International Journal of Advanced Trends in Engineering and Management
(ISSN 2583-7052)

Vol. V, No. 05, pp. 11-19, May

Advanced Traffic Flow Prediction Using Transformer-BiLSTM Networks with Particle Swarm Optimization

Ramya R R[3]

[3]Assistant Professor, Department of Computer Science and Engineering,
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,
Avadi, Tamil Nadu, India.

ABSTRACT: Accurate Traffic Flow Prediction (TFP) is very important for Intelligent Transportation Systems (ITS), but it faces challenges of missing data, inconsistent feature values and inefficient model optimization. This paper proposes an Advanced TFP framework using Transformer-BiLSTM Networks with Particle Swarm Optimization (PSO) and evaluate it on the GTFS Traffic Prediction Dataset. The data pre-processing includes the removal of inconsistencies and null entries. The correlation analysis includes the identification of significant feature relationships. The Min–Max Scaler scales the input data to improve the stability of training. The Transformer-BiLSTM network effectively learns the global dependencies of features and temporal patterns of traffic for accurate congestion prediction. The novelty of the proposed framework is to integrate the Transformer-BiLSTM with PSO to acquire improved feature learning, parameter optimization and accurate traffic flow prediction. The network parameters are optimized using the PSO to improve convergence and predictive performance. The experimental outcomes show that proposed model attains 98% accuracy, precision, recall and F1-score, which are better than CNN (92%), LSTM (90%) and LR (92%). This makes it a reliable solution for intelligent traffic management.

Keywords: Data Pre-processing, Correlation analysis, Min-Max Scaling, Transformer BiLSTM, PSO, DL.

1. Introduction

A smart city is emerged bit by bit in modern cities. Rapid urbanization and a fast-growing population pose serious issues for transportations in modern cities. ITS is an important part of smart city, and traffic forecasting is one of its components [1-2]. Accurate forecasting of traffic is of great importance for a number of applications. For instance, forecasting traffic flow help to reduce congestion; and forecasting car-hailing demand informs sharing companies of the necessity to place their cars in places of high demand. The increase in available traffic data provides us with an opportunity to investigate the problem more deeply [3-4].

Figure 1: *Intelligent Traffic Management system*

The issue of traffic congestion in urban areas has arisen as a result of rapid urbanization and the rapid development of vehicles [5]. Consequently, urban areas experience economic losses, environmental pollution, and unaccountable loss of quality of life for their residents. It has led to the broad implementation of smart city technologies, which include Internet of Things (IoT) mechanisms, AI, and Deep Learning (DL) techniques in order to upgrade urban transportation systems. The most significant of smart city technologies is the prediction of traffic congestion [6].

The traditional DL techniques like Convolutional Neural Network (CNN) and GRU are common neural networks employed in traffic prediction solutions. CNN generates spatial features from traffic information by convolution and pooling operations, making it possible for CNN to capture spatial traffic activities. However, fails to capture long-range dependencies of time series information. GRU approaches the problem of sequence learning by having the update gate and reset gate, which enables GRU to model the sequential traffic data with less computational complexity. Yet, GRU still cannot capture long counterpart dependencies, affecting the performance of traffic prediction under complex traffic conditions [7-8]. To overcome these issues proposed Transformer–BiLSTM network that combines the global attention mechanism of the Transformer with bidirectional learning properties of BiLSTM to deliver accurate traffic prediction. The use of the optimization algorithm help to find optimal hyperparameters, achieve better convergence of the model, and improve quality of traffic flow prediction.

The Ant Colony Optimization (ACO) is an algorithm for hyperparameter optimization that imitates the foraging behaviour of ants guided by the pheromone. It achieves global exploration, but the algorithm is slow in convergence and has a high computational cost when applied to large-scale traffic data. The Genetic Algorithm (GA) also optimizes parameters using evolutionary steps of selection, crossover, and mutation in order to optimize model functionality. GA is good at conducting global exploration and maintaining diversity in its solutions but takes a lot of time to reach convergence and is impractical for real-time traffic flow prediction [9-10]. Thus, proposed in this paper PSO addresses the aforementioned difficulties and enables fast detection of optimal hyper parameters for Transformer-BiLSTM model.

1.1 Research Gap

Although there has been progress in existing traffic prediction methods, many models still struggle to capture both global feature dependencies and temporal traffic patterns, as well as to optimize parameters efficiently. This research addresses these challenges by integrating Transformer, BiLSTM and PSO into a unified framework to increase prediction accuracy, convergence speed and model reliability.

2. Recent Works

Rusul et al (2025) [11] have proposed BiLSTM for traffic prediction in short-term by learning bidirectional temporal features. But it has a high computational complexity and limited ability to capture global traffic patterns. Duan et al (2025) [12] have developed an associated road based SVR for short-term traffic flow prediction. It capture nonlinear traffic behaviours and improve accuracy of prediction. However, it requires careful parameter tuning and has high computational cost for large datasets and cannot effectively capture complex temporal dependencies. Tao et al (2025)[13] have introduced ML algorithms. The model uses historical data to predict traffic patterns, to improve traffic management. But it is data-driven, and is unable to well model complex spatial-temporal traffic dynamics. Kumar et al (2022) [14] have proposed An HBI-LSTM model using HV-ABC for predicting road traffic and choosing alternative paths. It helps to predict traffic and optimize routes. However, it has high computational complexity and longer optimization time for large-scale traffic networks. Zayed et al (2026) [15] have developed A FA optimized hybrid DL structure for Intrusion Detection (ID) in IoT locations. It enhances efficiency of ID via parameter optimization. But the complexity is large and the optimization time is longer.

- To pre-process input dataset by cleaning inconsistencies and handling missing values.

- To learn feature develop relationships using correlation analysis.
- To normalize input data using Min-Max Scaler.
- To develop Transformer-BiLSTM network for traffic flow prediction.
- To optimize the classification model using PSO.
- To improve accuracy, reliability of the proposed framework.

3. Proposed Work Explanation

The proposed framework starts with the Input Data where the collected dataset is provided for analysis. The first step involves Data Pre-processing where the data quality is enhanced by cleaning the data and dealing with missing values to ensure the dataset is consistent and complete. Then Data Visualization is done on the pre-processed data. Correlation analysis is done to analyse the relations between the different features and to find the significant pattern in the dataset

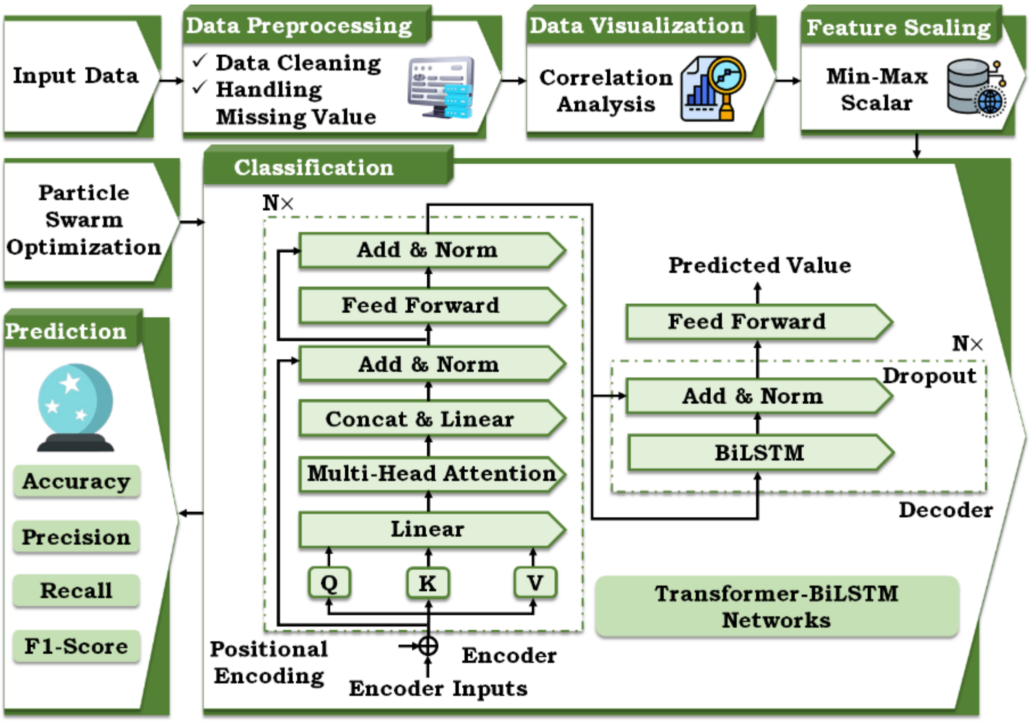

Figure 2: *Design of Proposed System*

Next, the processed data is fed to the Feature Scaling stage where Min–Max Scaler is used to normalize feature values to a common range. This normalization eliminates the differences in the magnitudes of features and assists the learning model to be more efficient and to converge better. Then the scaled features are fed to the Transformer-BiLSTM Networks for classification. The Transformer learns the global dependencies in the attention mechanism and the BiLSTM extracts the forward and backward sequential information to accurately learn the features and classification. PSO is employed to optimize the network parameters, which helps improve classification accuracy and convergence speed of the model. Finally, the optimized model generates the Prediction which guarantees reliable and accurate classification outputs.

4. Proposed System Modelling

4.1 Data processing

In the initial phase, pre-processing is applied to the input dataset to refine the data and optimize effectiveness of classification model. First, the data is cleaned to remove duplicate records, correct inconsistencies, and handle missing values. The missing values are imputed using appropriate imputation methods to complete the dataset.

4.2 Data visualization

Then for data visualization, correlation analysis is achieved to search relationships between input features. The analysis helps to minimize feature redundancy and gives insights on the dataset distribution.

4.3 Feature scaling

The selected features are then normalized using Min–Max Scaling techniques and which scales their values [0, 1], ensuring balanced feature representation and faster model convergence

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where X is original feature value and X_{norm} is normalized value respectively. Then, the normalized data is input into Transformer-BiLSTM classifier for prediction

4.4 Transformer-BiLSTM network

The proposed Transformer-BiLSTM network learns global feature dependencies and temporal traffic patterns from normalized data to predict the traffic flow. The Transformer encoder extracts high-level traffic features, with the positional encoding preserving the temporal order of the input sequence.

The positional encoding mechanism is defined as

$$PE(pos, 2i) = sin\left(\frac{pos}{10000^{2i}/_{dmodel}}\right) \quad (2)$$

$$PE(pos, 2i+1) cos\left(\frac{pos}{10000^{2i}/_{dmodel}}\right) \quad (3)$$

Here pos represents sequence position, i denotes embedding index, and d_{model} is the model measurement.

The normalized traffic data are immediately fed into the Transformer encoder to capture the global feature dependencies via multi-head self-attention. The extracted features are fed to BiLSTM layer to learn bidirectional temporal patterns to predict the traffic flow accurately.

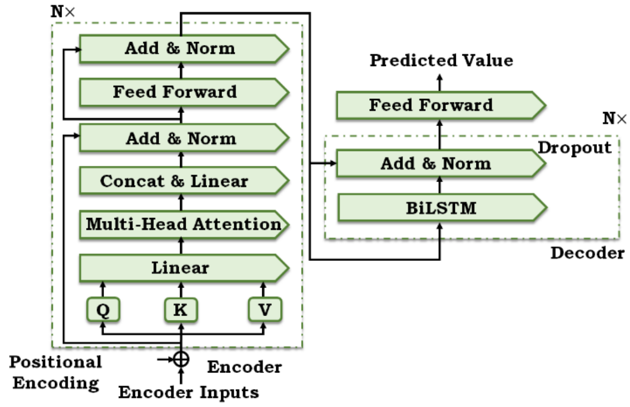

Figure 3: *Transformer-BiLSTM network*

The Transformer encoder derives informative traffic representations using the Query, Key, and Value matrices. The attention mechanism is provided as

$$Attention\ (Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

Here d_k refers to dimensional size of key vectors.

Next, the derived features are delivered through the BiLSTM model in bidirectional mode to learn temporal relationships. The hidden states are determined by the following equation

$$\vec{h}_t = f\left(W_{\vec{x}}x_t + W_{\vec{h}}\vec{h}_{t-1} + \vec{b}\right) \quad (5)$$

$$\overleftarrow{h}_t = f\left(W_{\overleftarrow{x}}x_t + W_{\overleftarrow{h}}\overleftarrow{h}_{t+1} + \overleftarrow{b}\right) \quad (6)$$

The final hidden representation is generated by concatenating both hidden states

$$y_t = [\vec{h}_t, \overleftarrow{h}_t] \quad (7)$$

Where $\vec{h}_t$ represents forward hidden state, while $\overleftarrow{h}_t$ corresponds to backward hidden states. The joint feature representation effectively capture the complex temporal traffic patterns and improve the accuracy of TFP.

4.5 PSO Optimisation

PSO is a swarm optimization technique modelled on bird flocking behaviour. In the proposed TFP framework, PSO is used to optimize the parameters of the Transformer-BiLSTM by updating each particle based on its $pbest$ and $gbest$ solutions.

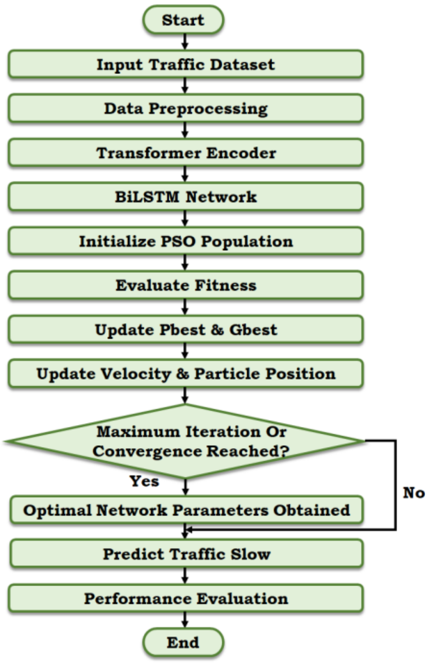

Figure 4: *Flowchart of PSO optimization*

The personal and global optimal position are represented as:

$$pbest_i = arg \min_{k=1,\dots,t} f\left(P_i(k)\right) \quad (8)$$

$$gbest = arg \min_{i=1,\dots,N} [f(pbest_i)] \quad (9)$$

In these equations, $f(\cdot)$ indicates fitness function, p_i is the position of i^{th} particle and N denotes swarm size. The particle velocity and position are computed as

$$V_i(t+1) = \omega V_i(t) + c_1 r_1\left(pbest_i - P_i(t)\right) + c_2 r_2\left(gbest - P_i(t)\right) \quad (10)$$

$$P_i(t+1) = P_i(t) + V_i(t+1) \quad (11)$$

Where V_i and P_i denote velocity and position of the i^{th} particle, and r_1 and r_2 are random numbers in the range [0,1]. PSO iteratively searches optimal network parameters to improve prediction performance of traffic flow.

5. Result and Discussion

The proposed Transformer-BiLSTM with PSO is evaluated on the GTFS Traffic Prediction Dataset. The results show that it achieves high accuracy and outperforms other methods in predicting traffic congestion.

Table 1: Dataset details and train-test splits

Target(Y)	[‘Degree of congestion’]
Train And Test Splitting	80% and 20%
Dataset Shape	66913 rows and 9 columns
Train Counts	53530
Test Counts	13383

Table 1 shows size of the dataset, the target variable and the 80:20 split train-test. It consists of 53,530 training samples and 13,383 testing samples.

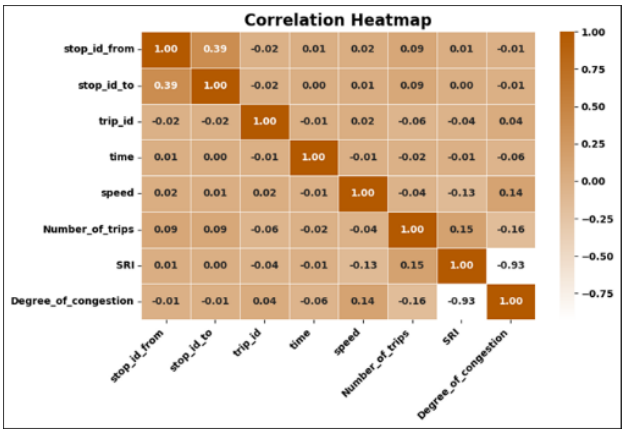

Figure 5: *Correlation heatmap of traffic flow features*

Figure 5 represents the correlation between features of traffic flow is shown in the figure. It helps to know the important feature relationships for the accurate TFP.

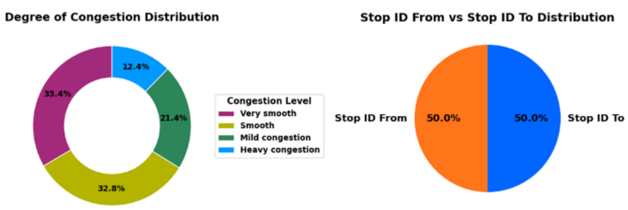

Figure 6: *Distribution of congestion levels and stop IDs*

Figure 6 shows the distribution of traffic congestion level and the ratio of Stop ID From and Stop ID To. It displays a summary of traffic conditions and data distribution per stop.

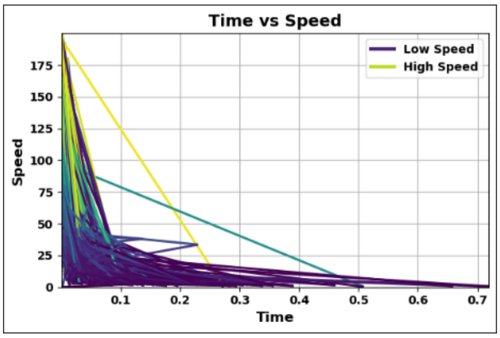

Figure 7: *Time vs speed analysis*

Figure 7 shows the variation of vehicle speed with time under different traffic conditions. It displays the speed trends used for traffic flow prediction.

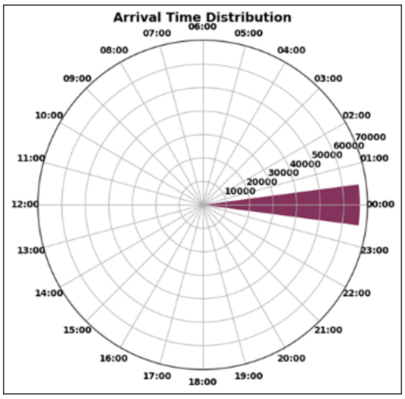

Figure 8: *Arrival time distribution of traffic data*

Figure 8 shows the distribution of vehicle arrivals in different time intervals. It shows the pattern of traffic flow over time. Used for prediction.

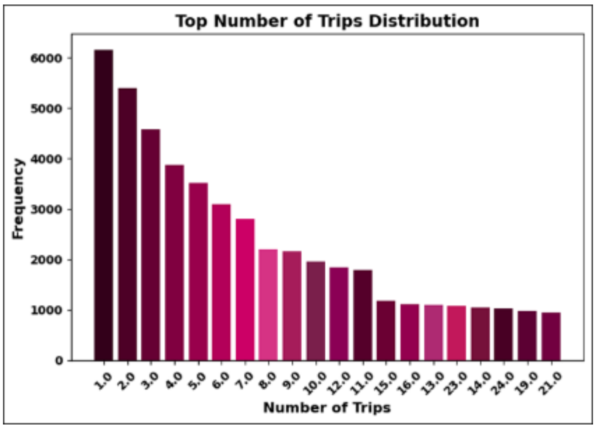

Figure 9: *Distribution of the top number of tips*

Figure 9 Frequency distribution of the most common trip counts. It concentrates on TFP analysis by travel patterns

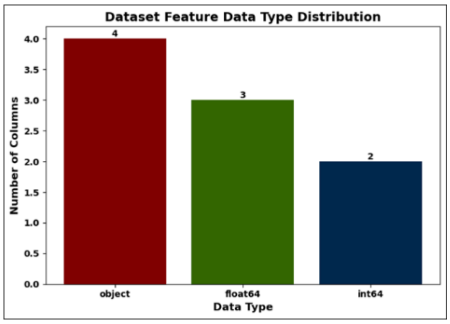

Figure 10: *Distribution of dataset feature Data type*

Figure 10 shows the distribution of feature data types in the traffic dataset. It examines the structure of the dataset used for traffic flow prediction.

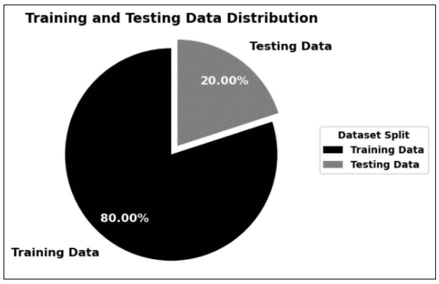

Figure 11: *Training and testing distribution*

Figure 11 represents the dataset is split into 80% training and 20% testing data. This split is used for training and evaluation of traffic flow prediction model.

Table 2: Performance evaluation metrics

Method	Value
Accuracy	98%
Precision	98%
Recall	98%
F1 score	98%

Table 2 represents that the model reaches 98% accuracy, precision, recall and F1-score, indicating stable and trustworthy results.

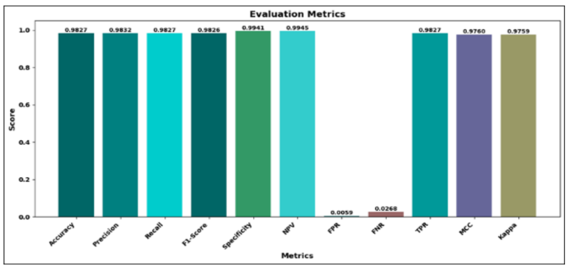

Figure 12: *Performance evaluation metrics of proposed TFP model*

Figure 12 represents performance measures of proposed model. It shows model attains high prediction accuracy and robust performance.

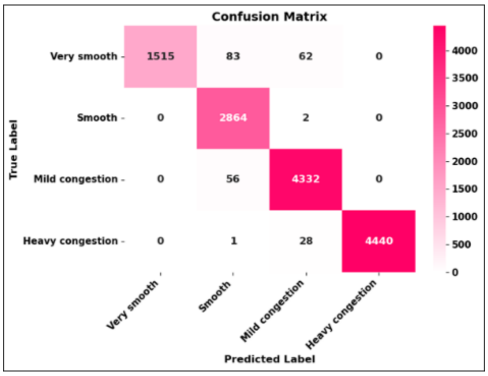

Figure 13: *Confusion matrix*

Figure 13 shows the actual and predicted traffic congestion levels. This implies that the model has a high classification accuracy with a low misclassification for all congestion classes.

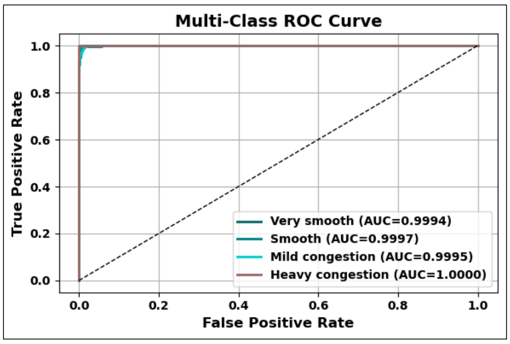

Figure 14: *Multi class ROC curve*

Figure 14 depicts the ROC curves and the AUC values for all the traffic congestion classes. It shows that the proposed model has a very good classification performance with AUC values close to 1.0.

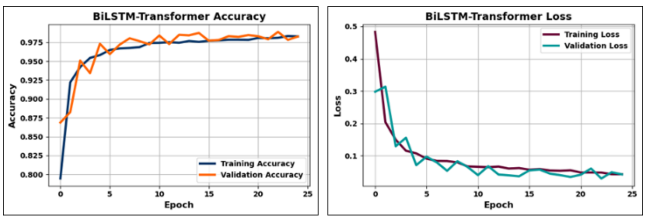

Figure 15: *Training and validation accuracy and loss*

Figure 15 represents that it reveals stable model convergence with improvement in accuracy and reduction in loss during training

Table 3: *Accuracy comparison of existing and proposed methods*

Method	Accuracy
CNN [7]	92%
LSTM [11]	90%
LR[13]	92%
Proposed	98%

Table 3 shows a comparison of accuracy of proposed model with existing methods. The proposed approach attains the highest accuracy of 98% which is better than CNN, LSTM and LR.

Table 4: *F1 Score comparison of existing and proposed methods*

Method	F1 Score
HB-LSTM [14]	96%
CNN-RNN [15]	91%
Proposed	98%

Table 4 represents proposed model achieves highest F1-score of 98% when compared with HB-LSTM (96%) and CNN-RNN (91%).

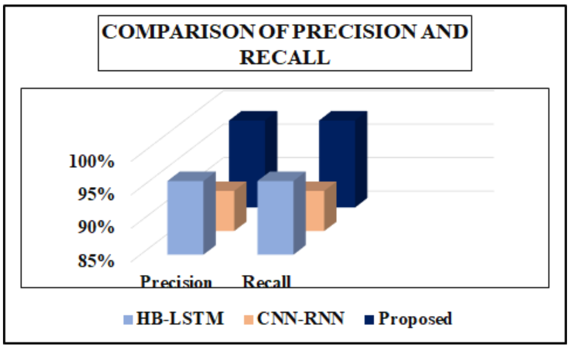

Figure 16: *Evaluation of precision and recall*

Figure 16 shows performance of HB-LSTM [14], CNN-RNN [15] and proposed models. The proposed method reaches best effectiveness among the three evaluation metrics.

6. Conclusion

This paper proposed an Advanced TFP model based on Transformer-BiLSTM Networks with PSO for accurate prediction of traffic congestion level. The Transformer successfully captured global traffic dependencies, while the BiLSTM learned temporal traffic patterns. The PSO optimized the model parameters to improve prediction performance. The proposed approach has been experimentally evaluated on a dataset of 66,913 records and demonstrated its effectiveness with 98%. The comparison results specify that the proposed model performs better than CNN, LSTM and LR methods. The confusion matrix and ROC curve with training-validation results further validated model's robustness, convergence and reliability making it a suitable solution for intelligent traffic management and smart transportation system. Future work focus on real-time traffic prediction exhausting IoT data and validating the proposed model on larger traffic datasets to improve scalability and prediction performance.

References

1. Xueyan Yin; Genze Wu; Jinze Wei; Yanming Shen; Heng Qi; Baocai Yin, Year: 2021, "Deep learning on traffic prediction: Methods, analysis, and future directions", IEEE Transactions on Intelligent Transportation Systems, Vol: 23, No: 6, pp. 4927-4943.
2. S. Narmadha; V. Vijayakumar, Year: 2023, "Spatio-Temporal vehicle traffic flow prediction using multivariate CNN and LSTM model", Materials today: proceedings, Vol: 81, pp. 826-833.
3. Fouzi Harrou; Abdelhafid Zeroual; Farid Kadri; Ying Sun, Year: 2024, "Enhancing road traffic flow prediction with improved deep learning using wavelet transforms", Results in Engineering, Vol: 23, pp. 102342.
4. Sura Mahmood Abdullah; Muthusamy Periyasamy; Nafees Ahmed Kamaludeen; S. K. Towfek; Raja Marappan; Sekar Kidambi Raju; Amal H. Alharbi; Doaa Sami Khafaga, Year: 2023, "Optimizing traffic flow in smart cities: Soft GRU-based recurrent neural networks for enhanced congestion prediction using deep learning", Sustainability, Vol: 15, No: 7, pp. 5949.
5. Noor Afiza Mat Razali; Nuraini Shamsaimon; Khairul Khalil Ishak; Suzaimah Ramli; Mohd Fahmi Mohamad Amran; Sazali Sukardi, Year: 2021, "Gap, techniques and evaluation: traffic flow prediction using machine learning and deep learning", Journal of Big Data, Vol: 8, No: 1, pp. 152.
6. Mahmuda Akhtar; Sara Moridpour, Year: 2021, "A review of traffic congestion prediction using artificial intelligence", Journal of Advanced Transportation, Vol: 2021, No: 1, pp. 8878011.
7. Mouna Zouari Mehdi; Habib M. Kammoun; Norhene Gargouri Benayed; Dorra Sellami; Alima Damak Masmoudi, Year: 2022, "Entropy-based traffic flow labeling for CNN-based traffic congestion prediction from meta-parameters", IEEE Access, Vol: 10, pp. 16123-16133.
8. Sura Mahmood Abdullah; Muthusamy Periyasamy; Nafees Ahmed Kamaludeen; S. K. Towfek; Raja Marappan; Sekar Kidambi Raju; Amal H. Alharbi; Doaa Sami Khafaga, Year: 2023, "Optimizing traffic flow in smart cities: Soft GRU-based RNN for enhanced congestion prediction using DL", Sustainability, Vol: 15, No: 7, pp. 5949.
9. Babalola Eyitemi Akilo; Samuel Abiodun Oyedotun; Godfrey Perfectson Oise; Onyemaechi Clement Nwabuokei; Nkem Belinda Unuigbokhai, Year: 2024, "Intelligent traffic management system using ant colony and deep learning algorithms for real-time traffic flow optimization", Journal of Science Research and Reviews, Vol: 1, No: 2, pp. 63-71.
10. Junxi Zhang; Shiru Qu; Zhiteng Zhang; Shaokang Cheng, Year: 2022, "Improved genetic algorithm optimized LSTM model and its application in short-term traffic flow prediction", PeerJ Computer Science, Vol: 8, pp. e1048.
11. Rusul Abduljabbar; Hussein Dia; Sohani Liyanage, Year: 2025, "Machine learning traffic flow

prediction models for smart and sustainable traffic management”, Infrastructures, Vol: 10, No: 7, pp. 155.

12. Ganglong Duan; Yutong Du; Yanying Shang; Hongquan Xue; Ruochen Zhang, Year: 2025, “Research on Support Vector Regression Short-Time Traffic Flow Prediction Model for Secondary Roads Based on Associated Road Analysis”, Appl. Sci., Vol: 15, pp. 1779.
13. Xingyu Tao; Lan Cheng; Ruihan Zhang; W. K. Chan; Huang Chao; Jing Qin, Year: 2024, “Towards Green Innovation in Smart Cities: Leveraging Traffic Flow Prediction with Machine Learning Algorithms for Sustainable Transportation Systems”, Sustainability, Vol: 16, pp. 251.
14. P. Kumar; M. Manur; A. K. Pani, Year: 2022, “Road Traffic Prediction and Optimal Alternate Path Selection Using HBI-LSTM and HV-ABC”, Indian J. Sci. Technol, Vol: 15, pp. 689-699.
15. Samar M. Zayed; Samah Alshathri; Walid El-Shafai, Year: 2026, “Firefly algorithm optimized hybrid deep learning framework for intrusion detection in IoT environments”, International Journal of Computational Intelligence Systems, Vol: 19, No: 1, pp. 158.