Analysis of Digital Maturity Components in Rail Transportation Industry: A Text Mining Approach

Samaneh Moradi[1], Mehrdad Hosseini Shakib[2,*], Ali Badizadeh [3]

Abstract

The assessment of digital maturity in rail transportation, a crucial infrastructure for sustainable development, has gained increasing importance. This research employs text mining to systematically analyze scientific literature exploring digital maturity components in the rail industry. 87 scientific articles published between 2016 and 2025 from reputable international databases were analyzed using advanced text mining techniques, including sentiment analysis, K-means clustering, LDA topic modeling, and n-gram algorithm in Python. Results identify five key areas: emerging technologies, digitalization challenges, cybersecurity, practical applications, and environmental sustainability as main transformation axes. Sentiment analysis reveals predominantly positive attitudes (65%) towards this transformation, with 15% negative and 20% neutral responses. The findings recommend rail industry leaders adopt an integrated approach to digital transformation management, emphasizing cybersecurity infrastructure, environmental sustainability, and digital skills development.

Keywords: digital maturity, rail transportation industry, text mining, railway 4.0, maturity assessment models

---

* Corresponding author. E-mail: Mehrdad.shakib@kiau.ac.ir

[1] Department of Industrial Management, Qa.C., Islamic Azad University, Qazvin, Iran

[2] Department of Industrial Management , Ka.C. , Islamic Azad University, Karaj, Iran

[3] Department of Industrial Management, Qa.C. , Islamic Azad University, Qazvin, Iran

Samaneh Moradi, Mehrdad Hosseini Shakib, Ali Badizadeh

1. Introduction

The rail transportation industry, as one of the main pillars of countries' economic and social development, has faced extensive transformations in recent decades. This strategic industry, which plays a vital role in passenger and cargo transportation, is now at the threshold of a fundamental paradigm shift (Ale F). The fourth industrial revolution and the emergence of new digital technologies have influenced all aspects of this industry, doubling the necessity to revise traditional models of management and operation of rail systems (Zunder Th, 2023).

The need for new technologies is essential as one of the factors for achieving innovation and its benefits (Khamseh, A. and Sadeghi, Y. 2018).Digital transformation in the rail industry, known as "Railway 4.0," extends beyond mere technological change. This transformation has created a profound shift in the design, construction, operation, and maintenance of rail systems, opening new horizons in safety, efficiency, and sustainability. Studies show that implementing smart solutions in the rail industry can lead to significant reductions in operational costs, increased reliability, and improved service quality (Awodele Ia).

Competitive pressures, changing customer expectations, and environmental requirements have driven the rail industry toward adopting new technologies. In this context, the concept of smart railway has emerged as a response to these challenges (Kumar S, 2025). According to recent studies, integrating digital technologies into rail infrastructure can achieve up to 35 percent energy savings and reduce travel delays by up to 45 percent (Kljaić Z, 2023).

However, the path of digital transformation in the rail industry faces multiple challenges. Technical barriers, infrastructure limitations, resistance to change, and management complexities are among the most significant challenges. Studies by Carranza et al. (Carranza G) indicate that over 60 percent of digital transformation projects in the rail industry face delays or failure, highlighting the necessity for a more precise understanding of success patterns and trends.

A review of the literature shows that numerous studies have been conducted on various aspects of digital transformation in the rail industry. For example, López-Aguilar et al. (López-Aguilar P, 2022) have examined information security and privacy challenges in railways. Bianchi et al. (Bianchi G) have studied new techniques for monitoring rail infrastructure. Additionally, Siddiqui et al. (Siddiqui A, 2024) have investigated the impact of fourth-generation technologies on transportation industry sustainability.

However, there remains a significant gap in the comprehensive understanding of digital maturity components in this industry. Existing studies have largely focused on specific aspects of digital maturity and have not provided a holistic and integrated picture of its components. This is while a deep understanding of transformation trends and patterns is essential for successful implementation of digital solutions.

The present research aims to fill this scientific gap by analyzing digital maturity components in the rail industry. The contribution of this research can be considered in several dimensions. First, this study takes a comprehensive and systematic approach to identifying and analyzing emerging trends in digital transformation of the rail industry for the first time. Second, the use of novel text mining methodology in analyzing the scientific literature in this field enables the discovery of hidden patterns and meaningful relationships. Third, this research helps bridge the gap between theory and practice by providing an integrated conceptual framework for understanding various dimensions of digital transformation in the rail industry. Fourth, identifying key challenges and opportunities in the path of digital transformation can serve as a

practical guide for policymakers and managers in the rail industry.

Given the importance of the topic and the existing research gap, this study seeks to answer the main question of what are the main components of digital maturity in the rail industry and what meaningful patterns can be identified in relation to these components? Answering this question can significantly help in better understanding the path of digital transformation and improving the decision-making process in this strategic industry. Next, after reviewing the research literature, the research method is introduced, and then the findings, conclusions, and discussion are presented.

2. Literature Review

2.1. Industry 4.0 and Digital Transformation

Industry 4.0 or the fourth industrial revolution, first introduced at the Hannover Fair in Germany in 2011, represents a profound transformation in production and service methods through the integration of digital and physical technologies (K. A. Zamany A, 2022). This concept aims to create intelligent systems where machines, products, and systems can autonomously exchange information with each other (Duan L, 2024). Kljaić et al. (Kljaić Z, 2023) in their research demonstrated that Industry 4.0 is characterized by key features such as vertical and horizontal system integration, advanced and intelligent automation, and real-time data-based decision making.

The core technologies forming the foundation of this transformation encompass a broad spectrum (K. A. I. S. Zamany A, 2024). Singh et al. (Singh P, 2022) identify the Internet of Things as one of the most important of these technologies, enabling the creation of a network of connected devices and data exchange. Tang et al. (Tang R & et al., 2022) emphasize the key role of artificial intelligence and machine learning in analyzing complex data and automated decision-making. Additionally, Li et al. (Li J, 2023) highlight the importance of cloud computing in storing and processing massive amounts of data.

Digital transformation extends beyond merely implementing new technologies. Babaei et al. (Babaei M) demonstrated in their research that this transformation requires fundamental changes in organizational culture, business models, and operational processes. Similarly, Asri et al. (Asri A) have emphasized the importance of integrated strategies in managing this transformation. Zunder et al. (Zunder Th, 2023) have also shown in their studies that success in digital transformation requires simultaneous attention to technical and technological infrastructure, development of organizational capabilities, and change management and culture building.

2.2. Smart Railway

The concept of smart railway or Railway 4.0, as the manifestation of Industry 4.0 in the rail transport sector, represents a fundamental transformation in this industry (Dalewska M, 2025). Laiton-Bonadiez et al. (Laiton-Bonadiez C, 2022) define smart railway as an integrated ecosystem where all components of the rail transportation system, from infrastructure to fleet and control systems, interact through digital technologies.

Modern technologies play a key role in realizing the smart railway concept. Singh et al. (Singh P, 2022) in their comprehensive study have shown that the Internet of Things enables real-time monitoring of tracks, wagons, and signaling systems. Carranza et al. (Carranza G) have examined the application of artificial intelligence in optimizing train movement planning and predicting maintenance needs. Additionally, Uhlenkamp et al. (Uhlenkamp Jf, 2022) have emphasized the importance of digital twins in simulating and optimizing railway operations.

Taraz Monfared et al. (Taraz Monfared N, 2023) in their research on implementing IoT in the rail freight transportation industry have

identified multiple challenges, including infrastructure issues, cybersecurity, and the need for workforce skill development. Similarly, López-Aguilar et al. (López-Aguilar P, 2022) have emphasized the importance of information security and privacy in smart railways.

However, the opportunities created by railway digitalization are significant. Tshabalala and Mpofu (Tshabalala S, 2023) have enumerated multiple benefits for smart railways, including increased safety, improved operational efficiency, cost reduction, and enhanced passenger services. Bohlooli et al. (Bohlooli N) have also shown in their study that intelligent customer knowledge management can lead to significant improvements in customer satisfaction and productivity in the rail transportation industry.

2.3. Digital Maturity Assessment Models in Rail Industry

Digital maturity assessment models are systematic frameworks for measuring organizations' progress in their digital transformation journey. Brezavšček and Baggia (Brezavšček A, 2025) in their recent research introduce these models as strategic tools that enable assessment of current status, identification of gaps, and planning for digital transformation. Harrod (Harrod, 2025) also emphasizes the importance of these models in assessing rail organizations' readiness for adopting new technologies and managing digital transformation.

In the transportation industry, digital maturity models have been developed considering the specific characteristics of this domain. Taifa and Twaha (Taifa Iw, 2025) in their study have presented a comprehensive framework for evaluating logistics service quality in rail transportation that encompasses various dimensions of digital maturity. Ogunmodede and Zincume (Ogunmodede O, 2024) have also emphasized the importance of developing digital skills alongside technological maturity assessment.

Research background shows that digital maturity assessment in the rail industry is of particular importance. Schumacher et al. (Schumacher A, 2016) presented a model for assessing Industry 4.0 readiness that included 9 dimensions for evaluating fourth industry maturity. Similarly, Comuzzi and Patel (Comuzzi M, 2016) proposed a framework for comprehensive maturity assessment that considers various dimensions such as technology, safety, sustainability, and customer satisfaction.

Chen et al. (Chen Zs) presented a stakeholder-based maturity assessment model that considers the perspectives and expectations of various stakeholders. Additionally, Kim et al. (Kim S, 2022) focused on the challenges facing the implementation of maturity assessment models and identified key barriers such as lack of data standardization, resistance to change, and limited stakeholder participation.

Fathi et al. (Fathi Mr) have examined key factors influencing investment and financing in the rail transportation industry. Nazari & Rostamkhani (Nazari M, 2023) have also developed and prioritized freight transportation pricing strategies using the SWOT technique, demonstrating the importance of a systematic approach to assessment and development in the rail industry.

3. Materials and Methods

This research employs a qualitative approach using text mining methods to analyze digital maturity components in the rail industry. This section describes the research implementation stages, population and sample, data collection and analysis methods, and validation methods.

3.1. Research Population

The research population consists of all scientific articles published in the field of digital transformation and Industry 4.0 in rail transportation. To obtain an appropriate sample, an extensive search was conducted in reputable international scientific databases. These databases included Scopus as the most

comprehensive citation database, Web of Science as the oldest scientific database, IEEE Xplore focusing on technical and engineering articles, and ScienceDirect as one of the most credible scientific publishers. The search timeframe was limited to 2016-2025 to cover the latest trends and developments in this field. To ensure search comprehensiveness, multiple specialized keywords were used, including "rail transportation," "digital transformation," "Industry 4.0," "smart railway," "digital maturity," "IoT in railway," and "digital technologies."

3.2. Article Screening and Selection Process

The article screening process was conducted based on precise scientific criteria. Selected articles had to be in English and published between 2016 and 2025. Additionally, their primary focus had to be on digital transformation and emerging technologies in the rail industry, and they had to be research articles, review papers, or case studies. After applying these criteria and removing duplicate articles, 87 articles with appropriate quality and topical relevance were ultimately selected for final analysis.

3.3. Data Preprocessing Process

For data preprocessing, Python programming infrastructure was used along with specialized natural language processing libraries NLTK and SpaCy. In the first stage, texts were cleaned of extraneous elements such as punctuation marks, numbers, and special characters. Then all letters were standardized to lowercase, and stop words and irrelevant vocabulary were removed. Finally, word spellings were standardized and potential spelling errors were corrected to prepare the text for analysis.

3.4. Data Analysis Methods

Data analysis was conducted with a multi-faceted approach using various text mining methods. In the first stage, the LDA model was used for keyword extraction, which identifies hidden topics and key concepts by analyzing the

probabilistic structure of texts. Subsequently, n-gram analysis was employed to discover meaningful patterns in texts, providing better understanding of concept context and background. Text sentiment analysis was performed using TextBlob to identify attitudes and approaches in scientific texts. For concept organization, the K-Means clustering algorithm was used, and finally, concept network analysis was conducted to understand relationships and semantic connections between key concepts.

3.5. Results Validation

Multiple measures were taken to ensure the validity and reliability of results. Content validity was ensured through evaluation and confirmation by rail industry experts. For convergent validity, results were compared with credible industry reports and recent review articles. The reliability of results was also examined through the use of various analysis methods and comparison of obtained results. Clustering quality was measured using the silhouette coefficient, and for ensuring research reproducibility, all analysis stages were carefully documented. This comprehensive approach to validation ensures the accuracy and precision of research findings.

4. Research Findings

Text mining analysis of articles in the field of digital maturity in the rail transportation industry has led to the identification of significant components and patterns. In this section, research findings are organized into several parts including descriptive analysis of articles, thematic content analysis, concept clustering, and conceptual network analysis.

4.1. Data Preprocessing Results

The data preprocessing was conducted in three main stages. In the first stage (text refinement), the data volume was reduced from 1,250,000 words to 1,200,000 words, indicating successful removal of extraneous elements such as numbers, excess punctuation marks, and irrelevant characters. In the text standardization stage, the data volume remained constant, but

text quality and consistency improved through converting all text to lowercase. Finally, with the removal of high-frequency words, the final data volume reached 980,000 words. This 21.6 percent reduction in data volume significantly helped improve the quality of subsequent analyses. **Error! Reference source not found.** shows the results of this stage.

Table 1. Results of Article Text Preprocessing and Refinement

Stage	Number of Words Before Processing	Number of Words After Processing
Text Refinement	1,250,000	1,200,000
Text Standardization	1,200,000	1,200,000
High-frequency Word Removal	1,200,000	980,000

4.2. Analysis of Key Word Frequency

The frequency distribution chart of 20 key and significant words in scientific texts can be seen in **Error! Reference source not found.**.
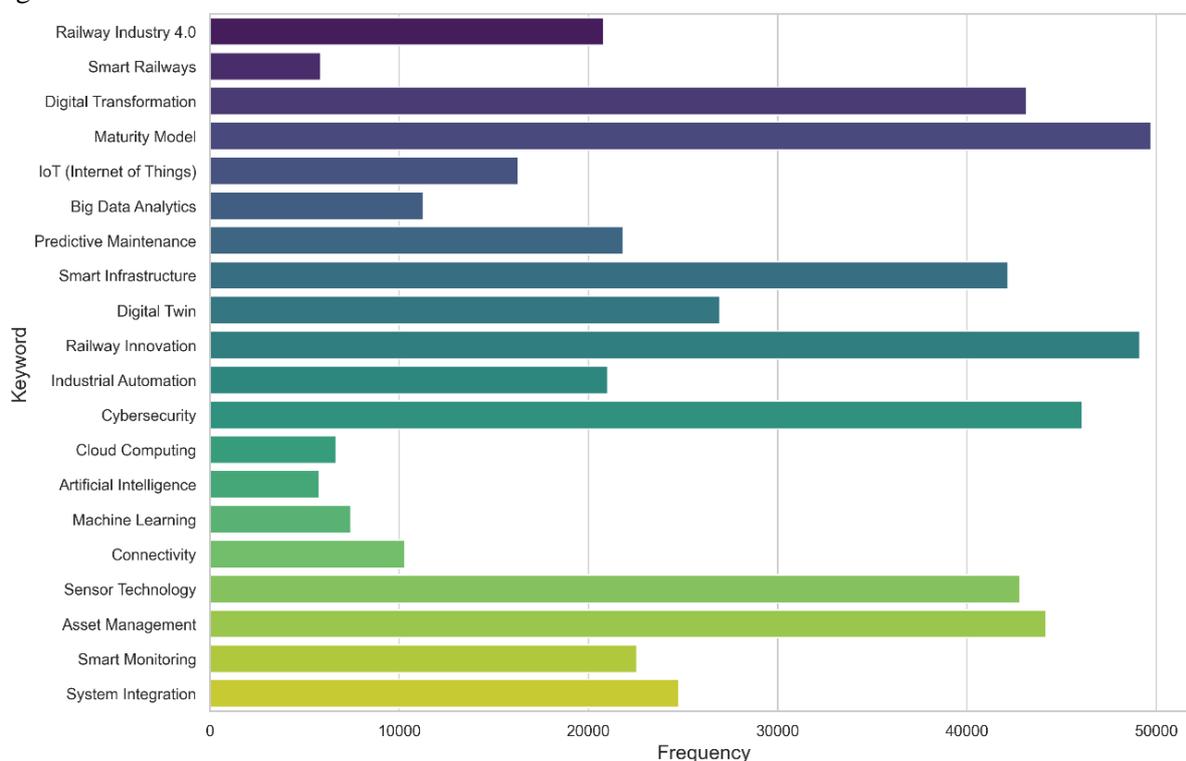
Figure 1. Frequency Distribution of 20 Most Recurring Words in Article Texts

Word frequency analysis shows the focus of studies on key concepts related to maturity models and innovation in the rail industry. Terms like "Maturity Model", "Railway Innovation", and "Cybersecurity" top the list of frequent words. The strong presence of terms such as "Smart Infrastructure" and "Sensor Technology" indicates special attention to smart infrastructure and emerging technologies in the rail industry. Additionally, the high frequency of words like "Energy Efficiency" and "Asset Management" indicates the importance of energy consumption optimization and asset management in the digital transformation of the rail industry.

The word cloud analysis presented in **Error! Reference source not found.** demonstrates that concepts related to smartification and digitalization of the rail industry have particular prominence. The size and position of words in this diagram indicate their relative importance in the literature. Interestingly, concepts related to cybersecurity and smart management also have a significant presence in the texts,

indicating a trend toward integrating new technologies with security considerations in the rail industry.

Figure 2. Word Cloud of High-Frequency and Important Terms in Article Texts

Thematic analysis using the LDA model led to the identification of top 30 keywords shown in **Error! Reference source not found.**. These words, which had the highest frequency and topical relevance in the texts, represent the main focus areas of the studies.

Samaneh Moradi, Mehrdad Hosseini Shakib, Ali Badizadeh

Figure 3. Top 30 Keywords and Their Frequencies

Analysis of LDA results shows that among the top 30 keywords, concepts related to maturity models and innovation in the rail industry have the highest frequency. "Maturity Model" with 49,732 occurrences, "Railway Innovation" with 49,131 occurrences, and "Cybersecurity" with 46,090 occurrences top this list, indicating the main focus of studies on maturity assessment and innovation in the rail industry with a security-oriented approach. Additionally, the strong presence of "Energy Efficiency" with 46,434 occurrences demonstrates researchers' special attention to energy consumption optimization in this industry.

Deeper examination of keywords reveals that concepts related to smartification and digitalization are of significant importance. The strong presence of concepts like "Smart Infrastructure", "Sensor Technology", and "Asset Management" with frequencies over 40,000 occurrences indicates a strong trend toward digitalization and smartification of rail infrastructure. Furthermore, the presence of concepts like "Autonomous Trains" and "Digital Twin" shows the industry's direction toward automation and digital simulation, which fully aligns with Industry 4.0 principles.

4.3. Identification of Textual Patterns

Using the n-gram technique, ten key patterns were identified in the texts, which are presented in **Error! Reference source not found.**. These patterns represent the dominant approach in studies that focus on digital transformation in the rail transportation industry.

Table 2. List of 10 High-frequency Phrases or Word Sequences with Their Frequencies and Analysis

No.	High-frequency Phrase	Frequency	Analysis
1	Digital transformation in railway industry	3,845	Indicates the importance of digital transformation in modern rail industry
2	Smart railway infrastructure management	3,562	Emphasizes smart management of rail infrastructure
3	Railway industry maturity assessment	3,241	Importance of maturity level assessment in rail industry
4	IoT-based predictive maintenance systems	2,987	Application of IoT in predictive maintenance
5	Real-time monitoring and control	2,856	Necessity of real-time monitoring and control of rail systems
6	Big data analytics in railways	2,734	Role of big data analytics in performance improvement
7	Cybersecurity for smart railways	2,645	Importance of cybersecurity in smart rail
8	Autonomous train operation system	2,532	Trend toward train operation automation
9	Energy efficient railway systems	2,423	Focus on energy consumption optimization
10	Digital twin technology implementation	2,318	Importance of implementing digital twin technology

Analysis of high-frequency textual patterns indicates that research primarily focuses on four key areas: digital transformation, smart infrastructure management, organizational maturity assessment, and predictive maintenance systems. These patterns indicate the rail industry's movement toward integrating new technologies with existing infrastructure. In particular, the high repetition of phrases related to digital transformation and smart management demonstrates the high priority of these topics in the development path of Industry 4.0 railways.

Furthermore, the strong presence of concepts such as cybersecurity, operation automation, and digital twin among frequent patterns indicates the increasing importance of safety and efficiency-related topics in the digital environment. Special attention to energy consumption optimization and real-time system monitoring also suggests that the rail industry, in its transformation path, is seeking greater sustainability and efficiency beyond smartification. The relationship graph presented in **Error! Reference source not found.** shows how different concepts are interconnected and form a coherent semantic network.
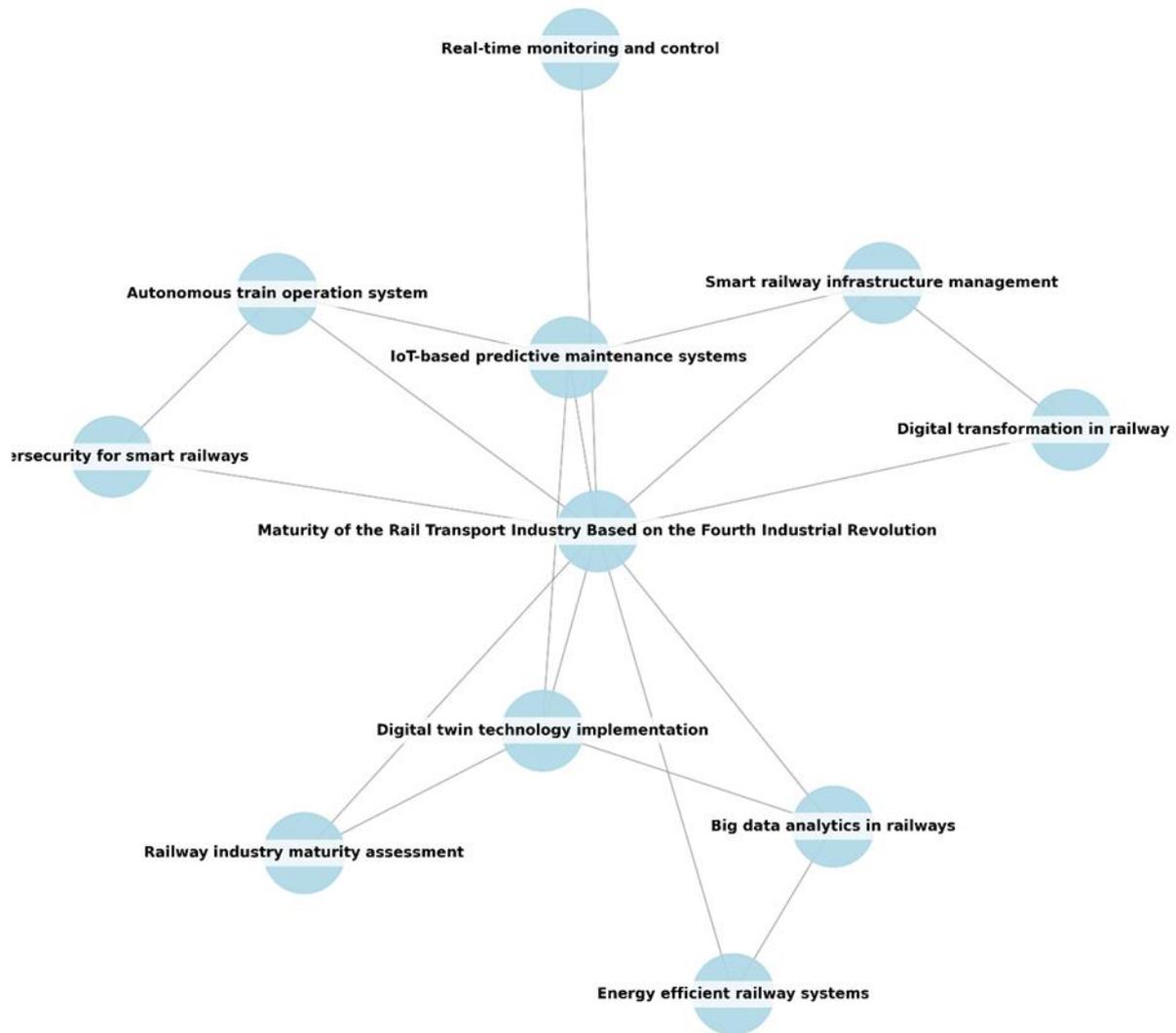
Figure 4. Connection Graph Between High-frequency Phrases

4.4. Text Sentiment Analysis

At this stage, using the TextBlob library, we performed content analysis and sentiment analysis of scientific texts. The main approach was to gain a deep understanding of researchers' attitudes and approaches in the field of Industry 4.0 implementation in rail transportation. **Error! Reference source not found.** shows the distribution of emotional content in texts related to fourth-generation rail industry.

Samaneh Moradi, Mehrdad Hosseini Shakib, Ali Badizadeh
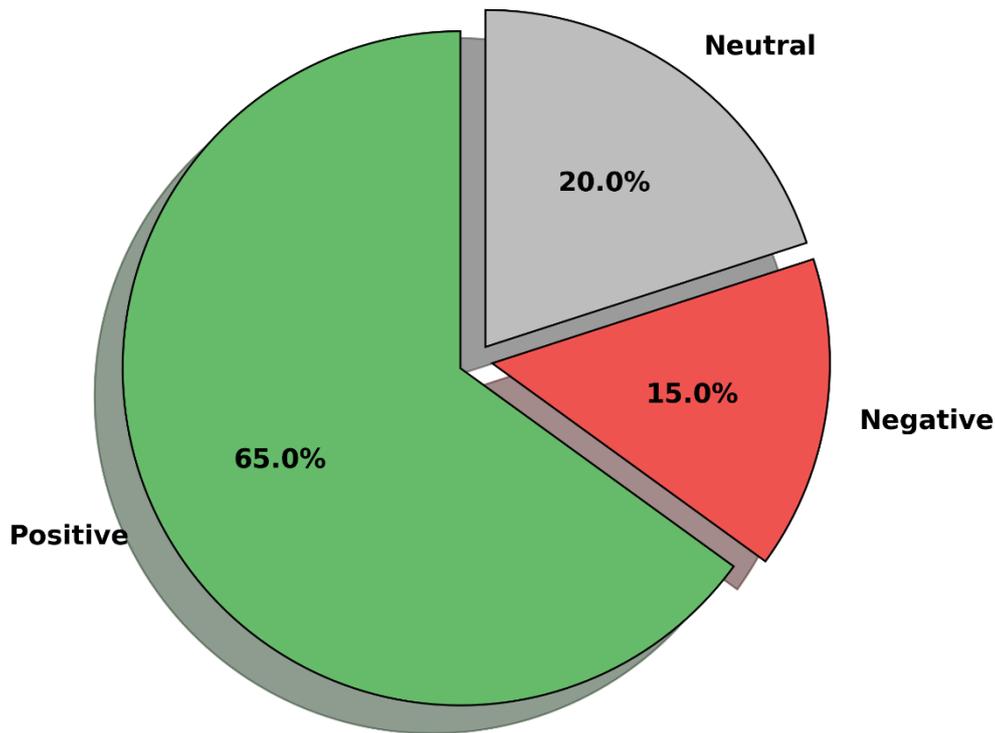
Figure 4. Distribution of Sentiment Content in Related Texts

Text sentiment analysis shows that 65 percent of texts have positive sentiment, 15 percent have negative sentiment, and 20 percent are neutral in terms of emotional approach. **Error!** **Reference source not found.** shows examples of sentences with the highest positive and negative sentiment content.

Table 3. Examples of Sentences with Highest Positive and Negative Sentiment

Sentence	Sentiment	Score
Integration of Industry 4.0 technologies has revolutionized railway operations with remarkable improvements in efficiency and safety	Positive	0.85
Smart railway systems demonstrate exceptional potential for sustainable transportation solutions	Positive	0.78
Cybersecurity threats pose significant risks to digitalized railway infrastructure	Negative	-0.72
Legacy system integration challenges create substantial barriers to digital transformation	Negative	-0.65

Analyzing the probable reasons for positive or negative sentiment, it can be said that positive attitudes are mainly related to the potential of new technologies, efficiency improvement, and increased safety. In contrast, negative attitudes are more related to cybersecurity challenges, implementation costs, and legacy system integration issues.

From the analysis conducted, it becomes clear that there are several key topics in attitudes toward the research subject, which are briefly mentioned in **Error! Reference source not found.**.

Table 4. Analysis of Attitudes Toward Digital Transformation in Rail Industry

No.	Attitude	Analysis
1	Innovation Opportunities	Positive view toward the potential of new technologies in performance enhancement

Analysis of Digital Maturity Components in Rail Transportation Industry: A Text Mining Approach

No.	Attitude	Analysis
2	Security Challenges	Concern about cyber threats and protection of critical infrastructure
3	Economic Issues	Duality between high initial costs and long-term benefits
4	Environmental Impacts	Very positive attitude toward pollution reduction and energy consumption optimization
5	Change Management	Challenges related to workforce training and organizational culture change

Sentiment analysis showed that despite significant challenges in the path of digital transformation in the rail industry, the dominant attitude in the scientific community toward this transformation is positive and encouraging. This indicates researchers' confidence in the potential of Industry 4.0 in enhancing the rail transportation industry.

4.5. Concept Clustering

Using the Scikit-learn library and implementing the K-Means clustering algorithm, we proceeded to categorize the extracted concepts. The main approach was to achieve meaningful groupings of concepts related to rail industry maturity based on Industry 4.0. **Error! Reference source not found.** presents the results of concept clustering.

Table 5. Clustering of Concepts Related to Digital Maturity in Rail Transportation Industry Based on Industry 4.0

Clusters	Concepts
Railway Industry 4.0 Technologies	IoT
	AI & ML
	Big Data Analytics
	Cloud Computing
	Digital Twin
	Smart Automation
	Cyber-Physical Systems
	Advanced Robotics
	AR/VR
Digitalization Challenges	Legacy System Integration
	Implementation Costs
	Change Resistance
	Digital Skills Gap
	Technical Complexity
	Organizational Change Management
	Standardization Challenges
	Maintenance Costs
	System Scalability
	Legal and Regulatory Challenges
	Infrastructure Limitations
	Investment Risks
Cybersecurity & Digital Risks	Critical Infrastructure Protection
	Network Security
	Threat Management
	Data Protection
	IoT Security
	Cryptography
	Incident Response
	Business Continuity

Clusters	Concepts
Practical Applications & Performance Improvement	Predictive Maintenance
	Route Optimization
	Smart Fleet Management
	Condition Monitoring
	Smart Traffic Management
	Automated Control Systems
	Passenger Experience Enhancement
	Fault Detection Systems
	Smart Asset Management
	Fuel Optimization
	Advanced Speed Control Systems
Sustainability & Environment	Carbon Reduction
	Renewable Energy
	Waste Management
	Green Transportation
	Infrastructure Sustainability
	Environmental Reporting
	Energy Optimization
	Green Design
	Material Recycling
	Noise Pollution Reduction
	Life Cycle Assessment
	Environmental Management Systems

Error! Reference source not found. shows the distribution of concepts in each cluster. The dimensions of each section of the diagram are designed proportionally to the number of concepts in that cluster.
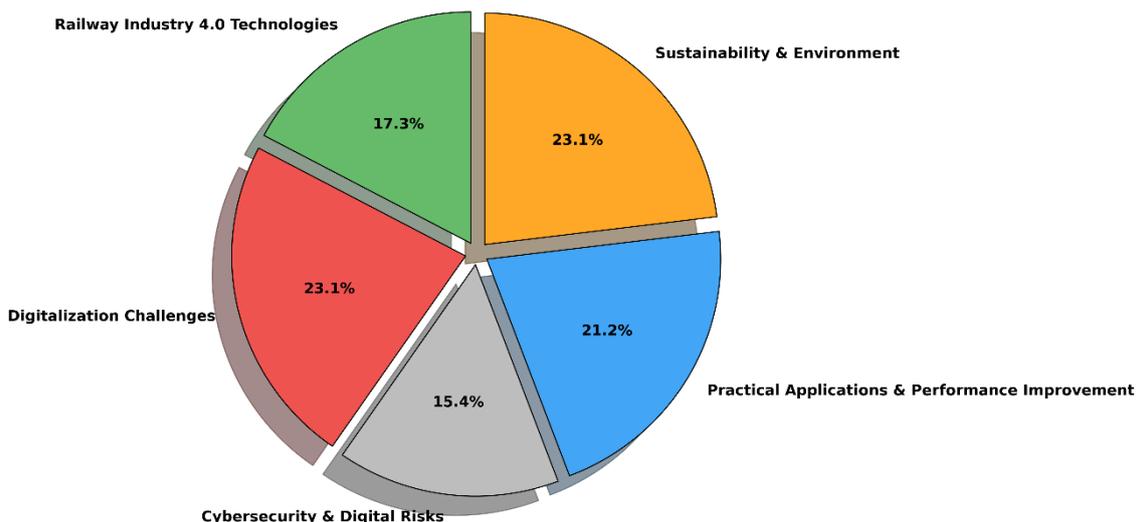
Figure 5. Distribution of Key Concepts in Five Clusters

Analysis of concept distribution in the five clusters shows that the highest conceptual focus is shared equally between two areas: digitalization challenges and environmental sustainability (each 23.1%). This indicates the high importance of implementation barriers and

environmental considerations in the fourth-generation rail transportation industry.

Practical applications and performance improvement with 21.2% ranks next, indicating special attention to practical and operational aspects. Industry 4.0 technologies and cybersecurity follow with 17.3% and 15.4% respectively. This balanced distribution shows that alongside attention to new technologies, security issues and implementation challenges are also of high importance.

The dispersion of concepts in the feature space, presented in **Error! Reference source not found.**, shows appropriate separation of clusters and meaningful relationships within each cluster.
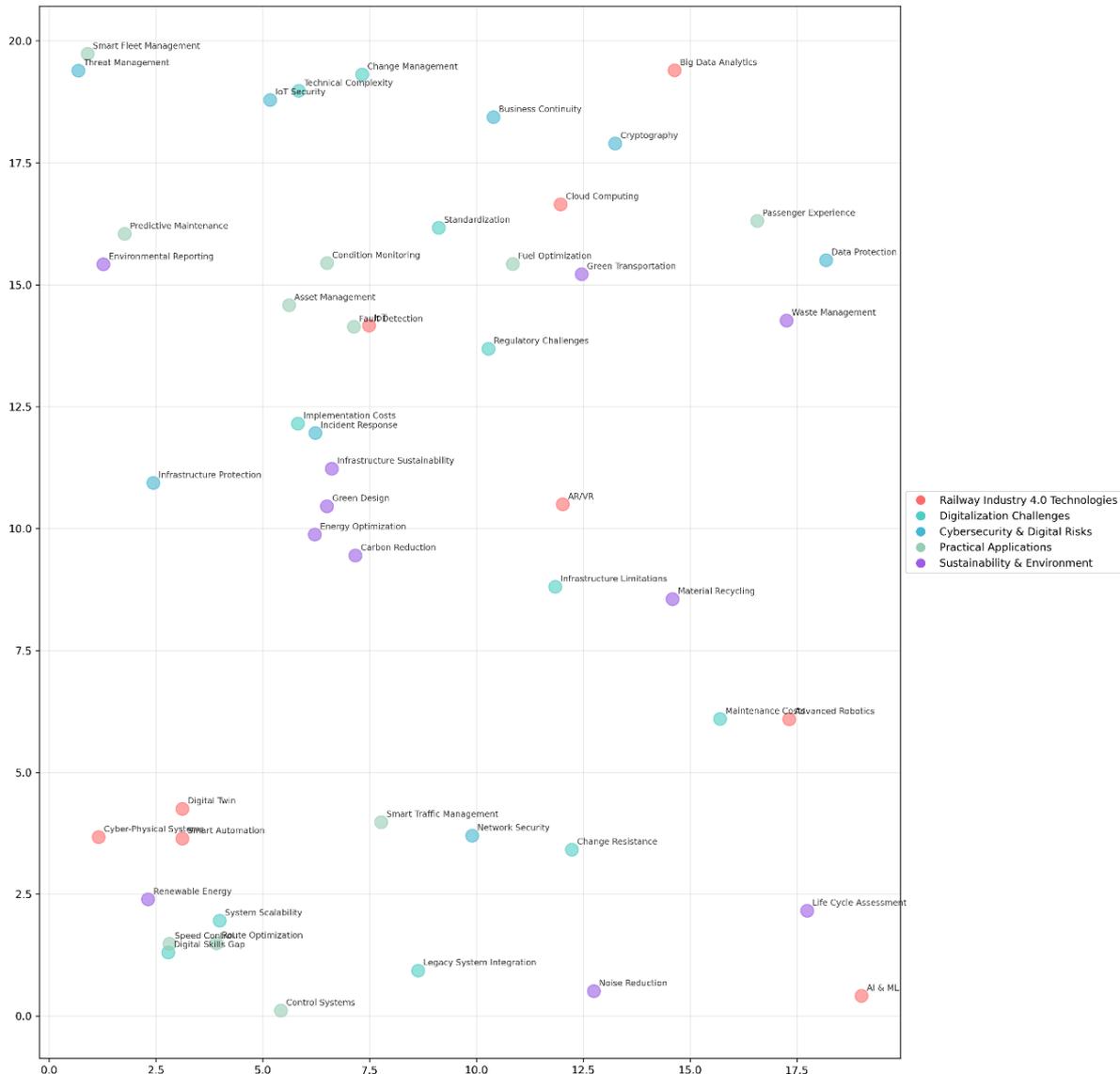
Figure 6. Dispersion of Concepts in Feature Space Based on Score and Weight

4.6. Thematic Analysis of Clusters

After clustering the concepts and identifying related groups, content analysis of the clusters was performed, and the results of this analysis are presented in **Error! Reference source not found.**.

Table 6. Content and Thematic Analysis of Five Clusters

Samaneh Moradi, Mehrdad Hosseini Shakib, Ali Badizadeh

|	Cluster	Main Axes	Key Sub-topics	Emerging Trends
1	Industry 4.0 Technologies	☐ Process Smartification ☐ Digital Integration ☐ Operation Automation	☐ Big Data Processing ☐ Self-learning Systems ☐ Digital Simulation	☐ Combination of New Technologies ☐ Integrated Platforms ☐ Innovative Solutions
2	Digitalization Challenges	☐ Technical Barriers ☐ Organizational Limitations ☐ Economic Challenges	☐ Change Management ☐ Infrastructure Development ☐ Cost Management	☐ New Management Solutions ☐ Gradual Approaches ☐ New Financing Models
3	Cybersecurity & Digital Risks	☐ Infrastructure Protection ☐ Risk Management ☐ Service Continuity	☐ Security Protocols ☐ Monitoring Systems ☐ Crisis Management	☐ Proactive Security ☐ Threat Intelligence ☐ Integrated Security
4	Practical Applications	☐ Operation Optimization ☐ Asset Management ☐ Service Enhancement	☐ Smart Maintenance ☐ Energy Management ☐ Automated Control	☐ Innovative Solutions ☐ Advanced Systems ☐ Smart Services
5	Sustainability & Environment	☐ Pollution Reduction ☐ Consumption Optimization ☐ Resource Management	☐ Clean Energy ☐ Material Recycling ☐ Green Transportation	☐ Clean Technologies ☐ Circular Economy ☐ Sustainable Innovations

Analysis of cluster contents shows significant convergence between new technologies and operational needs. For example, while the Industry 4.0 technologies cluster focuses on integrated platforms, the practical applications cluster implements these technologies in the form of advanced systems and smart services. Additionally, a meaningful connection is observed between challenges and new solutions, such that for each identified challenge, one or more innovative solutions have been proposed.

Another notable point is the strong presence of concepts related to sustainability and security across all clusters. This indicates that the rail industry, in its digital transformation journey, pays special attention to environmental and security issues. Specifically, the presence of concepts such as "circular economy" in the environmental sustainability cluster and "proactive security" in the cybersecurity cluster demonstrates the industry's preventive and forward-looking approach.

4.7. Evaluation of Clustering Quality
To assess the validity and efficiency of created clusters, the silhouette technique was employed. The results of this evaluation, presented in **Error! Reference source not found.**, with a score above 0.7 for all clusters, indicates strong separation and appropriate concept categorization.
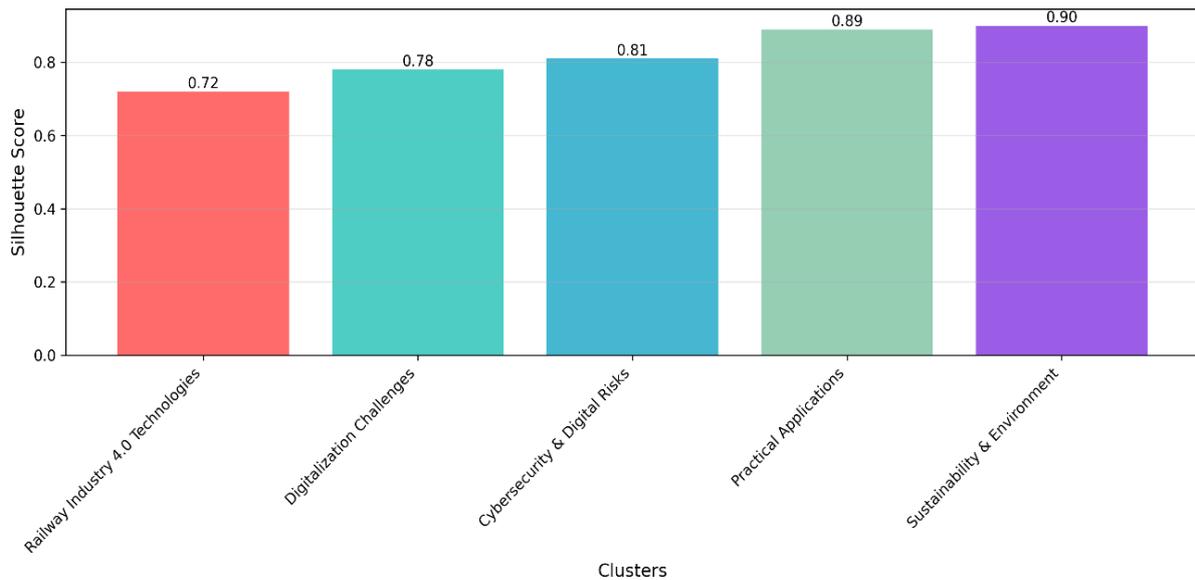
Figure 7. Results of Clustering Quality Assessment Using Silhouette Coefficient

The scientific validity of research results was also ensured through implementing the following steps:

- Content Review: A group of smart rail transportation industry experts reviewed and confirmed the concepts and classification logic. These experts, with practical experience in fourth-generation railways, evaluated the accuracy and relevance of the findings.

- Convergence Assessment: Results were compared with the latest specialized studies and reports in the field of smart rail transportation. This comparative analysis showed alignment of findings with global trends.

- Reliability Assurance: The stability of results was tested through applying various clustering and text analysis methods. Additionally, all analysis stages, from data preparation to algorithm implementation, were carefully documented to enable result reproduction in future research.

5. Discussion

The present research aimed to analyze digital maturity components in the rail transportation industry by using a text mining approach to analyze 87 scientific articles in this field. The results of the conducted analyses revealed significant findings about various dimensions of digital maturity in the rail industry, which are discussed below.

Text mining analysis showed that digital maturity components in the rail transportation industry can be categorized into five main clusters. The balanced distribution of concepts across these clusters, with similar percentages (23.1% for digitalization challenges and environmental sustainability, 21.2% for practical applications, 17.3% for Industry 4.0 technologies, and 15.4% for cybersecurity) indicates the equal importance of these dimensions in achieving digital maturity. This finding differs from previous studies that mainly focused on technical aspects and shows that digital maturity is a multifaceted concept that cannot be considered solely from a technological perspective.

A deeper analysis of clusters shows significant overlap between different components. For example, cybersecurity-related concepts are present in all clusters, but the intensity and nature of this presence varies. In the Industry 4.0 technologies cluster, cybersecurity is presented as a technical capability, while in the digitalization challenges cluster, its managerial and organizational aspects are more prominent.

Samaneh Moradi, Mehrdad Hosseini Shakib, Ali Badizadeh

This finding indicates that developing digital maturity requires a systemic and integrated approach.

The results of text sentiment analysis also revealed noteworthy findings. The dominance of positive attitudes (65%) toward digital transformation indicates the scientific community's optimism about the potential in this field. However, the presence of 15% negative attitudes, mainly related to security challenges and implementation costs, shows that the path to achieving digital maturity faces serious obstacles. This finding aligns with the results of Carranza et al. (Carranza G) study, which reported a 60 percent failure rate in digital transformation projects.

One of the important findings of this research is identifying the strong connection between environmental sustainability and digital maturity. The strong presence of concepts like "Energy Efficiency" among keywords (with 46,434 occurrences) and the allocation of 23.1% of concepts to the sustainability and environment cluster shows that the rail industry pays special attention to environmental considerations in its digital transformation journey. This finding can be considered as a key indicator in evaluating digital maturity.

Analysis of technological components in this research showed that the rail industry is transitioning from traditional approaches toward smart and integrated systems. The presence of concepts like "Smart Infrastructure" and "Digital Twin" among frequent words, and the high frequency of phrases like "IoT-based predictive maintenance systems" (with 2,987 occurrences) and "Real-time monitoring and control" (with 2,856 occurrences) demonstrates this reality. However, the notable point is that success in this transition is not limited to implementing new technologies. Deeper analysis of the challenges and barriers to digitalization cluster shows that issues such as organizational change management, developing digital skills, and legacy system integration play

a key role in the success or failure of digital transformation projects.

Another important finding of this research is identifying the high importance of cybersecurity in the rail industry's digital maturity. The presence of the word "Cybersecurity" with 46,090 occurrences among top keywords and the allocation of an independent cluster to security issues shows that cybersecurity is not just a technical requirement but a strategic component in digital maturity. This finding gains special importance considering the critical nature of rail infrastructure and the increase in cyber threats in recent years, and aligns with the findings of López-Aguilar et al. (López-Aguilar P, 2022) regarding the importance of information security in smart railways.

The results of this research also showed that digital maturity in the rail industry is linked with a customer-centric approach. The presence of concepts such as "passenger experience enhancement" and "smart fleet management" in the practical applications cluster shows that the rail industry pays special attention to service quality improvement and customer satisfaction in its digital transformation journey. This finding aligns with the results of Bianchi et al. (Bianchi G) study, which emphasized the importance of a customer-centric approach in smart railway development.

Based on the research findings, it can be said that achieving digital maturity in the rail industry requires a balanced and comprehensive approach where simultaneous attention to technological, organizational, security, operational, and environmental dimensions is essential. This conclusion aligns with Zunder et al. (Zunder Th, 2023) study, which emphasized the necessity of an integrated approach in digital transformation.

Based on the findings of this research, practical recommendations can be presented for various stakeholders in the rail industry. Cluster analysis results show that rail industry managers and policymakers should adopt a

balanced approach in developing digital capabilities. For example, findings related to the challenges cluster show that investing solely in new technologies, without attention to developing employees' digital skills and organizational change management, will have a low probability of success. Additionally, the strong presence of sustainability-related concepts in the findings shows that decision-makers should consider environmental considerations as one of the main criteria in evaluating digital transformation projects.

From an operational perspective, research findings indicate that implementing IoT-based predictive maintenance systems and real-time data analysis should be prioritized. The high frequency of concepts related to this area indicates its key role in enhancing digital maturity. However, text sentiment analysis results warn that this implementation should be done with special attention to cybersecurity issues and integration with existing systems.

One of the limitations of this research was its focus on scientific texts published in academic databases. Although this approach enabled systematic literature analysis, it might not have covered some practical experiences and tacit knowledge existing in the industry. Additionally, limiting the study timeframe to 2016-2025, while enabling focus on recent trends, might have overlooked some fundamental concepts that were introduced in earlier periods.

For future research, comparative analysis of digital maturity components in rail industries of different countries could provide valuable insights. Additionally, case studies of successful rail organizations in digital transformation could help better understand the practical implementation of identified components. Furthermore, future research could contribute to completing the digital maturity assessment framework in the rail industry by focusing on developing quantitative indicators for measuring each identified component.

The results of this research can be applicable to a wide range of stakeholders in the rail industry. Policymakers can utilize the findings of this study in developing digital transformation strategies. Operational managers and specialists can gain a better understanding of future trends and upcoming challenges. Additionally, technology and innovation companies can identify new business opportunities.

Overall, this research has provided a comprehensive picture of digital maturity components in the rail industry, establishing a theoretical framework for better understanding this concept. The research findings can serve as a guide for rail industry policymakers and managers in designing and implementing digital transformation programs. Additionally, identifying key components and their relationships can provide a basis for developing digital maturity assessment models in future studies.

6. Declaration of Interest Statement

The authors report there are no competing interests to declare.

7. References

− Ale F, D. I., Aderoba O, Adediran AA. An artificial intelligence-based integrated framework for lean and smart manufacturing: a case study of the rail industry. In: 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG); 2024 Apr. IEEE. p. 1-10. doi:10.1109/SEB4SDG60871.2024.10629752.

− Asri A, H. M., Khamseh A. Investigating and identifying the factors affecting freight pricing in the multimodal transportation market in MENA region: a qualitative research. J Transp Res. 2023;20(4):231–248. doi:10.22034/TRI.2022.331135.3025.

− Awodele Ia, M. M., Municio AMG, Chan AP, Darko A, Taiwo R, et al.. Awareness, adoption readiness and challenges of railway 4.0 technologies in a developing economy. Heliyon. 2024;10(4):e25934. doi:10.1016/j.heliyon.2024.e25934.

− Babaei M, K. A., HosseiniShakib M. Sustainable technology transfer model in the rail transport industry with mixed approach. J Transp Res. 2023;20(2):77–102. doi:10.22034/TRI.2022.312583.2973.

− Bianchi G, F. C., Freddi F, Giuliani F, LaPlaca A. Systematic review railway infrastructure monitoring: from classic techniques to predictive maintenance. Adv Mech Eng. 2025;17(1):16878132241285631. doi:10.1177/16878132241285631.

− Bohlooli N, E. M., Ramazani M, SangiNoorPour A. Designing a model to measure the maturity of customer knowledge management in the rail transportation industry (I.I.R. Railway Company). Interdiscip Stud Strateg Knowl. 2022;6(22):197-232. doi:20.1001.1.74672588.1401.6.22.7.6.

− Brezavšček A, B. A. (2025). Recent trends in information and cyber security maturity assessment: a systematic literature review. Systems. 2025;13(1):52. doi:10.3390/systems13010052.

− Carranza G, A. O., de la Rua O. Analysis of the challenges faced by the rail sector: understanding the rail industry of the future through the incorporation of technology and digitisation. Open J Bus Manag. 2023;11(4):1558-1576. doi:10.4236/ojbm.2023.114086.

− Chen Zs, Z. M., Chin KS, Darko A, Wang XJ, Pedrycz W. Optimized decision support for BIM maturity assessment. Autom Constr. 2023;149:104808. doi:10.1016/J.AUTCON.2023.104808.

− Comuzzi M, P. A. (2016). How organisations leverage big data: a maturity model. Ind Manag Data Syst. 2016;116(8):1468–1492. doi:10.1108/IMDS-12-2015-0495.

− Dalewska M, M. B. (2025). Innovative transport: environmental, social, and economic aspects. In: Human Perspectives of Industry 4.0 Organizations. CRC Press; 2025. p. 224-235.

− Duan L, L. J. (2024). Smart composite materials and IoT: revolutionizing real-time railway health monitoring. MRS Commun. 2024;1-17. doi:10.1557/s43579-024-00667-9.

− Fathi Mr, M. M., Moghaddam H. Futures investment and financing of rail transportation industry (machinery and equipment). Future Stud Manag. 2018;29(113):33–54. Available from: https://jmfr.srbiau.ac.ir/article_13021.html.

− Harrod, S. S. (2025). Railway signal digitalization with the European Rail Traffic Management System and positive train control: industry 4.0 expectations and reality. Transp Res Rec. 2025;03611981241265841. doi:10.1177/03611981241265841.

− Khamseh, A. and Sadeghi, Y. (2018). The Development of Innovation Management Model in Petrochemical Companies Producing Polyethylene Products in Iran and prioritize the factors of using Analytic network process (ANP) fuzzy. Journal of Management Accounting and Auditing Knowledge, 7(27), 163-180.

− Kim S, P.-C. R. C. I. L. D. (2022). Organizational process maturity model for IoT data quality management. J Ind Inf Integr.

2022;26:100256.
doi:10.1016/J.JII.2021.100256.

− Kljaić Z, P. D. C. M. T. M. M. T. J. N. M. (2023). An overview of current challenges and emerging technologies to facilitate increased energy efficiency, safety, and sustainability of railway transport. Future Internet. 2023;15(11):347. doi:10.3390/fi15110347.

− Kumar S, K. P. B. R. M. S. S. S. (2025). Smart solution for railways using IoT. AIP Conf Proc. 2025;3159(1):1-7. doi:10.1063/5.0247175.

− Laiton-Bonadiez C, B.-B. J. W. Z.-C. J. P.-S. E. A.-S. M. (2022). Industry 4.0 technologies applied to the rail transportation industry: a systematic review. Sensors. 2022;22(7):2491. doi:10.3390/s22072491.

− Li J, H. M. S. N. J. W. J. Z. (2023). Methods and applications for artificial intelligence, big data, Internet of Things, and blockchain in smart energy management. Energy AI. 2023;11:100208.
doi:10.1016/J.EGYAI.2022.100208.

− López-Aguilar P, B. E. M.-B. A. S. A. (2022). Information security and privacy in railway transportation: a systematic review. Sensors. 2022;22(20):7698. doi:10.3390/s22207698.

− Nazari M, R. A. (2023). Identification and prioritization of freight transportation pricing strategies, using SWOT and QSPM techniques (case study: railway). J Transp Res. 2023;20(1):213–226.
doi:10.22034/TRI.2022.293048.2926.

− Ogunmodede O, Z. P. N. (2024). Railway skill development for the Fourth Industrial Revolution: a systematic review of developing countries. In: 2024 IEEE International Conference on Engineering, Technology, and Innovation (ICE/ITMC); 2024 Jun. IEEE. p. 1-

7.
doi:10.1109/ICE/ITMC61926.2024.10794310.

− Schumacher A, E. S. S. W. (2016). A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. Procedia CIRP. 2016;52:161–166. doi:10.1016/J.PROCIR.2016.07.040.

− Siddiqui A, K. M. R. R. R. M. K. M. A. (2024). Industry 4.0 adoption in transportation: does Industry 4.0 adoption enhance sustainability? A systematic literature review. Int J Supply Oper Manag. 2024;11(2):231-249. doi:10.22034/IJSOM.2024.110058.2852.

− Singh P, E. Z. M. V. K. P. J. D. M. A. (2022). Internet of Things for sustainable railway transportation: past, present, and future. Cleaner Logist Supply Chain. 2022;4:100065. doi:10.1016/j.clscn.2022.100065.

− Taifa Iw, T. I. (2025). Development of the logistics service quality framework for railway transportation in Tanzania. Benchmarking. 2025. doi:10.1108/BIJ-04-2024-0280.

− Tang R, D. D. L. B. i. N. F. F. G. R. M. P. L. Z., & et al. (2022). A literature review of artificial intelligence applications in railway systems. Transp Res C Emerg Technol. 2022;140:103679.
doi:10.1016/J.TRC.2022.103679.

− Taraz Monfared N, S. A. R. G. A. (2023). The effect of Internet of Things (IoT) implementation on the rail freight industry: a futures study approach. Iran J Inf Commun Technol. 2023;15(55):191-207. doi:20.1001.1.27170411.1402.15.55.10.1.

− Tshabalala S, M. K. (2023). Opportunities presented by industrial 4.0 revolution to revitalize the railway sector: a review. In: Smart, Sustainable Manufacturing in an Ever-

Changing World: Proceedings of International Conference on Competitive Manufacturing (COMA'22); 2023 Mar. Cham: Springer International Publishing. p. 243-250. doi:10.1007/978-3-031-15602-1_19.

− Uhlenkamp Jf, H. J. B. B. E. L. M. F. M. T. K. D. (2022). Digital twins: a maturity model for their classification and evaluation. IEEE Access. 2022;10:69605–69635. doi:10.1109/ACCESS.2022.3186353.

− Zamany A, K. A. (2022). Identification of influential dimensions and components of technology transfer with a focus on digital transformation. J Technol Dev Manag. 2022;10(3):57-90.
doi:10.22104/jtdm.2023.5698.3032.

− Zamany A, K. A. I. S. (2024). Analyzing the dimensions and components of technology transfer in the context of Industry 5.0: a technological entrepreneurship approach. J Entrep Res. 2024;3(1):1-18. doi:10.22034/jer.2024.2025559.1091.

− Zunder Th, K. C. C. M. M. A. (2023). Recommendations for a transition to 'Railway 4.0 for Europe'. Transp Res Procedia. 2023;72:852-859.
doi:10.1016/j.trpro.2023.11.477.