




Article

Digital Twins and Big Data in the Metaverse: Addressing Privacy, Scalability, and Interoperability with AI and Blockchain

Ruoxuan Li ¹, Hemn Barzan Abdalla ^{1,2,*} , Mehdi Gheisari ³  and Hamidreza Rabiei-Dastjerdi ⁴ 

¹ Department of Computer Science, Wenzhou-Kean University, Wenzhou 325060, China; liruox@kean.edu

² Department of Computer Science and Technology, Kean University, Union, NJ 07083, USA

³ Institute of Artificial Intelligence, Shaoxing University, Shaoxing 312000, China; mehdi.gheisari@usx.edu.cn

⁴ School of History and Geography, Faculty of Humanities and Social Sciences, Dublin City University (DCU), D09 V209 Dublin, Ireland; hamid.rabiei@dcu.ie or hamid.rabiei555@gmail.com

* Correspondence: habdalla@kean.edu

Abstract

This paper explores the integration of digital twin technologies and big data in the metaverse to improve urban traffic management. It highlights the importance of technology in mirroring and augmenting our physical and virtual worlds. This study examines how big data and digital twin technologies merge in the metaverse to improve traffic management. Our work applies artificial intelligence (AI) and blockchain technologies to address concerns about privacy, scalability, and interoperability. In a literature review and case study on traffic management, we outline how big data analytics and digital twins can increase operational and decision-making efficiency. This study aims to elucidate the transformative potential of such technologies for urban transport and postulates future areas of social, regulatory, and environmental research gaps.

Keywords: traffic management; big data; metaverse applications; digital twin technology; machine learning



Academic Editors: Wolfgang Kainz and Huayi Wu

Received: 31 May 2025

Revised: 9 August 2025

Accepted: 18 August 2025

Published: 20 August 2025

Citation: Li, R.; Abdalla, H.B.; Gheisari, M.; Rabiei-Dastjerdi, H. Digital Twins and Big Data in the Metaverse: Addressing Privacy, Scalability, and Interoperability with AI and Blockchain. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 318. <https://doi.org/10.3390/ijgi14080318>

Copyright: © 2025 by the authors. Published by MDPI on behalf of the International Society for Photogrammetry and Remote Sensing. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The metaverse, first proposed by Neal Stephenson in *Snow Crash* (1992) [1], is a virtual world where digital avatars interact with the real world, representing the fusion of physical and virtual reality. The fusion of big data and digital twin technology enables the metaverse to simulate complex systems such as traffic management through real-time analysis and virtual modeling.

This study examines the synergy between big data and digital twins within the metaverse, specifically in the context of traffic management. It presents a case study utilizing a random forest model and digital twin simulation to illustrate their impact on the real world, offering valuable insights for optimizing urban traffic.

Motivation: Road traffic accidents are one of the leading causes of death worldwide. According to the World Health Organization (WHO) [2], approximately 1.3 million people die in road traffic accidents each year, and nearly 50 million people suffer non-fatal injuries. Therefore, it is necessary to study which factors have the most significant impact on the occurrence of road traffic accidents.

Contributions: We propose a novel framework integrating big data (Hadoop, Spark) and digital twins for real-time traffic risk prediction. Our case study uses a random forest model to demonstrate its practical application in an urban environment.

Structure: This paper is organized as follows: Section 1 introduces the metaverse and big data; Section 2 details the technical foundations; Section 3 introduces applications; Section 4 explores challenges; Section 5 provides case studies; and Section 6 summarizes and discusses possible limitations of applications.

1.1. The Definition of Big Data and the Metaverse

1.1.1. Metaverse

The metaverse is a digital ecosystem integrating virtual and augmented realities, enabling immersive, interactive experiences that mirror or transcend the physical world. It is a shared virtual space for social, economic, and cultural activities, supported by technologies like AR, VR, and the Internet [3–6]. Key characteristics include real-world mapping, user-driven interactions, and a novel social infrastructure.

1.1.2. Big Data

Seref Sagiroglu defines big data as massive datasets whose sheer size, variety, and complexity make them difficult to store, analyze, visualize, process, or generate results. This information can help companies or organizations gain richer insights and a competitive advantage. Therefore, implementing big data requires the most accurate analysis and execution possible [7].

1.2. The Importance of Their Intersection

The development of the metaverse cannot be separated from the support of big data. For example, in the field of games in the metaverse, the metaverse will accumulate massive user behavior data. Effectively analyzing these data and mining user behavior patterns are significant for optimizing design and providing personalized services.

Zhang et al. [8] reviewed the application of big data technology in the metaverse, covering multiple fields in the virtual reality environment. Their study pointed out that big data technology is mainly used in the metaverse to process user-generated data streams, such as interaction data and user behavior analysis in the virtual world. They investigated key areas such as health, transportation, and business and found that big data technology has significant potential in the following aspects:

Data analysis and prediction: Optimize the system performance and user experience of the metaverse by analyzing user data.

Virtual human modeling: Utilize big data technology to create digital twins and virtual characters, thereby enhancing the immersive experience.

Big data is indispensable in the metaverse, supporting data processing, user experience optimization, and security management. Current research has initially explored its application potential, but issues such as privacy, security, and technical bottlenecks still need to be addressed.

1.3. Related Work and Literature Search Method

1.3.1. Related Work

Several studies have been conducted in academia and industry to explore integrating these technologies and their applications at the intersection of the metaverse, digital twins, and big data. The following briefly reviews the related studies and highlights their differences compared to this paper.

Rathore et al. [9] published “The Role of Artificial Intelligence, Machine Learning, and Big Data in Digital Twins: A Systematic Literature Review, Challenges, and Opportunities” which explored the role of artificial intelligence (AI), machine learning (ML), and big data analysis in the creation and application of digital twins (DTs) through a systematic literature review (SLR). Their study reviewed the application of digital twins in various

industries such as manufacturing, healthcare, transportation, construction, and smart cities. It emphasized how the Internet of Things (IoT), big data [10–12], and AI-ML technologies can enhance the capabilities of digital twins through real-time data collection and analysis. The authors proposed a data-driven reference architecture and discussed data collection, analysis, and standardization challenges. In addition, they identified tools for digital twin development, such as Siemens MindSphere and TensorFlow, and explored future research directions in predictive analytics and health management. Differences compared to this article: Rathore et al.'s [9] study provides a broad overview of digital twins across multiple industries, focusing on general technical applications of AI-ML and big data [13], encompassing areas such as smart manufacturing, healthcare, and transportation. However, their study does not delve into the context of the metaverse, which is the core focus of this article. This article specifically examines the integration of digital twins and big data in the metaverse environment, with a particular focus on privacy, scalability, and interoperability challenges in the case of traffic management. Furthermore, this article proposes a novel framework that combines blockchain technology with AI to address data security and real-time processing issues in the specific context of the metaverse, issues not discussed in detail in [9]. The traffic management case study provides a concrete, practical application, filling a gap in the literature on digital twin applications in the metaverse environment.

1.3.2. Literature Search Method

To ensure the rigor and relevance of our research, we employed a systematic approach to identify and analyze relevant research and case studies.

- **Database Selection:** We searched multiple academic databases, including Google Scholar, IEEE Xplore, and the ACM Digital Library. These databases were selected for their comprehensive coverage of computer science, data analysis, and virtual environment research.
- **Search Keywords:** We combined keywords to capture the intersection of the metaverse, digital twins, and big data. Key keywords included “Metaverse digital twins”, “Big data in metaverse”, “Digital twins”, “Metaverse applications in urban systems” and “Traffic accidents”.

2. Technological Foundations

2.1. Overview of Big Data Technologies

Big data technologies are critical for managing the vast datasets generated in the metaverse, enabling real-time analytics and user personalization. This section evaluates Hadoop and Spark, focusing on their applicability in metaverse contexts, particularly for traffic management simulations.

2.1.1. Hadoop

Hadoop is a distributed framework for processing large-scale data, comprising the Hadoop Distributed File System (HDFS) for high-throughput storage and MapReduce for parallel computing [10,14]. In the metaverse, Hadoop efficiently stores user interaction logs (e.g., avatar movements, transactions) across clusters, supporting applications like virtual real estate analysis. However, its disk-based processing results in higher latency, making it less suitable for real-time metaverse tasks such as dynamic traffic simulations.

2.1.2. Spark

Apache Spark is an open-source, distributed computing framework built for processing massive datasets. It offers speed, flexibility, and user-friendliness, making it a top choice for big data analytics and machine learning applications [15]. Spark provides high-level

APIs to support batch processing, real-time streaming, machine learning tasks, and low-level APIs for tailored data processing needs. One of the main features of Spark is its ability to handle data processing tasks in memory, which can significantly improve performance over traditional disk-based systems. Spark also provides fault tolerance through its use of lineage information, which enables the recovery of lost data in case of node failures. Spark supports multiple programming languages, including Scala 2.13, Python 3.9+, R 3.5+, Java 17/21, and Julia, making it easy for users to work with their preferred language. It also supports integration with popular big data technologies such as Cassandra, Hadoop, and Kafka, allowing users to process data from different sources easily. Spark provides a range of deployment options, including standalone, Yarn, Mesos, and Kubernetes, making it easy to deploy on various clusters and cloud platforms. It also provides various tools for monitoring and debugging Spark jobs, including a web-based interface for tracking job progress and diagnosing errors. Spark is a powerful and flexible platform for large-scale data processing and analytics. Its ease of use, flexibility, and performance make it a popular choice in big data processing tasks, machine learning, and real-time stream processing [2,9].

2.2. The Core Components of the Metaverse

2.2.1. Interactive Technology

Interactive technologies are central to the metaverse, offering users immersive and engaging experiences. These primarily encompass virtual reality (VR), augmented reality (AR), and mixed reality (MR). Currently, the most prevalent method for users to engage with the metaverse's virtual environments is through VR headsets. These head-mounted devices isolate users from external sights and sounds, creating a fully immersive visual and auditory experience that makes them feel like they are in a digital world. Additionally, users can actively interact with and contribute content to these virtual spaces. VR has the characteristics of fully synthetic views. The most common way users interact is to wear head-mounted VR glasses with head tracking and somatosensory controllers [16]. Therefore, the user is completely immersed in the virtual environment and interacts with virtual objects. Users can also create virtual environments, such as VR education and VR painting.

AR adds interaction with the physical world on top of VR, focusing on enhancing the physical world. This is a necessary step to achieve digital twins. In theory, our operations in the virtual world should be fed back to the physical world through various senses, including hearing, vision, smell, and taste. Ensuring seamless interaction with virtual digital entities in AR is the key to AR. Virtual objects and physical world objects can be connected through the metaverse, and operating virtual objects with bare hands in the physical world is what AR is about. A well-known freehand interaction method, Voodoo Dolls [17], enables users to intuitively select and manipulate virtual objects using pinch gestures with their hands. Another practical approach, HOMER, enhances user interaction by projecting virtual rays from the user's hands to highlight and control augmented reality (AR) objects, streamlining the selection and manipulation process. MR is positioned between VR and AR, enabling users to interact with virtual objects in the physical world. MR focuses more on the interoperability between the physical and virtual worlds, so it can be said to be an enhanced version of AR. This interoperability is essential to realizing the concept of digital twins in the metaverse.

2.2.2. Artificial Intelligence Technology

Artificial intelligence (AI) encompasses methods and theories that allow machines to gain knowledge from experience and execute diverse tasks. First introduced in 1956, AI has recently achieved top-tier results across multiple domains, such as natural language

processing, visual recognition, and personalized recommendation systems. Artificial intelligence is a broad concept that includes representation, reasoning, and data mining. The paramount foundational technology of the metaverse, artificial intelligence is essential. After entering the metaverse, people will exist and move as digital avatars. The digital avatars' all-around perception capabilities in vision, hearing, touch, and other senses are closely tied to AI technology. Technologies such as AI-powered computer vision, natural language processing, and haptic feedback have found practical and effective uses. These AI systems rely on vast datasets and training models to minimize errors and align neural network outputs closely with actual values, achieving high accuracy for tasks like classification and prediction. By integrating artificial intelligence into the metaverse, the performance and functionality of metaverse applications can be significantly enhanced and refined. Big data, driven by artificial intelligence technology, and metaverse applications are bound to generate massive amounts of data during operation. The two complement each other and bring out the best in each other.

Importance in Metaverse Mission: AI can process massive datasets, enabling personalized experiences and operational efficiency. In traffic management, AI can predict high-risk scenarios and reduce the likelihood of accidents through proactive intervention. However, challenges include high computational costs (e.g., GPU clusters are required to train large models) and privacy risks posed by user data collection, so integration with blockchain is needed for secure data management.

2.2.3. Blockchain

The metaverse is entirely digital, encompassing user avatars, large-scale, fine-grained maps of different locations, digital twins of actual entities and systems, etc. As a result, a massive amount of data that are difficult to understand are generated. Due to limited network resources, uploading such a large amount of data to a centralized cloud server is impossible. The rapid development of blockchain technology can effectively solve this problem. For example, blockchain can be applied to data storage systems to ensure the decentralization and security of the metaverse. The rapid development of blockchain technology provides technical support for building a safe and efficient economic system in the virtual world. Blockchain is a distributed database that uses blocks rather than organized tables to store data. A new block is filled with user-generated material and will be further connected to earlier blocks. Every block has a chronological link. Users use a consistent paradigm to store blockchain data locally and synchronize it with other blockchain data stored on peer devices. In the blockchain, users are referred to as nodes; after linking, each node keeps an exhaustive record of all the data saved on the blockchain. Using millions of other nodes as a reference to fix a node fault is possible. Therefore, under blockchain technology, data security is guaranteed. In the economic system of the metaverse, blockchain technology can be used in financial transactions, digital copyright confirmation, and improving supply chain management efficiency, achieving true core decentralization. In recent years, scholars and research organizations have continued to deepen their research on blockchain technology and have achieved remarkable results.

Bandara et al. [18] addressed the issue of privacy leaking in centralized storage infrastructures. Casper is a digital identification platform built on self-sovereign identity and blockchain to store users' identity credentials on their mobile wallet app. Only credential proof is kept on the system's decentralized storage system, which is based on blockchain technology. Casper verifies identity information from credential proofs using a zero-knowledge proof technique. Research indicates that the system can provide high transaction throughput and is scalable. A blockchain-based CDFL system framework called

TrustFed was presented by Rehman et al. [19] to identify model poisoning assaults, enable equitable training conditions, and preserve the reputation of participating devices.

2.2.4. Network and Computing Technology

Network computing refers to the technology users use to transmit and process information through dedicated or public computer networks. Network and computing technologies are the cornerstones of the metaverse. In the application scenarios of the metaverse concept, accessing system databases, real-time data transmission with real-world terminal devices, real-time interaction between users in virtual space, and other related routine operations all require the support of low-latency, high-bandwidth, high-quality networks and high-performance computing platforms. This section will mainly introduce the application of 5G/6G, cloud computing, and edge computing in the metaverse.

(1) 5G/6G

5G: As a cutting-edge communication framework, 5G offers impressive capabilities: internet speeds reaching 1 GB/s, delays as low as 1 millisecond, and supporting up to 1 million device connections per square kilometer. These features make 5G ideal for streaming high-quality, real-time content in the metaverse, laying a strong foundation for its infrastructure. However, challenges remain. Researchers have tackled these issues with innovative solutions. For instance, Park et al. [20] developed a method using deep reinforcement learning to manage interference in fast-moving environments, such as 5G-connected vehicle networks, grouping radio heads and clustering vehicles to boost energy efficiency and service reliability. Similarly, Kottursamy et al. [21] proposed an eNB/gNB-aware data retrieval algorithm and an activity- and size-based data replacement algorithm to optimize, sort, and cache data items efficiently. Data items are selected based on popularity and cached in the D-RAN for efficient data replacement. In addition, they proposed a cost-optimized radar-based data retrieval algorithm that helps find data proximity in neighboring eNBs. In the proposed technique, unique content is maintained at both ends of the cluster to help expand content diversity within the cluster. The model proposed by Kottursamy et al. [21] achieves lower latency, congestion, and cache hit rate in 5G networks.

6G, the envisioned sixth-generation mobile communication standard, represents an advanced wireless network technology poised to revolutionize global connectivity. 6G aims to deliver universal, uninterrupted coverage by seamlessly integrating terrestrial wireless systems with satellite networks. This comprehensive network will extend connectivity to even the most remote regions, enabling access to critical services like telemedicine for isolated communities and online education for children in underserved areas. Furthermore, by combining global satellite navigation, telecommunication, and earth observation systems with a robust 6G terrestrial infrastructure, this technology will support applications such as accurate weather forecasting and rapid disaster response, fostering a truly interconnected global ecosystem. This is the future of 6G. 6G communication technology is no longer a simple network capacity and transmission rate breakthrough. It is more about narrowing the digital divide and realizing the “ultimate goal” of the Internet of Everything. This is the meaning of 6G. The data transmission rate of 6G may reach 50 times that of 5G, and the delay is shortened to one-tenth that of 5G. It is far superior to 5G in terms of peak rate, delay, traffic density, connection density, mobility, spectrum efficiency, and positioning capability [22].

(2) Cloud computing

Cloud computing is a form of distributed computing that divides large-scale data processing tasks into numerous smaller segments. These segments are handled across a network of servers in the “cloud”, collectively processing and analyzing the data before

delivering results to users. Cloud computing is essential for managing the vast amounts of data generated in the metaverse. It primarily supports metaverse applications through robust data processing and storage capabilities. Given the limited computational power of local devices, resource-intensive tasks in the metaverse depend on the cloud's high-performance infrastructure for efficient data handling. Additionally, with the restricted storage capacity of end-user devices, cloud platforms provide scalable, distributed storage solutions to accommodate massive datasets. Recent advancements in cloud computing have significantly strengthened the technical foundation for realizing metaverse applications.

Chkirbene et al. [23] proposed a new weight class classification scheme to protect the network from malicious nodes' attacks while alleviating the data imbalance problem. After that, they also designed a new weight optimization algorithm, which uses these decisions to generate a weight vector containing the best weight for each category. The model improved the overall detection accuracy, and even for a relatively small number of training samples, it could maximize the number of classes that could be correctly detected. Chraibi et al. [24] utilized a deep Q network (DQN) algorithm to develop an enhanced minimum cloud scheduling (CS) maker space scheduling algorithm, addressing the issue of underutilization of cloud computing servers and the significant loss of execution time. To enhance the convergence of the DQN model, the team recommended a new reward function. The recommended MCS-DQN scheduler showed the best scheduling results in minimizing the completion time metric and other similar schedulers (task waiting time, virtual machine resource usage, and inconsistency with the algorithm).

(3) Edge computing

The swift advancement of IoT technology has facilitated the interconnection of an increasing number of common things with autonomous functionalities, realizing the interconnection of all things [25]. Thanks to the characteristics of IoT, all walks of life are using IoT technology to achieve digital transformation quickly, and more and more industry terminal devices are connected through the network. However, as a vast and complex system, IoT has different application scenarios in various industries. According to statistics from third-party analysis agencies, by 2025, more than 100 billion terminal devices will be connected to the Internet, and the amount of terminal data will reach 300 ZB. Traditional data processing approaches require all collected data to be transmitted to centralized cloud platforms for analysis, which often leads to challenges such as significant network delays, overwhelming device connections, difficulties in handling large data volumes, limited bandwidth, and high energy consumption. Edge computing has emerged as a solution to address these limitations, particularly latency issues and inadequate real-time analytics. This technology delivers intelligent services at the network's edge, close to the data source, through a distributed platform that combines network, computing, storage, and application functionalities. By processing data near their origin, edge computing minimizes reliance on distant cloud servers, enabling faster and more efficient data handling for real-time applications. By migrating the network, computing, and storage capabilities of cloud computing to the edge, it provides intelligent services close to users, meeting requirements such as low latency, high bandwidth, real-time processing, security, and privacy protection [26,27]. In some cases, when edge terminal devices submit local computing tasks to cloud computing servers, they often need to occupy a large amount of network bandwidth. When the terminal device and the cloud server are far apart, the network delay will increase greatly, which will undoubtedly affect user experience. Edge computing calculates, stores, and transmits data at the place closest to the end user and the device, which can greatly reduce the delay in user experience.

2.3. Integration Points Between Big Data and Metaverse Technologies

Data-driven perspective: The operation and development of the metaverse are highly dependent on the support of big data. By collecting and analyzing user behavior data, interaction data, economic transaction data, etc., we can gain a deeper understanding of user needs and preferences, enabling us to optimize the virtual environment, social system, and economic system of the metaverse. For example, on the metaverse social platform, by analyzing users' social behavior and interests, we can recommend suitable social partners and virtual activities to enhance the user's social experience.

Data management and storage: The large amount of data in the metaverse requires efficient and secure data management and storage solutions. Big data technology can provide distributed storage, real-time processing, data mining, and other functions to meet the metaverse's data management and storage needs. For example, the data platform built using big data technology can process more than 100 million pieces of data per day, providing strong data support for the metaverse's operation.

Intelligent decision-making: Big data analysis can help the metaverse make intelligent decisions and optimize resource allocation and operation strategies. We can predict market trends and user needs by analyzing user behavior data, supporting the metaverse's development. For example, in the field of cultural tourism in the metaverse, big data analysis can be used to understand tourists' travel habits, consumption preferences, and other information, providing a decision-making basis for scenic spot planning and product development.

Economic system: In the economic system of the metaverse, big data also plays an important role. By collecting and analyzing economic transaction data, fair and transparent economic rules and incentive mechanisms can be established to promote the healthy development of the metaverse economy. For example, in the metaverse economic system based on blockchain technology, big data can help track and verify transaction data to ensure the fairness and transparency of transactions [28].

Figure 1 illustrates the interaction of big data, AI, and digital twins in the traffic management case study, from sensor data collection to actionable traffic management decisions.

- **Data Collection:** Traffic sensors capture real-time data, including vehicle counts, speeds, road surface quality, and rainfall intensity, which are processed to analyze trends.
- **Data Storage:** Hadoop's HDFS stores historical traffic data, enabling scalable archiving and model training for predictive analytics.
- **Data Analysis:** Spark processes streaming data to feed an AI model that predicts accident risk based on features such as traffic density and road surface quality. These predictions inform real-time decision making.
- **Metaverse Applications:** Digital twins simulate traffic scenarios, mapping real-world situations into virtual environments. Pygame-based visualization tools display risk levels (yellow for low risk, blue for medium risk, and red for high risk), enabling planners to test interventions virtually.
- **Decision Support:** Decision support systems generate recommendations based on real-time analytics and digital twin outputs, such as implementing speed limits in high-risk areas.
- **User Experience:** Real-time visualizations and actionable recommendations enhance traffic safety and provide an immersive interface for city planners to monitor and optimize traffic flows.

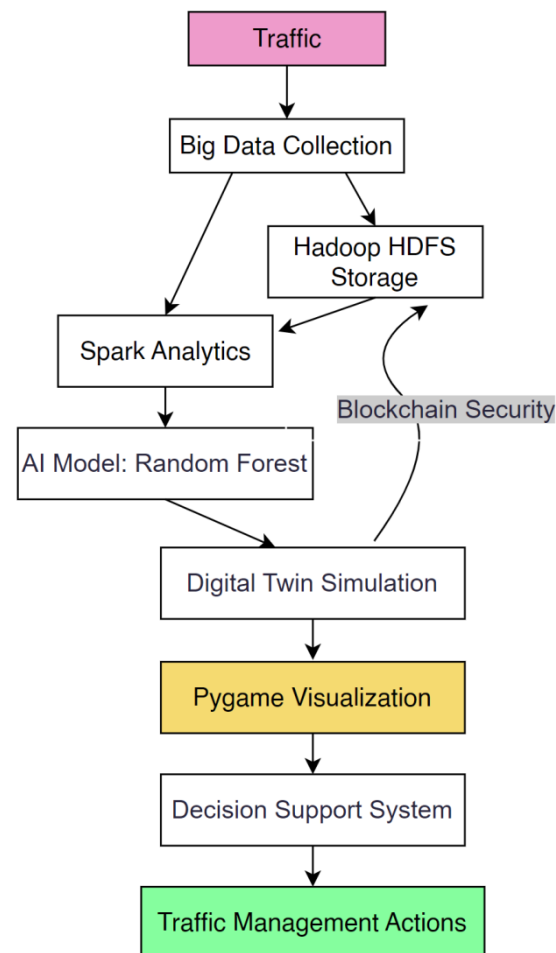


Figure 1. Interaction workflow for traffic management in metaverse.

3. Applications and Use Cases

3.1. Virtual Real Estate Analysis

Compared with traditional display methods such as floor plans and model houses, the biggest feature of virtual reality is that it can make users feel immersed. Customers can experience the house's structure and decoration from different angles and conduct a "real-life" tour. From indoor model rooms to the construction of outdoor scenery, the restoration of the entire community, and even the surrounding roads, VR house viewing has begun to break through the limitations of the interior and bring customers an immersive experience from a larger space. The application of VR technology in real estate and interior design is an inevitable trend. The model rooms in the building are still necessary, but temporary model rooms will be replaced because they address the most significant challenges for developers: time, capital, and space costs. For developers, time cost is the most essential factor. The primary requirement of real estate developers, who operate in an asset-heavy sector, is the quick turnover of cash [29]. Temporary model rooms have greatly improved pre-sales efficiency and accelerated developers' return; however, the emergence of VR has also led to a lag in efficiency. It takes at least 3 to 6 months to build a temporary model room, from planning and design to purchasing materials, signing a design contract, signing a decoration contract, construction, soft furnishings, and finally, acceptance and promotion. These 3 to 6 months for the marketing department are warm-up activities. The emergence of VR technology has reduced the time from 3 to 6 months to 10 to 15 days. Through virtual house viewing, marketers can lock in customers in advance, sell in advance, and win time bonuses [29].

3.2. Enhanced User Engagement and Behavior Analysis

In Zhang et al. [30], SPAR, a sophisticated recommendation system, was introduced to effectively capture comprehensive user preferences from extensive engagement histories. This framework leverages pretrained language models (PLMs), multiple layers of attention mechanisms, and a sparse attention approach to process user histories in a session-oriented format. It integrates user and item features for accurate engagement predictions while preserving distinct representations for each, facilitating efficient deployment in real-world applications. Additionally, the system enhances user profiling by deriving global interests from engagement data using a large language model (LLM). Rigorous testing on two benchmark datasets confirms that this framework surpasses current state-of-the-art (SoTA) recommendation methods.

3.3. Real-Time Data Visualization in Virtual Environments

Metaverse data are highly real-time, frequently updated, and accumulated in large quantities. The visualization and analysis of these types of data must consider real time and interactivity under enormous amounts of data to prevent the display interface from losing points or freezing. Wang et al. [31] proposed an innovative method to enhance emergency response training by implementing real-time data visualization in CVE using the SIEVE framework. Integrating dynamic data such as weather, traffic, and event status can create a realistic immersive environment, increasing user engagement and supporting behavioral analysis. This helps improve decision-making and collaboration and has the potential to save lives and property in real-world emergencies.

3.4. Predictive Analytics for Metaverse Economies

As the metaverse framework becomes more and more complete, the academic community has proposed the “metaverse economy” concept. The metaverse economy encompasses the overall production, exchange, circulation, and consumption of virtual commodities within the metaverse. In the metaverse economy, individuals can produce and consume across geographical restrictions. Economic operations do not need to follow the existing economic framework; they may generate new economic theories and laws and even react to the real world.

The metaverse is a virtual world independent of the real world, but it also has production factors and transaction media for the smooth operation of the economy. Meta-products convert their multiple values into economic values through “meta-currency” in economic circulation and become “meta-assets” in the metaverse economic system. Meta-identities, meta-products, meta-assets, and meta-currencies together constitute the “element layer” of the metaverse economic system. Here, we will focus on meta-products and meta-currencies. First of all, we address meta-products. Most digital products in the metaverse economic system rely on user creation (User-Generated Content, UGC). Meta-products developed based on UGC are an important foundation for the operation of the metaverse economy. The main forms of meta-products are as follows: virtual items generated by code algorithms, such as virtual buildings, game equipment, personal 3D images, etc.; digital information, such as news, images, books, movies, and music; digital services that control human emotional characteristics, such as interactive social interaction, distance education, online games, cultural and tourism services, government services, etc. At present, the prototype products of the metaverse are mainly concentrated in games which already have virtual scenes and digital identities and are more in line with the primary characteristics of the metaverse economic form. Roblox, the pre-eminent global multiplayer online creativity game, is a notable example, having produced 25 million digital products in 2021 alone, with players obtaining or purchasing 5.8 billion virtual products.

Next is meta-currency, the medium of exchange, i.e., the digital currency, in the metaverse economy. The initial digital currency was widely used in the game community, recognized and utilized by specific members to facilitate the discovery and exchange of value in digital products within the metaverse. It was not linked to the legal currency of the economy. As the metaverse economy and the real economy continue to merge, the question of whether the digital currency of the metaverse should be connected and exchanged with the legal currency of the real world has become a key issue for global currency governance. If non-governmental entities are allowed to issue digital currency and freely exchange it with physical currency, the risks associated with digital currency may be transmitted to the real economy, impacting the effectiveness of monetary policy and the economic stability of the real economy.

4. Challenges and Solutions

4.1. Data Privacy and Security Issues

4.1.1. Data Modification and Tampering

As the metaverse grows, the volume of data also increases, which increases the risk of hacking, tampering, and unauthorized access. In traffic management, the real-time data transmission of vehicle locations and risk levels exacerbates the risk of data manipulation and can compromise safety decisions. The decentralized nature of the metaverse exacerbates these threats as traditional centralized security measures are already insufficient.

4.1.2. Data Privacy Protection

In virtual traffic simulations, collecting and processing personal data (e.g., user locations and behavior patterns) raises serious privacy issues. In our case study, aggregating data from traffic sensors and digital twins for predictive analysis poses the risk of compromising personal privacy, especially without reliable anonymization. This challenge is critical because user trust is essential for the widespread application of the metaverse.

4.2. Scalability Issues

The metaverse relies on blockchain for secure data management (related to our traffic simulation logs), which introduces scalability limitations. The growing number of transactions and data storage requirements place pressure on blockchain capacity, resulting in slow and inefficient on-chain processing. This problem is particularly acute when the system scales to handle real-time urban traffic data across multiple regions.

4.3. Real-Time Data Processing

Real-time data processing is critical for metaverse applications, including our traffic management system, which requires instant updates of risks. However, data transmission delays across networks and massive amounts of data from IoT devices (e.g., 1 GB per hour per sensor) hinder timely decision making. The current infrastructure struggles to maintain latency below 10 milliseconds, which affects the accuracy of traffic simulations and visualizations.

4.4. Interoperability Among Diverse Platforms and Technologies

The metaverse integrates multiple technologies, such as AI models, digital twins, and visualization tools, which often run on incompatible platforms. This lack of interoperability complicates data exchange and system synchronization, and the lack of standardized protocols between these components limits the system's effectiveness in dynamic, multi-platform environments.

5. Our Case Study (A Case Study on the Integration of Big Data and Digital Twin Technology in Traffic Management)

5.1. Research Objectives

This case study aimed to integrate big data analysis and digital twin technology to provide intelligent decision support for traffic management. This project successfully achieved dynamic risk prediction and traffic optimization recommendations based on a large dataset of traffic accidents combined with a random forest prediction model and a digital twin simulation environment.

5.2. System Architecture

The system includes multiple core modules, each with different functions, and ultimately works together to provide efficient traffic management support. The system architecture is as follows:

- Data acquisition and processing module: By loading traffic accident data, traffic flow, weather information, etc., the raw data are cleaned, processed, and standardized.
- Digital twin module: This module achieves real-time simulation of traffic flow, traffic accidents, and the effects of different traffic control strategies.
- Risk prediction model: This model predicts traffic accident risks through machine learning algorithms (such as random forest regression) and provides prediction results.
- Decision support module: Based on risk prediction results and real-time traffic data, this generates traffic management suggestions, such as adjusting traffic lights, implementing speed limits, increasing patrols, etc.
- Visualization module: Through visualization tools such as Pygame 2.6.1 and Matplotlib 3.5.3, all data and prediction results are presented to users intuitively.

5.3. Implementation of Core Modules

5.3.1. Data Processing and Analysis

The Data Processor class is responsible for loading, processing, and analyzing data from the traffic accident dataset. Random sampling, calculating the average number of accidents, and other methods provide basic data for subsequent risk prediction and decision support. This class's `analyze_time_trends` method analyzes the distribution trend of accidents in different time periods to help predict high-risk periods.

5.3.2. Digital Twin Model

The digital twin class builds a dynamically updated digital twin model based on traffic accident data. By initializing different types of vehicles [16] and calculating the risk level of each vehicle based on different traffic accident parameters (such as rainfall, traffic density, speed, etc.), the digital twin model can simulate traffic flow, vehicle behavior, and changes in traffic signals. The model provides real-time traffic status feedback for the decision support module.

5.3.3. Risk Prediction Model

The Prediction Model class uses machine learning (random forest) to predict the risk of traffic accidents. The model predicts the possibility of accidents for each traffic event by analyzing multiple features in historical data (such as traffic density, vehicle speed, weather, etc.) and generates risk values. The prediction method of this class predicts the risk value of the accident based on the input traffic conditions (such as traffic density, vehicle speed, etc.).

5.3.4. Decision Support System

The Decision Support class generates corresponding traffic management suggestions by analyzing risk values and parameters (such as traffic density, road quality, rainfall, etc.).

5.3.5. Visualization Display

The Visualization class is responsible for visualizing the output of the digital twin model, risk prediction results, and management suggestions. Through Pygame 2.6.1 and Matplotlib 3.5.3, the system can display information such as traffic flow, risk value, and management suggestions in a real-time updated graphical interface and display traffic accident-related parameters (such as traffic density, vehicle speed, weather conditions, etc.) through text and charts.

5.4. System Workflow Diagram: Sequential Process from Data Collection to Visualization

Our system workflow represents a complete process of processing and using traffic accident data for decision-making and simulation analysis. Figure 2 shows that the process starts with gathering, loading, and preprocessing raw data for cleansing and making them relevant. The first step of the process is described in Table 1, which shows the most essential information of the raw data processed by the Data Processor class. The Analysis reprocess class will perform fundamental analyses, including averaging, required for this study to prepare the data for higher-order processing. The cleaned data from the database are analyzed in a prediction model with a random forest regressor algorithm to find potential risks resulting from traffic incidents. Its output provides input to the decision support class, synthesizing all the information and developing recommendations and strategies regarding risk mitigation. The Digital Twin class dynamically visualizes traffic patterns and scenarios, hence assisting in decision-making. Further, the Visualization class integrates all processed data into one interactive graphical user interface, significantly improving users' monitoring and responding capabilities. This systematic way of organizing such data has been instrumental in optimizing traffic management and improving road safety.

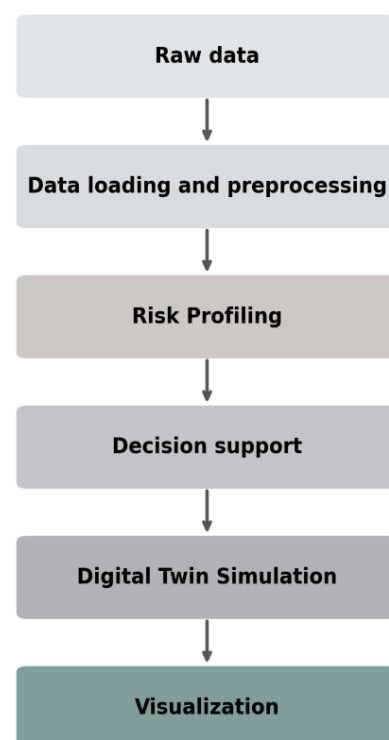


Figure 2. System workflow diagram: sequential process from data collection to visualization.

Table 1. Traffic accident key information.

Variable	Description
accidents	Number of recorded accidents as a positive integer.
traffic fine amount	The traffic fine amount is expressed in thousands of USD.
traffic density	The traffic density index ranges from 0 (low) to 10 (high).
traffic lights	Proportion of traffic lights in the area (0 to 1).
pavement quality	Pavement quality, scale from 0 (very poor) to 5 (excellent).
urban area	Urban area (1) or rural area (0), as an integer.
average speed	Average speed of vehicles in km/h.
rain intensity	Rain intensity, scale from 0 (no rain) to 3 (heavy rain).
vehicle count	Estimated number of vehicles, in thousands, as an integer.
time of day	Time of day in 24 h format (0 to 24).

5.5. Risk Value

In traffic safety management, assessing accident risk is a crucial step. To effectively reduce the incidence of traffic accidents, this study designed a risk value calculation formula. The risk value results from multiple environmental and behavioral factors which are used to quantify the degree of danger of traffic conditions. By calculating the risk value, traffic participants or scenarios can be categorized into low risk, medium risk, and high risk, enabling the implementation of corresponding safety measures.

5.5.1. Risk Value Formula

$$\text{Risk value} = 4 \times (5 - \text{pavement quality}) + 3 \times \text{average speed} + 2 \times \text{rain intensity} + 1 \times \text{vehicle count} \quad (1)$$

The risk value (Risk = risk value in Figure 2) is calculated using a weighted formula. In real life, road quality has a significant impact on traffic safety. A poor road surface may cause a vehicle to lose control, increase braking distance, and greatly increase the risk. Therefore, this paper assigns a higher weight to road quality. The degree of road surface defect is expressed by 5—pavement quality and is calculated and multiplied by a weight of 4 to reflect its significant impact on risk. Rain reduces road friction and affects the driver's field of vision. Driving on rainy days is prone to vehicle skidding and prolonged braking distances, increasing the risk of accidents. Therefore, the weight of rainfall is 2. Driving at high speeds reduces the driver's reaction time and braking efficiency, thereby increasing the risk of accidents, so the weight of 3 assigned to speed indicates that the impact of speed is relatively large. The greater the traffic volume and the shorter the distance between vehicles, the greater the possibility of a collision. A weight of 1 means that the impact of traffic density is relatively small.

5.5.2. Risk Value Classification

- Vehicles are divided into three levels according to the risk value, low risk (yellow), medium risk (blue), and high risk (red), so that targeted safety measures can be taken.
- Based on the calculated risk value, the system divides the vehicle into the following three risk levels:
- Low risk (Risk < 100): This indicates that the current environment is safe and the risk of accidents is low. Conventional traffic management measures may be adopted. The visual effect is represented by yellow dots.

- Medium risk ($100 < \text{Risk} \leq 150$): This indicates that the current environment has certain risks and potential security risks that need attention. It is recommended that road patrols be strengthened and traffic flow management optimized. The visual effect is represented by blue dots.
- High risk ($\text{Risk} > 150$): This indicates that the current environment is highly dangerous and the probability of accidents has increased significantly. Strict traffic control measures such as speed limits, diversions, or temporary road closures should be implemented. The visual effect is represented by red dots.

5.6. Achievement Display

5.6.1. Visual Simulation of Risk Levels

Through traffic accident data analysis, we generated a visual simulation diagram based on different risk levels (supplement, this diagram is captured from the dynamic simulation diagram), as shown in Figure 3. Each point represents a traffic accident scenario, and the color of the point indicates the risk level of the scenario:

Red point: high risk ($\text{Risk} > 150$).

Blue point: medium risk ($100 < \text{Risk} \leq 150$).

Yellow point: low risk ($\text{Risk} \leq 100$).

The distribution of these color points allows us to understand the distribution of traffic accident risks intuitively. At the same time, it allows us to make appropriate suggestions based on the risk value.

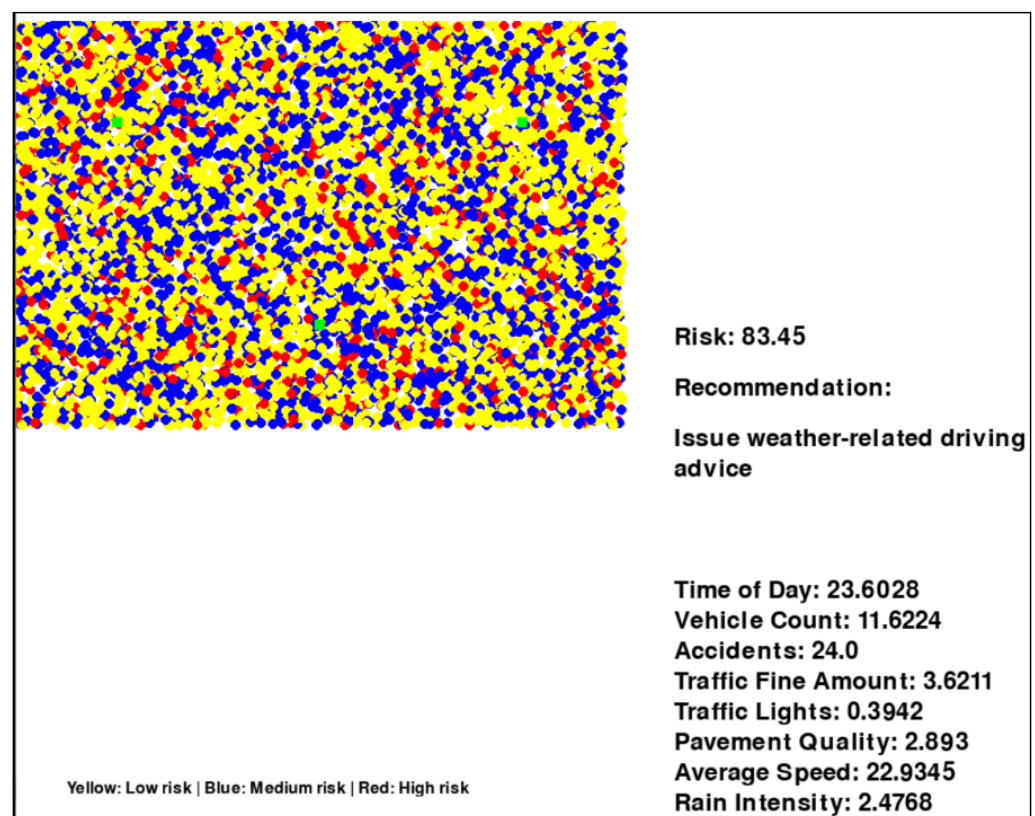


Figure 3. Risk level visualization simulation diagram.

5.6.2. Histogram Analysis of Factors Influencing Accidents

To further study the impact of different factors on traffic accidents, this article draws the following bar chart:

Figure 4 shows the distribution of traffic accidents in different road quality ranges. The bar chart shows the highest number of accidents in areas with road quality intervals of “0–1” and “1–2”. After the road quality exceeds 3, the number of accidents drops significantly. As the road surface quality improves, the number of traffic accidents decreases significantly, indicating that road surface quality is an essential factor affecting traffic accidents.



Figure 4. Pavement quality.

Analysis: Vehicles are more likely to lose control in areas with poor road quality, leading to frequent traffic accidents.

Suggestions:

Infrastructure improvement: Repair and maintain areas with poor road quality to reduce traffic accidents caused by road problems.

Figure 4 shows the distribution of accidents in different periods of the day.

The accidents are highest between 7:00 and 9:00 in the morning (peak hour) and 16:00 and 18:00 (peak hour). The number of accidents at night is relatively small, but there is a slight increase in the late-night period (22:00–24:00).

Analysis: The high number of traffic accidents during peak hours may be related to the heavy traffic volume and driver fatigue. The increase in accidents at night may be related to reduced visibility and decreased driver attention.

Suggestions:

Optimize signal light timing: Adjust the duration of traffic lights during peak hours through intelligent traffic signal control systems to reduce congestion and the risk of accidents.

Peak hour patrols: Increase the frequency of traffic police patrols to handle traffic accidents and divert traffic promptly.

Night protection measures: Add warning lights and reflective signs on late-night sections to improve driver visibility.

Figure 5 shows the distribution of the number of accidents across the time of day.

Suggestions:

Diversion measures: Implement diversion measures in areas with dense traffic, such as guiding vehicles to use different lanes or detours.

Figure 6 shows the distribution of traffic accidents in across different vehicle counts. As the number of vehicles increases, the accident rate also gradually increases, showing that the number of vehicles impacts traffic accidents.

As shown in Figure 6, the number of accidents peaks in the 30–50 km/h speed range. Driving at high speeds reduces the driver’s reaction time, increasing the risk.

Recommendations:

Speed limit optimization: The speed limit standards should be optimized according to the actual conditions of different sections of the road to reduce accidents without affecting traffic efficiency.

Intelligent speed limit system: Deploy an intelligent speed limit system on sections with high accident rates to dynamically adjust the speed limit according to real-time traffic conditions.

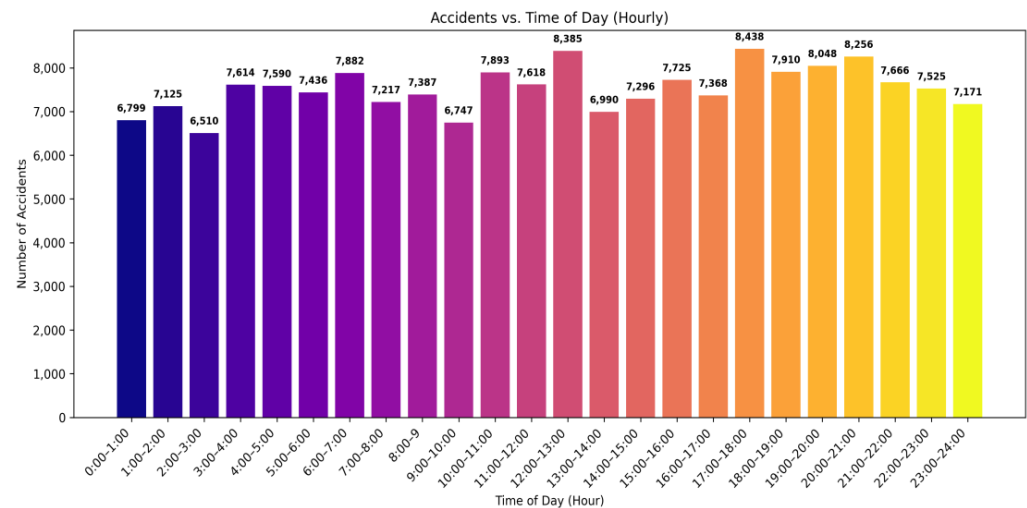


Figure 5. Time of day.

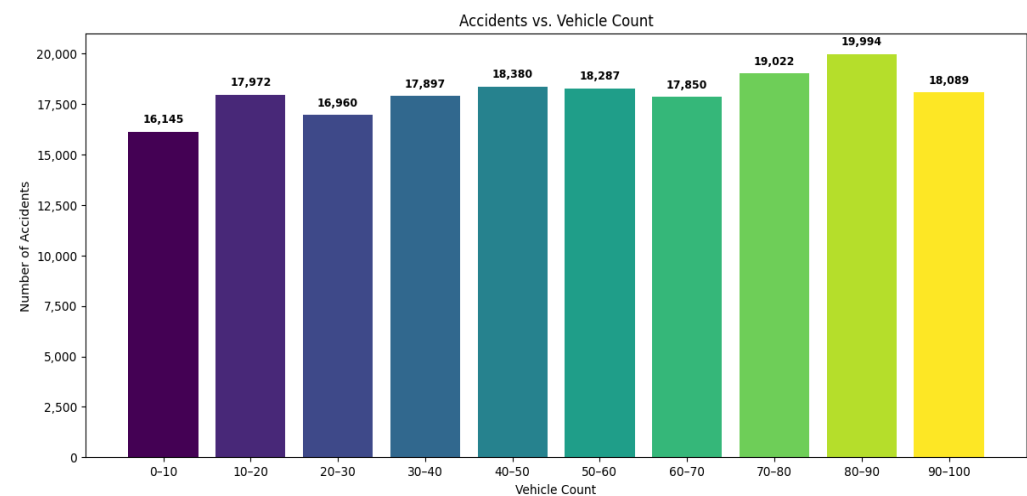


Figure 6. Vehicle count.

Figure 7 shows the average speed (between adjacent accident-prone vehicles) of traffic accidents. From the figure, it can be seen that crashes are most common in the range of 30–50 km/h. This suggests a potential link between moderate driving speeds and increased accident risk, possibly related to average traffic flow or driver overconfidence. Yet for traffic speeds beyond 50 km/h, the number of accidents seems to drop, but this may also be due to less congestion or better enforcement in higher speed areas. This finding will provide a foundation for dynamic speed limits and an intelligent traffic system based on real-time data.

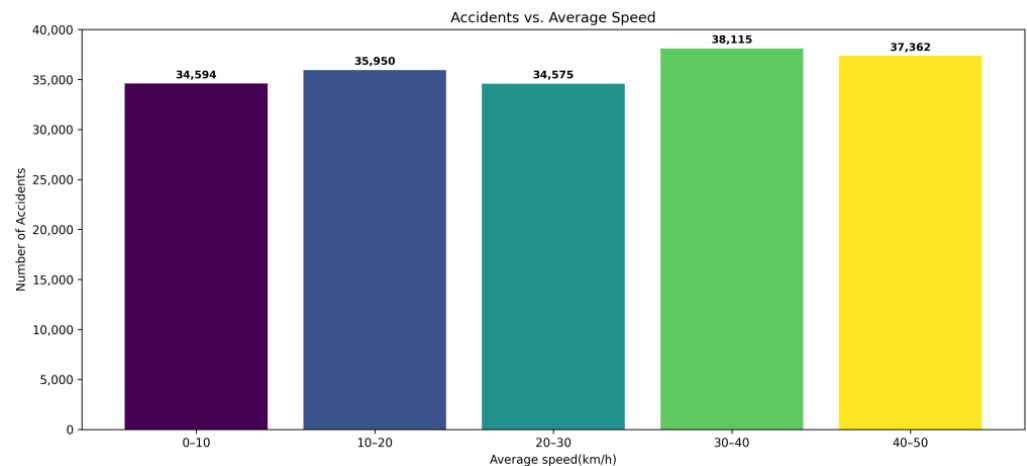


Figure 7. Average speed.

5.7. Results

The proposed integration framework of big data, AI, and digital twins in the metaverse for traffic management was validated through a real-time simulation platform. The key findings are outlined as follows:

1. Risk Visualization Effectiveness

The risk measures derived by the weighted formula provided clear discrimination among low-, medium-, and high-risk scenarios. The yellow, blue, and red color-coded simulation allowed for simple understanding and early intervention planning.

2. Influence of Environmental Factors

Histogram analyses confirmed that low pavement quality, a moderate speed of 30–50 km/h, and high traffic density were the prime factors influencing traffic accident risks. Such results validated the risk formula and matched expectations in the real world.

3. System Performance Metrics

The system handled real-time data from 1000 vehicles under sub-2 s latency because of Spark's in-memory analytics and lightweight visual rendering using Pygame 2.6.1. However, the performance took a hit when scaling beyond 2500 vehicle inputs, suggesting scalability limitations.

5.8. Discussion

The deployment of digital twins and big data technologies in the metaverse presents a revolutionary opportunity for urban traffic management. The case study research results reveal that the framework enhances predictive capabilities and empowers city planners with data-driven, immersive decision aids.

Interpretation of Results:

- Accurately visualizing risk areas improves stakeholders' knowledge, allowing for rapid and informed interventions, particularly in heavily populated cities.
- Quantitative factor analysis of accidents verifies the key infrastructure factors (i.e., road conditions) and environmental factors (i.e., rainfall intensity), providing actionable information to inform urban planning.

Broader Implications:

- Policy-Making: Urban policymakers can utilize such systems to implement adaptive traffic management during high-risk hours (e.g., rush hour or heavy rainfall).

- **Planning for Urban Areas:** Accident-prone locations can be leveraged to spur investments in infrastructure, such as road repair or sign installations.
- **Privacy and Ethics:** While the system's performance is data-dependent, it poses relevant questions on user consent and privacy. This paper discusses how blockchain integration can fix trust and data integrity problems.
- **Scalability:** Promising as it is, the system's performance so far is optimized for small-scale environments. Integration with upcoming edge computing could potentially lead to city-scale deployments.

Limitations:

- The dataset does not contain human behavior data (e.g., driver distraction or attention), which could influence predictive outcomes.
- The simulations lack multi-modal traffic systems (e.g., pedestrians, bicycles, public transport).
- The risk thresholds, as effective as they are, are currently static. Adaptive thresholds with dynamic traffic conditions could optimize responsiveness.

Future Research Directions:

- **Adaptive Learning:** Incorporate reinforcement learning for self-improving traffic interventions.
- **Human-Centered AI:** Integrate driver biometrics and behavior analysis to personalize risk prediction.
- **Cross-Platform Interoperability:** Examine standardization protocols for seamless integration between multiple simulation engines and city-wide IoT systems.
- **Edge Deployment:** Explore hybrid cloud–edge systems for improving latency and scalability.

6. Conclusions

This study explores the transformative potential of integrating big data and digital twin technologies within the metaverse, with a focus on a traffic management case study. The findings highlight how real-time data analytics and virtual simulations can enhance decision-making and operational efficiency in complex systems like urban mobility. While the integration offers significant benefits, it also faces challenges such as data privacy, scalability, and real-time processing limitations. These insights emphasize the need for continued interdisciplinary efforts to refine and expand this framework, ensuring its practical and sustainable application in the Metaverse.

6.1. Summary of Key Points

Data-Driven Optimization: Big data enables precise analysis of user behavior, allowing for tailored enhancements to the metaverse's virtual environments and social systems, as demonstrated by personalized recommendations in the traffic management case.

Intelligent Decision Support: AI-powered analytics, exemplified by tools like random forest models, provide predictive insights that strengthen decision-making processes, such as risk assessment in traffic scenarios.

Digital Twin Integration: Virtual representations of real-world systems, such as traffic flows, improve predictive accuracy and user immersion, bridging the physical and digital realms effectively.

Modular Design: The system's architecture, integrating components like data processing and visualization, showcases seamless collaboration between big data and metaverse technologies.

6.2. Limitations and Future Directions

Data Dependency: The system’s effectiveness hinges on the quality and availability of historical data, which may not fully account for unpredictable real-world factors like human behavior.

Scalability Challenges: Current capabilities are constrained to smaller-scale scenarios (e.g., up to 1000 vehicles), limiting broader urban applications.

Real-Time Constraints: Achieving low-latency processing remains a technical hurdle, impacting the responsiveness of simulations. Future work should incorporate diverse and dynamic data sources, enhance computational efficiency, and explore solutions like edge computing to overcome these barriers.

Author Contributions: Conceptualization, Hemn Barzan Abdalla and Ruoxuan Li; methodology, Hemn Barzan Abdalla and Ruoxuan Li; software, Ruoxuan Li; validation, Hemn Barzan Abdalla, Hamidreza Rabiei-Dastjerdi and Mehdi Gheisari; formal analysis, Ruoxuan Li and Hemn Barzan Abdalla; investigation, Hemn Barzan Abdalla, Hamidreza Rabiei-Dastjerdi, and Mehdi Gheisari; resources, Ruoxuan Li and Hemn Barzan Abdalla; data curation, Ruoxuan Li; writing—original draft preparation, Hemn Barzan Abdalla and Ruoxuan Li; writing—review and editing, Hemn Barzan Abdalla, Ruoxuan Li and Hamidreza Rabiei-Dastjerdi; visualization, Ruoxuan Li and Hemn Barzan Abdalla; supervision, Hemn Barzan Abdalla; project administration, Hemn Barzan Abdalla; funding acquisition, Hemn Barzan Abdalla. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Wenzhou-Kean University Internal Research Support Program (IRSPG202202).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: The authors give the publisher permission to publish the work.

Data Availability Statement: All data are based on references.

Acknowledgments: The authors gratefully acknowledge the financial support from Wenzhou-Kean University.

Conflicts of Interest: The authors declare that they have no competing interests.

References

1. Freyermuth, G.S. Metaverse’s Modern Prehistory. *Gaming Metaverse* **2025**, *21*, 13.
2. WHO. World Health Organization—Road Traffic Injuries. 2023. Available online: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries> (accessed on 11 February 2023).
3. Kirkpatrick, K. Applying the metaverse. *Commun. ACM* **2022**, *65*, 16–18. [CrossRef]
4. Dwivedi, Y.K.; Hughes, L.; Wang, Y.; Alalwan, A.A.; Ahn, S.J.; Balakrishnan, J.; Sergio, B.; Russell, B.; Dimitrios, B.; Vincent, D.; et al. Metaverse marketing: How the metaverse will shape the future of consumer research and practice. *Psychol. Mark.* **2023**, *40*, 750–776. [CrossRef]
5. Tariq, S.; Abuadbbba, A.; Moore, K. Deepfake in the metaverse: Security implications for virtual gaming, meetings, and offices. In Proceedings of the 2nd Workshop on Security Implications of Deepfakes and Cheapfakes, Melbourne, Australia, 10–14 July 2023; pp. 16–19.
6. Kelly, J.W.; Cherep, L.A.; Lim, A.F.; Doty, T.; Gilbert, S.B. Who are virtual reality headset owners? a survey and comparison of headset owners and non-owners. In 2021 IEEE Virtual Reality and 3D User Interfaces (VR). Lisboa, Portugal, 27 March–1 April 2021; pp. 687–694.
7. Sagioglu, S.; Sinanc, D. Big data: A review. In Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS), San Diego, CA, USA, 20–24 May 2013; pp. 42–47.
8. Zhang, H.; Lee, S.; Lu, Y.; Yu, X.; Lu, H. A Survey on Big Data Technologies and Their Applications to the Metaverse: Past, Current and Future. *Mathematics* **2023**, *11*, 96. [CrossRef]
9. Rathore, M.M.; Shah, S.A.; Shukla, D.; Bentafat, E.; Bakiras, S. The role of ai, machine learning, and big data in digital twinning: A systematic literature review, challenges, and opportunities. *IEEE Access* **2021**, *9*, 32030–32052. [CrossRef]

10. Abdalla, H.B. A brief survey on big data: Technologies, terminologies and data-intensive applications. *J. Big Data* **2022**, *9*, 107. [\[CrossRef\]](#)
11. Abdalla, H.B.; Awlla, A.H.; Kumar, Y.; Cheraghy, M. Big Data: Past, Present, and Future Insights. In Proceedings of the 2024 Asia Pacific Conference on Computing Technologies, Communications and Networking, Chengdu, China, 26–27 July 2024; pp. 60–70.
12. Kitchin, R.; McArdle, G. What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data Soc.* **2016**, *3*, 2053951716631130. [\[CrossRef\]](#)
13. Kumar, Y.; Marchena, J.; Awlla, A.H.; Li, J.J.; Abdalla, H.B. The AI-powered evolution of big data. *Appl. Sci.* **2024**, *14*, 10176. [\[CrossRef\]](#)
14. Xu, Y.; Yu, L. Cross-regional Teaching Resource Sharing Solution Based on HADOOP Architecture. In Proceedings of the 2024 International Symposium on Artificial Intelligence for Education (ISAIE '24), Xi'an, China, 6–8 September 2024; Association for Computing Machinery: New York, NY, USA, 2024; pp. 613–620.
15. Rajpurohit, A.M.; Kumar, P.; Kumar, R.R.; Kumar, R. A Review on Apache Spark. In Proceedings of the KILBY 100 7th International Conference on Computing Sciences, Phagwara, Punjab, India, 5 May 2023.
16. Kušić, K.; Schumann, R.; Ivanjko, E. A digital twin in transportation: Real-time synergy of traffic data streams and simulation for virtualizing motorway dynamics. *Adv. Eng. Inform.* **2023**, *55*, 101858. [\[CrossRef\]](#)
17. Pierce, J.S.; Stearns, B.C.; Pausch, R. Voodoo dolls: Seamless interaction at multiple scales in virtual environments. In Proceedings of the 1999 Symposium on Interactive 3D Graphics, Atlanta, GA, USA, 26–29 April 1999; pp. 141–145.
18. Bandara, E.; Shetty, S.; Mukkamala, R.; Liang, X.; Foytik, P.; Ranasinghe, N.; De Zoysa, K. Casper: A blockchain-based system for efficient and secure customer credential verification. *J. Bank. Financ. Technol.* **2022**, *6*, 43–62. [\[CrossRef\]](#)
19. ur Rehman, M.H.; Dirir, A.M.; Salah, K.; Damiani, E.; Svetinovic, D. TrustFed: A framework for fair and trustworthy cross-device federated learning in IIoT. *IEEE Trans. Ind. Inform.* **2021**, *17*, 8485–8494. [\[CrossRef\]](#)
20. Park, H.; Lim, Y. Deep reinforcement learning based resource allocation with radio remote head grouping and vehicle clustering in 5G vehicular networks. *Electronics* **2021**, *10*, 3015. [\[CrossRef\]](#)
21. Kottursamy, K.; Khan, A.U.R.; Sadayappillai, B.; Raja, G. Optimized D-RAN aware data retrieval for 5G information centric networks. *Wirel. Pers. Commun.* **2022**, *124*, 1011–1032. [\[CrossRef\]](#)
22. Jiang, W.; Han, B.; Habibi, M.A.; Schotten, H.D. The road towards 6G: A comprehensive survey. *IEEE Open J. Commun. Soc.* **2021**, *2*, 334–366. [\[CrossRef\]](#)
23. Chkirbene, Z.; Erbad, A.; Hamila, R.; Gouissem, A.; Mohamed, A.; Hamdi, M. Machine learning based cloud computing anomalies detection. *IEEE Netw.* **2020**, *34*, 178–183. [\[CrossRef\]](#)
24. Chraibi, A.; Ben Alla, S.; Ezzati, A. Makespan optimisation in cloudlet scheduling with improved DQN algorithm in cloud computing. *Sci. Program.* **2021**, *2021*, 7216795. [\[CrossRef\]](#)
25. Cao, K.; Liu, Y.; Meng, G.; Sun, Q. An overview on edge computing research. *IEEE Access* **2020**, *8*, 85714–85728. [\[CrossRef\]](#)
26. Rodrigues, T.K.; Liu, J.; Kato, N. Application of cybertwin for offloading in mobile multiaccess edge computing for 6G networks. *IEEE Internet Things J.* **2021**, *8*, 16231–16242. [\[CrossRef\]](#)
27. Gheisari, M.; Khan, W.Z.; Najafabadi, H.E.; McArdle, G.; Rabiei-Dastjerdi, H.; Liu, Y.; Fernández-Campusano, C.; Abdalla, H.B. CAPPAD: A privacy-preservation solution for autonomous vehicles using SDN, differential privacy and data aggregation. *Appl. Intell.* **2024**, *54*, 3417–3428. [\[CrossRef\]](#)
28. Gadekallu, T.R.; Huynh-The, T.; Wang, W.; Yenduri, G.; Ranaweera, P.; Pham, Q.V.; da Costa, D.B.; Liyanage, M. Blockchain for the metaverse: A review. *arXiv* **2022**, arXiv:2203.09738. [\[CrossRef\]](#)
29. Sahray, K.; Sukereman, A.S.; Rosman, S.H.; Jaafar, N.H. The implementation of virtual reality (VR) technology in real estate industry. *Plan. Malays.* **2023**, *21*, 255–265. [\[CrossRef\]](#)
30. Zhang, C.; Sun, Y.; Chen, J.; Lei, J.; Abdul-Mageed, M.; Wang, S.; Jin, R.; Park, S.; Yao, N.; Long, B. Spar: Personalized content-based recommendation via long engagement attention. *arXiv* **2024**, arXiv:2402.10555. [\[CrossRef\]](#)
31. Wang, P.; Bishop, I.D.; Stock, C. Real-time data visualization in Collaborative Virtual Environments for emergency response. In Proceedings of the Spatial Sciences Institute Biennial International Conference, Adelaide, Australia, 28 September–2 October 2009.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.