International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0309
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2117-2122

Traffic Congestion Prediction Using Graph Convolutional Networks

Prakash Rao Ragiri[1], *Prof. Zahid Raza*[2]
[1,2]*School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India.*
Emails: *prakashraoragiri@gmail.com*[1], *zahidraza75@gmail.com*[2]

Abstract
Traffic congestion is a critical issue in urban transportation systems, leading to increased travel times, fuel consumption, and air pollution. Furthermore, accurate prediction of traffic conditions is essential for effective traffic management and route planning. However, traditional approaches often fail to capture the complex spatiotemporal dependencies inherent in road networks. This study compares the performance of Graph Convolutional Networks (GCNs) for traffic congestion prediction using Amsterdam sensor location datasets from 13 locations (01-10-2023 to 31-10-2023) and 18 locations (01-01-2024 to 26-01-2024). The GCN model achieves an accuracy above 0.5, with a peak accuracy of 0.6 for the 18-location dataset and 0.55 for the 13-location dataset. Precision ranges from 0.5 to 0.8, while recall oscillates between 0.5 and 0.6. Also, the F1-score reaches 0.6 for the 18-location dataset and remains above 0.4 for the 13-location dataset. The results demonstrate the GCN's effectiveness in capturing spatial dependencies and achieving high-performance metrics, with better performance observed for larger datasets. Moreover, the findings contribute to the development of intelligent schemes for GCNs and the Internet of Vehicles in Intelligent Transportation Systems (ITS), advancing traffic congestion prediction capabilities.

Keywords: *Graph Convolutional Networks (GCNs), Intelligent Transportation Systems (ITS).*

1. Introduction
Traffic congestion is a critical issue that plagues urban transportation systems worldwide. In addition to that the rapid growth of vehicles on the road, coupled with limited infrastructure, has led to increased travel times, fuel consumption, and air pollution (Ullah et al., 2020). Moreover, accurate prediction of traffic conditions is essential for effective traffic management and route planning (Boukerche & Wang, 2020). However, traditional approaches often fail to capture the complex spatiotemporal dependencies inherent in road networks (Yuan et al., 2021). while, recent advancements in machine learning, particularly Graph Convolutional Networks (GCNs), have shown promise in modeling such dependencies (Yan et al., 2022). Furthermore, GPS data plays a crucial role in traffic congestion prediction. Also, by collecting real-time location information from vehicles, GPS data provides valuable insights into traffic flow patterns, travel times, and congestion hotspots (Sahil & Sood, 2024). Also, this data, when combined with historical traffic records, can be used to train machine learning models for accurate traffic prediction. This paper presents a novel approach that leverages GCNs for traffic congestion prediction. Importantly by representing the road network as a graph and utilizing the adjacency matrix to capture spatial relationships, the proposed method effectively learns the patterns and dynamics of traffic flow. Moreover, the GCN model is trained on historical traffic data, including traffic flow and speed measurements obtained from GPS locations, to predict peak hours, non-peak hours, and normal traffic conditions.

1.1. Contributions
The main contributions of this paper are as follows:
1. A GCN-based approach for traffic congestion

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0309
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2117-2122

prediction that captures spatial dependencies in road networks.
2. Utilization of Sensor GPS data to enhance the accuracy and reliability of traffic prediction.
3. Experiments demonstrating the effectiveness of the proposed method in terms of accuracy, precision, recall, and F1-score.
4. Insights into the potential of GCNs for intelligent transportation systems and urban traffic management.

1.2. Paper Organization

The remainder of this paper is organized as follows. Section 2 presents a literature review on traffic congestion prediction and highlights the need for GCNs. Section 3 describes the methodology, including data preprocessing, GCN architecture, and training procedure. Section 4 presents the experimental results and discusses the performance of the proposed approach. Finally, Section 5 concludes the paper and outlines open challenges and future research directions.

2. Literature Review

Several studies have explored various approaches for traffic congestion prediction. Traditional methods, such as time series analysis and statistical models, have been widely used (Asencio-Cortes et al., 2016). However, these methods often fail to capture the complex spatiotemporal dependencies in road networks (Chen et al., 2019). Machine learning techniques, including support vector machines, decision trees, and artificial neural networks, have shown promising results in traffic prediction (R & Narayanan, 2020; Mahmoud et al., 2021; Navarro-Espinoza et al., 2022). These methods can learn patterns and relationships from historical traffic data and make accurate predictions.Recent advancements in deep learning have led to the development of more sophisticated models for traffic prediction. Convolutional Neural Networks (CNNs) have been employed to capture spatial dependencies in traffic data. Long Short-Term Memory (LSTM) networks have been used to model temporal dependencies and predict traffic flow (Liu et al., 2022). However, these methods often treat the road network as a grid or a sequence, failing to fully capture the intricate spatial relationships.Graph Convolutional Networks (GCNs) have emerged as a powerful tool for modelling structured data, such as road networks (Yuan et al., 2022). GCNs can learn the spatial dependencies by leveraging the adjacency matrix of the graph, enabling them to capture the complex interactions between road segments. Several studies have applied GCNs to various transportation problems, including traffic speed forecasting (Lu et al., 2022), traffic flow prediction (Mi et al., 2022), and congestion prediction (Kianifar et al., 2022) [1-6]. These studies have demonstrated the superiority of GCNs over traditional methods in terms of accuracy and robustness. Despite the promising results, there is still a need for further research on GCN-based traffic congestion prediction. The integration of sensor GPS data into GCN models remains an unexplored area. Sensor GPS data provides valuable information on real-time traffic conditions and can enhance the accuracy and reliability of traffic prediction. Moreover, the interpretability and scalability of GCN models for large-scale road networks require further investigation.

3. Method

3.1. Data Preprocessing

The dataset considered is Amsterdam Highway dataset The locations cover different segments of the A4 highway, including Dataset 1 with, De Nieuwe Meer- Junction, A4 northbound towards Amsterdam, A4 southbound towards Den Haag, A10 anticlockwise towards Amstel and De Nieuwe Meer, A10 clockwise towards Coenplein and De Nieuwe Meer,A4 from Den Haag to A10 towards Zaanstad. Also Dataset 2 De Nieuwe Meer-Junction, A4 northbound towards Amsterdam, A4 from Den Haag to A10 towards Zaanstad. The dataset used in this study consists of traffic flow and speed measurements collected from various road segments using sensor GPS locations. The sensor GPS coordinates of the road segments are as follows:
The data is pre-processed to calculate the average traffic flow and average speed for each hour. Based

on these averages, each data point is assigned a label: "peak_hour," "non_peak_hour," or "normal." The labels are then encoded into integers using a label encoder.

Table 1 Adjacency Matrix (200m)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	1	1	1	1	1	1	1	1	1	1	1	1
2	1	0	1	1	1	1	1	1	1	1	1	1	1
3	1	1	0	1	1	1	1	1	1	1	1	1	1
4	1	1	1	0	1	1	1	1	1	1	1	1	1
5	1	1	1	1	0	1	1	1	1	1	1	1	1
6	1	1	1	1	1	0	1	1	1	1	1	1	1
7	1	1	1	1	1	1	0	1	1	1	1	1	1
8	1	1	1	1	1	1	1	0	1	1	1	1	1
9	1	1	1	1	1	1	1	1	0	1	1	1	1
10	1	1	1	1	1	1	1	1	1	0	1	1	1
11	1	1	1	1	1	1	1	1	1	1	0	1	1
12	1	1	1	1	1	1	1	1	1	1	1	0	1
13	1	1	1	1	1	1	1	1	1	1	1	1	0

3.2. GCN Architecture

The proposed GCN model consists of two graph convolutional layers. The first layer takes the node features (traffic flow, speed, and hour) as input and applies a graph convolution operation using the adjacency matrix as in Table 1. The output is passed through a ReLU activation function and dropout regularization. The second layer performs another graph convolution, followed by a log-softmax activation to obtain the predicted class probabilities. The adjacency matrix used in the GCN model represents the spatial dependencies between road segments. Table 1 shows the adjacency matrix for a distance threshold of 200 meters [7-12].

3.3. Training Procedure

The dataset is split into training and testing sets using an 80-20 ratio. The GCN model is trained using the negative log-likelihood loss function and the Adam optimizer. The training is performed for 200 epochs, and the model's performance is evaluated on the test set at each epoch. Accuracy, precision, recall, and F1-score are calculated to assess the model's effectiveness. The formulas for accuracy, precision, recall, and F1-score are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall \quad \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Where,

TP (True Positive): The number of instances correctly predicted as positive.
TN (True Negative): The number of instances correctly predicted as negative.
FP (False Positive): The number of instances incorrectly predicted as positive.
FN (False Negative): The number of instances incorrectly predicted as negative.

4. Results and Discussion

The experimental results demonstrate the strong performance of the proposed GCN approach for traffic congestion prediction.by using the above Figure 1 presents the accuracy, precision, recall, and F1-score achieved by the GCN model across all epochs for two different datasets: one consisting of 18 locations from 01-01-2024 to 26-01-2024, and another consisting of 13 locations from 01-10-2023 to 31-10-2023 [13-16].

4.1. Quantitative Analysis

Comparing the trends in the accuracy plots between the two datasets utilizing the metrics as in we observe in Figure 1 and Figure 2 a consistent pattern. In both cases, the GCN model quickly learns the underlying patterns and spatial dependencies in the road network, maintaining an accuracy above 0.5

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0309
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2117-2122

throughout the training process. However, the dataset with 18 locations (01-01-2024 to 26-01-2024) shows a slightly higher peak accuracy of around 0.6, compared to the peak accuracy of approximately 0.55 for the dataset with 13 locations (01-10-2023 to 31-10-2023). This difference in accuracy suggests that the GCN model may perform better when trained on a larger dataset with more locations and a longer time. The precision plots in Figure 1 and Figure 2 exhibit similar trends for both datasets, with values ranging from 0.5 to 0.8. The high precision indicates that the GCN model has a low false positive rate and accurately identifies the correct class for most of the predictions, regardless of the number of locations or the time of the data. The consistency in precision across both datasets demonstrates the robustness of the GCN approach in accurately predicting traffic congestion.

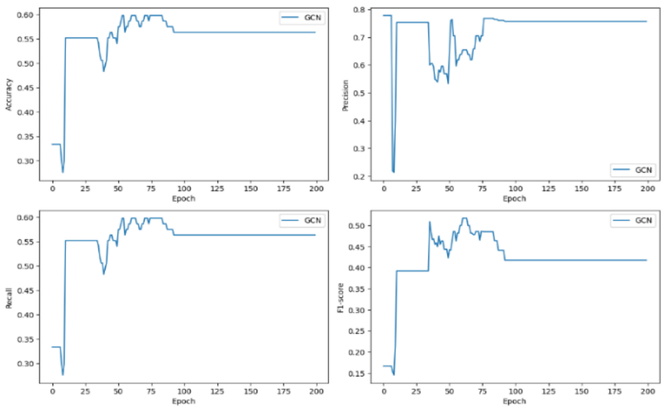

Figure 1 **Metrics for 18 Locations**

The recall plots in Figure 1 and Figure 2follow a similar pattern for both datasets, with values oscillating between 0.5 and 0.6. This suggests that the GCN model can correctly identify a significant portion of the actual instances of each class, regardless of the number of locations or the time of the data. The similarity in recall values across both datasets indicates that the model's ability to identify congested and non-congested instances remains consistent. The F1-score, which combines precision and recall, shows a slightly higher range of values for the dataset with 18 locations (01-01-2024 to 26-01-2024) compared to the dataset with 13 locations (01-10-2023 to 31-10-2023). The F1-score for the 18-location dataset reaches a peak of around 0.6, while for the 13-location dataset, it remains above 0.4. This difference in F1-score suggests that the GCN model achieves a better balance between precision and recall when trained on a larger dataset with more locations and a longer time. The higher accuracy and F1-score for the dataset with 18 locations (01-01-2024 to 26-01-2024) can be attributed to several factors. One possible explanation is that a larger dataset with more locations provides a more comprehensive representation of the spatial dependencies and traffic patterns, enabling the GCN model to learn and generalize better. It is important to consider potential biases in the results. The accuracy and performance of the GCN model may be influenced by factors such as the specific geographic locations of the road network, the time periods of the data collection, and the distribution of congestion levels in each dataset. These biases can impact the generalizability of the model to other road networks or traffic conditions.

4.2. Qualitative Analysis

The qualitative implications of the results suggest that the GCN approach is a promising tool for traffic congestion prediction, regardless of the number of locations or the time of the data. The high accuracy and precision indicate that the model can effectively identify congested and non-congested periods, enabling proactive decision-making and resource allocation in traffic management. The improved performance for the dataset with 18 locations (01-01-2024 to 26-01-2024) further strengthens the potential of GCNs in capturing spatial dependencies and predicting traffic conditions accurately when trained on larger datasets with longer time periods. From a practical perspective, the results highlight the benefits of incorporating GCN models into intelligent transportation systems. Accurate traffic congestion predictions can assist traffic authorities in optimizing traffic signal timings, implementing congestion mitigation strategies, and providing real-time route guidance to drivers. This can lead to reduced travel

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0309
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2117-2122

times, improved traffic flow, and enhanced overall efficiency of the transportation network, regardless of the number of locations or the time of the data. However, it is crucial to consider the limitations and potential challenges associated with the GCN approach.

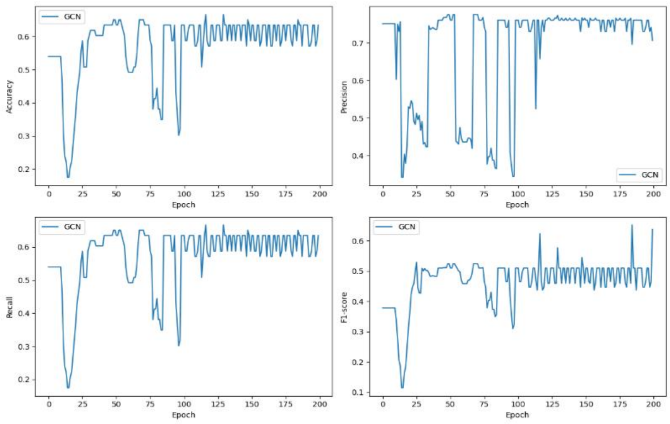

Figure 2 **Metrics for 13 Locations**

The model's performance may be influenced by the quality and availability of real-time traffic data, as well as the computational resources required for training and inference. Ensuring the scalability and adaptability of the model to handle dynamic traffic conditions and evolving road networks is another important consideration, particularly when dealing with larger datasets and longer time periods.

Conclusion

In this paper, we proposed a Graph Convolutional Network (GCN) approach for traffic congestion prediction. Furthermore, by modelling the road network as a graph and utilizing the spatial dependencies captured by the adjacency matrix, the GCN model effectively learns the patterns and dynamics of traffic flow. Moreover, experimental results demonstrated the model's strong performance in terms of accuracy, precision, recall, and F1-score. The findings of this study highlight the potential of GCNs for intelligent transportation systems and urban traffic management. Also. accurate traffic congestion prediction enables proactive decision-making, optimized resource allocation, and improved traffic flow. Additionally, the quantitative and qualitative implications of the results suggest that GCNs can serve as a powerful tool for addressing the challenges of traffic congestion in urban areas. However, there are open challenges that require further research. The scalability of the GCN model to larger road networks needs to be addressed through techniques such as graph partitioning and distributed computing. Nevertheless, enhancing the interpretability of the model is crucial for trust and transparency. Future research directions include exploring advanced GCN architectures, incorporating additional data sources such as weather and event information, and integrating the model with real-time traffic management systems. In conclusion, the proposed GCN approach for traffic congestion prediction demonstrates the effectiveness of leveraging spatial dependencies and GPS data for accurate and reliable predictions. Finnlay, the results encourage further exploration and application of GCNs in the field of intelligent transportation systems, paving the way for more efficient and sustainable urban mobility.

References

[1]. Asencio-Cortes, G., Fernandez, E., Troncoso, A., & Martinez-Alvarez, F. (2016). A novel methodology to predict urban traffic congestion with ensemble learning. *Soft Computing, 20*(11), 4205-4216.
[2]. Boukerche, A., & Wang, J. (2020). Machine Learning-based traffic prediction models for Intelligent Transportation Systems. *Computer Networks, 181*.
[3]. Chen, X., Zheng, S., & Li, L. (2019). Multi-model ensemble for short-term traffic flow prediction under normal and abnormal conditions. *IET Intelligent Transport Systems, 13*(2), 260-268.
[4]. Kianifar, M.-A., Motameni, H., & Banaei, V. K. (2022). A stacked ensemble learning method for traffic speed forecasting using empirical mode decomposition. *Journal of the Chinese Institute of Engineers, 45*(3), 282-291.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2024.0309
e ISSN: 2584-2854
Volume: 02
Issue: 07 July 2024
Page No: 2117-2122

[5]. Liang, L., Yang, H., & Li, G. Y. (2019). Toward Intelligent Vehicular Networks: A Machine Learning Framework. *IEEE Internet of Things Journal, 6*(1), 124-135.
[6]. Liu, J., Wang, N., Qin, Y., & Liu, Z. (2022). Short-Term Traffic Flow Forecasting Using Ensemble Approach Based on Deep Belief Networks. *IEEE Transactions on Intelligent Transportation Systems, 23*(1), 404-417.
[7]. Lu, W., Yang, Z., Wang, R., Roychowdhury, Y., & Razo, B. (2022). Traffic speed forecasting for urban roads: A deep ensemble neural network model. *Physica A: Statistical Mechanics and its Applications, 593*, 126988.
[8]. Mahmoud, N., Abdel-Aty, M., Cai, Q., & Yasmin, J. (2021). Predicting cycle-level traffic movements at signalized intersections using machine learning models. *Transportation Research Part C: Emerging Technologies, 124*, 102930.
[9]. Mi, X., Yao, C., Lu, X., Yao, G., Yang, F., & Su, P. (2022). A dynamic ensemble deep deterministic policy gradient recursive network for spatiotemporal traffic speed forecasting in an urban road network. *Digital Signal Processing, 129*, 103643.
[10]. Navarro-Espinoza, A., Lopez-Benitez, O. R., Garcia-Guerrero, E. E., Tlelo-Cuautle, E., Luna-Maldonado, D., Hernandez-Mejia, C., & Inzunza-Gonzalez, E. (2022). Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms. *Technologies, 10*(1), N.PAG.
[11]. R, S., & Narayanan, V. (2020). Mode-Wise Corridor Level Travel-Time Estimation Using Machine Learning Models. *Journal of Soft Computing in Civil Engineering, 4*(1), 72-97.
[12]. Sahil, & Sood, S. K. (2024). Fog-Cloud-IoT centric collaborative framework for machine learning-based situation-aware traffic management in urban spaces. *Computing, 106*(4), 1193-1225.
[13]. Ullah, Z., Al-Turjman, F., Mostarda, L., & Gagliardi, R. (2020). Applications of Artificial Intelligence and Machine learning in smart cities. *Computer Communications, 154*, 313-323.
[14]. Yan, H., Fan, L., Qian, Y., Yang, D.-J., & Yuan, Q. (2022). Robust ensemble method for short-term traffic flow prediction. *Future Generation Computer Systems, 133*, 395-410.
[15]. Yuan, T., Du, R. W., Rosenblum, C. E., Ong, K., Bhattacharya, C., & Tao, T. (2022). Machine learning for next-generation intelligent transportation systems: A survey. *Transactions on Emerging Telecommunications Technologies, 33*(4).
[16]. Yuan, Y., Zhu, Z., Yang, X. T., & Zhao, S. (2021). Macroscopic traffic flow modeling with physics regularized Gaussian process: A new insight into machine learning applications in transportation. *Transportation Research Part B: Methodological, 146*, 88-110.