International Journal on Orange Technologies (IJOT)

e-ISSN: 2615-8140
p-ISSN: 2615-7071

Volume: 8 Issue: 3 | September 2026
https://journals.researchparks.org/index.php/IJOT

An Analysis of The Utilization of Digital Twin and Big Data Analytics in Road Engineering for The Early Prediction of Urban Accident Hotspots

Zulfadhli [1], Defry Basrin [2], Rajib Muammar [3], Syarwan [4]

1. Samudra University, Indonesia
2. Samudra University, Indonesia
3. Samudra University, Indonesia
4. Lhokseumawe State Polytechnic, Indonesia

* Correspondence: zulfadhli@unsam.ac.id

Abstract: This study addresses the critical issue of escalating traffic accident rates in Indonesian urban areas, which recorded 101,904 cases as of September 2024. Current conventional methods for identifying accident-prone areas (blackspots) remain primarily reactive and lack the capability for early risk prediction. Conversely, while Digital Twin and Big Data Analytics technologies offer transformative potential for real-time road modeling, their integration into road engineering for the proactive prediction of accident blackspots in Indonesian urban contexts has not been systematically elaborated. This research employs a qualitative-descriptive library research approach, utilizing data sources that include academic books, scientific journal articles, and research reports. The collected data were analyzed using thematic content analysis techniques. The findings of this study reveal four primary conclusions. First, Digital Twin technology enables a virtual representation of road networks integrated with real-time sensor data, thereby creating a cyber-physical system capable of simulating diverse traffic scenarios. Second, Big Data Analytics possesses the capacity to process vast volumes of heterogeneous data including GPS, telematics, CCTV footage, and social media inputs to identify accident risk patterns that typically elude conventional statistical methods. Third, the integration of Digital Twin and Big Data Analytics facilitates the development of a risk prediction platform capable of proactively identifying accident blackspots with significantly higher accuracy than traditional approaches. Fourth, the implementation of these technologies in Indonesian urban areas encounters substantial challenges, notably limited data infrastructure, insufficient human resource capacity, and inadequate regulatory frameworks. This study concludes that the utilization of Digital Twin and Big Data Analytics in road engineering constitutes a highly promising strategy for the early prediction of urban accident blackspots. The implications of these findings underscore the necessity for targeted investments in data infrastructure, human capital development, and the formulation of regulatory frameworks that support the gradual and context-specific adoption of these technologies.

Keywords: Digital Twin, Big Data Analytics, Road Engineering, Accident Blackspots, Urban Areas.

Citation: Zulfadhli, Basrin, D, Muammar, R &, Syarwan. An Analysis of The Utilization of Digital Twin and Big Data Analytics in Road Engineering for The Early Prediction of Urban Accident Hotspots. International Journal on Orange Technologies (IJOT) 2026, 8(3), 79-92.

Received: 13th Jul 2026
Revised: 28th Aug 2026
Accepted: 05th Sep 2026
Published: 20th Oct 2026

Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

Traffic accidents continue to constitute a primary cause of mortality and morbidity in Indonesia, with data from Statistics Indonesia (BPS) recording 152,008 incident cases throughout 2023 and 101,904 cases up to September 2024. Urban areas account for a substantial proportion of these statistics, exhibiting distinct risk characteristics compared to rural regions, including high traffic density, multimodal interactions, and complex land-use patterns. Conventional methodologies for identifying accident-prone locations, which

predominantly rely on historical accident data and statistical analyses, are inherently reactive; they only identify problematic sites post-occurrence. This limitation underscores an urgent necessity for a more proactive and predictive paradigm to assess accident risks prior to the occurrence of casualties [1], [2].

Recent advancements in digital technology present novel opportunities for the transformation of highway engineering through the application of Digital Twin technology and Big Data Analytics. A Digital Twin serves as a virtual representation of physical assets specifically, roadways and their auxiliary infrastructure integrated with real-time data derived from sensors and monitoring systems. This technological framework facilitates the simulation of diverse traffic scenarios, the execution of what-if analyses, and the dynamic prediction of risks. Concurrently, Big Data Analytics enables the processing of vast and heterogeneous datasets, encompassing Global Positioning System (GPS) data, vehicle telematics, surveillance footage, and crowdsourced information, to uncover underlying patterns and trends that remain imperceptible through conventional analytical approaches. The synergistic integration of these two technologies holds substantial potential for the development of an accident risk prediction system that is markedly more accurate, real-time, and proactive [3], [4].Recent literature has demonstrated the efficacy of Digital Twin technology in enhancing road safety. Research published in Sustainability [5] introduced an artificial intelligence (AI)-driven Digital Twin framework that consolidates heterogeneous data sources comprising static infrastructure mapping, historical traffic logs, telematics, and visual camera feeds into a cohesive cyber-physical system. This platform employs AI algorithms to facilitate data fusion, anomaly detection, and predictive analytics. Furthermore, it enables the generation of dynamic safety heatmaps, congestion predictions, and scenario-based decision-making support. Empirical trials conducted on the Madison Flex Lane corridor evidenced the system's proficiency in processing real-time data, reconstructing traffic incidents, anticipating collision probabilities, and regulating eco-driving behaviors. These results substantiate the potential of Digital Twins to act as the foundational architecture for intelligent road safety management systems [5].

Concurrently, Big Data Analytics has emerged as a robust mechanism for the evaluation of traffic collisions. An investigation featured in Big Data and Cognitive Computing [6] applied Multivariate Adaptive Regression Splines (MARS) to evaluate the primary determinants impacting crash severity within the industrial regions of Thailand. The analysis revealed that specific variables, notably vehicle classification, road surface conditions, and the temporal context of the incident, exert a substantial influence on the degree of accident severity. Utilizing a Big Data Analytics methodology facilitates the recognition of intricate interactions among variables that traditional regression frameworks often fail to detect. Consequently, these findings illustrate that Big Data Analytics yields profound insights into the underlying causes of accidents, thereby facilitating the development of highly targeted and effective accident prevention strategies [6].

The integration of Digital Twin technology and Big Data Analytics within road engineering for the early prediction of accident-prone zones represents a rapidly advancing research frontier. A study published in Scientific Reports [7] investigated the application of Digital Twins in traffic safety analysis by incorporating vehicle dynamics and environmental variables into microscopic traffic simulations. The results demonstrated that Digital Twins, when augmented with real-time data, can significantly enhance the accuracy of accident risk predictions. Nevertheless, the research also highlighted inherent challenges regarding model calibration and the necessity for comprehensive datasets to guarantee predictive validity. These findings underscore that while the potential for integrating Digital Twins and Big Data Analytics is substantial, its practical implementation demands rigorous consideration of technical, methodological, and contextual dimensions [7].

In the context of Indonesia, the adoption of Digital Twin and Big Data Analytics in road engineering remains in its nascent phase. Preliminary initiatives have emerged, such as the deployment of a Smart City framework predicated on a Road Safety Policy Algorithm in Yogyakarta by the Indonesian National Police, which is designed to leverage big data and artificial intelligence to monitor, analyze, and respond to traffic conditions in real time. However, such efforts are currently sporadic and lack integration into a cohesive, systematic road engineering framework. Furthermore, research concerning the implementation of machine learning within the Indonesian transportation sector indicates that technological adoption is persistently constrained by data deficiencies, limited human resource capabilities, and inadequate digital infrastructure. Consequently, comprehensive studies are imperative to ascertain optimal strategies for adapting Digital Twin and Big Data Analytics to the specificities of Indonesian urban environments [8], [9].

The imperative to investigate the application of Digital Twin and Big Data Analytics in road engineering for the early prediction of accident-prone areas has intensified, driven by rapid urbanization and the escalating complexity of urban transportation systems. Conventional blackspot identification methods are no longer sufficient to address the road safety challenges prevalent in major Indonesian cities. These urban areas exhibit unique traffic characteristics, including the predominance of motorcycles, widespread non-compliance with traffic regulations, and infrastructural limitations. Conversely, Digital Twin and Big Data Analytics present a more proactive, predictive, and data-driven paradigm for identifying accident risks and formulating effective interventions. By leveraging these technologies, municipal governments can optimize resource allocation and cultivate a safer road environment for all users [3], [4], [5].

Consequently, this study aims to comprehensively analyze the integration of Digital Twin and Big Data Analytics within road engineering for the early prediction of urban accident blackspots. Through a systematic literature review, this research seeks to establish a profound understanding of the concepts, methodologies, and applications of these technologies, alongside the challenges and opportunities associated with their implementation in Indonesia. This paper posits that the incorporation of Digital Twin and Big Data Analytics in road engineering constitutes a highly promising strategy for enhancing urban road safety in Indonesia, provided it is underpinned by substantial investments in data infrastructure, human capital development, and a robust regulatory framework [6], [7].The fundamental problem addressed in this research stems from the high rate of traffic accidents in Indonesian urban areas, which is inadequately counterbalanced by early risk prediction capabilities. Concurrently, despite their substantial potential, Digital Twin and Big Data Analytics technologies have yet to be systematically elaborated within the context of Indonesian road engineering. The analytical gap lies in the absence of comprehensive studies that integrate Digital Twin, Big Data Analytics, and the prediction of accident-prone areas into a single, coherent conceptual framework. The urgency of this research is predicated on the critical need to generate theoretical insights and practical recommendations that can expedite the adoption of these technologies, thereby enhancing road safety across Indonesian cities. Ultimately, the objective of this study is to analyze the concepts and applications of Digital Twin and Big Data Analytics in road engineering, identify the associated challenges and opportunities for their implementation, and formulate a contextualized integration model for the early prediction of accident hotspots in urban Indonesia.

2. Materials and Methods

This study employs a library research design utilizing a qualitative-descriptive approach, specifically structured to conduct an in-depth examination of the application of Digital Twin and Big Data Analytics in road engineering for the early prediction of urban

accident hotspots. The qualitative-descriptive framework was selected because the primary objective is to develop a comprehensive understanding of the associated concepts, methods, and technological applications, rather than to measure or statistically generalize the phenomena. As a library-based study, the research relies on journal articles, academic books, research reports, and policy documents as its primary data sources. This methodological characteristic aligns with the fundamental nature of library research, which positions the researcher as a critical reader tasked with extracting multidimensional meanings from the available literature [10], [11], [12].

The data sources utilized in this study are classified into two principal categories: primary and secondary data sources. The primary data sources encompass foundational textbooks on Digital Twin, Big Data Analytics, and highway engineering. Specifically, these include the works of Wang et al. [3] on Digital Twin Technologies in Transportation Infrastructure Management, Ardakani and Cheshmehzangi [4] on Big Data Analytics for Smart Transport and Healthcare Systems, and Hardiyatmo [13] on Geosintetik untuk Rekayasa Jalan Raya. Furthermore, the primary data incorporate policy documents, such as Law Number 22 of 2009 concerning Traffic and Road Transportation, as well as land transportation statistical reports issued by Statistics Indonesia (Badan Pusat Statistik). Conversely, the secondary data sources comprise national and international scientific journal articles, research reports, and official publications from institutions including the World Bank, the World Health Organization (WHO), and the Ministry of Transportation, published between the years 2021 and 2025. Additionally, the principal methodological references guiding this research consist of the works authored by Creswell and Creswell [12], Zed [10], and Hasanah [11] [3], [4], [12], [13].

Data collection for this literature review was conducted utilizing the documentation method, a systematic procedure designed to identify, compile, and organize literature sources pertinent to the research objectives. The process commenced with a comprehensive literature search employing predetermined keywords, including "Digital Twin road engineering," "Big Data Analytics road safety," "accident-prone area prediction," "digital twin traffic safety," "big data analytics accident prediction," "prediksi titik rawan kecelakaan Indonesia," "rekayasa jalan digital," and "smart city road safety." This search was executed across multiple academic databases, such as Google Scholar, Garuda (Garba Rujukan Digital), PubMed, ScienceDirect, Springer, Taylor & Francis, and IEEE Xplore, in addition to the official websites of the Central Bureau of Statistics (Badan Pusat Statistik) and the Ministry of Transportation (Kementerian Perhubungan).

Upon identifying potentially relevant sources, a rigorous selection process was implemented based on predefined inclusion criteria. Specifically, the selected literature was required to have been published between 2021 and 2025, demonstrate a direct relevance to the research topic, originate from credible and accredited repositories, and provide data or analytical insights concerning Digital Twins, Big Data Analytics, and accident prediction. This systematic selection procedure culminated in a data corpus comprising journal articles, books, and policy documents that fulfilled the requisite criteria for in-depth analysis [11], [14], [15].

The data analysis technique employed in this study is content analysis utilizing a thematic approach. This analytical process was executed through a series of interconnected stages. The initial phase, data reduction, involved the systematic selection and simplification of relevant information extracted from the collected sources. Following this, the second phase comprised data presentation, wherein the distilled information was structured into a thematic narrative corresponding to the research sub-headings. The final stage entailed verification and conclusion drawing, which required the interpretation and synthesis of the findings to formulate a coherent understanding regarding the application of Digital Twins and Big Data Analytics in road engineering. Throughout these procedures, a concerted effort was made to remain receptive to diverse perspectives and

to mitigate any selection bias that might compromise the integrity of the findings. Ultimately, the thematic approach facilitated the identification of salient patterns within the literature, allowing for the development of a comprehensive synthesis that forms the basis for subsequent policy and practical recommendations [15], [16], [17].

To ensure the validity of the data in this literature review, several strategies were implemented to guarantee the credibility, transferability, dependability, and confirmability of the findings. First, source triangulation was conducted by comparing information across diverse literature including journal articles, books, research reports, and policy documents to identify both consistencies and divergences within the findings. Second, theoretical triangulation was applied by integrating multiple perspectives from road engineering, data science, and traffic safety. This approach enriches the interpretation and mitigates the risk of a singular theoretical bias. Third, an audit trail was established through the transparent documentation of the entire research process, from the initial literature search to the formulation of conclusions, thereby enabling subsequent verification by other scholars. Fourth, peer review was facilitated through discussions with colleagues and experts in civil engineering, transportation, and data science to acquire critical feedback on the interpretations and research findings. Fifth, researcher reflexivity was maintained by consciously acknowledging personal positions and assumptions that might influence the analytical process, actively striving to minimize uncontrolled subjective biases [12], [16], [18]. Consequently, the literature review methodology employed in this study is anticipated to yield a profound and rigorously validated understanding of the utilization of Digital Twins and Big Data Analytics in road engineering, while providing a conceptual contribution to the advancement of road safety policies and practices in Indonesia.

3. Results and Discussion

The Concept and Architecture of Digital Twins in Road Engineering: A Cyber-Physical System Foundation for Traffic Safety

In the domain of road engineering, a Digital Twin (DT) constitutes a virtual representation of physical assets encompassing roads, bridges, tunnels, and ancillary infrastructure integrated with real-time data derived from sensors and monitoring systems. Rather than functioning as a mere static three-dimensional model, a DT operates as a dynamic framework perpetually updated with empirical data, thereby facilitating the continuous simulation, analysis, and prediction of asset behavior under diverse conditions. Within this engineering context, the DT serves as a critical interface between the physical and digital realms, establishing a cyber-physical system (CPS) characterized by a bidirectional flow of data between the tangible asset and its virtual counterpart. The underlying architecture of a DT for road networks typically comprises three primary strata: the physical layer (incorporating sensors, cameras, and Internet of Things devices), the network layer (facilitating data transmission), and the digital layer (encompassing computational models, analytics, and visualization). The synergistic integration of these three layers enables the real-time monitoring of road conditions, the early detection of structural degradation, and the prognostication of maintenance requirements [3], [5].

Furthermore, the architecture of a Digital Twin in road engineering can be systematically categorized according to its maturity and inherent complexity. The initial level, a static DT, exclusively delineates the geometric and attributive properties of the road network devoid of real-time data connectivity. The second level, a dynamic DT, is interfaced with sensor data to facilitate the real-time monitoring of road conditions. The third level, a predictive DT, integrates artificial intelligence models and advanced simulations to forecast future asset behavior, thereby supporting informed decision-making processes. The fourth and most advanced level, an autonomous DT, possesses the

capability to execute independent interventions based on predictive analytics and machine learning. At present, the majority of DT implementations within road infrastructure are situated at the second or third maturity levels, with contemporary research progressively advancing toward the realization of the fourth level. Specifically, within the context of forecasting accident hotspots, the third and fourth tiers of DT exhibit paramount relevance owing to their capacity for the proactive identification of risks [3], [5], [7].

The fundamental components of a Digital Twin for roadways comprise geometric, semantic, and dynamic models. The geometric model delineates the physical configuration and dimensions of the road, encompassing horizontal and vertical alignments, lane widths, and intersection topographies. The semantic model incorporates supplementary data concerning road attributes, such as surface classifications, pavement conditions, road markings, traffic signage, and safety infrastructure. Furthermore, the dynamic model captures the temporal behavior of both the roadway and its traffic, integrating variables such as traffic volume, velocity, density, and movement trajectories. The synthesis of these three models empowers the Digital Twin to simulate diverse traffic scenarios and assess the implications of infrastructural or policy modifications on road safety. Specifically, in the context of forecasting accident-prone locations, the dynamic model is of paramount importance, as it facilitates the identification of traffic conditions that potentially elevate collision risks [3], [4], [5].

The technological framework underpinning roadway Digital Twins encompasses the Internet of Things (IoT), Building Information Modeling (BIM), Geographic Information Systems (GIS), and Artificial Intelligence (AI). IoT devices, which include speed, volume, and weather sensors, alongside surveillance cameras, supply real-time data regarding roadway environments and traffic dynamics. BIM offers a comprehensive, information-dense digital representation of road assets, detailing their geometry, material composition, and functional characteristics. GIS establishes a spatial infrastructure essential for assimilating heterogeneous data sources and analyzing the spatial distribution of traffic accidents. Moreover, AI specifically machine learning and deep learning methodologies is employed to process data, discern underlying patterns, and generate predictive insights. The convergence of these technologies enables the Digital Twin to operate as a cohesive platform for the continuous monitoring, analysis, and prediction of road safety outcomes [3], [5].

Within the domain of road engineering, Digital Twin technology presents several primary advantages for enhancing traffic safety. Firstly, it facilitates the real-time monitoring of road conditions, encompassing the early detection of pavement degradation, the deterioration of road markings, and the impairment of safety infrastructure. Such early detection mechanisms permit more expedited and precise maintenance interventions, thereby mitigating the risk of accidents precipitated by substandard road conditions. Secondly, Digital Twins afford the capability to simulate diverse traffic scenarios, including modifications to road geometry, the integration of supplementary safety facilities, and alterations in traffic policies. These simulations allow for the empirical evaluation of intervention impacts prior to their real-world execution, consequently minimizing both risks and financial expenditures. Thirdly, the technology supports the prognostication of accident risks through the comprehensive analysis of historical and real-time data, thus enabling the preemptive identification of accident-prone zones. Fourthly, Digital Twins foster enhanced collaboration among diverse stakeholders comprising road engineers, transportation planners, law enforcement, and the general public via a unified platform that ensures symmetrical access to critical information [3], [5], [7].

Conversely, the deployment of Digital Twin technology within road infrastructure is impeded by a multitude of challenges. Technical constraints encompass the requisite for voluminous, high-fidelity data, the integration of heterogeneous data sources, and the

formulation of rigorously accurate and validated models. From an organizational perspective, obstacles include the necessity for inter-agency coordination, the enhancement of human resource capacities, and the cultivation of an organizational culture oriented toward data-driven decision-making. Financial impediments involve substantial initial capital investments requisite for sensor deployment, network infrastructure, and software acquisition, compounded by recurrent operational and maintenance expenditures. Furthermore, regulatory challenges persist due to the current absence of explicit standards and guidelines governing Digital Twin implementation in road infrastructure, alongside pertinent concerns regarding data privacy and security. Ameliorating these multifaceted challenges necessitates a holistic, phased methodology that actively engages relevant stakeholders while meticulously accounting for specific local contexts [3], [4], [5].

Consequently, it can be concluded that the Digital Twin serves as a crucial foundation for transforming road engineering toward a more proactive and predictive approach to traffic safety. The architecture of a Digital Twin comprising geometric, semantic, and dynamic models, and supported by IoT, BIM, GIS, and AI technologies facilitates real-time monitoring, scenario simulation, and accident risk prediction. Although its implementation encounters technical, organizational, financial, and regulatory challenges, the potential benefits, including accident reduction, maintenance efficiency, and enhanced collaboration, position the Digital Twin as a strategic investment for urban road safety. In Indonesia, the adoption of Digital Twins for road infrastructure remains in its nascent stages; however, developmental momentum has emerged in tandem with smart city initiatives and the digital transformation of the transportation sector [3], [5], [7].

Big Data Analytics for Accident Risk Modeling and Prediction: Unveiling Hidden Patterns in Heterogeneous Data

Big Data Analytics (BDA) encompasses the processes of collecting, storing, processing, and analyzing high-volume, high-velocity, and diverse datasets to generate actionable insights. Within the context of road safety, BDA facilitates the integration and analysis of data from multifaceted sources including historical accident records, real-time traffic conditions, meteorological data, road geometry, and driver behavior to identify risk factors, forecast accident occurrences, and inform preventive strategies. The primary characteristics that distinguish BDA from conventional statistical analyses include its capacity to process unstructured data (such as text, images, and video), discern non-linear patterns and complex interactions, and dynamically update models as novel data is acquired [4], [6], [7].

The principal data sources utilized for Big Data Analytics (BDA) in the domain of road safety encompass police collision reports, traffic metrics derived from sensors and cameras, vehicular telematics, GPS coordinates from navigation applications, meteorological data, topographical road geometry from surveys and mapping systems, alongside crowdsourced information extracted from social media and community reporting platforms. The amalgamation of these heterogeneous data streams facilitates a highly comprehensive analysis of the multifaceted variables contributing to traffic incidents. For instance, the synthesis of collision records with meteorological data can elucidate the impact of atmospheric conditions on accident probabilities at specific geographic coordinates. Similarly, integrating traffic flow metrics with road geometry parameters can delineate the influence of structural road characteristics on driver behavior and subsequent accident risks. Furthermore, the convergence of telematics and GPS data provides critical insights into how individual driving behaviors and routing choices correlate with the likelihood of vehicular collisions [4], [5], [6].

Methodologically, BDA in road safety employs a spectrum of analytical techniques, including descriptive statistics, regression analysis, machine learning, deep learning, and

network analysis. Descriptive statistics serve to synthesize and graphically represent fundamental data patterns, such as the spatiotemporal and typological distributions of traffic accidents. Regression models are applied to ascertain the correlations between independent variables (e.g., environmental and infrastructural conditions) and dependent variables (e.g., collision severity). Machine learning algorithms notably decision trees, random forests, support vector machines, and gradient boosting are implemented to construct predictive models capable of stratifying accident risk levels or forecasting potential incidents derived from historical datasets. Deep learning architectures, encompassing artificial neural networks and recurrent neural networks, are particularly leveraged for the processing of sequential data, such as temporal traffic series and vehicular surveillance footage. Finally, network analysis is instrumental in deciphering the complex interdependencies among diverse risk factors, thereby isolating the causal trajectories that precipitate traffic accidents [6], [7], [19].

Recent studies have demonstrated the efficacy of Big Data Analytics (BDA) in predicting accident risks. Research published in Sustainability [19] utilized machine learning techniques to forecast traffic collisions by analyzing historical data from diverse sources. The findings indicated that ensemble models, which integrate multiple machine learning algorithms, achieved higher predictive accuracy than individual models. The most critical variables for these predictions included the time of occurrence, meteorological conditions, road geometry, and traffic volume. Another study featured in IEEE Xplore [20] leveraged data from vehicles equipped with Advanced Driver Assistance Systems (ADAS) to identify accident-prone zones. By employing Kernel Density Estimation (KDE), the research successfully detected high-risk areas previously unidentified by conventional methods, thereby highlighting the potential of BDA to uncover latent risks [6], [19], [20].

Furthermore, BDA facilitates real-time analysis for accident risk prediction. By integrating data from roadside sensors and surveillance cameras, BDA systems can continuously monitor traffic conditions and identify potentially hazardous situations, such as anomalous congestion, excessive speeds, or aggressive driving behaviors. These systems are capable of transmitting early warnings to drivers via navigation applications or variable message signs, as well as alerting traffic operators to initiate interventions, such as adjusting traffic signal timings or deploying personnel. This real-time analytical capability represents a significant advantage of BDA over traditional, retrospective analytical methodologies [4], [5], [7].

Within the Indonesian context, the adoption of Big Data Analytics (BDA) for road safety continues to encounter several multifaceted challenges. Primary obstacles include inadequate data infrastructure, characterized by a scarcity of road-mounted sensors and surveillance cameras, alongside insufficient data storage and processing capabilities. A further impediment is the institutional fragmentation of data; accident records are managed by the police department, traffic data by the transportation agency, and road infrastructure data by the public works department, thereby complicating data integration efforts. Furthermore, there is a significant deficit in human resource capacity, particularly a shortage of data science and analytics experts who possess a comprehensive understanding of the transportation sector. Nevertheless, several emerging initiatives have demonstrated considerable potential. These include the utilization of crowdsourced data from navigation applications to identify accident-prone locations, as well as the leveraging of social media platforms for the early detection of traffic incidents. Provided there is appropriate investment in infrastructure, human capital, and regulatory frameworks, BDA can serve as a highly effective instrument for enhancing road safety in Indonesia [4], [8], [9].

Consequently, it can be deduced that Big Data Analytics offers a transformative approach to the modeling and prediction of traffic accident risks. Its capacity to integrate heterogeneous datasets, discern complex patterns, and perform real-time analysis

facilitates a more accurate and proactive identification of accident hotspots compared to conventional methodologies. Although its implementation in Indonesia is currently hindered by infrastructural, data-related, and human resource constraints, the potential benefits ranging from accident reduction and optimized resource allocation to comprehensive improvements in road safety render BDA a strategic investment for urban traffic management. Furthermore, the integration of BDA with Digital Twin technology will significantly augment predictive capabilities, thereby enabling a more holistic approach to road safety management [4], [5], [6], [20].

The Integration of Digital Twin and Big Data Analytics for the Early Prediction of Accident-Prone Areas

The convergence of Digital Twin (DT) and Big Data Analytics (BDA) establishes a robust synergy for the early forecasting of accident hotspots within urban environments. The DT framework offers a virtual paradigm that dynamically represents roadway infrastructure and its surrounding environment in real time, whereas BDA facilitates the processing and analysis of the massive and heterogeneous data streams channeled into this platform. The amalgamation of these technologies enables the development of a risk prediction mechanism that transcends reactive responses to historical incidents, operating proactively to discern latent hazards prior to the occurrence of collisions. Consequently, this system synthesizes historical accident records, real-time traffic metrics, meteorological conditions, road geometric parameters, and driver behavioral patterns to formulate precise, actionable risk assessments [3], [5], [7].

The architectural framework of the DT-BDA integration for forecasting accident-prone zones can be delineated into five hierarchical layers. The foundational data layer involves the acquisition of data from diverse inputs, including Internet of Things (IoT) sensors, closed-circuit television (CCTV) cameras, vehicular Global Positioning Systems (GPS), meteorological stations, and historical accident repositories. Subsequently, the data integration layer executes the preprocessing, transformation, and consolidation of these disparate data sources into a standardized format. The third stratum, the Digital Twin layer, encompasses the virtual replication of the road network, which is continuously updated in real time utilizing inputs from the preceding integration layer. The fourth stratum, the analytics layer, incorporates machine learning and deep learning algorithms trained on both historical and concurrent data to evaluate and predict collision probabilities. Finally, the visualization and interface layer provides interactive dashboards and spatial maps to communicate risk prognoses to end-users. The cohesive integration of these five layers empowers the system to preemptively identify accident-prone locations and formulate evidence-based recommendations for strategic intervention [3], [5], [7].

Recent literature has demonstrated the efficacy of integrating Digital Twin and Big Data Analytics (DT-BDA) for the prediction of traffic accident risks. A study published in Sustainability [5] developed an artificial intelligence (AI)-driven Digital Twin framework that consolidates multi-source data including static infrastructure maps, historical traffic logs, telematics, and surveillance camera feeds into a unified cyber-physical system. This platform employs AI models to execute data fusion, anomaly detection, and predictive analytics, while concurrently supporting dynamic safety heatmaps, congestion forecasting, and scenario-based decision-making. Empirical trials conducted on the Madison's Flex Lane corridor evidenced the platform's proficiency in processing real-time data, reconstructing traffic incidents, forecasting collision probabilities, and regulating eco-driving protocols. These findings substantiate that the convergence of DT and BDA can yield highly accurate, real-time risk prediction systems [5], [7].

Furthermore, research featured in Scientific Reports [7] investigated the application of Digital Twins in traffic safety assessments by embedding vehicular dynamics and environmental variables within microscopic traffic simulations. The results indicated that augmenting Digital Twins with real-time data substantially elevates the precision of

accident risk predictions. The developed model demonstrates the capacity to simulate diverse operational scenarios encompassing fluctuations in traffic volume, meteorological conditions, and roadway geometry to evaluate their subsequent impacts on accident susceptibility. Ultimately, these insights corroborate that DT-BDA integration facilitates comprehensive what-if analyses and delivers superior risk prediction accuracy when compared to conventional methodologies [3], [7].

The integration of Digital Twin and Big Data Analytics (DT-BDA) facilitates the proactive identification of accident-prone areas by analyzing risk patterns that emerge from data. For instance, the system can pinpoint road segments where the convergence of high velocities, traffic density, and adverse weather conditions engenders a substantial risk of accidents, even if no prior collisions have been recorded at those specific locations. Additionally, the system is capable of detecting transient accident hotspots, such as construction zones or special event venues, thereby issuing early warnings to road users. This capacity for proactive prediction constitutes a fundamental advantage of DT-BDA integration over conventional methodologies, which predominantly rely on historical data [3], [5], [7].

Within the context of Indonesia, the integration of DT-BDA for the prediction of accident hotspots remains in its nascent stages; however, several initiatives have demonstrated considerable potential. The inauguration of a Smart City initiative predicated on the Road Safety Policy Algorithm in Yogyakarta by the Indonesian National Police exemplifies the preliminary utilization of big data and artificial intelligence (AI) for road safety. This initiative is engineered to harness big data and AI to monitor, analyze, and respond to traffic conditions in real time. Nevertheless, such endeavors remain sporadic and have yet to be incorporated into a comprehensive Digital Twin framework. Consequently, more systematic efforts are imperative to cultivate an integrated DT-BDA platform, which necessitates the engagement of diverse stakeholders and the exploitation of heterogeneous data sources [4], [8], [9].

The primary challenges associated with the integration of Digital Twin (DT) and Big Data Analytics (BDA) for predicting accident-prone areas in Indonesia encompass inadequate data infrastructure such as sensors, cameras, and network connectivity institutional data fragmentation, a deficit in human resource capacity within the fields of data science and artificial intelligence (AI), and insufficient regulatory frameworks governing data management and privacy. Furthermore, model validation presents a significant hurdle, as predictive models trained on specific contextual data may lack generalizability when applied to disparate environments. Consequently, a contextual and adaptive methodology is imperative, one that accounts for localized variables including the high prevalence of motorcycles, distinct driving behaviors, and existing infrastructure conditions. By systematically mitigating these impediments, DT-BDA integration can serve as a robust framework for enhancing urban road safety across Indonesia [5], [7], [8], [9].

In conclusion, the integration of Digital Twin and Big Data Analytics constitutes a transformative paradigm for the early prediction of urban traffic accident hotspots. The synergy between real-time virtual platforms and extensive data analytics facilitates a mechanism for risk identification that is substantially more accurate, proactive, and actionable compared to conventional methodologies. Although its implementation in Indonesia is constrained by infrastructural, data-related, personnel, and regulatory challenges, the prospective advantages namely the mitigation of traffic collisions, the optimization of resource allocation, and the comprehensive advancement of overall road safety render DT-BDA integration a highly strategic investment for urban environments. Through a phased, context-specific, and collaborative framework, Indonesia can effectively harness these advanced technologies to cultivate a safer road ecosystem for all users [3], [4], [5], [7].

Challenges and Implementation Strategies of Digital Twin and Big Data Analytics in Urban Road Engineering in Indonesia

The implementation of Digital Twin (DT) and Big Data Analytics (BDA) in road engineering for the early prediction of accident-prone hotspots in Indonesian urban areas encounters multifaceted challenges that necessitate systematic resolution. These impediments encompass aspects of data infrastructure, human capital capacity, regulatory frameworks, financial constraints, and organizational culture. Addressing these obstacles requires a holistic, phased, and contextualized strategy that engages a diverse array of stakeholders, including central and local governments, academia, industry, and the general public. Furthermore, implementation strategies must account for the heterogeneous baseline conditions across various Indonesian municipalities, alongside the imperative to ensure the sustainability and scalability of the proposed solutions [3], [8], [9].

A primary challenge lies in the inadequacy of existing data infrastructure. The effective operationalization of DT and BDA heavily relies on the availability of massive, high-fidelity, and real-time datasets. In numerous Indonesian cities, the requisite data collection infrastructure such as traffic sensors, closed-circuit television (CCTV) cameras, and meteorological stations remains sparse and unevenly distributed. Existing data is frequently fragmented across various institutions, characterized by inconsistent formatting and restricted accessibility. Additionally, data storage and processing systems are often insufficient to manage substantial data volumes. Strategic interventions to mitigate these infrastructure deficits include targeted investments in sensor and network technologies, the development of unified data platforms capable of integrating disparate sources, and the utilization of crowdsourced and open-source data to augment existing repositories. Furthermore, governmental bodies should foster public-private partnerships to advance data infrastructure development and establish standardized data protocols to facilitate seamless integration [4], [5], [8].

The second challenge pertains to human resource capacity. The implementation of Digital Twins (DT) and Big Data Analytics (BDA) necessitates specialized expertise in data science, machine learning, artificial intelligence (AI), and software engineering, competencies that are currently scarce in Indonesia. This deficiency in human capital is evident not only at the technical echelon but also across managerial and policy-making levels. Strategic interventions to mitigate this limitation encompass the establishment of targeted training and certification programs to augment the existing workforce, the alignment of higher education curricula with contemporary industrial demands, and the recruitment of international talent. Furthermore, synergistic collaborations involving academia, industry, and governmental bodies can significantly facilitate knowledge transfer and capacity building. The cultivation of communities of practice and knowledge-sharing platforms can also expedite the processes of learning and technological adoption [3], [8], [9].

The third challenge involves the insufficiency of the existing regulatory framework. At present, Indonesia lacks specific regulations governing the application of DT and BDA within the context of road infrastructure. Critical issues, including data governance, privacy, cybersecurity, and the legal liabilities associated with AI utilization, remain inadequately addressed. Strategies to resolve these regulatory gaps necessitate the formulation of a comprehensive legal framework that incorporates data standards, security protocols, and ethical guidelines. Furthermore, such regulations must clearly delineate the distribution of responsibilities between central and regional governments, as well as between the public and private sectors. The policy development process should incorporate public consultations and integrate international best practices, while judiciously adapting them to the local socio-political context [3], [5], [7].

The fourth challenge pertains to funding. The initial capital investment required for the implementation of Digital Twins (DT) and Big Data Analytics (BDA) encompassing

sensors, network infrastructure, software, and personnel training is substantial, whereas local government budgets are frequently constrained. Strategies to mitigate this financial barrier include the development of innovative financing models, such as public-private partnerships, infrastructure bonds, and dedicated road safety funds. Furthermore, the central government can provide fiscal incentives and technical assistance to municipalities that demonstrate a commitment to adopting these technologies. Additionally, a phased approach initiating with pilot projects in selected cities prior to gradual expansion can reduce risks and alleviate initial costs, thereby facilitating empirical learning and system refinement preceding full-scale deployment [4], [8], [9].

The fifth challenge involves organizational culture. The deployment of DT and BDA necessitates a paradigm shift from intuition-based to data-driven decision-making, alongside a requirement for enhanced cross-sectoral collaboration. Within numerous governmental entities, data is frequently perceived as a proprietary asset to be guarded rather than a resource to be shared, and decision-making processes remain heavily influenced by hierarchical structures and subjective experience. Strategies to overcome these institutional barriers encompass visionary leadership, the cultivation of a data-centric culture through professional training and incentive mechanisms, and the establishment of specialized units dedicated to data analytics. Cross-sectoral collaboration can be further facilitated through the implementation of shared digital platforms, standardized data-sharing protocols, and regular coordination forums. Ultimately, such cultural transformation requires considerable time and a sustained, long-term commitment from organizational leadership [3], [5], [8].

Within a broader context, the strategic implementation of Digital Twins (DT) and Big Data Analytics (BDA) for road safety in Indonesia necessitates the consideration of unique local characteristics. These encompass the high prevalence of motorcycles, heterogeneous driver behaviors, and disparate infrastructural conditions. Predictive models developed in foreign contexts may not be directly transferable to Indonesia without rigorous calibration and contextual adaptation. Consequently, localized research and development initiatives engaging academia, industry practitioners, and policymakers are imperative to formulate contextually relevant and empirically valid models. Furthermore, it is critical to ensure that these technological advancements are not exclusively advantageous to major metropolitan areas but are also accessible to smaller municipalities and rural regions, taking into account inherent constraints in resources and infrastructure [7], [8], [9].

Consequently, it can be deduced that the integration of DT and BDA within road engineering for the early prediction of accident blackspots in Indonesian urban environments encounters complex and multidimensional challenges. Nevertheless, these impediments can be mitigated through a holistic, phased, and context-specific strategy that engages diverse stakeholders and leverages public-private partnerships. Strategic investments in data infrastructure, human capital, regulatory frameworks, financial mechanisms, and organizational culture serve as fundamental prerequisites for successful implementation. Driven by robust commitment and collaborative efforts, Indonesia can harness the potential of DT and BDA to cultivate a safer, more efficient, and sustainable urban road environment. Ultimately, these technologies do not represent an end in themselves, but rather function as instrumental tools to achieve an overarching objective: preserving human lives and enhancing the overall quality of life for citizens [3], [5], [8], [9].

4. Conclusion

This study concludes that the utilization of Digital Twin technology and Big Data Analytics in road engineering for the early prediction of urban traffic accident hotspots represents a transformative and highly promising paradigm. Digital Twin technology furnishes a real-time virtual platform that accurately replicates road networks and their surrounding environments, thereby facilitating the comprehensive monitoring,

simulation, and forecasting of traffic dynamics. Concurrently, Big Data Analytics provides the robust capability to process and analyze vast and heterogeneous datasets, elucidating underlying risk patterns that frequently elude conventional methodologies. The amalgamation of these two technologies engenders a profound synergy, enabling the proactive identification of accident-prone zones with significantly greater precision than traditional techniques. Contemporary studies have consistently demonstrated the efficacy of this integration in enhancing road safety, evidenced by its capacity to forecast risks in real-time, delineate obscured accident hotspots, and substantiate scenario-based decision-making processes. Nevertheless, the implementation of these advanced technologies within Indonesian urban contexts encounters substantial challenges. These impediments primarily encompass deficiencies in data infrastructure, limited human resource capacities, inadequate regulatory frameworks, constrained financial allocations, and entrenched organizational cultures.

The implementation of these research findings necessitates a collective commitment from all stakeholders to address the associated challenges systematically and progressively. Both central and regional governments must invest in critical data infrastructure including sensors, surveillance cameras, communication networks, and data processing platforms while simultaneously establishing regulatory frameworks that govern data management, privacy, and cybersecurity. Furthermore, the Ministry of Public Works and Housing, the Ministry of Transportation, and the National Police Traffic Corps must collaborate to integrate heterogeneous data sources and develop a unified Digital Twin platform. Higher education and research institutions are required to foster localized research and development initiatives to generate predictive models that are contextually relevant and empirically valid for Indonesian conditions. The technology sector should act as a strategic partner by providing cost-effective hardware and software solutions tailored to these specific operational requirements. Additionally, the public must be actively engaged as both end-users and data contributors, with a strict emphasis on privacy protection and inclusive participation. Through these collaborative endeavors, Digital Twin technology and Big Data Analytics can serve as highly effective instruments for enhancing urban road safety in Indonesia, significantly reducing traffic accident rates, and ultimately preserving human lives.

REFERENCES

[1] Badan Pusat Statistik, *Statistik Transportasi Darat 2023*. BPS, 2024.

[2] World Bank, *Road Safety in Indonesia: Challenges and Opportunities*. World Bank, 2023.

[3] W. Wang, Q. Zaheer, S. Qiu, W. Wang, C. Ai, J. Wang, S. Wang, and W. Hu, *Digital Twin Technologies in Transportation Infrastructure Management*. Springer, 2024, doi: 10.1007/978-981-99-5804-7.

[4] S. P. Ardakani and A. Cheshmehzangi, *Big Data Analytics for Smart Transport and Healthcare Systems*. Springer, 2023, doi: 10.1007/978-981-99-5920-3.

[5] K. Long, C. Ma, H. Li, Z. Li, H. Huang, H. Shi, Z. Huang, Z. Sheng, L. Shi, P. Li, and S. Chen, "AI-enabled digital twin framework for safe and sustainable intelligent transportation," *Sustainability*, vol. 17, no. 10, Art. no. 4391, 2025, doi: 10.3390/su17104391.

[6] M. Seefong, P. Wisutwattanasak, C. Se, K. Theerathitichaipa, S. Jomnonkwao, T. Champahom, V. Ratanavaraha, and R. Kasemsri, "Big data analytics with the multivariate adaptive regression splines to analyze key factors influencing accident severity in industrial zones of Thailand: A study on truck and non-truck collisions," *Big Data and Cognitive Computing*, vol. 7, no. 3, Art. no. 156, 2023, doi: 10.3390/bdcc7030156.

[7] G. Xu, J. Chen, Z. Wang, A. Zhou, M. Schrader, J. Bittle, and Y. Shao, "Enhancing traffic safety analysis with digital twin technology: Integrating vehicle dynamics and environmental factors into microscopic traffic simulation," *Scientific Reports*, vol. 15, Art. no. 44404, 2025, doi: 10.1038/s41598-025-31930-1.

[8] Garuda: Garba Rujukan Digital, "The transportation industry in Indonesia has experienced rapid development, driven by technological advancements, particularly in the application of machine learning," 2025.

[9] BacaJogja, "Yogyakarta pelopor smart city berbasis algoritma untuk menekan kecelakaan," Aug. 15, 2024.

[10] M. Zed, *Metode Penelitian Kepustakaan*. Yayasan Obor Indonesia, 2025.
[11] N. Hasanah, *Metode Penelitian Kepustakaan: Konsep, Teori, & Desain Penelitian*. PT Literasi Nusantara Abadi Grup, 2023.
[12] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 6th ed. SAGE Publications, 2023.
[13] H. C. Hardiyatmo, *Geosintetik untuk Rekayasa Jalan Raya: Perancangan dan Aplikasi*, 3rd ed. Gadjah Mada University Press, 2023.
[14] Booth, A. Sutton, M. Clowes, and M. Martyn-St James, *Systematic Approaches to a Successful Literature Review*, 3rd ed. SAGE Publications, 2022.
[15] M. L. Pan and J. S. Craven, *Preparing Literature Reviews: Qualitative and Quantitative Approaches*, 6th ed. Routledge, 2024.
[16] U. Flick, *An Introduction to Qualitative Research*, 7th ed. SAGE Publications, 2023.
[17] H. Hasan et al., *Metode Penelitian Kualitatif*. Yayasan Tri Edukasi Ilmiah, 2025.
[18] S. J. Tracy, *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact*, 3rd ed. Wiley, 2024.
[19] S. P. Ardakani, X. Liang, K. T. Mengistu, R. S. So, X. Wei, B. He, and A. Cheshmehzangi, "Road car accident prediction using a machine-learning-enabled data analysis," *Sustainability*, vol. 15, no. 7, Art. no. 5939, 2023, doi: 10.3390/su15075939.
[20] R. S. Mishra, D. S. Thakur, A. Subramanian, M. Advani, S. Velmurugan, J. Jose, C. V. Jawahar, and R. K. Sarvadevabhatla, "Enhancing road safety: Predictive modeling of accident-prone zones with ADAS-equipped vehicle fleet data," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, Jeju Island, Republic of Korea, Jun. 2024, pp. 2583–2590, doi: 10.1109/IV55156.2024.10588591.