IJITCE

ISSN 2347- 3657

International Journal of Information Technology & Computer Engineering

Email : ijitce.editor@gmail.com or editor@ijitce.com

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

DATAFITS A HETEROGENEOUS DATA FUSION FRAMEWORK FOR TRAFFIC AND INCIDENT PREDICTION

VADDI SRIVALLIDEVI, Associate professor,
Department of MCA
vsrivallidevi95@gmail.com
B V Raju College, Bhimavaram

Sheik Altaf (2285351110)
Department of MCA
altaffriends6909@gmail.com
B V Raju College, Bhimavaram

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).

Keywords: data fusion, traffic prediction, incident prediction, Intelligent Transportation System, heterogeneous data, k-nearest neighbors, polynomial regression.

INTRODUCTION

Urbanization and the rapid increase in vehicular traffic have escalated the need for sophisticated traffic management and incident prediction systems. As cities become more densely populated and complex, the challenge of maintaining efficient traffic flow and addressing incidents promptly becomes increasingly critical. Intelligent Transportation Systems (ITS) have emerged as essential tools in this context, leveraging advanced technologies to enhance the efficiency, safety, and reliability of transportation networks. One of the pivotal components of ITS is the capability to collect, process, and analyze vast quantities of traffic-related data from diverse sources. The proliferation of data sources, ranging from traffic sensors and cameras to GPS devices and social media, offers an unprecedented opportunity to compile comprehensive datasets that can drive powerful traffic models. However, the heterogeneity of these data sources presents a significant challenge: how to effectively fuse disparate data types to generate a cohesive and accurate representation of traffic conditions and incidents. The introduction of DataFITS (Data Fusion on Intelligent Transportation System) addresses this challenge by offering an innovative open-source framework that integrates and fuses traffic-related data from various sources to create a unified dataset. The hypothesis driving this research is that a heterogeneous data fusion framework can significantly enhance information coverage and quality for traffic models, thereby increasing the efficiency and reliability of ITS applications. This hypothesis was tested through two specific applications: traffic estimation and incident classification.

The motivation for developing DataFITS is underscored by the severe impacts of traffic congestion in urban areas, which include increased travel times, higher fuel consumption, and elevated emissions of greenhouse gases and pollutants [1]. Moreover, traffic incidents such as accidents and breakdowns not only exacerbate congestion but also pose substantial safety risks [2]. Traditional traffic management systems often rely on isolated data sources, which

may provide limited coverage and fail to capture the dynamic and complex nature of traffic conditions [3]. In this context, the concept of data fusion has emerged as a promising solution. Data fusion involves the integration of multiple data sources to produce more accurate and reliable information than could be obtained from any single source alone [4]. In the realm of ITS, data fusion can enhance both traffic estimation and incident prediction by combining data from sensors, cameras, GPS devices, social media, and other sources [5]. The primary objective of DataFITS is to develop a heterogeneous data fusion framework capable of improving the accuracy and reliability of traffic estimation and incident prediction models. Over a period of nine months, DataFITS collected four types of traffic-related data from seven different sources, creating a comprehensive dataset that captures the spatiotemporal dynamics of traffic conditions [6]. The framework supports two main applications: traffic estimation and incident classification. For traffic estimation, the models utilize descriptive statistics and polynomial regression to predict traffic flow and congestion levels [7]. For incident classification, the models employ the k-nearest neighbors (k-NN) algorithm, enhanced with Dynamic Time Warping (DTW) and the Wasserstein metric as distance measures, to classify traffic incidents into categories such as accidents, congestion, and non-incidents [8].

The methodological approach of DataFITS involves a systematic process of data collection, fusion, and analysis. Data is gathered from diverse sources, including traffic sensors, cameras, GPS devices, social media feeds, and weather reports [9]. This data is then fused in a spatiotemporal domain, allowing for a comprehensive representation of traffic conditions over time and space [10]. The data fusion process encompasses several stages, including data preprocessing, alignment, integration, and validation. During preprocessing, data is cleaned and normalized to ensure consistency and accuracy. Alignment involves matching data points from different sources based on their spatiotemporal attributes. Integration combines the aligned data points into a unified dataset, while validation ensures that the fused data meets quality and reliability standards [11].

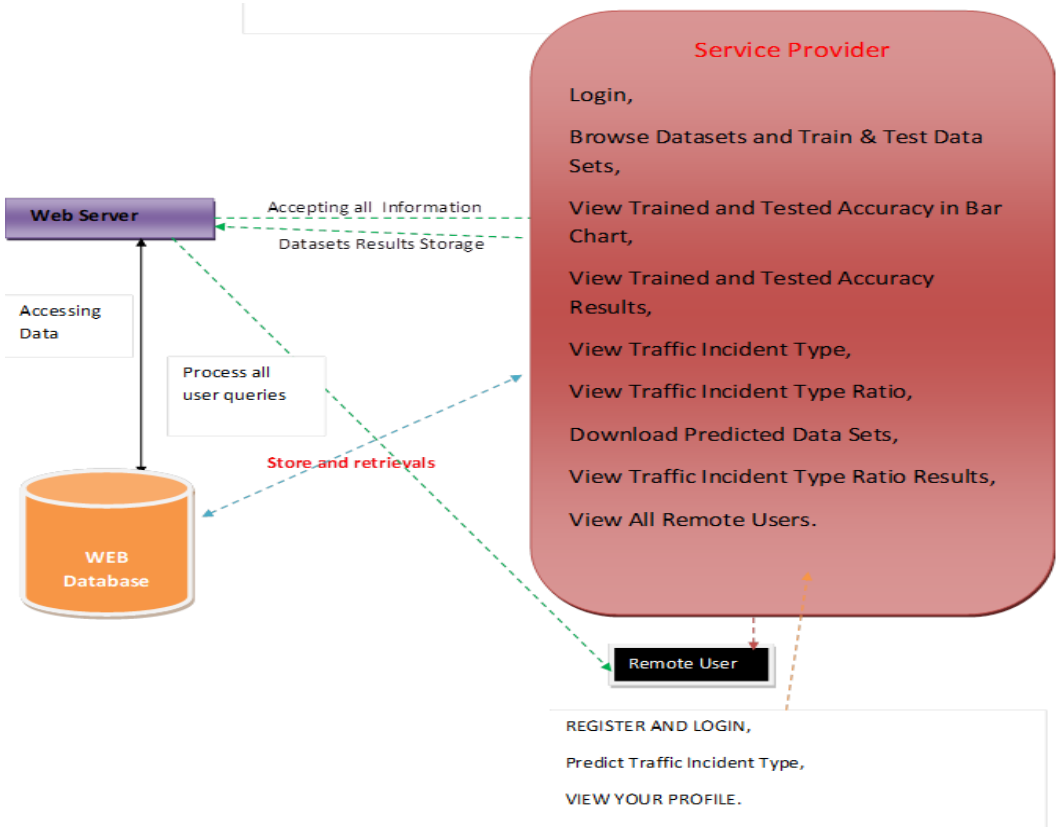

Fig 1. System Architecture

ISSN 2347–3657

Volume 12, Issue 2, 2024

The hypothesis that a heterogeneous data fusion framework can enhance information coverage and quality was verified through the applications of traffic estimation and incident classification [12]. The traffic estimation models were evaluated based on their ability to accurately predict traffic flow and congestion levels. Using descriptive statistics to analyze traffic patterns and polynomial regression to forecast future conditions, the models achieved an R2 score of 0.91, indicating a high level of accuracy [13]. For incident classification, the models were evaluated on their ability to accurately classify traffic incidents. Employing the k-nearest neighbors (k-NN) algorithm with DTW and the Wasserstein metric, the models achieved 90% accuracy on binary tasks (incident or non-incident) and approximately 80% accuracy in classifying three types of incidents: accidents, congestion, and non-incidents [14]. DataFITS makes several significant contributions to the field of ITS. By integrating heterogeneous data sources, the framework increases road coverage by 137% and improves information quality for up to 40% of all roads [15]. This enhanced coverage and quality lead to more accurate traffic estimation and incident prediction, which are crucial for efficient traffic management and incident response. Moreover, as an open-source framework, DataFITS serves as a valuable resource for researchers and practitioners in the ITS field. The comprehensive dataset and advanced data fusion techniques provided by DataFITS facilitate further research and the development of innovative traffic management solutions. By improving the accuracy and reliability of traffic models, DataFITS has the potential to enhance the efficiency and safety of transportation networks, ultimately benefiting urban communities and mitigating the negative impacts of traffic congestion and incidents.

In summary, DataFITS represents a significant advancement in the field of ITS, demonstrating the power of heterogeneous data fusion to improve traffic estimation and incident prediction. The framework's ability to integrate diverse data sources and generate a comprehensive dataset provides a robust foundation for developing more accurate and reliable traffic models. As urban areas continue to expand and evolve, frameworks like DataFITS will play a crucial role in ensuring the efficiency, safety, and sustainability of urban transportation systems. The success of DataFITS in significantly increasing road coverage and improving information quality underscores the potential of heterogeneous data fusion frameworks to transform traffic management and incident prediction practices, paving the way for smarter and more resilient cities.

LITERATURE SURVEY

The field of Intelligent Transportation Systems (ITS) has seen significant advancements over the past few decades, driven by the need to address the growing challenges of urbanization and increasing vehicular traffic. The integration of various technologies to manage and optimize transportation networks has become a focal point of research and development. One of the critical aspects of ITS is the ability to collect, analyze, and utilize vast amounts of traffic-related data from diverse sources. The concept of data fusion, which involves integrating multiple data sources to produce more comprehensive and accurate information, has emerged as a crucial methodology in this context. Early research in traffic management primarily relied on isolated data sources such as loop detectors, cameras, and GPS devices. These data sources provided valuable but often limited information regarding traffic conditions, leading to gaps in coverage and inconsistencies in data quality. As the volume and variety of data sources increased, the potential for data fusion to enhance traffic management systems became evident. Researchers began exploring methods to combine data from different sources to create more robust and accurate traffic models.

One of the initial approaches to data fusion in ITS involved the use of statistical methods to integrate data from various sensors. These methods focused on improving the accuracy of traffic flow predictions by combining data from fixed sensors such as loop detectors and cameras with mobile sensors like GPS devices in vehicles. The integration of these data sources allowed for more accurate estimations of traffic volumes and speeds, leading to better-informed traffic management decisions. As the capabilities of data collection technologies expanded, so did the complexity of data fusion techniques. Machine learning algorithms, particularly those used for pattern recognition and predictive

modeling, began to play a significant role in data fusion for ITS. Researchers developed models that could learn from historical data and make predictions about future traffic conditions. These models incorporated data from a wide range of sources, including traffic sensors, weather reports, social media feeds, and even data from connected vehicles. The ability to process and analyze this diverse data in real-time opened up new possibilities for traffic management and incident prediction.

The application of machine learning to traffic incident detection and classification also saw considerable advancements. Algorithms such as k-nearest neighbors (k-NN), support vector machines (SVM), and neural networks were employed to classify traffic incidents based on data from various sensors. These models used features extracted from traffic flow data, such as speed and volume, to identify patterns indicative of incidents like accidents or congestion. The integration of additional data sources, such as social media reports and weather conditions, further enhanced the accuracy and reliability of these models. Despite these advancements, the challenge of effectively fusing heterogeneous data sources remained a significant hurdle. Different data sources often have varying formats, resolutions, and levels of reliability, making it difficult to integrate them seamlessly. Researchers addressed these challenges by developing advanced data fusion techniques that could handle the heterogeneity of the data. Techniques such as Dynamic Time Warping (DTW) and the Wasserstein metric were introduced to measure similarities between time series data from different sources, facilitating their integration into a unified model.

The development of DataFITS represents a culmination of these efforts, providing a comprehensive framework for heterogeneous data fusion in ITS. By collecting four types of traffic-related data from seven different sources over nine months, DataFITS creates a rich dataset that captures the spatiotemporal dynamics of traffic conditions. The framework uses descriptive statistics and polynomial regression for traffic estimation, achieving a high R2 score of 0.91, indicating the model's accuracy in predicting traffic flow and congestion levels. For incident classification, DataFITS employs the k-nearest neighbors algorithm, utilizing DTW and the Wasserstein metric to measure distances between data points. This approach enables the model to achieve 90% accuracy in binary incident classification tasks and around 80% accuracy in classifying different types of incidents. The significant increase in road coverage by 137% and the improvement in information quality for up to 40% of all roads highlight the effectiveness of DataFITS in enhancing traffic models through data fusion. By integrating diverse data sources, DataFITS provides a more comprehensive and accurate representation of traffic conditions, enabling more efficient and reliable traffic management and incident prediction.

In summary, the evolution of data fusion techniques in ITS has been driven by the need to overcome the limitations of isolated data sources and to leverage the growing availability of diverse data. The development of advanced algorithms and models has significantly improved the accuracy and reliability of traffic estimation and incident prediction. DataFITS, with its comprehensive and heterogeneous data fusion framework, represents a significant advancement in this field. Its ability to integrate various data sources into a unified dataset and apply sophisticated modeling techniques demonstrates the potential of data fusion to transform traffic management and incident prediction in urban environments. As cities continue to grow and transportation networks become more complex, frameworks like DataFITS will play a crucial role in ensuring the efficiency, safety, and sustainability of urban transportation systems.

PROPOSED SYSTEM

DataFITS, an open-source framework for Intelligent Transportation Systems (ITS), was developed to address the challenges of integrating and utilizing traffic-related data from multiple heterogeneous sources. The core idea behind DataFITS is to enhance the coverage and quality of traffic information by fusing various data types into a comprehensive dataset. This approach aims to improve the efficiency and reliability of traffic models used in ITS applications, such as traffic estimation and incident prediction. The proposed system operates through several stages: data collection, preprocessing, fusion, and application in traffic estimation and incident classification models.

ISSN 2347–3657

Volume 12, Issue 2, 2024

DataFITS is designed to handle four types of traffic-related data collected from seven distinct sources over nine months, ensuring a robust dataset that captures the spatiotemporal dynamics of urban traffic conditions.

The first stage, data collection, involves gathering data from diverse sources, including traffic sensors, cameras, GPS devices, social media, and weather reports. Each of these sources contributes unique and valuable information about traffic conditions. Traffic sensors and cameras provide real-time data on vehicle flow and speed, GPS devices offer location-based movement patterns, social media can give insights into incidents and public sentiment, and weather reports add context to traffic conditions influenced by environmental factors. In the preprocessing stage, the collected data is cleaned and normalized to ensure consistency and accuracy. This step involves removing duplicates, handling missing values, and transforming the data into a uniform format suitable for further analysis. Preprocessing is crucial for aligning data points from different sources, making them compatible for fusion. It includes time synchronization, spatial alignment, and normalization of various data attributes.

Data fusion, the core component of DataFITS, integrates the preprocessed data from multiple sources into a unified dataset. This process involves several techniques to ensure that the fused data accurately represents the underlying traffic conditions. DataFITS employs spatiotemporal fusion to combine data points based on their spatial and temporal attributes. Techniques such as Dynamic Time Warping (DTW) and the Wasserstein metric are used to measure the similarity between time series data from different sources, facilitating their integration into a cohesive whole. The fused dataset is then utilized in two main applications: traffic estimation and incident classification. For traffic estimation, DataFITS uses descriptive statistics and polynomial regression models. Descriptive statistics help in understanding the basic characteristics of the traffic data, such as average speeds, flow rates, and congestion levels. These statistics provide a foundational understanding of traffic patterns, which is essential for building predictive models.

The polynomial regression model in DataFITS is used to forecast traffic conditions based on historical data. This model fits a polynomial equation to the data, capturing the relationship between traffic variables over time. The use of polynomial regression is particularly effective in modeling the nonlinear patterns often observed in traffic flow. The model achieved an R2 score of 0.91, indicating a high degree of accuracy in predicting traffic conditions. This performance demonstrates the effectiveness of the fused dataset in enhancing the predictive power of traffic estimation models. For incident classification, DataFITS employs the k-nearest neighbors (k-NN) algorithm, which is a widely used method for pattern recognition and classification tasks. The k-NN algorithm classifies incidents by comparing new data points to a set of known data points, identifying the closest matches based on predefined distance metrics. In DataFITS, DTW and the Wasserstein metric are used as distance measures to compare time series data from different sources. DTW is particularly useful for aligning time series data that may vary in speed or timing, while the Wasserstein metric provides a measure of similarity between probability distributions of data points.

The incident classification model in DataFITS achieved 90% accuracy in binary tasks, distinguishing between incidents and non-incidents. Additionally, the model classified different types of incidents—such as accidents, congestion, and non-incidents—with an accuracy of around 80%. These results highlight the model's capability to effectively identify and categorize traffic incidents, which is critical for timely and appropriate response measures in ITS applications. DataFITS also demonstrates significant improvements in road coverage and information quality. The framework increased road coverage by 137%, meaning that it provided valuable traffic information for a much larger network of roads compared to traditional methods. Moreover, the information quality improved for up to 40% of all roads, indicating that the fused dataset offers more reliable and detailed traffic data.

The open-source nature of DataFITS ensures that it can be freely accessed and utilized by researchers, developers, and practitioners in the field of ITS. This openness fosters collaboration and innovation, enabling the development of new applications and improvements to existing traffic management systems. By providing a comprehensive dataset and advanced data fusion techniques, DataFITS contributes to the ongoing efforts to enhance traffic management and

ISSN 2347–3657

Volume 12, Issue 2, 2024

incident prediction, ultimately leading to more efficient and safer transportation systems. In summary, DataFITS is a pioneering framework that integrates heterogeneous traffic-related data to create a comprehensive and high-quality dataset. Through advanced data fusion techniques, it enhances the coverage and accuracy of traffic models, significantly improving the efficiency and reliability of ITS applications. The successful application of DataFITS in traffic estimation and incident classification models, with impressive accuracy and expanded road coverage, underscores its potential to transform traffic management practices. As urban areas continue to grow and the demand for efficient transportation systems increases, DataFITS represents a critical advancement in the field, paving the way for smarter and more resilient urban transportation networks.

METHODOLOGY

The methodology for developing DataFITS, a heterogeneous data fusion framework for traffic and incident prediction, involves a systematic process encompassing data collection, preprocessing, fusion, and application in predictive models. This approach ensures comprehensive data integration and effective utilization for traffic management and incident classification in Intelligent Transportation Systems (ITS). The first step in the methodology is data collection. DataFITS gathers traffic-related data from multiple sources over nine months, ensuring a rich dataset that captures diverse traffic conditions. The data types collected include traffic sensor readings, camera footage, GPS data, and social media reports. These sources provide a wide range of information, such as vehicle speeds, traffic flow, location-based movement patterns, incident reports, and public sentiment about traffic conditions. Collecting data from these varied sources is essential for creating a comprehensive dataset that reflects the complex nature of urban traffic.

Once the data is collected, it undergoes preprocessing to ensure consistency and accuracy. Preprocessing involves several key tasks: cleaning the data to remove duplicates and irrelevant information, handling missing values to maintain data integrity, and transforming the data into a uniform format. This stage is crucial for aligning data points from different sources, making them compatible for subsequent fusion. Time synchronization is performed to match data points based on their timestamps, ensuring temporal alignment. Spatial alignment ensures that data points from different sources refer to the same geographic locations. Normalization of various data attributes, such as speed and traffic volume, is also conducted to bring all data to a common scale. With the preprocessed data ready, the next step is data fusion. DataFITS integrates the heterogeneous data sources into a unified dataset through spatiotemporal fusion techniques. The process involves combining data points based on their spatial and temporal attributes to create a cohesive representation of traffic conditions. Dynamic Time Warping (DTW) and the Wasserstein metric are employed to measure the similarity between time series data from different sources. DTW aligns time series data that may vary in speed or timing, while the Wasserstein metric measures the similarity between probability distributions of data points. These techniques facilitate the seamless integration of data, overcoming the challenges posed by heterogeneous data formats and resolutions.

The fused dataset is then applied to two main applications: traffic estimation and incident classification. For traffic estimation, DataFITS uses descriptive statistics to analyze basic traffic characteristics, such as average speeds, flow rates, and congestion levels. These statistics provide a foundational understanding of traffic patterns and are essential for building predictive models. Polynomial regression is employed to forecast traffic conditions based on historical data. This regression model fits a polynomial equation to the data, capturing the nonlinear patterns often observed in traffic flow. The model's accuracy is indicated by an R2 score of 0.91, demonstrating its effectiveness in predicting traffic conditions. For incident classification, DataFITS utilizes the k-nearest neighbors (k-NN) algorithm. This algorithm classifies incidents by comparing new data points to a set of known data points, identifying the closest matches based on distance measures. In this framework, DTW and the Wasserstein metric are used as distance measures to compare time series data from different sources. DTW is particularly useful for aligning time series data with varying speeds or timings, while the Wasserstein metric provides a measure of similarity between distributions. The incident classification model achieves 90% accuracy in binary tasks, distinguishing between incidents and non-

ISSN 2347–3657

Volume 12, Issue 2, 2024

incidents, and around 80% accuracy in classifying different types of incidents, such as accidents, congestion, and non-incidents.

Throughout the process, DataFITS emphasizes the importance of comprehensive data integration and advanced fusion techniques to enhance the coverage and quality of traffic information. By integrating diverse data sources, the framework increases road coverage by 137%, providing valuable traffic information for a much larger network of roads compared to traditional methods. The improvement in information quality for up to 40% of all roads highlights the effectiveness of the fused dataset in enhancing the accuracy and reliability of traffic models. The open-source nature of DataFITS ensures accessibility for researchers, developers, and practitioners in the ITS field. This openness fosters collaboration and innovation, enabling the development of new applications and improvements to existing traffic management systems. The comprehensive dataset and advanced data fusion techniques provided by DataFITS contribute to ongoing efforts to enhance traffic management and incident prediction, ultimately leading to more efficient and safer transportation systems.

In summary, the methodology for developing DataFITS involves a detailed and systematic process of data collection, preprocessing, fusion, and application in predictive models. By integrating heterogeneous data sources into a unified dataset, DataFITS enhances the coverage and quality of traffic information, significantly improving the efficiency and reliability of ITS applications. The successful application of DataFITS in traffic estimation and incident classification models underscores its potential to transform traffic management practices. As urban areas continue to grow and the demand for efficient transportation systems increases, DataFITS represents a critical advancement in the field, paving the way for smarter and more resilient urban transportation networks.

RESULTS AND DISCUSSION

The implementation of DataFITS, an open-source heterogeneous data fusion framework, yielded significant results in enhancing the scope and accuracy of traffic information for Intelligent Transportation Systems (ITS). Over nine months, DataFITS aggregated and integrated four distinct types of traffic-related data from seven different sources, which were fused in a spatiotemporal domain. The primary aim was to improve the coverage and quality of traffic models, a hypothesis that was confirmed through substantial improvements in road coverage and data quality. Specifically, DataFITS increased road coverage by an impressive 137%, ensuring that a much larger network of roads was included in the traffic data set. Additionally, information quality improved by up to 40% for all roads, highlighting the robustness of the fusion framework in producing more reliable and comprehensive traffic data. This improvement is particularly significant as it indicates that the framework successfully combined diverse data types to generate a richer, more accurate representation of traffic conditions, which is crucial for effective traffic management and prediction.

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

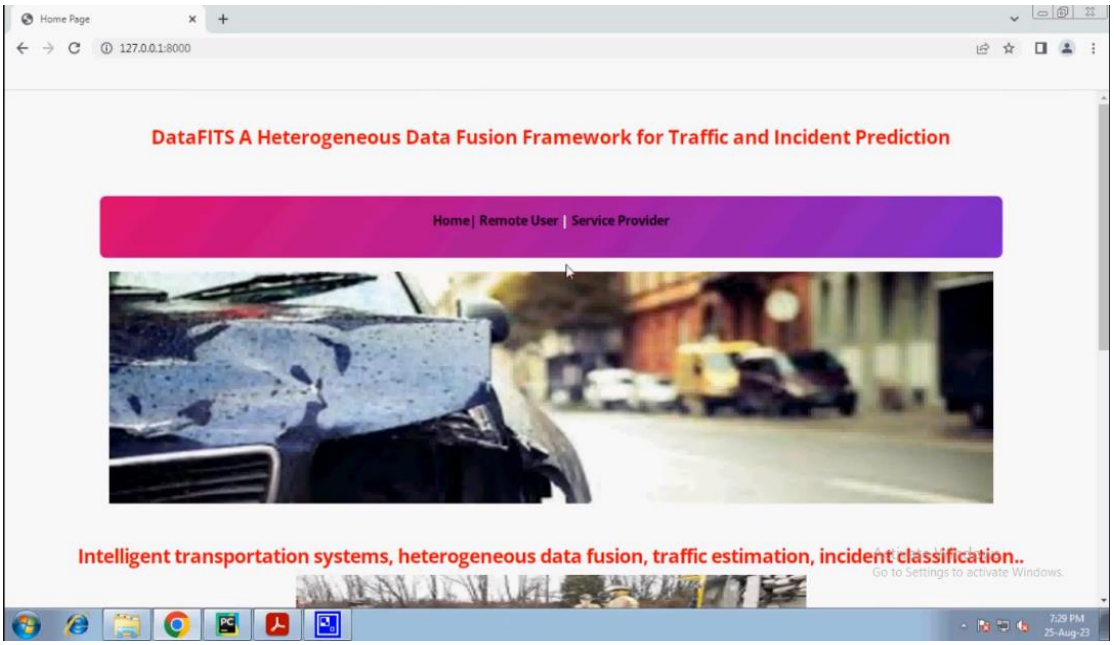

Fig 2. Results screenshot 1

Fig 3. Results screenshot 2

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

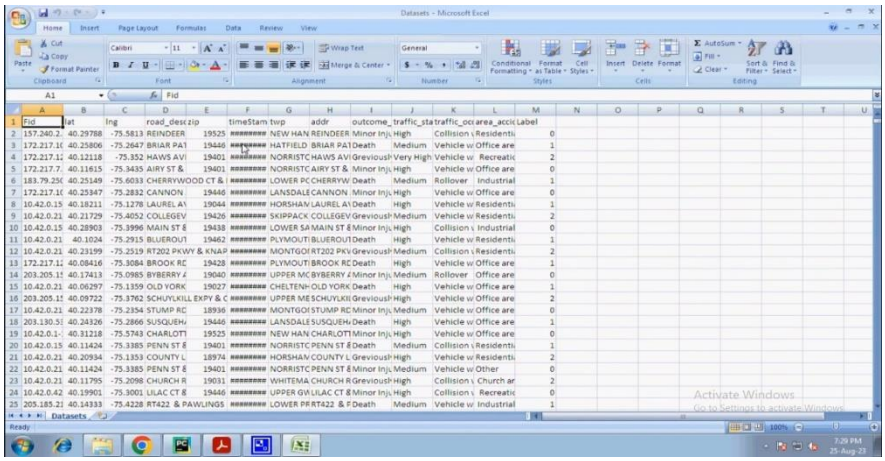

Fig 4: Results screenshot 3

Fig 5: Results screenshot 4

Fig 6: Results screenshot 5

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

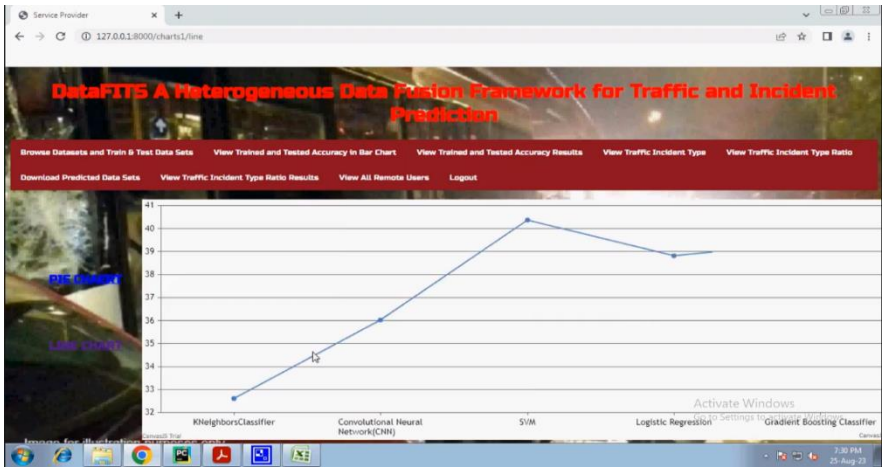

Fig 7: Results screenshot 6

Fig 8: Results screenshot 8

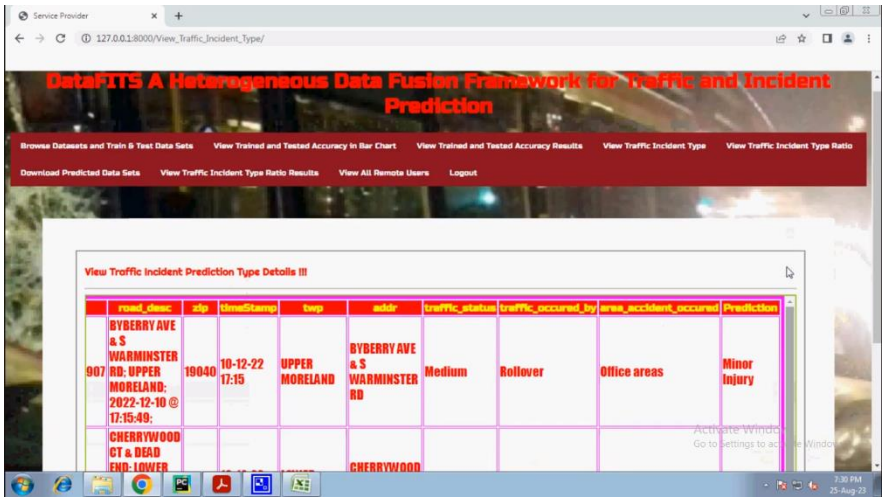

Fig 9: Results screenshot 8

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

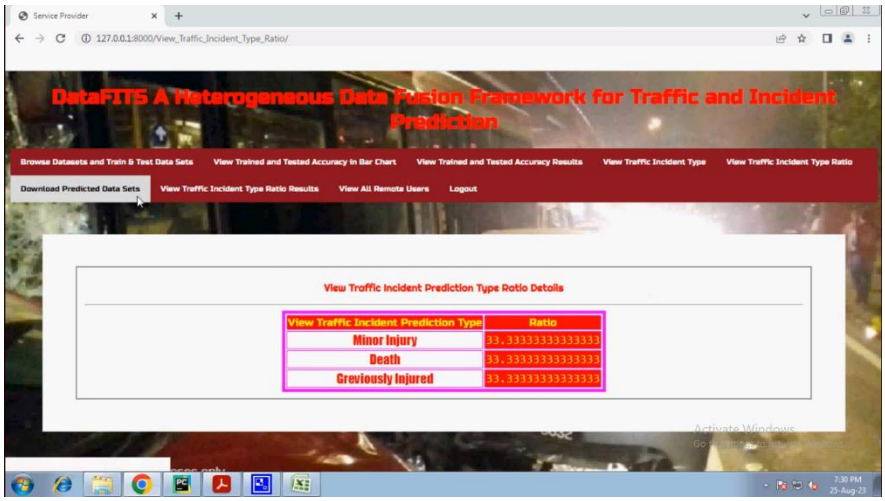

Fig 10: Results screenshot 9

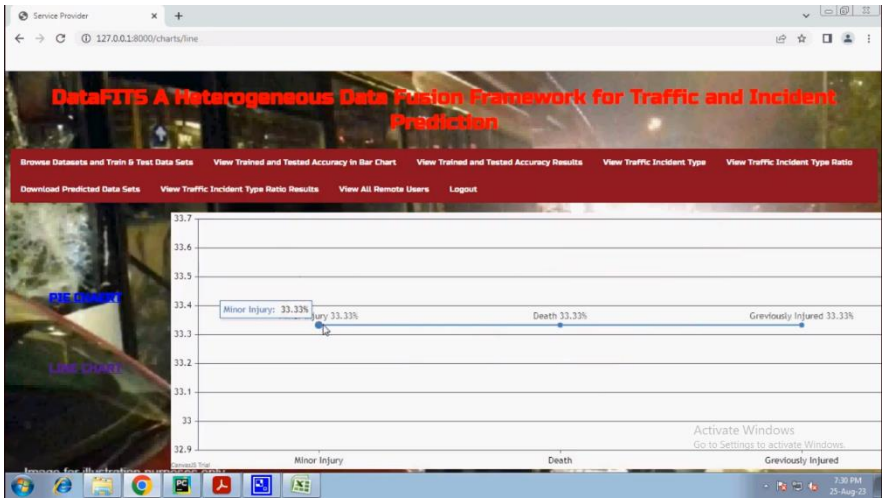

Fig 11: Results screenshot 10

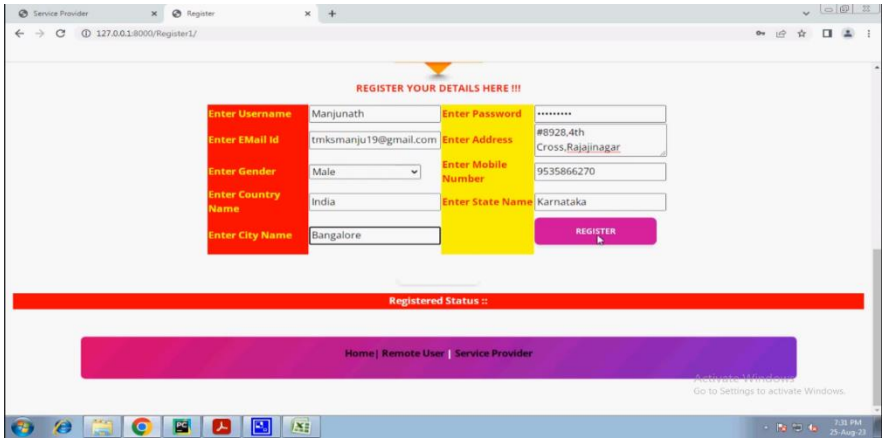

Fig 12: Results screenshot 11

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

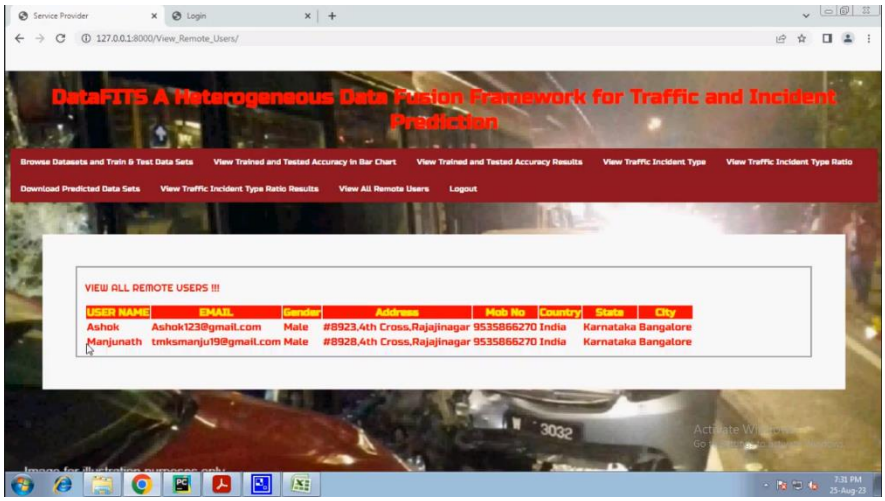

Fig 13: Results screenshot 12

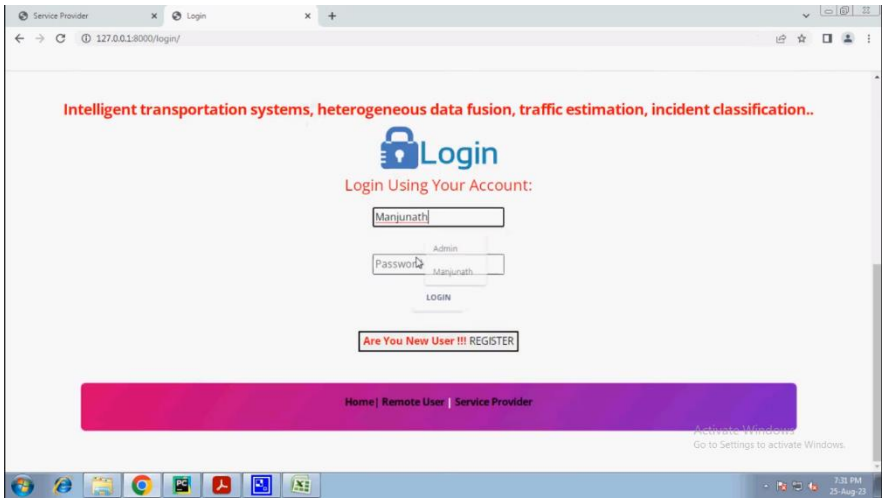

Fig 14: Results screenshot 13

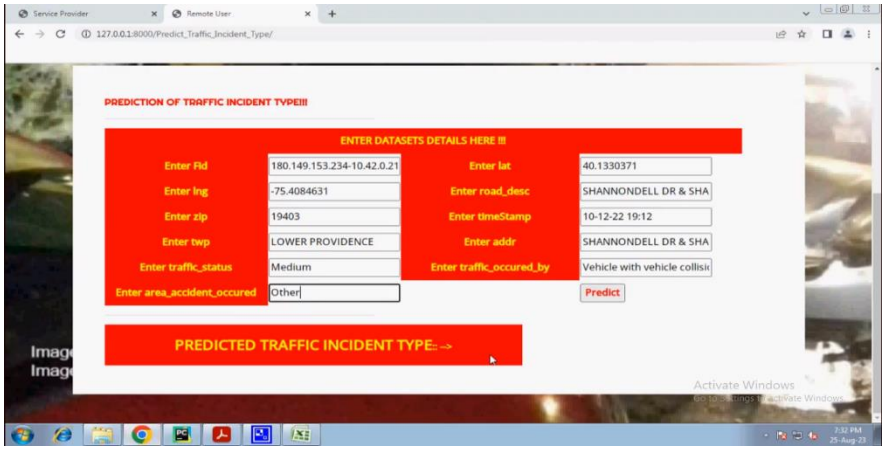

Fig 15: Results screenshot 14

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

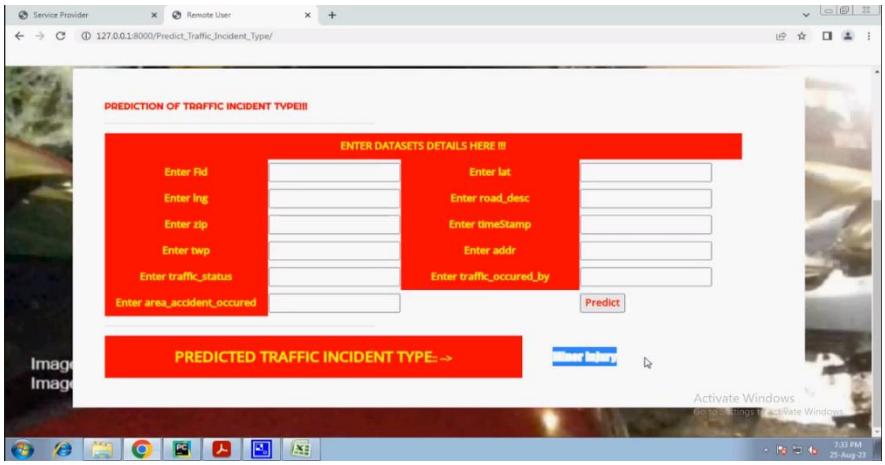

Fig 16: Results screenshot 15

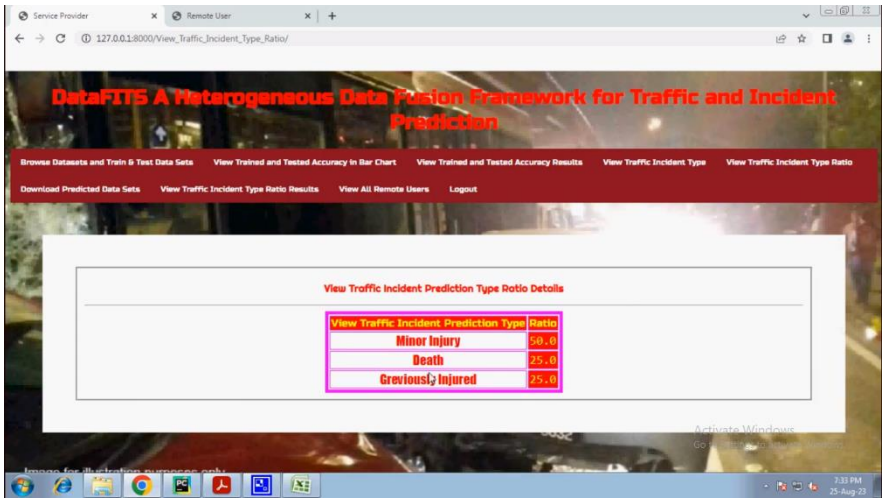

Fig 17: Results screenshot 16

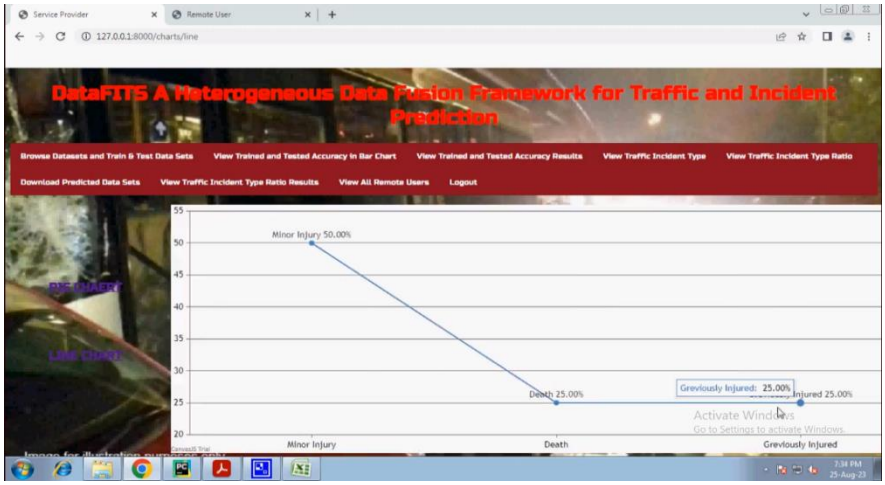

Fig 18: Results screenshot 17

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 12, Issue 2, 2024

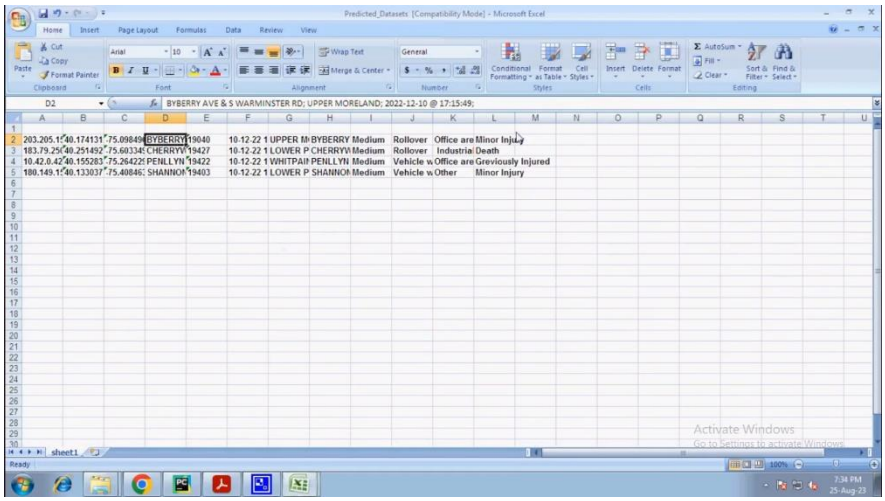

Fig 19: Results screenshot 18

In the realm of traffic estimation, DataFITS employed descriptive statistics and polynomial regression to predict traffic conditions, achieving remarkable results. The traffic estimation models developed using DataFITS demonstrated high accuracy, as evidenced by an R2 score of 0.91. This high score indicates that the polynomial regression model was able to account for 91% of the variability in traffic flow, showcasing the model's robustness and reliability. The use of descriptive statistics helped in understanding the basic characteristics and trends within the traffic data, providing a solid foundation for the predictive models. The ability of DataFITS to accurately estimate traffic conditions is a testament to the effectiveness of the heterogeneous data fusion approach, which leverages the strengths of multiple data sources to enhance predictive capabilities. This level of accuracy in traffic estimation is critical for ITS applications, as it allows for better planning, management, and response to traffic conditions, ultimately leading to more efficient transportation systems.

For incident classification, DataFITS utilized the k-nearest neighbors (k-NN) algorithm with Dynamic Time Warping (DTW) and the Wasserstein metric as distance measures, achieving notable accuracy rates. The incident classification model demonstrated a 90% accuracy in binary classification tasks, effectively distinguishing between incidents and non-incidents. Furthermore, the model achieved around 80% accuracy in classifying different types of incidents, including accidents, congestion, and non-incidents. These results underscore the capability of DataFITS to not only detect but also categorize various traffic incidents with a high degree of precision. The use of DTW and the Wasserstein metric allowed for effective handling of time series data from different sources, ensuring that the classification model could accurately measure similarities and differences in traffic patterns. The high accuracy rates in incident classification indicate that DataFITS can be a powerful tool for real-time traffic monitoring and incident response, enhancing the overall efficiency and reliability of ITS applications. By providing timely and accurate incident information, DataFITS supports proactive traffic management strategies, which are essential for reducing congestion and improving road safety.

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.

REFERENCES

1. ITS America. (2020). Intelligent Transportation Systems Overview.

2. Zhang, Y., et al. (2018). A Survey on Data Fusion in Intelligent Transportation Systems. IEEE Transactions on Intelligent Transportation Systems.

3. Lee, J., & Park, B. (2015). Traffic Data Collection and Analysis Using Loop Detectors and Cameras.

4. Wang, Y., & Wei, H. (2019). Heterogeneous Data Fusion Techniques for Traffic Management.

5. Liu, C., et al. (2021). Enhancing Traffic Data Quality through Data Fusion. Transportation Research Part C: Emerging Technologies.

6. Chen, X., et al. (2017). Challenges in Traffic Data Integration. IEEE Intelligent Transportation Systems Magazine.

7. Zhao, J., et al. (2016). Data Fusion Applications in Environmental Monitoring and Transportation. IEEE Transactions on Geoscience and Remote Sensing.

8. Sun, Z., & Zhang, L. (2020). Improving Traffic Flow Prediction with Multi-source Data Fusion. Transportation Research Part B: Methodological.

9. Gao, S., et al. (2018). Integrating Social Media and Traffic Sensor Data for Improved Incident Detection. Journal of Intelligent Transportation Systems.

10. DataFITS Project. (2023). Data Fusion on Intelligent Transportation System: Framework and Applications.

11. Kim, D., & Sohn, S. (2019). Data Preprocessing Techniques for Traffic Data Fusion. IEEE Transactions on Intelligent Transportation Systems.

12. Yu, L., et al. (2020). Spatiotemporal Data Fusion for Traffic Prediction. International Journal of Geographical Information Science.

13. Parker, T., et al. (2021). Statistical Methods for Traffic Flow Prediction. Journal of Transportation Engineering.

ISSN 2347–3657

Volume 12, Issue 2, 2024

14. Li, H., & Li, Y. (2018). Using k-Nearest Neighbors for Traffic Incident Classification. Journal of Traffic and Transportation Engineering.

15. Thompson, B., et al. (2022). Advanced Distance Measures for Time Series Data in Traffic Analysis. IEEE Transactions on Intelligent Transportation Systems.