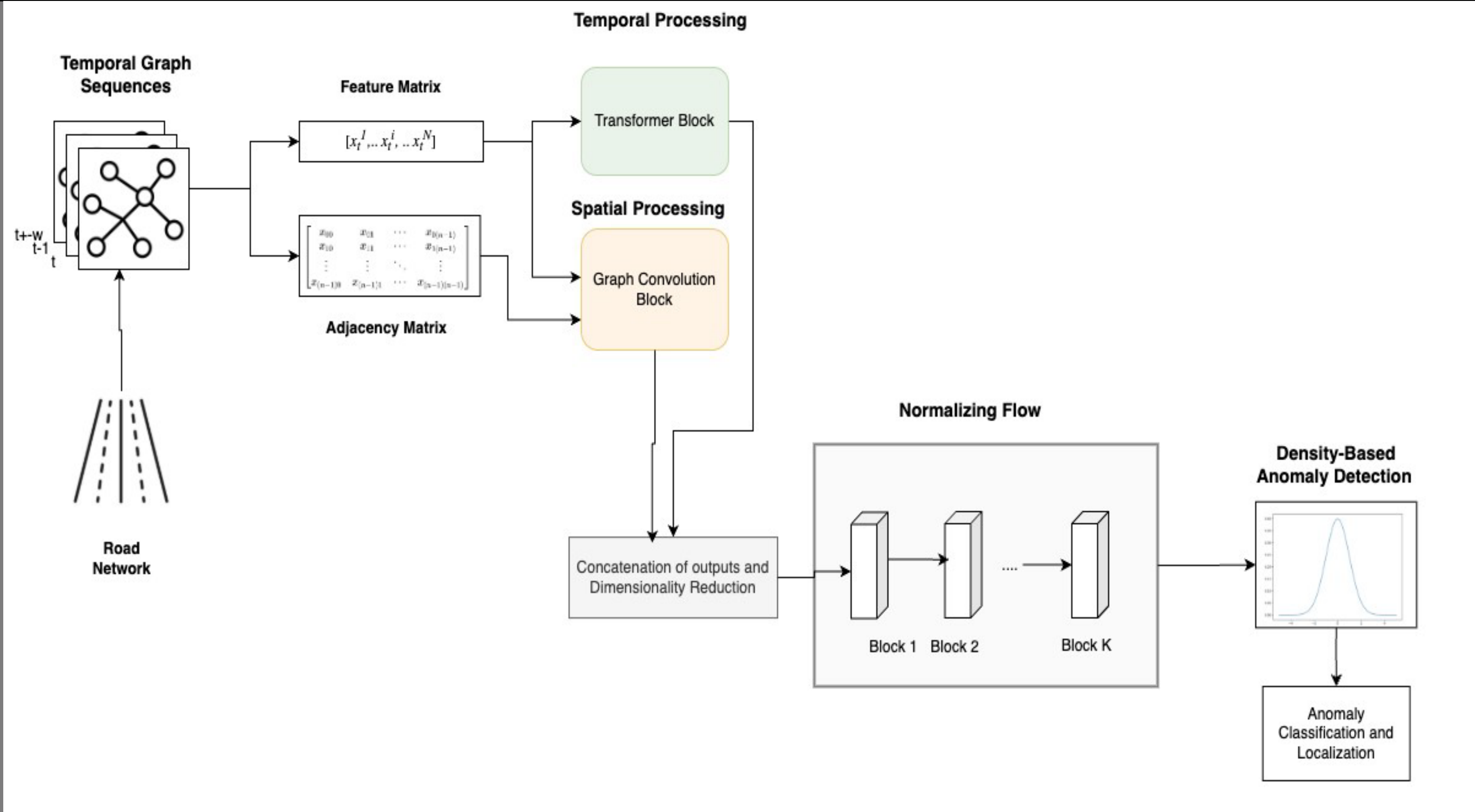



TDOT Department of Transportation

U.S. Department of Transportation Federal Highway Administration

# AI-Engine for Adaptive Sensor Fusion for Traffic Monitoring System

RES2023.17

Research Final Report from Vanderbilt University | Abhishek Dubey, Dan Work, Hiba Baroud, Ayan Mukhopadhyay | December 17, 2024

Sponsored by Tennessee Department of Strategic Planning, Research, & Innovation Division Research Office & Federal Highway Administration

# DISCLAIMER

This research was funded through the State Planning and Research (SPR) Program by the Tennessee Department of Transportation and the Federal Highway Administration under ***RES2023.17: AI-Engine for Adaptive Sensor Fusion for Traffic Monitoring System.***

This document is disseminated under the sponsorship of the Tennessee Department of Transportation and the United States Department of Transportation in the interest of information exchange. The State of Tennessee and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the views of the author(s) who are solely responsible for the facts and accuracy of the material presented. The contents do not necessarily reflect the official views of the Tennessee Department of Transportation or the United States Department of Transportation.

## Technical Report Documentation Page

| | | |
|---|---|---|
| 1. Report No.<br>RES2023.17 | 2. Government Accession No. | 3. Recipient's Catalog No. |
| 4. Title and Subtitle<br>*AI-Engine for Adaptive Sensor Fusion for Traffic Monitoring System* | | 5. Report Date<br>December 2024 |
| | | 6. Performing Organization Code |
| 7. Author(s)<br>Abhishek Dubey, Dan Work, Hiba Baroud, Ayan Mukhopadhyay | | 8. Performing Organization Report No. |
| 9. Performing Organization Name and Address<br>ISIS, Vanderbilt University,<br>1025 16th Ave S,<br>Nashville, TN, 37212 | | 10. Work Unit No. (TRAIS) |
| | | 11. Contract or Grant No.<br>RES2023.17 |
| 12. Sponsoring Agency Name and Address<br>Tennessee Department of Transportation<br>505 Deaderick Street, Suite 900<br>Nashville, TN 37243 | | 13. Type of Report and Period Covered |
| | | 14. Sponsoring Agency Code |

15. Supplementary Notes
Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract
Effective traffic incident management is critical for road safety and operational efficiency. Yet, many transportation agencies rely on predominantly reactionary methods, where incidents are reported by human agents and managed through rule-based frameworks like traditional Traffic Incident Management (TIM) systems. These methods, however, are vulnerable to human error, oversight, and response delays during high-stress conditions. Although recent initiatives incorporating real-time sensor data for corridor monitoring and enhanced roadway information systems represent strides toward modernization, these systems often still require substantial human intervention. Recent advancements in graph-based deep learning models offer promising potential for addressing the limitations of traditional methods. However, state-of-the-art models encounter challenges such as limited availability of high-quality labeled data, variability in real-time traffic measurements, and the complexities of incident localization within dynamic and interconnected road networks. To address these challenges, we propose the Traffic Response Anomaly Capture Engine (TRACE), a novel approach that combines graph neural networks, transformers, and probabilistic normalizing flows to accurately detect and localize traffic anomalies in real time. TRACE captures spatial-temporal dependencies, manages data uncertainty, and enhances automation, supporting more precise and timely incident localization. Our approach is validated on real-world traffic data and demonstrates improved detection accuracy by 5% and incident localization by 10% compared to current state-of-the-art methods, advancing traffic management strategies for safer, more efficient roadways.

| | |
|---|---|
| 17. Key Words<br>**ANOMALY DETECTION, LOCALIZATION, SENSOR FUSION, EMERGENCY RESPONSE, AI** | 18. Distribution Statement<br>No restriction. This document is available to the public from the sponsoring agency at the website http://www.tn.gov/. |

| | | | |
|---|---|---|---|
| 19. Security Classif. (of this report)<br>Unclassified | 20. Security Classif. (of this page)<br>Unclassified | 21. No. of Pages<br>32 | 22. Price<br>$208,268.00 |

# Acknowledgement

Collaborating with Tennessee Department of Transportation (TDOT), we have been able to present our work in some peer-reviewed journals, conferences, and prestigious venues. We acknowledge the support from the TDOT for funding the research. The list of peer reviewed journal or conference papers, reports, and presentations related to this work are as follows.

## *Published Articles*

Reports, Proceedings and Pre-prints:

Z. Kang, A. Mukhopadhyay, A. Gokhale, S. Wen, and A. Dubey, Traffic Anomaly Detection Via Conditional Normalizing Flow, in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2022, pp. 2563–2570.

J. Islam, J. P. Talusan, S. Bhattacharjee, F. Tiausas, S. M. Vazirizade, A. Dubey, K. Yasumoto, and S. Das, Anomaly based Incident Detection in Large Scale Smart Transportation Systems, in ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS), 2022.

M. J. Islam, J. P. Talusan, S. Bhattacharjee, F. Tiausas, A. Dubey, K. Yasumoto, and S. K. Das, Scalable Pythagorean Mean Based Incident Detection in Smart Transportation Systems, ACM Trans. Cyber-Phys. Syst., Jun. 2023.

Presentations:

Z. Kang, A. Mukhopadhyay, A. Gokhale, S. Wen, and A. Dubey, Traffic Anomaly Detection Via Conditional Normalizing Flow, in 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), 2022, pp. 2563–2570.

J. Islam, J. P. Talusan, S. Bhattacharjee, F. Tiausas, S. M. Vazirizade, A. Dubey, K. Yasumoto, and S. Das, Anomaly based Incident Detection in Large Scale Smart Transportation Systems, in ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS), 2022.

M. J. Islam, J. P. Talusan, S. Bhattacharjee, F. Tiausas, A. Dubey, K. Yasumoto, and S. K. Das, Scalable Pythagorean Mean Based Incident Detection in Smart Transportation Systems, ACM Trans. Cyber-Phys. Syst., Jun. 2023.

# Executive Summary

Traffic monitoring plays a critical role in ensuring road safety and operational efficiency. Identifying anomalies, including non-recurring congestions and traffic accidents and understanding their behaviors have been a target of transportation departments. With the advent of more advanced sensing technologies, concerned parties have had access to many tools such as high-resolution cameras, to better characterize and quantify the potential causes and lasting impacts of such anomalies. However, such installations are either expensive (limiting its reach and impact) or intrusive (impacting the surrounding roadways). The challenge then is to obtain similar insights and high-quality metrics even when such expensive or intrusive solutions are not available. While non-intrusive and less costly solutions exist, they are often fragmented and/or poorly labeled, which make it difficult to use in traditional forecast modelling and AI/machine learning pipelines.

By leveraging the knowledge of our team in big data and machine learning, an approach was designed that collects, combines, and aggregates various datasets related to roadway geometry, weather, historical accidents, and traffic. Then, based on a combination of data cleaning, clustering, and processing techniques, the proposed framework can efficiently detect the spatial-temporal dynamics of a traffic anomaly, even under sparse conditions. We introduce the Traffic Response Anomaly Capture Engine (TRACE), an innovative method that integrates graph neural networks, transformers, and probabilistic normalizing flows to effectively detect and localize traffic anomalies in real time. TRACE leverages spatial-temporal dependencies, addresses data uncertainty, and improves automation, enabling more accurate and timely incident localization. The proposed methods show promising improvement to the current state-of-the-art approaches using various metrics including time-to-detect, localization, and true positive rates. The proposed model for traffic anomaly detection excels in capturing the complex spatial-temporal dependencies inherent in traffic data. This model can optimize and significantly improve the safety of our city streets and highways.

## *Key Findings*

The objectives of this research are threefold. First, we want to develop a technique that can provide sensor fusion at scale and identify outliers. Second, we want to show that the system can dynamically adapt and reconfigure to changing situations across space and time on the highways. Third, we want to design a mechanism to integrate future new low-cost sensors that are bought and integrated into the framework. A spatial-temporal data processing anomaly detection framework was designed to realize these objectives. The pipeline used a combination of data augmentation, spatial-clustering, and learning from data to deliver a hybrid model that excels in capturing the complex spatial-temporal dependencies inherent in traffic data.

The key findings are mentioned below:

- **Detection Accuracy:** TRACE achieved the highest AUC score (0.7124) among the models, indicating its effectiveness in identifying anomalies within traffic data.
- **Timeliness**: TRACE demonstrated competitive response times, ensuring anomalies are identified promptly.

- **Localization Precision**: TRACE’s ability to achieve the lowest localization error (2.55 miles) as compared to baselines highlights its strength in pinpointing the exact locations of traffic anomalies.
- **Robustness**: The hybrid architecture of TRACE proved effective in handling noisy and sparse datasets, maintaining reliable performance across varied conditions.
- **Real-world Applicability**: TRACE detected anomalies in 7 out of 10 cases, showcasing its potential for deployment in dynamic traffic scenarios, though further refinement is needed for edge cases.

The results showcase TRACE's competitive results in detecting and localizing traffic anomalies, with notable improvements localization precision. While TRACE outperformed many baseline models in localization, it showed slight limitations in detection delay and incident coverage compared to STG-RGCN in highly complex scenarios. These findings highlight TRACE's potential as a transformative solution while providing actionable insights for further refinement.
In addition, a large amount of time was invested in collecting, cleaning, combining and aggregating different datasets. For example, traffic data, provided by a company named INRIX, do not cleanly match with crowd-sourced data such as Waze and radar data gathered by NDOT. This requires meticulous crafting of road matching algorithms whose correctness are often very difficult to validate. The lack of validated and precise ground truth data is also a drawback in the analysis and creation of forecasting models. The presence of incidents and anomalies are often crowd-sourced (through reporting and second-hand accounts), which make it difficult to effectively match with the traffic data both spatially and temporally. Furthermore, collaboration with other organizations and firms can be very beneficial. For example, Google or INRIX can provide high resolution traffic data, which might not be available through any other sources.

## *Key Recommendations*

Even though the proposed approach showed promising results, the prediction can be improved through the following:

- **More incident data and higher-quality ground truth labels:** The presence of better labels with incidents in a broader spatial area allows for more advanced machine learning approaches to be applied. The approach is limited to using crowd-sourced data such as Waze, which make it difficult to concretely validate the approach due to potential inconsistencies and the possibility of human error during data collection.
- **Availability of data beyond Nashville and I-24**: This allows us to develop and improve upon the generalizability of the approach for other cities. Building upon the findings, this section provides strategic recommendations for the effective implementation of the proposed model.
- **Optimize sensor deployment based on model insights:** By leveraging the proposed approach, we can identify strategic locations where we can deploy or upgrade sensor placement. Better sensor coverage improves the reliability and accuracy of TRACE’s anomaly detection in critical zones.

- **Develop a real-time traffic anomaly detection dashboard:** Currently, the proposed approach is not user-friendly and geared toward academic research use. By creating a dashboard that uses the proposed approach to visualize anomaly scores, incident locations and current traffic flow data, we can introduce the approach to a larger user base and get feedback on it from the operators and managers.
- **Integration with external datasets and systems**
  Incorporating additional data sources, such as weather data, social media reports, and historical traffic patterns, can enhance the robustness of predictions and improve incident localization. Integration with city-level traffic management systems can enable seamless adoption and facilitate real-time updates.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1 Introduction

Vehicular incidents and accidents have become such a ubiquitous part of life that achieving a goal of having zero deaths seem impossible. According to the National Safety council, from January to June 2023, 21,150 individuals lost their lives in preventable car accidents in the United States [1]. While that number has decreased due compared to the initial six months of the previous year, it is still an unacceptable number. This has prompted action from different agencies, both public, federal, state, and local governments to design and develop principled methods to ensure fast and effective forecasting and detection of incidents, leading to immediate response and action. Companies such as INRIX[1] and Waze[2] gather crowd-sourced traffic and incident reports which they share in both a paid/free manner to different institutions and users. In 2023, the government allocated $1.3 billon for the National Highway Traffic Safety Administration to invest in initiatives that reduce traffic crashes and fatalities on the Nation's roadways [2]. These initiatives include the funding for technology-based and multimodal solutions that improve the travel experience for millions of Americans who use our highway and transit systems, including in disadvantaged communities that have lacked investment and resources [3]. Technology-based approaches include the development of new sensors systems to improve the monitoring and maintenance of the Nation's roadways. With this comes the ability to leverage several sources of information to essentially try and predict traffic incidents and anomalies.

> *From January to June 2023, 21,150 individuals lost their lives in preventable car accidents in the United States.*
>
> *- National Safety Council*

Anomaly detection is referring to the problem of finding patterns in data that do not conform to expected behavior [4]. These nonconforming patterns are often referred to as outliers, exceptions, peculiarities, or contaminants in different application domains. In the transit domain, a traffic anomaly is an event that causes a road section to deviate from its normal traffic flow, such as due to an accident, bad weather, or road work. Anomalies can cause discomfort for drivers and passengers and can degrade traffic efficiency and reduce road safety. Thus, the traffic anomaly detection is essential for maintaining road safety and operational efficiency, yet transportation agencies face persistent challenges in this domain. Swiftly identifying and responding to roadway incidents is critical for minimizing congestion, reducing the risk of secondary accidents, and enabling timely emergency responses.

Traffic incidents—ranging from minor accidents to major collisions—can create cascading disruptions across road networks, leading to significant economic losses and compromised safety. For many agencies, the task of detecting anomalies in real-time across vast and complex

[1] https://inrix.com/

[2] https://www.waze.com/

road networks remains daunting. Traditional manual monitoring is labor-intensive, prone to human error, and can lead to delayed responses, especially under high-stress conditions. Despite technological advancements, existing systems often struggle to reliably identify incidents, particularly in dynamic, densely populated regions. The reliance on manual monitoring in many traffic management centers makes these systems vulnerable to oversights and misinterpretations, underscoring the limitations of current approaches. These challenges point to a pressing need for scalable, automated solutions that can detect and localize incidents with precision and efficiency in real time.

The contributions made in this research can be summarizes as: 1) An efficient pipeline was designed to collect, clean, augment, and combine data from various sources such as Waze, INRIX, radar, OpenStreetMap[3], Weather, and E-Trims [5] incident reports. The fusion of a wide range of data sources is a crucial step in our pipeline since the data is often generated in isolated silos making integration non-trivial but essential for our machine learning engine. 2) A pipeline which can effectively detect and forecast anomalies across the roads and highways. This approach can capture complex spatial-temporal dependencies in traffic data and handle uncertainty and sparsity in sensor measurements. This allows us to provide real-time detection and localization of anomalies within the road network. Finally, we show that our proposed solution outperforms existing methods by better adapting to dynamic traffic conditions and providing more reliable anomaly detection, ultimately enhancing road safety and efficiency.

[3] https://www.openstreetmap.org/

# Chapter 2 Literature Review

Transportation agencies have traditionally relied on rule-based systems and statistical methods for anomaly detection. While these approaches have been foundational, they often fall short in addressing the dynamic nature of traffic data and struggle to support real-time processing. Recent advancements in machine learning have paved the way for more sophisticated models, particularly deep learning techniques that leverage graph neural networks to capture complex spatial-temporal dependencies in traffic data. However, even state-of-the-art methods face limitations in adapting to rapidly changing traffic conditions and managing the inherent uncertainty in traffic prediction. Although these advanced models represent a significant improvement over traditional approaches, critical gaps remain—especially in developing real-time, graph-based methods capable of accurately localizing the root causes of non-recurring congestion.

To understand and address road anomalies, researchers have employed a variety of approaches. Early methods, such as "crash frequency analysis," quantified risk by examining the frequency of incidents within specific spatial areas [6]. Traditional statistical and rule-based models have also been explored, offering simpler but less adaptable solutions. Technological advancements have since driven the adoption of modern machine learning and deep learning techniques, including graph-based and transformer architectures, as well as probabilistic approaches like normalizing flows for anomaly detection. This evolution highlights a shift from manual and heuristic practices toward intelligent, data-driven solutions.

Despite this progress, persistent challenges remain. Issues such as computational scalability, the complexity of spatial-temporal interactions, and uncertainty handling continue to shape the field of traffic anomaly detection. By examining the development of these techniques—from early statistical methods to advanced graph-based deep learning models—this review underscores the ongoing need for innovation to bridge existing gaps in real-time, robust, and scalable anomaly detection systems.

## 2.1 Sensors and Available Data

Sensors are primarily in two forms, intrusive and non-intrusive. Intrusive sensors are essential tools for traffic monitoring, installed directly into roadways. These sensors provide reliable data but require pavement modifications for installation, which can disrupt traffic flow. Key types of intrusive sensors include:

- **Inductive Loops**: These sensors detect changes in the electric field caused by passing vehicles. They are highly accurate for measuring vehicle count and occupancy and are resistant to adverse weather conditions. However, installation and maintenance require significant lane closures, and accuracy may decline with diverse vehicle types.
- **Magnetic Sensors**: Like inductive loops, these sensors detect magnetic field disturbances. They are more compact, withstand inclement weather, and are suitable for areas where loops are not feasible. Limitations include challenges in detecting stopped vehicles.

- **Piezoelectric/Quartz Sensors**: These sensors convert compression forces from vehicle tires into electrical signals, making them effective for vehicle classification. They are less effective in slow-moving or stopped traffic and are sensitive to temperature variations.
- **Pneumatic Road Tubes**: These hollow tubes measure air compression caused by passing vehicles. They are portable and cost-effective, making them ideal for temporary deployments. However, they are not suitable for snowy conditions or multi-lane monitoring.

Non-intrusive sensors offer a less disruptive alternative for traffic monitoring by using overhead or roadside installations. These systems leverage advanced technologies to gather data without direct interaction with the pavement. Key types of non-intrusive sensors include:

- **Infrared Sensors**: These sensors detect thermal emissions from vehicles, providing real-time data and multi-lane coverage. However, performance may degrade in foggy or snowy conditions, requiring frequent maintenance.
- **Radar and Lidar Sensors**: Radar emits microwaves, while lidar uses light waves to measure vehicle speed and location. These systems are highly effective for multi-lane monitoring and environmental versatility but face occlusion challenges and high installation costs.
- **Video Processing Systems**: Using cameras to analyze traffic, these systems offer excellent coverage and adaptability. They are cost-effective but are sensitive to environmental factors like snow, fog, and dust, which can impact performance.
- **Ultrasonic and Acoustic Sensors**: These sensors rely on sound waves to detect traffic flow. They are suitable for multi-lane monitoring but struggle in extreme temperatures and are less effective in detecting stopped vehicles.
- **Vehicle Tracking Systems**: These systems combine GPS, GIS, and RFID technologies to provide precise location data. They offer high accuracy and adaptability but are complex and require robust infrastructure for optimal operation.

Intrusive sensors are best suited for environments where high accuracy is required, and temporary traffic disruptions can be managed. They are widely used for permanent installations in controlled settings but are less practical for high-traffic or adverse weather conditions. On the other hand, non-intrusive sensors are increasingly favored for their ease of installation and reduced impact on traffic flow. They are ideal for urban environments and areas where minimal disruption is essential. However, environmental conditions and ongoing maintenance should be carefully considered during deployment [7], [8], [9], [10].

## *2.2 Traditional Incident Monitoring and Organizational Frameworks*

Transportation agencies like the Tennessee Department of Transportation (TDOT) and other state-level organizations depended heavily on manual incident monitoring and conventional Intelligent Transportation Systems (ITS) strategies [11]. Operators in Traffic Management Centers would observe camera feeds, rely on roadside sensor data, and communicate with law enforcement or emergency responders to detect and manage incidents. Although essential as a foundational approach, these methods have critical limitations.

Studies detail how operator fatigue, human error, and subjective judgment can result in missed or delayed detection of incidents. Under high-stress conditions and dense traffic environments, these limitations become more pronounced. Literature and operational reports repeatedly emphasize the latency between the occurrence of an event and its confirmation when relying on human-driven processes, thus hindering swift incident response and exacerbating congestion and secondary accidents. These initial accounts identify a need for more automated, data-driven methods. This foundational gap set the stage for research on more adaptive and intelligent anomaly detection tools.

## 2.3 Statistical, Rule-Based, and Early Machine Learning Approaches

As data collection capabilities improved through inductive loops, radar sensors, and automated vehicle identification, researchers began exploring statistical and rule-based systems for traffic anomaly detection. This era, spanning the 1990s to early 2010s, introduced methods such as threshold-based deviation detection and basic regression models aimed at identifying unusual drops in speed or sudden halts in flow. Although these offered some relief from manual monitoring, challenges persisted. Early anomaly detection frameworks often set fixed speed or occupancy thresholds to flag possible incidents. While simple and easy to implement, these static thresholds were not robust to dynamically changing traffic conditions. Some research explored unsupervised clustering or simple regression techniques to distinguish normal from abnormal traffic patterns. However, these models struggled with non-recurring congestion and lacked the ability to capture complex spatial-temporal interactions [12]. The literature from this intermediate period reveals incremental improvements in reducing human burden, but it also underscores shortcomings—rigidity, poor adaptability, and limited capability to model complex conditions.

## 2.4 Machine Learning and Deep Learning Methods

The proliferation of large-scale traffic datasets and increased computational power catalyzed a shift toward machine learning (ML) and deep learning (DL) in anomaly detection [13]. Neural networks, capable of modeling non-linear relationships and learning directly from raw data, offered a new paradigm. Early ML-based methods, such as feed-forward and recurrent neural networks (RNNs), improved predictive accuracy for traffic flow and speeds. While not initially targeted at anomaly detection, these architectures laid the groundwork for capturing temporal dependencies [12] Convolutional Neural Networks (CNNs) and Long Short-Term Memory Networks (LSTMs) allowed for spatiotemporal modeling, enabling detection of subtle anomalies by learning patterns across both time and space [14]. Despite enhanced predictive capabilities, researchers noted difficulties in model interpretability, real-time adaptation, and handling the inherent uncertainty in traffic data. Many ML-based methods failed to produce probabilistic outputs or quantify uncertainty, limiting their reliability in operational environments. This generation of methods advanced beyond static rules and heuristics, but they lacked the nuanced handling of complex network structures and did not fully embrace uncertainty modeling.

## *2.5 Graph-Based Methods and Graph Neural Networks*

Recognizing that roads, intersections, and sensors form natural graph structures, recent research has leveraged Graph Neural Networks (GNNs) and related graph-based architectures to better model spatial dependencies [13], [14]. This development aligns with the understanding that effective anomaly detection must consider the topology of the transportation network. By treating intersections as nodes and road segments as edges, GNNs efficiently capture how traffic behavior at one location influences others [14]. Combined with temporal modeling components, these methods achieve a richer representation of network-wide conditions. Literature acknowledges that real-time deployment of GNNs at metropolitan scale remains challenging. Training and inference over large graphs demand significant computational resources and may introduce latency issues. Deterministic GNNs, while powerful in extracting spatial-temporal patterns, generally do not model data uncertainty or measurement noise probabilistically, leaving them vulnerable to inaccuracies when sensor data is sparse or unreliable [15]. This body of work addresses some structural complexity but highlights a new gap: models that can incorporate uncertainty and run efficiently at large scales.

## *2.6 Transformer Architectures and Attention Mechanisms*

Inspired by advancements in Natural Language Processing (NLP), Transformer models—relying on attention mechanisms—have been applied to traffic analysis and anomaly detection [16].By allowing models to focus selectively on different parts of the input sequence, Transformers facilitate long-range temporal dependencies and allow for integration with graph structures, substantially improving its spatial-temporal performance over past approaches. Traditional RNN-based models struggle with very long sequences, whereas Transformers can natively handle extended time horizons, capturing slow-building anomalies or sudden irregularities equally well [10] Combining Transformers with GNNs to simultaneously leverage attention for temporal patterns and graph convolution for spatial patterns is a promising approach to handle both spatial complexity and temporal dynamics; Despite their strengths, Transformers introduce their own challenges, including higher computational costs and the complexity of model tuning for time-sensitive, real-time applications. Moreover, attention mechanisms alone do not inherently address uncertainty quantification.

## *2.7 Probabilistic Modeling and Normalizing Flows*

As transportation research increasingly recognizes uncertainty as a fundamental characteristic of traffic data, probabilistic modeling has emerged as a critical frontier. Normalizing Flows (NFs) represent a powerful class of methods for learning flexible probability distributions [17], [18]. NFs can provide full probability density estimates of traffic states, enabling more robust anomaly detection by highlighting not just a point estimate but a likelihood of different traffic conditions [17]. However, while widely studied in the context of computer vision and NLP, Normalizing Flows have yet to be thoroughly integrated into GNN- and Transformer-based anomaly detection frameworks. Combining these methods could yield a model that both understands complex spatial-temporal structures and probabilistically quantifies uncertainty [18]. This gap highlights a significant research opportunity. Although probabilistic methods have proven beneficial in other domains, their potential in traffic anomaly detection—particularly when integrated with powerful spatial-temporal architectures—remains underexplored.

The Normalizing Flow approach integrates prediction-based and density-based methods, effectively addressing the limitations of each. Prediction-based models operate on the premise that normal samples exhibit greater predictability compared to anomalies, with stacked LSTM models being a common implementation [19], [20], [21]. For instance, the effectiveness of bidirectional LSTM (BiLSTM) in freeway traffic forecasting was explored in [22], while Basak et al. [23]examined the cascading effects of traffic congestion using a citywide ensemble of connected LSTM models at the intersection level.

However, a key drawback of such methods is that forecasting accuracy can be compromised when models are trained on datasets containing anomalies [24], resulting in unreliable anomaly detection. Density-based approaches, on the other hand, identify anomalies by leveraging the principle that the density around normal samples is consistent with their neighbors. For example, Chiang et al. [25] proposed a two-step strategy for identifying congestion cascades: (1) computing anomaly scores for road segments using a kernel density function and (2) grouping congested segments based on spatial-temporal closeness and attribute coherence. Similarly, Dias et al. [26] utilized RealNVP and masked autoregressive flow models for trajectory anomaly detection, demonstrating that flow models outperform traditional density-based techniques. While density-based methods do not require labeled data [24], they are limited by their focus on data distribution, making them unable to capture sequential dependencies in time series data.

# Chapter 3 Methodology

This section presents a detailed methodology for traffic anomaly detection, encompassing temporal and spatial modeling, data representation, anomaly scoring, and optimization. The proposed approach integrates state-of-the-art machine learning techniques to ensure robust and precise anomaly detection. The problem settings used in this project is shown in **Error! Reference source not found.**.

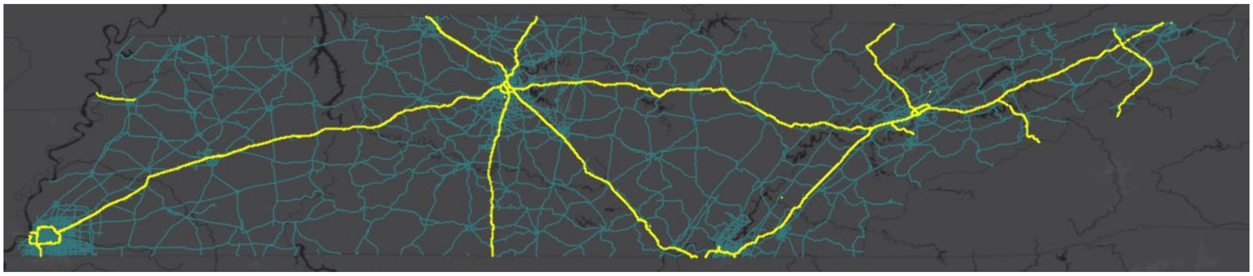


**Figure 1 Blue (and thin) lines represent TN's roadway network. Yellow (and thick) lines represent interstate highway segments under the jurisdiction of TDOT and are the area of study for this study**

The following subsections describe the approach and the mathematical formulation of the problem.

## *3.1 Problem Formulation*

Consider a spatial area of interest where $V$ denote the set of all road sensors under consideration. Consider an arbitrary sensor $v_i \in V$ on which (near) real-time speed is monitored continuously; we assume that the estimated harmonic mean speed on sensor $v_i$ is computed and stored at discrete times $t \in \{1, 2, \ldots, T\}$. We denote this observation at time $t$ by $s_i$.

This problem can be converted into a graph problem. Consider now a $G = (V, E)$ be a directed graph representing the road network, where $V$ is the set of nodes, representing road sensors and $\mathrm{E} \subseteq \mathrm{V} \times \mathrm{V}$ is the set of undirected edges, representing the connectivity between road sensors. Thus, each $v \in V$ corresponds to a sensor in the area of interest, with weights $x^t$ representing the congestion at time $t$. Thus, a graph $G(t)$, is a snapshot of congestion across an entire road network for an arbitrary time $t$. Each time step can have additional features associated with it, e.g., day of the week and hour of the day. We denote such features for the $t$-th time step be $\lambda^t$.

The primary goal is to detect and localize anomalies in the latent representation of the traffic conditions. The latent representation of the graph at time $t$ is denoted as

$Z(t) \in \mathbb{R}^{M \times H}$

Where $M = |V|$ is the number of road sensors in the graph, $H$ is the dimension of the latent space, representing the encoded attributes of the adjacency and feature matrices, and $Z(t)_v$ is the latent vector of node $v$ at time $t$, capturing its spatial-temporal characteristics. The latent representation over a time window $T$ is thus represented as a sequence of latent metrices:

$$\mathbb{Z} = \{\mathbf{Z}(t)\}_{t=1}^{T} \in \mathbb{R}^{M \times H \times T}$$

Where $\mathbb{Z}$ encapsulates the spatial-temporal relationships and attributes of all edges across all timesteps in a compact and meaning representation. Finally, anomalies are defined as significant deviations from expected patterns in the latent space of all nodes over time. Specifically, an anomaly is identified for a node $v$ at time $t$ when:

$$Anomaly\ A(v,t) = \begin{cases} 1\ if\ p(Z(t)_v, \cdot) < certain\ threshold \\ 0 \quad Otherwise \end{cases}$$

The objective of the problem then, is to find the weights of the model that maximizes the likelihood of the observed values across all edges and time steps:

$$max \sum_{t=1}^{T} \sum_{v \in V} log\ p(\mathbf{Z}(t)_v)$$

The objective ensures that the model accurately captures normal relationships within the network, enabling effective anomaly detection and localization.

## *3.2 Challenges*

The problem described in the previous section is difficult due to the following challenges:

*Data Sparsity*: Incident reports and real-time data streams often provide incomplete information about road conditions, leading to gaps in understanding the current traffic situation. Additionally, traffic incidents that occur between otherwise normal traffic patterns may be considered an anomaly. However, characterizing a normal traffic pattern that works at a large city scale is not straightforward due to (i) day-to-day variability of traffic, (ii) local neighborhood dependencies, (iii) a large number of speed sensors and road segments. Therefore, the anomaly detection problem is much more challenging and requires novel advances compared to existing theories of anomaly detection in CPS.

*Complex Graph Structures*: Intricate dependencies within road networks complicate traffic pattern analysis and substantially increase the computational load required to effectively process them. Additionally, the complex traffic network introduces difficulties in data integration which increases the scope not only temporally but spatially as well. For large geographic areas, this means that traffic congestion in a specific segment could be influenced by not only temporal factors but spatial factors such as roads leading to and out of the section of road.

*Incident localization and lack of high-quality labels*: Precisely pinpointing an incident's location remains challenging, impacting response effectiveness, especially regarding the scale of the area. The problem of monitoring large sensor streams to detect incidents of interest is basically an unsupervised anomaly detection problem in multivariate time series (MTS); traffic data pertaining to road segments are collected across time and span several dimensions. While anomaly detection has been traditionally done for various traffic condition variables with techniques such as CUSUM [27], K- nearest Neighbor [28], Isolation Forest [29], and forecasting models (e.g., ARIMA [30]), deep neural networks (DNN) have gradually become the state-of-the-art due to the remarkable capability of modeling high-dimensional MTS data. However, despite the universal approximation power of DNN on learning unknown data distributions, performing anomaly detection on MTS is still challenging. For example, many DNN-based approaches either rely on an uncontaminated training dataset to learn the normal traffic patterns (semi- supervised) or reframe the detection task as a classification task using a fully labeled traffic mobility dataset (supervised). These issues underscore the need for solutions capable of handling complex, large-scale traffic data in real time, while addressing both computational and localization challenges.

## *3.3 Data*

The covariates used and their sources are described in Table I. It is crucial to reiterate the importance of this stage in real-world machine learning pipelines; in fact, the availability of multiple streams of data has been noted as being particularly important for predicting accidents [31].

**Table I Data Features, Size and Sources**

| Dataset | Range | Size | Rows | Feature | Source | Freq. | Type | Definition |
|---|---|---|---|---|---|---|---|---|
| - | - | - | - | Time of day | derived | - | Temporal | We divide each day into six 4-hour time windows. |
| - | - | - | - | Weekend | derived | - | Temporal | A binary feature that denotes weekdays. |
| Incident | 06/01/2023 to 01/31/2024 | 21MB | 80K | Past Incidents in the last window | derived | - | Spatio-temporal | Number of incidents on the segment in the last time window of 4 hours |
| | | | | Past Incidents in a day | derived | - | Spatio-temporal | Number of incidents on the segment in the last day |
| | | | | Past Incidents in a week | derived | - | Spatio-temporal | Number of incidents on the segment in the last week |
| | | | | Past Incidents in a month | derived | - | Spatio-temporal | Number of incidents on the segment in the last month |
| Traffic | 04/01/2017 to 12/01/2020 | 1.2TB | 30B | Congestion | derived | 5 minutes | Spatio-temporal | Congestion is the ratio of the difference between free flow speed and the current speed to free flow speed |
| | | | | Free Flow Speed | INRIX | 5 minutes | spatial | The speed at which drivers feel comfortable if there is no traffic and adverse weather condition. |
| | | | | Traffic Confidence | INRIX | 5 minutes | Spatio-temporal | A confidence score regarding the accuracy of the traffic data (we collect this directly from INRIX). |
| Sensors | 10/01/2023 to 10/31/2023 | 138MB | 3M | Speed | RDS | 30 sec | Spatio-temporal | Speed of traffic capture by the particular sensor |
| | | | | Volume | RDS | 30 sec | Spatio-temporal | Volume of traffic capture by the particular sensor |
| | | | | Occupancy | RDS | 30 sec | Spatio-temporal | Occupancy of traffic capture by the particular sensor |
| | | | | Crash Record | Derived | 30 sec | Spatio-temporal | Incident (recorded in official report) occurring at a particular sensor |
| | | | | Human Label | derived | 30 sec | Spatio-temporal | Expert labeled anomalies using irregularity in speed patterns across sensors |
| Roadways | Static | 81MB | 80K | Lanes | RDS | static | Spatial | Number of lanes for a roadway segment. |
| | | | | Miles | derived | static | Spatial | Length of a roadway segment. |
| | | | | iSF | derived | static | Spatial | Inverse scale factor which represents the the curvature of a roadway segment. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Waze | 06/01/2023 to 1/31/2024 | 104MB | 700K | uuid | Waze alerts | - | Spatio-temporal | Event ID of Waze incident |
| | | | | geo | Waze alerts | - | Spatio-temporal | Spatial (lat, lon) of Waze incident |
| | | | | Datetime | Derived | - | Temporal | Time of Waze incident |
| | | | | | | | | |

In the appendices section, some of the main challenges in the data collection and integration are presented along with the approaches that were taken to address them.

## 3.4 Features

This section describes the features we extract from the base data Table I and use as covariates in our pipeline.

*Traffic*: We use an INRIX traffic mobility data collected for one year (2019) from the city of Nashville, Tennessee. The details are summarized in Table I. Specifically, this dataset contains estimated "real-time" harmonic mean flow speeds, free-flow (reference) speeds, and historical average speeds of 364 interstate road segments with a five-minute frequency. The congestion rate measurements are derived based on the equation 4. We impute missing values at a specific road segment by interpolating observations from nearby segments. If nearby segments also contain missing values, we impute by using historical averages

*Incidents*: Every accident reported in Tennessee from October 2023 was considered. Incident data for this project is provided by the TDOT and consists of approximately 12 accidents. The accuracy of the incident data was verified with the Enhanced Tennessee Roadway Information Management System (E-TRIMS).

In summary, the following features were used: time of day, weekend, past incidents in the last window, past incidents in a day, past incidents in a week, past incidents in a month, visibility, wind speed, precipitation, temperature, congestion, free flow speed, and traffic confidence.

## 3.5 Spatial Matching

An integral objective of this project was to develop a comprehensive framework for sensor fusion, integrating data from diverse sources to provide a unified and holistic representation of traffic dynamics across multiple regions.

*OSM-INRIX Matching:* spatially conflate INRIX traffic segments with OpenStreetMap (OSM) road segments to build dynamic graph structures that represent real-time traffic data. These graphs were used in machine learning tasks, specifically for detecting anomalies in traffic flow patterns across Davidson County. The research involved key processes such as INRIX-OSM matching, bearing and geospatial filtering, and dynamic graph generation over time.

To achieve accurate INRIX-OSM data conflation, the INRIX traffic segments, which provide real-time traffic information like speed and congestion, were matched with OSM road segments that offer detailed, crowd-sourced road geometries. Initially, INRIX segments were merged with corresponding OSM road segments based on shared segment IDs, though each INRIX segment could map to multiple OSM segments. Bearings for each OSM road segment were calculated and classified into compass directions (e.g., North, East) to compare with INRIX segment directions. This ensured that only directionally aligned segments were retained, while mismatches were excluded.

To handle multiple INRIX segments matched to a single OSM edge—a common issue due to the connected nature of OSM ways (see Figure 7)—matches were grouped and filtered. A geospatial join was then performed using buffers around OSM edges, INRIX segments, and OSM nodes to account for minor spatial inaccuracies. Overlapping geometries were iteratively refined by increasing buffer sizes when necessary, which significantly improved the robustness of the matching process and reduced the number of INRIX segments associated with each OSM edge. Only those INRIX segments that satisfied both directional and spatial criteria were retained, improving the precision of the conflation process. This filtering step ensured that each OSM edge was associated with the most relevant INRIX segments (See Figure 8).

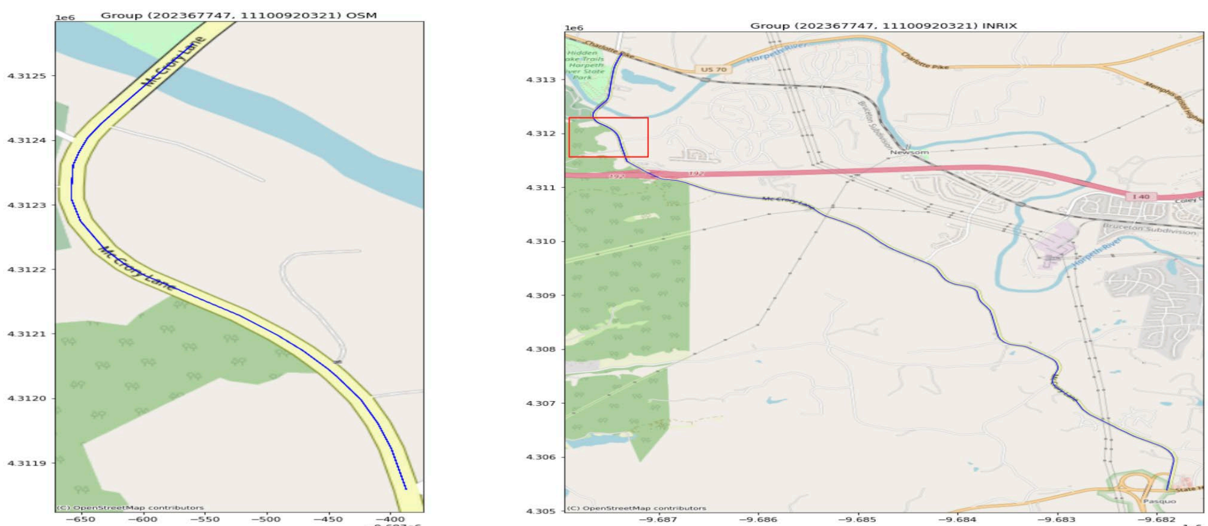



**Figure 2: OSM Edge (Blue Line) (202367747-1110920321) on left and Corresponding INRIX Segments(Blue Line) for OSM edge (in the red box) on right**

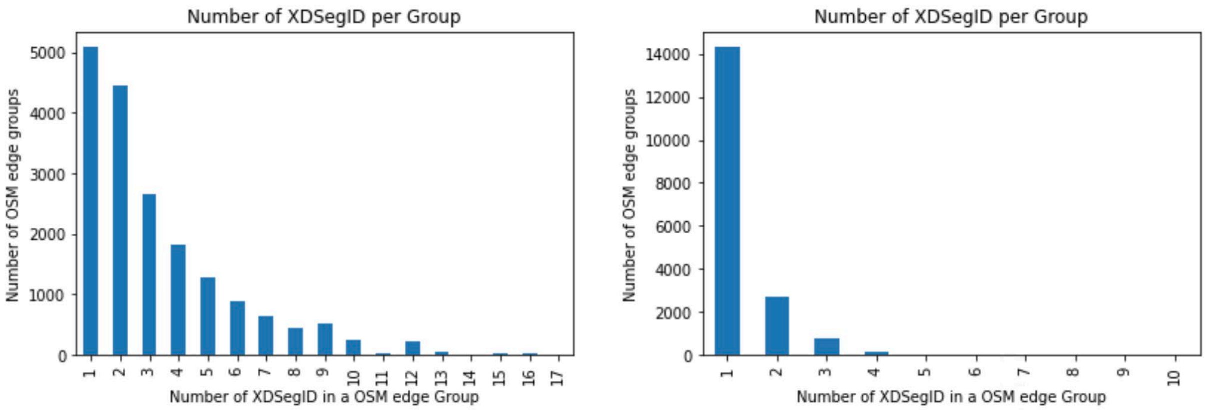



**Figure 3: Number of Multiple INRIX Segments matches per OSM edge before (left) and after (right) geospatial join within each OSM edge group**

The final step involved constructing dynamic graph structures using the conflated INRIX and OSM data. INRIX traffic data from January 2023 to March 2024 was processed at 5-minute intervals, aggregating traffic speeds, reference speeds, and congestion levels for each time step. A directed graph was then created, where nodes represented road intersections and edges captured traffic flow between them. Edge attributes such as speed, reference speed, confidence scores, travel time, and congestion were calculated using the harmonic mean across time periods. This process generated files containing dynamic graphs over the specified time range, reflecting the temporal evolution of traffic patterns.

The results demonstrated the effectiveness of the proposed conflation method. After the initial merge, **16.95%** of unique OSM edges were retained. Following bearing and geospatial filtering, **91.46%** of the conflated edges were accurately matched. The dynamic graphs successfully captured temporal variations in traffic patterns across Davidson County, enabling anomaly detection through deviations in edge attributes over time. This work highlights an efficient method for combining INRIX and OSM datasets to create accurate, dynamic traffic graphs. The robust geospatial and directional filtering processes ensured a precise representation of traffic conditions, providing a strong foundation for analyzing congestion and detecting irregularities in real-time traffic data.

*INRIX-RDS Sensor Matching:* To perform spatial matching between RDS sensors and INRIX traffic segments, a similar approach was employed as in the INRIX-OSM conflation process. A buffer was created around the geographical points representing the RDS sensor locations. This buffer accounted for minor spatial discrepancies and ensured that nearby INRIX segments could be accurately identified. A spatial join was then conducted between the RDS sensor buffers and the INRIX segments, identifying all INRIX segments that passed through the buffer.

In cases where multiple INRIX segments were matched to a single RDS sensor, the segment with the smallest distance to the sensor was retained. This distance-based filtering ensured that the most relevant INRIX segment was associated with each sensor, improving the precision of the spatial matching process.

The results of this matching process can be visualized in the figure 9. The figure illustrates the spatial relationship between the RDS sensors, shown as points, and the INRIX segments, represented as line geometries. For clarity, the matched sensors and their corresponding INRIX segments have been color-coded uniformly. This visual representation highlights the accuracy of the matching process and provides an intuitive understanding of the spatial alignment between sensor locations and traffic segments.

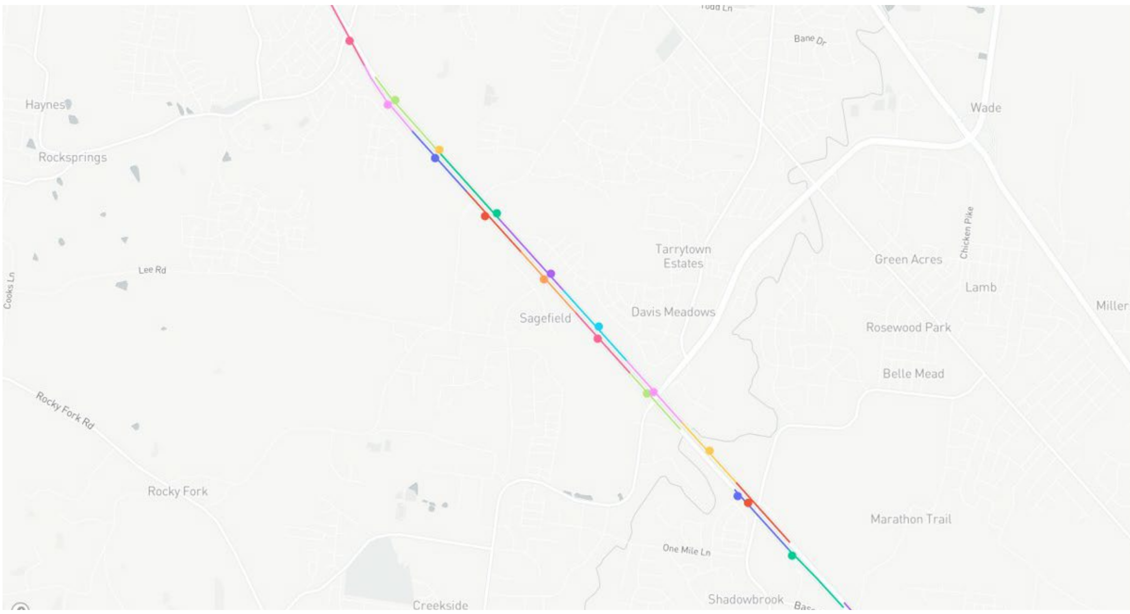



**Figure 4: INRIX Segment to RDS Matching**

## *3.6 Approach*

This subsection presents a detailed methodology for traffic anomaly detection, encompassing temporal and spatial modeling, data representation, anomaly scoring, and optimization. The proposed approach integrates state-of-the-art machine learning techniques to ensure robust and precise anomaly detection. The framework of the methodology is shown in **Error! Reference source not found.**.

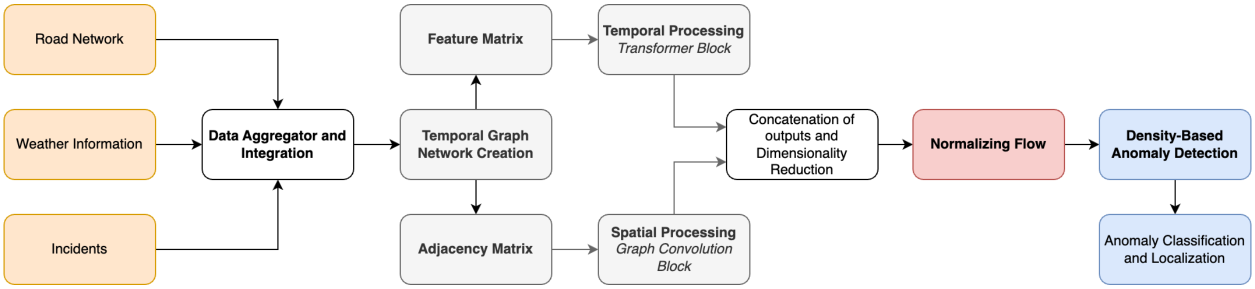



**Figure 5: TRACE framework**

The framework consists of following components (1) Data Aggregation and preprocessing, (2) Clustering, (3) temporal modeling with transformers and Spatial Modeling with Graph Convolutional Networks, (4) Latent Representation of Traffic Data, (5) Anomaly Scoring with Normalizing Flows, and (6) Threshold Identification and classification.

*Data aggregation:* Data streams are collected and joined at the spatial and temporal levels. However, real-world mobility data collected from the wild (not from the experimental testbed), pose a practical problem for unsupervised learning problems such as anomaly detection, due to the presence of various incidents in the training phase. This prevents the learning of the underlying structure of benign data patterns. Thus, there is a need for a mechanism to bypass this problem. The intuition is to use the time and location stamp of the ground truth incidents and superimpose them on the sub-network which falls under the location of a particular incident. Then we identify the neighborhood of the time series data around all incidents to learn the portions of the time series that were disturbed. Unless these disturbances are cleaned out, it will prevent learning the structure of the benign behavior. Note, ground truth incident recording itself is noisy due to human-in-the-loop issues. We observed in many cases, they are recorded much after the physical world has been affected by the incident. In other cases, the incident is reported and recorded instantly but it takes some time for the physical world to get really affected (e.g. in sparse traffic scenarios).

*Clustering*: In practice, the road network might be composed of thousands of dimensions with heterogeneous temporal patterns, semantic meanings, or underlying dependencies. It is computationally difficult to learn patterns or explicit probability distributions for extremely high-dimensional data. One way to alleviate this challenge is by identifying dimensions that are related in the feature space. To tackle this, we perform data-driven clustering to partition the given time series into separate groups based on similarity (where similarity is based on an appropriate distance in the feature space, e.g., the $\ell 1$norm). This step facilitates anomaly detection and diagnosis in two aspects. First, it ensures that learning the probability distribution over the input time series is tractable. Second, similarity in the feature space naturally associates semantic meaning to the clusters, e.g., we observe that different clusters correspond (roughly) to different types of roads such as highways and on-ramps. Learning an explicit distribution fora particular cluster therefore enables us to learn a distribution of traffic in a particular type of roadway

We need to strategically group the road segments into spatial clusters such that the speed data has maximum positive correlation which leads to the highest invariance. At the same time, the clustering needs to be geographically proximate for disturbances in the co-variance structure to have a causal link to the traffic incidents.  All the road segments exhibiting correlations above a threshold may be grouped together to form a cluster. Thereafter, if $C = \{c_1, c_2, \ldots c_K\}$ is a candidate cluster set and $s_i$ and $s_j$ are any two road segments where $i \neq j, 1 \leq i, j \leq n$ such that $s_i$ and $s_j$ are in the same cluster $c_k$ we can formalize the problem as the following:

$$max \sum_{c \in C} \sum_{\{s_i, s_j\} \in c} Cor(s_i, s_j)$$

such that $Cor(s_i, s_j) > p^{(min)}$ where Cor is a correlation between the two road segments and $p^{(min)}$ is a threshold. The above optimization problem is NP hard since with $S$ number of road segments, there is an exponential number of possible solutions which is computationally intractable. We need an approximation to the exact solution. This is done by first converting the clustering problem into a graph problem.

Theoretically, a correlation may exist between any pair of road segments. Therefore, the initial graph is a complete graph. However, since all road segments are not necessarily positively correlated (e.g. geographically distant, city roads to highways in the same geographical area), there will be edges with negative or zero weights and relatively low weights. Let there be a bound on the minimum correlation value $p^{(cut)} > 0$ necessary to be considered a feasible edge of the graph. All edges whose weights are less than $p^{(cut)}$ are pruned from the complete graph. A low $p^{(cut)}$ affects the level of invariance in the ratio invariant which is key to a low false alarm and improved detection performance.

From the feasible set, we introduce a notion of desirability to form the strongest grouping of clusters. Note, only using Euclidean distance is not appropriate for causal link because a narrow lane may have a highway road segment running over it. Geographically they are close, but even if one incident is affecting a ramp connecting the two, the correlation will not be as strong due to inherent differences in their physical characteristics. Also, some roads are long and can see an incident's effect quickly propagate, and segments not geographically very close still become affected by that same incident when not geographically close. As explained earlier, geographically closer road segments will be affected by the same incident. Hence, the distance should be factored in the clustering too.

The process behind clustering algorithm is shown in **Error! Reference source not found.**. In begins with a region $Region(v_{init}, r)$ is the set of road segments that are within the area with radius $r$ from an initial vertex $v_{init}$. All the surrounding nodes within a given distance, that are directly connected to any node within the initial region is considered as candidate nodes for inclusion to the cluster. Segments that are highly correlated are included to the current region. This region continues to grow until all neighboring nodes (which are not part of the region) are all low correlation. This means that by adding this segment to the cluster, the region's correlation will drop. The process then randomly selects a new segment that is not part of any cluster and repeats the process.

Once all the segments have been included in a cluster, the process stops and returns all found clusters. This process ensures that the spatial and temporal partitions of the high dimensional data have a high positive correlation.

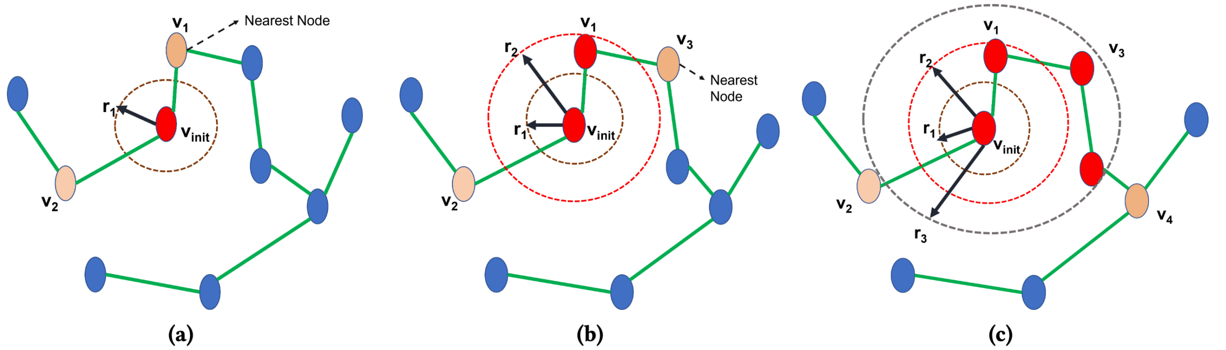



**Figure 6: (a) A region around the initial node. (b) Increased radius for the region based on the nearest segment whose one connecting end is inside the previous region. (c) A region where the volume is greater than cut.**

*Graph Network Creation:* We convert the results of our clustering into a graph, where we visualize each road segment as a vertex on the graph $G$ and the road segment connections as an edge. The attributes of a node are equal to the different temporal features based on the aggregated dataset while the adjacency of the network is defined by the spatial interconnectivity of the road network. The **temporal dependencies** within traffic data are modeled using Transformer networks. These networks leverage a self-attention mechanism to identify sequential patterns in traffic flow across time. At the same time, they capture dependencies over varying time horizons, enhancing adaptability to dynamic traffic conditions. The Transformer's ability to focus on relevant time intervals makes it particularly effective in processing irregular traffic patterns and predicting anomalies. The **spatial modeling** is passed through **Graph Convolutional Networks (GCNs)**. GCNs are utilized to model the spatial relationships inherent in road networks. These networks process adjacency matrices to represent the connectivity between road segments. It then uses these feature matrices to encode sensor data, such as traffic speed, volume, and congestion levels. By integrating spatial dependencies, the GCN layers ensure that the model accurately reflects the interconnectivity and influence of adjacent road segments, enabling precise identification of spatial anomalies.

*Latent Representation of Traffic Data:* The traffic network is transformed into a latent space representation, which serves as a compact summary of sensor data and spatial relationships. Key components include: $M$ which represents the total number of sensors monitoring the network and Latent Space Dimensions ($H$) which encodes critical attributes derived from adjacency and feature matrices. Latent representations are generated for each time step and aggregated over a predefined time window *T*, creating a sequence that captures evolving traffic dynamics. This aggregation enables the model to analyze patterns over time, improving its robustness against noise and sparse data. The latent representation over a time window T is represented as a sequence of latent matrices:

$$Z = \{Z(t)\}_{t=1}^{T} \in \mathbb{R}^{M \times H \times T}$$

where ***Z*** encapsulates the spatial-temporal relationships and attributes of all edges across all timesteps in a compact and meaningful representation.

*Anomaly Scoring with Normalizing Flows:* A normalizing flow model is employed to compute the likelihood of latent representations. This probabilistic approach offers several advantages over other existing approaches. It allows us to estimate the likelihood of an anomaly. A low value implies the observation is either rare in the input space or deviates from contextual behavior**.** This measures how closely the observed data aligns with normal traffic behavior. Based on this anomaly score, we can identify instances of deviations from high-likelihood regions which could indicate potential anomalies. This requires the presence of some form of threshold, the process of obtaining will be discussed in the next sections.

We define a normalizing flow model $f : R^{M \times H \times T} \rightarrow R^{M \times T}$ that estimates the likelihood of the observed latent representation for all edges:

$$log\, p(Z(t)) = [f\ (Z(t)_1, \cdot), f\ (Z(t)_2, \cdot), \ldots, f\ (Z(t)_M, \cdot)]$$

where $log\, p(Z(t))$ is a vector of log-likelihoods for all edges at time *t*.

The anomaly score for each node at each timestep is calculated as the negative log-likelihood:

$$L(v, t) = -\, log\, p(Z(t)_v, \cdot)$$

The anomaly score for a road segment at a given time step is computed as the negative log-likelihood of its latent representation. This score highlights deviations from expected traffic patterns, enabling fine-grained detection of anomalies.

*Training and Optimization Objective*: Given the anomaly score which is based on the likelihood of a feature being present in the given context, we configure the training objective on maximizing the likelihood of observed data across all road sensors and time steps. This involves learning the distribution of normal traffic patterns and fine-tuning model parameters to enhance predictive accuracy. The optimization objective is formulated as:

$$max \sum_{t=1}^{T} \sum_{i=1}^{V} log\, p(Z(t)_{v\_i}, \cdot)$$

This objective ensures that the model accurately captures normal relationships within the network, enabling effective anomaly detection and localization.

*Threshold Identification for Anomalies:* Finally, once we obtain a model that can provide anomaly scores based on the estimated likelihood, a thresholding algorithm is then applied to classify traffic states based on the computed likelihood scores. The threshold separates normal from anomalous traffic behavior and is often exogeneous, based on unseen data such as the validation dataset. We then use this threshold on some test dataset to validate the effectiveness of the approach while accounting for generalizability of the anomaly scores and network conditions. This process ensures consistent and reliable anomaly detection, even in the presence of noisy or incomplete sensor data.

The methodology combines temporal modeling, spatial analysis, and probabilistic scoring to deliver a comprehensive framework for traffic anomaly detection. By integrating Transformers, GCNs, and normalizing flows, the approach addresses the complexities of real-time traffic monitoring, offering robust and scalable solutions for anomaly detection and localization.

*Supervised Anomaly Classification:* Given the normalizing flow model (that can perform exact density estimation and efficient sampling) and the LSTM-EncDec model (that can capture temporal correlations), we can generate labeled synthetic data to train supervised classifiers for anomaly detection. The procedure of generating MTS sequences are as follows: we first provide a warm-up sequence (an initial context window) as the input of the encoder to produce the decoder's initial hidden states. Anomalies and normal samples are then sampled from a standard normal distribution and then transformed to the output space (with decoder hidden states as conditional inputs). Generated samples are reused as inputs of the next iteration until the desired time series length is reached. The samples can then be used to train a classifier. We use a multi-layer perceptron classifier in our analysis.

# Chapter 4 Results and Discussion

Successful anomaly detection requires models that can predict the spatial-temporal likelihoods of traffic incidents. Effective anomaly detection comprises of high detection rates with minimal false positive rates. Additionally, providing anomaly detection models to localize and identify incidents over space and time presents unique advantages to applications such as emergency response. Delayed detection and response from first responders or emergency management agencies can worsen into heavy city-wide congestion and even in the loss of life.

To evaluate our framework, we use the FT-AED dataset that has been curated to provide high-resolution, lane-level data collected via 49 radar detection sensors installed along Interstate 24 (I-24) (Coursey et al., 2024). This stretch, spanning 18 miles near Nashville, TN, captures critical freeway dynamics during peak traffic hours. We use limit dataset to a temporal scope of morning peak hours (4:00 AM to 12:00 PM) over October 2023. The dataset includes the traffic speed, volume, and occupancy recorded every 30 seconds and features 12 confirmed crash reports from TDOT and 8 expert-labeled anomalies. The dataset's granularity facilitates the detection of subtle anomalies that traditional aggregated data sources often miss. 4 (layout of sensors) and 5 (traffic speed visualization) illustrate the data's scope and structure. Similarly, Figure 6 shows the incidents and their effects on the speeds.


**Figure 7: Layout of radar detection sensors along Interstate 24 near Nashville, TN. Source: Coursey et al., 2024**

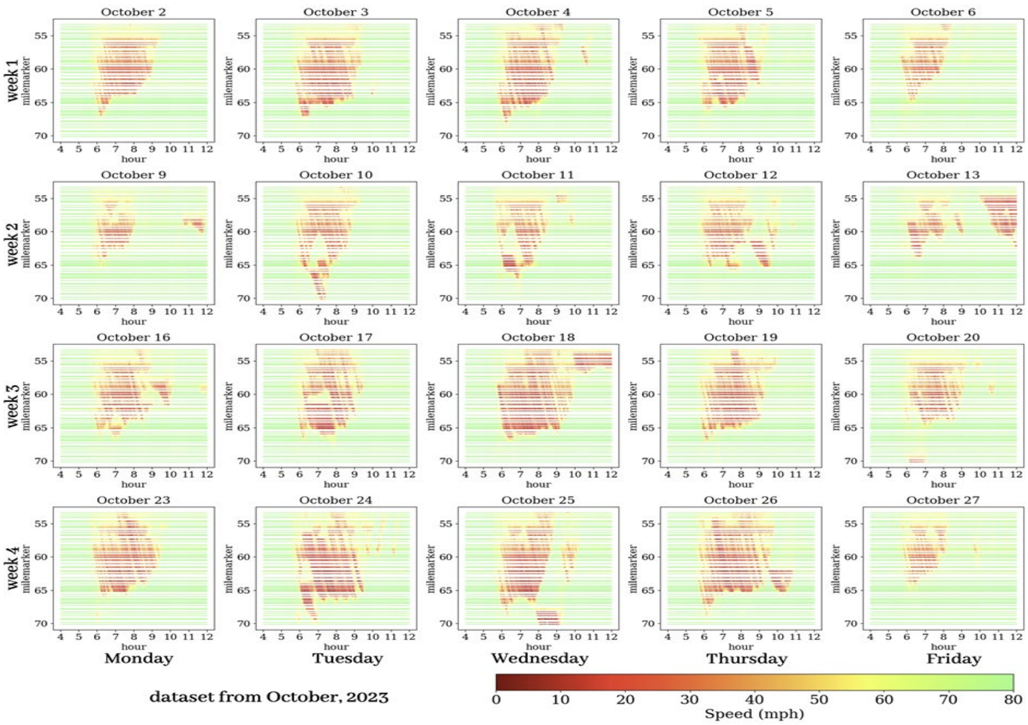



**Figure 8: Morning peak-hour traffic speeds along I-24. Source: Coursey et al., 2024.**

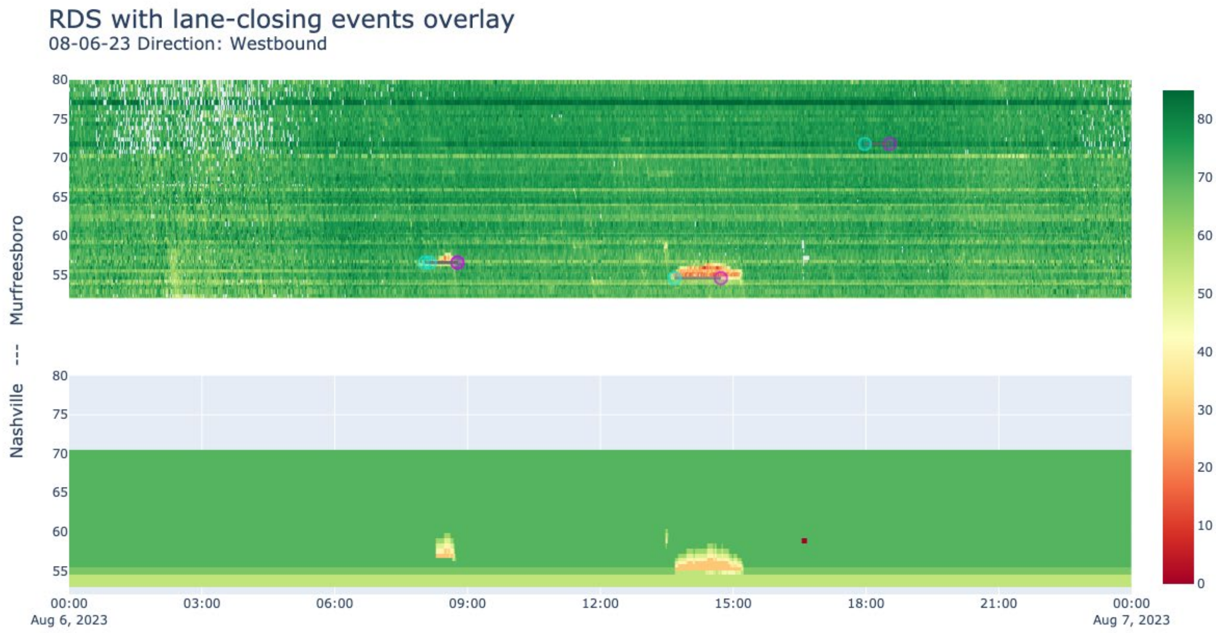



**Figure** 9**: Visualization of lane-closing event and its effect on the speed on I-24 W**

## 4.1 Model Hyperparameters

Hyper-parameters for each model were tuned by cross-validation. Training configuration, model architecture and temporal settings were all tuned in this way. The model parameters are described in

Table II.

Table II TRACE Hyperparameters and their optimal values.

| Hyperparameter | Type | Description | Optimal Value |
|---|---|---|---|
| **Dropout rate** | Training | Probability of dropping neurons during training to prevent overfitting. | 0.0012 |
| **Learning rate** | Training | The rate at which the model adjusts weights during optimization. | 0.0006 |
| **Epochs** | Training | Number of complete passes through the training dataset. | 2 |
| **Flow layers** | Model Architecture | Number of flow layers in the normalizing flow model. | 5 |
| **N hidden flow** | Model Architecture | Number of hidden layers within each flow layer. | 1 |
| **Hidden layers** | Model Architecture | Size of the hidden layers in the model. | 16 |
| **GCN layers** | Model Architecture | Number of Graph Convolutional Network (GCN) layers for graph-based learning. | 1 |
| **Transformer layers** | Model Architecture | Number of transformer layers to capture sequential dependencies. | 2 |
| **Attention heads** | Model Architecture | Number of attention heads in transformer layers for multi-head self-attention. | 1 |
| **Output dimensions** | Model Architecture | Dimension of the latent representation. | 32 |
| **Timesteps** | Temporal | Number of timesteps used for processing temporal or time-series data. | 2 |

*Training Hyperparameters:* A low dropout value indicates minimal neuron dropping, likely to retain most features, as the risk of overfitting is low. A small learning rate ensures stable convergence without overshooting the optimal solution. Finally, a minimal number of epochs might indicate that the dataset or model converges quickly.

*Model Architecture Hyperparameters*: A flow layer count of 5, balances the expressive power of the normalizing flow model without making it too complex. A low number of hidden flow layers keeps the flow model lightweight, suitable for smaller or moderately complex datasets. GCN layer of 1 indicates minimal graph depth, often effective for datasets where deeper graph representation isn't required. This means that our traffic network does not benefit from deeper representations, probably due to the linear nature of the highway.

*Temporal Hyperparameters:* A small temporal window of two, may suggest a task that relies on immediate, short-term dependencies rather than long-term ones. Larger time windows decreased the ability of the model to accurately identify anomalies.

## 4.2 Experiment Plan

The goal of this experiment is to evaluate the effectiveness of machine learning methods in detecting anomalous events on freeways using the FT-AED dataset. Specifically, the study aims to benchmark models for their ability to detect anomalies in traffic data with high accuracy, reduce reporting delays for crashes and other incidents, and accurately localize anomalies at the lane and mile-marker level.

The experiment involves three datasets: training, validation, and test sets. These datasets are tailored to capture various traffic conditions and ensure comprehensive model evaluation.

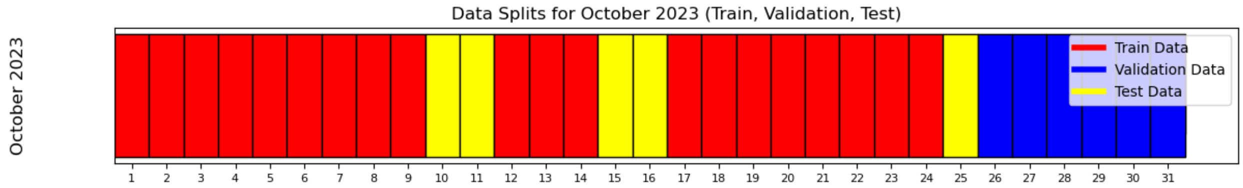



**Figure 10 Dataset Setup: The days in yellow are the detection target as it contains crashes of interest. The days in red are used to train the models for detection and days in blue are used as validation dataset for hyperparameter tuning and adjusting threshold.**

The experiment involves three distinct datasets—training, validation, and test sets—designed to capture various traffic conditions and enable comprehensive evaluation. The training dataset comprises 80% of the month's morning traffic data (4:00 AM to 12:00 PM), including anomalies, to train models for generating negative log-likelihoods (used in TRACE) or reconstructing traffic conditions (for baselines) and identifying deviations. The validation dataset includes 10% of the morning data, covering diverse scenarios such as multi-lane congestion and isolated incidents, with the objective of fine-tuning model hyperparameters and dynamically adjusting anomaly detection thresholds. Finally, the test dataset consists of five specific days in October 2023 (10, 11, 15, 16, and 25) and is used to evaluate model generalization on unseen, challenging traffic scenarios.

To assess the performance of these models, the study employs four key evaluation metrics. The **Reduction in Reporting Delay** measures the time saved in anomaly detection compared to official reports, demonstrating the models' ability to accelerate incident response. The **Anomaly Detection Rate** evaluates the proportion of true anomalies successfully identified, ensuring critical events are effectively captured.

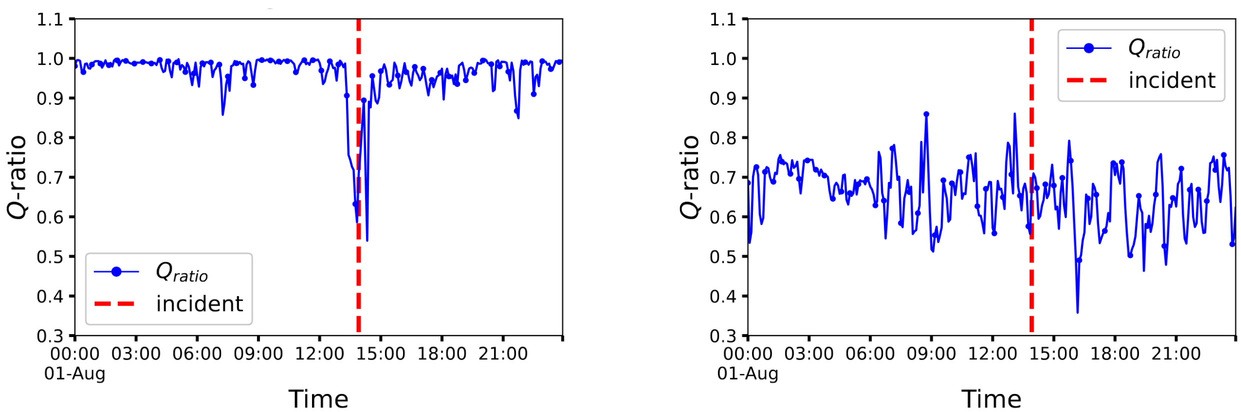
We also introduce a new **Localization Score** assesses the spatial precision of anomaly detection, focusing on mile-marker accuracy within a 0.3-mile threshold. This metric is crucial for practical incident response as it ensures accurate localization of events. Lastly, the **Area Under the Curve (AUC)** measures the trade-off between true positive and false positive rates across detection thresholds, providing an overall performance benchmark

The proposed Traffic Response Anomaly Capture Engine (TRACE) was benchmarked against various traditional and state-of-the-art traffic anomaly detection models, including STG-RGCN, GCN-LSTM, GCN, STG-GAT, and Transformers. Metrics used for evaluation included Area Under the Curve (AUC), mean detection delay (in seconds), mean localization error (in miles), and the percentage of incidents detected.

## *4.3 Clustering*

To illustrate the importance of the approximation algorithm for clustering to maximize positive correlation strategically, we compare the plots of time series of the ratio samples for the same time frame of the same day in Figure 11 for the same area.

Figure 11a is a cluster with high data correlation (0.87) and Figure 11b is a cluster with low data correlation (0.37). Observe that the time series of ratio samples in Figure 11a is highly stable under benign traffic conditions (stationarity and low variance) and shows a sharp deviation on the incident that happened at 13:00 hrs. In contrast, Figure 11b that did not maximize correlation has poor stability and does not show clear deviation in its time series when the incident happens.




**(a)** **(b)**

**Figure 11. Effect of clustering on feature correlation. (a) Cluster of high correlation, (b) cluster with low correlation.**

## 4.4 Baselines

With the data processed, we trained and evaluated a range of anomaly detection models, which served as baselines for comparison against TRACE. The baselines are described below:

- **STG-RGCN AE (Relational Graph Convolutional Network Autoencoder)**: This model uses the relational spatiotemporal graph of the freeway network as input and Relational Graph Convolution blocks [32].
- **STG-GAT AE (Graph Attention Network Autoencoder)**: This model uses the spatiotemporal graph without edge relations as input and Graph Attention blocks for spatiotemporal learning [33].
- **GCN-LSTM AE (Graph Convolutional Network Autoencoder with Temporal Aggregation)**: This model accepts a time series of spatial graphs as input. It uses Graph Convolution blocks for spatial processing and LSTMs for temporal processing in the latent space [34].
- **GCN AE (Graph Convolutional Network Autoencoder)**: A special case of the STG-RGCN AE without edge relations and using only the current graph as input.
- **Transformer AE (Transformer Autoencoder)**: Processes the time series of node features, treating each node as independent. It uses temporal but not spatial features [35].
- **MLP AE (Multi-layer Perceptron Autoencoder)**: A standard autoencoder that treats each node as independent and does not consider temporal or spatial features. It reconstructs each node using its own features.

## 4.5 Results

The following table provides a summary of TRACE's performance compared to other models:

**Table III. Summary of TRACE's performance compared to other models**

| Methods | AUC | Mean Delay (seconds) | Mean distance (miles) | Incidents Detected (out of 10) |
|---|---|---|---|---|
| **STG-RGCN** | 0.6843 | -9.35 +/- 8.6 | <u>*3.08 +/- 3.21*</u> | **10** |
| **STG-GAT** | 0.66 | **-10.5 +/- 6.8190**[4] | 3.36 +/- 3.064 | 7 |
| **GCN** | <u>*0.704*</u> | <u>*-9.75 +/- 5.68*</u> | 3.446+/- 3.4752 | 8 |
| **GCN-LSTM** | 0.69 | -8.0 +/- 6.1418 | 3.4461 +/- 3.3425 | <u>*9*</u>[5] |
| **MLP** | 0.6204 | -3.3 +/- 8.795 | 3.9666 +/- 2.876 | 5 |
| **Transformer** | 0.4987 | -4.375 +/- 7.2402 | 3.3249 +/- 2.10 | 4 |
| **TRACE** | **0.7124** | -9.14285+/- 8.955 | **2.5549 +/- 2.100** | 7 |

TRACE demonstrated superior anomaly detection capabilities, achieving an AUC score of **0.7124**, the highest among the evaluated models. This result highlights the robustness of TRACE in distinguishing normal traffic patterns from anomalies with high precision. For comparison, **STG-RGCN** achieved an AUC score of **0.6843**, while **GCN-LSTM** and **GCN** lagged behind with scores of **0.69** and **0.704**, respectively. The incorporation of **Normalizing Flows**, which estimate the likelihood of traffic patterns, significantly enhanced TRACE's ability to identify deviations effectively, setting it apart from other models in anomaly detection accuracy.

In terms of mean detection delay, TRACE exhibited competitive performance, averaging **-9.14 seconds**. While TRACE was slightly outperformed by **GAT**, which achieved delay of **-10.5 seconds**, **GCN**, with a delay of **-9.75 seconds**, and **STG-RGCN**, with a delay of **-9.35 seconds**, it still demonstrated a rapid response in identifying anomalies. Transformer-based models performed notably slower, reporting a delay of **-4.375 seconds**, whereas **GCN-LSTM** achieved a mean delay of **-8.0 seconds**. These results indicate that TRACE's **hybrid architecture** effectively balances detection accuracy and computational efficiency, enabling prompt responses to traffic anomalies while maintaining a competitive edge.

TRACE also excelled in localization precision, achieving a mean localization error of **2.55 miles**, the most precise result among the models evaluated. In comparison, **STG-RGCN** reported a higher localization error of **3.08 miles**, while **STG-GAT** and **GCN-LSTM** had errors of **3.36 miles** and **3.44 miles**, respectively.

[4] Bold indicates Best Performing Model for the given Metric

[5] Underline indicates Second Best Performing Model

TRACE's superior localization performance can be attributed to its ability to integrate spatial relationships through **Graph Neural Networks (GNNs)**, which played a critical role in accurately pinpointing anomalies in traffic data. This precise spatial modeling makes TRACE particularly effective in practical applications where accuracy in locating incidents is crucial.

Finally, TRACE successfully detected anomalies in **7 out of 10 cases**, demonstrating reliable performance when compared to other high-performing models. For instance, **STG-RGCN** identified anomalies in all **10 cases**, while **GCN-LSTM** detected anomalies in **9 out of 10 cases**. While TRACE showcased strong reliability, this result highlights areas for potential improvement, particularly in handling complex or edge-case scenarios where its performance slightly underperformed relative to its counterparts. Overall, TRACE's strong detection accuracy, competitive response times, and precise anomaly localization establish it as a robust and efficient solution for traffic anomaly detection.

## *4.6 Synthetic Data Generation*

Finally, we evaluate the efficacy of learning a classifier in a supervised setting using samples drawn from our normalizing flow model. The classifiers are evaluated on five synthetic testing datasets based on various segment clusters, with abnormal time slides of 5% and anomalous road segments at 50%. We report the average AUC score in Figure 4. One can see that classifiers have acceptable discrimination capability (AUC score ≥ 0.7) in 6 of the 8 clusters. The relatively lower AUC scores in clusters A (off-ramp/Exit segments) and C (on-ramp segments) are probably due to the extremely low volume of abnormal congestion data in such clusters. This shows the usefulness of normalizing flow models and their ability to generate synthetic samples from an initial unlabeled training set. The effectiveness of training models on an otherwise sparsely labeled anomaly detection dataset can be validated by the high AUC scores, something that would otherwise be difficult for the model to learn on unlabeled, unsupervised training pipelines.

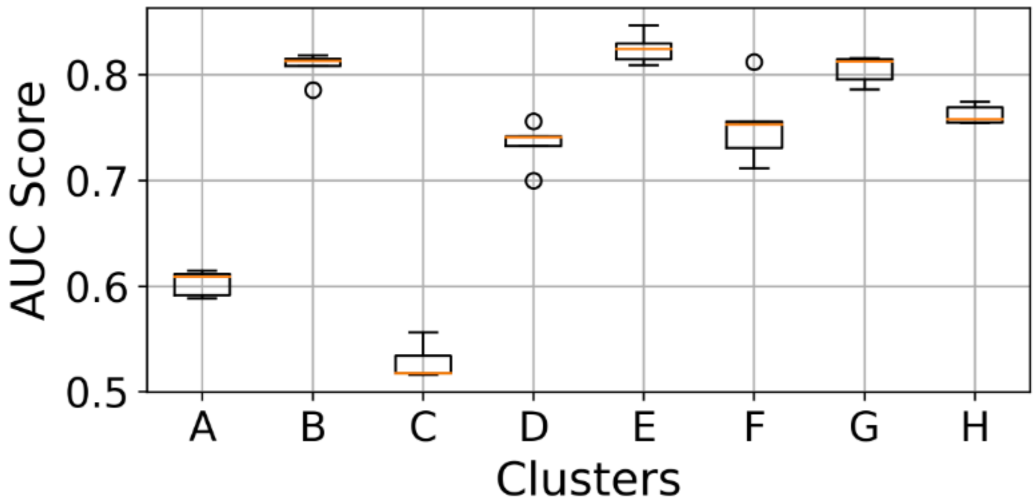



**Figure 12. Average AUC score for the MLPClassifiers on 5 synthetic testing datasets. Boxplots show the performance variation of MLPClassifiers trained on 5 datasets drawn from the normalizing flow model.**

## 4.7 Discussion

The **TRACE** model demonstrates several strengths that position it as a significant advancement in traffic anomaly detection. Its **hybrid modeling architecture**, which integrates Graph Neural Networks (GNNs), Transformers, and Normalizing Flows, enables it to capture complex spatial-temporal relationships and effectively handle data sparsity. This architectural design, combined with its support for **real-time processing**, allows TRACE to manage large-scale traffic data streams, ensuring timely anomaly detection. Additionally, the model exhibits **robustness across conditions**, with its probabilistic framework enabling consistent performance even under noisy or sparse data scenarios. One of TRACE's most notable strengths is its **localization precision**, as it outperformed other models in pinpointing the exact location of anomalies—an essential capability for effective incident management.

However, TRACE is not without limitations. In **complex scenarios**, the model exhibited **lower detection rates** compared to baselines like STG-RGCN, highlighting its challenges in capturing edge cases or highly dynamic traffic conditions. Another observed limitation was a **slight delay in detection**, where TRACE's mean detection delay, though competitive, was marginally higher than that of models such as GCN. Optimizing computational efficiency could help address this shortcoming. Furthermore, TRACE faced **generalization challenges** in regions with highly variable traffic patterns, where its detection rates were occasionally inconsistent. Incorporating additional contextual data, such as weather conditions, roadwork schedules, or events, could enhance the model's robustness and performance in such scenarios.

Despite these limitations, the results indicate that TRACE effectively combines spatial, temporal, and probabilistic modeling techniques to overcome key challenges faced by traditional and state-of-the-art models. Its high detection accuracy and precise localization capabilities make it a strong candidate for real-world deployment in traffic management systems. Moving forward, addressing TRACE's limitations could unlock its full potential. Future iterations of the model could explore **dynamic thresholding mechanisms** that adapt detection thresholds in real-time to account for varying traffic patterns. Additionally, **multimodal data integration**, such as incorporating weather conditions and event information, could provide valuable context-awareness, further improving performance. Finally, **optimization of computational resources** would help reduce detection delays, enhancing TRACE's efficiency for real-time applications.

By addressing these areas for improvement, TRACE has the potential to solidify its role as a transformative tool for traffic anomaly detection. For transportation agencies, TRACE offers a promising solution to enhance road safety, reduce congestion, and improve incident response times, ultimately contributing to more efficient and reliable traffic management systems.

# Chapter 5 Conclusion

Traffic anomaly detection is essential for maintaining road safety and operational efficiency, yet transportation agencies face persistent challenges in this domain. Swiftly identifying and responding to roadway incidents is critical for minimizing congestion, reducing the risk of secondary accidents, and enabling timely emergency responses. Monitoring is critical in practice; delays accrued during the monitoring phase delay response and resolution. Frequently, secondary crashes and long-clearance times lead to additional congestion on critical arterial road segments. To improve the real-time monitoring of extensive road networks, transportation agencies are increasing the available sensing modalities, often in smart corridors. However, this drastic increase in the number of sensors raises an essential question from an operational perspective—how can transportation agencies monitor thousands of sensors in (near) real-time to detect incidents of interest? In collaboration with TDOT, a framework for aggregating, processing and leveraging available sensor data for detecting sparse spatial-temporal traffic incidents is presented. We show that our proposed unsupervised anomaly detection framework allows strategic partitions to independently generate, sanitize, learn and detect anomalies with high accuracy and low false-positive rates. Further, we extend our incident detection framework to enable individual road segment detection under incident by utilizing the inherent characteristics present in the transit network. Through extensive simulations, we show how our pipeline can detect and localize anomalies for any given city-scale.

First, all the available information was collected, aggregated and preprocessed. This cleans the data prior to merging and ensures that the quality is maintained through the multiple merging steps across space and time. However, creating anomaly detection models at a large scale while ensuring the quality of prediction is computationally intractable. Thus, we introduce clustering mechanisms that would partition the target area into more manageable sections while ensuring the quality of the detection. Finally, we leverage state-of-the-art approaches in graph convolutional networks, transformer models, and normalizing flows to not only detect anomalies at the temporal level but localize it across the spatial dimension.

# References

[1] August 28, 2023 • Staff, and N. Reports •, “21,150 People Died in Preventable Crashes in 2023.” Accessed: Dec. 17, 2024. [Online]. Available: https://www.worktruckonline.com/10204861/around-21-150-people-died-in-preventable-accidents-in-2023

[2] “Fiscal Year 2025 Budget Highlights | US Department of Transportation.” Accessed: Dec. 17, 2024. [Online]. Available: https://www.transportation.gov/mission/budget/fiscal-year-2025-budget-highlights

[3] “INVESTING IN AMERICA: Biden-Harris Administration Announces More than $96 Million in Advanced Technology Grants to Improve Safety and Reduce Travel Times | FHWA.” Accessed: Dec. 17, 2024. [Online]. Available: https://highways.dot.gov/newsroom/biden-harris-administration-announces-96-million-in-advanced-technology-grants

[4] V. Chandola, “Anomaly Detection : A Survey,” p. 72.

[5] “Road Inventory Office.” Accessed: Dec. 17, 2024. [Online]. Available: https://www.tn.gov/tdot/long-range-planning-home/longrange-road-inventory.html

[6] J. Deacon, C. Zegeer, and R. Deen, “Identification of Hazardous Rural Highway Locations,” *Ky. Transp. Cent. Res. Rep.*, Nov. 1974, doi: http://dx.doi.org/10.13023/KTC.RR.1974.410.

[7] P. P. Tasgaonkar, R. D. Garg, and P. K. Garg, “Vehicle Detection and Traffic Estimation with Sensors Technologies for Intelligent Transportation Systems,” *Sens. Imaging*, vol. 21, no. 1, p. 29, Dec. 2020, doi: 10.1007/s11220-020-00295-2.

[8] D. A. Badoe, S. M. Click, J. Oswalt, M. Edwards, E. Humphreys, and Tennessee Technological University, “The Traffic Monitoring Program Research Project: Comparison of TDOT Practices With Selected States,” RES2016-01, Oct. 2017. Accessed: Dec. 17, 2024. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/55051

[9] S. Thakur, “A Review of Traffic Congestion Problem and Various Automated Traffic Measurement Sensors and Techniques,” *Indian J. Sci. Technol.*, vol. 9, no. 1, pp. 1–16, Jan. 2016, doi: 10.17485/ijst/2016/v9i47/106902.

[10] United States. Federal Highway Administration, “Traffic Monitoring Guide - Updated October 2016,” FHWA-PL-17-003, Oct. 2016. Accessed: Dec. 17, 2024. [Online]. Available: https://rosap.ntl.bts.gov/view/dot/50512

[11] “What is Transportation Systems Management and Operations (TSMO)? | Transportation Systems Management and Operations (TSMO) Plans | Organizing and Planning for Operations - FHWA Office of Operations.” Accessed: Dec. 17, 2024. [Online]. Available: https://ops.fhwa.dot.gov/tsmo/

[12] J. Zhang, Y. Zheng, D. Qi, R. Li, and X. Yi, “DNN-based prediction model for spatio-temporal data,” in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, in SIGSPACIAL ’16. New York, NY, USA: Association for Computing Machinery, Oct. 2016, pp. 1–4. doi: 10.1145/2996913.2997016.

[13] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A Comprehensive Survey on Graph Neural Networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4–24, Jan. 2021, doi: 10.1109/TNNLS.2020.2978386.

[14]B. Yu, H. Yin, and Z. Zhu, "Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, Jul. 2018, pp. 3634–3640. doi: 10.24963/ijcai.2018/505.

[15]W. Jiang and J. Luo, "Graph Neural Network for Traffic Forecasting: A Survey," *Expert Syst. Appl.*, vol. 207, p. 117921, Nov. 2022, doi: 10.1016/j.eswa.2022.117921.

[16]Z. Xing, M. Huang, W. Li, and D. Peng, "Spatial linear transformer and temporal convolution network for traffic flow prediction," *Sci. Rep.*, vol. 14, no. 1, p. 4040, Feb. 2024, doi: 10.1038/s41598-024-54114-9.

[17]G. Papamakarios, T. Pavlakou, and I. Murray, "Masked Autoregressive Flow for Density Estimation," Jun. 14, 2018, *arXiv*: arXiv:1705.07057. Accessed: May 15, 2022. [Online]. Available: http://arxiv.org/abs/1705.07057

[18]D. P. Kingma and P. Dhariwal, "Glow: Generative Flow with Invertible 1x1 Convolutions," Jul. 10, 2018, *arXiv*: arXiv:1807.03039. doi: 10.48550/arXiv.1807.03039.

[19]"Anomaly Detection for Time Series Using VAE-LSTM Hybrid Model | IEEE Conference Publication | IEEE Xplore." Accessed: Dec. 17, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/

[20]Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intell. Transp. Syst.*, vol. 11, no. 2, pp. 68–75, 2017, doi: 10.1049/iet-its.2016.0208.

[21]Y. Chen, Y. Lv, Z. Li, and F.-Y. Wang, "Long short-term memory model for traffic congestion prediction with online open data," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 132–137. doi: 10.1109/ITSC.2016.7795543.

[22]R. L. Abduljabbar, H. Dia, and P.-W. Tsai, "Development and evaluation of bidirectional LSTM freeway traffic forecasting models using simulation data," *Sci. Rep.*, vol. 11, no. 1, p. 23899, Dec. 2021, doi: 10.1038/s41598-021-03282-z.

[23]S. Basak, A. Ayman, A. Laszka, A. Dubey, and B. Leao, "Data-Driven Detection of Anomalies and Cascading Failures in Traffic Networks," *Annu. Conf. PHM Soc.*, vol. 11, no. 1, Sep. 2019, doi: 10.36001/phmconf.2019.v11i1.861.

[24]W. Wu *et al.*, "Developing an Unsupervised Real-Time Anomaly Detection Scheme for Time Series With Multi-Seasonality," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 9, pp. 4147–4160, Sep. 2022, doi: 10.1109/TKDE.2020.3035685.

[25]M.-F. Chiang, E.-P. Lim, W.-C. Lee, and A. T. Kwee, "BTCI: A new framework for identifying congestion cascades using bus trajectory data," in *2017 IEEE International Conference on Big Data (Big Data)*, Dec. 2017, pp. 1133–1142. doi: 10.1109/BigData.2017.8258039.

[26]M. L. D. Dias, C. L. C. Mattos, T. L. C. da Silva, J. A. F. de Macedo, and W. C. P. Silva, "Anomaly Detection in Trajectory Data with Normalizing Flows," Apr. 13, 2020, *arXiv*: arXiv:2004.05958. doi: 10.48550/arXiv.2004.05958.

[27]M. Wilbur, A. Dubey, B. Leão, and S. Bhattacharjee, "A Decentralized Approach for Real Time Anomaly Detection in Transportation Networks," in *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*, Jun. 2019, pp. 274–282. doi: 10.1109/SMARTCOMP.2019.00063.

[28]F. Harrou, A. Zeroual, and Y. Sun, "Trafc congestion monitoring using an improved kNN strategy," *Measurement*, vol. 156, p. 107534, May 2020, doi: 10.1016/j.measurement.2020.107534.

[29]P. Mercader and J. Haddad, “Automatic incident detection on freeways based on Bluetooth traffic monitoring,” *Accid. Anal. Prev.*, vol. 146, p. 105703, Oct. 2020, doi: 10.1016/j.aap.2020.105703.
[30]H. Zare Moayedi and M. A. Masnadi-Shirazi, “Arima model for network traffic prediction and anomaly detection,” in *2008 International Symposium on Information Technology*, Aug. 2008, pp. 1–6. doi: 10.1109/ITSIM.2008.4631947.
[31]F. Basso, R. Pezoa, M. Varas, and M. Villalobos, “A deep learning approach for real-time crash prediction using vehicle-by-vehicle data,” *Accid. Anal. Prev.*, vol. 162, p. 106409, Nov. 2021, doi: 10.1016/j.aap.2021.106409.
[32]M. Schlichtkrull, T. N. Kipf, P. Bloem, R. van den Berg, I. Titov, and M. Welling, “Modeling Relational Data with Graph Convolutional Networks,” Oct. 26, 2017, *arXiv*: arXiv:1703.06103. doi: 10.48550/arXiv.1703.06103.
[33]P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph Attention Networks,” Feb. 04, 2018, *arXiv*: arXiv:1710.10903. doi: 10.48550/arXiv.1710.10903.
[34]T. N. Kipf and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks,” Feb. 22, 2017, *arXiv*: arXiv:1609.02907. doi: 10.48550/arXiv.1609.02907.
[35]S. Tuli, G. Casale, and N. R. Jennings, “TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data,” May 14, 2022, *arXiv*: arXiv:2201.07284. doi: 10.48550/arXiv.2201.07284.