Abstract

Urban traffic congestion has always been a pressing issue in urban mobility, fuel, and environmental sustainability. This study proposes a data-centred approach for analysing traffic congestion, utilizing computer vision techniques and ML models. The system analyses traffic congestion in terms of vehicle count and speed metrics extracted from surveillance video feeds across times and days. The research focuses on the temporal analysis of traffic trends, the generation of predictive models for classifying congestion, and identifying critical bottlenecks in the urban road network. This was followed by a comparative analysis of a multitude of machine learning models for the accurate classification of congestion levels, especially Random Forest, Logistic Regression, Support Vector Classifier (SVC), and Gradient Boosting. Among these, Gradient Boosting Classifier surpassed the rest with 100% accuracy, ROC AUC score of 0.9999, and macro-averaged precision, recall, and F1 score all equal to 1.00; Random Forest did well, whereas Logistic Regression was comparatively moderate in classification with lower recall and F1 scores. The study emphasizes that congestion modelling with the integration of vehicle speed data enhances the classification accuracy and real-time traffic flow analysis. The identification of congestion hotspots further enables data-backed interventions for urban traffic management and optimization. This means that the results will trigger.

Chapter 1 Introduction

As certainly all of you will agree, the most important challenge cities of the world are facing today is vehicular congestion, which has increased at unprecedented levels due to urbanization and highly populous areas. Traffic congestion significantly affects travel time and the consumption of fuel apart from air quality, economic productivity, and general urban liveability. As cities grow, the conventional traffic management measures-Monitor manually, timing signals for fixed time, and infrastructure changes-reactively will soon be inadequate to deal with the dynamic and complex nature of urban traffic flow. Thus, the data driven technology application in traffic study and management has created some opulence to this phenomenon. Among those is the use of computer vision, which refers specifically to the use of real-time video surveillance and integrated optical traffic camera views to provide an automatic vehicle activity detection and processing capability aimed at detection, tracking and analysis of movement of vehicles at scale (Faherty, M. et.al, 2024). Then this paves the possibility for traffic metrics extraction like vehicle number counts, flow rates, and speed, among others, that can be modelled for high temporal and spatial resolution.

This research (Tanwar, R et.al 2025) utilizes computer vision and machine learning to study patterns of traffic congestion in cities and derive useful insights regarding them. This study will analyse the variations in counts of vehicles over times of the day, days of the week, and so on, and relate them to patterns of recurrent traffic congestion over time. Some additional depth will be gained from the fact that some vehicle speed data are included. Various machine learning methods were used to evaluate their usefulness in classifying congestion levels-Random Forest, Logistic Regression, Support Vector Classifier (SVC), and Gradient Boosting. The results show that models, such as Gradient Boosting and Random Forest, are superior to other models in precision, recall, and overall classification accuracy.

This study, therefore, presents a scalable framework for smart traffic analyses and urban mobility optimizations-a combination of real-time computer vision data and predictive modelling. The approach is designed to empower city planners, transportation authorities, and smart city initiatives to make informed and data-driven choices to alleviate congestion and improve urban life.

1.1 Background and Scope

The phenomenal growth of urban populations coupled with increasing vehicular ownership have contributed to increased traffic congestion in cities across the globe (Castiglione, M., et.al 2023). This perennial problem has caused an increase in travel times, fuel consumption, carbon emissions, and stress among commuters. Conventional traffic control systems dependent on fixed-time signal controls and manual observations are grossly inadequate to deal with the dynamic environment of modern traffic. However, in recent years, enormous advancements in machine learning and computer vision have opened up new avenues for real-time monitoring of traffic and predictive analysis of traffic congestion. Cameras operated for traffic surveillance might now be turned to serve intelligent transportation systems (ITS) by obtaining information through video feeds on vehicle counts, traffic density, and movement patterns. The traffic video analysis of computer vision algorithms provides the basis for a large-scale automation of traffic analysis.

This study (Sreelekha, M. et.al 2025) is conducted under the umbrella of data-driven urban mobility optimization with particular concern to traffic congestion pattern analysis with computer vision. Besides this, machine learning models are used to predict the level of congestion, and the study explores the impacts of vehicle speed data on congestion classification and flow efficiency. The project aims to put in place a solution by identifying peak congestion periods and critical road segments for recommendations on traffic signal operation.

Scope of the study:

1. Investigation into temporal and spatial traffic patterns based on traffic count and speed data recorded from the data.
2. developing and evaluating predictive models for the classification of congestion levels.
3. Comparative performance studies on different machine-learning algorithms.
4. Identification of congestion hotspots and provision of recommendations for data-driven traffic management.
5. This research narrows down to measurable traffic parameters and real-world applicability in the bigger scope of creating smart, sustainable, and responsive urban public transport systems.

1.2 Research Questions

RQ1. What are the effects of vehicle speed data on congestion classification models in terms of accuracy and reliability?

1.3 Research Objective

The primary task at hand is to design a data-driven analysis aimed at studying and optimizing parallel traffic congestion with statistical analysis and various machine-learning techniques. The specific objectives include:

2 To analyse congestion patterns based on evaluating traffic flow, that is, variations in counts, density, and speed of vehicles at different times within a day and between days of a week.

3 To develop predictive models for congestion levels based on statistical and machine-learning techniques including Random Forest, Logistic Regression, and Gradient Boosting.

4 To evaluate the influence of vehicle speed data on the classification of congestion and general assessment of traffic flow, thereby improving the outlook for congestion model prediction.

5 To analyse specific points of congestion in the urban road network and develop data-driven approaches towards better traffic management to optimize urban mobility.

1.4 Thesis Outline

The thesis comprises the following chapters:

Chapter 1 Introduction

This chapter is meant to present the background with respect to study, research objectives, research questions, and the limitations of the research. It also underlines the importance of traffic congestion analysis and urban mobility optimization-the foundation for understanding them.

Chapter 2 Literature Review

This chapter contains a review of the existing related records or literature on the subject-matter of discussion on traffic congestion analysis, machine learning techniques for traffic prediction, and the role of vehicle speed data in congestion classification. It highlights specific research gaps and tracks how the methods evolved as they were applied in urban traffic management.

Chapter 3 Methodology

This chapter is on the methodology of the research; data collection methods, preprocessing stages, and the machine learning models that were applied for congestion prediction have all been described herein. This chapter also presents the metrics through which the performance of the models was evaluated.

Chapter 4 Data Analysis and Model Development

The actual analysis of traffic data will result in the identification of congestion patterns, modelling, and development. The model evaluation will be reported in a chapter focusing on the findings, identifying the most accurate algorithms for congestion classification.

Chapter 5 Results and Discussion

This chapter presents the results of the analysis, along with the discussion of the model evaluation. Comparing the performance of various machine learning models with the interpretations of results will also be done in this chapter.

Chapter 6: Discussion

Examines the findings that relate to the research questions and objectives concurrently. Pertains to the strength, limitations, and practical applicability of the advocated optimization approach.

Chapter 7: Conclusion and Future Work

Summarizes the main results of the study, brings out the contributions of the research field on intelligent transportation systems, and discusses recommendations that may be useful for future enhancements and real-world implementation.

2. Literature Review

2.1 Introduction

On a microcosmic level, urban traffic congestion is a global concern that affects all cities, resulting in various negative effects such as increased travel time, fuel usage, and environmental pollution. Historically, solutions to traffic congestion have been static road signal systems and infrastructure planning. Traffic congestion now calls for more dynamic and intelligent solutions than old solutions because of rapid urbanization and changing mobility patterns. In recent times, advances in data analytics and computational methods have been instrumental in shifting attention from refractive measures to proactive, data-driven strategies for traffic monitoring and control by both researchers and policymakers (Caroleo, B.,et.al 2025). Machine learning algorithms have become quite popular in the last few years in addressing the issue of traffic congestion. The models are highly flexible and able to identify complex relations to identify high congestion given different traffic features, such as several vehicles, flow rate, or speed-an input. Such algorithms include Logistic Regression, which provides understandable linear classifications; for high accuracy by so-called ensemble methods such as Random Forest and Gradient Boosting; and in addition, Support Vector Classifier (SVC) to be deployed since it is so performing in spaces of high dimensions and is also suited to non-linear classification problems.

Many researchers have shown the influence of historical and real-time traffic data training models predicting traffic congestion. Some of the features include time of day, day of the week when the vehicles are dense, and speed and ability to handle non-linear classification problems. Training congestion predictive models since historical and real-time traffic data has been shown through various studies to be of various importance. These include features such as time of day, day of the week, vehicle density, and speed, which can broadly be analysed for capturing the temporal and spatial trend of congestion. The reliability of the models is evaluated using metrics of accuracy, ROC-AUC score, precision, recall, and F1 score. This chapter surveys the literature on conventional and modern traffic congestion analytical approaches. It underlines the apparent role of machine learning in urban traffic prediction, reviews the strengths and limitations of different algorithms, and points out gaps in contemporary methodologies. The chapter thus builds a theoretical basis for the current study, aimed at building a stable quantitative model for urban mobility optimization.

2.2 Identification of the problems

Noteworthy amounts of research are presently available about traffic congestion; however, a few challenges and gaps still exist in urban mobility optimization approaches-mediated by tradition or data. This segment succinctly describes the zero problems identified clearly in the existing literature, which warrant the need for this study.

1. Limited Adaptability of Classical Methods

Conventional management techniques such as fixed-time signal control and manual data collection fail to adapt well to dynamic conditions of traffic. If not all, these systems have been formulated to fail in capturing unexpected congestions, which event could have been triggered by an event, a weather change, or even a failure of infrastructure; consequently, an inefficient traffic flow and increased delay come to the fold.

2. Inadequate Real-Time and Historical Data Utilization

Many of these works depend on historical data since they lack real-time monitoring to understand the ongoing traffic behaviour. Thus, prediction models may fail to reflect the on-road reality as they do not hold good relevance in high-density urban environments.

3. Speed Underutilized as a Feature in Congestion Prediction

While vehicle count and density are usually applied, speed yet remains the most neglected variable-though critical in detecting the flow efficiency-absent or poorly related to the machine learning system. This missing out is enough to lessen the contextual and accurate relevance of congestion classification systems.

4. No Paired Comparisons of Machine Learning Algorithms

Quite a few existing works resulted into a single machine learning technique, using a test scenario against mathematical or uniformity.

2.3 Previous research

Recent research on urban transportation has examined the increasing utilization of real-time data and machine learning to predict the impact of traffic congestion, accidents, and weather on bus delays. (Kanagamalliga S et.al 2024) Existing approaches relying on static schedules and historical data lack the responsiveness needed to respond to dynamically changing conditions. The new advanced algorithms, such as Random Forest and Gradient Boosting, along with GPS, traffic sensors, and weather data, contribute greater accuracy. These models unravel complex patterns within large amounts of data and run timely delay forecasts. Research has also looked at how such data are visible in automated infrastructures to provide better urban mobility and more intelligent transit planning. Yet, there is still a challenge related to data integration and model scalability.

The recent investigations into Intelligent Transportation Systems (ITS) focus on integrating machine learning (ML) and artificial intelligence (AI) in alleviating urban mobility and traffic problems. High sophistication in ML: Random Forest, Gradient Boosting, and LSTM algorithms predict traffic flow with high accuracy by utilization of historical and real-time data. Models in question apply data from Internet of Things (IoT) sensors, cameras, and connected vehicles for dynamic traffic signal control and route optimization. (Sreelekha, M et.al 2024) The literature also points out the positive impact of AI on enhancing public transport efficiency through adaptive scheduling and multimodal connectivity. Nevertheless, some hurdles are still witnessed regarding data integration between systems, scaling of the model, and aligning with the infrastructure. Continuous research in this field is the need of the hour to exploit the maximum capabilities of AI-based ITS in carving out smarter and sustainable cities.

The integration of autonomous vehicles (AVs) into urban transportation systems has been widely studied for its potential to enhance traffic efficiency, safety, and sustainability (Louati, 2024). Reinforcement learning, particularly actor–critic models, has emerged as a key approach for optimizing AV decision-making in dynamic traffic environments (Louati, 2024). Multi-agent frameworks further improve inter-vehicle cooperation, especially in mixed traffic involving human-driven vehicles (Louati, 2024). Models like IDM and MOBIL are frequently used to simulate realistic traffic behaviour. Recent research emphasizes collaborative learning and reward optimization to achieve smoother traffic flow and reduced congestion (Louati, 2024). Sustainable cities are complex systems marked by dynamic changes and wicked problems that challenge traditional planning methods (Bibri, 2021). Recent studies highlight the role of smart city technologies and big data in addressing these challenges (Bibri, 2021). Urban computing and intelligence offer powerful tools for data-driven decision-making, enabling strategic and integrated urban planning (Bibri, 2021).

(Kumar, A et.al 2024) AI-based traffic flow management systems are being seen as novel solutions for urban transport puzzles in fast-evolving urban environments. Studies cite machine learning algorithms, such as Random Forest, Gradient Boosting, and LSTM, to serve real-time traffic predictions and dynamic routing and optimized traffic signal control. These systems, using data from sensors, cameras, and connected vehicles, enhance the efficiency of traffic, avoid congestion, and reduce environmental impacts. Much still needs to be done concerning interoperability, scalability, and ethical issues such as data privacy violations. Nevertheless, AI-supported systems are seen to have a great role in shaping the sustainable smart cities through helping with urban mobility and traffic management optimization. Urban transportation systems face frequent disruptions due to traffic congestion, accidents, and weather conditions, affecting the reliability of public bus services (Kanagamalliga et al., 2024).

Existing research has explored the use of real-time data and predictive analytics to enhance public transit efficiency (Kanagamalliga et al., 2024). Studies show that integrating data from GPS, traffic sensors, and weather sources can significantly improve delay forecasting (Kanagamalliga et al., 2024).

Urban mobility management is increasingly leveraging digital twin technologies to enhance traffic optimization and reduce environmental impact (Xu et al., 2023). Prior studies highlight the integration of IoT sensors and cyber-infrastructure for real-time traffic monitoring and control (Xu et al., 2023). Platforms like digital twins enable situational awareness, traffic prediction, and performance evaluation using big data analytics (Xu et al., 2023). Traditional fundamental diagram (FD) models have long been used to describe traffic flow, but their limitations in handling multidimensional data and supply-side factors have prompted ongoing research (Liu et al., 2023). Recent studies have explored more flexible, data-driven approaches such as Gaussian Process (GP) models for real-time traffic prediction (Liu et al., 2023). GP models offer the ability to capture complex, nonlinear relationships among traffic variables (Liu et al., 2023).

Urban traffic management systems are increasingly adopting computer vision and AI to enhance vehicle control and navigation (Dash et al., 2024). Studies have shown the effectiveness of models like YOLO for real-time vehicle detection and classification (Dash et al., 2024). Integration with OCR tools and national databases enables accurate identification and prioritization of emergency vehicles (Dash et al., 2024). Autonomous vehicles (AVs) are reshaping transportation, but validating their performance remains a major challenge (Chen et al., 2023). Data-driven microscopic traffic simulation has emerged as a key method for AV testing due to its scalability, realism, and access to high-fidelity traffic data (Chen et al., 2023). Studies highlight the importance of diverse datasets and robust evaluation metrics (Chen et al., 2023). Traffic congestion forecasting and management have gained significant attention due to their impact on urban mobility (Lee et al., 2023). Recent studies have utilized machine learning models like Long Short-Term Memory (LSTM) for accurate congestion predictions (Lee et al., 2023). Visual analytics systems are increasingly used to allow real-time exploration of congestion causes and propagation (Lee et al., 2023).

With rapid urbanization and the growth of smart cities, conventional traffic management systems struggle to cope with increasing congestion and density (Kumar et al., 2024). AI-driven systems have emerged as a solution to optimize traffic flow, reduce congestion, and improve overall efficiency (Kumar et al., 2024). Recent studies highlight AI traffic prediction algorithms, real-time data integration, and dynamic routing schemes as key components (Kumar et al., 2024). As urbanization and technology advance, AI-driven innovations are increasingly seen as key solutions to urban mobility challenges, particularly traffic congestion (Bahamazava, 2023). Research highlights the potential of autonomous vehicles and intelligent traffic management systems to optimize traffic flow and reduce congestion (Bahamazava, 2023). Studies employing mathematical models like Ordinary Differential Equations (ODEs) have been used to quantify the impact of AI adoption on traffic dynamics (Bahamazava, 2023).

Urbanization and motorization are soaring, and traditional transport systems barely keep pace with the demands of modern urban life. Artificial Intelligence (AI) and Machine Learning (ML) are seen as breakthroughs into which intelligent mobility systems can be ushered. Research suggests that AI/ML might drive urban mobility improvements-in sustainability, efficiency, and equity-and key studies on these technologies are conducted in a variety of urban contexts and demonstrate potential applications for optimizing traffic management, public transport, and multimodal connectivity. However, achieving equitable results will require governance-ethical-social justice integrated approaches, which this literature underlines as critical in demonstrating the important roles that the establishment of future transport systems-means-free, green, and sustainable access-will face. The European transportation sector is facing major challenges, including congestion, safety issues, and environmental impacts. Mitsakis, E. *et al.* (2025) Emerging disruptive technologies are reshaping traffic management and multimodal transport systems. SYNCHROMODE proposes an innovative data-driven decision support system that integrates multimodal network optimization, traffic prediction, and real-time event management. Key elements of this system include data quality assessment, cooperative dashboards, and the development of novel Key Performance Indicators (KPIs). Case studies in Thessaloniki, the Netherlands, and Madrid demonstrate the potential of SYNCHROMODE’s toolbox to enhance traffic management and multimodal transport efficiency, offering a new paradigm for Europe’s transport systems.

AI application in dynamic traffic flow management will very soon gain its utmost importance for urban intersections. Recent studies highlight using models like YOLOv8 for real-time vehicle recognition and categorization to allow priority passage for emergency vehicles. Integration with OCR technologies for number plate recognition and along with databases like NIC ensures proper identification of emergency vehicles. Algorithms such as A*, Dijkstra, and BFS are integral to optimizing traffic routing regarding congestion without external mapping services. (Dash, S et.al 2024) Such advances are developing a scalable, self-sustaining, security solution for urban traffic management with efficiency and emergency response performance improvements. Improving the quality or extent over space in freeway traffic state estimations is achieved with recent efforts of combining various sensor data streams. (Zhang, J et.al 2024) It has been proved that the coverage and granularity of traffic monitoring are bettered when ETC data were mixed with traffic detector data. The use of probabilistic models, which recognize traffic state dependences between adjacent road segments, has proved effective in real-world applications. Likelihood-based and maximin likelihood optimizations have become even more prevalent in bringing together the distributions of traffic flows. In addition, large-scale freeway networks are handled by decomposition and heuristic algorithms adopted by different techniques. Such case studies, like that of China's G92 freeway, bring proof to these models with extremely low error rates both in peak and off-peak hours. Urban mobility research today strongly emphasizes the need for data-driven approaches in identifying and improving key infrastructure elements about pedestrian and cyclist movement. (Sanchez-Sepulveda et.al 2024). Some studies have investigated the influence of open urban data and clustering algorithms on measures of walkability and cyclability, revealing important correlations between urban planning and active mobility trends. Strong analytics pipelines have been used to assess accessibility and sustainability opportunities in super-dense urban contexts in the interest of health and minimizing impacts. Framing such questions illustrates how geospatial analysis and multi-mode datasets can support the investments in policy and infrastructure. However, there is still a gap in the integration of affordability and equity in mobility solutions. Examples of such methodology applications in cities such as Barcelona illustrate the scalability of these approaches in supporting global targets of net-zero and liveability.

2.4 Limitations of Deep Learning and ML models

In recent times, machine learning and deep learning have been increasingly adopted in urban traffic forecasting and delay prediction. Despite this growing popularity, their limitations significantly hinder their effectiveness and scalability. In addition, an imposition of large, high-quality datasets at the centre of the assessment proves to be a challenge since they generally cannot be collected because of poor sensor coverage or restrictions on privacy. Also, many such models have shown resistance to real-time adaptability, especially with sudden traffic disruptions, accidents, or unforeseen weather. Also, black-box nature of deep learning models reduces the interpretability which might make it difficult for urban planners and authorities to comprehend the decision logic and, hence, trust the outcomes. (Makanadar, A., et.al 2024)

Furthermore, they usually require very high computational resources to be employed during training or deployment, which is not possible for low-resource urban areas. Problems such as overfitting on historical patterns and data imbalance, where rare events such as severe congestion are not represented, further limit the generalization of models. Finally, integrating heterogeneous pieces of data, such as GPS, traffic sensors, weather APIs, and incident reports, is still an elusive goal and often involves sophisticated preprocessing and fusion techniques. Such challenges indicate the need for hybrid models and more interpretable and resource-efficient AI systems in urban mobility management. AI and ML solutions for urban transportation research are considerably focused on congestion mitigation, accident reduction, and environmental conservation. Earlier studies had shown the use of predictive models such as ARIMA and ANNs for accident trend prediction and energy optimization. AI-assisted traffic systems are showing the way in adaptive signal control, hotspot detection, and eco-routing. Conventional traffic management systems are relatively unsatisfactory in scalability and adaptability, but emerging ML frameworks provide real-time, localized, and data-driven solutions. Integration of the models in existing infrastructures is still a challenge, especially for the fast-developing cities like AlKharj. Present studies suggest that AI will be a transformational force in evolving safer and greener mobility systems for the future. (Louati, A., 2025)

2.5 Table

Table.1 Comparison of Recent Studies in Urban Transportation and Traffic Congestion Management

Author(s)	**Focus Area**	**Methodology / Model**	**Technology / Data Used**	**Key Contribution**
Kanagamalliga et al. (2024)	Bus delay prediction	Machine Learning (RF, GB)	GPS, traffic sensors, weather data	Real-time bus delay forecasting with improved accuracy
Louati (2024)	Autonomous vehicle (AV) optimization	Reinforcement Learning (Actor–Critic),	AV sensors, multi-agent data	Enhanced AV cooperation, decision-making, and

		Multi-Agent Systems		congestion reduction
Bibri (2021)	Smart sustainable cities and planning	Urban computing, Big data analytics	Smart city data, IoT	Strategic urban planning and decision-making using smart technologies
Xu et al. (2023)	Traffic optimization using digital twins	IoT integration, Real-time traffic monitoring	IoT sensors, digital twin platforms	Real-time traffic prediction, situational awareness, and performance evaluation
Liu et al. (2023)	Traffic flow modelling	Gaussian Process (GP), FD models	Multidimensional traffic datasets	Real-time nonlinear traffic forecasting
Dash et al. (2024)	Vehicle detection and navigation using computer vision	YOLO, OCR, AI-based detection	CCTV, National Vehicle Database	Real-time emergency vehicle detection and smart traffic control
Chen et al. (2023)	AV validation and simulation	Microscopic traffic simulation	High-fidelity traffic datasets	Scalable AV testing in realistic environments
Lee et al. (2023)	Congestion forecasting and visualization	LSTM, Visual Analytics	Real-time traffic feeds, congestion logs	Accurate prediction and interactive visualization of

				congestion causes
Kumar et al. (2024)	AI in traffic management	AI algorithms, dynamic routing schemes	Real-time data, traffic sensors	Optimized traffic flow and congestion reduction
Bahamazava (2023)	AI impact on urban mobility	ODE modelling, traffic flow analysis	AI traffic systems, autonomous mobility data	

2.6 Research Gap

While traffic congestion prediction has made remarkable advances through machine learning, deep learning, and smart data integration over the years, several glaring gaps remain. Firstly, most models are historically biased and mostly work with historical, structured data with very few exceptions where multi-source real-time data streams are considered. The inclusion of such real-time information is deemed to make a substantive difference in enhancing adaptive responses to dynamic traffic conditions.

Second, interpretability and transparency of the modelling remain poorly studied. The other approach could be a black box with minimal insight into how predictions are arrived at, as far as usability in traffic decision-making is concerned. Moreover, many studies concentrate on single technologies or data sources instead of adopting an integrated approach for an automated infrastructure scalable in varied urban contexts. Also, research on adaptive models capable of self-improvement with new data or those reacting to abrupt disruptions such as natural disasters, protests, or system failures is highly limited. Furthermore, very few projects have matured into deployment and validation stages in real-life settings, and most are still at proof-of-concept level without enough empirical evaluation in live traffic scenarios. These gaps highlight a need for a much more robust, explainable, and adaptive predictive system that incorporates diverse sources of information and has proven to work well for urban traffic management. While most of them have proven AI and machine learning capabilities in prediction of traffic states and urban mobility optimization, there are several gaps left unexplored. Most existing systems depend heavily on an external mapping service, hence, lack full routing and decision autonomy. The merging of various sources of data (e.g., ETC, detector data) is underused or limited in extent across large networks. Advanced algorithms like YOLOv8 and optimization models have demonstrated quite a lot in many things; however, few studies are directed towards real-time implementation and deployment while the traffic environment undergoes continuous changes. Finally, the ethics, privacy, and interoperability aspects concerning large-scale sensor data collection and priority for emergency vehicles are very few yet to be explored. The addressing of the above gaps will indeed pave way for proximity development of a more robust, scalable, and intelligent urban traffic management system.

2.7 Conclusion

There is increasing evidence in the literature concerning the growing trend of employing data-driven and machine learning techniques in combating urban traffic congestion. Such advanced modelling techniques, as Random Forest, Gradient Boosting, and LSTM, have produced promising outcomes when tested for traffic predictions and public transport delays, especially when supported by real-time data gathered from other supplemental sources such as GPS, traffic sensors, and weather. Innovations like digital twins or reinforcement learning provide intelligent control of dynamic traffic and autonomous vehicle integration. Their few limitations, such as the requirement of very quality data, a lack of model interpretability, increased scalability challenges, and inadequate deployments in the real world, continue presenting ongoing challenges to their eventual adoption in practice. The gap in research lies in the development of systematic traffic prediction systems that are robust, yet transparent and adaptive in nature, capable of integrating heterogeneous data sources and yet operating reliably under varying urban situations. This will go a long way in ensuring that there are further developments aimed at making mobility smart and better planning for urban transport as regards smart cities.

Chapter 3 Methodology

It is intended by this research project to form and develop a predictive model of traffic congestion management for historical traffic data. The purpose of this project is to learn from previously collected traffic data using machine learning and statistic techniques to predict the current traffic condition and the possibility of managing traffic congestion. This study's methodology approach involves the preparing and processing steps in a structured aura collection, preprocessing, modelling, evaluation, and deployment in urban computing and machine learning principles. Data processing procedures and model applications are followed through different time frames to derive sound practical output applicable in traffic management.

3.1 Data Collection

The first step in this methodology involves aggregating traffic data from two datasets, one of one-month and the other of two months. Both datasets comprise counts of cars, bikes, buses, and trucks in addition to metadata like traffic conditions: heavy, normal, and low, day in the week, and time of day. These datasets are also from traffic monitoring systems, and they are important in understanding traffic characteristics. Vehicle counts are taken in the interval of 15 minutes, and traffic condition will be the target variable for prediction. The main challenge at this step is to assure the data is: broad, uniform and covers a vast range of possible traffic conditions, Data (Kaggle website https://www.kaggle.com/datasets/hasibullahaman/traffic-prediction-dataset/data).

3.2 Data Pre-Processing

Data preprocessing is a critical stage in any data science project, and this research is no exception. It involves preparing the data for machine learning modelling through cleaning, transforming, and normalizing it. Initially, datasets were merged along with an extra column titled 'Source', that shows whether that data is a one-month or two-month data. Outliers were detected and eliminated by Interquartile Range (IQR) methods to avoid any slope of model performance due to lone data points. Missing values and duplicates were also checked, to ensure the dataset is complete and accurate.

Following cleaning, the normality normalizing techniques were enforced in compression of the vehicle counts toward a normal distribution using Quantile Transformer, which rendered biasing of the data distribution as uniform. Thus, allowing the model upgrade for machine learning models that assume the normality of input features, like support vector machines (SVM). In addition, the One-Hot Encoding method transformed categorical columns like 'Traffic Situation', 'Day of Week', 'Source', which is enabled by machine learning algorithms for vigorous handling of categorical variables analysis. Additionally, there are temporal features like time and date, extracted from time for hours, which serve as an important feature for time-dependent models.

3.3 Exploratory Data Analysis (EDA)

Prior to the analysis of the dataset, exploratory data analysis (EDA) was performed on the data to reveal all the underlying patterns pertinent to the dataset. EDA is useful for mapping relationships in the dataset and gives insight into traffic conditions at different times of the day, different days of the week, and for different traffic situations. It was visualized in many ways-histograms, box plots, and pie charts-to show the distribution of vehicle counts and traffic situations. Box plots were also helpful in viewing changes in vehicle counts related to source, traffic situation, and day of the week, thereby providing richer traffic data. A correlation analysis heat map was created to examine relations between vehicle types, which can give significant knowledge pertinent to developing models.

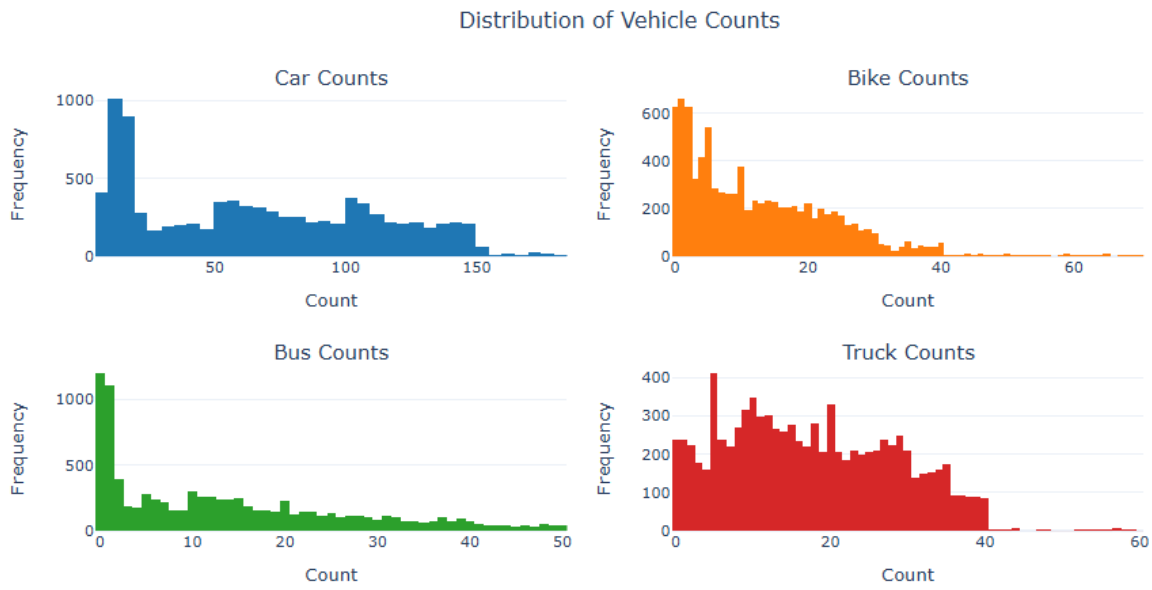

Fig.1 Distribution of Vehicle Counts.

This set of four histograms illustrates the distribution of vehicle counts by type: cars, bikes, buses, and trucks. Each subplot shows the frequency of vehicle counts recorded in certain count intervals.

The Car Counts histogram (top-left) shows that the distribution is right-skewed, with most car count values concentrated below 50. However, the secondary peaks around 60 and 110 depict moderately high volumes in those counts. The Bike Counts histogram (top-right) is also right-skewed with a sharp peak near zero, indicating that low bike counts are most common. The frequency gradually declines as the count increases, with very few exceptions above 40. The Bus Counts histogram (bottom-left) indicates that frequencies drop sharply after very low counts, peaking between 0-5, showing that the buses are the least frequent vehicle type, with increased counts being rare. Finally, the Truck Counts histogram (lower right) has a flatter peak compared to the other count types, showing a relatively uniform distribution with counts between 5 and 30, though the frequencies do decline for counts above 35. Overall, the visualization depicts that cars have any traffic to speak of, followed by bikes, with buses and trucks often found in small numbers. All the histograms being right skewed imply that high counts of vehicles are relatively rare among all types.

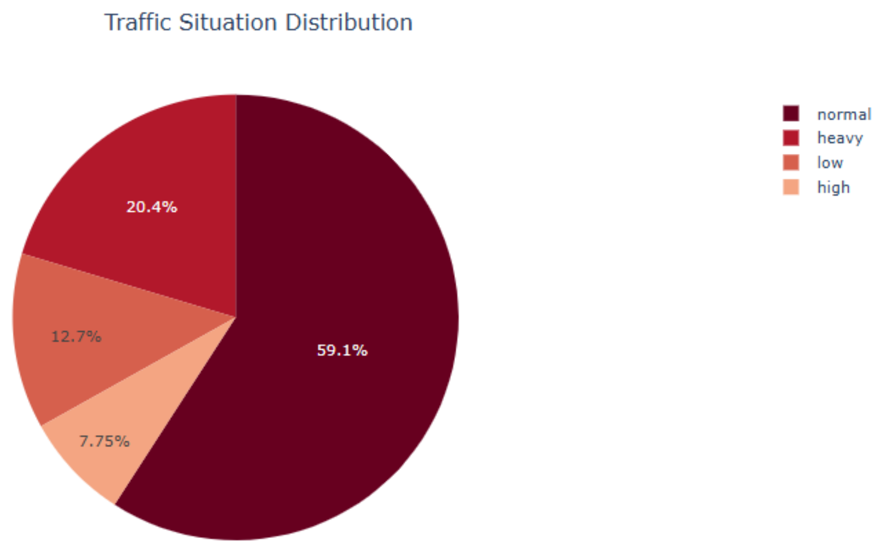

Fig.2 Traffic Situation Distribution

The pie chart represents the distribution of traffic situations across four categories: normal, heavy, low, and high. The 'normal' traffic condition dominates the dataset, making up 59.1% of all recorded instances, indicating that most traffic flows without major congestion. 'Heavy' traffic follows, accounting for 20.4%, reflecting significant congestion periods during peak hours or busy routes. The 'low' traffic condition, with 12.7%, suggests times of reduced vehicle movement, likely during early mornings or off-peak hours. 'High' traffic, which makes up only 7.75%, represents slightly elevated traffic levels but not as severe as 'heavy' congestion. Overall, the data suggests that traffic is generally manageable, with a smaller share of time experiencing disruption. This distribution can guide traffic control strategies by focusing interventions during the 28% of time when traffic is either heavy or high. It also highlights the importance of maintaining conditions that keep traffic within the 'normal' range.

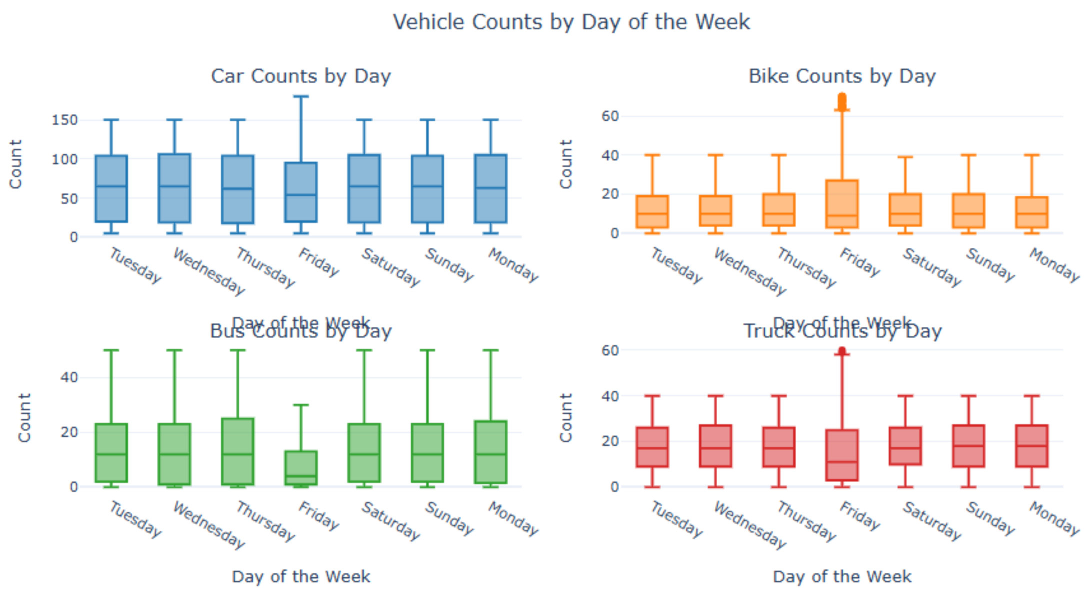

Fig.3 Vehicle Counts By Day of the week

This image has made visible four box plots that show counts of the number of vehicles by days of the week for cars, exercises, buses, and trucks. From the car plots, we can see that the distribution pictured is consistent across all days, with median values representing around 75 to 100. The distribution also spans a wide range, which was higher slightly during Friday. It can be said that generally, the counts for bikes are lower than those for cars, and Friday was observed to showcase a very high variability and higher maximum count; therefore, it can be inferred that sometimes bikers crowd this day. During the week, bus counts seem to be steady from day to day, except for Friday, which is characterized by the noticeable low median and reduced variability: indicating fewer buses or more steady counts that day. Truck counts are seen following a similar and steady pattern going through different days, but it is noted that even here Friday again has a special distinction since it carries a high outlier which possibly indicates that freight movement increases on this day. It may be noted from the analysis that

Friday is the most variable day across all vehicle types, specifically for bikes and trucks, possibly due to some changes in commuting or freight patterns as the workweek comes to an end.

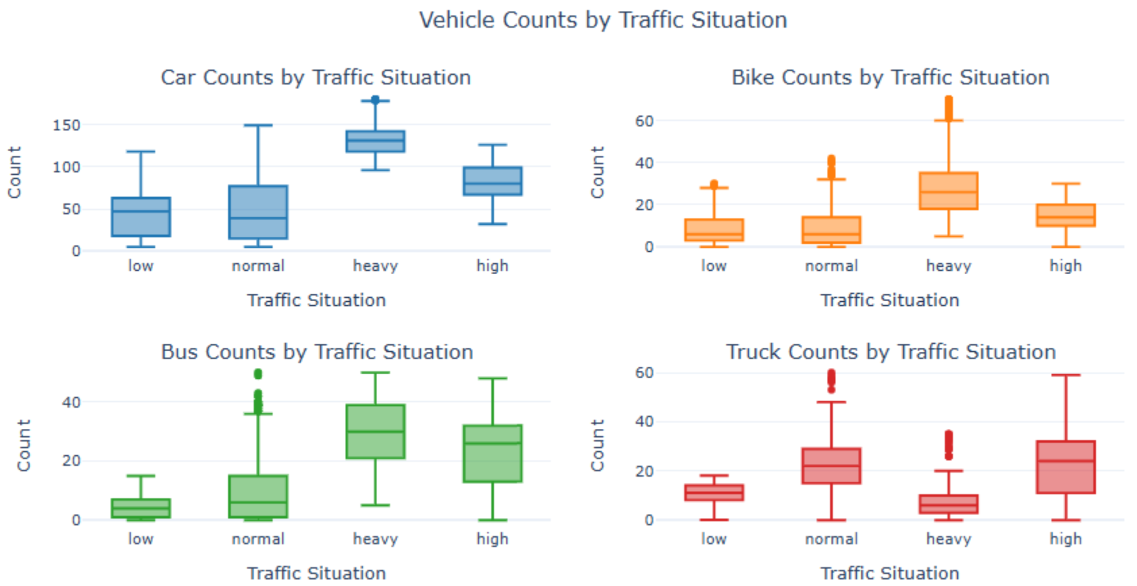

Fig.4 Vehicle Counts by Traffic Situation

Four box plots present vehicle counts categorized by traffic situations of low, normal, heavy, and high-and the counts by cars, bikes, buses, and trucks. For the counts of cars, it can be seen that counts increase considerably with an increase in heavy traffic conditions, recording the highest median and range values, proving that cars are a major source of congestion. For bikes, counts surely increase during heavy traffic conditions, but with a great deal of variability thereby indicating that bike use may possibly be an option when the roads are clogged with traffic. During normal and low traffic, bus counts are low but pull a substantial increase once the traffic gets heavy and high, and that goes on to show that buses are, at least somewhat, mass-transit during the period of crises. Unavoidably, truck counts show a parallel improvement, with accesses exhibiting higher counts and increased variation under normal and high traffic, though counter-intuitively lower in heavy traffic-perhaps because of time restrictions or rerouting during peak congestion. In summary, it shows that heavy traffic counts have been correlated with increased counts for all four classes of vehicles, particularly cars and buses, which is putting huge pressure on urban congestion.

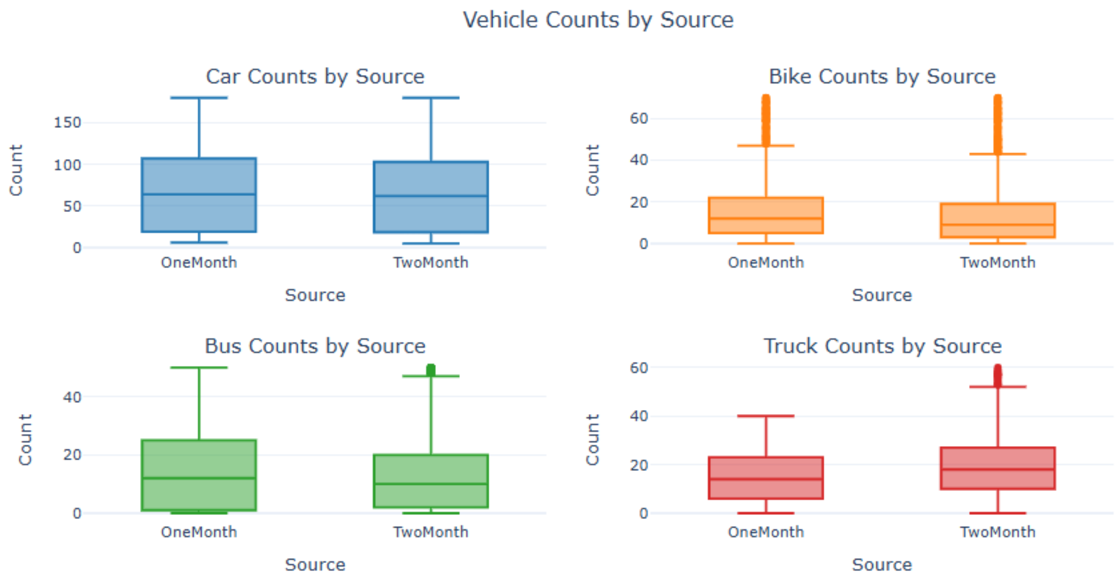

Fig.5 Vehicle Counts by source

The four box plots contrasting vehicle counts help compare the two data collection sources., One Month and Two months., for different types of vehicles: cars, bikes, buses, and trucks. For cars, the distribution and spread of counts registered with both sources are almost the same, with a wide range and a higher concentration around the median indicating a consistent volume of traffic over time. Bike counts appear almost similar from both sources; however, the Two Month has quite a few more outliers, which might mean occasional spikes in bike traffic. Bus counts imply a slightly lower median for the Two Month data, which might indicate slightly less or variability in bus movement. Truck counts give a higher mean and range for Two Month data, which means it could be increased freight activity. In short, the patterns in vehicle counts being compared remain essentially the same across both data collection sources. This shows the reliability and stability of the observed trends.

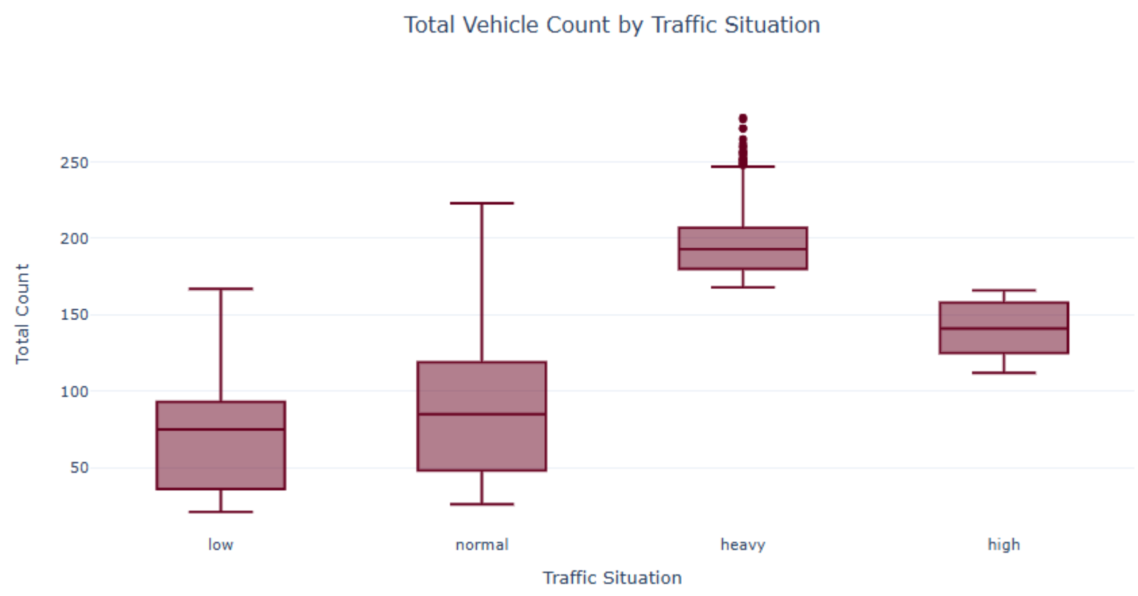

Fig.6 Total Vehicle Count by Traffic Situation

The box plot visualizes the total vehicle count across different traffic situations: low, normal, heavy, and high. The data clearly indicates that heavy traffic situations have the highest median and overall vehicle count, suggesting a strong correlation between congestion and high traffic volume. There is also a significant spread and presence of outliers in heavy traffic, implying occasional spikes. Normal traffic shows moderate vehicle counts with a wide distribution range. High traffic, despite its label, has fewer vehicles than heavy traffic but more consistency with less variability. Low traffic situations have the smallest vehicle counts and narrowest spread, reflecting quieter periods on the road. This plot highlights how total vehicle volume varies substantially depending on the traffic condition.

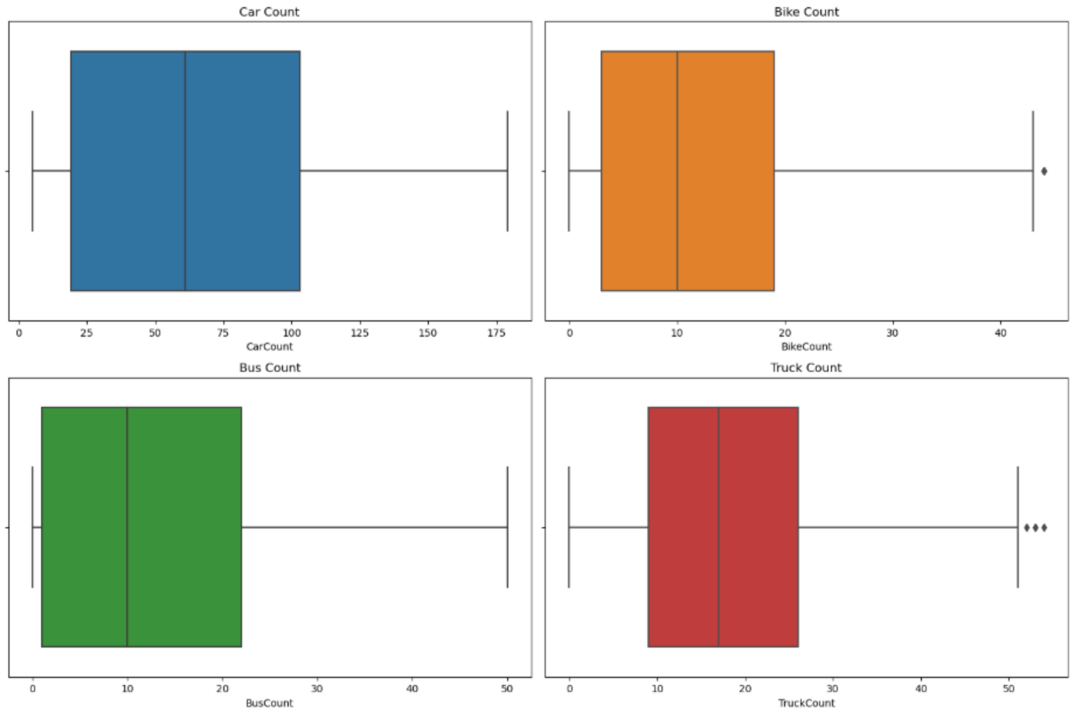

Fig.7 Outlier of the images

The box plots present a comparative overview of the vehicle-count distributions for cars, bikes, buses, and trucks. The car count has an extremely high range with a median of about 65, which shows high and variable traffic, while the distribution indicates that it is symmetric with no apparent outliers; hence the car flow is constant in the given dataset. Opposingly, the low median count of bikes, at around 12, indicates that bike traffic is generally low; however, some spikes occur occasionally. With very little variability, the bus count has shown very homogeneous behaviour, with the interquartile range being narrow, so that the median is around 10. Stability thus indicates uniformity of bus traffic. The truck count, is fairly spread out with a median around 20 but has several high-value outliers exceeding 50, indicating that truck traffic sometimes faces substantial upswings. To sum up, cars have the highest variability, buses are the most stable, with bikes and trucks having occasional high-count anomalies.

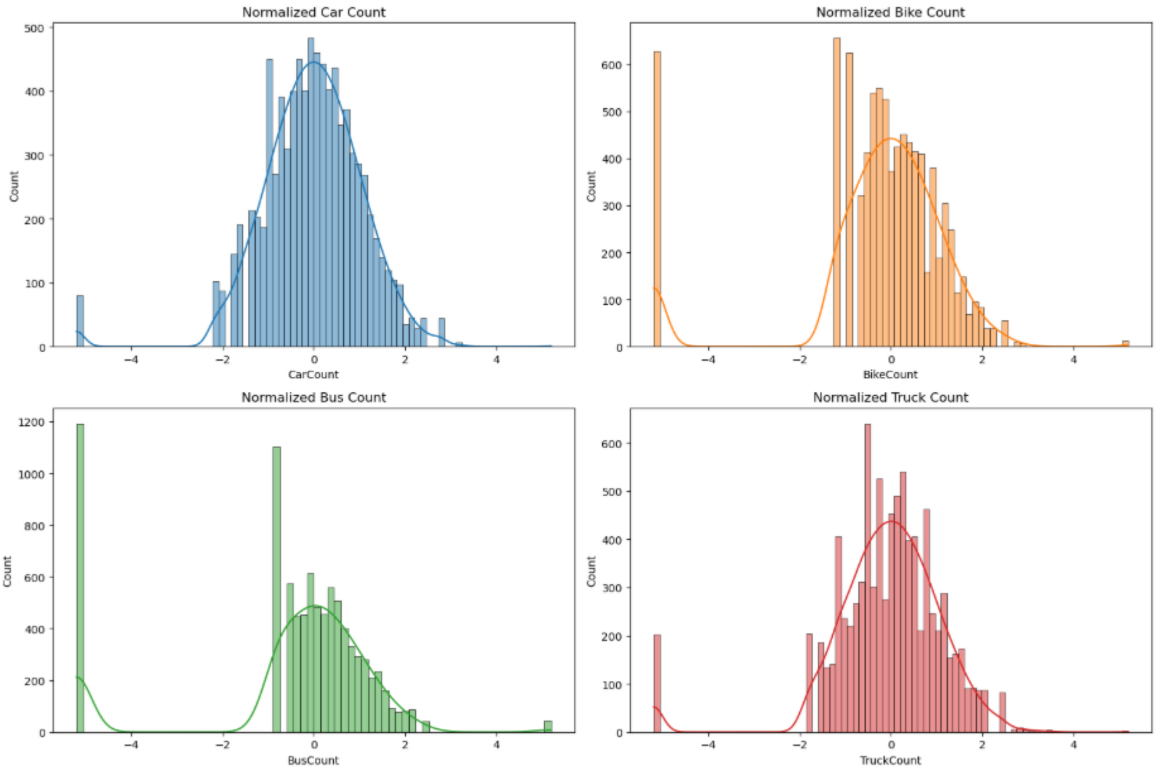

Fig.8 Normalization

The normalized distribution plots are the standardized frequency distributions for each vehicle type. The normalized car count is closely associated to the bell-shaped graph which indicates the near normal distribution of most of values clustering around the mean. This may also mean that car traffic is typical with hardly any extremes. The normalized count of bikes has slightly different story. It has a right-skewed distribution where most of its values are found at the left-side and white appearance taper off to the right. This corresponds to low counts of bicycles being much more common than the occasional high spikes. The normalized count of buses is such that it could be said to represent multimodal and irregular data having massive number of peaks and good number of outliers, mostly on its left side. Thus, they indicate erratic and inconsistent data on the number of buses taking traffic patterns. A similar phenomenon has been noticed in the normalized truck count that has a slight skew but could be considered roughly unimodal, centralized at the mean along with small deviations. Thus, though not perfectly normal, it indicates moderately consistent truck traffic behaviour that is also not outliers compared to bikes and buses. Car count, then, is the most normally distributed. On the other hand, bike counts and bus counts exhibit skewness and multimodal values, indicating variability in their traffic.

Feature Engineering

Feature engineering is very useful in making machine learning model predictions better. In this article, the authors created many new features from raw data, such as aggregating the total number of vehicles and generating time-based features like hour of the day along with support on either the fact that it is a weekend or weekday. Thusly, these features enhance the context of the models for precise predictions. As far as the dataset is considered, it has been split with only 80 percent in model training and 20 percent in testing. This way, the model will work on unknown data to give a more realistic performance estimate.

Model Selection

powerful Random Forest, logistic regression, support vector machine, and gradient boosting classifier-to see which among them can yield better results in traffic prediction. These algorithms are chosen because of their capacity to work with large datasets, handle non-linear relationships, and yield interpretable results. Random Forest and gradient boosting are ensemble methods known to excellently handle classification problems with complex data patterns, while SVM is known for its capacity to work in high dimensional space, which is also true for symptoms in this dataset; hence it suits the situation. Logistic regression serves as a benchmark model against which complex algorithms will be examined less rigorously.

The modelling proper kicked off by configuring the pre-processing pipeline such that numerical features and categorical features are handled. The pre-processing involved standardization of numerical features, such as vehicle counts and total vehicle count, using a standard scaler to ensure that they are on the same scale, which is an important condition for algorithms such as SVM to work properly. Categorical features were processed through one-hot encoding so that they can be let into machine learning models. A pipeline was formed for each algorithm, where the preprocessing and model fitting is done in a linear way to facilitate experimentation, ensuring that results of every model were being compared on the same dataset with different setups.

Support Vector Classifier

```
SVC Model Performance:
Model accuracy: 0.92

Confusion Matrix:
[[ 327    3    0    3]
 [   4   86    0   32]
 [   0    0  190   48]
 [  17   13   13 1008]]

Classification Report:
              precision    recall  f1-score   support

       heavy       0.94      0.98      0.96       333
        high       0.84      0.70      0.77       122
         low       0.94      0.80      0.86       238
      normal       0.92      0.96      0.94      1051

    accuracy                           0.92      1744
   macro avg       0.91      0.86      0.88      1744
weighted avg       0.92      0.92      0.92      1744

ROC AUC Score: N/A
```

Fig.9 SVC Result Analysis

The efficacy of the Support Vector Classifier (SVC) model is apprised through various evaluative measures, such as model accuracy, a confusion matrix, and a classification report; these measures render an introductory into the modelling ' s capabilities regarding prediction. The following sections elaborate further on the measures involved:

$$Accuracy = Total\ SamplesTrue\ Positives + True\ Negatives$$

Model Accuracy:

In fact, the model has an accuracy score of 0.92. This means that the classifier maladaptive correctly predicted the class labels for 92% of all samples present in the dataset. This is a simple metric calculated as follows: For the "Heavy" class, precision is rated at 0.94; therefore, 94% of the predicted heavy instances were indeed correct. In comparison, recall is rated at 0.98, meaning 98% of the true heavy instances were correctly recognized. The F1-score-the balancing of precision and recall-is rated at 0.96. For the "High" class, precision is rated at 0.84 and recall at 0.70, suggesting that although a fair number of high-class predictions are correct, the model seems to have more trouble capturing all high instances (low recall). This sets the F1-score at 0.77, which accounts for this imbalance. For the "Low" class, 0.94 precision suggests good precision, but there is some room for improvement in recalling low-class instances, while the 0.86 F1-score is showing good balance. For the "Normal" class, performance appears to be okay for both precision and recall, at 0.92 and 0.96, respectively; thus, it has produced a strong F1-score of 0.94. The overall performance across the classes is summarized by means of the macro average and weighted average. The macro average computes the unweighted mean of precision, recall, and F1 score across the classes, treating each class equally regardless of the number of instances in that class. The weighted average, on the other hand, gives influence according to the number of instances in that class. Therefore, this average becomes more even regarding class distribution.

Gradient Boosting

Gradient boosting refers to the general method of boosting, where models are created in a sequence. Each new model tries to correct the errors from the previous ones. With this, the model gradually learns to minimize the error better. It's called "gradient" boosting because gradient descent is employed for minimizing the loss function (log loss for classification, mean square error for regression) The Gradient Boosting Classifier demonstrated excellent, near-perfect learning ability in predicting traffic condition classes, with only 2 misclassifications in 1744 samples. Perfect or near-perfect performances in all key metrics: accuracy, precision, recall, F1-score and ROC AUC. Therefore, the model has been successfully applied, with good generalization for this classification problem.

```
GradientBoostingClassifier Model Performance:
Model accuracy: 1.00

Confusion Matrix:
[[ 333    0    0    0]
 [   0  122    0    0]
 [   0    0  236    2]
 [   0    0    0 1051]]

Classification Report:
              precision    recall  f1-score   support

       heavy       1.00      1.00      1.00       333
        high       1.00      1.00      1.00       122
         low       1.00      0.99      1.00       238
      normal       1.00      1.00      1.00      1051

    accuracy                           1.00      1744
   macro avg       1.00      1.00      1.00      1744
weighted avg       1.00      1.00      1.00      1744

ROC AUC Score: 0.9999968997957427
```

Fig.10 Gradient Boost Result Analysis

Formula

For a loss function $F(x) = m = 1\sum M\gamma mhm(x)$

At each stage mmm, the model updates based on the gradient of the loss function:

$$rim = -[\partial F(xi)\partial L(yi, F(xi))]F(x) = Fm - 1(x)$$

Gradient boosting is when a series of models are created with each model correcting the previous model's errors based on a gradient descent procedure on the loss function. Gradient boosting is a powerful tool in supervised learning and frequently put to use in various fields, including fraud detection, churn prediction, and traffic classification facilities.

Logistic Regression

The Logistic Regression model showed moderate classification performance with an accuracy of 0.86, meaning that 86% of the predictions matched the true class labels. Logistic Regression is a kind of linear model employed for binary or multi-class classification whereby the probability of class membership is estimated using the logistic function sigmoid.

$$Accuracy = Number\ of\ Correct\ Predictions/Total\ Number\ of\ Samples$$

```
LogisticRegression Model Performance:
Model accuracy: 0.86

Confusion Matrix:
[[323   5   0   5]
 [  6  64   1  51]
 [  0   0 142  96]
 [ 10  26  49 966]]

Classification Report:
              precision    recall  f1-score   support

       heavy       0.95      0.97      0.96       333
        high       0.67      0.52      0.59       122
         low       0.74      0.60      0.66       238
      normal       0.86      0.92      0.89      1051

    accuracy                           0.86      1744
   macro avg       0.81      0.75      0.78      1744
weighted avg       0.85      0.86      0.85      1744

ROC AUC Score: 0.9586639315265979
```

Fig.11 Logistic Regression result Analysis

The Logistic Regression model is showing a reasonable baseline performance, especially for the "heavy" and "normal" classes. The "high" and "low" classes, however, are giving it some trouble because of misclassifications owing to its linear nature. Nevertheless, the high ROC AUC score implies some degree of good ranking performance. One could improve upon this by looking at feature engineering or moving on to non-linear models like Random Forest or Gradient Boosting, which deal better with class overlaps and non-linear decision boundaries.

Random Forest

The Random Forest Classifier is at its best, giving the classification task an overall accuracy of 0.99, which means it correctly predicted 99 percent of the 1744 traffic condition samples. Random Forest is a powerful method that operates by joining a set of decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees for the result.

```
RandomForestClassifier Model Performance:
Model accuracy: 0.99

Confusion Matrix:
[[ 333    0    0    0]
 [    0  122    0    0]
 [    0    0  230    8]
 [    2    3    0 1046]]

Classification Report:
              precision    recall  f1-score   support

       heavy       0.99      1.00      1.00       333
        high       0.98      1.00      0.99       122
         low       1.00      0.97      0.98       238
      normal       0.99      1.00      0.99      1051

    accuracy                           0.99      1744
   macro avg       0.99      0.99      0.99      1744
weighted avg       0.99      0.99      0.99      1744

ROC AUC Score: 0.9993652758780449
```

Fig.13 Random Forest Result

To start things off, we will discuss Random Forest Classifiers, which are defined as avowedly accurate, robust, and balanced Classify Constructors. Random Forest Classifiers can accept both categorical and numerical features in data. This algorithm provides built-in feature importance analysis to rank variables by their contribution to prediction. Random Forests are robust and not sensitive to noise and overfitting due to the averaging mechanism. Compared with Logistic Regression and SVC, Random Forest deals much better with the nonlinear patterns and outperforms these two, notably on complex imbalanced and noisy datasets, as observed in traffic condition prediction.

Model Evaluation

After training the models on the training data available up until October 2023, the performance had been evaluated using various metrics, including generating confusion matrices and classification reports on accuracy to show how good the model was overall and within classes of traffic situations. The ROC AUC score was calculated for estimating the discrimination of models between different traffic situations. Models were assessed basing on their scores in the test set to get the best model for further proceedings. Also, this score has been important in evaluating the multi-class classification performance, since traffic prediction involves cross-conditions of predicting different traffic situations because it is obtained from real-world scenarios.

	Model	Accuracy	ROC AUC Score	Precision (macro avg)
0	RandomForestClassifier	0.99	0.9994	0.99
1	LogisticRegression	0.86	0.9587	0.81
2	SVC	0.92	NaN	0.91
3	GradientBoostingClassifier	1.00	0.9999	1.00

	Recall (macro avg)	F1-Score (macro avg)	Precision (weighted avg)
0	0.99	0.99	0.99
1	0.75	0.78	0.85
2	0.86	0.88	0.92
3	1.00	1.00	1.00

	Recall (weighted avg)	F1-Score (weighted avg)
0	0.99	0.99
1	0.86	0.85
2	0.92	0.92
3	1.00	1.00

Fig.14 Model Evaluation and comparison Analysis

Model Deployment and Real-Time Predictions

In addition to the traditional performance measures, the effect of overfitting was measured through attempts to use both the training and test sets. A model performing well against training data but poorly on test data would be overfitting and will not, therefore, generalize well to other new data. This was crucial especially with very large datasets that could be contaminated by noise or anomalies. Cross-validation techniques could also be employed to solidify robustness assessment of the models built. Now the model is deployed for real-time prediction. Traffic prediction models become useful opportunities when integrated into existing traffic management systems. For instance, traffic light timings can be adjusted based on the predictions of the model, commuters can receive real-time traffic alerts, and traffic flow can be optimized during peak times. The model will be tested in this study in a simulated environment that predicts traffic conditions using real-time data inputs. This will help assess whether the model can predict with both timeliness and accuracy within real-world constraints.

Furthermore, there would be continuous monitoring of the performance of the model in real-time scenarios against actual traffic data to ensure its accuracy. In instances where the prediction made by the model would prove inaccurate, it would be retrained with updated data, and parameter fine-tuning would be done for better performance of the model. The iterative nature of this process gives confidence that the model will keep performing with efficacy as traffic patterns change with time.

Conclusion

The methodology presented in this research project is, in conclusion, a thorough and systematic analysis of traffic congestion forecasting. The complete procedure starts with data collection, where the second step is extensive preprocessing and exploratory analysis to discover the characteristics of the data set. Model performance is enhanced through feature engineering, and multiple machine learning algorithms are applied and evaluated. A best-performing model is deployed for prediction in real-time, having been continuously monitored for accuracy. This methodology therefore provides a strong framework for predicting traffic conditions while being applicable to other urban computing and traffic management projects.

Future work and Discussion

For future work, involving other sources like real-time GPS data, average vehicle speed, and intersection-level traffic flow would considerably improve traffic condition classification on a more localised level. Adding multimodal traffic, such as bicycles, public transport, and pedestrian flow, to the dataset could add significant richness to the analysis. Besides, adding time-series forecasting models, like ARIMA or LSTM, would ensure that future traffic trends are predicted rather than simple classifications of existing states. Such continuously active sensors would work without central systems and would provide real-time responsiveness by deploying optimized models on edge devices or IoT-enabled traffic sensors. The synthetic data generation techniques such as SMOTE or GANs could also help improve prediction accuracy for classes that have limited representation by addressing class imbalance. In addition, explainable AI (XAI) techniques like SHAP or LIME should also be part of it, providing justification for the decisions of the model which will make it applicable to practical deployment in smart city infrastructure. These would further make it possible to develop highly adaptive, predictive, and intelligent traffic management systems in the future.

References

Sreelekha, M. and Midhunchakkaravarthy, 2024, January. Intelligent Transportation System for Sustainable and Efficient Urban Mobility: Machine Learning Approach for Traffic Flow Prediction. In *International Conference on Multi-Strategy Learning Environment* (pp. 399-412). Singapore: Springer Nature Singapore.

Kumar, A., Batra, N., Mudgal, A. and Yadav, A.L., 2024, March. Navigating Urban Mobility: A Review of AI-Driven Traffic Flow Management in Smart Cities. In *2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 1-5). IEEE.

Makanadar, A., Shahane, S. (2024). Smart Mobility and Cities 2.0: Advancing Urban Transportation Planning Through Artificial Intelligence and Machine Learning. In: Manoj, M., Roy, D. (eds) Urban Mobility Research in India. UMI2023 2023. Lecture Notes in Civil Engineering, vol 551. Springer, Singapore. https://doi.org/10.1007/978-981-97-8116-4_2

Mitsakis, E. *et al.* (2025). A Data-Driven Decision Support System for Multimodal Network and Traffic Management – SYNCHROMODE. In: Nathanail, E.G., Gavanas, N., Adamos, E. (eds) Climate Crisis and Resilient Transportation Systems. CSUM 2024. Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham. https://doi.org/10.1007/978-3-031-82818-8_8

Dash, S., Dewangan, S., Mishra, B.S.P., Biswas, R., Mishra, S. and Mishra, S.S., 2024, December. Optimizing Urban Mobility: A Comprehensive Approach to Dynamic Traffic Flow Management. In *2024 6th International Conference on Computational Intelligence and Networks (CINE)* (pp. 1-8). IEEE.

Zhang, J., Huang, D., Liu, Z., Zheng, Y., Han, Y., Liu, P. and Huang, W., 2024. A data-driven optimization-based approach for freeway traffic state estimation based on heterogeneous sensor data fusion. *Transportation Research Part E: Logistics and Transportation Review*, *189*, p.103656.

Bahamazava, K., 2023. AI-driven scenarios for urban mobility: Quantifying the role of ODE models and scenario planning in reducing traffic congestion. Journal Name.

Bibri, S.E., 2021. Data-driven smart sustainable cities of the future: urban computing and intelligence for strategic, short-term, and joined-up planning. Comput.Urban Sci., 1, pp. 8.

Available at: https://doi.org/10.1007/s43762-021-00008-9.

Chen, D., Zhu, M., Yang, H., Wang, X., and Wang, Y., 2023. Data-Driven Traffic Simulation: A Comprehensive Review. IEEE Transactions on Intelligent Vehicles.

Kumar, A., Batra, N., Mudgal, A., and Yadav, A.L., 2024. Optimizing Traffic Flow using AI and Traffic Management Systems. 11th International Conference on Reliability, Infocom Technologies and Optimization.

Lee, C., Kim, Y., Jin, S., Kim, D., Maciejewski, R., Ebert, D., and Ko, S., 2023. Real-Time Traffic Flow Optimization Using AI. IEEE Transactions on Visualization and Computer Graphics.

Liu, Z., Lyu, C., Wang, Z., Wang, S., Liu, P., and Meng, Q., 2023. A Gaussian-Process-Based Data-Driven Traffic Flow Model and Its Application in Road Capacity Analysis. IEEE Transactions on Intelligent Transportation Systems.

Louati, A., Louati, H., Kariri, E., Neifar, W., Hassan, M.K., Khairi, M.H.H., Farahat, M.A., and El-Hoseny, H.M., 2024. Sustainable Smart Cities through Multi-Agent Reinforcement Learning-Based Cooperative Autonomous Vehicles. Sustainability, 16(5), pp. 1779. Available at: https://doi.org/10.3390/su16051779.

Xu, H., Berre, A., Yoginath, S.B., Sorensen, H., Nugent, P.J., Severino, J., Tennille, S.A., Moore, A., Jones, W., Sanyal, J., and Berres, A., 2023. Smart Mobility in the Cloud: Enabling Real-Time Situational Awareness and Cyber-Physical Control Through a Digital Twin for Traffic. IEEE Transactions on Intelligent Transportation Systems.

Sanchez-Sepulveda, M.V., Navarro-Martin, J., Fonseca-Escudero, D., Amo-Filva, D. and Antunez-Anea, F., 2024. Exploiting urban data to address real-world challenges: Enhancing urban mobility for environmental and social well-being. *Cities*, *153*, p.105275.

Louati, A., 2025. Machine learning framework for sustainable traffic management and safety in AlKharj city. *Sustainable Futures*, *9*, p.100407.

Yogi, K.S., Sharma, A., Gowda, V.D., Saxena, R., Barua, T. and Mohiuddin, K., 2024, March. Innovative Urban Solutions with IoT-Driven Traffic and Pollution Control. In *2024 International Conference on Automation and Computation (AUTOCOM)* (pp. 136-141). IEEE.

Kan, Z., Liu, D., Yang, X. and Lee, J., 2024. Measuring exposure and contribution of different types of activity travels to traffic congestion using GPS trajectory data. *Journal of Transport Geography*, *117*, p.103896.

Fathima, M.D., Hariharan, R. and S, E.R., 2024, December. Smart Street: AI-Powered Traffic Flow Enhancement with Adaptive Signal Control. In *International Conference on Advanced Network Technologies and Intelligent Computing* (pp. 144-158). Cham: Springer Nature Switzerland.

Bin Asad, K.M. and Yuan, Y., 2023. The impact of scale on extracting urban mobility patterns using texture analysis. *Computational Urban Science*, *3*(1), p.33.

Wang, J., Zhang, Y., Xing, X., Zhan, Y., Chan, W.K.V. and Tiwari, S., 2024. A data-driven system for cooperative-bus route planning based on generative adversarial network and metric learning. *Annals of Operations Research*, *339*(1), pp.427-453.

Caroleo, B., Sadeghi, J., Botta, C., Nikneshan, S. and Arnone, M., 2024, December. A Data-Driven Integrated Framework for Virtual Testing of Autonomous Vehicles in Mixed Traffic Scenarios. In *International Conference on Intelligent Transport Systems* (pp. 293-310). Cham: Springer Nature Switzerland.

Castiglione, M., Nigro, M. and Sacco, N., 2023, June. Multi-source Data-driven Procedure for Traffic Analysis Zones Definition. In *2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* (pp. 1-6). IEEE.

Tanwar, R. and Agarwal, P.K., 2025. Multimodal integration in India: Opportunities, challenges, and strategies for sustainable urban mobility. *Multimodal Transportation*, p.100210.

Faherty, M., McArdle, G. and Tafidis, P., 2024. Envisioning sustainable urban mobility in Dublin: Stakeholder perspectives and system dynamics. *Transportation Research Interdisciplinary Perspectives*, *28*, p.101280.