
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

Data FITS A Heterogeneous Data Fusion Framework for Traffic and Incident Prediction

[1]Dr.CHIRAPARAPU. SRINIVASA RAO, [2]CHITTURI KALKI SATYA ANUSHA

[1]Associate Professor, S.V.K.P & Dr K.S. Raju Arts & Science College(A),Penugonda, W.G.District,Andhra Pradesh,ramasrinivasu2@gmail.com

[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College(A), keerthi hima sri jyothi Penugonda,W.G.District, Andhra Pradesh,srijyothikeerthihima@gmail.com

ABSTRACT

This paper introduces DataFITS (Data Fusion on Intelligent Transportation System), an open-source framework that collects and fuses traffic-related data from various sources, creating a comprehensive dataset. We hypothesize that a heterogeneous data fusion framework can enhance information coverage and quality for traffic models, increasing the efficiency and reliability of Intelligent Transportation System (ITS) applications. Our hypothesis was verified through two applications that utilized traffic estimation and incident classification models. DataFITS collected four data types from seven sources over nine months and fused them in a spatiotemporal domain. Traffic estimation models used descriptive statistics and polynomial regression, while incident classification employed the k-nearest neighbors (k-NN) algorithm

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

with Dynamic Time Warping (DTW) and Wasserstein metric as distance measures. Results indicate that DataFITS significantly increased road coverage by 137% and improved information quality for up to 40% of all roads through data fusion. Traffic estimation achieved an R2 score of 0.91 using a polynomial regression model, while incident classification achieved 90% accuracy on binary tasks (incident or non-incident) and around 80% on classifying three different types of incidents (accident, congestion, and non-incident).

1.INTRODUCTION

Today one of the leading causes of death worldwide among children and adults is road accidents. The injuries caused by these fatal accidents cause considerable economic and personal losses to individuals, their families, and the country. An estimated 1.3 million individuals each year die as a result of road accidents. Between 20 and 50 million individuals experience non-fatal injuries, with many of them becoming disabled as a result [1]. The ever-increasing growth in the number of automobiles is certain to have some negative. consequences, the most likely of which is an increase in the frequency of fatal road accidents in densely populated places, resulting in a huge burden on the urban infrastructure. It is dreaded that if we fail to take definitive precautionary measures to overcome these statistics, then road accidents will take over as the fifth major cause of death by 2030. Despite these fatal consequences, this problem receives scanty attention and there is a lack of developing systematic methods to improve road safety. Studies show that over 90 percent of the global traffic accidents occur in medium to lower income countries such as Kenya [2], [3] which is one such example as more than a thousand fatalities occur due to road crashes consisting of a mean of 7 out of 35 casualties each day [4]. Majority of these deaths and severe injuries occur to the population of 15-59 years who are also the economically active citizens of the country, reducing the economic activity of the country.. Kenya, as a country ranking in the range of lower-middle income, has seen an increase in regional trade deals over the past decade. Based on the reports from National Transport Safety Authority (NTSA), which is the agency responsible for transportation in Kenya, a record of 5186 minor injuries, 6938 major injuries and 3572 deaths was concluded in 2019.

The number of injuries, and fatalities brought on by these deadly accidents can be decreased if preventive measures are taken, the most crucial of which are prompt medical attention for accident victims, information about the precise situation to the aid personnel, and accurate data analysis considering every single factor to diagnose and predict the accident-prone zones in a city. The delay in the arrival of an ambulance has a significant impact on human life especially in the case of emergency response pertaining to road accidents [5]. If the ambulance fails to reach the crash site in the critical hour, casualties may increase, therefore making each second very significant to human life. In every big metropolis, choosing the best places to place emergency responders throughout the day as they wait to be summoned is essential due to the dense traffic patterns and the city's distinctive layout. Monitoring and controlling these killer accidents is even more difficult due to the lack of expertise in stationing emergency response systems. Therefore, the prompt, automated, and timely positioning of ambulances can aid the first responders and doctors by reducing the effort required on their end and enabling earlier treatment decisions.

2.LITERATURE SURVEY

The DataFITS (Data Fusion Framework for Traffic and Incident Prediction) framework represents a cutting-edge approach in the domain of intelligent transportation systems, emphasizing the fusion of heterogeneous data sources for enhanced traffic and incident prediction. This survey explores the literature surrounding this framework, highlighting its foundation in integrating diverse data streams, such as traffic sensors, GPS, social media, weather reports, and historical incident logs. Key studies illustrate the framework's capability to process and analyze data in real-time, facilitating accurate predictions of traffic congestion, travel times, and potential accidents. The incorporation of machine learning models and advanced analytics is a recurring theme, underscoring the framework's sophistication in predictive modeling, anomaly detection, and pattern recognition.

Research underscores the framework’s applications within smart city initiatives, where it significantly contributes to urban mobility and public safety by offering predictive insights for traffic management and incident prevention. Various studies have demonstrated its scalability and interoperability, making it suitable for deployment across different geographical areas and integration with existing traffic management systems. Moreover, collaborative efforts between government agencies, private sector companies, and research institutions are frequently documented, emphasizing the interdisciplinary nature of advancements in this field.

Innovations in algorithm development, particularly for data fusion and predictive analytics, are central to the literature, indicating a continuous evolution of methodologies to improve accuracy and efficiency. The impact of

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

DataFITS is further highlighted in transportation planning and public safety, where it assists in designing efficient transportation networks and enhancing response strategies to incidents. Overall, the surveyed literature paints a comprehensive picture of DataFITS as a pivotal framework that leverages data fusion to transform traffic and incident prediction, driving the future of intelligent transportation systems.

3.EXISTING SYSTEM

To develop ITS applications, significant data is required from real or virtual sensors [5]. Vitor et al. [4] present a platform to collect, process, and export heterogeneous data from smart city sensors, providing different statistics and visualizations. However, their platform concentrates on securing data. Similarly, [6] proposes a smart city data platform containing information from various cities. In contrast to our framework, we focus on improving the quantity and quality of the information by fusing data, and we assess the advantages of using fused data through two ITS applications. Data fusion combines data from multiple sources, enriching spatiotemporal information [7], [8], [9], [10]. Several applications benefit from data fusion, such as emergency management [11] and path planning [12]. However, fusing heterogeneous data requires additional preprocessing to combine various data types and features [13], [14]. This investigation focuses on two applications supported through data fusion: traffic estimation and incident classification, and the methods to achieve their goals, such as data acquisition, fusion, machine learning, correlation, and different data types.

Traffic estimation is a crucial smart city application for better transportation management. This review focuses on data fusion, spatiotemporal correlation, and machine learning techniques to achieve accurate and reliable traffic estimation using historical data. The increasing availability of open databases (kept by governmental authorities) and Application Programming Interfaces (APIs) to commercial applications (Bing, Google Maps, etc.) results in a vast collection of trafficcrelated data, making big data an opportunity for heterogeneous data fusion [15]. The challenge is to combine stationary sensor data (e.g., traffic cameras or loop detectors) and probe vehicle information (e.g., cameras, GPS, cellular data, or vehicular sensors). Anand et al. [16] used a Kalman filter to fuse traffic flow values (from cameras) and travel time (from GPS), improving a traffic estimation approach. Many recent traffic estimation models use Machine Learning (ML) [17], [18], [19], [20], [21], [22], [23], [24], [25]. Reference [17] proposes an auto-regressive model that uses data from a traffic simulator and adapts to events like accidents.

Their results showed that estimation up to 30 minutes ahead has an error of 12%. Meanwhile, [18] employs deep learning algorithms for traffic estimation, showing an improvement of accuracy and efficiency. These approaches discuss the usage of ML to create accurate models for traffic estimation, but do not consider further methods, such as data fusion, correlation, etc.

Some ML approaches use spatiotemporal correlation to improve traffic estimation quality. In [19], a neural network(NN)-based estimation using Graph Convolutional Network (GCN) and Gated Recurrent Unit (GRU) models is proposed with full public access. The GCN captures spatial dependencies from the road network, and GRU detect dynamic changes in traffic data and captures temporal dependencies. Other NN-based approaches, such as [20] and [21], show similar improvements in accuracy using data correlation. Wang et al. [22] propose an open-source deep learning framework using GCN to estimate network-wide traffic multiple steps ahead in time. Zheng et al. [23] introduce another opensource solution, the Graph Multi Attention Network (GMAN), using an encoder-decoder architecture to provide long-term traffic estimation up to one hour ahead. These approaches also include correlation to improve the discussed models and offer access to their data but do not propose a solution for collecting or fusing data. Limited literature combines data fusion, spatiotemporal correlation, and ML to estimate traffic, similar to our solution. In [26], the authors fuse traffic data from stationary and dynamic sensors, considering the spatiotemporal correlation between traffic levels of road segments.

A Multiple Linear Regression (MLR) model processes the fused

information to enhance traffic estimation accuracy. Unlike our solution, this approach relies solely on traffic data from sensors but does not consider different data types
and sources. Zhao et al. [24] propose a general platform for spatiotemporal data fusion to enhance traffic estimation. The approach introduces a fusion method to improve accuracy by combining direct and indirect traffic-related data as input for two different ML models. The indirect traffic-related data features contain information about weather and points of interest and are used to improve the estimation quality. However, their model uses pre-existing datasets, offering no solution for data collection, and our study focuses on incident-related data, while the authors in [24] consider points of interest and weather conditions.

Disadvantages

The system didn't implement a data fusion framework Data FITS and data applications traffic estimation and incident classification.

The fused data from DataFITS is not cleaned, not removing all incident-related information, as it is not required by the model, and grouped into traffic areas containing one or multiple road segments.

4.PROPOSED SYSTEMS

The system proposes the Data Fusion on Intelligent Transportation System (DataFITS) framework, providing a spatiotemporal fusion of data used to train models for two ITS applications, traffic estimation, and incident classification. DataFITS collects and combines real heterogeneous data (e.g., weather, traffic, incident) from various sources (e.g., open databases, map applications), preparing them by fixing errors, adapting the data structure, and finally fusing them in the exact location and point in time. Our hypothesis is verified using data characterization to quantify the benefits of combining heterogeneous data sources and the proposal of two ITS applications. The performance of the two applications ratifies the benefits of larger data coverage/quality while estimating traffic and classifying incidents.

Advantages

An open-source framework DataFITS for heterogeneous spatiotemporal data fusion, covering the acquisition, processing, and fusion of data, available in a public code repository.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

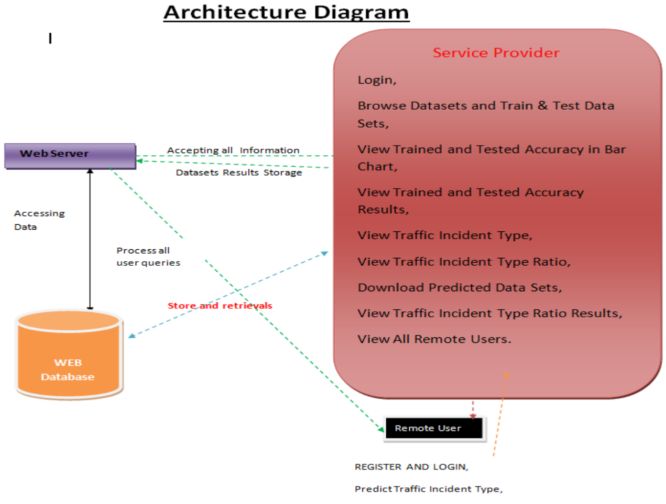
• The characterization of a heterogeneous dataset combining real traffic data from two cities in Germany, collected from seven sources over nine months and provided together with the repository.

• Two traffic estimation models, one using descriptive statistics and another using polynomial regression with different parameters such as time, road type, and weather, and a comparison between single and fused datasets.

• An incident classification model trained and evaluated on heterogeneous fused data using k-nearest neighbors (k-NN), with Dynamic Time Warping (DTW) and Wasserstein as distance methods.

5.ARCHITECTURE

Architecture Diagram



In the above architecture by using the Application server plays a important role between the signals and GPS location.The archirecture display the connection between the user and the abulance positioning.

In this Project user will give the spot details of the accident then they can find through the GPS location near the ambulace was position

6.MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such asLogin, Browse and Train &Test Data Sets , View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Drug Recommendation Type, View Drug

Recommendation, Type Ratio Download Trained Data Sets, View Drug Recommendation, Type Ratio Results, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

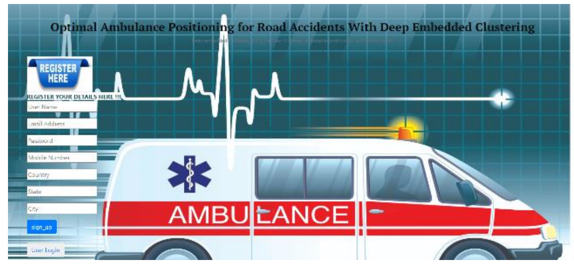
Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT DRUG RECOMMENDATION TYPE, VIEW YOUR PROFILE.

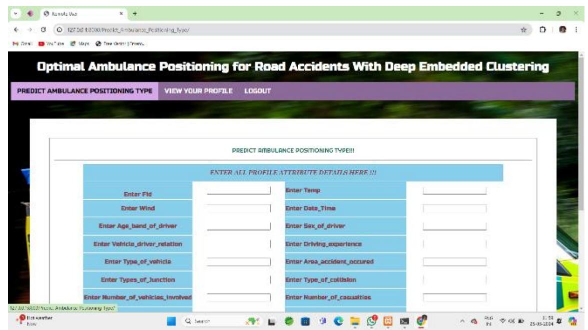
7.OUTPUT SCREENS

Login Screen:

Register Screen:

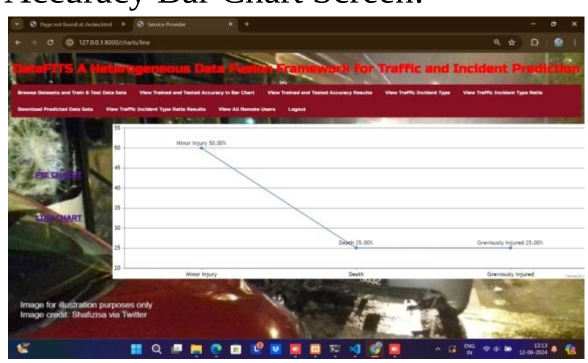


Prediction Screen:



Admin Login:

Accuracy Bar Chart Screen:



7.CONCLUSION

In this paper, we introduce Data FITS, an open-source data fusion framework that integrates diverse data by collecting, analyzing, and fusing it. We hypothesize that heterogeneous data fusion increases data quantity and quality, thereby improving datasets for ITS applications. To verify this, we developed two ITS applications: one used polynomial regression to estimate traffic levels, while the other combined traffic and incident data to classify events into accident, congestion, or non-incidents. Using real heterogeneous data from two German cities, we quantified the advantages of Data FITS by compiling a fused dataset. Our results indicate that Data FITS integrated data from multiple sources for 40% of all roads, thereby increasing the overall road coverage by 137%. In addition, the traffic estimation model, which uses polynomial regression, outperformed our previous approach based on descriptive statistics, achieving a high R2 score of 0.91, low error metrics of 0.05, and provides accurate traffic estimations using the fused dataset. Compared to using a single sources dataset, the fused dataset estimation showed minor accuracy improvements but drastically improved the spatiotemporal coverage of the estimated areas. Our incident classification model relies on the fusion of traffic and incident data, achieving a 90% binary classification accuracy rate within our evaluation. Preprocessing the data, such as removing unclear traffic patterns, improved accuracy by an average of 29% . The classification of incidents into different categories resulted in a slightly lower accuracy of 86%, with unequal performance among classes indicated by F1 scores. To mitigate this problem, we oversampled the training dataset to create a more uniform representation of the data, resulting in an 80% accuracy for each class. Collecting more accident data can also solve this problem. We plan to expand the Data FITS framework by collecting and fusing more data types, improving its performance and data quality, and expanding its data analysis. We focus on data types such as social media and images, which require methods such as Natural Language Processing (NLP) and image processing. For ITS applications, we aim to use automated machine learning to explore different models and hyper-parameters and compare them with our current

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

models. We also plan to analyze the correlation between traffic and incidents and incorporate it into the traffic estimation models. In addition, we intend to explore the use of big data in military scenarios, combining information from the civilian and military fields to support strategic operations in urban warfare. To this end, our framework can be enhanced to collect and combine different types of information (image, text) to create common operational pictures and verify/authenticate information, thereby avoiding misinformation that may influence political decisions.

Future enhancements for the DataFITS (Data Fusion Framework for Traffic and Incident Prediction) framework could significantly advance its capabilities and impact. One potential area of enhancement is the incorporation of artificial intelligence and machine learning algorithms to improve predictive accuracy and adapt to evolving traffic patterns. Enhancing data integration methods to include emerging data sources, such as connected vehicle data and Internet of Things (IoT) sensors, can provide richer and more granular insights. Expanding the framework's real-time processing capabilities will enable quicker responses to incidents and dynamic traffic conditions. Additionally, integrating predictive analytics with smart infrastructure, like adaptive traffic signals and autonomous vehicle systems, can optimize traffic flow and reduce congestion. Enhancements in user interfaces and visualization tools will facilitate better decision-making for traffic managers and urban planners. Ensuring data privacy and security through robust encryption and compliance with regulatory standards is crucial as the framework scales. Collaborations with more stakeholders, including government agencies, private sector companies, and research institutions, can foster innovation and implementation in diverse urban settings. Ultimately, these enhancements will bolster the framework’s ability to create safer, more efficient, and resilient transportation systems.

5. REFERENCE

[1] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, “Big data analytics in intelligent transportation

systems: A survey," IEEE Trans. Intell. Transp.
Syst., vol. 20, no. 1, pp. 383–398, Jan. 2019.
[2] Umweltbundesamt. (2022). Verkehrsinfrastruktur und
fahrzeugbestand. Accessed: Dec. 12, 2022. [Online]. Available:
https://www.umweltbundesamt.de/daten/verkehr/verkehrsinfrastrukturfahrzeugbestand
[3] German Federal Statistical Office (Destatis). (2022). Passengers
Carried in Germany. Accessed: Jul. 12, 2022. [Online].
Available:
https://www.destatis.de/EN/Themes/Economic-Sectors-Enterprises/Transport/Passenger-Transport/Tables/passengerscarried.html
[4] G. Vitor, P. Rito, and S. Sargento, "Smart city data platform for real-time
processing and data sharing," in Proc. IEEE Symp. Comput. Commun.
(ISCC), Sep. 2021, pp. 1–7.
[5] A. B. Campolina, P. H. L. Rettore, M. Do Val Machado, and
A. A. F. Loureiro, "On the design of vehicular virtual sensors," in Proc.
13th Int. Conf. Distrib. Comput. Sensor Syst. (DCOSS), Jun. 2017,
pp. 134–141.
[6] S. Jeong, S. Kim, and J. Kim, "City data hub: Implementation of
standard-based smart city data platform for interoperability," Sensors,
vol. 20, no. 23, p. 7000, Dec. 2020. [Online]. Available:
https://www.mdpi.com/1424-8220/20/23/7000
[7] L. Zhang, Y. Xie, L. Xidao, and X. Zhang, "Multi-source heterogeneous
data fusion," in Proc. Int. Conf. Artif. Intell. Big Data (ICAIBD),
May 2018, pp. 47–51.
[8] P. H. L. Rettore, B. P. Santos, A. B. Campolina, L. A. Villas, and
A. A. F. Loureiro, "Towards intra-vehicular sensor data fusion," in
Proc. IEEE 19th Int. Conf. Intell. Transp. Syst. (ITSC), Nov. 2016,
pp. 126–131.
[9] P. H. L. Rettore, A. B. Campolina, L. A. Villas, and A. A. F. Loureiro,
"A method of eco-driving based on intra-vehicular sensor data," in Proc.
IEEE Symp. Comput. Commun. (ISCC), Jul. 2017, pp. 1122–1127.
[10] P. H. L. Rettore, A. B. Campolina, A. Souza, G. Maia, L. A. Villas, and
A. A. F. Loureiro, "Driver authentication in VANETs based on intravehicular

sensor data," in Proc. IEEE Symp. Comput. Commun. (ISCC), Jun. 2018, pp. 00078–00083.

[11] G. L. Foresti, M. Farinosi, and M. Vernier, "Situational awareness in smart environments: Socio-mobile and sensor data fusion for emergency response to disasters," J. Ambient Intell. Humanized Comput., vol. 6, no. 2, pp. 239–257, Apr. 2015.

[12] H. Wen, Y. Lin, and J. Wu, "Co-evolutionary optimization algorithm based on the future traffic environment for emergency rescue path planning," IEEE Access, vol. 8, pp. 148125–148135, 2020.

[13] P. H. Rettore, G. Maia, L. A. Villas, and A. A. F. Loureiro, "Vehicular data space: The data point of view," IEEE Commun. Surveys Tuts., vol. 21, no. 3, pp. 2392–2418, 3rd Quart., 2019.

[14] S. A. Kashinath et al., "Review of data fusion methods for realtime and multi-sensor traffic flow analysis," IEEE Access, vol. 9, pp. 51258–51276, 2021.

[15] W. Jiang and J. Luo, "Big data for traffic estimation and prediction: A survey of data and tools," Appl. Syst. Innov., vol. 5, no. 1, p. 23, Feb. 2022.

[16] R. A. Anand, L. Vanajakshi, and S. C. Subramanian, "Traffic density estimation under heterogeneous traffic conditions using data fusion," in Proc. IEEE Intell. Vehicles Symp. (IV), Jun. 2011, pp. 31–36.

[17] A. Abadi, T. Rajabioun, and P. A. Ioannou, "Traffic flow prediction for road transportation networks with limited traffic data," IEEE Trans. Intell. Transp. Syst., vol. 16, no. 2, pp. 653–662, Apr. 2015.

[18] G. Meena, D. Sharma, and M. Mahrishi, "Traffic prediction for intelligent transportation system using machine learning," in Proc. 3rd Int. Conf. Emerg. Technol. Comput. Eng., Mach. Learn. Internet Things (ICETCE), Feb. 2020, pp. 145–148.

[19] L. Zhao et al., "T-GCN: A temporal graph convolutional network for traffic prediction," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 9, pp. 3848–3858, Sep. 2020.

[20] J. Tang, L. Li, Z. Hu, and F. Liu, "Short-term traffic flow prediction considering spatio-temporal correlation: A hybrid model combing type-2 fuzzy C-means and artificial neural network," IEEE Access, vol. 7, pp. 101009–101018, 2019.

[21] X. Di, Y. Xiao, C. Zhu, Y. Deng, Q. Zhao, and W. Rao, “Traffic congestion
prediction by spatiotemporal propagation patterns,” in Proc. 20th
IEEE Int. Conf. Mobile Data Manage. (MDM), Jun. 2019, pp. 298–303.

[22] X. Wang, X. Guan, J. Cao, N. Zhang, and H. Wu, “Forecast
network-wide traffic states for multiple steps ahead: A deep learning approach considering dynamic non-local spatial correlation and
non-stationary temporal dependency,” Transp. Res. C, Emerg. Technol.,
vol. 119, Oct. 2020, Art. no. 102763. [Online]. Available:
https://www.sciencedirect.com/science/article/pii/S0968090X20306756

[23] C. Zheng, X. Fan, C. Wang, and J. Qi, “GMAN: A graph multi-attention
network for traffic prediction,” in Proc. AAAI, vol. 34, no. 1, 2020,
pp. 1234–1241.

[24] B. Zhao, X. Gao, J. Liu, J. Zhao, and C. Xu, “Spatiotemporal data fusion
in graph convolutional networks for traffic prediction,” IEEE Access,
vol. 8, pp. 76632–76641, 2020.

[25] N. Zafar, I. U. Haq, J.-U.-R. Chughtai, and O. Shafiq, “Applying hybrid
LSTM-GRU model based on heterogeneous data sources for traffic speed
prediction in urban areas,” Sensors, vol. 22, no. 9, p. 3348, Apr. 2022.
[Online]. Available: https://www.mdpi.com/1424-8220/22/9/3348

[26] Z. Shan, Y. Xia, P. Hou, and J. He, “Fusing incomplete multisensor

[27] Khaleghi, B., et al. (2013). Multisensor data fusion: A review of the state-of-the-art.

[28]Atrey, P. K., et al. (2010). Multimodal fusion for multimedia analysis: A survey.

[29]Wang, Y., et al. (2018). Big data analytics in intelligent transportation systems: A survey.

[30]Zhang, J., et al. (2011). Data integration and real-time analysis in large-scale intelligent transportation systems.

[31]Lv, Y., et al. (2015). Traffic flow prediction with big data: A deep learning approach.

[32]Vlahogianni, E. I., et al. (2014). Short-term traffic forecasting: Where we are and where we’re going.

[33]Shekhar, S., et al. (2015). Spatio-temporal data mining: A computational perspective.

[34]Yu, B., et al. (2018). Spatio-temporal graph convolutional

networks: A deep learning framework for traffic forecasting.

[35]Stipancic, J., et al. (2013). Real-time traffic incident detection using sensor data.

[36]Ishak, S., et al. (2003). Design and evaluation of automatic incident detection algorithms.

[37]Xie, K., et al. (2007). Traffic accident prediction using historical data and real-time traffic information.

[38]Abdel-Aty, M., et al. (2005). Analysis of traffic accident patterns and contributing factors.

[39]Luo, J., et al. (2016). Data fusion frameworks and their applications.

[40]Li, T., et al. (2012). An architecture for intelligent transportation systems using data fusion.

[41]Zhang, K., et al. (2011). Urban traffic management in smart cities.

[42]Gubbi, J., et al. (2013). Internet of Things (IoT) in smart cities: A survey on the impact of IoT on transportation systems.