPUs. The following highlights form the overall prediction process; we can see in figure 11 that processed traffic plots are reshaped and passed into the model at the input phase. The model is tuned to observe six (6) hour traffic data (24 consecutive plots, representing 15minutes interval for a 6hour period) and then finally predicts the next fifteen-minute traffic data thus giving predicted traffic images for a given space and time. The model's performance is finally evaluated using some renown evaluation metrics for traffic flow prediction.

4.3.1 Plotting Traffic Data into Space Time Images

This phase demonstrates preparation of initial inputs for the prediction model. We translated traffic data into space time images were average traffic speed values are characterised by gradation colours showing the spatial intensity of the values for each 15-minute timestep of every day. The heatmaps are generated to represent the spatial dimension and temporal dimension of the traffic data for each day (Figure 9). The spatial dimension is in this case a graph depicting the spatial intensity of the average speed values using the same color-scale type to represent the hierarchy of the traffic data. The temporal dimension represents the timestep for which each is plot is generated. Given that we considered 6 hours, 24 images are

generated to represent the fifteen (15m) minute interval across one day. All geohashes (only those traversed by road segments), were decoded to come up with a total of 47 unique latitudes and 54 longitudes. This facilitated the resultant plotting of 47X54 grids of traffic average speed data. The plots are initially split by day and corresponding timesteps, thus, for a given day and each geohash 47X54 grids traffic data plots are created. The traffic dataset was split using a portion 80% as training data and while using 20% as test data. A visualisation traffic data maps is shown in figure (10) below.
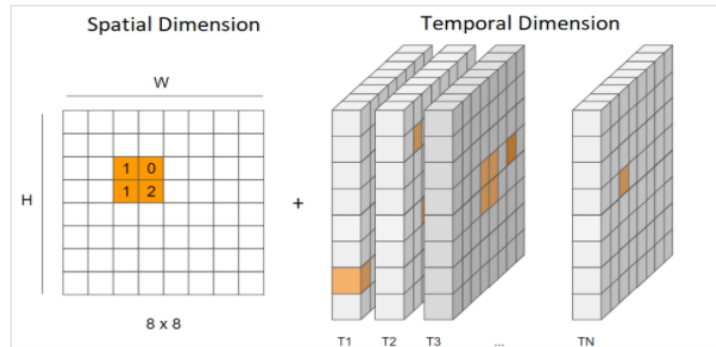
Figure 9: Spatial and temporal dimension of heatmap
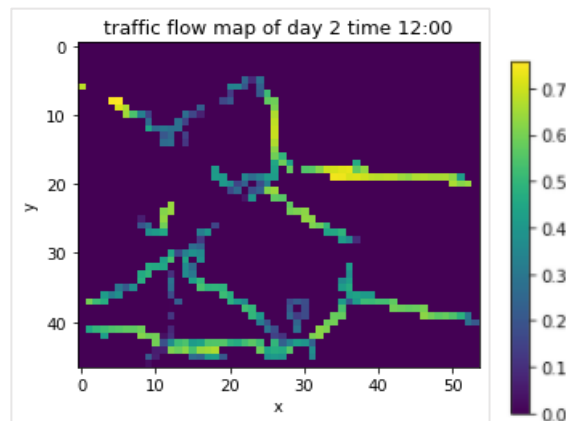
Figure 10: Example of traffic images

4.3.2   Encoding - Predicting Structure

The Conv-LSTM can be deployed differently to design a building block for varied prediction complex structures but for this particular spatiotemporal sequence traffic forecasting scenario, we applied the structure described in section 2.6. In the Conv-LSTM, intuitive features are extracted as feature vectors by CNN's potent feature extraction capability, while the LSTM is applied to replace the pooled layer usually used in the traditional CNN model to perform traffic prediction. The structure consists of both the encoding network and prediction network that are all made by stacking several convolutional layers. Similar to the approach employed in [88], the forecasting network copies its first states and cell outputs from the last state of the

result of the encoding structure. As our goal is to have the prediction output in the same dimensionality as the input features, initial input layers are stacked by the convolutional LSTM models and finally apply a three-dimensional (3) convolution (Conv3D) to produce the expected output layer. Throughout the encoding and prediction process, the LSTM captures temporal patterns and dynamics embedded in the traffic data, while the convolutional operation facilitates the extraction and inclusion of spatial dependencies embedded in the traffic patterns.

The performance of the Conv-LSTM was enhanced by applying the batch normalization method. Batch normalisation implements a normalization and regularization effect and so we applied it at every layer transition. Since Conv-LSTM has few parameters requiring less regularization, we did not use dropout. Due to the spatial relations encoded in feature plots or maps, activations can become highly correlated thus rendering dropout ineffective. The resultant model is a 4 layered 2D Convolutional LSTM each having 32 filters, with batch normalization after each layer. The model gives an output layer as a Convolutional 3D layer, producing traffic plot predictions in a similar format as input 96 x 47 x 54 x 1.

4.3.3   Tuning Model Hyperparameters:

Hyperparameters are cognizant parameters whose values influence the learning process that would have an impact on the result of the model. Hyperparameter tuning is essential for achieving best model performance on a given dataset [89]. In implementing the Conv_LSTM model, we explored a combination of different hyperparameters while analysing the performance results. This was done to achieve a suitable combination of hyperparameters that would simulate the learning process to enable us achieve good accuracy for the prediction task [90]. For our modelling framework, major hyperparameters tested include; The kernel size, learning rate, filters, batch size and number of layers, (see table 1). Setting the hyperparameters is done by tuning and interacting hyperparameters while measuring what performance accuracy can be earned by tuning it. The hyperparameters are manually tuned or rather adjusted while trying different hyperparameter combinations in machine learning platforms which in our case was TensorFlow with Keras API's . The hyperparameters that showed substantial impact on the training process and are easily tuned to our prediction goal were selected and implemented as final hyperparameters for our final model. In table (2) we give a summary of hyperparameters values as well as the performance values.

As a strategy to improving model performance, we tried adding more filters as well layers. Adding more filters aids the detection of more features embedded in the inputs thus helping to capture more traffic flow patterns in the road networks. So, we increased filters from 32 to 64 to see if there is improvement in the model prediction performance. Also, increasing more

layers was a strategy to aiding the extraction of abstract information from the traffic data, though in our implementation using more layers and filters showed no noticeable influence on the prediction performance as each accuracy results were almost the same. So, we finally used four layers.

The Kernel size depicts the width and height of the convolution window, adjustments to increase the kernel size facilitates the observation of hidden information in Conv_LSTM blocks (grids). We tested both the 5*5 and 3*3 kernel size in the performance to see if this would improve our prediction performance. Both kernel sizes showed almost similar results and thus we used a 3*3 size as being sufficient for our model.

Hyper Parameters	Value
Number of Layers	3 & 4
Kernel size	(3,3) & (5,5)
Filters	32 &64
Learning rate	0.001
Batch size	2

Table 1 Hyperparameters Tested.

No.	Kernel size	Layers	Filters	Batch size	Learning Rate	Accuracy %
1	3*3	3	64	2	0.001	94.83
2	3*3	4	64	2	0.001	94.8
3	3*3	4	32	2	0.001	94.88
4	3*3	3	32	2	0.001	94.75
5	5*5	4	64	2	0.001	94.78
6	5*5	3	32	2	0.001	94.89
7	5*5	3	64	2	0.001	94.8
8	5*5	4	32	2	0.001	94.83

Table 2 Showing hyperparameters combinations and accuracies achieved.

We employed Adam as the optimization function while maintaining its learning rate for all the model hyper tuning. Recent studies assert that Adam optimizer exhibits powerful performance in accelerating model performance. Since optimizers together with their learning rates are key to achieving good prediction results, we maintained Adam as the optimizer.

4.3.4   Performing Traffic Prediction

The plotted space time images gave a total of 24 images for each day (24 distinct time steps) and a total of 120 plots over 5 days. This forms our dataset for the prediction process and dividing it into 80% as training set and 20% as testing set. These traffic images (plots) are

then packed into 24 consecutive plots per training sample, meaning for every training sample is the tensor shape is given by 24 x 47 x 54 x 1 (24 consecutive traffic flow images in a single channel). They were then propagated to the Conv_LSTM neural network to train the model for precedent prediction of next plot see figure 11.

Given a time T, the prediction approach we are implementing is utilizing the model to predict the traffic flow plot or pattern at T+1, in the event that its input data is comprised of 24 previous traffic plots (T-23 to T, equivalent to 6 hours day's data in 15-minute intervals). Using the input data, the model should predict T+1 plot. Next, the T+1 traffic plot is added to the input dataset and we remove the T-23 plot. The resultant input data is then applied to predict the T+2 (the next 30 minutes) traffic flow plot. The process is then repeated until T+4 is predicted. Figure 11 below gives a simpler visualisation of how we achieve our prediction results after space time traffic plots are fed into the neural network. The operations to perform prediction are as explain in section 2.5.
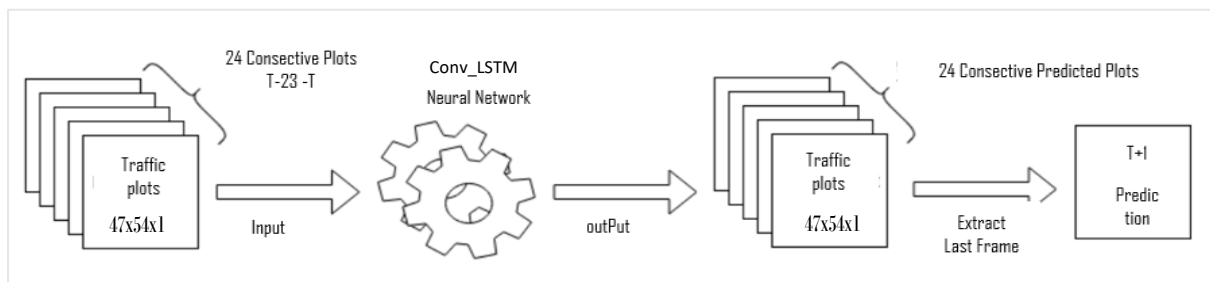
Figure 11: Workflow for implementing traffic prediction.

4.4 Prediction with Different Models

In order to analyse the performance of the Conv_LSTM model, we chose three statistical model algorithms and one neural network to act as comparison bases for our proposed model. The statistical algorithms include Random Forest (F), Linear regression(LR) and XgBoost(XGB) while the neural network includes Convolutional Neural Networks. Random forest (RF) performs predictions based on individual decision tree branches, each tree makes a class prediction vote and predictions are achieved from the class having most votes [91]. Linear regression (LR) analyses the relationship between variables(response and explanatory variables) and performs predictions by making linear approximations [92]. Xgboost (XGB) also referred to as Extreme Gradient Boosting creates new weaker models and successively integrates their predictions to improve model performance [93]. Convolutional neural networks (CNN) architectures comprise a convolution layer, pooling layer and fully connected layer and these form a unique framework for extracting critical features in data [94]. The CNN was autotuned to get best model hyperparameters while for the statistical methods appropriate

parameters were defined for better model performance. We performed Linear regression, number of estimators used for the random forest regressor were 50, while we used a number of 30 estimators for the Xgboost regressor. We applied testing data to all comparison models and results from the prediction processes were recorded inform of accuracy measures of Mean Squared Error and Root Mean Squared Error.

4.5 Model Evaluation

In [95], [96] they mention that the standard evaluation metric for traffic prediction are Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Most studies assert that RMSE is the standard metric for forecasting model errors [97]. This study only used MSE (Mean Square Error) and RMSE as basis for evaluating our model performance. The RMSE can be described as the square root of the mean of the squares of differences between actual and predicted values. The MSE and RMSE error evaluations are good measures of accuracy that help to tell how well the model can predict a given aspect.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|e_i| \qquad\qquad\qquad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}e_i^2} \qquad\qquad\qquad (7)$$

5. RESULTS

This chapter presents results obtained from implementing our methodology workflow. The source code for works implemented in python software can be found **here**.

5.1 Traffic Pattern Analysis

5.1.1 Spatial Autocorrelation Investigation

Spatial autocorrelation was investigated for one hour aggregated average speed (AvgSPD60) and thirty-minute aggregated average speed (AvgSPD30). Results for the Global Moran´s I statistic show a statistically significant Moran´s index that is equal to 0.447236 and 0.0.532554 for the AvgSPD60 and AvgSPD30 Variables, respectively. A P-value = 0.000000 and 0.00000 for AvgSPD60 and AvgSPD30 respectively is lesser than the accepted 0.05 alpha level which shows that the spatial autocorrelation is statistically significant for both temporal intervals.
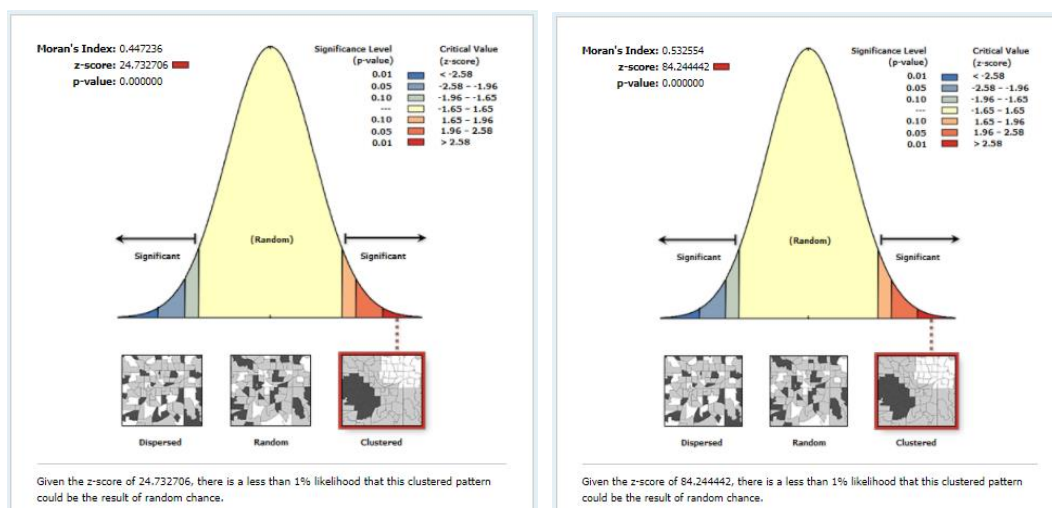
Figure 12: Spatial autocorrelation analysis

The resultant Z-score = 24.732706 and 84.244442 for the AvgSPD60 and AvgSPD30 respectively that are greater than +1.96 indicate the presence of a clustered pattern for the traffic flow speeds. A p-value can represent the probability of having the observed spatial pattern to be random, therefore, a P-value = =0.000000 and 0.0000 for AvgSPD60 and AvgSPD30 respectively that is lower than the accepted alpha level of 0.05 is considered low meaning that there is a low probability that the observed spatial pattern is random which justifies the existence of a global cluster of the traffic flow speed variables in the study area. With the resultant p-values, we may conclude that there is a 95% confidence level to reject the complete spatial randomness hypothesis and that the spatial distribution of the AvgSPD60 and AvgSPD30 is not created by random spatial process. These results indicate that there is

a high degree of spatial dependence of traffic flow speed across the study area and this proves that across different times of the day, there is spatial dependence in traffic patterns across road networks.

5.1.2 Temporal Autocorrelation Investigation

Basing on a fifteen-minute interval of aggregated traffic data, temporal autocorrelation in the dataset was analysed. The relationship between one observation and fifteen minutes prior was tracked across the whole traffic dataset. The resultant correlation in the data is plotted as bar charts between +1 and -1. Looking at figure(13) in the autocorrelation curve, statistically significant observations are above 0.2. We can thus see at the start that there is a statistically significant positive autocorrelation to a time lag of about 20 minutes. The Partial auto correlation graph also shows that there are statistically significant relationships for a lag of upto 1,2,3 and 4 (four) periods. This indicates there is temporal autocorrelation embedded in traffic patterns across the different day time intervals.
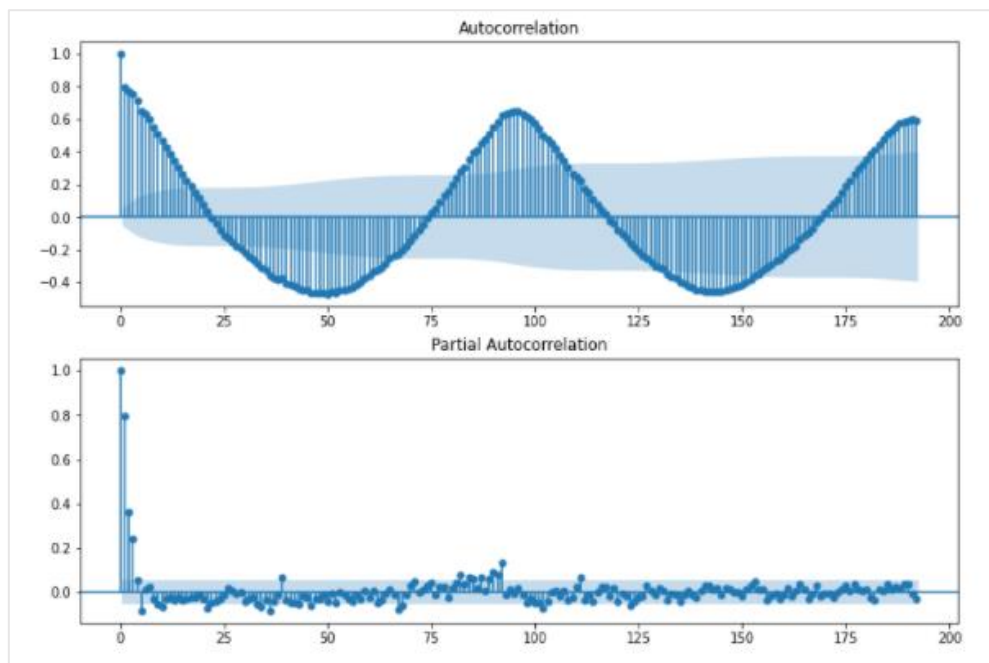
Figure 13: Showing Autocorrelation and Partial Autocorrelation in dataset

5.1.3 Spatial Temporal Analysis

Figure 11 shows hourly time of day and week-day variations of traffic patterns across the road network. During working days (day 1-5, and day 8-12) starting from 8:00- 22:00, we can observe low speed patterns across all days. Severe congestion patterns can be observed at in the rush hours for example between 7:00 to 10:00 morning period and 17:00 and 19:00

evening period with flow speeds dropping significantly. The patterns on weekends vary especially on Sundays for Day 7 and Day 14 were we see an increase in speed across the day hours compared to other days. Day 12 was a public holiday on Friday and therefore we see less congestion compared to other days. Fridays and Saturdays do not vary much from other normal working days which could mean that most people like to rest on Sundays. The traffic patterns observed during day hours indicate that movement speeds are highly influenced by daily commuting behaviour of the population. Figure (15) shows the maximum speeds observed across days and there is no significant variation in daily maximum speed observed. This could be due to speed limit guidelines in the study area. High flow speed patterns can be seen in early day mornings beginning the hour of mid night to about 6:00am.
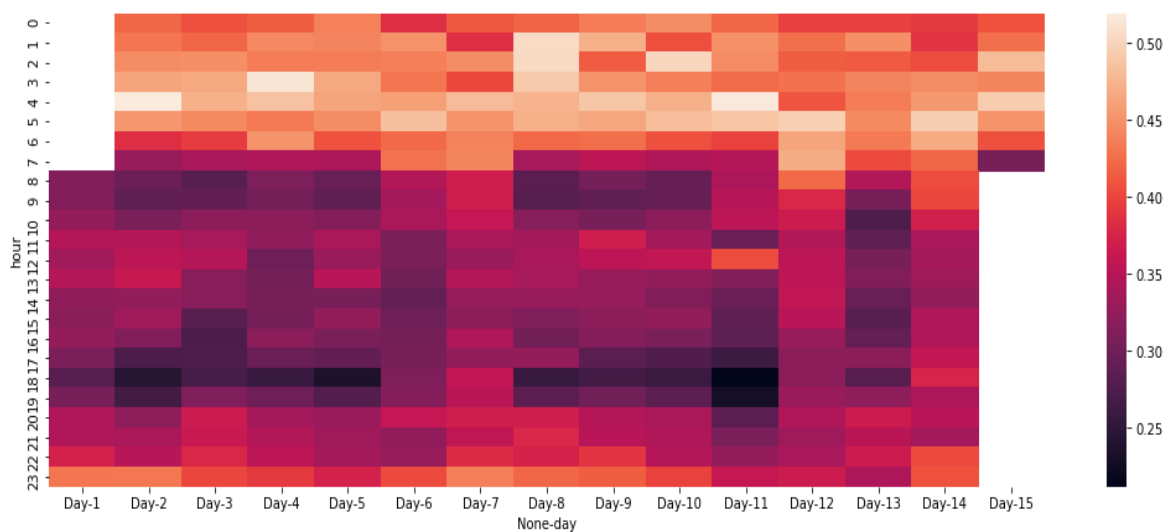
Figure 14: Day hour heatmap
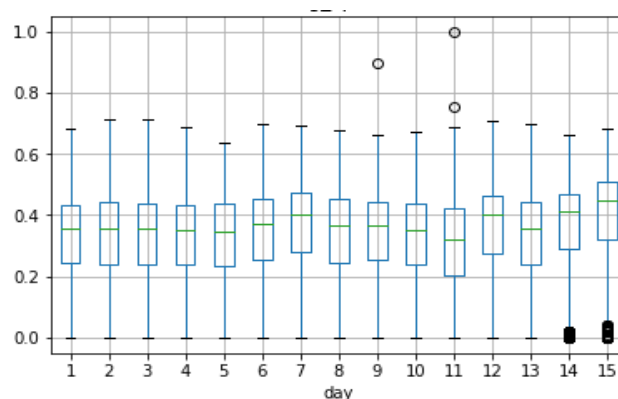
Figure 15: Box plot showing daily average speeds.

Figure 15 is a box plot showing average daily traffic flow speeds across the road network. The box plots showing many outliers for given days and big interquartile ranges in speed observations indicate that there can be extreme variations in traffic flow speeds not only in

different times of day but also total aggregated daily speeds. The variations could be as a result of changes in the road network such as road accidents, big events, or weather.

With insights given from figure (11&12), we went ahead to investigate location traffic patterns by hour. We try to observe how traffic patterns evolve across the study area by analysing spatial distributions of average speeds at midnight (00:00), mid-day (12:00) and evening (18:00) as shown in figures (1.2.3. respectively). We used the aggregate traffic speed for all the days considered in the study.

| (a) 00:00 Hour | (b)   12:00 Hour | (c) 18:00 Hour |
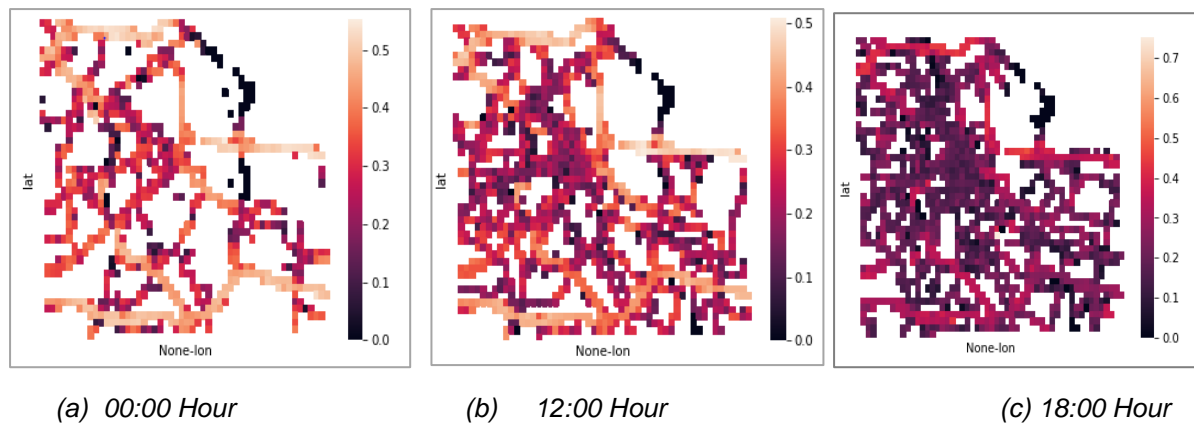
Figure 16: Traffic spatial distribution at different times.

From Figure 16 we can see that the slowest speeds are experienced in the evenings almost across the wholes study area. Congestion levels are more pronounced in the evening hours than in the afternoon hours. The slow speeds seem not to be evolving from a given point, this shows that the is some degree of spatial connectivity in the study and that people may have the same travel behaviour across locations. Though in Figure 16 a. we can see a relief in the traffic speed, we also learn that congestion levels can extend past normal commuting hours. This gives insight into how congestion evolves in urban areas and it could guide precedent traffic measures.

5.2 Traffic Modelling and Prediction

During the training process, we tried a set of hyper parameters and for each set we recorded the resulting metrics. We Increased the number of Conv_LSTM layers and Kernel size to allow more extraction of features though results after every implementation was the same. The resultant model is a 4 layered 2D Convolutional LSTM each having 32 neurons, with batch normalization after each layer. This is because it showed a lot of stability in learning the input features see table 2 .

The prediction process was to predict traffic flows at T+1 which is next 15 minutes as well T+2, T+3, T+4, which is in the next 30 minutes, 45 minutes, and 60 minutes time respectively from the first instance using 15-minute aggregated traffic data. The prediction was performed for a

time period 6:00am to 12:00pm giving us 24 timesteps. Prediction performance of the model yielded promising results as we can see in table 3, but most noticeable is that the prediction performance of the model decreases as prediction time interval increases.

Time T	MSE	RMSE
T+1 (15 minutes)	0.0189	0.1375
T+2 (30 minutes)	0.0281	0.1677
T+3 (45 Minutes)	0.0284	0.1685
T+4 (60 minutes)	0.0288	0.1687

Table 3: Prediction Results
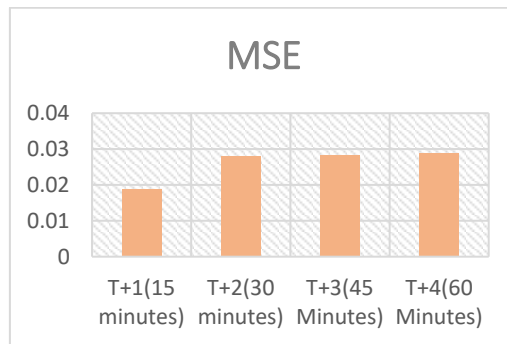
Figure 17: A Plot for MSE prediction results
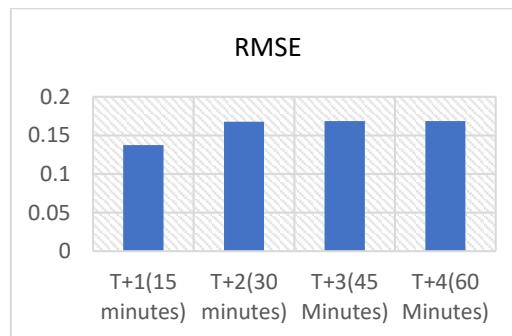
Figure 18: A Plot for RMSE prediction results

As the model performs prediction, when the time interval goes from T+1 to T+2, which is predicting the next 30 minutes, the MSE increases from 0.0189 to 0.0281, and the RMSE increases from 0.1375 to 1.677. Thereafter, we see the MSE increase in the next interval from 30 minutes to 45 minutes where the MSE increases by 0.0003 and the RMSE increases by 0.01 similar observation is observed between the 45 minutes and 60 minutes interval.
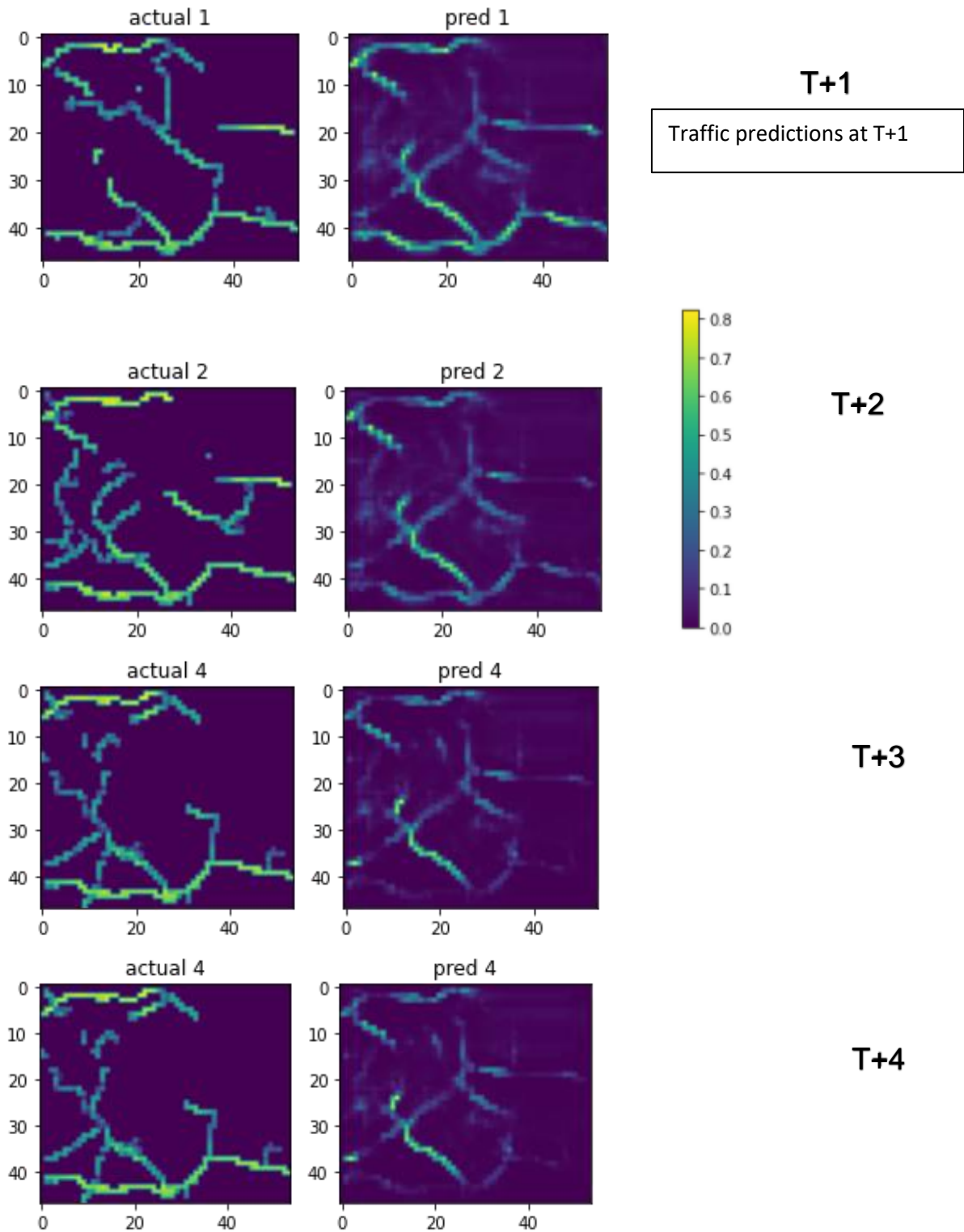
Figure 19: Actual and predicted traffic flow images

The model performs very well for the first instance (T+1). The model works by predicting T+1 future plot, and then the predicted plot is fed back into the model to predict successive plots in the next time intervals. This model method gives an accurate result at T+1 which is in this case is next 15 minutes, but the accuracy gradually reduces as it makes prediction at next

time intervals. This is due to the fact that the MSE and RMSE errors get amplified for each prediction iteration as predicted plots are re-fed into the prediction process.

Figure 19 Shows resultant traffic flow plots, all images both predicted and actual are plotted to a similar hierarchical scale color scale. we can see the predictions accuracy gradually reduce as we predict up to T+4. Looking at plot T+1, the results are promising for predicting traffic flow. Regions were we see low traffic speeds predicted are at the central part of the study area that in the preliminary findings we saw that those areas experienced low speeds especially in morning hours which in this case T is at 6:00am. The model was trained using five-day data, if the model is trained with more day data, the model would achieve even more great results especially at T+1.

5.3 Comparison with Other Models

Here we compare results obtained from the Conv_LSTM model with other models: With the Conv_LSTM, we considered both the spatial and temporal aspect, here we compare with models that predict traffic flow without accounting for both spatial and temporal characteristics.

Time T	MSE	RMSE
Conv_LSTM	0.0260	0.161
CNN	0.0289	0.168
Random Forest	0.0293	0.1712
XGboost	0.0330	0.181
Linear Regression	0.0335	0.183

Table 4: Model comparison results
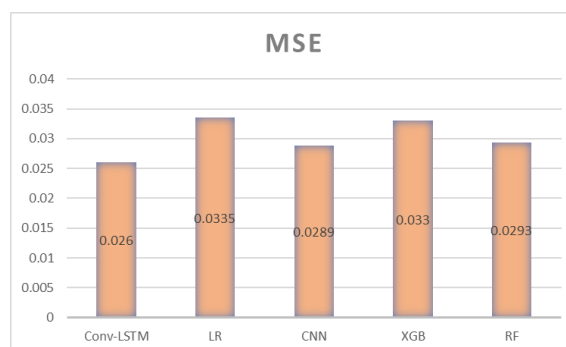
Figure 20: Comparison of models (MSE).

30
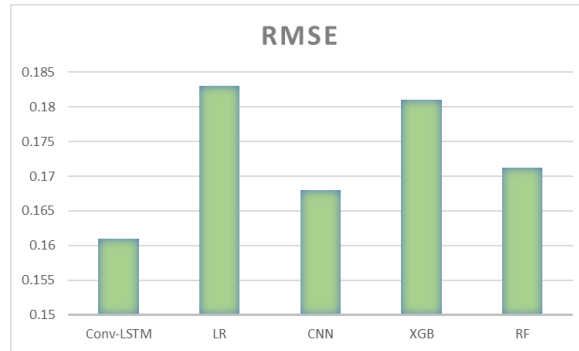
Figure 21: Comparison of models (RMSE)

Table 3 gives an overview of results from predicting 15minutes traffic flow by using other prediction algorithms. The CNN models the spatial component to some extent, and It can be seen that CNN MSE value of 0.029 is quite close to the MSE of the Conv_LSTM (0.026). The CNN too has a lower RMSE compared to other algorithms. The second-best algorithm that achieves an MSE of 0.029 is Random forest., we completely ignored the spatial component as it's really hard to work with the random forest algorithm. Linear regression and xgboost achieved the lowest accuracy with their MSE's at 0.0335 and 0.0330. respectively. We can conclude that CNN too outperformed the other parametric algorithms, and with proper hyper parameterisation it could achieve even better results. Overall, the Conv_LSTM outperformed all the comparison algorithms achieving MSE and RMSE values of 0.026 and 0.161 respectively. This shows that the inclusion of spatial dependence in traffic modelling improves model performance.

6. DISCUSSION

6.1 Discussion

To recap the research questions formulated for this study, they include; How can we quantify spatial temporal patterns embedded in trajectory traffic data? How can we develop traffic flow prediction models by considering both spatial and temporal dimensions of traffic data? How can we assess the significance of integrating traffic spatial features on model performance? The results obtained from this research provide resourceful answers to the questions.

The exploration and analysis of traffic flow spatial temporal patterns enlightened us about spatial and temporal dependencies in traffic data. Results from performing spatial autocorrelation investigation by using inclusive average speed variables of 30minutes and one hour aggregation achieved p-values less than 0.05 indicating existence of spatial dependencies in the road network. This affirms that patterns embedded in traffic data observations (trajectories) are as a result of underlaying spatial processes inherent in the road network. Analysis of temporal autocorrelation achieved statistically significant positive autocorrelation to a time lag of about 20 minutes. This means that for a given scenario were traffic congestion is increasing at a given location, there is high probability that the congestions will continue to increase for the next 5 minutes, and slightly high probability in the next 10 minutes all the way to 20 minutes. Authors in [32] studied the autocorrelation structure of traffic across day periods especially across different peak and non-peak hours and results showed that traffic patterns vary in each time periods and this is indicated by the ACF and PACF graphs were we see positive and negative statistically significant autocorrelations.

The spatial-temporal patterns observed in the analysis are dynamic and vary both in space and time across the study area and this can be attached to the daily human behaviour in making travel decisions [20]. Though the study showed that traffic patterns are almost the same during peak hours, traffic patterns during mid days slightly vary and this can be linked to the population being busy with work. It is evident in figure 16.b that the city of Singapore experiences severe congestion during evening peak hours, Singapore's transport website shows that it has put in place some traffic management systems such as road pricing as a way of discouraging private movements, this analysis can be used to further guide on how to manage or ease traffic movement. The location of road links significantly contributes to traffic volumes, depending on whether the links are located in the Central Business District, in residential areas, or just highways. The central business district will have high traffic volumes due to its characteristic high travel demand. This was also studied in [49] [60] were they discovered the relationships between traffic volume and street locations and results showed that some links majorly contributed to hourly evolving traffic patterns.

This study put focus on incorporating both the spatial and temporal dimension as it is a major research focus in current transportation studies. Also, we had a sub goal to predict city wide flow since most studies model road traffic by considering a single road as a continuous location. Traffic studies assert that traffic related data depicted by both space and time dimensions should reciprocally be considered in predicting city-wide traffic patterns . This study transformed traffic variables into matrix elements (Conv_LSTM blocks) in which the elements are space and time traffic data traits. The generated space time matrix is viewed as a channel of images in which every pixel inhibits a corresponding value in the matrix. The resultant images have M (Number of unique latitudes) pixels width and N (Number of unique longitudes) pixels height. In our case, we had 47 unique Latitudes and 54 unique longitudes. M and N form the final two dimensions for the space time matrix. Using concepts introduced in [77] in which they design model structures to learn traffic as images, we implemented one of the most recent machines learning algorithms (Conv_LSTM) [78] to model traffic flow by developing a modelling structure that takes traffic flow space -time images as inputs to perform short term traffic flow. Final model selection was majorly done by selecting the model that showed best hyperparameter combination and as well showed stability in learning the training data. Model results are promising for predicting city wide flow especially if predicting traffic flow for the next one step (T+1). We got more accurate results for T+1 (15minutes prediction) and the prediction accuracy kept reducing in attempts to predict the next 30 minutes, 45 minutes and so on. Some studies have come across the same challenges for example, in [37] were the RMSE value increased at each next interval prediction. An alternative solution to this challenge would be to reprocess the training data in a way that the model can learn all four (T+1-T+4) future plots at once. If considering a 24-hour time step (6hours), it means that we will feed T-24 to T as input data, then we predict to get results as T-19 to T+5 giving us T-24. This method would possibly alleviate the recurring prediction error problem. However, due to long modelling and training times i will could not try out the solution before the deadline.

The model was trained with only 5 days data, increasing the temporal extent for the input data to the model would provide more samples for the prediction model to learn from. This would improve prediction performance for areas with less observation trajectory data. we observed one thing that the model performs best for locations having adequate trajectory observations for each 15minute interval across the day, in other words roads frequently traversed by taxis will be highly predicted compared to those with insufficient trajectory data for some time intervals. We take an example of "Outram road" see figure 22 below.
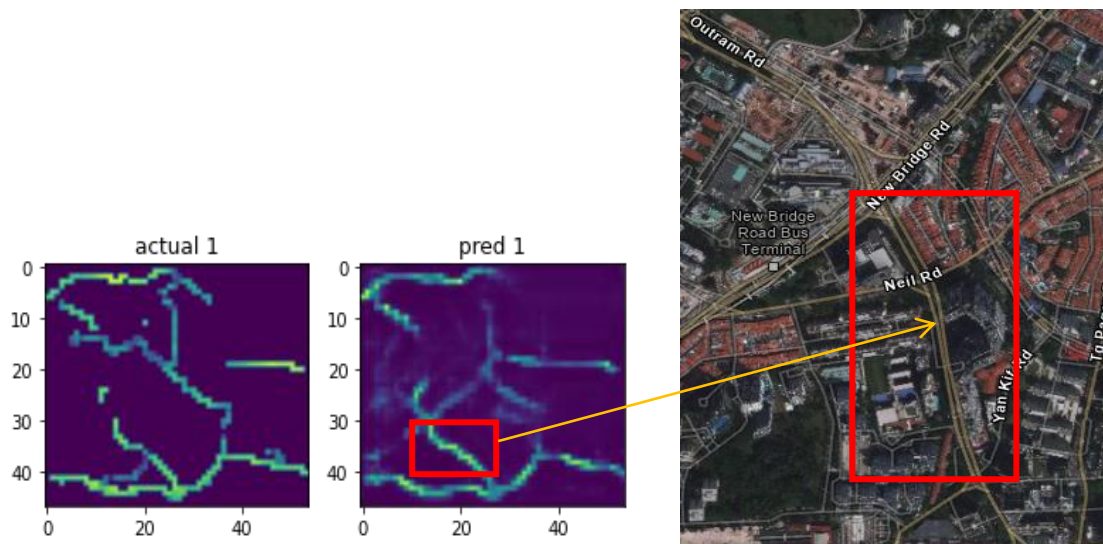
Figure 22: Frequently traversed road.

The highted road section (Outram Road), had trajectory data for almost all time intervals including late nights as well as early mornings. This shows taxis frequently use this part of the road network. For all predictions conducted T+1 to T+4, the road stands out in being a representation of both the actual and predicted traffic values. Even as the error affected T+1 to T+4, its traffic flow speed was predicted to some accuracy. Also, we can see that it has high traffic speed, it could be a major road connecting people in and out of the city centre. The city centre is predicted to have low speeds, and this is because we are predicting morning peak hours and low speeds are expected at that time. More training samples had low values for that time interval. Overall insights drawn from this scenario are that trajectory data can be used to predict Urban traffic flow and its effective in predicting frequently traversed routes. Rich analysis and modelling can be gained if we have more vehicles capturing trajectory data or if different taxi companies can come together to form a resourceful collection of trajectory data for all their cars thus achieving wide traffic data coverage. In our case, we were able to obtain data from one company as other companies do not collect trajectory data.

The level of aggregation of traffic data may affect the level of prediction result one would get. As we can see in Figures 23a-b, all maps represent the same hour but due to the level of aggregation, the traffic patterns look different. The one-hour speed map shows great coverage, this is because all other intervals are aggregated to one hour and within one hour almost all roads have been traversed by vehicles capturing trajectory data. The 30-minute aggregation flow map is almost similar to the one-hour map. And this is because more observations are collected as the hour goes on. The 15-minute map shows very little coverage. Predictions in lower timeframes may experience low performance especially for roads not frequently traversed by taxis due to low training samples.

a) *One-hour average speed Map*

b) *30-minute average speed map*

b) *15 -minute average speed map*
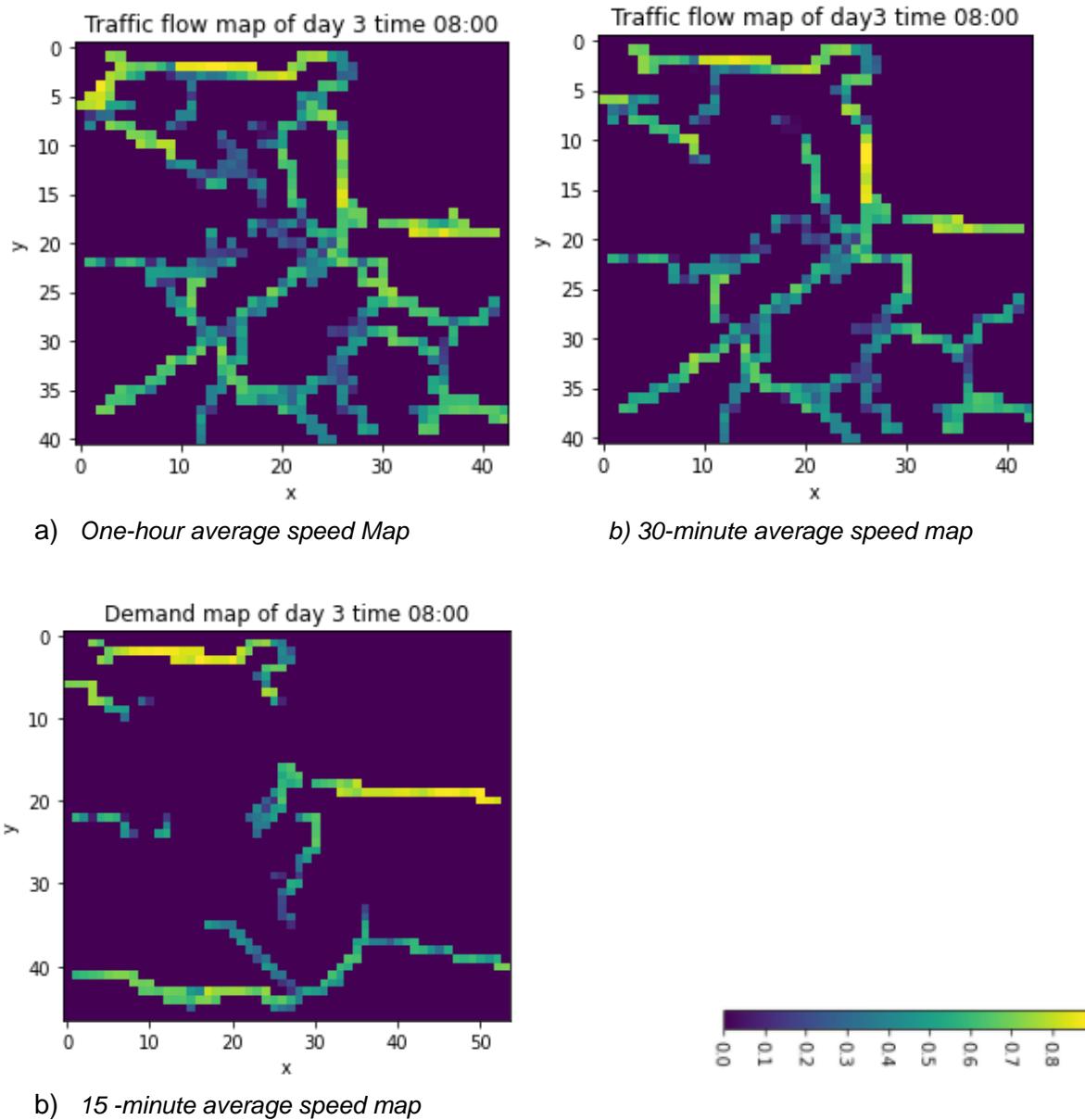
Figure 23(a-c):Traffic flow maps for different aggregation levels

We also assessed the benefit of including both spatial and temporal features in the prediction process. Figure 20, 21 and table 4 demonstrate results obtained from other prediction algorithms when applied to the traffic data in the prediction tasks. When applied on testing data, the Conv_LSTM model achieves better accuracy than other prediction algorithms, this indicates that the CONV_LSTM is more efficient in learning and adapting to new traffic data samples. A possible reason as to why LR, RF, XGB and CNN are outperformed in the prediction task is that they handle traffic flow speeds in a given link section as independent occurrences and also assume that the link speeds are self-affected. This assumption disregards inherent spatial and temporal relationships among network links thus ignoring significant mutual effects of neighbouring roads or deeper traffic flow characteristics. Some

existing deep learning frameworks, such as; RNN, and LSTM NN still remain inferior to Conv_LSTM and CNN as they cannot amicably incorporate spatial aspects from a road network perspective and yet high correlations exist in most urban congestion bottlenecks [98]. The choice of a deep learning machine learning Neural Network algorithm for our implementation approach provided possibilities for (i) simultaneously processing high-dimensional numerous time series data; (ii) handle nonlinearity instincts of the traffic data; and (iii) tacitly capturing both spatial and temporal correlations embedded in city traffic flow patterns.

A comparison of the Root Mean Squared error (RMSE) results shown in table 4 show that the Conv_LSTM model outperformed other models by a small margin i.e. Conv_LSTM = 0.161, CNN = 0.168, RF = 0.171, LR = 0.183, XGB, = 0.181. A look at results from other traffic forecasting studies that aimed at modelling both spatial and temporal characteristics indicate that deep learning models achieve better results without varying so much from traditional prediction methods. For example, in [99] they employed a Deep Convolutional Neural Network (DCNN) to forecast traffic flow and RMSE results slightly out competed baselines. The same situation is seen in [100] where their approach combines a one-dimensional Convolutional Neural Networks and Gated Recurrent Units (GRU) to capture space and time facets and the model achieved upto 11% accuracy better than the other baseline models.

The performance of machine learning models is significantly affected by insufficient training samples [62], the quality of prediction results is dependent on representative input data. For our approach, we had a good spatial coverage, but the temporal domain had gaps as some time intervals had no traffic data and so our prediction results were mainly highted by roads that are frequently traversed by taxis considered in this study. A rich trajectory dataset with a temporal span of over a month for most intervals, our model performs even better for most road sections.

One aspect in the performance of the Conv_LSTM is the training time, depending on how many epochs you set, the Conv_LSTM takes longer time, this is because of the convolutional operations that perform to capture spatial and temporal patterns embedded in traffic patterns. The CNN and Random forest regressor take more training time than linear regression and XgBoost algorithms and they achieve second - third best prediction MSE accuracy.

6.2 Limitations

One of the limitations in this project were to do with unsuccessful map-matching tasks, we managed to conduct map matching for initial attempts but as we progressed, map matching over 12 million observations kept on bringing errors and this made us drop some approaches. This is because we needed to rightly link trajectory data to segments. We opted for a grid generation approach such as that applied in [28] to enable a spatial aggregation between the road network and the data. This facilitated further data aggregations to time intervals of 15minutes, 30 minutes, 60 minutes for analysis at those time scales. Ongoing research if successful with big data map matching, can similarly implement our approach for traffic prediction though data reduction techniques should be applied to improve input training data quality.

There are several open-source trajectory datasets most of which are over 5 years. Though these datasets have significantly supported initial transport related studies, they had many gaps especially to do with low sampling rates. Present day trajectory data capturing produces more enriched trajectory data with sampling rates of up to one second, however, these datasets are not freely available. Attempts to getting more comprehensive datasets were fruitless, we were able to acquire a trajectory dataset with a high sampling rate but spanning only two weeks. More temporal coverage of traffic datasets would give better intuition of traffic patterns and traffic tasks. Also, more access of trajectories would enable testing our model on other cities to further eulogise the performance of our model . While there is a huge amount of traffic data collected daily by different parties, limitations lie in accessing this data.

7. CONCLUSION AND FUTURE WORKS

This research performed short- term traffic flow forecasting by using a modelling approach that extracts both spatial and temporal traffic patterns from the traffic data to model traffic flow. A four-layer Convolutional LSTM structure was achieved and to the urban traffic prediction problem, the Conv-LSTM provides the following potent properties: (a) convolutional operations aid the automatic extraction of spatial-temporal traffic patterns thus bypassing manual feature extraction; and (b) inputs to the Conv_LSTM represents city wide traffic information inform of high-level space time images that are employed to perform traffic flow speeds predictions for a wide road network, thus enabling prediction for various road at once instead of handling one road link at a time. Model results gave promising contribution in efforts for traffic flow prediction especially for the immediate time step (T+1). We got more accurate results for T+1 (15minutes prediction) and though prediction accuracy keeps reducing in attempts to predict the next 30 minutes, 45 minutes and so on. Overall, the model achieved the best prediction accuracy. In comparison to other models such as, CNN, RF, XGB and LR results show that the Conv_LSTM model out-performs other models and as we can see in table 4 there is accuracy promotion in both prediction accuracy, MSE and RMSE. This shows that the incorporation of both spatial and temporal characteristics significantly improves urban traffic prediction accuracies. Furthermore, the exploration of traffic patterns revealed that there are strong spatial and temporal autocorrelations in road network traffic patterns, and that these autocorrelations vary across different day time periods.

The proposed model implemented in this study can be applied to other cities though its important that the model be trained with traffic data specific to a given city as different cities have different traffic patterns. It is important to note that traffic patterns have considerable relationship with prevailing urban growth and development rates. Depending on whether urban growth is slowly or highly increasing, it would affect prevailing traffic patterns. For example, today a city may be experiencing peak movement by 9:00am, but in the next months,   peak movement may begin by 7:00am; it is important to keep on training the model depending on how traffic patterns change so as achieve accurate prediction results for a given season.

While this study tried to predict traffic flow, there is need to research more about some aspects related to our model approach. Such as:

1. The grid size chosen for this research was a 100x100m grid, grid size influences the level of aggregation. We used a 100X 100m grid to minimize the generalisation of the traffic data. There is no clear literature about appropriate grid sizes for aggregating traffic data and so more research would guide on how best to select grid size thus improving traffic prediction efforts.

2. The issue of considering network topology in traffic prediction is still a hurdle. Though some studies are trying to do it, they are not explicitly doing it. It may involve some mathematical modelling other than machine learning. We tried it to some extent though it was so complex. The integration of both mathematical modelling techniques and machine learning techniques would provide for topology analysis and modelling in traffic flow forecasts.

3. The other issue with the approach used in this study is that it does not account for road hierarchy, this may be considered to be under network topology. For example, some cities have flyover interlinked road networks and so analysing how each hierarchy performs is a great analysis for traffic improvement efforts.

4. This study has shown the potentials of modelling traffic flow at a large-scale using trajectory data, today, most studies are centred on using sensor traffic data. With sensor data, one is limited to modelling only roads that have sensors and yet there is a lot of traffic transitions taking place in road networks. With vehicle trajectory data, we are able to get wider view of overall movement at different locations. Increase in interest of using trajectories would drive more organizations to collect this data thus driving more resourceful traffic predictions.

5. Further, more future works can look into aspects of incorporating ancillary data such as (weather, holidays, events) in Using Conv_LSTM models to predict Traffic flow.

8. REFERENCES

[1]   W. He, D. Li, T. Zhang, L. An, M. Guo, and G. Chen, "Mining regular routes from GPS data for ridesharing recommendations," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 79–86, 2012.

[2]   A. Loder, L. Ambühl, M. Menendez, and K. W. Axhausen, "Understanding traffic capacity of urban networks," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019.

[3]   P. Xie, T. Li, J. Liu, S. Du, X. Yang, and J. Zhang, "Urban flow prediction from spatiotemporal data using machine learning: A survey," *Inf. Fusion*, vol. 59, no. February, pp. 1–12, 2020.

[4]   A. S. Kumarage, "URBAN TRAFFIC CONGESTION : THE PROBLEM & SOLUTIONS URBAN TRAFFIC CONGESTION : THE PROBLEM & SOLUTIONS Paper Published in the Economic Review , Sri Lanka Transportation Engineering Division ," no. January 2004, p. 10, 2004.

[5]   E. Necula, "Analyzing traffic patterns on street segments based on GPS data using R," *Transp. Res. Procedia*, vol. 10, no. July, pp. 276–285, 2015.

[6]   J.-P. Rodrigue, C. Comtois, and B. Slack, "Transportation and the spatial structure," *Geogr. Transp. Syst.*, pp. 49–94, 2019.

[7]   X. Kong, Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang, "Urban traffic congestion estimation and prediction based on floating car trajectory data," *Futur. Gener. Comput. Syst.*, vol. 61, pp. 97–107, 2016.

[8]   T. Afrin and N. Yodo, "A survey of road traffic congestion measures towards a sustainable and resilient transportation system," *Sustain.*, vol. 12, no. 11, pp. 1–23, 2020.

[9]   G. Sagl, M. Loidl, and E. Beinat, "A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic," *ISPRS Int. J. Geo-Information*, vol. 1, no. 3, pp. 256–271, 2012.

[10]   S. V. Kumar and L. Vanajakshi, "Short-term traffic flow prediction using seasonal ARIMA model with limited input data," *Eur. Transp. Res. Rev.*, vol. 7, no. 3, pp. 1–9, 2015.

[11]   T. Hasegawa, "Chapter 5-Intelligent Transport System," *Traffic Saf. Sci. - Interdiciplinary Wisdom IATSS*, p. 212, 2015.

[12]   J. Andersen and S. Sutcliffe, "Intelligent Transport Systems (ITS) - An Overview,"
        IFAC Proc. Vol., vol. 33, no. 18, pp. 99–106, 2000.

[13]   J. Kim and H. S. Mahmassani, "Spatial and temporal characterization of travel
        patterns in a traffic network using vehicle trajectories," *Transp. Res. Procedia*, vol. 9,
        pp. 164–184, 2015.

[14]   M. X. Hoang, Y. Zheng, and A. K. Singh, "FCCF: Forecasting citywide crowd flows
        based on big data," *GIS Proc. ACM Int. Symp. Adv. Geogr. Inf. Syst.*, vol. 2996934,
        2016.

[15]   A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic Flow Prediction for Road
        Transportation Networks With Limited Traffic Data," *IEEE Trans. Intell. Transp. Syst.*,
        vol. 16, no. 2, pp. 653–662, 2015.

[16]   W. Serrano, "Digital Systems in Smart City and Infrastructure: Digital as a Service,"
        Smart Cities, vol. 1, no. 1, pp. 134–153, 2018.

[17]   C. Queen and C. Albers, "Forecasting traffic flows in road networks: A graphical
        dynamic model approach.," *… Symp. Forecast. …*, no. August 2008, pp. 1–24, 2008.

[18]   P. Minev, S. Srinivasan, and P. N. Vabishchevich, *Flux formulation of parabolic
        equations with highly heterogeneous coefficients*, vol. 340. 2018.

[19]   Y. Zheng, "Trajectory data mining: An overview," *ACM Trans. Intell. Syst. Technol.*,
        vol. 6, no. 3, pp. 1–41, 2015.

[20]   T. Sobral, T. Galvão, and J. Borges, "Visualization of urban mobility data from
        intelligent transportation systems," *Sensors (Switzerland)*, vol. 19, no. 2, 2019.

[21]   G. ATLURI, A. KARPATNE, and V. KUMAR, "Spatio-temporal data mining: A survey
        of problems and methods," *arXiv*, vol. 51, no. 4, 2017.

[22]   J. Zupan, "Introduction to artificial neural network (ANN) methods: what they are and
        how to use them," *Acta Chim. Slov.*, vol. 41, no. September, pp. 327–327, 1994.

[23]   W. Qu *et al.*, "Short-Term Intersection Tra ffi c Flow Forecasting," 2020.

[24]   F. Guo, "Short-term traffic prediction under normal and abnormal conditions," *Transp.
        Res. Board 91rd Annu. Meet.*, no. July, p. 245, 2013.

[25]   O. Mohammed and J. Kianfar, "Flow Prediction : A Case Study of Interstate 64 in
        Missouri," *2018 IEEE Int. Smart Cities Conf.*, pp. 1–7, 2018.

[26]   M. Karimpour, A. Karimpour, K. Kompany, and A. Karimpour, "Integral Methods in

Science and Engineering, Volume 2," *Integr. Methods Sci. Eng. Vol. 2*, no. September, 2017.

[27]   H. Yao, X. Tang, H. Wei, G. Zheng, and Z. Li, "Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction," *33rd AAAI Conf. Artif. Intell. AAAI 2019, 31st Innov. Appl. Artif. Intell. Conf. IAAI 2019 9th AAAI Symp. Educ. Adv. Artif. Intell. EAAI 2019*, no. March, pp. 5668–5675, 2019.

[28]   P. Xie, T. Li, J. Liu, S. Du, X. Yang, and J. Zhang, "Urban flow prediction from spatiotemporal data using machine learning: A survey," *Inf. Fusion*, vol. 59, no. Tianrui Li, pp. 1–12, 2020.

[29]   R. Herring *et al.*, "Using mobile phones to forecast arterial traffic through statistical learning," *89th Transp. Res. Board Annu. Meet.*, pp. 10–2493, 2009.

[30]   Q. Wang, M. Lu, and Q. Li, "Interactive, multiscale urban-traffic pattern exploration leveraging massive gps trajectories," *Sensors (Switzerland)*, vol. 20, no. 4, pp. 1–16, 2020.

[31]   S. Du, T. Li, X. Gong, and S. J. Horng, "A hybrid method for traffic flow forecasting using multimodal deep learning," *Int. J. Comput. Intell. Syst.*, vol. 13, no. 1, pp. 85–97, 2020.

[32]   T. Cheng, J. Haworth, and J. Wang, "Spatio-temporal autocorrelation of road network data," *J. Geogr. Syst.*, vol. 14, no. 4, pp. 389–413, 2012.

[33]   S. Gao, Y. Wang, Y. Gao, and Y. Liu, "Understanding urban traffic-flow characteristics: A rethinking of betweenness centrality," *Environ. Plan. B Plan. Des.*, vol. 40, no. 1, pp. 135–153, 2013.

[34]   T. Bogaerts, A. D. Masegosa, J. S. Angarita-Zapata, E. Onieva, and P. Hellinckx, "A graph CNN-LSTM neural network for short and long-term traffic forecasting based on trajectory data," *Transp. Res. Part C Emerg. Technol.*, vol. 112, no. December 2019, pp. 62–77, 2020.

[35]   X. Cheng, R. Zhang, J. Zhou, and W. Xu, "DeepTransport: Learning Spatial-Temporal Dependency for Traffic Condition Forecasting," *arXiv*, 2017.

[36]   E. I. Vlahogianni, M. G. Karlaftis, and J. C. Golias, "Short-term traffic forecasting: Where we are and where we're going," *Transp. Res. Part C Emerg. Technol.*, vol. 43, pp. 3–19, 2014.

[37]   G. Dai, C. Ma, and X. Xu, "Short-term traffic flow prediction method for urban road

sections based on space-time analysis and GRU," *IEEE Access*, vol. 7, no. September, pp. 143025–143035, 2019.

[38]   L. Liu *et al.*, "Dynamic Spatial-Temporal Representation Learning for Traffic Flow Prediction," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–15, 2020.

[39]   A. Bull, *Traffic Congestion - The Problem and How to Deal with it?* 2004.

[40]   M. Éva, A. Konstantinos, A. Roberto, E. Antonio, E. Gianotti, and L. Iorio, *Intelligent Transport Systems for sustainable mobility.* 2007.

[41]   J. Tang, J. Zeng, Y. Wang, H. Yuan, F. Liu, and H. Huang, "Traffic flow prediction on urban road network based on License Plate Recognition data: combining attention-LSTM with Genetic Algorithm," *Transp. A Transp. Sci.*, 2020.

[42]   D. L. Gerlough and M. J. Huber, *Traffic Flow Theory: A Monograph.* 1975.

[43]   FHWA, "Traffic Flow Theory Revised. A State-of-the-Art Report (2001 revision)," *TRB Spec. Rep. 165 - Revis. Monogr. Traffic Flow Theory*, pp. 1-1,12-6, 2001.

[44]   S. Sun, J. Chen, and J. Sun, "Traffic congestion prediction based on GPS trajectory data," *Int. J. Distrib. Sens. Networks*, vol. 15, no. 5, 2019.

[45]   R. Akçelik, M. Besley, and D. Thompson, "Microsimulation and analytical methods for modelling urban traffic," *Traffic*, no. 4/2000, 2000.

[46]   J. Salotti, S. Fenet, R. Billot, N. E. El Faouzi, and C. Solnon, "Comparison of traffic forecasting methods in urban and suburban context," *Proc. - Int. Conf. Tools with Artif. Intell. ICTAI*, vol. 2018-November, pp. 846–853, 2018.

[47]   Y. Zhang, W. H. Shi, and Y. C. Liu, "Comparison of several traffic forecasting methods based on travel time index data on weekends," *J. Shanghai Jiaotong Univ.*, vol. 15, no. 2, pp. 188–193, 2010.

[48]   B. L. Smith, B. M. Williams, and R. Keith Oswald, "Comparison of parametric and nonparametric models for traffic flow forecasting," *Transp. Res. Part C Emerg. Technol.*, vol. 10, no. 4, pp. 303–321, 2002.

[49]   W. H. K. Lam, Y. F. Tang, K. S. Chan, and M. L. Tam, "Short-term hourly traffic forecasts using Hong Kong Annual Traffic Census," *Transportation (Amst).*, vol. 33, no. 3, pp. 291–310, 2006.

[50]   G. A. Gravvanis, A. I. Salamanis, and C. K. Filelis-Papadopoulos, *Advanced parametric methods for short-term traffic forecasting in the era of big data*, vol. 149.

Springer International Publishing, 2019.

[51] L. Seymour, P. J. Brockwell, and R. A. Davis, *Introduction to Time Series and Forecasting.*, vol. 92, no. 440. 1997.

[52] A. Stathopoulos and M. G. Karlaftis, "A multivariate state space approach for urban traffic flow modeling and prediction," *Transp. Res. Part C Emerg. Technol.*, vol. 11, no. 2, pp. 121–135, 2003.

[53] B. M. Williams, "Flow Prediction Evaluation of ARIMAX Modeling," no. 01, pp. 194–200.

[54] P. Kuan Hoong, "BitTorrent Network Traffic Forecasting With ARMA," *Int. J. Comput. Networks Commun.*, vol. 4, no. 4, pp. 143–156, 2012.

[55] A. Andreoni and M. N. Postorino, "a Multivariate Arima Model To Forecast," 2006.

[56] Y. Zou, X. Hua, Y. Zhang, and Y. Wang, "Hybrid short-term freeway speed prediction methods based on periodic analysis," *Can. J. Civ. Eng.*, vol. 42, no. 8, pp. 570–582, 2015.

[57] Y. Zhang, A. Haghani, and X. Zeng, "Component GARCH Models to Account for Seasonal Patterns and Uncertainties in Travel-Time Prediction," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 719–729, 2015.

[58] R. Chrobok, O. Kaumann, J. Wahle, and M. Schreckenberg, "Different methods of traffic forecast based on real data," *Eur. J. Oper. Res.*, vol. 155, no. 3, pp. 558–568, 2004.

[59] Y. Zhang and Y. Liu, "Comparison of parametric and nonparametric techniques for non-peak traffic forecasting," *World Acad. Sci. Eng. Technol.*, vol. 39, no. 3, pp. 242–248, 2009.

[60] T. Alghamdi, K. Elgazzar, M. Bayoumi, T. Sharaf, and S. Shah, "Forecasting traffic congestion using ARIMA modeling," *2019 15th Int. Wirel. Commun. Mob. Comput. Conf. IWCMC 2019*, no. June, pp. 1227–1232, 2019.

[61] M. Fouladgar, M. Parchami, R. Elmasri, and A. Ghaderi, "Scalable deep traffic flow neural networks for urban traffic congestion prediction," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2017-May, pp. 2251–2258, 2017.

[62] J. Rzeszótko and S. H. Nguyen, "Machine learning for traffic prediction," *Fundam. Informaticae*, vol. 119, no. 3–4, pp. 407–420, 2012.

[63] S. Indolia, A. K. Goswami, S. P. Mishra, and P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," *Procedia Comput. Sci.*, vol. 132, pp. 679–688, 2018.

[64] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Phys. D Nonlinear Phenom.*, vol. 404, no. March, pp. 1–43, 2020.

[65] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, 2016.

[66] H. Drucker, C. J. C. Surges, L. Kaufman, A. Smola, and V. Vapnik, "Support vector regression machines," *Adv. Neural Inf. Process. Syst.*, vol. 1, pp. 155–161, 1997.

[67] J. Ahn, E. Ko, and E. Y. Kim, "Highway traffic flow prediction using support vector regression and Bayesian classifier," *2016 Int. Conf. Big Data Smart Comput. BigComp 2016*, pp. 239–244, 2016.

[68] J. A. Bullinaria, "Recurrent Neural Networks Neural Computation : Lecture 12," pp. 1–20, 2015.

[69] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, "Traffic flow prediction using LSTM with feature enhancement," *Neurocomputing*, vol. 332, pp. 320–327, 2019.

[70] N. Ramakrishnan and T. Soni, "Network Traffic Prediction Using Recurrent Neural Networks," *Proc. - 17th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2018*, pp. 187–193, 2019.

[71] A. Ermagun and D. Levinson, "Spatiotemporal traffic forecasting: review and proposed directions," *Transp. Rev.*, vol. 38, no. 6, pp. 786–814, 2018.

[72] I. Okutani and Y. J. Stephanedes, "Dynamic prediction of traffic volume through Kalman filtering theory," *Transp. Res. Part B*, vol. 18, no. 1, pp. 1–11, 1984.

[73] "ToblersFirstLawofGeography---23." .

[74] J. Li, "A critical review of spatial predictive modeling process in environmental sciences with reproducible examples in R," *Appl. Sci.*, vol. 9, no. 10, 2019.

[75] E. Ko, J. Ahn, and E. Y. Kim, "3D markov process for traffic flow prediction in real-time," *Sensors (Switzerland)*, vol. 16, no. 2, 2016.

[76] Y. Liu, H. Zheng, X. Feng, and Z. Chen, "Short-term traffic flow prediction with Conv-LSTM," *2017 9th Int. Conf. Wirel. Commun. Signal Process. WCSP 2017 - Proc.*, vol.

2017-Janua, pp. 1–6, 2017.

[77] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang, and Y. Wang, "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction," *arXiv*, 2017.

[78] X. Shi, Z. Chen, H. Wang, D. Y. Yeung, W. K. Wong, and W. C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *Adv. Neural Inf. Process. Syst.*, vol. 2015-Janua, pp. 802–810, 2015.

[79] Y. Ma, Z. Zhang, and A. Ihler, "Multi-Lane Short-Term Traffic Forecasting with Convolutional LSTM Network," *IEEE Access*, vol. 8, pp. 34629–34643, 2020.

[80] J. Burge, M. Bonanni, M. Ihme, and L. Hu, "Convolutional LSTM neural networks for modeling wildland fire dynamics," *arXiv*, 2020.

[81] D. Wang, Y. Yang, and S. Ning, "DeepSTCL: A Deep Spatio-temporal ConvLSTM for Travel Demand Prediction," *Proc. Int. Jt. Conf. Neural Networks*, vol. 2018-July, no. 2016, pp. 1–8, 2018.

[82] J. Ke, H. Zheng, H. Yang, and X. (Michael) Chen, "Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach," *Transp. Res. Part C Emerg. Technol.*, vol. 85, pp. 591–608, 2017.

[83] T. Bogaerts, A. D. Masegosa, and J. S. Angarita-zapata, "Analysis of TRANSFOR19 forecasting competition," pp. 29–31, 2019.

[84] S. Santhi, "The Economic Development of Singapore: A Historical Perspective," no. July, 2020.

[85] K.-K. Chin, "Road Pricing – Singapore's 30 Years of Experience," *CESifo DICE Rep.*, vol. 03, no. 3, pp. 12–16, 2005.

[86] X. Huang *et al.*, "Grab-posisi: An extensive real-life GPS trajectory dataset in southeast Asia," *Proc. 3rd ACM SIGSPATIAL Int. Work. Predict. Hum. Mobility, Predict. 2019*, pp. 1–10, 2019.

[87] A.-K. Reid and N. Allum, "Learn About Time Series ACF and PACF in Stata With Data From the USDA Feed Grains Database (1876–2015)," *Learn About Time Ser. ACF PACF Stata With Data From USDA Feed Grains Database*, 2020.

[88] Z. Lin, M. Li, Z. Zheng, Y. Cheng, and C. Yuan, "Self-Attention ConvLSTM for Spatiotemporal Prediction," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 07, pp. 11531–11538, 2020.

[89] F. Pedregosa, "Hyperparameter optimization with approximate gradient," *33rd Int. Conf. Mach. Learn. ICML 2016*, vol. 2, pp. 1150–1159, 2016.

[90] F. Matthias and H. Frank, "Chapter1: Hyperparameter Optimization," pp. 3–38, 2018.

[91] A. Cutler, D. R. Cutler, and J. R. Stevens, "Ensemble Machine Learning," *Ensemble Mach. Learn.*, no. February 2014, 2012.

[92] T. M. H. Hope, "Linear regression," *Mach. Learn. Methods Appl. to Brain Disord.*, pp. 67–81, 2019.

[93] S. Malik and R. Harode, "XGBoost : A Deep Dive into Boosting," no. February, 2020.

[94] A. Patil and M. Rane, "Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition," *Smart Innov. Syst. Technol.*, vol. 195, pp. 21–30, 2021.

[95] Y. Rong, X. Zhang, X. Feng, T. K. Ho, W. Wei, and D. Xu, "Comparative analysis for traffic flow forecasting models with real-life data in Beijing," *Adv. Mech. Eng.*, vol. 7, no. 12, pp. 1–9, 2015.

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

[97] W. D. Sunindyo and A. S. M. Satria, "Traffic congestion prediction using multi-layer perceptrons and long short-term memory," *EECCIS 2020 - 2020 10th Electr. Power, Electron. Commun. Control. Informatics Semin.*, pp. 209–212, 2020.

[98] B. S. Kerner, H. Rehborn, M. Aleksic, and A. Haug, "Recognition and tracking of spatial-temporal congested traffic patterns on freeways," *Transp. Res. Part C Emerg. Technol.*, vol. 12, no. 5, pp. 369–400, 2004.

[99] S. Sun, H. Wu, and L. Xiang, "City-wide traffic flow forecasting using a deep convolutional neural network," *Sensors (Switzerland)*, vol. 20, no. 2, 2020.

[100] Y. Wang, S. Xu, and D. Feng, "A New Method for Short-term Traffic Flow Prediction Based on Multi-segments Features," *ACM Int. Conf. Proceeding Ser.*, pp. 34–38, 2020.

9. ANNEXES

9.1 Conv_LSTM 2D structure

```
Model: "sequential"

_____
Layer (type)                 Output Shape              Param #
=================================================================
conv_lst_m2d (ConvLSTM2D)    (None, None, 47, 54, 32)  38144
_____
batch_normalization (BatchNo (None, None, 47, 54, 32)  128
_____
conv_lst_m2d_1 (ConvLSTM2D)  (None, None, 47, 54, 32)  73856
_____
batch_normalization_1 (Batch (None, None, 47, 54, 32)  128
_____
conv_lst_m2d_2 (ConvLSTM2D)  (None, None, 47, 54, 32)  73856
_____
batch_normalization_2 (Batch (None, None, 47, 54, 32)  128
_____
conv_lst_m2d_3 (ConvLSTM2D)  (None, None, 47, 54, 32)  73856
_____
batch_normalization_3 (Batch (None, None, 47, 54, 32)  128
_____
conv3d (Conv3D)              (None, None, 47, 54, 1)   865
=================================================================
Total params: 261,089
Trainable params: 260,833
Non-trainable params: 256
_____
```

9.2 Correlation Matrix of traffic data variables

	day	avg_speed	latitude	longitude	normalizedTime	normalizedDayTime	Xcoord	Ycoord	dayOfWeek
day	1.000000	0.186821	0.029122	-0.002970	-0.015077	0.999993	-0.002970	0.029122	1.000000
avg_speed	0.186821	1.000000	-0.009169	0.084411	-0.185418	0.186136	0.084411	-0.009169	0.186821
latitude	0.029122	-0.009169	1.000000	0.249615	-0.053730	0.028922	0.249615	1.000000	0.029122
longitude	-0.002970	0.084411	0.249615	1.000000	-0.039633	-0.003118	1.000000	0.249615	-0.002970
normalizedTime	-0.015077	-0.185418	-0.053730	-0.039633	1.000000	-0.011334	-0.039633	-0.053730	-0.015077
normalizedDayTime	0.999993	0.186136	0.028922	-0.003118	-0.011334	1.000000	-0.003118	0.028922	0.999993
Xcoord	-0.002970	0.084411	0.249615	1.000000	-0.039633	-0.003118	1.000000	0.249615	-0.002970
Ycoord	0.029122	-0.009169	1.000000	0.249615	-0.053730	0.028922	0.249615	1.000000	0.029122
dayOfWeek	1.000000	0.186821	0.029122	-0.002970	-0.015077	0.999993	-0.002970	0.029122	1.000000