SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Data Governance Frameworks for Large-Scale Data Processing and Optimization

Nisha Pandey[1], Brijesh Vala[2], Jyoti Shekhawat[3], Dr. Swarna Swetha Kolaventi[4], Vinitha M[5], Hadasha Nobel Tune[6], Mukesh Rajput[7], Vishal Ambhore[8]

[1] Assistant Professor, Department of Computer Science & Engineering, Noida International University, India. Email: nisha.pandey@niu.edu.in
[2] Assistant Professor, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Parul Institute of Technology, Parul University, Vadodara, Gujarat, India. Email: Brijesh.vala@paruluniversity.ac.in
ORCID: 0000-0001-7094-7871
[3] Assistant Professor, Department of Computer Science & Engineering, Vivekananda Global University, Jaipur, India. Email: jyoti.shekhawat@vgu.ac.in ORCID: 0009-0002-9459-8818
[4] Assistant Professor, Department of uGDX School of Technology, Atlas SkillTech University, Mumbai, India. Email: swarna.kolaventi@atlasuniversity.edu.in ORCID: 0000-0001-9892-847X
[5] Assistant Professor, Department of Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: vinitham@maher.ac.in
[6] Assistant Professor, Department of Commerce, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, India. Email: nobel@maher.ac.in
[7] Assistant Professor, Department of Computer Application, Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand, India. Email: socse.mukesh@dbuu.ac.in ORCID: 0009-0006-0714-5685
[8] Assistant Professor, Department of E&TC Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra 411037, India. Email: vishal.ambhore@vit.edu

Abstract
The rapid expansion of smart city ecosystems has resulted in the generation of large-scale, heterogeneous traffic data, requiring effective data governance and optimized processing frameworks. However, existing cloud-based methods often lack robust governance mechanisms, leading to issues such as poor data quality, high latency, and unreliable traffic prediction. This research addresses these limitations by proposing a data governance framework integrated with an Artificial Butterfly Optimized Back Propagation Neural Network (ABO-BPNN) for traffic congestion prediction. The proposed framework incorporates data standardization, quality assurance, metadata management, and secure data access within a scalable architecture to ensure efficient large-scale data handling. The framework is trained on a traffic dataset collected from IoT sensors, including vehicle speed, traffic density, weather conditions, and incident-related information. Data preprocessing is performed using Min-Max normalization to remove noisy data. Feature extraction is performed with Principal Component Analysis (PCA) to reduce dimensionality and enhance significant traffic patterns. The ABO algorithm optimizes the weights of the BPNN, therebyimproving convergence and prediction capability. The framework is implemented using Python with TensorFlow, and experimental results demonstrate enhanced prediction 96.21% accuracy, 95.30% precision, 94.78% recall, and a 94.95% F1-score. The research concludes that integrating data governance with optimization-driven neural networks significantly improves large-scale traffic data processing in smart city environments.

Keywords: Data Governance, Congestion Prediction, Smart City, Artificial Butterfly Optimized Back Propagation Neural Network (ABO-BPNN).

1. Introduction

Traffic congestion remains a problem across the world, and its primary cause lies in fast urbanization, vehicle ownership, and increasing transport demand. The effects of traffic congestion include increased fuel use, high transportation expenses for both people and cargo, increased frequency of road accidents, and pollution [1]. With urban development continuing and intensifying, traffic congestion becomes not just a transport matter but also an economic and social efficiency indicator. Smart city programs around the globe are taking active steps towards developing solutions to deal with traffic congestion [2].It is observed that with the increased income level and better living standards, the use of private automobiles becomes very common in both developing and developed countries. The increased usage of automobiles puts strain on the roads, which leads to the problem of traffic congestion, which is very hard to avoid in any city environment. Therefore, the Intelligent Transportation Systems (ITS) technology attracted lots of interest in modeling and predicting the congestion pattern [3, 4]. Prediction of traffic congestion is identified as an important aspect of research that has practical applications in traffic management systems within cities. Prediction of congestion helps to plan traffic control to avoid congestion, as predicting traffic congestion is very helpful since avoiding it can be quite hard owing to unforeseen changes in traffic situations [5, 6].Besides, a proper congestion forecast is essential for saving money due to saving time, avoiding fuel costs, and increasing productivity. Moreover, congestion forecasts play a crucial role in environmental protection since they help reduce the amount of vehicle emissions due to the smoothness of traffic flows and minimal idling times. The creation of advanced congestion forecast models that are able to offer alternative routes and optimal traffic management is of high importance for smart cities of the modern era [7, 8].

Research Aim: This research focuses on the design of a highly functional data governance structure that incorporates the ABO-BPNN framework to achieve precise traffic congestion predictions in the smart city environment. This data governance approach increases the efficiency of large volumes of traffic data from IoT sensors via data standardization, data quality control, metadata management, and secure data access.

Research Organization: Section 1 is the background of the research. Section 2 highlights the review of literature. Section 3 is related to the acquisition of the dataset, its preprocessing, and feature extraction, as well as the proposed methodology. Section 4 comprises results and discussion. The performance, limitations, and future work of the research are discussed in Section 5.

2. Related Work

An approach to predicting traffic congestion using various methods to analyze video frames, such as camera calibration, defining the region of interest, and enhancing frames, was researched in [9]. They used the Logistic Regression (LR) method for traffic congestion prediction, and their performance accuracy of 78%. A smart traffic congestion management system using a fusion-based technique for controlling the speed of vehicles in vehicular networks was discussed in [10]. It is shown that the accuracy level achieved through FITCCS-VN is 95%. This system is, however, hindered by the low performance of the nodes used in the system and the lack of security in information transfer. Improvement of the prediction of traffic flow on roads via deep learning algorithms to minimize travel time and congestion was explored in [11]. Using the Long Short-Term Memory (LSTM) with multiscale wavelet filtering for traffic flow prediction, the developed system shows the 70% prediction accuracy. Nevertheless, the efficiency of the system is restricted due to the limitations of the sensors' data acquisition. The efficiency of traffic speed and flow control on road sections to reduce the amount of air pollution was analyzed in [12]. In the experiment, LSTM Network models with a feature series are used for the prediction of traffic congestion. In this method, the Mean Average Error (MAE) value is decreased by 12.90%. The problems related to smart cities, such as heavy traffic congestion and air pollution, were researched in [13]. In this case, using the Deep Deterministic Policy Gradient (DDPG) as a benchmarking model for predicting traffic congestion, and achieves high accuracy obtained from this method. The complexity of the architecture problems is overfitting and non-interpretable models, emerging. The design of an intelligent transport system using vehicular networks for solving traffic congestion problems in smart cities is analyzed in [14]. In the designed system, Internet of Vehicles (IoV) is used to forecast traffic congestion and attained very accurate results for predicting speed and traffic flows. Nevertheless, the system uses database size limitations to compress large volumes of data, hence affecting its performance. Traffic congestion prediction using analysis of travel times for buses with the objective of decreasing traffic flow in smart cities was explored in [15]. It involved the use of the Time-Based LSTM model for traffic congestion prediction, which attained 80% accuracy. EfficientNet, in combination with the EOA, was used in [16] for analyzing the optimization of ensemble learning models to optimize traffic congestion prediction for proactive traffic management. This model gave a good performance of 94.17% accuracy. However, this model still needs improvements regarding the full integration with traffic data collected by IoT sensors.

3. Methodology

Figure 1 shows the data governance process for predicting traffic congestion in a smart city environment. In the dataset, there is a vast quantity of traffic data captured by IoT sensors, which includes vehicle speed, traffic density, weather, number of vehicles, and road incidents. The Min-Max normalization technique is used in the data pre-processing stage to minimize noise. Then, the extracted features are done through PCA, and this helps in lowering the dimensionality of the data. The proposed ABO-BPNN technique is used to optimize the weights of the neural network.

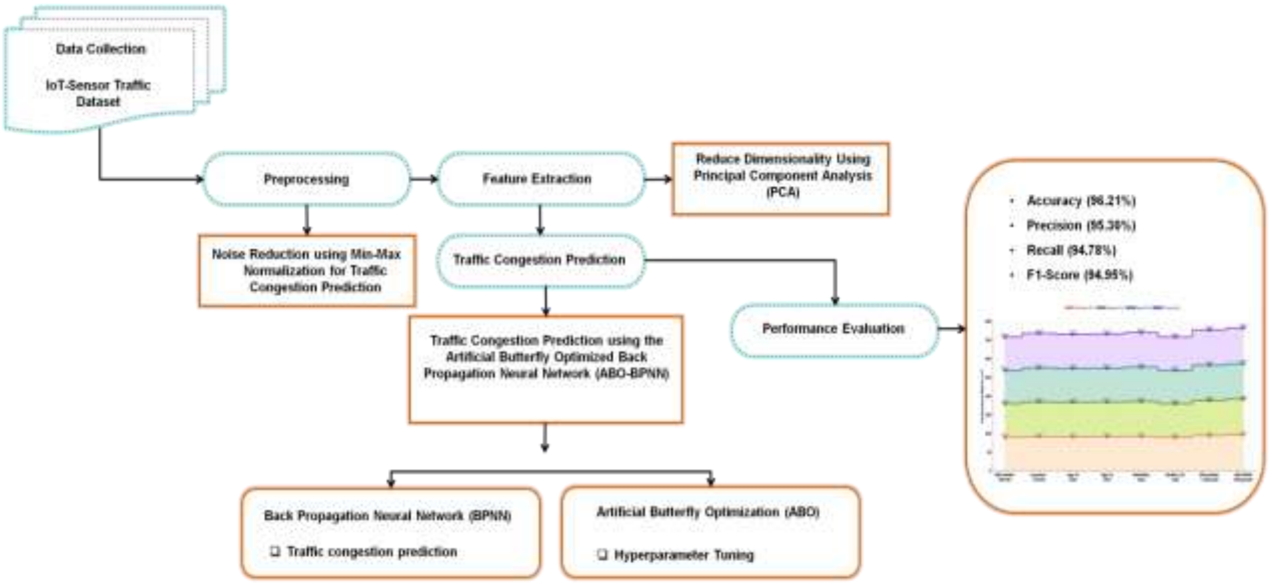


Figure 1: Methodology Workflow for Traffic Congestion Prediction in Smart City Environment

3.1 Data Collection

The process of collecting data involved the use of the Urban Traffic Flow dataset sourced by Kaggle (https://www.kaggle.com/datasets/ziya07/urban-traffic-flow-dataset). The dataset involves IoT-based traffic sensor recordings to facilitate smart cities analysis. The dataset includes 200 observations and 7 features, where numerical variables include Vehicle Count, Vehicle Speed, Congestion Level, and Target Vehicle Count. Data is divided into80% training dataand 20% testing data in developing an efficient ABO-BPNN for traffic congestion prediction in smart cities.

3.2Data preprocessing using Min-Max Normalization

Reducing noise is vital in enhancing Predicting traffic congestion in a smart city by ensuring the quality and uniformity of IoT sensor data used. In this research, the Min-Max normalization technique is employed in the preprocessing of traffic features, including traffic speed, density, and congestion. It involves converting raw data into a normalized form expressed in Equation (1).

$$Y_{new} = \frac{Y - Y_{min}}{Y_{max} - Y_{min}} \qquad (1)$$

Where Y_{new} is the normalized value of the input traffic attribute scaled within a consistent range (usually 0 to 1), Y refers to the raw value of the input traffic attribute obtained from IoT devices like speeds, densities, and congestions for remove noise traffic data, $Y_{min}$and Y_{max}refers to the minimum and maximum values observed for that particular traffic attribute in the dataset, which helps with lower-bound normalization.

3.3Reduce Dimensionality Using Principal Component Analysis (PCA)

PCA based feature extraction is essential in the proposed smart city traffic congestion prediction framework. Initially, the high-dimensional IoT traffic dataset is standardized to normalize variables such as vehicle speed, congestion level, and environmental conditions. PCA then transforms correlated factors into uncorrelated principal components while maintaining the highest possible variance. The greatest variance is captured by the first main component, as represented in Equation (2).

$$X_{ji} = \frac{d_{ji}-\overline{d_i}}{\sigma_i} \quad (2)$$

Where, X_{ji} stands for the normalized version of the j^{th} data point corresponding to the i^{th} traffic characteristic, including the characteristics such as vehicle speed, which enables standardized scaling of the data for dimensionality reduction, while d_{ji} represents the raw version of the i^{th} traffic characteristic corresponding to the j^{th} data point obtained from the IoT sensors. $\overline{d_i}$ stands for the average value of the i^{th} traffic characteristic taken over all the traffic data points, and σ_i stands for the standard deviation of the i^{th} traffic characteristic, which enables normalization of the traffic data to facilitate accurate prediction of traffic congestion.

3.4Traffic Congestion Prediction using Artificial Butterfly Optimized Back Propagation Neural Network (ABO-BPNN) Model

ABO-BPNN for the Traffic Congestion Prediction system uses the ABO method to achieve high accuracy and fast convergence. By using this framework, the weights and biases of neural networks can be optimized in order to minimize the training process and avoid getting trapped in local minima. The framework is able to analyze traffic trends on a large scale, ensure accurate predictions for future traffic, and enable intelligent management of transport systems in smart cities.

Back Propagation Neural Network (BPNN) for traffic congestion prediction: The BPNN is extensively used for predicting traffic congestion due to the capacity of modeling non-linear relationships between different traffic factors and conditions. Within the proposed architecture, the BPNN operates as the main prediction framework that is trained on the basis of traffic information gathered by means of IoT. The training procedure follows a least mean squared error (LMSE) minimization strategy, where output prediction errors are propagated backward through the network to iteratively adjust the connection weights. Initially, random weight values are assigned before training begins. The neural network architecture comprises input, hidden, and output layers, where hidden layer computation is expressed in Equation (3).

$$net_i = \sum_{j=0}^{m} u_{ji}, w_j, i = 1, 2, \dots, n \quad (3)$$

Where net_i represents the total of the weights associated with inputs for learning the traffic data pattern from IoT information like speed, density, weather, vehicle count, and accidents. $\sum_{j=0}^{m} u_{ji}$ refers to the weight associated with the j^{th} traffic input neuron and the i^{th} hidden neuron, signifying the effect of each of the traffic data points towards predicting congestion. w_j denotes the value of the traffic data point as an input collected from IoT devices. nrepresents the Number of input neurons as per the number of traffic parameters in the dataset, which is expressed in Equation (4).

$$z_i = e_G(net_i), \; i = 1, 2, \dots, n \quad (4)$$

z_i represents the output produced by the i^{th} hidden neuron, indicating traffic feature transformation for predicting congestion. $e_G$depicts the nonlinear activation function used for calculating the nonlinear relationship of traffic congestion, which is expressed in Equation (5).

$$e_G(w) = \frac{1}{1+\exp(-w)} \quad (5)$$

Where $e_G(w)$ represents the sigmoid activation function output to be applied to normalizing neuron behavior in traffic congestion simulation. w refers to the sum of input weights indicating the overall impact of traffic variables. $\exp(-w)$ denotes the exponential factor dictating the non-linear mapping of traffic congestion patterns. The congestion is predicted by aggregating the outputs from the hidden layer as expressed in Equation (6).

$$P = e_p(\sum_{i=0}^{n} \omega_{il} z_i) \quad (6)$$

P represents the predicted traffic congestion level utilized in making decisions in smart cities, $e_p$is the output activation function which transforms the output of the neural network into the predicted traffic congestion level, and $\sum_{i=0}^{n} \omega_{il}$ denotes the connection weight between the i^{th} hidden neuron and output neuron, which determines the significance of traffic information learned by the network. z_i represents the output of the hidden neurons containing information about the learned traffic. n stands for the number of hidden neurons required for modeling the traffic congestion level.

Hyperparameter Tuning with Artificial Butterfly Optimization (ABO): In developing the proposed traffic congestion forecasting model, hyperparameter optimization is carried out to improve the performance of the BPNN by determining optimal learning parameters, such as the learning rate, the number of hidden neurons, and the weight ranges. For this purpose, the ABO algorithm is utilized because of its effective global and local search capabilities. Inspired by the fragrance-based communication behavior of butterflies in nature, ABO searches the hyperparameter space efficiently. Each butterfly emits a scent representing the fitness value of a candidate solution, mathematically expressed in Equation (7).

$$e = dJ^{b} \tag{7}$$

InHyperparameter Search Space, e represents the intensity of the scent, denoting traffic flow to find solutions of traffic congestion optimization by using the candidate hyperparameters in the BPNN algorithm. d refers to the scaling constant used to specify the intensity of the evaluation process. J^{b} denotes the value of the objective function based on speed, density, weather, vehicle count, accidents, and traffic evaluation sensitivity. In the global search stage, butterflies get attracted to the best solution available in the current generation. This phenomenon is expressed in Equation (8).

$$w_j^{s+1} = w_j^{s} + \left(q^2 \times h^{*} - w_j^{s}\right) \times e_j \tag{8}$$

In Global Search Stage, w_j^{s+1} represents the updated hyperparameter j (such as learning rate) for the next iteration to improve traffic congestion prediction. w_j^{s} refers to the value of the hyperparameter j at iteration s during the ABO algorithm optimization. q^2 denotes the exploration parameter for random traffic in ABO, which influences global search diversity. h^{*} represents the optimal solution obtained up to now, indicating the best BPNN structure for IoT traffic. e_j refers to the scent concentration of butterfly j for optimal performance on traffic congestion prediction. In the local search stage, improvement is done by utilizing the neighbor solutions to keep diversity and prevent premature convergence defined in Equation (9).

$$w_j^{s+1} = w_j^{s} + (q^2 \times w_i^{s} - w_l^{s}) \times e_j \tag{9}$$

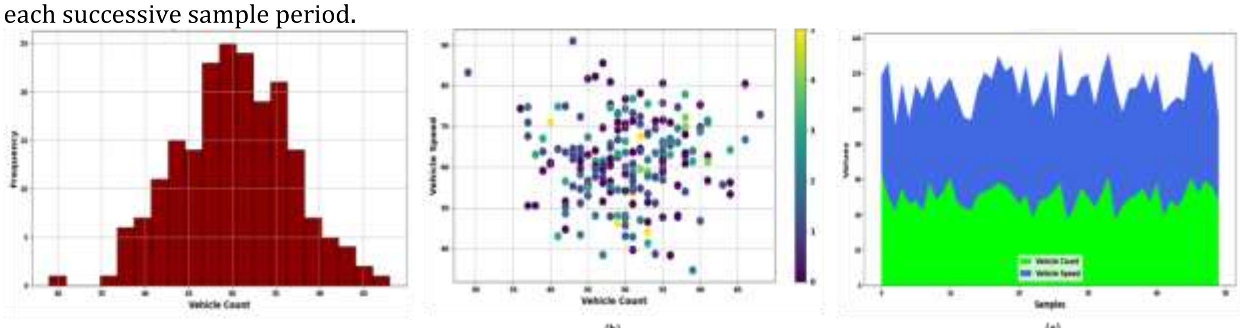
In Local Search Stage, w_i^{s} represents the Value of the hyperparameter associated with a random neighboring butterfly's traffic solution. w_l^{s} refers to the Value of the hyperparameter of another random neighboring butterfly's traffic solution. e_j denotes the level of scent that drives local search according to IoT traffic-based objective function performance.

Hyperparameter of Proposed ABO-BPNN: The proposed model to predict traffic congestion is established using optimized hyperparameters used in establishing the ABO-BPNN model, the Population size (30 - 60), iterations (50 - 150), switching probability (0.7 - 0.9), Sensory modality (0.01 - 0.1), and power exponent for optimization (0.1 - 0.3). Some other parameters of the BPNN framework used include learning rate (0.001 - 0.01), Hidden neurons (64 - 256), Batch size (16 - 64), Epochs (50 - 200), and optimizer. It enhances the traffic data and improves prediction performance within smart city environments.

4. Result

The proposed Data Governance model integrated with the ABO-BPNN framework significantly improves traffic congestion prediction in large-scale smart city environments. Proposed strategy also enhances the efficiency of computing, scalability, and forecasting accuracy for intelligent traffic control systems. This experimental setting is based on GPU-based systems, whereas preprocessing, training, and evaluation tasks are performed using Python, TensorFlow, and Scikit-learn libraries.

Exploratory Data Analysis: Figure 2(a) presents a detailed analysis of the context of traffic congestion prediction, with consideration of baseline traffic flow, traffic deviations from the baseline, and outliers such as extremely high traffic congestion or no traffic at all. In Figure 2(b), the dependency of vehicle traffic and speed on other factors, such as traffic density or time, is demonstrated using the color bar.This is an illustration of the basic relationship between speed and traffic flow, showing the key point at which speed drops off rapidly as traffic flow increases. The graph shown in Fig. 2(c) shows the variation in speed and number of vehicles for each successive sample period.



(a) (b) (c)

Figure 2: Analysis of (a) Distribution of Vehicle Flow for traffic Congestion, (b) Bivariate Correlation between Number of Vehicles and Speed with Density Mapping, and (c) Trends in Time Series for Vehicle Volume and Vehicle Speed Data in a Smart City Environment

Evaluation Metrics: Evaluations of the Data Governance framework with the proposed ABO-BPNN model for predicting traffic congestion are based on classification measurements such as accuracy, a precision, a recall, and a f1-Score. Accuracy represents how accurately the prediction framework is at predicting traffic congestion in an IoT traffic environment. Precision represents how accurate the predictions of traffic congestion are compared to all predictions of traffic congestion. Recall indicates how many traffic congestion cases are detected by the algorithm and F1-Score represents a combination of a precision and a recall in an attempt to achieve optimal results in a diverse smart city traffic environment.

Performance Evaluation: Table 1 and Figure 3 present the effectiveness of the suggested model in relation to the Data Governance Framework with integration of using the ABO-BPNN model for traffic congestion prediction, which is determined relative to some DL models. Some traditional DL models, such as EfficientNet-B0 [16] method is achieves the low accuracy, the Inception-V3 [16] method shows a good accuracy compare to EfficientNet-B0, The VGG-16 [16] method achieves good accuracy relevant to Inception-V3, The Vgg-19 [16] method also shows a same accuracy level compare to Vgg-16, The MobileNet [16] method shows high accuracy, and the ResNet-152 [16] method perform moderately, with an low accuracy rate. On the other hand, models such as EfficientNet + EOA [16] give improved performance with an accuracy of 94.17%. The proposed ABO-BPNN model performs with the highest accuracy 96.21% accuracy, 95.30% precision, 94.78% recall, and a 94.95% F1-score.

Table 1: Comparison Analysis of Existing and Proposed Methods

Methods	Accuracy	Precision	Recall	F1-Score
EfficientNet-B0 [16]	89.46	90.21	88.21	89.31
Inception-V3 [16]	91.82	92.28	91.12	91.76
Vgg-16 [16]	91.12	92.09	90.43	91.13
Vgg-19 [16]	91.68	92.24	90.57	91.71
MobileNet [16]	92.51	92.95	91.40	92.17
ResNet-152 [16]	89.32	90.73	88.21	89.27
EfficientNet + EOA [16]	94.17	94.28	93.76	93.93
ABO-BPNN [Proposed]	96.21	95.30	94.78	94.95

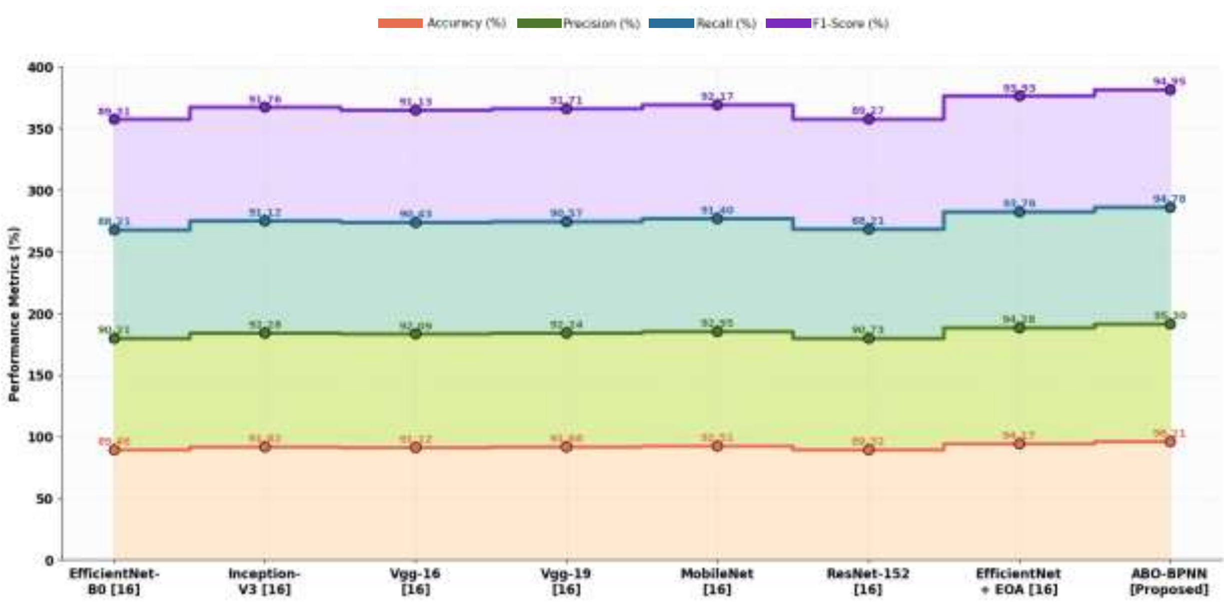


Figure 3: Comparative Evaluation of Conventional and Proposed Methods for the Traffic Congestion Prediction

Table 2 displays the analysis of the comparative performance of different BPNN framework variants for traffic prediction. The standalone BPNN achieved moderate classification performance, while the integration of normalization and PCA progressively improved F1-score, recall, accuracy, and precision by enhancing feature quality and reducing redundancy, the optimal ABO-BPNN full framework with a higher performance level than any other framework, with 96.21% accuracy, 95.30% precision, 94.78% recall, and a 94.95% F1-score.

Table 2: Ablation study Analysis of the Proposed Data Governance framework Integrated with ABO-BPNN

Model Variant	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
BPNN Only	85.42	84.10	83.75	83.92
BPNN + Normalization	88.36	87.20	86.85	87.01
BPNN + PCA	91.78	90.65	90.10	90.37
BPNN + PCA + Normalization	92.85	91.70	91.25	91.46
ABO-BPNN [Full Proposed model]	96.21	95.30	94.78	94.95

Discussion: An efficient Data Governance framework with the integration of the ABO-BPNN framework that ensures accurate prediction of traffic congestion in smart cities characterized by densely populated urban areas, andthe EfficientNet-B0 [16] model poses challenges due to its low compatibility and lack of scalability towards the IoT large-scale traffic data. The Inception-V3 [16] model is limited with respect to weather conditions and varying traffic infrastructures. The Vgg-16 [16] model presentschallengesin terms of efficiency in dealing with varying traffic situations. Moreover, the Vgg-19 [16] model offers limited scalability towards larger networks. The MobileNet [16] algorithm faces challenges in its generalization to other regions or traffic states because it needs some form of adjustment or training to fit into such different scenarios. The ResNet-152 [16] algorithm is not a full integration of the traffic information from the IoT devices. Limitations faced by the EfficientNet + EOA [16] method can be attributed to its complexity as well as limited computational capacity, particularly while dealing with large data sets. In contrast, the proposed ABO-BPNN model addresses existing constraints through effective data management and optimized learning strategies, particularly when dealing with large-scale IoT traffic data. As a result, the ABO-BPNN framework offers effective congestion predictions in varying traffic conditions.

5. Conclusion

The proposed data governance Framework involves the ABO-BPNN model, which can be employed for making predictions in relation to traffic congestion effectively. IoT traffic data on a large scale is utilized to guarantee data quality using noise removal and Min-Max Normalization. Through feature dimension reduction using PCA, an efficient data framework was obtained. Through the use of the ABO algorithm, and 96.21% accuracy, 95.30% precision, 94.78% recall, and a 94.95% F1-score was achieved. Limitations associated with this research include reliance on one dataset containing information about the impact of IoT on traffic, thus restricting its applicability to other urban areas. Whether the system scales well under extreme circumstances is yet another aspect that needs to be further examined. Moving forward, efforts should concentrate on making use of data generated in various cities and implementing transformer algorithms.A further important area includes expanding the present data management architecture by incorporating congestion detection algorithms and edge computing technology.

References

1. Liu, Y., Gong, J., Huang, T. and Cao, J., 2025. Traffic congestion predictor: A spatiotemporal graph neural network with congestion-conditional adaptive mechanism and optimization algorithm. Expert Systems with Applications, 277, p.127200. https://doi.org/10.1016/j.eswa.2025.127200
2. Qi, Y. and Cheng, Z., 2023. Research on traffic congestion forecast based on deep learning. Information, 14(2), p.108. https://doi.org/10.3390/info14020108
3. Zhang, J., Chen, Y., Wang, T., Xie, C.Z.T. and Tian, Y., 2025. Mixture of Spatial–Temporal Graph Transformer Networks for urban congestion prediction using multimodal transportation data. Expert Systems with Applications, 268, p.126108. https://doi.org/10.1016/j.eswa.2024.126108

4. Pandey, S.C. and P, V.K., 2025. Quantum-Enhanced Transformer Network Model for Traffic Congestion Prediction in Smart Transportation Systems. International Journal of Intelligent Transportation Systems Research, 23(2), pp.951-971. https://doi.org/10.1007/s13177-025-00494-9
5. Cheng, W., Li, J.L., Xiao, H.C. and Ji, L.N., 2022. Combination predicting model of traffic congestion index in weekdays based on LightGBM-GRU. Scientific reports, 12(1), p.2912. https://doi.org/10.1038/s41598-022-06975-1
6. Pan, Y.A., Guo, J., Chen, Y., Cheng, Q., Li, W. and Liu, Y., 2024. A fundamental diagram based hybrid framework for traffic flow estimation and prediction by combining a Markovian model with deep learning. Expert Systems with Applications, 238, p.122219. https://doi.org/10.1016/j.eswa.2023.122219
7. Wang, X., Zeng, R., Zou, F., Liao, L. and Huang, F., 2023. STTF: An efficient transformer model for traffic congestion prediction. International Journal of Computational Intelligence Systems, 16(1), p.2. https://doi.org/10.1007/s44196-022-00177-3
8. Asencio-Cortés, G., Florido, E., Troncoso, A. and Martínez-Álvarez, F., 2016. A novel methodology to predict urban traffic congestion with ensemble learning. Soft Computing, 20(11), pp.4205-4216.https://doi.org/10.1007/s00500-016-2288-6
9. Kumar, V., Tiwari, S., Sharma, R.K., Sinha, A. and Tejani, G.G., 2025. A real time video sliced frame image based intelligent traffic congestion monitoring system using faster CNN. Journal of optics, pp.1-16. https://doi.org/10.1007/s12596-025-02606-w
10. Saleem, M., Abbas, S., Ghazal, T.M., Khan, M.A., Sahawneh, N. and Ahmad, M., 2022. Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques. Egyptian Informatics Journal, 23(3), pp.417-426. https://doi.org/10.1016/j.eij.2022.03.003
11. Harrou, F., Zeroual, A., Kadri, F. and Sun, Y., 2024. Enhancing road traffic flow prediction with improved deep learning using wavelet transforms. Results in Engineering, 23, p.102342. https://doi.org/10.1016/j.rineng.2024.102342
12. Kumar, M. and Kumar, K., 2024. Traffic congestion prediction using feature series LSTM neural network and a new congestion index. International Journal on Artificial Intelligence Tools, 33(02), p.2350067. https://doi.org/10.1142/S0218213023500677
13. Wu, P., Zhang, Z., Peng, X. and Wang, R., 2024. Deep learning solutions for smart city challenges in urban development. Scientific Reports, 14(1), p.5176. https://doi.org/10.1038/s41598-024-55928-3
14. Prakash, J., Murali, L., Manikandan, N., Nagaprasad, N. and Ramaswamy, K., 2024. RETRACTED ARTICLE: A vehicular network based intelligent transport system for smart cities using machine learning algorithms. Scientific reports, 14(1), p.468. https://doi.org/10.1038/s41598-023-50906-7
15. Huang, Z., Xia, J., Li, F., Li, Z. and Li, Q., 2019. A peak traffic congestion prediction method based on bus driving time. Entropy, 21(7), p.709. https://doi.org/10.3390/e21070709
16. Vatambeti, R., Reddy, N.R., Ibrahim, S.H.S., Kocherla, R., Shareef, S.K., Yerragudipadu, S. and Jagadesh, B.N., 2025. Enhancing urban traffic congestion prediction through efficientnet and optimized ensemble learning models. Scientific Reports, 15(1), p.40224. https://doi.org/10.1038/s41598-025-24012-9