

FORECASTING BOARDING PASSENGER COUNT USING LONG SHORT-TERM MEMORY (LSTM) NETWORKS

Madhuri Patel¹
Vidisha Thakar
Sunil Padhiyar
Nitin Bathani

Received 23.07.2024.
Revised 23.12.2024.
Accepted 22.01.2025.

Keywords:

*Intelligent Transport System,
Time-Series Forecasting, LSTM*

ABSTRACT

To run smoothly, city transport systems must forecast boarding passenger flow from a station. This aids resource management, scheduling, and passenger comfort. However, forecasting abnormal passenger flow—such as sudden passenger increases or decreases—is difficult. These abnormalities are rare, thus there is little historical data to learn from. Deep learning models must handle weekdays, weekends, holidays, and timeslots. LSTM neural network model is used to solve this problem. LSTM neural networks retain long-term dependencies in sequential data, making them ideal for time series prediction. They need good training to recognize anomalous passenger flow for boarding. This uses data patterns including daily, weekly, and monthly cycles to predict. Periodic features help the model predict passenger flow.

For the same, we have gathered data received from e-ticket machines for the month of January 2022 to November 2023 pertaining to Thane Municipal Transport system. In this study we have documented four various types (i.e. huge demand to minimal demand) of boarding stations with detail analysis. The LSTM model is initially trained with hyper-parameter tuning on the entire passenger flow data, including normal and anomalous patterns related to time-slots, holidays and day of week. The model learns general trends and periodic fluctuations. December 2023 passenger hourly boarding counts have been forecasted. Such models have been evaluated using RMSE, MAE, R-Square, and necessary graphical representations. The outcome of evaluation parameters suggests that models are accurate and compatible for deployment and results of testing data is in-line with real count. In summary, the time-series models efficiently handles the difficulty of anticipating irregular passenger flow by using the capabilities of LSTM approaches, together with a clever combination strategy based on real-time fluctuations. This improves passenger boarding forecasting, which are essential for intelligent city transport systems.



¹ Corresponding author: Madhuri Patel
Email: patel.madhuri1991@gmail.com

1. INTRODUCTION

In context of public transportation system, infrastructural upgrades and service quality is always crucial parameters for administrators and various stakeholders as reported by various researchers (Toque et al. 2017; Pelletier et al. 2011; Moreira-Matias et al., 2015; Furth et al., 2006). This allows to match the capital and resource management for city transport system. For various decisions like expansion of station, planning route for particular directions, resource management are very important component of intelligent transportation systems (ITS) (Liang et al., 2019; Liyanage et al. 2020; Liyanage et al. 2019; Smith et al., 2002). ITS can give valuable reference data to administrators and travellers, allowing for more informed decision-making and helping to the development of smart cities (Zhou et al., 2013; Ma et al., 2014; Tirachini et al., 2013; Patel et al., 2022). Besides, passengers will also focuses on situation of availability of boarding from a particular station at particular time slot. Therefore, building an effective framework for modelling and accurately predicting short-term bus passenger flow is critical and necessary (Patel et al., 2024; Patel et al., 2022). By looking dynamic situation, various kind of prediction and forecasting is available in field of data science to take any decision.

Traditionally, short-term forecasting models have used a variety of statistical and machine learning (ML) techniques. These methods suggested by Guo et al. (2021), Sarker (2021), Singh et al. (2021), Mitra et al. (2022) include regression analysis, time-series models, support vector machines, artificial neural networks, Bayesian methods, gradient boosting methods, and KNN-based methods. Despite their utility, classical models have limits, principally because they cannot interpret raw datasets. Constructing an ML-based model necessitates rigorous feature engineering, which includes creating a feature extractor to convert raw data into appropriate internal representations. This transformation allows the learning subsystem to discover temporal connections in the input data. In the age of big data, feature engineering has gotten more sophisticated and labour-intensive.

Deep learning (DL) has been presented as a remedy to the problems caused by previous approaches. A typical deep learning model may accept raw input data and automatically detect relevant features at each level, considerably simplifying the feature engineering process. This skill has resulted in significant increases in traffic forecasting accuracy. Among deep learning models, the Long Short-Term Memory (LSTM) network (Hochreiter & Schmidhuber, 1997), a specific type of deep recurrent neural network (RNN), stands out. LSTMs are designed to handle dynamic temporal dependencies in time series data using dynamic feedback connections, which feed the previous step's output back into the current step's input layer. This feedback mechanism makes LSTMs ideal for representing the temporal dependencies inherent in time

series data. As a result, various LSTM-based models have been devised, which outperform classic forecasting approaches and are now widely used in traffic research (Yue & Ma, 2023; Vlahogianni et al., 2004; Wu et al., 2018, Liu et al., 2018; Cho et al., 2014; Peng et al. 2020; Nagaraj et al., 2022; Ouyang et al., 2020). However, many research focus solely on the application of LSTM to traffic forecasting, frequently overlooking the optimization process, which is critical for improving model performance. Optimization is an important phase in the deep learning process. During training, the model optimizer adjusts parameters that affect model training and output in order to approximate or achieve optimal values. This approach entails optimizing an objective function by pursuing the steepest descent direction, as defined by the negative gradient. Adaptive optimization approaches, such as adaptive moment estimation (Adam), have gained popularity due to their competitive performance and low tuning requirements (Chang et al., 2018; Sakinah et al., 2019). As a result, Adam has become the default algorithm for many deep learning systems, including traffic forecasting. When using the LSTM model for transport forecasting, Chang et al. (2019) inadequate generalization might result in greater forecast errors and compromise model stability.

In this paper, we have analysed the passenger boarding from various stations of Thane Municipal Transport services from 2022 and 2023. The public transport has more than 480 boarding points where major 18 boarding points caters more than 50% of total passenger flow. We have analyzed the pattern of boarding of passenger flow on daily basis with hourly timeslot. The effect of holidays, weekdays, and timeslot have also been identified. Based on the same, by properly tackling the issues of feature engineering and optimization in deep learning, our approach considerably improves the predictive capacities of intelligent transportation systems with LSTM model, contributing to the progress of smart cities.

2. LITERATURE SURVEY

Analyzing passenger flow patterns is critical for city transport system planning. Both descriptive and predictive statistical models are essential for understanding usage dynamics. Historically, purely statistical models were the norm, but the advent of machine learning has shifted the focus towards more sophisticated engineering solutions. This transition has been fuelled by the abundance of data generated by modern transportation systems, particularly from e-ticketing, smart card and RFID usage. The data collected from such machines is used to develop services like trip purpose identification, route choice estimation, and evidence-based transport planning, highlighting the growing interest of researchers and solution developers in this area.

Ma et al. (2019) introduced convolutional bidirectional LSTM neural networks for network-wide metro ridership prediction. Their approach leverages the power of deep learning algorithms to predict metro ridership by utilizing spatial and temporal features in parallel through two independent networks: a Convolutional Neural Network (CNN) and a Bidirectional LSTM (Bi-LSTM) network (Ma et al., 2019). The spatial features are derived from metro ridership data transformed into images, with the CNN extracting spatial features from these images and the Bi-LSTM handling time series data.

Zhao et al. (2017) proposed a method for spatio-temporal analysis of passenger travel patterns using massive data. This analysis informs policy making and public-private partnership policy by revealing both spatial (origin-destination) and temporal (passenger-trip) patterns. Their approach uses predefined rules to identify regularities and anomalies in the data. In the Shenzhen metro system, an average of 2.8 million transactions occur daily, providing a clear indication of data growth within metro systems.

Gan et al. (2020) examined mobility patterns and travel distances to explore the relationship between urban land use and average travel distances (ATDs), using Nanjing metro smart card data as a case study. Smart card data is an excellent source for point-to-point travel information, making travel patterns easily decipherable. Their analysis employed quantile regression to assess urban land use across different spatial scales (0.8 km to 5 km). The study focused on various time frames throughout the day, including both weekdays and weekends, and found spatial correlations between different time frames. These findings can be utilized for metro planning, with the analysis based solely on in-flow and out-flow using the quantile regression method.

With the ongoing advancement of data collection capacities, the significance of predictive algorithms in handling large volumes of data is growing. Multiple types of data, including cab trajectory data, cell phone signaling data, and travel application data, can now be employed for traffic forecast analysis. Scientists have investigated the utilization of deep learning methods to improve the implementation of neural network models on extensive traffic data. Methods such as recurrent neural networks, extended short-term memory networks (, and gated recurrent units (Rui et al., 2016), along with several hybrid models (Xiao et al., 2014; Zhang et al., 2020; Cui et al., 2019), have been utilized. The researchers have obtained noteworthy outcomes by utilizing substantial volumes of traffic data and commonly employed algorithms such as LSTM and GRU (Li et al., 2017; Zhizhen et al., 2022; Cui et al., 2019). These achievements encompass predictions of traffic station flow, transportation mode flow, and traffic networks across extended, intermediate, and short time periods (Yao et al., 2019; Reza et al., 2022). The transformer technique has been successfully utilized in predicting

traffic time-series data. It enhances model training by accurately capturing the spatial and temporal patterns of traffic data and efficiently handles the challenges of processing lengthy series data with interdependencies.

Overall, the literature indicates a growing trend towards utilizing advanced machine learning and deep learning techniques for analyzing and predicting passenger flow patterns. These methods offer significant improvements over traditional statistical models, providing more accurate and actionable insights for metro system planning and optimization.

3. PASSENGER FLOW FORECASTING BASED ON LSTM

Long Short-Term Memory (LSTM) networks are a more sophisticated version of Recurrent Neural Networks (RNNs). They are specifically designed to accurately record long-range connections, something standard RNNs struggle with. Various research (Liu et al., 2017; Ke et al., 2017; Kim et al., 2016; Sun et al., 2015; Sun et al., 2014; Ding et al., 2016; Li et al., 2017; Gers & Schmidhuber, 2000) have proven the use and effectiveness of Long Short-Term Memory (LSTM) models in predicting passenger flow. The defining characteristic of LSTM units is their specific memory cell architecture. Within the realm of forecasting passenger flow, this layer of memory cells assumes a pivotal role. LSTMs have the ability to incorporate temporal sequence data and station encoding into their input space, which improves their predicting accuracy.

3.1 Layered Architecture for Sequential LSTM in Passenger Flow forecasting for City Transport

LSTM units are crucial components of sophisticated neural networks. Gers & Schmidhuber (2000), Lippi et al. (2013) describes that LSTM (figure 1) comprise the forget gate, input gate, and output gate, each serving specific purposes that are vital for effectively handling and preserving information over long durations. Within the realm of passenger flow forecasting for urban transportation, these gates serve vital functions in the processing and anticipation of intricate temporal patterns.

The forget gate utilizes a sigmoid function to selectively discard information from the cell state, ensuring that only pertinent past data is preserved. In the context of predicting passenger flow, it is crucial to consider that certain past events, such as holidays or special events, may not have as much relevance in present forecasting compared to more recent trends.

The input gate regulates the influx of information into the cell state by employing the tanh activation function. This function takes in input data, such as recent passenger counts and external elements like weather or public events, and converts it into a numerical range between -1

and +1. The input gate guarantees that the network can precisely evaluate the importance of new information in relation to the past context.

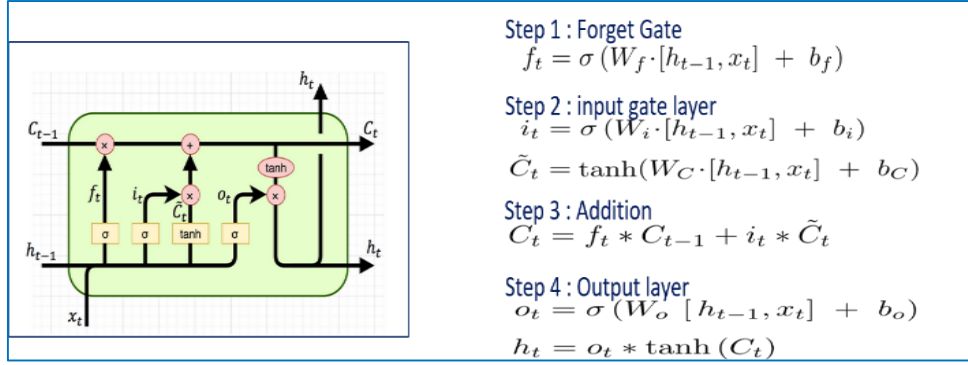


Figure 1. LSTM Architecture and the steps of working of various functions

The output gate transfers the processed information from the current cell state to the next, hence easing the sequential forecasting process. The LSTM network utilizes its distinctive memory cell architecture to handle the cell state information by means of the input gate in a sequence of critical stages that are essential for precise forecasting of passenger flow.

- 1) Sigmoid Regulation: The sigmoid function controls the values that flow through the cell by regulating the forget gate. In this stage, the data from the input vector (x_t) (which can include of current passenger counts, weather conditions, and station-specific data) and the hidden state (h_t) (which holds information from past time steps) is processed and refined. This guarantees that only relevant data has an impact on the prediction process.
- 2) tanh Function: The tanh function is used in the input gate to transform all inputs into a range of $[-1, +1]$ and generate an output vector. This stage guarantees that the incoming information is suitably adjusted in scale, allowing the network to effectively manage input of different magnitudes and dynamics.
- 3) State Update: The products of the sigmoid and tanh functions are multiplied and then added together to create the subsequent cell state. The

revised cell state integrates pertinent historical data and the latest input information, enabling the network to generate well-informed forecasts for future passenger flows.

By incorporating these mechanisms, the LSTM network can effectively capture the intricate temporal relationships seen in passenger flow data, which is frequently influenced by daily, weekly, and seasonal patterns. LSTM networks are highly suitable for this application due to their capacity to effectively process and retain long-range dependencies. This enables them to accurately learn and predict variations in passenger numbers across different time periods and stations, thereby enhancing the precision and dependability of city transport forecasts.

4. EXPERIMENTAL TECHNIQUES

The objective of this study, as depicted in figure 2, is to forecast forthcoming data based on the available passenger flow data. Our goal is to create a transformation that can effectively capture the temporal relationship between past data and make precise predictions.

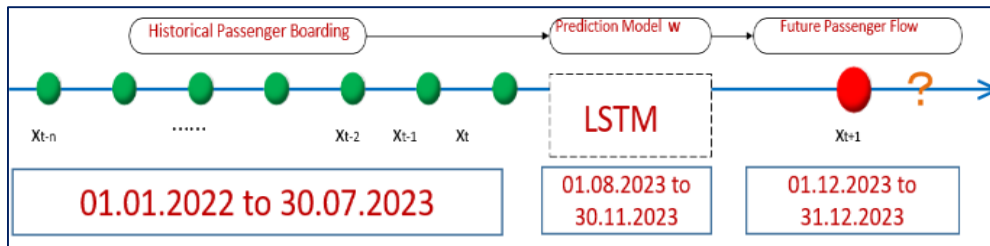


Figure 2. Problem definition of short-term passenger flow prediction.

Hence, the prediction problem may be precisely described as the given equation (1):

$$x_{t+1} = f(x_{t-k}, x_{t-k+1}, \dots, x_t, W) \quad \dots (1)$$

The prediction objective at time interval $t + 1$ is denoted as $x_{(t+1)}$. The prediction model to be developed is represented by f . The sets of historical observations are $x_{t-n}, x_{t-n+1}, \dots, x_t$. All parameters to be learned are

denoted by W . A transformation function, denoted as f , acquires knowledge of the temporal relationship (W) by analyzing past datasets and utilizes this knowledge to create predictions based on new incoming datasets.

4.2 Exploratory Data Analysis

This study presents the recorded count of passengers boarding each hour at one of the four main stations of Thane Municipal Transport (TMT) services. The ticket data utilized in this investigation is gathered from the ticket machines allocated for buses at different stations within the bus transit system. The data employed in this study include transaction time, ticket number, boarding and outgoing station codes and names, and inbound and outbound times. The study presents the boarding stations that were reported, along with their corresponding passenger contribution, as shown in Figure 3. The data pre-processing have been carried out in following steps.

- Step: 1. Preparation of dictionary which maps Marathi name to English name for source and destination and mapping of the same.
- Step: 2. Check count of Source_Name and Destination_Name and remove outliers (due to manual entry or route change)
- Step: 3. Removal of unessential attributes
- Step: 4. Fill all missing timeslots with passenger count = 0 to form an input data suitable for time series implementation
- Step: 5. Insertion of dependent features and other parameters
 - insertion of attribute month, year and weekday
 - Insertion of attribute is_holiday with value 0/1 based on Maharashtra holiday
 - Insertion of attribute Holiday_Type with value 1 to 27 based on type of holiday
 - Insertion of attribute OneHourSlot with value 1 to 24 from Date_time hour

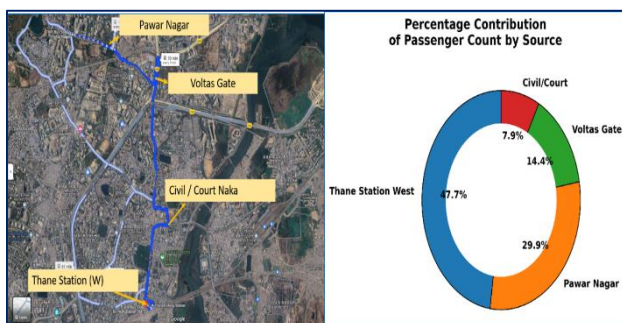


Figure 3. Boarding stations under this report and their contribution

After preprocessing, the processed passenger flow data are summarized, and a portion of the hourly passenger flow data are shown in figure 4.

	Date_Time	Source	Sid	Is_Holiday	Holiday_Type	Month	Year	Weekday	Passenger_Count	hour
0	2022-01-01 00:00:00	Civil_Court	5.0	0	0	1	2022	5	1.0	0
1	2022-01-01 01:00:00	Civil_Court	5.0	0	0	1	2022	5	1.0	1
2	2022-01-01 02:00:00	Civil_Court	5.0	0	0	1	2022	5	0.0	2
3	2022-01-01 03:00:00	Civil_Court	5.0	0	0	1	2022	5	0.0	3
4	2022-01-01 04:00:00	Civil_Court	5.0	0	0	1	2022	5	0.0	4
...	...	...	...	...	...	...	...	...	...	...
63691	2023-11-30 19:00:00	Voltas_Gate	12.0	0	0	11	2023	3	83.0	19
63692	2023-11-30 20:00:00	Voltas_Gate	12.0	0	0	11	2023	3	58.0	20
63693	2023-11-30 21:00:00	Voltas_Gate	12.0	0	0	11	2023	3	36.0	21
63694	2023-11-30 22:00:00	Voltas_Gate	12.0	0	0	11	2023	3	41.0	22
63695	2023-11-30 23:00:00	Voltas_Gate	12.0	0	0	11	2023	3	29.0	23

63696 rows × 10 columns

Figure 4. Final data table after pre-processing

For understanding of pattern of commute, the daily passenger flow for one boarding station is depicted in figure 5. The deviation of the number during weekday and weekend is also easily visible. For the some days, the decrement in passengers than usual days are due to holiday.

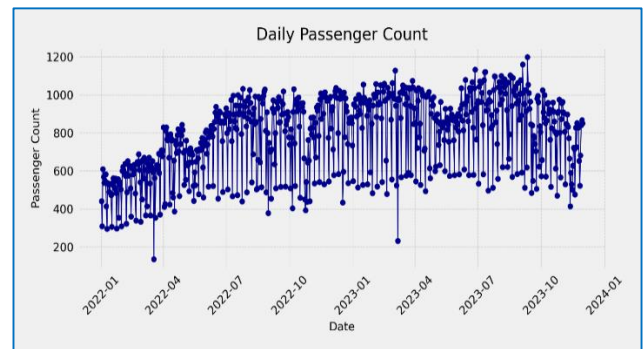


Figure 5. Daily passenger flow for one station

The details of variation in passenger flow during working day and holiday is given in figure 6, 7 and 8. The time slot wise average passenger flow shows significant changes in commute during normal working day and holiday on the same weekday. However, the effect is less on Saturday and significantly less deviation on Sunday.

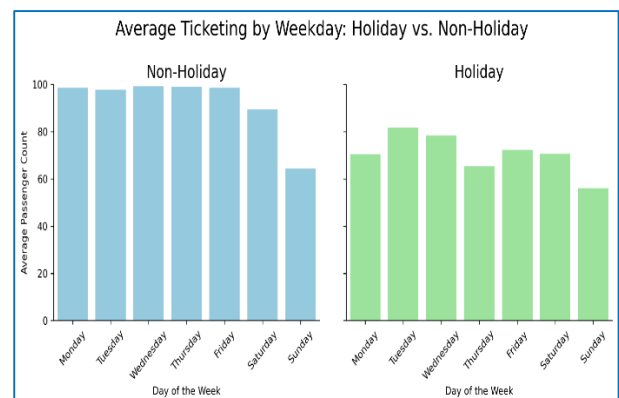


Figure 6. Relative comparison of passenger flow on normal day and holiday

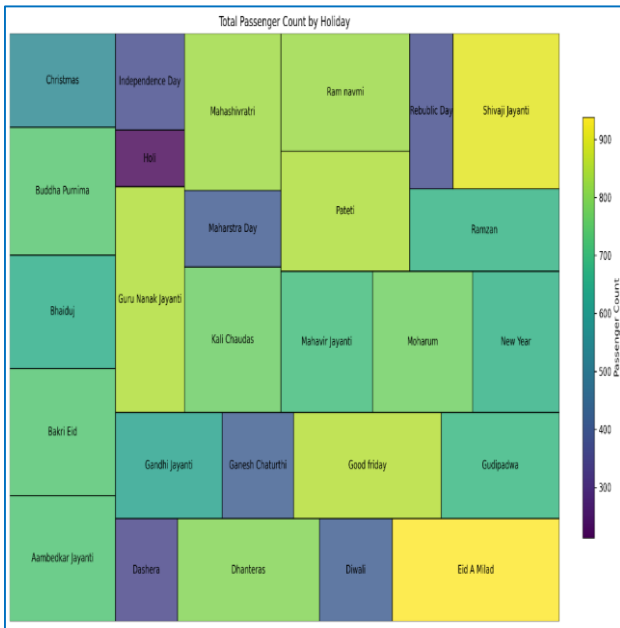


Figure 7. Heat-map of passenger flow on holiday

Even the effect of individual holiday is different due to regional religious belief. We can observe from fig (7), the holidays like Eid-e-Milad, Ambedkar Jayanti and Pateti has lower decrement in passenger boarding but the sudden decrement is observed on Holi, Maharashtra Day etc. Hence, the individual holiday is also been considered as separate feature and therefore, holiday type have been accommodated.

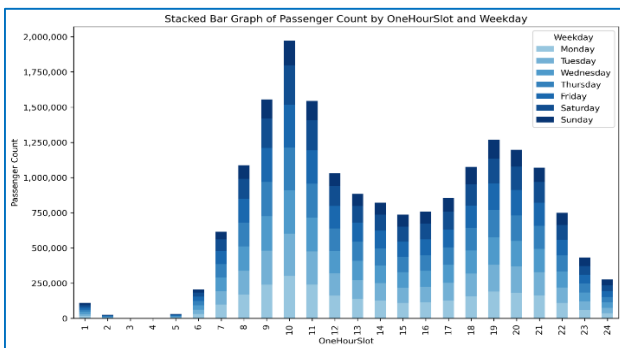


Figure 8. Hourly passenger flow for studied station on various weekday

On the same way, the passenger commute during various time slot changes on weekday and weekend wise. The peak hours are morning 9 to 11 and evening 6 to 9. The commute effect during weekday on peak hours and weekend and non-peak hours has significant divergence. The hourly slot have been assigned as feature for training of the model. With the same, the final dataset have been prepared for the model and the structure of input data is listed in table (1).

The data for boarding stations reported in this study is depicted in figure 9. The data is considered from Jan-2022 to August-2023 for training the model. The prediction have been tested from September-2023 to November-2023 and for December-2023 the same have

been predicted based on past datas et The model was implemented in Python 3.5, using Keras by Gulli & Pal (2017) and TensorFlow Abadi et al. (2015) as the deep learning packages (table 1).

Table 1. Features and its description for passenger flow

Feature	Description
Date_Time	Date of travel
Source Id	Station ID (unique)
Source	Station Name
Is_Holiday	If holiday = 1& Non-holiday=0
Holiday_Type	Assigned number to various holiday as per fig.
Onehourslot	Hour of travel (between 0 to 24)
Month	Month of travel (between 1 to 12)
Year	Year of travel (2022,2023)
Weekday	Day of Week (0 to 6)
Passenger_Count	Number of passenger boarded

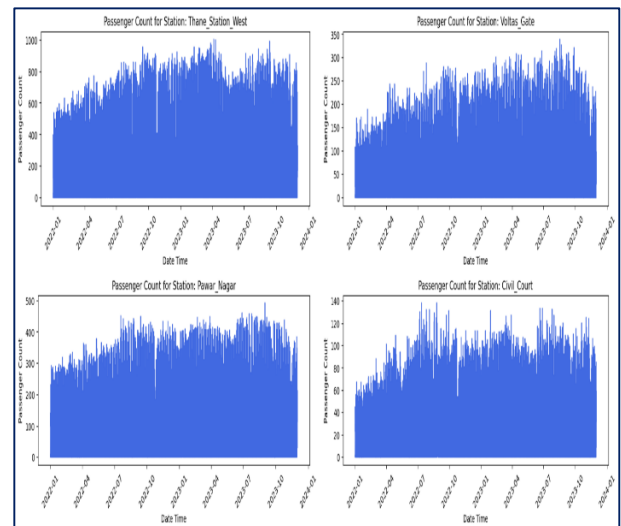


Figure 9. Passenger flow for studied station

Bharti et al. (2023) suggests that optimizing deep learning models depends critically on hyperparameter tuning. We set the time leg feature to 1,24,(7*24), which corresponds to the total number of time slices in a day, previous day, and week thereby faithfully depicting the whole daily bus passenger flow. By means of large-scale hyperparameter exploration, we have deployed adam optimizer that ensures the accurate performance. Deep learning makes extensive use of the Adam (adaptive moment estimate) optimizer. It aggregates the benefits of two other well-known optimizers: RMSProp and AdaGrad. Adam modifies the learning rate for every parameter separately in adaptive learning. This lets the optimizer manage various gradients and parameter magnitudes, hence facilitating more effective convergence. The bias correction terms assist to stabilize the early phases of the training process, so producing more consistent convergence. Adam is fit for deep learning models and large-scale machine learning challenges since it is computationally efficient and

requires minimum memory. Commonly seen in deep learning and reinforcement learning, Adam is widely applicable and suited for a range of non-stationary goals and stochastic gradient estimations. Among the most often utilized optimization techniques in the deep learning community, Adam is one of them because of these advantages. The activation functions of LSTM are basically sigmoid and Tanh. The best result is obtained with learning rate of 0.001. Conventionally, we stop the training procedure if the loss of validation dataset does not decrease after five iterations. For the same the minimum delta value is considered as 0.001. Additionally, we train our models by minimizing the mean square error over 200 epochs.

Performance of train and test dataset have been carried out between predicted number of passenger for particular station and timeslot with truth count for the same. The mean absolute error (MAE) and root mean squared error (RMSE) are chosen as the evaluation metrics. Lower values of these metrics indicate higher prediction accuracy and better model performance. The performance and adaptation of model have been evaluated based on R-Score. The definitions are provided in Equations (2)–(4).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \dots \dots (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \dots \dots (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \dots \dots (4)$$

Where y_i and y'_i represents the actual and predicted test values respectively, and n is the size of samples.

5. RESULT AND DISCUSSION

After multiple attempts, the best result is produced for all stations and the same have been further analysed. The parameters that gives best performing result have been listed in table 2.

By training the LSTM-based model with the historical data the results are exceptionally good. The value of R-score is ensuring the models are not over fitted and it gives best result with 0.912 score for high passenger station like Thane Station (W) – Satis. Compared with the overall passenger flow, lower value of MAE and RMSE suggest that predicted values matches with real passenger counts for testing periods of time. The details like dropout ratio, LSTM layers and batch size is also given.

Table 2. Evaluation parameters for various boarding stations and best parameters

Name of Station	LSTM	Dropout	RMSE	R-score	MAE
Civil_Court	226	0.4	8.86	0.87	6.27
Pawar_Nagar	470	0.1	30.0	0.89	22.1
Thane_Station_West	341	0.4	48.2	0.91	32.6
Voltas_Gate	214	0.2	21.3	0.82	14.8

The LSTM model for Civil Court station shows a strong performance with a relatively low RMSE of 8.865034 and a high R-score of 0.870947, indicating that the model explains about 87% of the variance in the passenger flow data. The MAE of 6.275776 further supports the accuracy of model in predicting hourly passenger counts. The best result is obtained with dropout rate of 0.4 helps prevent over fitting, and the batch size of 469 provides a good balance between training speed and model convergence. The model for Pawar Nagar station has a RMSE of 30.00629 that also indicates lower prediction errors with reference of actual passenger numbers. The model still explains about 90% of the variance in the data. The MAE of 22.14161 indicates substantial deviation in predictions. The lower dropout rate of 0.1 suggests the model might be prone to overfitting, and a smaller batch size of 247 was used.

Overall, the LSTM models demonstrate excellent performance across different stations, with the highest accuracy. The choice of hyperparameters, such as the number of LSTM units, dropout rate, and batch size, play a crucial role in model performance, influencing the trade-off between over fitting and generalization.

As discussed, the models are performing on the reliable basis. Fig. (10) depicts the handling capacity of model during peak time, off-peak time even handling weekends. The dates from 11.11.2023 to 14.11.2023 are Diwali holidays that have been also exceptionally handled. It is to be noted that unlike other holidays, the regional holidays are based on moon activities and the month of such holidays also changes. Based on the same, the forecasting of the passenger flow for December, 2023 have also been carried out. The same is depicted in figure 11.

Overall the accuracy of predicting passengers (figure 10, Appendix) is usually evaluated based on three types of interdependencies suggested by Zhang et al. 2019:

- Temporal dependencies: The passenger demand exhibits a pronounced periodic pattern, with high demand expected during morning and evening peaks, and low demand during late evening or early morning hours. Additionally, the short-term passenger demand is influenced by the trend of the most recent historical passenger demand. The models have dealt with

the same in very accurate manner for all boarding stations.

- Spatial dependencies: Ouyang et al. (2020) discovered that the passenger demand in a particular zone is not solely influenced by the factors within that zone, but is also influenced by all the variables in the entire network. In general, the variables in close proximity have a more pronounced impact compared to those in remote areas. This highlights the necessity for a sophisticated model that can accurately reflect local spatial relationships. The Thane West Statis is key boarding point which caters maximum passengers have also been predicted in very accurate way with LSTM.

- Exogenous dependencies: This refers to external factors that might have significant impacts on short-term passenger demand. Examples of such variables are the holidays and its effect on boarding. The exogenous variables also exhibit temporal and spatial interdependence. The model developed for all stations have predicted with very lower errors.
- The overall integration and simplest method of execution have enabled the work for real-time implementation. The work have been carried out for all boarding stations of TMT network.

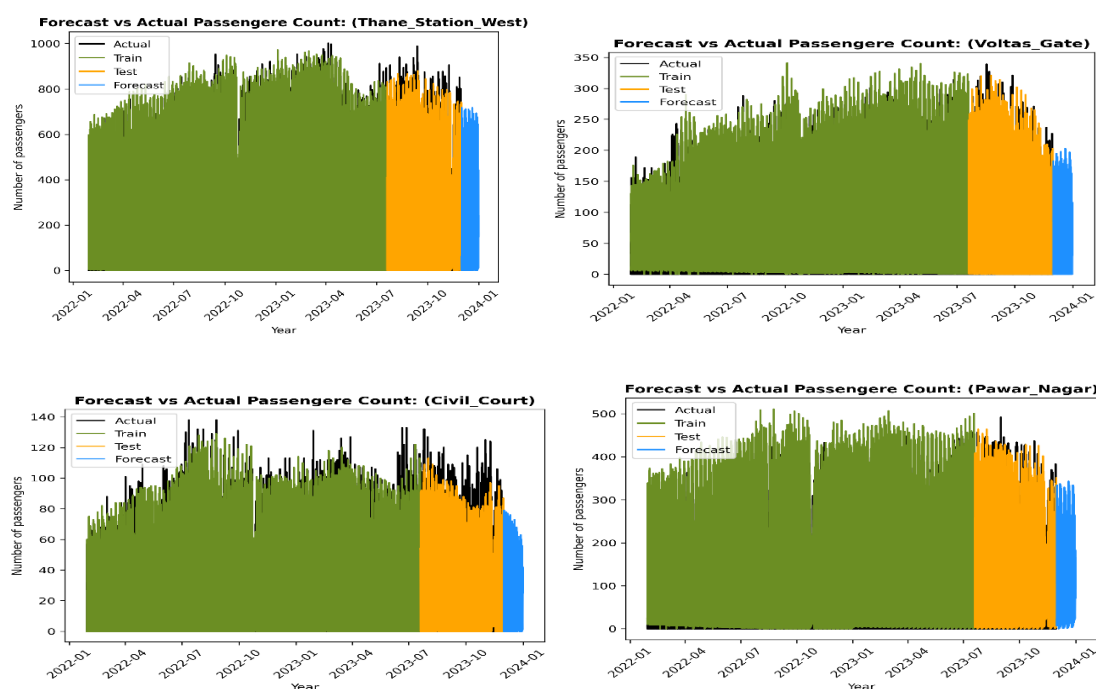


Figure 11. Training, testing and forecasting of passenger flow through LSTM

6. CONCLUSION

This work provides a deep learning method called the Long Short-Term Memory (LSTM) for predicting passenger flow boarding from particular station. The suggested architecture consists of numerous LSTM layers, and convolutional operators. It is fed with various explanatory variables, such as historical passenger flow with journey data. Factors that can affect a situation include boarding station, the specific time of day, the day of the week, and the holidays of region. Exploratory data analysis and its feature extraction have been successfully carried out. The pre-processing and data cleaning have been performed. The models have been trained with Adam

optimization and learning rate of 0.001. The results demonstrate that LSTM model is able to predict and forecast the passenger commute based on hourly basis in terms of RMSE, Rsquare, and MAE in large datas et This indicates that the proposed approach is more effective in capturing the spatio-temporal characteristics for short-term passenger boarding forecasting. It is commonly understood that enhancing the statistical characteristics of usage distribution for each station can result in improved prediction and forecasting accuracy. One important future approach for expanding the research is to incorporate additional factors such as weather, events, the existence of commercial or other facilities, etc. when modelling usage patterns.

References:

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467.

- Bharti, P., Redhu, P., & Kumar, K. (2023). Short-term traffic flow prediction based on optimized deep learning neural network: PSO-Bi-LSTM. *Physica A: Statistical Mechanics and its Applications*, 625, 129001. doi.org/10.1016/j.physa.2023.129001
- Chang, Z., Zhang, Y., & Chen, W. (2018). Effective Adam-optimized LSTM neural network for electricity price forecasting. In *2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)* (pp. 245–248). Beijing, China. doi.org/10.1109/ICSESS.2018.8663710
- Chang, Z., Zhang, Y., & Chen, W. (2019). Electricity price prediction based on hybrid model of Adam optimized LSTM neural network and wavelet transform. *Energy*, 187, 115804. doi.org/10.1016/j.energy.2019.07.134
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv*, arXiv:1406.1078.
- Cui, Z., Henrickson, K., Ke, R., & Wang, Y. (2019). Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 21, 4883–4894. doi.org/10.1109/TITS.2019.2950416
- Ding, C., Wang, D., Ma, X., & Li, H. (2016). Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. *Sustainability*, 8(11), 1100. doi.org/10.3390/su8111100
- Furth, P. G., Hemily, B., Muller, T. H., & Strathman, J. G. (2006). *Using archived AVL-APC data to improve transit performance and management*. Transportation Research Board.
- Gan, Z., Yang, M., Feng, T., & Timmermans, H. J. P. (2020). Examining the relationship between built environment and metro ridership at station-to-station level. *Transportation Research Part D: Transport and Environment*, 82, 102332. doi.org/10.1016/j.trd.2020.102332
- Gers, F. A., & Schmidhuber, J. (2000). Recurrent nets that time and count. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* (Vol. 3, pp. 189–194). Como, Italy. doi.org/10.1109/IJCNN.2000.857923
- Gulli, A., & Pal, S. (2017). *Deep learning with Keras*. Packt Publishing.
- Guo, W., Che, L., Shahidehpour, M., & Wan, X. (2021). Machine-learning based methods in short-term load forecasting. *The Electricity Journal*, 34(1), 106884. doi.org/10.1016/j.tej.2020.106884
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. doi.org/10.1162/neco.1997.9.8.1735
- Ke, J., Zheng, H., Yang, H., & Chen, X. M. (2017). Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transportation Research Part C: Emerging Technologies*, 85, 591–608. doi.org/10.1016/j.trc.2017.10.016
- Kim, S., & Shin, D. H. (2016). Forecasting short-term air passenger demand using big data from search engine queries. *Automation in Construction*, 70, 98–108. doi.org/10.1016/j.autcon.2016.06.009
- Li, L., Wang, Y., Zhong, G., Zhang, J., & Ran, B. (2017). Short-to-medium term passenger flow forecasting for metro stations using a hybrid model. *KSCE Journal of Civil Engineering*, 22, 1937–1945. doi.org/10.1007/s12205-017-1822-7
- Liang, X., Wang, G., Min, M. R., Qi, Y., & Han, Z. (2019). A deep spatio-temporal fuzzy neural network for passenger demand prediction. In *Proceedings of the 2019 SIAM International Conference on Data Mining* (pp. 100–108). SIAM.
- Lippi, M., Bertini, M., & Frasconi, P. (2013). Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning. *IEEE Transactions on Intelligent Transportation Systems*, 14(2), 871–882. doi.org/10.1109/TITS.2013.2247040
- Liu, L., & Chen, R.-C. (2017). A novel passenger flow prediction model using deep learning methods. *Transportation Research Part C: Emerging Technologies*, 84, 74–91. doi.org/10.1016/j.trc.2017.08.001
- Liu, Y., Qin, Y., Guo, J. Y., Cai, C. J., Wang, Y. G., & Jia, L. M. (2018). Short-term forecasting of rail transit passenger flow based on long short-term memory neural network. In *2018 International Conference on Intelligent Rail Transportation (ICIRT)* (pp. 1–5). Singapore. doi.org/10.1109/ICIRT.2018.8641605
- Liyanage, S., & Dia, H. (2020). An agent-based simulation approach for evaluating the performance of on-demand bus services. *Sustainability*, 12, 4117. doi.org/10.3390/su12104117
- Liyanage, S., Dia, H., Abduljabbar, R., & Bagloee, S. A. (2019). Flexible mobility on-demand: An environmental scan. *Sustainability*, 11, 1262. doi.org/10.3390/su11051262
- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54, 187–197. doi.org/10.1016/j.trc.2015.03.014

- Ma, X., Zhang, J., Du, B., Ding, C., & Sun, L. (2019). Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(6), 2278–2288. doi.org/10.1109/TITS.2018.2867042
- Ma, Z., Xing, J., Mesbah, M., & Ferreira, L. (2014). Predicting short-term bus passenger demand using a pattern hybrid approach. *Transportation Research Part C: Emerging Technologies*, 39, 148–163. doi.org/10.1016/j.trc.2013.12.008
- Mitra, A., Jain, A., & Kishore, A. (2022). A comparative study of demand forecasting models for a multi-channel retail company: A novel hybrid machine learning approach. *Operations Research Forum*, 3, 58. doi.org/10.1007/s43069-022-00166-4
- Moreira-Matias, L., Mendes-Moreira, J., Sousa, J. F., & de Gama, J. (2015). Improving mass transit operations by using AVL-based systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 16(4), 1636–1653. doi.org/10.1109/TITS.2014.2376772
- Nagaraj, N., Gururaj, H. L., & Swathi, B. H. (2022). Passenger flow prediction in bus transportation system using deep learning. *Multimedia Tools and Applications*, 81, 12519–12542. doi.org/10.1007/s11042-022-12306-3
- Ouyang, Q., Lv, Y., Ma, J., & Li, J. (2020). An LSTM-based method considering history and real-time data for passenger flow prediction. *Applied Sciences*, 10(11), 3788. doi.org/10.3390/app10113788
- Patel, M., Patel, S. B., Swain, D., & Raval, P. (2022). Visual analysis of Thane intelligent transport system. In *2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC)* (pp. 58–64). Bengaluru, India. doi.org/10.1109/IIHC55949.2022.10059995
- Patel, M., Patel, S. B., Swain, D., & Raval, P. (2024). Navigating the roadmap of intelligent transport in Indian smart cities: Sectorial analysis, challenges and key elements of data science enabled solutions. In *Smart Electric and Hybrid Vehicles: Design, Modeling, and Assessment by Industry 4.0 Approaches*. CRC Press.
- Patel, M., Patel, S. B., Swain, D., & Shah, S. (2024). Unleashing the potential of boosting techniques to optimize station-pairs passenger flow forecasting. *Procedia Computer Science*, 235, 32–44. doi.org/10.1016/j.procs.2024.04.004
- Pelletier, M.-P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557–568. doi.org/10.1016/j.trc.2010.12.003
- Peng, K., Bai, W., & Wu, L. (2020). Passenger flow forecast of railway station based on improved LSTM. In *2020 2nd International Conference on Advances in Computer Technology, Information Science and Communications (CTISC)* (pp. 166–170). Suzhou, China. doi.org/10.1109/CTISC49998.2020.00033
- Sakinah, N., Tahir, M., Badriyah, T., & Syarif, I. (2019). LSTM with Adam optimization-powered high accuracy preeclampsia classification. In *2019 International Electronics Symposium (IES)* (pp. 314–319). Surabaya, Indonesia. doi.org/10.1109/ELECSYM.2019.8901536
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2, 160. doi.org/10.1007/s42979-021-00592-x
- Singh, U., Rizwan, M., Alaraj, M., & Alsaïdan, I. (2021). A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments. *Energies*, 14(16), 5196. doi.org/10.3390/en14165196
- Smith, B. L., Williams, B. M., & Oswald, R. K. (2002). Comparison of parametric and nonparametric models for traffic flow forecasting. *Transportation Research Part C: Emerging Technologies*, 10, 303–321. doi.org/10.1016/S0968-090X(02)00009-8
- Sun, Y., Leng, B., & Guan, W. (2015). A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing*, 166, 109–121. doi.org/10.1016/j.neucom.2015.03.085
- Sun, Y., Zhang, G., & Yin, H. (2014). Passenger flow prediction of subway transfer stations based on nonparametric regression model. *Discrete Dynamics in Nature and Society*, 2014, Article 397154. doi.org/10.1155/2014/397154
- Tirachini, A., Hensher, D. A., & Rose, J. M. (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. *Transportation Research Part A: Policy and Practice*, 53, 36–52. doi.org/10.1016/j.tra.2013.06.005
- Toqué, F., Khouadjia, M., Come, E., Trepanier, M., & Oukhellou, L. (2017). Short & long term forecasting of multimodal transport passenger flows with machine learning methods. In *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 560–566). Yokohama, Japan. doi.org/10.1109/ITSC.2017.8317939
- Vlahogianni, E. I., Golias, J. C., & Karlaftis, M. G. (2004). Short-term traffic forecasting: Overview of objectives and methods. *Transport Reviews*, 24, 533–557. doi.org/10.1080/0144164042000195072
- Wu, Y., Tan, H., Qin, L., Ran, B., & Jiang, Z. (2018). A hybrid deep learning based traffic flow prediction method and its understanding. *Transportation Research Part C: Emerging Technologies*, 90, 166–180. doi.org/10.1016/j.trc.2018.03.001

- Xiao, Y., Liu, J. J., Hu, Y., Wang, Y., Lai, K. K., & Wang, S. (2014). A neuro-fuzzy combination model based on singular spectrum analysis for air transport demand forecasting. *Journal of Air Transport Management*, 39, 1–11. doi.org/10.1016/j.jairtraman.2014.03.004
- Yao, H., Tang, X., Wei, H., Zheng, G., & Li, Z. (2019). Revisiting spatial-temporal similarity: A deep learning framework for traffic prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 5668–5675). Honolulu, HI, USA. doi.org/10.1609/aaai.v33i01.33015668
- Yue, M., & Ma, S. (2023). LSTM-based transformer for transfer passenger flow forecasting between transportation integrated hubs in urban agglomeration. *Applied Sciences*, 13(1), 637. doi.org/10.3390/app13010637
- Zhang, J., Chen, F., & Guo, Y. (2020). Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit. *IET Intelligent Transport Systems*, 14, 1210–1217. doi.org/10.1049/iet-its.2020.0027
- Zhang, Z., Werner, F., Cho, H.-M., Wind, G., Platnick, S. E., Ackerman, A. S., Di Girolamo, L., Marshak, A., & Meyer, K. (2017). A framework for quantifying the impacts of sub-pixel reflectance variance and covariance on cloud optical thickness and effective radius retrievals based on the bi-spectral method. In *Radiation Processes in the Atmosphere and Ocean (IRS2016): Proceedings of the International Radiation Symposium (IRC/IAMAS)* (Vol. 1810, p. 030002). Auckland, New Zealand. doi.org/10.1063/1.4975502
- Zhao, J., Qu, Q., Zhang, F., Xu, C.-Z., & Liu, S. (2017). Spatio-temporal analysis of passenger travel patterns in massive smart card data. *IEEE Transactions on Intelligent Transportation Systems*, 18(11), 3135–3146. doi.org/10.1109/TITS.2017.2679179
- Zhou, C., Dai, P., & Li, R. (2013). The passenger demand prediction model on bus networks. In *2013 IEEE 13th International Conference on Data Mining Workshops* (pp. 1069–1076). IEEE. doi.org/10.1109/ICDMW.2013.128

Madhuri Patel

Department of Information Technology,
L.D. College of Engineering,
Navrangpura,
Ahmedabad.
India.
patel.madhuri1991@gmail.com
ORCID 0009-0005-9178-297X

Vidisha Thakar

Department of Information Technology,
L.D. College of Engineering,
Navrangpura,
Ahmedabad.
India.
vidisha.apit2018@gmail.com
ORCID 0009-0009-4148-3364

Sunil Padhiyar

Department of Rubber Technology,
L.D. College of Engineering,
Navrangpura,
Ahmedabad.
India.
sunil.rubtech@gmail.com
ORCID 0009-0006-1010-1632

Nitin Bathani

Department of Electronics and
Communication
Govt. Engineering College Modasa,
Modasa.
India.
nitinbathani@gmail.com
ORCID 0009-0006-1010-1621

Appendix:

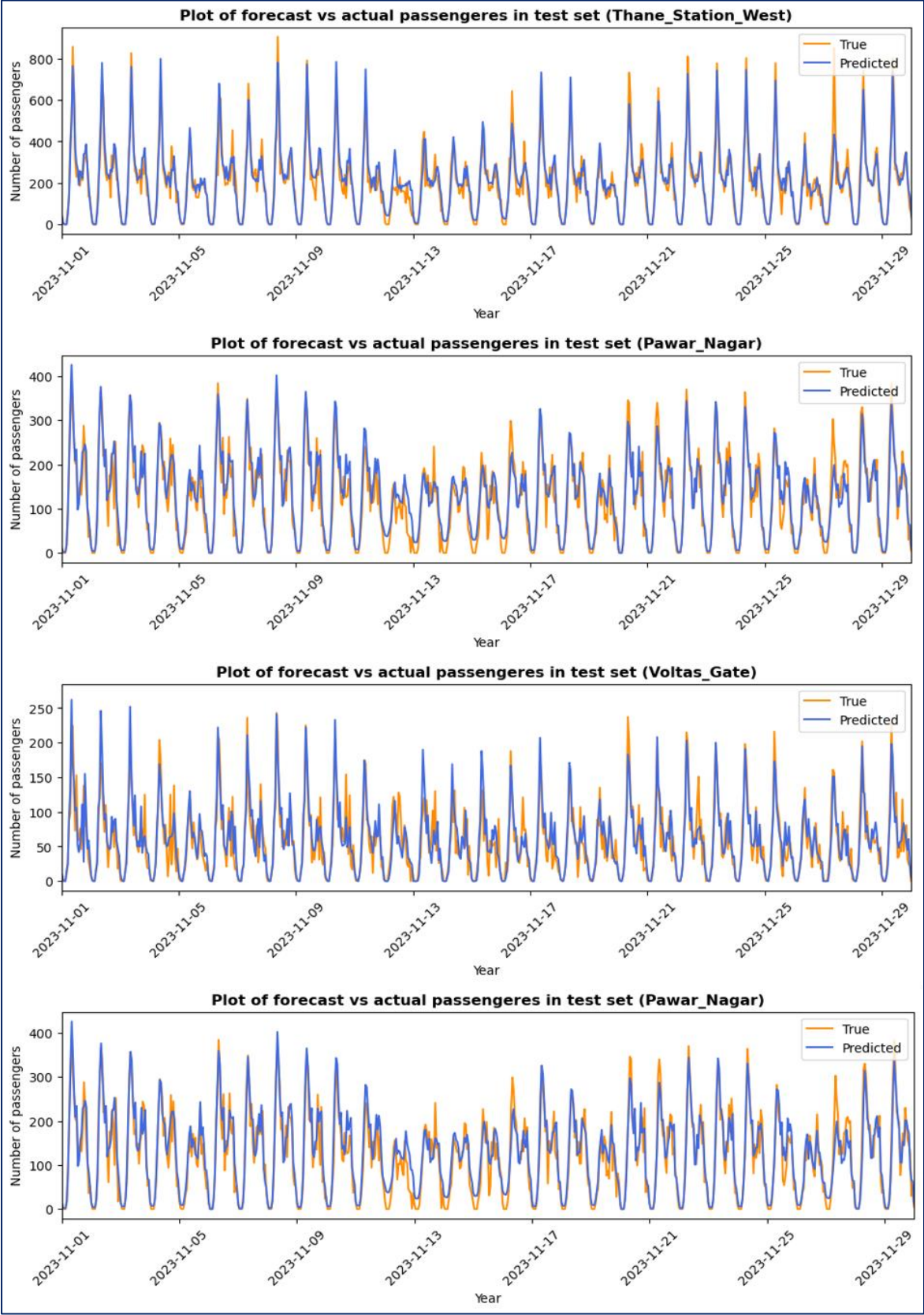


Figure 10. Predicted and actual passengers for various station