Intelligent Transportation Systems: A Survey on Data Engineering

Safa Batita[a], Achraf Makni[b] and Ikram Amous[c]

National School of Electronics and Telecommunications (ENET'Com), University of Sfax, MIRACL Laboratory, Airport Road, km 4, B.P. 1088, 3018 Sfax, Tunisia

Keywords: Intelligent Transportation Systems, Databases, Artificial Intelligence, Advanced Driver Assistance Systems.

Abstract: This paper presents an examination of data engineering within Intelligent Transportation Systems (ITS), focusing on integrating advanced technologies such as Real-Time Databases (RT-DBs), Graph Databases (GDBs), and Artificial Intelligence (AI) to improve ITS capabilities. The decision to focus on database systems and AI in this paper is based on their crucial roles in shaping modern transportation systems and offers a comprehensive view of the technological framework influencing ITS. Through an extensive review of existing literature, the paper explores how these solutions synergistically contribute to data collection, organization, processing, and extraction of value from various ITS data. The paper analyzes the transformative impact of real-time data management in connected vehicle systems and the efficacy of GDBs in capturing complex relationships within intelligent transportation networks. Additionally, it assesses the adaptability of AI in various ITS applications, including traffic prediction, driver assistance, and accident analysis. Despite their benefits, the paper discusses persistent challenges related to system complexity, interoperability, data management, and model accuracy, which impact the widespread deployment of ITS. Furthermore, the paper presents recommendations for addressing these challenges and emphasizes research directions that require further exploration, underscoring the importance of intelligent and efficient transportation worldwide.

1 INTRODUCTION

In the structure of modern civilization, transportation systems are deeply integrated into daily human activities. With an estimated 40% of the world's population spending at least one hour on the road every day, it's clear that transportation is central to our collective existence. This widespread dependence has increased significantly in recent years, a trend that reflects the rapid urbanization and globalization seen around the world. As a result, transportation infrastructure is at an intersection, facing multiple opportunities for innovation and development while handling numerous challenges.

ITS are advanced applications that offer innovative services for different modes of transport and traffic management. ITS use information and communication technologies to improve transportation system efficiency, safety, and environmental performance. These systems, supported by sophisticated computational models and databases, are central to the transition towards autonomous vehicles and interconnected urban traffic networks. This paper comprehensively analyzes emerging research on data engineering for ITS applications relating to the integration of advanced technologies like RT-DBs, GDBs, and IA.

The survey examines the role of RT-DBs in enhancing Advanced Driver Assistance Systems (ADAS), a key ITS application for vehicle safety and autonomy. It explores the use of GDBs in effectively managing the complex relationships inherent in ITS networks, including connected vehicles, smart infrastructure, and transportation systems. Additionally, the transformational impact of AI and Machine Learning (ML) techniques is analyzed across various ITS implementations encompassing traffic prediction, congestion avoidance, driver behavior modeling, and accident analysis. These results underscore the significance of integrating advanced database systems and AI technologies in shaping the future of intelligent transportation.

The rest of this paper is organized as follows: Section 2 presents ITS integrating databases. Section 3 delves into the ITS using AI. Section 4 offers a comprehensive discussion on the current state of ITS technology, accompanied by recommendations and

[a] https://orcid.org/0009-0005-2988-6516
[b] https://orcid.org/0000-0002-6992-5824
[c] https://orcid.org/0000-0002-5893-9833

Batita, S., Makni, A. and Amous, I.
Intelligent Transportation Systems: A Survey on Data Engineering.
DOI: 10.5220/0012857300003756

In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 169-179
ISBN: 978-989-758-707-8; ISSN: 2184-285X

research challenges for future initiatives. The paper concludes with Section 5.

2 ITS INTEGRATING DATABASES

ITS rely on advanced technologies to handle the vast volume and velocity of data produced by connected vehicles, transit networks, and smart infrastructure. Effective data storage and processing capabilities are crucial for extracting value and enabling enhanced transportation analytics and services. This section analyzes two key data management technologies applied across various ITS implementations: RT-DBs and GDBs.

This section explores research that integrates real-time relational databases and enhances ITS representation with GDBs.

2.1 ITS Enhanced with Real-Time Relational Databases

The modern ITS face the significant challenge of managing increasingly voluminous data, much of which is generated in real-time. This presents a critical challenge for ensuring data consistency and validity. To address this challenge, integrating databases into ADAS has been proposed in several studies. RT-DBs are especially instrumental in ADAS, empowering reliable analysis of dynamic vehicle data from various onboard sensors. By consistently integrating with ADAS, RT-DBs ensure efficient management of constantly evolving driving variables such as vehicle position, speed, and direction. This integration improves the overall effectiveness and responsiveness of ADAS.

In their pursuit of advancing autonomous ADAS, (Marouane et al., 2016) focused on leveraging pattern reuse methods. They introduced the integration of a real-time database system and formulated three distinct design patterns. These patterns were crafted to articulate real-time constraints and manage real-time data, considering both structural and dynamic aspects of the system. Continuing this line of research, (Marouane et al., 2018) proposed a specialized UML (Unified Modeling Language) profile, named UML-RTDB2 (UML-RealTimeDataBase), to express pattern variability and real-time constraints and convey non-functional properties. This further refined the framework for developing sophisticated autonomous ADAS.

However, the objective of (Elleuch et al., 2023) aimed to enhance the performance of Cooperative Advanced Driver Assistance Systems (C-ADAS) with a focus on improving road safety. This enhancement addressed various scenarios such as collision avoidance at intersections, obstacle and dangerous area detection (Elleuch et al., 2021), and management of overtaking maneuvers (Elleuch et al., 2019). The key aspect of their approach involved effectively managing the significant amount of sensor data transmitted through Vehicle-To-Vehicle (V2V) and Vehicle-To-Infrastructure (V2I) communications. To address this challenge, the authors proposed integrating real-time database into C-ADAS within the framework of Vehicular Ad-hoc NETworks (VANETs), which would be managed by a suitable Real-Time Database Management System (RT-DBMS). The incorporation of RT-DB has proven to enhance communication efficiency by reducing the number of messages sent, exchanged, and lost. Additionally, it improves response times by establishing new formulas, conditions, and rules based on the information stored in the real-time database.

In summary, the innovative approaches cited above in these studies signify considerable progress in the development of ADAS and C-ADAS technologies. By emphasizing the reuse of patterns, integrating real-time databases, and formulating specialized UML profiles, these approaches announce a transformative move towards autonomous driving systems that are not only more intricate and capable but also safer and more efficient.

Note that these approaches are based on the real-time Relational Database (RDB) model proposed by (Idoudi et al., 2008). According to this model, real-time data is defined as a quadruple: d = (d $_{value}$, d $_{stamp}$, d $_{avi}$, mde).

Here, d $_{value}$ represents the current value of the data, d $_{stamp}$ denotes the timestamp of the value update, d $_{avi}$ signifies the absolute validity interval, and mde refers to the maximum permissible error between the actual and stored values. This model is implemented to ensure precise data management and processing in autonomous driving systems.

Despite the benefits of the real-time aspect, incorporating RDB presents challenges like implementation intricacies, data processing complexities, and scalability and adaptability issues. As this domain progresses, imminent research is expected to address these challenges, aiming for autonomous driving systems that are more adaptable and flexible while achieving a balance between technological sophistication and practical usability.

2.2 ITS Empowered by Graph Databases

Graph databases are increasingly significant in ITS due to their exceptional ability to handle complex and interconnected data, which is essential for modern transportation networks. This technology is pivotal for managing and analyzing vast amounts of data generated by various components of ITS, including traffic flow, transportation networks, and vehicle communications.

Several studies have delved deeply into this area to illustrate the significance of GDBs in ITS further. For instance, the research by (Oberoi et al., 2018) proposes a structured time-varying graph (TVG) model for understanding dynamic road traffic environments, which integrates both spatial and temporal dimensions. The purpose of this model is to record and analyze the interactions and changes within urban traffic systems while considering static and dynamic elements such as vehicles and road infrastructures. By integrating time-varying node and edge presence and labeling functions, the model improves the precision of traffic flow analysis in urban environments. The paper details the theoretical foundation for the TVG and establishes it with existing work on spatial graph modeling. The goal of this study is to establish the groundwork for future application development, with a focus on using these models to develop graph algorithms that can analyze and interpret traffic patterns. The incorporation of real-world traffic data, gathered by CEREMA in Rouen, France, will simplify the implementation and evaluation of the proposed models and algorithms.

(Wirawan et al., 2019) proposes a unique database design for multimodal transportation, focusing on Semarang. They developed this model using an Oriented Entity-Relationship Diagram (O-ERD), later converting it into a graph database schema executed on the Neo4j graph database. The model includes three main nodes: Shelter, Angkot Stopper, and Closer Place, representing Bus Rapid Transit (BRT) shelters, city transportation, and nearby locations. A unique feature is the "Angkot Stopper" node, symbolizing angkots with flexible stopping points. The model's efficiency was tested through Cypher query language search queries, particularly using the "collect" function to enhance path formation. This approach differs from previous studies, in that it integrates routing algorithms within the graph database system, simplifying route construction and improving route discovery based on passenger destination proximity.

(Bhogaram et al., 2020) highlights the utility of the graph database Neo4j for analyzing transportation networks. They used centrality algorithms to identify important nodes and critical paths within the transportation network, improving resilience against challenges like heavy traffic and natural disasters. This study demonstrates the efficiency of GDBs in analyzing and optimizing transportation systems.

(Chandra et al., 2020) introduced GraphRQI, an advanced algorithm for classifying driving behaviors by analyzing movement paths. This method employs a supervised learning algorithm and spectral analysis of traffic graphs to enhance computational efficiency. In the GraphRQI model, drivers are depicted as nodes in a sparse, undirected, and unweighted, with their interactions indicated by edges. It classifies behaviors such as aggressive or conservative driving based on interactions within traffic graphs. GraphRQI effectively classifies driving behaviors by capturing traffic graph interactions. Its characteristic value algorithm computes the traffic graph with double the speed of previous methods. Tests using traffic videos and autonomous driving datasets, particularly in urban areas, showed a 25% improvement in accuracy over existing driver behavior classification methods. However, its accuracy depends on the reliability of the tracking technology used to monitor road agent positions.

(Zhang et al., 2021) presented a novel method to analyze the structure and behavior of Autonomous Transportation Systems (ATS). ATS represents an advanced form of intelligent transportation, known for its self-organizing and autonomous features. The team's approach involved creating a knowledge graph network to represent the ATS, categorizing it into five distinct nodes: Technology, Demand, Service, Function, and Component. Each of these nodes captures different aspects of the ATS. The research used Neo4j to store structured data, forming a comprehensive knowledge graph of the transportation system network. This graph is composed of two layers: the model layer and the data layer. The model layer outlines the relationships between various entities in the ATS, according to the five elements and their attributes. This creates a structural framework for the system. The data layer, meanwhile, uses Neo4j's capabilities to store and visually present data related to the ATS.

(Bollen et al., 2021) explored Neo4j for managing data in sensor-equipped transportation networks, centering on spatial and temporal data querying using Cypher with custom functions and procedures. Their study aims to bridge the gap between spatiotemporal data and queries, laying advanced analytical improvements in ITS.

(García et al., 2022) introduced the interoperable graph-based Local Dynamic Map (iLDM) for autonomous and connected vehicles. This local database effectively integrates both static and dynamic data from multiple sources employing Neo4j and OpenLABEL, ensuring adaptability in the rapidly changing vehicle technology sector. A thorough performance testing process, involving a vehicle discovery service function, showcased iLDM's superiority over other LDM implementations, making it highly practical for the real-time development of advanced driver assistance systems.

(Maduako et al., 2022) introduced a novel approach for distinguishing high-risk traffic accident locations, incorporating Neo4j to illustrate the dynamic relationship between accidents and the road network as a space-time-varying graph. By analyzing network connectivity through graph analytics metrics such as degree centrality and PageRank, the research identifies high-risk areas for urban planners, enabling proactive accident prevention.

(Zhang et al., 2022) presented a comprehensive analysis of traffic accident data and constructed a knowledge graph to enhance traffic safety management. By integrating multidimensional factors such as people, vehicles, roads, and the environment, the knowledge graph facilitates the acquisition and reuse of valuable insights within structured case data. Through visualization analysis, including accident portraits, classifications, statistics, and correlation paths, the knowledge graph provides complex relationships among accident elements. This helps both researchers and traffic management departments better understand accident characteristics and implement effective measures to avoid accidents and improve overall safety.

(Yuan et al., 2023) focused on developing a knowledge graph for traffic safety management using Neo4j. The study addresses the complexity and scattered nature of traffic safety data by integrating various data types into a structured knowledge graph. It includes creating node and relationship entities to represent different aspects of traffic safety, like illegal acts, vehicle failures, and emergency responses. This study moreover discusses the implementation of query functions using Cypher and rule matching for effective data analysis and decision-making in traffic safety management. It highlights the potential of using Neo4j for organizing and analyzing complex data in the context of traffic safety.

Note that the previous studies collectively underscore the importance of GDBs, through Neo4j, as indispensable tools for transportation network analysis and administration, driving advancements in safety, efficiency, and resilience across various transportation domains. So, each study presented aimed to highlight the unique capabilities and applications of graph databases through Neo4j. For instance, (Oberoi et al., 2018) demonstrated that using a graph representation in Neo4j facilitated effective modeling and analysis of the dynamic spatial and temporal aspects of the urban intersection scenario. While (Oberoi et al., 2018) concentrated on developing theoretical graph models, the experimental findings showcased the practical utility of using Neo4j's graph queries for real-time collision detection. Similar to (Wirawan et al., 2019) and (Bhogaram et al., 2020), the experiments utilized Neo4j's capabilities for route optimization and identifying critical paths, further validating the benefits of GDBs for transportation network analysis. In addition, the integration of AI techniques, like neural networks, aligns with (Chandra et al., 2020) and (Maduako et al., 2022), which utilized Neo4j to analyze driving behaviors and mitigate risks within the transportation systems, emphasizing its role in improving accuracy and facilitating proactive risk identification. Other studies, including (Zhang et al., 2021), (García et al., 2022), (Zhang et al., 2022), and (Yuan et al., 2023) further investigate the development of comprehensive knowledge graphs for ITS, with a focus on real-time data management, spatial queries, and vehicle trajectory prediction. Combining these functions within a unified knowledge graph could pave the way for future advancements, using Neo4j's capabilities to improve safety, efficiency, and robustness across transportation networks. Each cited research highlights the importance of GDBs in ITS, demonstrating their value in improving network operations, analyzing accidents, managing risks proactively, and optimizing traffic flow in various transportation settings.

Table 1 offers a comprehensive comparison of various ITS that have integrated GDBs. Each approach is evaluated based on its objective, use of real-world and real-time data, decision-making, reliance on data quality, and complexity. The table shows variations among approaches, with some excelling in leveraging real-world and real-time data, whereas others distinguish themselves in decision-making. In any case, it's essential to recognize that each approach has its own set of strengths and weaknesses.

For instance, methodologies heavily dependent on real-world and real-time data may yield more precise and timely experiences, essentially improving ITS. However, managing and processing such extensive data volumes can pose challenges.

In conclusion, while Table 1 gives an extensive overview of these various approaches and their contributions to ITS using GDBs, it‘s significant to consider the strengths and weaknesses of these approaches.

3 ITS USING AI

Advancements have significantly influenced the evolution of intelligent transportation systems in artificial intelligence. These innovations are revolutionizing our strategies concerning urban mobility, traffic management, and vehicle safety. Within this dynamic environment, numerous studies have emerged, each investigating various applications and approaches of artificial intelligence to enhance the efficiency and efficacy of transportation systems. This comprehensive review delves into various significant research endeavors that highlight the integration of these technologies in managing and improving traffic flow, predicting driver actions, and optimizing safety protocols at intersections and urban roads. The incorporation of these technologies is not just improving current systems but also laying the way for future advancements in transportation safety and efficiency.

(Meena et al., 2020) introduced a novel tool to accurately and timely predict traffic flow considering diverse environmental factors that can affect traffic like traffic signals, accidents, and road maintenance. Given the recent exponential increase in traffic data and the move toward big data concepts for transportation, today's traffic prediction methods that rely on traffic models are still insufficient for real-world applications. To analyze the vast amounts of data transportation system data with less complexity, the authors intend to use machine learning, genetic algorithms, soft computing, deep learning algorithms, and image processing techniques for traffic sign recognition. The proposed algorithm showcased improved complexity concerns and showed greater accuracy than previous algorithms.

(Hu et al., 2020) simulated driver behavior at signalized intersections under diverse traffic scenarios involving many vehicle types like cars, buses, and motorized three-wheelers. This research collects real world GPS data and video data from vehicles approaching intersections with red signals in Delhi and Mumbai, India. It examines the acceleration and deceleration patterns of these vehicles to determine the impact zone of the intersection - the distance from the intersection where drivers commence decelerating after observing the red signal. The research aims to categorize drivers into categories such as aggressive, normal, and timid based on their acceleration and deceleration behavior. Nevertheless, it finds that drivers cannot be easily classified, and their behavior is better represented by a continuous normal distribution rather than discrete classes. The analysis, which does not employ machine learning or deep learning techniques, emphasizes the complexity of modeling driver behavior in diverse traffic conditions and the dependency on high-quality GPS and video data.

(Lv et al., 2020) used Deep Learning (DL) to address safety issues in ITS. The research examines various aspects such as data transmission performance, prediction accuracy, and route change strategies. In the analysis of the system's data transmission performance, it is found that when the probability of successful transmission is 100% and the λ value between 0.01 and 0.05, it is closest to the actual result, and the data delay is the smallest. In the analysis of prediction accuracy, and using the Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) algorithms, the authors found that in different types of cases, the improved system has the best prediction performance with increasing iterations. After further analysis of the system's route guidance strategy, it is found that the route guidance strategy in this study can effectively inhibit congestion propagation in the face of congested road sections, and achieve the effect of timely evacuation for traffic congestion. As a result of this study, the improved ITS can significantly reduce system data transmission delay, improve prediction accuracy, and effectively change the path in the face of congestion to suppress congestion propagation, providing an experimental reference for further transportation.

(Olayode et al., 2021) compared the Markov Chain Model (MCM) and the Artificial Neural Network (ANN) model for predicting vehicle traffic flow at signalized intersections. Traffic datasets were obtained from South African highways, roads, and intersections, courtesy of the South African Department of Transport. This traffic information was obtained using sophisticated traffic monitoring equipment and techniques, such as inductive loop detectors, video cameras, and GPS- controlled equipment stationed throughout the road. In the ANN model, 100 sets of traffic data were considered, 70% for learning, 15% for testing, and 15% for validation. According to the results obtained in this study, the best traffic dataset training performance was obtained when the number of hidden neurons was 9, giving a good coefficient of determination of 0.96304.

(Karri et al., 2021) aimed to enhance safety at signalized intersections by employing machine learning to address the challenges drivers face in the dilemma zone, the critical moment when a traffic light turns yellow. The research used Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) to classify driver decisions as safe or unsafe. It analyzed behav-

Table 1: Comparative table of ITS that have integrated GDB.

Approach	Objectif	Real world data	Real time data analysis	Decision making	Reliance on data quality	Complexity
(Oberoi et al., 2018)	Develop dynamic traffic graph algorithms	yes	yes	no	yes	yes
(Wirawan et al., 2019)	Design multimodal transportation database	no	no	no	yes	yes
(Bhogaram et al., 2020)	Analyze Critical Transport Paths	yes	yes	no	yes	no
(Chandra et al., 2020)	Classify driving behaviors	yes	no	no	yes	no
(Zhang et al., 2021)	Create and analyze ATS networks	yes	yes	no	no	yes
(Bollen et al., 2021)	Optimize sensor network queries	yes	no	yes	no	yes
(García et al., 2022)	Create an interoperable LDM withOpenLABEL	no	no	no	no	yes
(Maduako et al., 2022)	Identify high-risk traffic locations	yes	yes	no	no	yes
(Zhang et al., 2022)	Analyze traffic accident data	no	no	yes	no	yes
(Yuan et al., 2023)	Develop traffic safety graphs	no	no	yes	no	no

iors from 49 drivers in varied environments. It found that except for the cubic SVM kernel, all SVM approaches predicted behavior with over 85% accuracy, with the linear SVM being the most precise. Although the coarse Gaussian SVM ranked second in accuracy, it demanded more computation time. Compared with KNN and Linear Discriminant Analysis, these also demonstrated high accuracy rates, with 90.1% and 89.4% respectively. The findings offer crucial insights into the efficacy of different ML techniques in predicting driver behavior and potentially reducing accidents at intersections.

(Bagheri et al., 2022) proposed an Artificial Neural Network-based simulation model for gap acceptance behavior. This model was developed through ANN simulations, leveraging real-world data from a comprehensive database collected at a stop-controlled intersection in New Jersey. The practicality of integrating this model into a microscopic simulation tool was evaluated using the Simulation of Urban MObility (SUMO) package's Application Programming Interface (API). The ANN model was trained to mimic drivers' gap acceptance decisions and subsequently implemented in SUMO through its API, enabling the simulation of driver behavior at intersections. This model was benchmarked against the standard SUMO settings and a calibrated version of SUMO based on waiting times and acceptable deviations of vehicles on the minor road approach. The comparative analysis revealed that the ANN-based model outperformed the default and calibrated SUMO models in terms of the selected output metrics. Furthermore, the study highlighted that the ANN model yielded a more accurate representation of vehicle driving behavior on the major road approach to the intersection, indicating its potential for enhancing the realism and accuracy of traffic simulations.

(Singh et al., 2022) distinguished between the Intersection Zone Of Influence (IZOI) and the middle of a block by analyzing the driver's acceleration and deceleration maneuvers. This behavioral data is captured using a Global Positioning System (GPS) in vehicles, particularly after drivers encounter a red signal at an intersection. Additionally, the study attempts to determine the optimal approach length for intersection simulations that affect driver behaviors, categorizing them as aggressive, normal, or shy based on their acceleration/deceleration patterns. This comprehensive approach provides for a nuanced understanding of how different zones at intersections influence driver behavior.

(Bharadiya, 2023) focused on exploring the pivotal role of ML and AI in the development of smart cities. Its primary goal is to understand how these technologies contribute to managing growing urban areas, enhancing economic growth, reducing energy consumption, and improving residents' living standards. Additionally, the study examines the information flow related to Information and Communication Technology (ICT) in smart cities. Methodologically, this research encompasses conducting surveys to identify typical technologies supporting communication in smart cities and systematically evaluating current trends in publications concerning ICT in these urban areas. ML and AI techniques are employed to analyze and interpret the data gathered. The findings reveal that ML and AI are instrumental in various aspects of smart city development, especially in ITS. These technologies are used for tasks like modeling and simulation, dynamic routing, congestion management, and intelligent traffic control, extending their use across different modes of transportation such as rail, and road travel.

(Sayed et al., 2023) provided a comprehensive review of ML and DL techniques utilized in traffic prediction, along with addressing the challenges inherent in applying ML and DL in this domain. The rapid expansion of the Internet of Things (IoT) has facilitated the emergence of smart cities, with ITS at their core, aiming to enhance transportation efficiency and mobility, particularly in addressing traffic congestion. With the increasing adoption of artificial intelligence approaches, the accuracy of traffic flow prediction models has improved significantly.

(Shaffiee Haghshenas et al., 2023) focused on predicting the Level of Road Crash Severity (LRCS) using ML methods applied to real-existing data from 1627 accidents on roads in Calabria, Italy. The main objectives include building accurate prediction models, comparing the performance of ANN and Convolutional Neural Networks (CNN), and identifying the most influential parameters through sensitivity analysis. Results indicate that while there is no significant difference in model accuracy, the CNN model outperforms the ANN model, achieving 68.4% accuracy compared to 61.7%. Sensitivity analysis reveals the number of vehicles and road elements as the most and least important factors affecting LRCS, respectively. The study concludes that these models offer valuable tools for predicting LRCS, with variations depending on specific case studies.

The advancements in AI have significantly influenced the evolution of ITS, evident in the various array of studies examining its applications. The challenges are various going from enhancing safety and traffic flow prediction to facilitating smart city development and crash severity prediction. Thus, (Lv et al., 2020) focused on improving safety within ITS through DL, highlighting enhancements in data transmission performance and congestion management. In contrast, (Olayode et al., 2021) conducted a comparative analysis between the MCM and ANN for traffic flow prediction at signalized intersections, with the ANN demonstrating superior performance. (Karri et al., 2021) aimed to enhance safety at signalized intersections by employing ML techniques to classify driver decisions during critical moments. Conversely, (Bagheri et al., 2022) proposed an Artificial Neural Network-based simulation model for gap acceptance behavior, surpassing standard simulation models in accuracy. Additionally, (Singh et al., 2022) differentiated between intersection zones' influences on driver behavior, providing nuanced insights into traffic safety measures. (Bharadiya, 2023) explored the role of ML and AI in smart city development, emphasizing their contributions to urban growth management and transportation efficiency. (Sayed et al., 2023) provided a comprehensive review of ML and DL techniques in traffic prediction, highlighting their increasing accuracy in addressing congestion challenges. Lastly, (Shaffiee Haghshenas et al., 2023) employed ML methods to predict road crash severity, with CNN outperforming ANN. These comparative analyses highlight the various applications of AI in addressing various challenges within ITS, from enhancing safety and traffic flow prediction to facilitating smart city development and crash severity prediction.

Table 2 presents a comprehensive overview of studies within ITS using AI, detailing their objectives as well as the strengths and weaknesses related to each approach. Each study leverages real-world data to address complex challenges in transportation safety and efficiency. Whereas these approaches harness the power of ML or DL algorithms to extract valuable insights from vast datasets, they also encounter challenges related to the complexity of modeling and the dependency on data quality. The use of AI in these studies offers promising advancements in understanding and managing traffic flow, driver behavior, and road safety. However, ensuring the reliability and representativeness of the data to maximize the efficiency of AI-powered solutions.

In conclusion, Table 2 provides a comprehensive look at the various studies in ITS that utilize AI, highlighting the importance of considering the strengths and weaknesses related to each approach to understand their contributions and potential challenges completely.

Table 2: Comparative table of ITS using AI.

Approach	Objectif	Real world data	Complexity	Dependency on data Quality	ML	DL
(Meena et al., 2020)	Enhance traffic flow prediction	yes	yes	no	yes	no
(Hu et al., 2020)	Analyze intersection driver behavior	yes	yes	no	no	yes
(Lv et al., 2020)	Improve safety in ITS	yes	yes	no	no	yes
(Olayode et al., 2021)	Predict vehicle trajectory	yes	no	no	no	yes
(Karri et al., 2021)	Classify driving behaviors	no	no	yes	yes	no
(Bagheri et al., 2022)	Determine the acceptable gap in intersections	no	no	yes	yes	no
(Singh et al., 2022)	Modeling driver behaviors at intersections	yes	no	no	no	no
(Bharadiya, 2023)	Optimize urban management	no	no	yes	yes	no
(Sayed et al., 2023)	Improve traffic prediction accuracy	no	no	yes	yes	yes
(Shaffiee Haghshenas et al., 2023)	Predict road crash	yes	yes	yes	no	yes

4 DISCUSSION AND FUTURE DIRECTIONS

This section starts by exploring the strengths and challenges associated with the integration of RT-DBs and GDBs in ITS. By combining the capabilities of timely data processing and efficient data structures, these databases can offer promising avenues to improve efficiency and decision-making processes within transportation networks. Then, it discusses the integration of AI in ITS and shows the gained benefits. This paragraph also shows the strong and relevant relationship between databases and AI.

Real-time databases offer significant advantages that make them highly suitable for applications in ITS. These databases handle, analyze, and store data with minimal delay, ensuring that the system always has access to the latest information. This is vital for making time-sensitive decisions that are related to optimizing routes, managing traffic, and responding to incidents. In real-time environments, where quick handling and action on data are crucial, RT-DBs can effectively manage the high-velocity data streams generated by sources such as traffic sensors, cameras, and GPS devices. So, we recommend implementing RT-DBs in ITS and leveraging their benefits by incorporating advanced analytics techniques for real-time data processing, and regularly updating and optimizing database infrastructure to manage increasing data volumes and changing system requirements. Additionally, encouraging collaboration among database engineers, transportation specialists, and AI experts can aid in the improvement of innovative solutions that harness the full capabilities of RT-DBs in enhancing transportation efficiency and safety.

It should be noted that relational databases are mainly used in the research work proposing RT-DBs. Within the relational model based on tabular format, modeling the intricate interconnections among entities such as vehicles, roads, and obstacles poses a significant challenge, resulting in computationally costly queries. This mismatch between the network-like structure of ITS data and the tabular format of relational databases can lead to challenges in handling complex relationships and querying interconnected data, which are common in ITS scenarios. So, this can negatively impact performance and efficient data retrieval. So, we think that the use of relational databases may not be the best choice for ITS applications.

In addition, relational databases struggle to keep up with the rapidly evolving data sources and requirements of ITS due to their rigid and static schema design. This often results in costly and disruptive restructuring efforts.

In contrast, NoSQL databases such as GDBs excel in managing interconnected data while using flexible data models to effectively represent complex relationships. They are characterized by great efficiency and scalability and are often considered a natural form of representation of ITS data. Thus, as the volume and complexity of ITS data continue to grow, relational databases are becoming increasingly impractical for advanced ITS applications, while NoSQL alternatives have architectural advantages that are better aligned with ITS requirements.

To check the performance of the graph database versus the relational database, we have conducted comparative studies by simulating many road situations with obstacles and executing useful queries to pass these situations without danger. The first kind of simulation used an RDB system whereas the second was based on a graph database. Figure 1 shows an example of four vehicles used in the simulation.

The target objective: “Vehicle1” must avoid the obstacle while taking into account vehicles coming in the opposite direction.

The useful queries to execute by “Vehicle1” are:

- Q1: First vehicle in the opposite direction
- Q2: Distance between two vehicles
- Q3: List of vehicles preceding “Vehicle1”
- Q4: List of vehicles in the opposite direction to “Vehicle1”

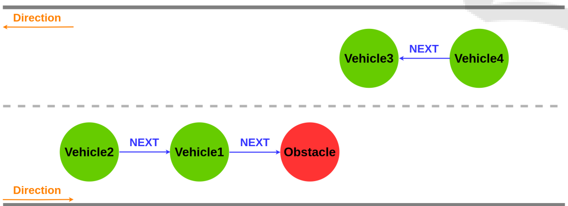

Figure 1: Example of four vehicles on the road.

The results presented in figure 2, figure 3, figure 4, and figure 5 indicate that the use of a graph database is more advantageous compared to RDB. Particularly, in certain cases, the execution time is reduced by a factor of four. Additionally, the quantity of data manipulated by a graph database does not impact the execution time for certain queries.

Therefore, we recommend exploring the integration of NoSQL databases, especially GDBs, in ITS applications to address the limitations of relational databases and improve system performance and scalability.

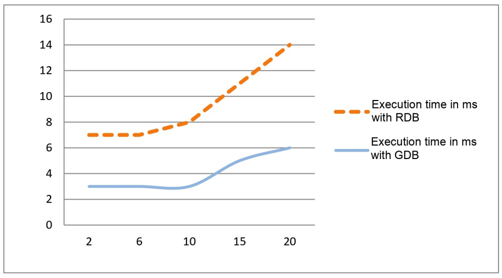

Figure 2: Query execution time of query 1.

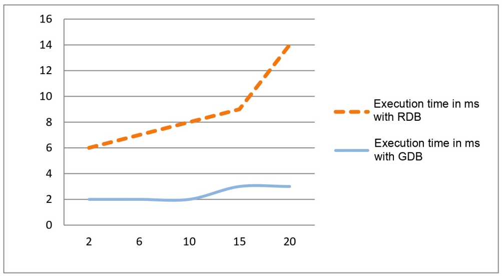

Figure 3: Query execution time of query 2.

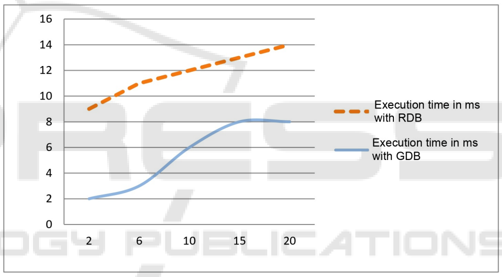

Figure 4: Query execution time of query 3.

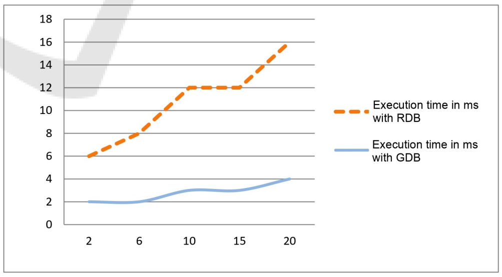

Figure 5: Query execution time of query 4.

Moreover, we believe that integrating real-time capabilities into GDBs could present a promising path forward, combining the benefits of real-time data processing with the flexibility and efficiency of GDBs structures. By using real-time data processing and analysis features, GDBs empower ITS to react promptly to changing traffic conditions and make informed decisions in real-time. Moreover, this integration facilitates accurate traffic pattern prediction and route planning optimization, improving traffic management and decreasing congestion.

On the other hand, the integration of AI within ITS brings various benefits to transportation systems. AI algorithms use extensive datasets from various sources like sensors and cameras to predict traffic flows with precision, optimize route planning, and improve real-time decision-making. Through the utilization of AI, ITS can improve traffic management, boost safety, and optimize transportation networks, ultimately leading to more efficient and sustainable urban mobility solutions.

The integration of AI within ITS underscores the importance of factors like data quality which can have a substantial impact on the performance and accuracy of AI algorithms. Thus, we can distinguish the relevant relationship between AI and databases and consider databases as the basic element of AI. Indeed, databases can deliver the timely and relevant data needed for training data sets. Furthermore, Performance and speed directly impact the ability to process data on time. The ability of the database to grow with data, known as scalability, is another important advantage. In addition, the use of databases allows building a pipeline that performs data-science-driven model hosting.

For more performance, NoSQL databases are required. Indeed, in addition to their high scalability, NoSQL databases support various data structures, which is beneficial for AI applications requiring flexibility in data modeling. Moreover, with NoSQL graph databases, a knowledge graph describes the meaning of relationships between two elements. This kind of graph provides a semantic view of data and can be an efficient way to model semantics which is important to getting pertinent results from AI.

Consequently, we recommend yet another time integration of NoSQL databases, especially GDBs, in ITS applications to build efficient support for AI.

Conversely, AI can have an impact on databases by providing innovative solutions to effectively manage the growing complexity and volume of data generated by applications. Thus, the incorporation of AI techniques into databases improves their performance, facilitating enhanced data processing, storage, and retrieval capabilities. This can increase the potential to significantly augment ITS capabilities, presenting promising avenues for future research and development in both ITS and database management. This enhanced performance plays a critical role in advancing ITS by guaranteeing that can manage the increasing needs for real-time data analysis and decision-making in transportation systems.

5 CONCLUSION

In conclusion, the research highlights the importance of combining ITS with advanced database management systems and AI technologies to revolutionize transportation systems. ADAS with RT-DBs excels in quick decision-making for safety, while ITS with GDBs effectively handles complex network relationships. The versatility of AI in various ITS applications, such as traffic prediction, driver assistance, and accident analysis, has been explored. However, challenges like system complexity, interoperability issues, and data handling constraints persist. These challenges need to be resolved to enable widespread implementation of ITS solutions.

Moving forward, future ITS research should concentrate on overcoming these obstacles to ensure the consistent integration of database systems and AI for more efficient, safe, and adaptable urban mobility solutions. By aligning these technologies, the future of intelligent transportation shows great potential for transforming transportation systems globally. Moreover, the recommendations presented in this research offer valuable insights into addressing these challenges and making the field of ITS.

REFERENCES

Bagheri, M., Bartin, B., and Ozbay, K. (2022). Simulation of vehicles' gap acceptance decision at unsignalized intersections using sumo. *Procedia Computer Science*, 201:321–329.

Bharadiya, J. (2023). Artificial intelligence in transportation systems a critical review. *American Journal of Computing and Engineering*, 6(1):34–45.

Bhogaram, P., Wu, X., He, M., and Okenwa, O. (2020). Optimal and critical path analysis of state transportation network using neo4j. *International Journal of Urban and Civil Engineering*, 14(10):312–317.

Bollen, E., Hendrix, R., Kuijpers, B., and Vaisman, A. (2021). Time-series-based queries on stable transportation networks equipped with sensors. *ISPRS International Journal of Geo-Information*, 10(8):531.

Chandra, R., Bhattacharya, U., Mittal, T., Li, X., Bera, A., and Manocha, D. (2020). Graphrqi: Classifying driver behaviors using graph spectrums. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4350–4357. IEEE.

Elleuch, I., Makni, A., and Bouaziz, R. (2019). Cooperative overtaking assistance system based on v2v communications and rtdb. *The Computer Journal*, 62(10):1426–1449.

Elleuch, I., Makni, A., and Bouaziz, R. (2021). An intelligent and efficient safe driving system. In *International Conference on Hybrid Intelligent Systems*, pages 181–193. Springer.

Elleuch, I., Makni, A., and Bouaziz, R. (2023). Cicaps: a cooperative intersection collision avoidance persistent system for cooperative intersection adas. *The Journal of Supercomputing*, 79(6):6087–6114.

García, M., Urbieta, I., Nieto, M., González de Mendibil, J., and Otaegui, O. (2022). ildm: An interoperable graph-based local dynamic map. *Vehicles*, 4(1):42–59.

Hu, J., Huang, M.-C., and Yu, X. (2020). Efficient mapping of crash risk at intersections with connected vehicle data and deep learning models. *Accident Analysis & Prevention*, 144:105665.

Idoudi, N., Duvallet, C., Sadeg, B., Bouaziz, R., and Gargouri, F. (2008). Structural model of real-time databases: An illustration. In *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, pages 58–65. IEEE.

Karri, S. L., De Silva, L. C., Lai, D. T. C., and Yong, S. Y. (2021). Classification and prediction of driving behaviour at a traffic intersection using svm and knn. *SN computer science*, 2:1–11.

Lv, Z., Zhang, S., and Xiu, W. (2020). Solving the security problem of intelligent transportation system with deep learning. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4281–4290.

Maduako, I., Ebinne, E., Uzodinma, V., Okolie, C., and Chiemelu, E. (2022). Computing traffic accident high-risk locations using graph analytics. *Spatial information research*, 30(4):497–511.

Marouane, H., Duvallet, C., Makni, A., Bouaziz, R., and Sadeg, B. (2018). An uml profile for representing real-time design patterns. *Journal of King Saud University-Computer and Information Sciences*, 30(4):478–497.

Marouane, H., Makni, A., Bouaziz, R., Duvallet, C., and Sadeg, B. (2016). Definition of design patterns for advanced driver assistance systems. In *Proceedings of the 10th Travelling Conference on Pattern Languages of Programs*, pages 1–10.

Meena, G., Sharma, D., and Mahrishi, M. (2020). Traffic prediction for intelligent transportation system using machine learning. In *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, pages 145–148. IEEE.

Oberoi, K. S., Del Mondo, G., Dupuis, Y., and Vasseur, P. (2018). Modeling road traffic takes time (short paper). In *10th International Conference on Geographic Information Science (GIScience 2018)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik.

Olayode, I. O., Tartibu, L. K., and Okwu, M. O. (2021). Traffic flow prediction at signalized road intersections: a case of markov chain and artificial neural network model. In *2021 IEEE 12th International Conference on Mechanical and Intelligent Manufacturing Technologies (ICMIMT)*, pages 287–292. IEEE.

Sayed, S. A., Abdel-Hamid, Y., and Hefny, H. A. (2023). Artificial intelligence-based traffic flow prediction: a comprehensive review. *Journal of Electrical Systems and Information Technology*, 10(1):13.

Shaffiee Haghshenas, S., Guido, G., Vitale, A., and Astarita, V. (2023). Assessment of the level of road crash severity: Comparison of intelligence studies. *Expert Systems with Applications*, 234:121118.

Singh, M. K., Pathivada, B. K., Rao, K. R., and Perumal, V. (2022). Driver behaviour modelling of vehicles at signalized intersection with heterogeneous traffic. *IATSS research*, 46(2):236–246.

Wirawan, P. W., Riyanto, D. E., Nugraheni, D. M. K., and Yasmin, Y. (2019). Graph database schema for multimodal transportation in semarang. *Journal of Information Systems Engineering and Business Intelligence*, 5(2):163–170.

Yuan, D., Zhou, K., and Yang, C. (2023). Architecture and application of traffic safety management knowledge graph based on neo4j. *Sustainability*, 15(12):9786.

Zhang, L., Jiang, S., Huang, K., Xiao, Y., You, L., and Cai, M. (2021). Knowledge graph-based network analysis on the elements of autonomous transportation system. In *2021 IEEE 21st International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, pages 536–542. IEEE.

Zhang, L., Zhang, M., Tang, J., Ma, J., Duan, X., Sun, J., Hu, X., and Xu, S. (2022). Analysis of traffic accident based on knowledge graph. *Journal of advanced transportation*, 2022.