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 3, November 2026
Impact Factor: 8.2

City-Wide Traffic Prediction Using Aggregated Mobile Network Activity Data

Krushna Shivaji Yadav[1] and Prof. Mrs. P. S. Gundpujari[2]
Department of Electronics & Telecommunication Engineering
Adsul's Technical Campus, Chas, Ahilyanagar, India

Abstract: *This manuscript presents a source-based analysis of city-wide traffic prediction using aggregated mobile-network activity, with York Region, Ontario, Canada, as the study context. It summarizes the Network Insight Model, which uses aggregated cellular activity, and the Integrated Mobility Model, which combines cellular activity with Bluetooth observations. The supplied reference article reports that the models use Network Activity Tensors (NAT), Bluetooth Graph Tensors (BGT), Graph Attention Networks (GAT), Gated Recurrent Units (GRU), and convolutional fusion to predict travel time and speed. Its reported York Region results are reproduced here as results of the cited study—not as experiments conducted by the present authors. For travel-time prediction, the reference reports MAE/RMSE/R²/MAPE of 15.55/27.70/0.83/14.79% for the Network Insight Model and 13.03/22.97/0.88/12.29% for the Integrated Mobility Model. For speed prediction, the corresponding Integrated Mobility Model results are 5.71/7.72/0.76/12.44%. This manuscript is therefore a reference-based technical review and does not claim an independent replication or new York Region measurements.*

Keywords: city-wide traffic prediction, aggregated mobile network activity, Network Activity Tensor, Bluetooth Graph Tensor, Graph Attention Network, GRU, York Region, literature-based analysis.

I. INTRODUCTION

Urban traffic conditions vary over time and are coupled across connected road segments. Forecasting these conditions can help traffic agencies anticipate peak demand, identify recurring congestion, and plan operational interventions. Conventional point sensors provide direct measurements but may be unavailable or unevenly distributed. Individual vehicle trajectories can offer detailed movement information, yet they raise data-governance concerns and may not provide complete coverage.

Aggregated mobile-network activity is a potential complementary mobility signal. When supplied by an operator under an appropriate data-sharing agreement, counts of active devices can describe broad changes in activity across network areas without exposing individual subscriber trajectories. Such activity is not a direct measurement of vehicles: device ownership, network coverage, device settings, and non-road activity can all affect the relationship between device counts and traffic. Consequently, any model using cellular activity must be calibrated and validated against independent traffic observations.

The reference study by Geng et al. describes a Network Insight Model based on aggregated cellular activity and an Integrated Mobility Model that combines cellular information with Bluetooth observations [1]. The supplied reference article summarizes this framework and reports experiments conducted in York Region, Canada [2]. The present work does not reuse those reported results as local evidence. Instead, it defines a protocol for investigating whether a computationally lighter spatio-temporal graph approach can be evaluated in York Region using locally collected aggregate traffic observations and, subject to authorization, cellular activity.

The exact title is retained at the user's request. The study area, dataset, implementation, and results must nevertheless be demonstrably distinct from the cited work. The title is identical to the supplied reference article

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40875

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 3, November 2026
Impact Factor: 8.2

title; before submission, the authors should check the target venue's policies and consider whether retaining an identical title could cause bibliographic confusion.

II. RELATED WORK

A. Aggregated Mobile-Network Activity

Geng et al. present cellular network activity as a city-scale mobility signal and describe a tensor-based representation of historical activity [1]. The approach is relevant where operator-level aggregation can provide broad spatial coverage. However, access to such data is not automatic, and an academic implementation must document aggregation level, spatial resolution, temporal resolution, consent or lawful basis, retention rules, and any disclosure controls. In the absence of operator-provided data, a study must not label manually counted vehicles or synthetic values as cellular activity.

B. Bluetooth and Road-Sensor Observations

Bluetooth detections can support travel-time estimation when sensors are deployed at known locations and observations can be matched across sites. The reference article cites work on Bluetooth traffic data for urban travel-time forecasting [2], [7]. Such measurements are subject to detection-rate bias and require careful treatment of missing matches. For the proposed York Region pilot, manual classified counts are a feasible initial ground-truth source; they do not, by themselves, provide Bluetooth detections or link-level travel times.

C. Graph-Based Spatio-Temporal Forecasting

Graph neural networks represent road networks through nodes and edges, allowing spatial relationships to be learned alongside temporal dynamics. Graph Convolutional Networks and Graph Attention Networks provide established mechanisms for neighborhood aggregation [3], [4]. Recurrent units such as GRUs can model sequential observations. A3T-GCN combines attention-based temporal processing with graph convolution for traffic forecasting [5]. These methods motivate a compact graph-recurrent baseline for a small urban network, but their performance must be tested on the collected York Region dataset rather than assumed from other cities.

III. RESEARCH OBJECTIVES AND QUESTIONS

The objective of this manuscript is to explain and critically summarize the city-wide traffic-prediction framework and the York Region results reported in the supplied reference. It focuses on the data representations, model architecture, evaluation metrics, and the reported comparison between cellular-only and cellular-plus-Bluetooth configurations. It does not claim to collect new observations or reproduce the experiments.

The contribution of this manuscript is a structured, attributed synthesis of the supplied York Region study, including its reported travel-time and speed metrics, dataset description, and limitations. Any future independent extension must be presented separately and supported by newly collected or lawfully obtained data and independently executed experiments.

IV. STUDY AREA AND DATASET DESIGN

A. Study Area and Junction Sampling

The study area described in the supplied reference is York Region, Ontario, Canada. The source reports using Bluetooth traffic observations covering approximately 1,300 road segments, together with aggregated cellular network activity from cell sites in the region [1], [2]. This manuscript does not claim access to the raw datasets.

B. Road-Network Data

Road geometry and connectivity may be obtained from OpenStreetMap (OSM), subject to its attribution and ODbL requirements [8]. The road network will be represented as a directed graph G=(V,E), where nodes represent selected junctions or observation points and edges represent directed road segments. Each edge may contain road class, length, direction, and connectivity attributes. OSM geometry is a map-derived representation and should be checked against field conditions before analysis.

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 3, November 2026
Impact Factor: 8.2

C. Manual Classified Traffic Counts

The supplied reference describes Bluetooth observations containing vehicle counts, average speed, travel time, origin and destination intersections, travel direction, and road-segment information. These are source-dataset attributes; the present authors have not independently collected or audited the underlying observations [1], [2].

D. Video and Public Records

Video-based counting may be used only when the camera owner or responsible authority has granted permission. The camera should be fixed at a safe location, and analysis should retain aggregate counts rather than identifiable faces or registration plates. Public traffic records may supplement the dataset only after their source, coverage, licensing, and measurement definitions have been verified. District-level statistics must not be treated as junction-level traffic counts.

E. Aggregated Cellular Activity

The source study reports obtaining aggregated counts of unique active devices from a cellular operator API at 15-minute observation intervals. It describes aggregating three five-minute API observations to estimate activity over each 15-minute period. Such data are not included with this manuscript, and no claim of operator access by the present authors is made [1], [2].

V. REFERENCE FRAMEWORK AND MODEL DESCRIPTION

A. Data Alignment and Quality Control

All observations will be converted to a consistent local time zone and aligned to fixed 15-minute bins. Duplicate records, impossible timestamps, missing intervals, observer interruptions, and inconsistent category labels will be flagged. Missing traffic counts will remain explicitly missing during quality assessment; they will not automatically be replaced with zero, because zero traffic and an unobserved interval are different conditions. Any imputation used for model inputs will be fitted using training data only and documented.

B. Feature Construction

For each observed road approach and time step, the traffic feature vector will include classified counts and total count. Calendar features may include hour-of-day, day-of-week, weekend indicator, and public-holiday indicator where verified. If operator data are supplied, aggregated activity features will be aligned to the same time bins and mapped to road-graph nodes using a documented spatial association procedure. Features will be normalized using statistics computed from the training partition only.

C. Road Graph and Lightweight GAT-GRU

The road graph will encode adjacency between connected observation points. Directionality will be preserved where the available road geometry and observations support it. A lightweight Graph Attention Network layer will aggregate neighboring node representations using learned attention coefficients. A GRU will process the sequence of recent graph features to capture temporal dependence. A compact prediction head will produce the next-step traffic-volume estimate for each monitored node or approach. The initial target is 15-minute-ahead total vehicle count; speed and travel-time outputs will be added only if independent, reliable measurements are collected.

For the multimodal experiment, the cellular activity vector will be concatenated with or fused into the node-level traffic representation after spatial and temporal alignment. Two configurations will be evaluated: (i) traffic-observation-only model and (ii) traffic plus authorized aggregated cellular activity. This ablation will isolate whether cellular activity contributes predictive information beyond local counts.

D. Baseline Models

The proposed model will be compared with a historical-average predictor, linear regression, Random Forest, XGBoost, and a GRU without graph attention. If computational resources and data volume permit, an established graph-temporal baseline such as A3T-GCN will also be implemented. All models will receive equivalent input histories and be evaluated on identical chronological test intervals.

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 3, November 2026
Impact Factor: 8.2

VI. SOURCE DATASET AND EXPERIMENTAL SETUP

A. Data Sources Reported in the Reference

The supplied reference reports a York Region dataset collected from December 2022 to April 2023. Bluetooth traffic observations were available at five-minute intervals and covered approximately 1,300 road segments. Aggregated cellular activity was obtained from a cellular operator API at 15-minute observation intervals [1], [2].

B. Experimental Design Reported in the Reference

The supplied PDF describes one-step-ahead forecasting for the next five-minute interval. It reports model training with a learning rate of 0.0001, batch size 8, and 500 epochs on a GPU-enabled server. The uploaded six-page summary does not provide enough detail to independently reconstruct every data split or training run; these details should be checked against the full original article before attempting replication [1], [2].

C. Evaluation Metrics

The source study reports MAE, RMSE, coefficient of determination (R^2), and MAPE for travel-time and speed prediction, and uses the Travel Time Index (TTI) to describe congestion. Metric units and zero-value handling should be interpreted using the original article's definitions [1].

TABLE I. EVALUATION METRICS USED IN THE SOURCE STUDY

Metric	Definition	Interpretation
MAE	Mean absolute prediction error	Lower is better; reported in the target variable's unit.
RMSE	Square root of mean squared error	Lower is better; reported in the target variable's unit.
WAPE	Sum of absolute errors divided by sum of actual counts	Lower is better; useful when some intervals have zero counts.
R^2	Proportion of test-set variance explained	Closer to 1 is better; interpret with sample size and variance.

VII. REPORTED RESULTS AND DISCUSSION

The numerical values in this section are transcribed from the supplied six-page IJARSCT PDF and attributed to the underlying York Region study by Geng et al. [1], as reproduced in the supplied document [2]. They are not results generated by the present authors. The supplied PDF describes data collected in York Region from December 2022 to April 2023, including Bluetooth observations at five-minute intervals over approximately 1,300 road segments and operator-provided aggregated cellular activity at 15-minute intervals [2].

A. Travel-Time Prediction. The supplied reference reports that the Network Insight Model achieved MAE = 15.55, RMSE = 27.70, R^2 = 0.83, and MAPE = 14.79%. The Integrated Mobility Model achieved MAE = 13.03, RMSE = 22.97, R^2 = 0.88, and MAPE = 12.29%. Random Forest recorded MAE = 18.54, RMSE = 35.32, R^2 = 0.75, and MAPE = 17.75%. The PDF states that Linear Regression and XGBoost produced MAE values above 20, RMSE values above 37, and MAPE values above 20%; it does not provide exact values for those two baselines [2].

TABLE II. REFERENCE-STUDY TRAVEL-TIME RESULTS (NOT OUR EXPERIMENTS)

Model (source study)	MAE	RMSE	R^2	MAPE
Network Insight Model	15.55	27.70	0.83	14.79%
Integrated Mobility Model	13.03	22.97	0.88	12.29%
Random Forest	18.54	35.32	0.75	17.75%
Linear Regression	>20	>37	Not stated	>20%
XGBoost	>20	>37	Not stated	>20%

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40875

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 3, November 2026
Impact Factor: 8.2

B. SPEED PREDICTION

B. Speed Prediction. The reference reports MAE = 6.52, RMSE = 8.72, R^2 = 0.70, and MAPE = 14.26% for the Network Insight Model. The Integrated Mobility Model records MAE = 5.71, RMSE = 7.72, R^2 = 0.76, and MAPE = 12.44%. Random Forest records MAE = 11.56, RMSE = 15.44, R^2 = 0.65, and MAPE = 18.15%. A3T-GCN records MAE = 8.10, RMSE = 11.20, R^2 = 0.36, and MAPE = 19.78% [2].

TABLE III. REFERENCE-STUDY SPEED-PREDICTION RESULTS (NOT OUR EXPERIMENTS)

Model (source study)	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%
A3T-GCN	8.10	11.20	0.36	19.78%

C. Additional Findings Reported by the Source.

The supplied PDF describes attention-based directional relationships between selected intersection pairs and reports a morning congestion peak around 07:00–09:00 and an evening peak around 15:00–18:00, with TTI values reported up to 1.83. It also reports that the Integrated Mobility Model performed better during the examined holiday and rush-hour periods. These are findings from the source study and are not independently verified in this manuscript [1], [2].

The study will collect aggregate traffic counts and road-network attributes. No personal identity, face, vehicle registration number, or individual mobile-device identifier is required for the proposed analysis. Any video collection requires permission and a documented retention/deletion policy. Cellular data must be supplied in aggregated form under an authorized agreement and must meet the provider's privacy and minimum-group-size safeguards. Data provenance, collection dates, observation protocols, code versions, preprocessing decisions, and model configurations should be recorded so the experiment can be reproduced.

IX. LIMITATIONS AND FUTURE WORK

A single-observer manual survey is resource-efficient but may be affected by fatigue, occlusion, dense mixed traffic, and classification error. A small number of junctions and short collection periods may limit generalization across the city and across seasons. Cellular activity may be unavailable, spatially coarse, or weakly correlated with road traffic; its predictive value cannot be assumed. Future work may expand the number of junctions, introduce independent observer validation, collect travel-time measurements, examine event and weather effects, and evaluate transferability to other Canadian cities.

X. CONCLUSION

The supplied reference describes a city-wide traffic-prediction framework that uses aggregated cellular network activity and, in its multimodal configuration, Bluetooth sensor observations. The reported York Region results indicate that the Integrated Mobility Model outperformed the Network Insight Model and the listed baselines on the reported travel-time and speed metrics [1], [2]. These findings are summaries of the cited study and must not be interpreted as new experiments by the present authors. A genuinely original extension would require independently obtained York Region data, a documented implementation, a leakage-free evaluation, and newly computed test results. The current document is suitable as a source-based technical review draft, not as an original experimental research paper.

ACKNOWLEDGMENT

The authors acknowledge the researchers whose published work is analyzed in this source-based review. No direct data contribution or institutional support is claimed in this manuscript.

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 3, November 2026
Impact Factor: 8.2

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. [Bibliographic details reproduced from the supplied reference article; verify against the publisher record before submission.]

[2] P. C. Kute, M. S. Dighe, P. A. Sawant, and H. B. Jadhav, "City-Wide Traffic Prediction using Aggregated Mobile Network Activity Data," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 6, no. 1, Nov. 2026, doi: 10.48175/IJARSCT-40527. [Details transcribed from the supplied PDF; verify publication metadata.]

[3] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. International Conference on Learning Representations, 2017. [Verify final proceedings metadata.]

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

[5] J. Bai et al., "A3T-GCN: Attention temporal graph convolutional network for traffic forecasting," ISPRS International Journal of Geo-Information, vol. 10, no. 7, 2021. [Verify page/article number and DOI.]