IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

City-Wide Traffic Prediction using Aggregated Mobile Network Activity Data

[1]Pranav Chandrakant Kute, [2]Prof. M. S. Dighe, [3]Prof. P. A. Sawant, [4]Dr. H. B. Jadhav
Student, Department of Computer Engineering[1]
Guide, Department of Computer Engineering[2]
Co-Guide , Department of Computer Engineering[3]
HOD, Department of Computer Engineering[4]
Adsul's Technical Campus, Chas, Ahilyanagar , India

Abstract: *Accurate city-wide traffic prediction is essential for intelligent mobility management, urban planning, and traffic operations. Conventional approaches often depend on fixed traffic sensors or individual vehicle trajectories, which may provide uneven spatial coverage. This paper presents a framework based on aggregated cellular network activity as a large-scale source of mobility information for predicting road travel time and speed. The source research introduces two complementary deep-learning models: the Network Insight Model, which uses aggregated cellular activity alone, and the Integrated Mobility Model, which combines cellular activity with Bluetooth sensor observations. Cellular activity is represented through Network Activity Tensors (NAT), while Bluetooth observations are represented through Bluetooth Graph Tensors (BGT). Graph Attention Networks (GAT), Gated Recurrent Units (GRU), and convolutional layers are used to learn spatial and temporal relationships and generate city-wide origin–destination predictions. The models are evaluated using real-world data from York Region, Canada, with MAE, RMSE, R^2, MAPE, and Travel Time Index (TTI). The reported results show that combining cellular and Bluetooth data improves both travel-time and speed prediction compared with the cellular-only model and the selected baselines.*

Key Words: Smart transportation, traffic prediction, aggregated cellular network activity, Bluetooth sensors, Network Activity Tensor, Graph Attention Network, GRU, Travel Time Index

1. INTRODUCTION

Traffic prediction is an important part of modern transportation systems because it supports urban mobility management, road safety, and transportation planning. Accurate prediction across an entire city is challenging due to the spatial and temporal dependencies in traffic flow. Traffic conditions show periodic behavior, such as morning and evening peak-hour congestion and lower traffic during off-peak hours. In addition, the road network is interconnected, so traffic conditions at one intersection can influence upstream and downstream road segments.

Traditional approaches such as historical averaging, Kalman filtering, and linear regression may not fully represent the nonlinear and stochastic nature of traffic. Deep-learning methods, including Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRUs), have been used to model spatial and temporal traffic dependencies.

Many existing traffic prediction methods depend on localized road sensors or individual trajectory data. These sources can have inconsistent coverage and may not support reliable prediction in areas with limited instrumentation. Aggregated cellular network activity offers an additional source of mobility information. Changes in the number of active mobile devices across cellular sites can reflect population movement and traffic-related patterns at a city scale.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

The source research proposes two models: the Network Insight Model, which uses cellular network activity, and the Integrated Mobility Model, which fuses cellular activity with Bluetooth sensor observations. Both models use attention mechanisms to identify influential spatial and temporal relationships.

II. RELATED WORK

A. Trajectory-Based Prediction

Trajectory-based methods estimate travel time or arrival time using the movement histories of individual vehicles or passengers. These methods use GPS traces, Bluetooth detectors, and real-time transit feeds. Machine-learning and deep-learning techniques such as k-Nearest Neighbors, Gaussian Mixture Models, LSTM, GRU, and deep neural networks have been applied to bus and private-vehicle travel-time prediction. Although these methods can perform well for particular trips or road segments, their dependence on detailed trajectory or sensor data can limit city-wide coverage.

B. City-Wide Prediction

City-wide traffic prediction forecasts travel time, speed, or traffic flow over large urban areas. Some methods divide a city into grids or regions, while others represent intersections and road segments as a graph. Graph-based and hybrid architectures combine spatial feature extraction with temporal sequence learning. Methods such as A3TGCN combine graph convolution and recurrent processing to model spatio-temporal dependencies.

C. Limitations of Existing Approaches

Dependence on dense sensor or trajectory data can result in inconsistent spatial coverage.
Models trained on fixed sensor layouts or city-specific trajectories may have limited generalizability to different urban environments.
Traffic is directional, and incoming and outgoing flows can affect intersections differently.
Many prediction approaches provide limited interpretability about which intersections or road segments influence the output.
The considered framework addresses these limitations by using aggregated cellular network activity, which is available through cellular infrastructure, and by combining it with Bluetooth data where available. Attention mechanisms are used to provide interpretable spatial and temporal information.

III. TRAVEL TIME PREDICTION SYSTEM

A. System Overview

The system receives two data sources: aggregated cellular network activity and Bluetooth sensor observations. A Tensor Generator converts these inputs into two structured representations: the Network Activity Tensor (NAT) and the Bluetooth Graph Tensor (BGT). Depending on the available data, the system uses the Network Insight Model or the Integrated Mobility Model. A Traffic Predictor then generates city-wide travel-time and speed estimates.
Cellular network activity provides broad spatial coverage and captures fluctuations in the number of active devices across cellular sites. Bluetooth sensors complement this information by recording anonymous detections of passing vehicles at selected intersections. These observations support the estimation of travel time, speed, and traffic volume along connecting road segments.

FIG. 1. SYSTEM WORKFLOW

Cellular network activity + Bluetooth data	Tensor Generator (NAT and BGT)	Network Insight / Integrated Mobility Model	Attention-based traffic prediction	Travel time, speed and TTI estimates

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

B. Network Activity Tensor

The cellular operator API reports the number of connected devices at 15-minute intervals. These measurements are transformed into a three-dimensional Network Activity Tensor. For each Bluetooth sensor location, the five nearest cellular sites are identified and their activity counts are mapped to that intersection. The tensor includes the sensor or intersection index, the nearest cellular sites, and historical time offsets.

The source study describes two NAT versions. NAT V1 uses nine recent and periodic time offsets, including the last 30 minutes, the previous hour, the previous day, and the previous week. NAT V2 extends the historical information to include one, two, and three weeks prior, helping the model learn recurring weekly patterns such as weekday commuting and weekend behavior.

C. Bluetooth Graph Tensor

Bluetooth sensors provide intersection-level traffic observations. The Bluetooth Graph Tensor represents traffic information such as travel time, vehicle speed, and incoming and outgoing car counts. The data are organized using the road-network graph, in which intersections are nodes and road segments are edges.

D. Data Alignment and Missing Values

The cellular and Bluetooth inputs are aligned by location and time to support multimodal learning. The source study reports that missing data are imputed with zero. Tensor normalization uses a seven-day moving mean and standard deviation so that values are treated consistently across intersections and time steps.

IV. DEEP-LEARNING MODEL ARCHITECTURE

A. Network Insight Model

The Network Insight Model uses only the Network Activity Tensor. It learns relationships between cellular-site activity and traffic conditions over time and predicts city-wide travel time and speed. This model is suitable for locations where Bluetooth sensor observations are sparse or unavailable.

B. Integrated Mobility Model

The Integrated Mobility Model combines the Network Activity Tensor and Bluetooth Graph Tensor. Cellular activity provides broad mobility information, while Bluetooth observations provide localized traffic measurements. By fusing these two sources, the model can learn complementary information and improve prediction accuracy.

C. Graph Attention Network

The road network is represented as a graph in which intersections are nodes and road segments are edges. Graph Attention Networks assign normalized attention coefficients to neighboring nodes. These coefficients indicate how strongly information from one intersection contributes to the representation of another. The attention mechanism can distinguish influential upstream and downstream intersections and can represent directional relationships.

D. Gated Recurrent Unit

A Gated Recurrent Unit processes historical measurements to capture temporal dependencies. Its gating mechanism controls how information is retained or forgotten over time. In the considered framework, the GRU summarizes historical traffic and network-activity sequences into temporal feature representations.

E. Graph Fusion and Convolution

The outputs of the GRU are reshaped and combined to represent origin–destination intersection pairs. In the source study, the resulting feature tensor has a 437 × 437 origin–destination structure. The Network Insight Model uses a feature depth of 20, while the Integrated Mobility Model uses a feature depth of 22. A final 1 × 1 convolution combines the learned features while preserving the origin–destination dimensions and generates the city-wide prediction matrix.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

V. EXPERIMENTS AND EVALUATION

A. Dataset

The study uses two datasets collected in York Region, Ontario, Canada, from December 2022 to April 2023. Bluetooth travel data are obtained from the York Region Open Data Portal at five-minute intervals and cover approximately 1,300 road segments. The records include the number of cars, average speed, average travel time, origin and destination intersections, travel direction, and road length.

Network activity data are obtained through a cellular operator API. The API reports the number of unique active devices in each cell for a specified time window. The data are aggregated and privacy-preserving by design. Measurements are collected at 15-minute observation intervals. To improve the estimate of active devices, the study invokes the API three times using five-minute intervals and aggregates the results.

B. Experimental Setup

The models and preprocessing pipeline are implemented in Python. The experiments use PyTorch and PyTorch Geometric for deep-learning and graph-learning components, and NumPy and Pandas for data preprocessing and tensor manipulation. The experimental server is equipped with an AMD Ryzen Threadripper 3960X processor and two NVIDIA GeForce RTX 3080 GPUs. The models are trained with a learning rate of 0.0001, batch size of 8, and 500 epochs. The forecasting task is one-step-ahead prediction for the next five-minute interval.

The model architecture uses five hidden layers with sizes 12, 10, 12, 10, and 18. Eight attention heads are used in the graph attention component, and Leaky ReLU is used as the activation function. The Integrated Mobility Model requires approximately 48 hours to converge in the reported GPU-enabled environment; inference requires a single forward pass.

C. Evaluation Metrics

The study evaluates travel-time and speed prediction using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), coefficient of determination (R^2), and Mean Absolute Percentage Error (MAPE). It also uses the Travel Time Index (TTI) to evaluate congestion patterns.

TABLE I. EVALUATION METRICS

Metric	Meaning	Preferred result
MAE	Average absolute difference between actual and predicted values.	Lower
RMSE	Square root of mean squared prediction error; penalizes larger errors.	Lower
R^2	Proportion of variance in actual values explained by the model.	Closer to 1
MAPE	Average absolute prediction error expressed as a percentage.	Lower
TTI	Ratio of prevailing travel time to free-flow travel time.	Closer to 1 indicates less congestion

At the road-segment level, TTI is calculated as the observed travel time divided by the corresponding free-flow travel time. City-wide TTI is aggregated across road segments. A higher TTI indicates greater congestion.

VI. RESULTS AND DISCUSSION

A. Travel-Time Prediction

The Network Insight Model, trained only on aggregated network activity data, achieves an MAE of 15.55, RMSE of 27.70, R^2 of 0.83, and MAPE of 14.79%. The Integrated Mobility Model improves these results, achieving an MAE of 13.03, RMSE of 22.97, R^2 of 0.88, and MAPE of 12.29%. The source study reports that the Linear Regression and XGBoost baselines have MAE values above 20, RMSE values above 37, and MAPE values above 20%.

TABLE II. TRAVEL-TIME PREDICTION RESULTS

Model	MAE	RMSE	R^2	MAPE
Network Insight Model	15.55	27.70	0.83	14.79%

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

Integrated Mobility Model	13.03	22.97	0.88	12.29%
Random Forest	18.54	35.32	0.75	17.75%
Linear Regression	Above 20	Above 37	—	Above 20%
XGBoost	Above 20	Above 37	—	Above 20%

B. Speed Prediction

For speed prediction, the Network Insight Model achieves an MAE of 6.52, RMSE of 8.72, R^2 of 0.70, and MAPE of 14.26%. The Integrated Mobility Model improves the results to an MAE of 5.71, RMSE of 7.72, R^2 of 0.76, and MAPE of 12.44%. The Random Forest baseline records an MAE of 11.56, RMSE of 15.44, R^2 of 0.65, and MAPE of 18.15%.

TABLE III. SPEED PREDICTION RESULTS

Model	MAE	RMSE	R^2	MAPE
Network Insight Model	6.52	8.72	0.70	14.26%
Integrated Mobility Model	5.71	7.72	0.76	12.44%
Random Forest	11.56	15.44	0.65	18.15%
A3TGCN	8.10	11.20	0.36	19.78%

C. Attention-Based Interpretation

The attention weights reveal how intersections influence one another. The study identifies strong directional relationships between intersection pairs 102 and 40, 386 and 87, and 8 and 47. The attention scores are asymmetric, meaning that the influence from one intersection to another can differ by direction. These findings can support the identification of important traffic corridors and provide information useful for traffic planning and signal coordination.

D. Congestion and Travel Time Index

The city-wide TTI analysis shows low and relatively stable traffic during night and early-morning hours (00:00–05:00), followed by a morning peak between approximately 07:00 and 09:00, when TTI exceeds 1.7. Congestion remains moderate around midday and reaches its highest level during the evening peak from approximately 15:00 to 18:00, with TTI values up to 1.83. This indicates travel times can be nearly 83% longer than free-flow travel time. The Integrated Mobility Model follows the historical congestion pattern more closely than the Network Insight Model.

E. Holiday and Rush-Hour Performance

The study also evaluates Christmas and New Year holidays and two rush-hour windows: 06:00–10:00 and 14:00–19:00. The Integrated Mobility Model consistently records lower MAPE than the Network Insight Model for travel-time and speed predictions during the examined peak periods. The study further reports that extended historical information improves the stability of travel-time prediction, while shorter-term information can be more effective for instantaneous speed variation.

The numerical results in Tables II and III are transcribed from the source research paper. Where the source text provides only a qualitative comparison for a baseline

VII. CONCLUSION

This paper has described a city-wide traffic prediction framework based on aggregated mobile network activity. The framework introduces the Network Insight Model, which uses cellular network activity alone, and the Integrated Mobility Model, which combines cellular activity with Bluetooth sensor data. Network Activity Tensors and Bluetooth Graph Tensors represent the two data sources, while Graph Attention Networks, GRUs, and convolutional layers learn spatial and temporal traffic patterns.

The reported experiments on York Region data show that cellular network activity provides a useful signal for predicting travel time and speed. Integrating Bluetooth observations improves prediction accuracy further. Attention weights also provide interpretable information about influential intersections and directional relationships. Future work

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 1, November 2026
Impact Factor: 8.2

identified in the source study includes optimizing the spatial granularity of cellular inputs, from pico to macro cells, and extending temporal data coverage beyond six months to improve robustness and long-term prediction performance.

ACKNOWLEDGMENT

The author acknowledges the researchers who developed the source study and the cellular operator and York Region data providers whose aggregated network activity and Bluetooth traffic data supported the original research. The technical content and reported experimental results in this paper are based on the cited research article.

REFERENCES

[1] H. Geng, N. Beigi-Mohammadi, D. P. Sharma, J. Liu, and A. Leon-Garcia, "City-Wide Traffic Prediction Using Aggregated Mobile Network Activity Data," IEEE Open Journal of Intelligent Transportation Systems, vol. 7, pp. 1712–1728, 2026, doi: 10.1109/OJITS.2026.3708478.

[2] J. Bai et al., "A3T-GCN: Attention temporal graph convolutional network for traffic forecasting," ISPRS International Journal of Geo-Information, vol. 10, no. 7, pp. 1–17, 2021.

[3] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. International Conference on Learning Representations, 2016.

[4] P. Veličković et al., "Graph attention networks," in Proc. International Conference on Learning Representations, 2018.

[5] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Advances in Neural Information Processing Systems, 2019, pp. 8024–8035.

[6] M. Fey and J. E. Lenssen, "PyTorch Geometric," Cornell University, Technical Report, 2019.

[7] S. Carrese, E. Cipriani, U. Crisalli, A. Gemma, and L. Mannini, "Bluetooth traffic data for urban travel time forecast," Transportation Research Procedia, vol. 52, pp. 236–243, 2021.

[8] York Region, "York Region Open Data," 2021. [Online]. Available: https://www.york.ca/york-region/statistics-and-data/open-data