# INTEGRATED TRAFFIC INSIGHT SYSTEM: PREDICTIVE ANALYSIS AND INCIDENT MONITORING

Kamireddy uday kiran[1],Santhosh kasam[2],Pagidipathi srinivas goud[3]

[1]Associate Professor,M.Tech.,BRILLIANT GRAMMAR SCHOOL EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS-INTEGRATED CAMPUS

Abdullapurmet (v), hayath nagar (m), r.r dt. Hyderabad

[2]Associate Professor,M.Tech.,BRILLIANT GRAMMAR SCHOOL EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS-INTEGRATED CAMPUS

Abdullapurmet (V), Hayath Nagar (M), R.R Dt. Hyderabad

Department of CSE,

[3]UG Students, BRILLIANT GRAMMAR SCHOOL EDUCATIONAL SOCIETY'S GROUP OF INSTITUTIONS-INTEGRATED CAMPUS

Abdullapurmet (V), Hayath Nagar (M), R.R Dt. Hyderabad

## 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 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).

## I. INTRODUCTION

DATA availability is a critical aspect in the design of modern Intelligent Transportation Systems

(ITSs), which implement models to understand better various patterns of the

transportation system [1], thus improving mobility and safety for

people and goods. With modern society depending heavily on efficient and reliable transportation, the importance of these systems has seen a rapid increase in significance over recent years. In Germany alone, both the number of registered cars and the number of carried passengers using public transportation have shown a substantial increase, reaching their all-time highs of 48.5 million cars (2022) and 12.7 billion carried passengers (2019, before the pandemic) [2], [3]. As a result, urban areas experience an increasing number of traffic-related incidents (e.g., congestion and accidents), increasing time delays, emissions, and fuel consumption [4].

For this reason, academia and industry have driven efforts to create the next generation of transportation systems that are eco-friendly, cost-efficient, and powered by data analysis and communication technology. We hypothesize that a heterogeneous data fusion framework can enhance the coverage and quality of information serving as input for traffic models, thus increasing the efficiency and reliability of ITS applications. Therefore, we propose the Data Fusion on Intelligent Transportation System (Data FITS) framework, providing a spatiotemporal fusion of data used to train models for two ITS applications, traffic estimation, and incident classification. Data FITS 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. Thus, the main contributions of this investigation are:

• An open-source framework Data FITS for heterogeneous spatiotemporal data fusion, covering the acquisition, processing, and fusion of data, available in a public code repository.1

• 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.

The rest of the paper is organized as follows. Section II reviews recent literature using data fusion to design applications like traffic estimation and incident classification and compares them against our solution. The design of Data FITS and the traffic data applications are described in Section III. Section IV evaluates the performance of our framework and the effectiveness of our traffic estimation and incident classification models using the heterogeneous fused data, verifying our hypothesis. Finally, we conclude this paper in Section V, highlighting open problems for future investigations.

**II. LITERATURE REVIEW**

DataFITS: A Heterogeneous Data Fusion Framework for Traffic and Incident Prediction ,Philipp Zißner , Paulo H. L. Rettore , Bruno P. Santos, Johannes F. Loevenich , and Roberto Rigolin F. Lopes , *Member, IEEE,*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 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).

**III.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 trafficrelated 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.

## IV.PROPOSED SYSTEM

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.

• 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.

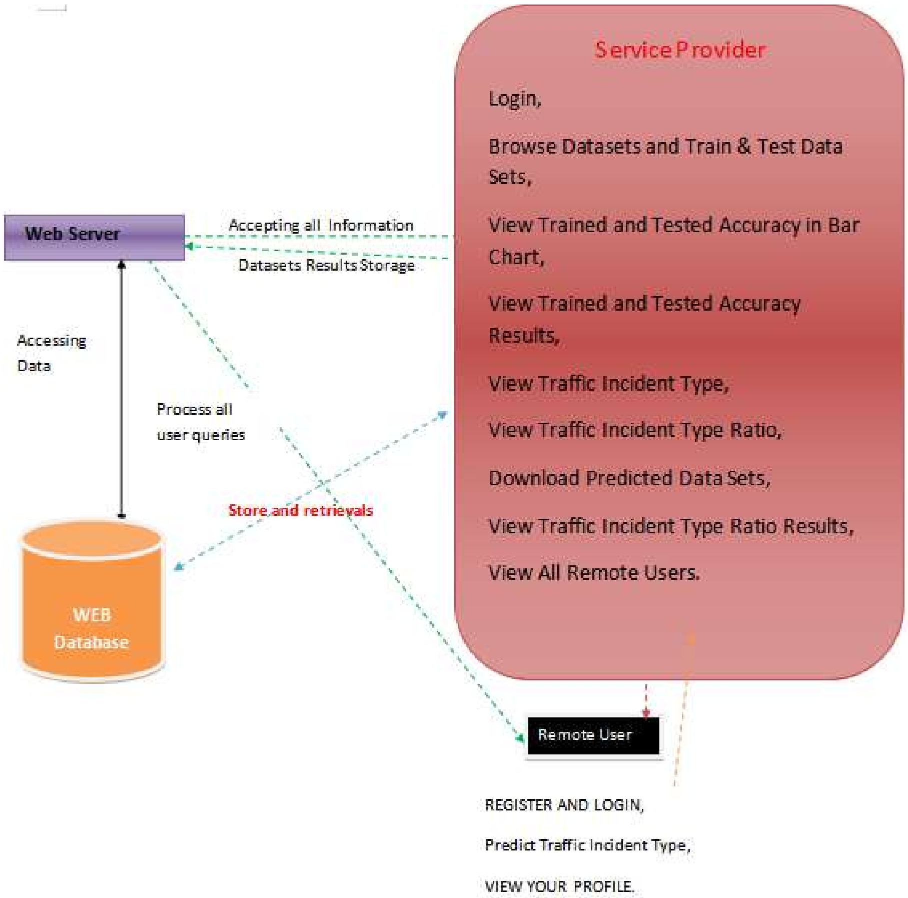


Fig: Architecture diagram

## V.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 as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Traffic Incident Type, View

Traffic Incident Type Ratio, Download Predicted Data Sets, View

Traffic Incident 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 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 traffic incident type, view your profile.

## VI. ALGORITHAMS:

### Decision tree classifiers

Decision tree classifiers are used successfully in many diverse areas. Their most important feature is the capability of capturing descriptive decision making knowledge from the supplied data. Decision tree can be generated from training sets. The procedure for such generation based on the set of objects (S), each belonging to one of the classes C1, C2, …, Ck is as follows:

Step 1. If all the objects in S belong to the same class, for example Ci, the decision tree for S consists of a leaf labeled with this class

Step 2. Otherwise, let T be some test with possible outcomes O1, O2,…, On. Each object in S has one outcome for T so the test partitions S into subsets S1, S2,… Sn where each object in Si has outcome Oi for T. T becomes the root of the decision tree and for each outcome Oi we build a subsidiary decision tree by invoking the same procedure recursively on the set Si.

### Gradient boosting

Gradient boosting is a machine learning technique used in regression and classification tasks, among others. It gives a prediction model in the form of an ensemble of weak prediction models, which are typically decision trees.[1][2] When a decision tree is the weak learner, the resulting algorithm is called gradient-boosted trees; it usually outperforms random forest.A gradient-boosted trees model is built in a stage-wise fashion as in other boosting methods, but it generalizes the other methods by allowing optimization of an arbitrary differentiable loss function.

**K-Nearest Neighbors (KNN)**

- Simple, but a very powerful classification algorithm
- Classifies based on a similarity measure
- Non-parametric
- Lazy learning
- Does not "learn" until the test example is given
- Whenever we have a new data to classify, we find its K-nearest neighbors from the training data

Example

- Training dataset consists of k-closest examples in feature space
- Feature space means, space with categorization variables (non-metric variables)
- Learning based on instances, and thus also works lazily because instance close to the input vector for test or prediction may take time to occur in the training dataset

**Logistic regression Classifiers**

*Logistic regression analysis* studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The name *logistic regression* is used when the dependent variable has only two values, such as 0 and 1 or Yes and No. The name *multinomial logistic regression* is usually reserved for the case when the dependent variable has three or more unique values, such as Married, Single, Divorced, or Widowed. Although the type of data used for the dependent variable is different from that of multiple regression, the practical use of the procedure is similar.

Logistic regression competes with discriminant analysis as a method for analyzing categorical-response variables. Many statisticians feel that logistic regression is more versatile and better suited for modeling most situations than is discriminant analysis. This is because logistic regression does not assume that the independent variables are normally distributed, as discriminant analysis does.

This program computes binary logistic regression and multinomial logistic regression on both numeric and categorical independent variables. It reports on the regression equation as well as the goodness of fit, odds ratios, confidence limits, likelihood, and deviance. It performs a comprehensive residual analysis including diagnostic residual reports and plots. It can perform an independent variable subset selection search, looking for the best regression model with the fewest independent variables. It provides confidence intervals on predicted values and provides ROC curves to help determine the best cutoff point for classification. It allows you to validate your results by automatically classifying rows that are not used during the analysis.

**Naïve Bayes**

The naive bayes approach is a supervised learning method which is based on a simplistic hypothesis: it assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature .

Yet, despite this, it appears robust and efficient. Its performance is comparable to other supervised learning techniques. Various reasons have been advanced in the literature. In this tutorial, we highlight an explanation based on the representation bias. The naive bayes classifier is a linear classifier, as well as linear discriminant analysis, logistic regression or linear SVM (support vector machine). The difference lies on the method of estimating the parameters of the classifier (the learning bias).

While the Naive Bayes classifier is widely used in the research world, it is not widespread among practitioners which want to obtain usable results. On the one hand, the researchers found especially it is very easy to program and implement it, its parameters are easy to estimate, learning is very fast even on very large databases, its accuracy is reasonably good in comparison to the other approaches. On the other hand, the final users do not obtain a model easy to interpret and deploy, they does not understand the interest of such a technique.

Thus, we introduce in a new presentation of the results of the learning process. The classifier is easier to understand, and its deployment is also made easier. In the first part of this tutorial, we present some theoretical aspects of the naive bayes classifier.

Then, we implement the approach on a dataset with Tanagra. We compare the obtained results (the parameters of the model) to those obtained with other linear approaches such as the logistic regression, the linear discriminant analysis and the linear SVM. We note that the results are highly consistent. This largely explains the good performance of the method in comparison to others. In the second part, we use various tools on the same dataset (Weka 3.6.0, R 2.9.2, Knime 2.1.1, Orange 2.0b and RapidMiner 4.6.0). We try above all to understand the obtained results.

**Random Forest**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees. For regression tasks, the mean or average prediction of the individual trees is returned. Random decision forests correct for decision trees' habit of overfitting to their training set. Random forests generally outperform decision trees, but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

The first algorithm for random decision forests was created in 1995 by Tin Kam Ho[1] using the random subspace method, which, in Ho's formulation, is a way to implement the "stochastic discrimination" approach to classification proposed by Eugene Kleinberg.

An extension of the algorithm was developed by Leo Breiman and Adele Cutler, who registered "Random Forests" as a trademark in 2006 (as of 2019, owned by Minitab, Inc.).The extension combines Breiman's "bagging" idea and random selection of features, introduced first by Ho[1] and later independently by Amit and Geman[13] in order to construct a collection of decision trees with controlled variance.

Random forests are frequently used as "blackbox" models in businesses, as they generate reasonable predictions across a wide range of data while requiring little configuration.

**SVM**

In classification tasks a discriminant machine learning technique aims at finding, based on an *independent and identically distributed* (*iid*) training dataset, a discriminant function that can correctly predict labels for newly acquired instances. Unlike generative machine learning approaches, which require computations of conditional probability distributions, a discriminant classification function takes a data point *x* and assigns it to one of the different classes that are a part of the classification task. Less powerful than generative approaches, which are mostly used when prediction involves outlier detection, discriminant approaches require fewer computational resources and less training data, especially for a multidimensional feature space and when only posterior probabilities are needed. From a geometric perspective, learning a classifier is equivalent to finding the equation for a multidimensional surface that best separates the different classes in the feature space.

SVM is a discriminant technique, and, because it solves the convex optimization problem analytically, it always returns the same optimal hyperplane parameter—in contrast to *genetic algorithms* (*GAs*) or *perceptrons*, both of which are widely used for classification in machine learning. For perceptrons, solutions are highly dependent on the initialization and termination criteria. For a specific kernel that transforms the data from the input space to the feature space, training returns uniquely defined SVM model parameters for a given training set, whereas the perceptron and GA classifier models are different each time training is initialized. The aim of GAs and perceptrons is only to minimize error during training, which will translate into several hyperplanes' meeting this requirement.

## VII. 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 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.

## IX.REFERENCES

[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