Japan Bilingual Publishing Co.

Transportation Development Research
https://ojs.bilpub.com/index.php/tdr

ARTICLE

Data–Driven Technology Foresight: Integrating ARIMA Forecasting and Network Analysis for Patent–Based Roadmapping in Korea's Railway Industry

Yong-Jae Lee [1*], *Tae-Seong Lee* [2]

[1] *Ai Lab, Cloud Group, TOBESOFT, Seoul 06083, South Korea*

[2] *Department of Business, Chung-Ang University, Seoul 06974, South Korea*

ABSTRACT

Traditional expert–driven technology foresight is often limited by subjectivity, low replicability, and high costs. To address these challenges, this study proposes a novel, data–driven framework for technology foresight that offers an objective, scalable, and multi–dimensional approach to identifying and prioritizing emerging technologies in the railway industry. Our tripartite methodology systematically integrates three analytical dimensions: (1) Temporal Forecasting using Autoregressive Integrated Moving Average (ARIMA) models to project the growth trajectories of technology keywords; (2) Structural Analysis using Social Network Analysis (SNA) to evaluate the systemic importance and influence of technologies within the innovation network; and (3) Semantic Analysis using BERT–based contextual embeddings to track the longitudinal evolution of technological concepts. By applying this framework to 4,199 patents from the Korean Intellectual Property Office (1990–2023), we demonstrate its efficacy. The results identify a portfolio of high–priority technologies, such as "sensor," "signal," and "control," which exhibit both strong growth momentum

*CORRESPONDING AUTHOR:
Yong–Jae Lee, Ai Lab, Cloud Group, TOBESOFT, Seoul 06083, South Korea; Email: yj11021@tobesoft.com

ARTICLE INFO
Received: 25 March 2025 | Revised: 17 April 2025 | Accepted: 26 April 2025 | Published Online: 10 May 2025
https://doi.org/10.55121/tdr.v3i1.572

CITATION
Lee, Y.–J., Lee, T.-S., 2025. Data–Driven Technology Foresight: Integrating ARIMA Forecasting and Network Analysis for Patent–Based Roadmapping in Korea's Railway Industry. Transportation Development Research. 3(1): 82–97. DOI: https://doi.org/10.55121/tdr.v3i1.572

COPYRIGHT
Copyright © 2025 by the author(s). Published by Japan Bilingual Publishing Co. This is an open access article under the Creative Commons Attri bu– tionl 4.0 International (CC BY 4.0) License (https://creativecommons.org/licenses/by/4.0/).

and high network centrality. Critically, our semantic drift analysis reveals a significant paradigm shift, exemplified by the term "device," which has evolved from representing simple mechanical components to denoting complex digital control systems. This integrated framework provides a robust, transparent, and replicable methodology for strategic technology roadmapping, offering actionable intelligence that moves beyond static analysis to capture the dynamic nature of technological evolution.

Keywords: Technology Roadmapping; Technology Forecasting; ARIMA Time–Series Analysis; Social Network Analysis; Railway Industry

1. Introduction

The Fourth Industrial Revolution is driving rapid technological change, making technology foresight—the systematic identification of future technological developments—a critical strategic imperative for securing national and corporate competitiveness [1]. The railway industry, a foundational pillar of infrastructure, is currently undergoing a profound transformation driven by digitalization, automation, and sustainability [2-4]. In this high–velocity environment, the ability to anticipate and strategically integrate emerging technologies is no longer just an advantage but a necessity.

However, traditional foresight methodologies, which heavily rely on qualitative, expert–based approaches like Delphi surveys and scenario planning, face significant limitations [5]. While expert knowledge is indispensable, these methods are often constrained by subjective biases, a lack of transparency and replicability, and significant time and resource expenditure [6]. These shortcomings highlight a critical gap and a pressing need for more objective, data–driven, and quantitative approaches that can complement expert judgment with empirical rigor [7]. This study addresses this methodological gap by developing and validating a novel, multi–dimensional quantitative framework for technology foresight, using a comprehensive dataset of railway industry patents (1990–2023). Our research design is motivated by the observation that a technology's future importance is a function of not one but three distinct dynamics: its growth trajectory, its position within the broader innovation ecosystem, and the evolution of its meaning [8]. To capture this complexity, our framework uniquely integrates three analytical lenses, as depicted in the research framework.

First, we employ temporal analysis using Autoregressive Integrated Moving Average (ARIMA) models to forecast the growth potential of technology keywords, identifying which are gaining momentum [9]. Second, we utilize structural analysis via Social Network Analysis (SNA) to map keyword co–occurrence networks, revealing the relational importance of technologies as hubs or bridges within the innovation landscape [10]. Third, we introduce a semantic analysis using BERT–based contextual embeddings to track the longitudinal evolution in the meaning of key terms, thereby uncovering subtle but critical paradigm shifts [11]. By synthesizing these temporal, structural, and semantic dimensions, our framework offers a holistic, objective, and scalable methodology for technology roadmapping. This research contributes not only a practical tool for the railway sector but also a robust analytical model applicable to any industry seeking to navigate complex technological transitions. This research, therefore, contributes an objective, replicable, and scalable approach to enriching technology roadmapping, with applicability extending beyond the railway sector to other complex industrial domains.

2. Literature Review

2.1. Patent Data–Driven Technology Forecasting Approaches

Patents are a well–established proxy for technological innovation and have been extensively utilized for trend analysis and foresight [12]. A significant body of literature employs patent analytics to map technological landscapes through descriptive methods, such as analyzing filing trends, citation networks, and patent classifications [13]. While these studies offer valuable retrospective insights into innovation patterns, they often lack a forward–looking, predictive component [14]. Their focus remains on describing "what has happened" rather than systematically forecasting "what is likely to happen next," which is the critical function of effective technology roadmapping.

To bridge this gap, some researchers have incorporated time–series models, with the Autoregressive Integrated Moving Average (ARIMA) model being a prominent choice for its efficacy in forecasting temporal data [15]. Within technology management, ARIMA has been successfully applied to predict patenting activity and the growth trajectories of specific technologies. However, this approach is inherently uni–dimensional. By evaluating a technology's momentum in isolation, it neglects its relational context. The future importance of a technology depends not only on its individual growth but also on its inter–dependencies and its role within the broader technological ecosystem—a dimension that time–series analysis alone cannot capture.

2.2. Network–Based Approaches to Technology Foresight

Social Network Analysis (SNA) offers a complementary perspective by mapping the relational architecture of an innovation system [16]. By constructing networks from data such as keyword co–occurrence, researchers can identify structurally important technologies. Centrality measures quantify a node's influence, revealing which technologies function as core hubs, bridges between clusters, or peripheral actors. However, conventional SNA–based approaches also present a critical limitation: they typically provide a static snapshot of relationships [17]. Such analyses often fail to capture the temporal dynamics of a technology's growth, meaning a technology that is central today could be in decline, a crucial insight that a static network analysis would miss.

This review of the literature reveals a clear, tripartite research gap. First, while temporal analysis (e.g., ARIMA) can forecast growth, it overlooks structural influence. Second, while structural analysis (e.g., SNA) can identify influential technologies, it often ignores their dynamic growth trajectories. Crucially, a third gap exists: both approaches treat technology keywords as static labels, disregarding the phenomenon of semantic drift, where the meaning and application of a term evolve over time [18]. For instance, the keyword "device" might refer to a simple mechanical component in 1990 but a complex, software–integrated sensor in 2020. Overlooking this semantic evolution can lead to profound misinterpretations of technological trends. This study addresses this tripartite gap by proposing a novel methodology that synthesizes: 1) temporal forecasting (ARIMA) to identify high–growth technologies; 2) structural analysis (SNA) to pinpoint systemically influential technologies; and 3) semantic analysis (BERT) to interpret the evolving nature of these key technologies. This integrated approach facilitates a more dynamic, nuanced, and accurate understanding of technological evolution, moving beyond subjective assessments to offer a robust, data–driven framework.

3. Methodology

3.1. Phase 1: Data Acquisition and Preprocessing

The methodology follows the three–phase research framework illustrated in the R visualization code and **Figure 1**. Each phase is designed to build upon the previous one, moving from raw data to integrated, strategic insights. The foundation of this study is a corpus of 4,199 patent documents (titles and abstracts) related to the railway industry, sourced from the Korean Intellectual Property Office (KIPO) for the period 1990–2023 [9]. To prepare this textual data for analysis, a systematic preprocessing pipeline was implemented in R. This involved: (1) tokenization into individual terms; (2) normalization via conversion to lowercase; (3) removal of standard English stopwords (e.g., "the," "is"); and (4) filtering of domain–specific, non–informative stopwords (e.g., "railway," "korea," "method," "present") to reduce noise and improve signal quality [19]. The processed corpus was then transformed into a Document–Term Matrix (DTM), a mathematical representation where rows correspond to documents and columns to terms, with cell values indicating term frequency [20]. This DTM serves as the primary input for the subsequent temporal and structural analyses.

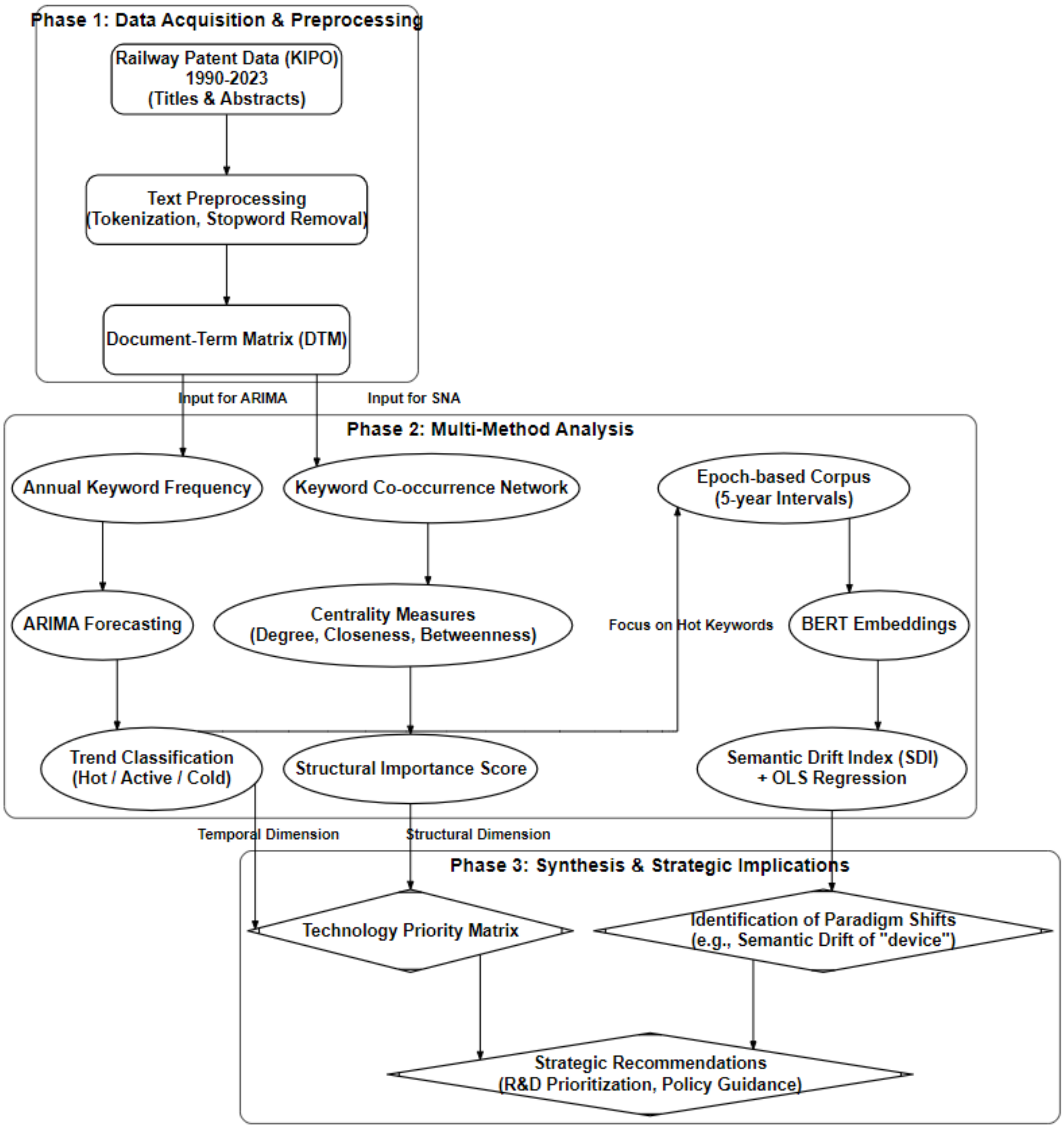

Figure 1. Research Framework: A schematic representation of the overall methodology used in this study.

3.2. Phase 2: Multi–Method Analysis

3.2.1. Temporal Analysis: Forecasting with ARIMA

To forecast the future growth potential of each technology keyword, we performed a time–series analysis on the annual frequency of the top–ranking keywords derived from the DTM. We employed the Autoregressive Integrated Moving Average (ARIMA) model, a robust and widely adopted forecasting method in econometrics and technology studies [21]. For each keyword's annual frequency data, an optimal ARIMA(p, d, q) model was automatically identified using the auto. The arima () function from the forecast package in R. This function systematically searches for the model that minimizes the Akaike Information Criterion (AIC), a measure that balances model fit and complexity [22]. Based on the five–year forecast generated by the optimal model, keywords were classified into three distinct categories: 'Hot' (statistically significant increasing trend), 'Active' (stable or flat trend), and 'Cold' (statistically significant decreasing trend), as illustrated conceptually in **Figure 2**. This stage provides a quantitative basis for identifying technologies with high growth momentum.

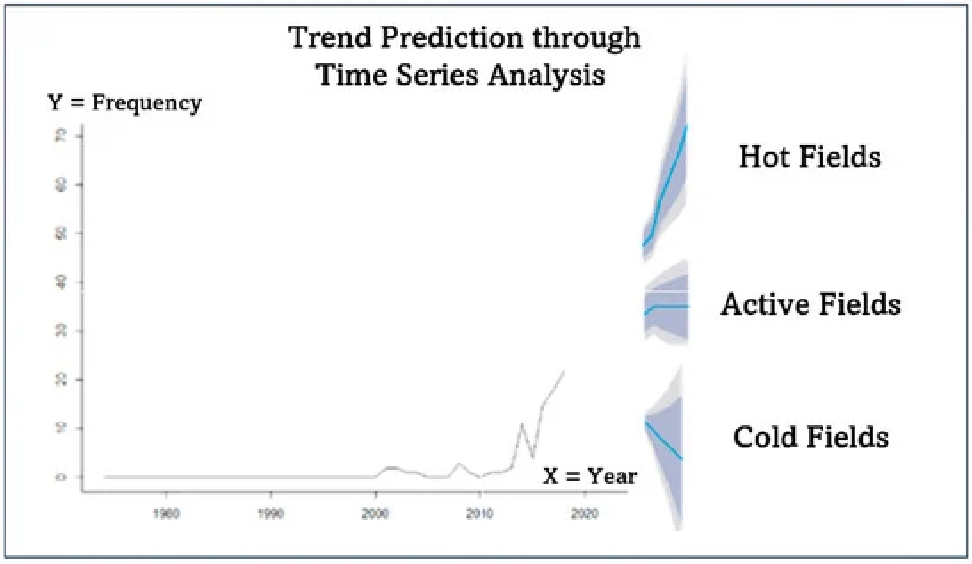

Figure 2. Example of Time–Series (ARIMA) Analysis.

3.2.2. Structural Analysis: Network Centrality with SNA

To evaluate the structural importance and relational dynamics of technologies within the innovation ecosystem, we constructed a keyword co–occurrence network. In this network, nodes represent technology–related keywords, while weighted edges capture the frequency with which two keywords co–appear in the same patent documents. This network–based approach provides a clear view of the relational architecture of the railway technology domain, highlighting how specific technologies interact, cluster, and evolve within the broader innovation landscape.

To quantify the influence of individual technologies within this structure, we applied three established centrality measures [23]. Degree centrality (Equation 1) reflects the overall connectivity of a technology, based on the number of direct links a node possesses. Closeness centrality (Equation 2) captures how efficiently a technology can spread information by measuring the inverse of the sum of its shortest path distances to all other nodes, thereby indicating its integration into the technological core. Betweenness centrality (Equation 3) measures the extent to which a technology functions as a bridge or “gatekeeper,” lying on the shortest paths between other nodes and connecting otherwise disparate clusters within the network.

$$D_i=\sum_{j=1}^{N} Z_{ij} \tag{1}$$

$$C_c=\Big[\sum_{j=1}^{N} d(i,j)\Big]^{-1} \tag{2}$$

$$C_B(v)=\sum_{s\neq v\neq t}\frac{\sigma_{st}(\mathrm{v})}{\sigma_{st}} \tag{3}$$

3.2.3. Semantic Analysis: Conceptual Evolution with BERT

To investigate how the meaning and application of key technologies evolve, we conducted a longitudinal semantic drift analysis [24]. This analysis focused on the ‘Hot’ keywords identified by the ARIMA models to determine if their growing prominence was accompanied by a conceptual shift. The patent dataset was partitioned into seven five–year epochs (1990–1994, 1995–1999, ..., 2020–2024). For each epoch, we generated contextual word embeddings for each occurrence of a target keyword using a pre–trained BERT model (bert–base–uncased) [25]. The aggregate semantic vector for a keyword in a given epoch (V) was computed by averaging the contextual embeddings of all its instances within that epoch (Equation 4).

$$V_e=\frac{1}{N_e}\sum_{i=1}^{N_e} v_{i,e} \tag{4}$$

We then quantified the magnitude of semantic change by calculating a Semantic Drift Index (SDI) for each epoch (e) as the cosine distance between its semantic vector (Ve>) and the vector of the baseline period ($V0$, 1990–1994), as shown in Equation 5. A higher SDI value signifies a greater divergence from the keyword's original meaning. To provide qualitative context for this quantitative shift, we also trained epoch–specific Word2Vec models to identify the nearest semantic neighbors of the target keyword in each period. Finally, to test for a systematic temporal trend in the semantic drift, we performed an Ordinary Least Squares (OLS) regression of the SDI against time (te), where a statistically significant coefficient (β) would confirm a consistent, directional drift in meaning (Equation 6) [26]. This final analytical stage allows us to not only identify a technology's rising prominence but also to understand *how* its role and function are evolving.

$$SDI_e = 1 - \frac{V_e \cdot V_0}{|V_e|\,|V_0|} \tag{5}$$

$$SDI_e = \alpha + \beta t_e + \epsilon_e \tag{6}$$

4. Results

4.1. Descriptive Analysis of Patent Landscape

The final dataset for this study consists of 4,199 railway–related patents filed with the Korean Intellectual Property Office (KIPO) between 1990 and 2023. An analysis of patenting activity over this period, as illustrated in **Figure 3**, reveals a pattern of steady growth from the 1990s, reaching a peak in the mid–2010s, followed by a modest decline. This trajectory suggests a maturation of the industry's foundational technologies and a potential shift in innovation focus. A line graph showing the number of patent applications per year, starting low in 1990, rising to a peak around 2015, and then showing a slight decline towards 2023.

Figure 3. Frequency of Railway Patent Applications (1990–2023).

The innovation landscape is notably concentrated among a few key entities (**Table 1**). The Korea Railroad Research Institute (KRRI) stands out as the most prolific applicant, responsible for 790 patents, which constitutes 18.8% of the total. The top five applicants—including Hyundai–Rotem Co., Daewoo Heavy Industries, Korea Railroad Corporation, and Hyundai Mobis Co. Ltd.—collectively account for 38.1% of all patents in the dataset. This concentration indicates that a small number of major players dominate the technological development within the South Korean railway industry.

Table 1. Top Five Patent Applicants in the Korea Railway Industry (1990–2023).

Rank	Applicant	Number of Patents Filed (Patent Share)
1	Korea Railroad Research Institute (KRRI)	790 (18.8%)
2	Hyundai–Rotem Co.	449 (10.7%)
3	Daewoo Heavy Industries	197 (4.7%)
4	Korea Railroad Corporation	101 (2.4%)
5	Hyundai Mobis Co. Ltd.	64 (1.5%)

4.2. Integrated Foresight: Synthesizing Temporal Trends and Network Structure

To identify high–potential areas for future technological investment, this study integrated time–series forecasting with network analysis. First, text mining of patent titles and abstracts was conducted to extract the 20 most frequent technology–related keywords. As visualized in the word cloud in **Figure 4**, terms such as “structure,” “rail,” “control,” “signal,” and “prevention” emerged as highly prominent, reflecting their foundational importance in railway innovation. These top 20 keywords collectively appeared in over half of the patents in the corpus, underscoring their central role.

Next, to forecast future trends, ARIMA models were applied to the annual frequency data of each keyword. This analysis enabled the classification of keywords into three categories: ‘Hot’ (increasing trend), ‘Active’ (stable trend), and ‘Cold’ (decreasing trend). As depicted in the representative results in **Figure 5**, keywords such as device, control, circuit, speed, sensor, and signal are projected to experience significant growth (‘Hot’). In contrast, foundational terms like track and manufacturing are expected to maintain stable relevance (‘Active’), while keywords such as entrance and test exhibit a declining trajectory (‘Cold’).

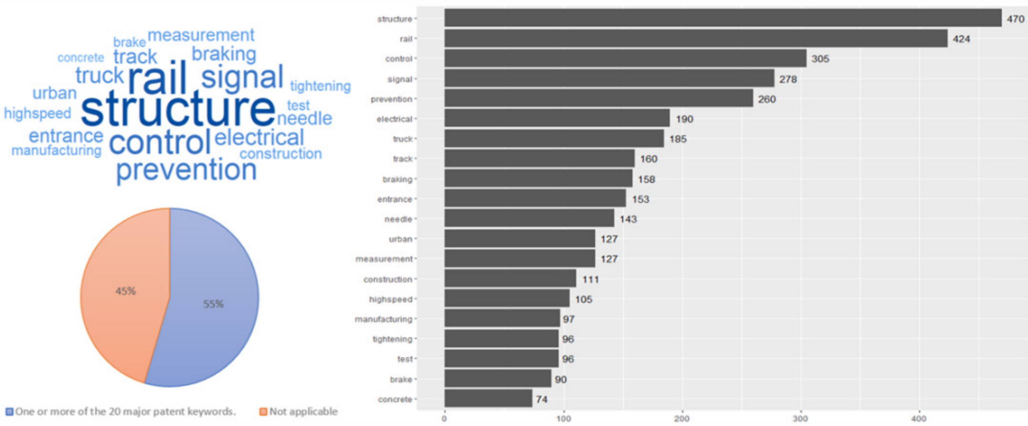

Figure 4. Visualization of Top 20 Technology Keywords.

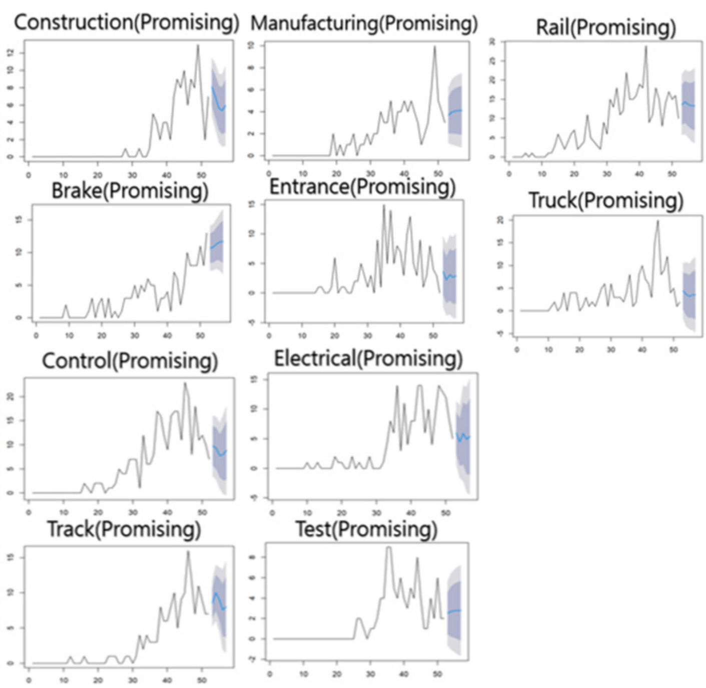

Figure 5. Predicting promising patent keywords through an optimal ARIMA model.

While ARIMA modeling effectively captures temporal momentum, it does not account for a technology's relational importance. To address this, an SNA was performed on a keyword co–occurrence network. Centrality measures were calculated to quantify each keyword's influence within this network. The network visualization in **Figure 6** reveals that keywords identified as 'Hot'—notably device, circuit, speed, sensor, and signal—also possess the highest centrality scores, identifying them as critical hubs and integrators within the railway technology ecosystem.

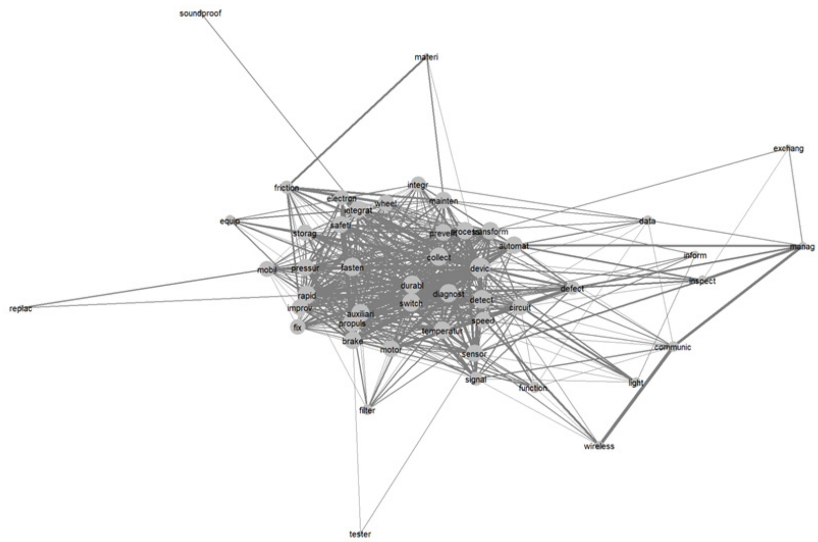

Figure 6. Visualization Results of Social Network Analysis.

By integrating temporal growth forecasts from ARIMA with structural influence metrics from Social Network Analysis (SNA), we developed a Technology Priority Matrix (**Table 2**). This matrix offers a data–driven framework for strategic prioritization by classifying technologies according to both their growth potential and their network centrality. At the top tier, Priority 1: Core & Growing technologies—including device, sensor, signal, circuit, and speed—emerge as the most critical. These areas not only show strong growth trajectories but also occupy central positions in the innovation network, positioning them as the core drivers of digitalization, automation, and high–speed systems. Complementing these, Priority 2: Emerging Connectors, such as control, represent rapidly growing technologies with medium centrality that are poised to become pivotal bridges linking established and emerging clusters.

Table 2. Centrality Measures of Technology Keywords in the Railway Industry.

Temporal Trend (ARIMA)	High (SNA)	Medium (SNA)	Low (SNA)
Hot	Priority 1: Core & Growing (device, sensor, signal, circuit, speed)	Priority 2: Emerging Connectors (control)	–
Active	–	Priority 3: Established Core (track)	Priority 4: Niche & Stable (manufacturing)
Cold	–	–	Priority 5: Declining (entrance, test)

In contrast, Priority 3 and 4 technologies—such as track (established core) and manufacturing (niche but stable)—are more mature and foundational, demanding sustained but less intensive innovation efforts. Finally, Priority 5: Declining or Legacy areas, including entrance and test, exhibit both low growth and weak structural influence, signaling a decline in strategic importance. Together, this integrated analysis provides an objective and multi–dimensional framework for guiding R&D investment and long–term planning, ensuring that resources are allocated toward technologies with the highest potential for shaping the future of the railway innovation ecosystem.

4.3. Establishment of Technology Management Strategy Based on Promising Detailed Technology Keywords

The integrated analysis pinpoints device, circuit, speed, sensor, and signal as the highest–priority technology domains that are shaping the industry's future. These keywords consistently rank highest in both network influence and growth potential, marking them as prime areas for strategic R&D investment, as summarized in **Table 3**. Concrete examples of innovation in these priority areas can be seen in patents from the Korea Railroad Research Institute (KRRI), as detailed in **Table 4**. These patents exemplify the strategic direction suggested by the data. The identified technologies highlight several key strategic thrusts for the railway industry.

Table 3. Priority Technology Keywords and Their Characteristics.

Technology Keyword	Social Network Analysis	Time Series Analysis	Priority
device, circuit, speed, sensor, signal	High	Hot	1
control	Medium	Hot	2
track	Medium	Active	3
manufacturing	Low	Active	4
entrance, test	Low	Cold	5

Table 4. Key Patents from Korea Railroad Research Institute (KRRI) Related to Priority Keywords.

Patent Title	Description	IPC Classification	Country of Application
Hybrid Heating Type Railroad Waste Sleeper Pyrolysis Equipment	A system for pyrolyzing waste wood sleepers using hybrid heating technology, including microwave application to reactor containers.	C10B19/00 (Heating of coke ovens by electrical means)	South Korea (KR), United States (US)
Method and Apparatus for Providing Automated Interlocking Logic for Railway Signaling Based on Moving Block	A railway signaling system utilizing automated interlocking logic based on a moving block system to enhance train path efficiency and safety.	B61L27/00 (Central railway traffic control systems; Trackside control; Communication systems)	South Korea (KR), United States (US), Japan (JP), World Intellectual Property Organization (WO)

First, investing in smart railway infrastructure is crucial for automation and operational efficiency. The KRRI patent for "Automated Interlocking Logic for Railway Signaling Based on Moving Block" (**Figure 7**) is a prime example. Such systems, which leverage sensor and signaling technologies, enable real–time optimization, predictive maintenance, and enhanced safety, directly aligning with the high–priority keywords. A diagram illustrating a centralized control system communicating with multiple trains on a track, optimizing their movement and spacing in real–time.

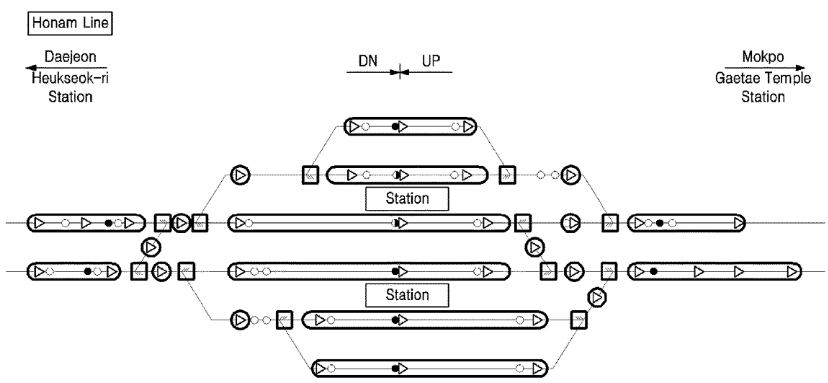

Figure 7. Automated Interlocking Logic System for Railway Signaling Using Moving Block Technology.

Second, a focus on sustainability and energy–efficient technologies is vital for addressing environmental concerns. The "Hybrid Heating Type Railroad Waste Sleeper Pyrolysis Equipment" developed by KRRI (**Figure 8**) demonstrates innovation in the circular economy by efficiently processing waste materials. Such eco–friendly solutions align with global sustainability goals and can create new economic opportunities. A schematic of an industrial system showing waste sleepers being fed into a reactor, which uses hybrid heating to convert them into other materials, contributing to a circular economy. Finally, the international applications for these patents underscore the importance of global patent expansion and strategic collaboration to secure technological leadership and protect intellectual property in key markets.

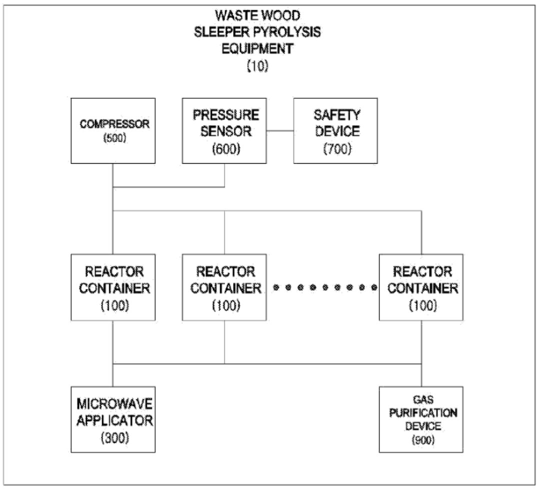

Figure 8. Hybrid Heating Type Railroad Waste Sleeper Pyrolysis Equipment.

4.4. Semantic Drift Analysis: Uncovering the Evolving Meaning of Device

To add a qualitative dimension to the foresight analysis, a semantic drift analysis was conducted on the "Hot" priority keywords using BERT–based contextual embeddings. This analysis aimed to determine if a keyword's meaning evolved as its prominence grew. The results revealed that among the high–growth keywords, device was the only one to exhibit a statistically significant semantic drift over time. The Ordinary Least Squares (OLS) regression results in **Table 5** confirm this finding. "Device" shows a strong model fit ($R^2 = 0.838$) and a highly significant positive coefficient ($p < 0.01$), indicating a systematic and directional shift in its meaning. In contrast, other priority keywords like 'speed', 'sensor', 'signal', and 'circuit' remained semantically stable.

Table 5. OLS Regression Results for Semantic Drift of "Hot" Keywords.

Keyword	R–squared	Adj. R–squared	Coefficient (x1)	*p*–value
Device	0.838	0.811	0.0012	0.001
Speed	0.328	0.194	0.0021	0.179
Sensor	0.458	0.322	0.0046	0.140
Signal	0.459	0.351	0.0089	0.095
Circuit	0.004	−0.196	0.0002	0.897

The nature of this evolution becomes clear when examining the top semantic neighbors of "device" across different time periods (**Table 6**). In the early 1990s, "device" was associated with mechanical concepts like friction and valves. By the 2000s, its context shifted to include electrical terms such as array and current. After 2010, its neighbors increasingly reflect digital and control–oriented concepts like sidelink, control, and location_indicator, signaling a transition toward smart, interconnected systems. This longitudinal shift is quantified by the Semantic Drift Index (SDI), which measures the change in meaning relative to the 1990–1994 baseline. As visualized in **Figure 9**, the SDI for "device" shows a consistent upward trend, confirming a systematic departure from its original mechanical context. A line chart with the Y–axis as SDI (cosine distance) and the X–axis as time epochs. The line shows a steady upward trend, with an R^2 value of 0.838 and $p < 0.01$ noted, indicating a strong positive correlation between time and semantic change.

Table 6. Top Semantic Neighbors of *Device* Across Five–Year Epochs.

Epoch	Top Semantic Neighbors of *Device*
1990–1994	extended, friction, vehicle_side, valve_vehicle
1995–1999	center, fastening, earthquake_damage, measuring
2000–2004	automatically_evaluating, lubricant_supply, equipment
2005–2009	pad_directly, array, current, conversion, smoke
2010–2014	vehicles, vehicle, power, electric, magnetic
2015–2019	comprising_one, information, electric, motor
2020–2024	thereof, electric, using, sidelink, control
2025–2029	location_indicator, rotating_lifting, blower, system

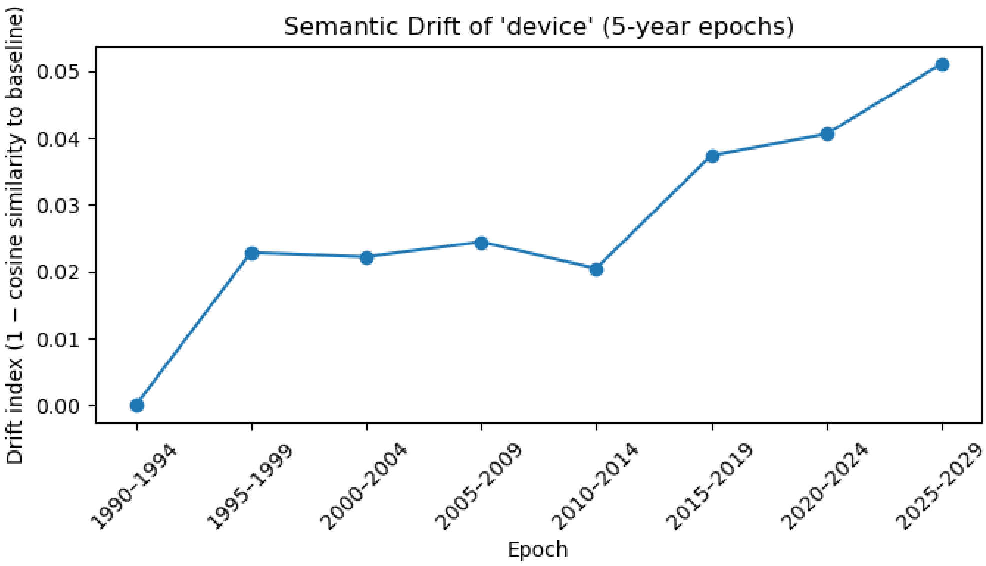

Figure 9. Semantic Drift Index (SDI) of Device (1990–2029).

The robustness of this trend is further validated by the detailed OLS regression model in **Table 7**, which yields a positive and statistically significant coefficient ($\beta = 0.0012$, $p < 0.01$). This result confirms that the meaning of "device" has consistently evolved, reflecting the convergence of mechanical, electrical, and digital technologies in modern railway innovation. This analysis demonstrates the value of examining semantic drift to uncover nuanced paradigm shifts that are often missed by frequency or network analyses alone.

Table 7. OLS regression results for the semantic drift of the device.

Variable	Coefficient	Std. Error	t–value	*p*–value	95% CI (Lower)	95% CI (Upper)
Constant	−2.3047	0.418	−5.513	0.001	−3.328	−1.282
Year	0.0012	0.000	5.578	0.001	0.001	0.002

5. Discussion

5.1. Limitations and Caveats

While this study's framework provides a robust approach to technology foresight, several limitations must be acknowledged to properly contextualize the findings and inform future research. First, the reliance on patent data as the sole proxy for innovation has inherent constraints [27]. Not all inventions are patented, as firms may opt for trade secrets to protect process innovations or proprietary software algorithms. Consequently, the observed decline in patent filings post–2010 might reflect not only industry maturation but also a strategic pivot toward other forms of intellectual property protection, potentially leading to an underestimation of innovation in certain domains like software. Second, the predictive power of ARIMA models is contingent on the assumption that future trends will follow historical patterns [28]. While effective for linear and seasonal trends, these models may be less reliable for forecasting sudden, non–linear technological disruptions that lack historical precedent. Therefore, our framework is better suited to identifying emerging trends within an established paradigm than to predicting the genesis of entirely new ones.

Third, the analytical methods themselves have limitations [29]. Keyword extraction, even when carefully curated, inevitably simplifies complex technical concepts. The SNA component effectively maps the associative structure between technologies but does not, by itself, reveal the causal or directional nature of these relationships (e.g., whether advances in sensor technology drive innovation in signal systems or vice–versa). Finally, this study's scope was intentionally confined to patent data from the Korean Intellectual Property Office to provide a deep longitudinal analysis of a single national industry. This national focus, however, limits the direct generalizability of our specific findings to the global railway industry, as innovation priorities and trajectories may differ across regions like Europe, North America, or Japan.

5.2. Avenues for Future Research

The limitations of this study highlight several opportunities for methodological refinement. Future research could improve predictive accuracy by developing hybrid forecasting models that combine ARIMA with advanced machine learning techniques such as Long Short–Term Memory (LSTM) networks. These hybrid approaches are capable of capturing both linear trends and non–linear complexities, thereby offering more robust forecasts for disruptive technological change. Further benchmarking against alternative models such as Prophet or Vector Autoregression (VAR) would not only validate their performance but also enhance methodological rigor.

Another important avenue lies in data triangulation. The foresight framework's robustness can be significantly strengthened by incorporating a wider variety of data sources. Expanding the analysis to include global patent repositories such as USPTO, EPO, and WIPO would enable meaningful cross–national comparisons, while integrating scientific publications, venture capital investments, market reports, and corporate press releases could provide earlier signals of emerging technologies. Such triangulation would help bridge the gap between early–stage innovation and its eventual commercialization, creating a more holistic and forward–looking analytical framework.

Finally, future studies should pursue qualitative–quantitative integration and scope expansion. A mixed–methods approach—where quantitative insights guide case studies and expert interviews in 'Hot' technology domains—would validate results, uncover causal mechanisms, and provide richer contextual narratives of innovation dynamics. Beyond the railway sector, applying the framework to industries such as renewable energy, aerospace, or biotechnology would test its generalizability and refine its methodological robustness. Within the railway industry itself, more granular analyses of sub–domains like high–speed rail, urban mobility, or freight logistics could generate highly targeted insights to inform domain–specific strategies.

6. Conclusions

This study proposed and validated a novel, data–driven framework for technology foresight, designed to overcome the limitations of traditional expert–based methods by providing an objective, replicable, and multi–dimensional approach. By systematically integrating temporal forecasting

(ARIMA), structural network analysis (SNA), and longitudinal semantic analysis (BERT), our research offers a more nuanced and dynamic understanding of technological evolution within the South Korean railway industry from 1990 to 2023.

6.1. Principal Findings and Implications

Our principal findings are that a technology's future strategic value is a composite of its growth momentum, structural influence, and conceptual evolution [30]. The integrated analysis, synthesized in the Technology Priority Matrix, successfully identified a portfolio of core technologies (e.g., *device, sensor, signal*) that are not only projected to grow ('Hot') but also function as central hubs in the innovation network [31]. This provides clear, evidence–based prioritization for R&D investment.

Most significantly, the semantic drift analysis uncovered a latent paradigm shift that would be invisible to conventional quantitative methods [32,33]. The statistically significant evolution of the term "device" from a mechanical component to a digital control system provides empirical proof of the railway industry's fundamental transition towards digitalization and intelligent systems. This finding implies that future innovation strategies must shift from an engineering–centric to a systems–integration and service–oriented perspective [34].

6.2. Theoretical and Practical Implications

The theoretical contribution of this study lies in presenting a holistic framework that bridges the tripartite gap in existing technology foresight research by integrating temporal, structural, and semantic dimensions [35,36]. This multidimensional approach advances the field beyond traditional uni–dimensional analyses, demonstrating that the significance of a technology is shaped not only by its growth trajectory but also by its systemic role and evolving meaning [37]. In particular, the identification of semantic drift as a measurable indicator of paradigm shifts provides technology management scholars with a novel analytical tool for capturing and interpreting shifts in innovation trajectories. The practical contribution of this framework is its ability to deliver an objective, evidence–based, and reproducible methodology for strategic decision–making. By grounding foresight in quantifiable trends and relational insights, stakeholders can allocate R&D resources more effectively, design targeted technology roadmaps, and formulate policy with greater confidence than when relying on intuition alone [38]. Furthermore, the framework's adaptability across different sectors underscores its value as a strategic intelligence tool for managing technological change in any industry where innovation processes are documented in textual data [36].

6.3. Strategic Recommendations for Industry Stakeholders

Based on the findings, several strategic recommendations can be advanced to strengthen the competitiveness and long–term sustainability of Korea's railway industry. First, stakeholders should prioritize investment in digital core technologies, particularly in Priority 1 and 2 domains such as devices, sensors, signal, and control. Concentrated R&D in AI–powered signaling, IoT–enabled predictive maintenance, and high–speed communication systems will accelerate the transition toward smart, automated, and efficient railway infrastructure. Equally important is the need to embrace a service–oriented innovation model. The semantic evolution of "device" illustrates a broader shift from engineering–centric development to user– and service–focused innovation. Future systems should therefore be designed to enhance passenger experience, deliver real–time information services, and integrate seamlessly with other mobility platforms.

In parallel, stakeholders should adopt a more proactive global intellectual property strategy and foster international collaboration. Securing patents in major global markets such as the US, EU, and Japan is critical for protecting technological advances and expanding industry influence. At the same time, forging alliances with leading technology firms and research institutes will accelerate co–development and knowledge transfer. Finally, industry must advance sustainable and circular technologies, with emphasis on eco–friendly innovations such as the waste sleeper pyrolysis system, improved energy efficiency, and circular economy practices. Pursuing these strategies in combination will enable the Korean railway industry to evolve into a sector that is not only technologically advanced but also globally competitive, user–centered, and environmentally sustainable.

Author Contributions

Conceptualization, Y.–J.L.; methodology, Y.–J.L.; software, Y.–J.L.; validation, Y.–J.L. and T.–S.L.; formal analysis, Y.–J.L. and T.–S.L.; investigation, Y.–J.L. and T.–S.L.; resources, Y.–J.L.; data curation, Y.–J.L.; writing—original draft preparation, Y.–J.L. and T.–S.L.; writing—review and editing, Y.–J.L. and T.–S.L.; visualization, Y.–J.L. and T.–S.L.; supervision, Y.–J.L.; project administration, Y.–J.L.; funding acquisition, Y.–J.L. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Data will be made available on request

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Pietrobelli, C., Puppato, F., 2016. Technology foresight and industrial strategy. Technological Forecasting and Social Change. 110, 117–125. DOI: https://doi.org/10.1016/j.techfore.2015.10.021

[2] Khabarov, V.I., Volegzhanina, I.S., 2020. Digital Railway as a precondition for industry, science and education interaction by knowledge management. IOP Conference Series: Materials Science and Engineering. 918(1), 012189. DOI: https://doi.org/10.1088/1757–899X/918/1/012189

[3] Phusakulkajorn, W., Núñez, A., Wang, H., et al., 2023. Artificial intelligence in railway infrastructure: current research, challenges, and future opportunities. Intelligent Transportation Infrastructure. 2, liad016. DOI: https://doi.org/10.1093/iti/liad016

[4] Love, P.E.D., Zhou, J., Matthews, J., et al., 2018. Managing rail infrastructure for a digital future: Future–proofing of asset information. Transportation Research Part A: Policy and Practice. 110, 161–176. DOI: https://doi.org/10.1016/j.tra.2018.02.014

[5] Meyer, T., Von Der Gracht, H.A., Hartmann, E., 2022. Technology foresight for sustainable road freight transportation: Insights from a global real-time Delphi study. FUTURES & FORESIGHT SCIENCE. 4(1), e101. DOI: https://doi.org/10.1002/ffo2.101

[6] Terk, E., Hunt, T., Saar, I., 2025. Future shifts in the competitiveness of freight transport: a Delphi method–based study. Estonian Discussions on Economic Policy. pp. 192–214. DOI: https://doi.org/10.15157/TPEP.V32I1–2.25523

[7] Xie, J., Huang, J., Zeng, C., et al., 2020. Systematic Literature Review on Data–Driven Models for Predictive Maintenance of Railway Track: Implications in Geotechnical Engineering. Geosciences. 10(11), 425. DOI: https://doi.org/10.3390/geosciences10110425

[8] Moretto, S.M., Palma, A.P., Moniz, A.B., 2012. Constructive technology assessment in railway: The case of high–speed train industry. International Journal of Railway Technology. 1(3), 73–95.

[9] Lee, Y.–J., Han, Y.J., Kim, S.–S., et al., 2024. Identifying the Technology Opportunities and the Technology Taxonomy for Railway Static Inverters With Patent Data Analytics. IEEE Access. 12, 17389–17403. DOI: https://doi.org/10.1109/ACCESS.2024.3360138

[10] Kwon, K., Jun, S., Lee, Y.–J., et al., 2022. Logistics Technology Forecasting Framework Using Patent Analysis for Technology Roadmap. Sustainability. 14(9), 5430. DOI: https://doi.org/10.3390/su14095430

[11] Feldhans, R., Wilke, A., Heindorf, S., et al., 2021. Drift Detection in Text Data with Document Embeddings. In: Yin, H., Camacho, D., Tino, P., et al. (eds.). Intelligent Data Engineering and Automated Learning – IDEAL 2021, Lecture Notes in Computer Science. Springer International Publishing: Cham, Switzerland. pp. 107–118. DOI: https://doi.org/10.1007/978–3–030–91608–4_11

[12] Archibugi, D., 1992. Patenting as an indicator of technological innovation: a review. Science and

Public Policy. 19(6), 357–388. DOI: https://doi.org/10.1093/spp/19.6.357
[13] Chakraborty, M., Byshkin, M., Crestani, F., 2020. Patent citation network analysis: A perspective from descriptive statistics and ERGMs. PLOS ONE. 15(12), e0241797. DOI: https://doi.org/10.1371/journal.pone.0241797
[14] Song, K., Kim, K., Lee, S., 2018. Identifying promising technologies using patents: A retrospective feature analysis and a prospective needs analysis on outlier patents. Technological Forecasting and Social Change. 128, 118–132. DOI: https://doi.org/10.1016/j.techfore.2017.11.008
[15] Sokmen, N., Petrov, A.E., 2021. Forecasting of Radar Technologies Using Patent Information. In Proceedings of the 15th International Conference on Application of Information and Communication Technologies (AICT), Baku, Azerbaijan, 13 Oectober 2021; pp. 1–5. DOI: https://doi.org/10.1109/AICT52784.2021.9620514
[16] Linares, I.M.P., De Paulo, A.F., Porto, G.S., 2019. Patent–based network analysis to understand technological innovation pathways and trends. Technology in Society. 59, 101134. DOI: https://doi.org/10.1016/j.techsoc.2019.04.010
[17] Li, S., Garces, E., Daim, T., 2019. Technology forecasting by analogy–based on social network analysis: The case of autonomous vehicles. Technological Forecasting and Social Change. 148, 119731. DOI: https://doi.org/10.1016/j.techfore.2019.119731
[18] Sarica, S., Luo, J., Wood, K.L., 2020. TechNet: Technology semantic network based on patent data. Expert Systems with Applications. 142, 112995. DOI: https://doi.org/10.1016/j.eswa.2019.112995
[19] Kalaivani, E.R., Marivendan, E.R., 2021. The effect of stop word removal and stemming in datapreprocessing. Annals of the Romanian Society for Cell Biology. 25(6), 739–746. Available from: http://annalsofrscb.ro/index.php/journal/article/view/5490 (cited 24 March 2025).
[20] Jun, S., Sung Park, S., Sik Jang, D., 2012. Technology forecasting using matrix map and patent clustering. Industrial Management & Data Systems. 112(5), 786–807. DOI: https://doi.org/10.1108/02635571211232352
[21] Chen, D.–Z., Lin, C.–P., Huang, M.–H., et al., 2012. Technology Forecasting Via Published Patent Applications and Patent Grants. Journal of Marine Science and Technology. 20(4). DOI: https://doi.org/10.6119/JMST–011–0111–1
[22] Xue, D., Hua, Z., 2016. ARIMA Based Time Series Forecasting Model. Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering). 9(2), 93–98. DOI: https://doi.org/10.2174/2352096509999160819164242
[23] Gilsing, V., Nooteboom, B., Vanhaverbeke, W., et al., 2008. Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. Research Policy. 37(10), 1717–1731. DOI: https://doi.org/10.1016/j.respol.2008.08.010
[24] Harrigian, K., Dredze, M., 2022. The Problem of Semantic Shift in Longitudinal Monitoring of Social Media: A Case Study on Mental Health During the COVID–19 Pandemic. In Proceedings of the 14th ACM Web Science Conference 2022, Barcelona Spain, 26 June 2022; pp. 208–218. DOI: https://doi.org/10.1145/3501247.3531566
[25] Miaschi, A., Dell'Orletta, F., 2020. Contextual and Non–Contextual Word Embeddings: an in–depth Linguistic Investigation, In the Proceedings of the 5th Workshop on Representation Learning for NLP, Association for Computational Linguistics, Online, 17 April; pp. 110–119. DOI: https://doi.org/10.18653/v1/2020.repl4nlp–1.15
[26] Giorgi, S., Nguyen, K.L., Eichstaedt, J.C., et al., 2022. Regional personality assessment through social media language. Journal of Personality. 90(3), 405–425. DOI: https://doi.org/10.1111/jopy.12674
[27] Ponta, L., Puliga, G., Manzini, R., 2021. A measure of innovation performance: the Innovation Patent Index. Management Decision. 59(13), 73–98. DOI: https://doi.org/10.1108/MD–05–2020–0545
[28] Smith, M., Agrawal, R.A., 2015. Comparison of Time Series Model Forecasting Methods on Patent Groups. MAICS. 1353, 167–173. Available from: https://ceur–ws.org/Vol–1353/paper_13.pdf (cited 24 March 2025).
[29] Noh, H., Jo, Y., Lee, S., 2015. Keyword selection and processing strategy for applying text mining to patent analysis. Expert Systems with Applications. 42(9), 4348–4360. DOI: https://doi.org/10.1016/j.eswa.2015.01.050
[30] Xiong, Q., Ma, Z., 2025. The Multi–dimensional Evaluation System Framework for New Equipment in Rail Transit. In: Zeng, X., Xie, X., Sun, J., et al. (eds.). International Symposium for Intelligent Transportation and Smart City (ITASC) 2025 Proceedings, Lecture Notes in Electrical Engineering. Springer Nature Singapore: Singapore, pp. 109–117.

DOI: https://doi.org/10.1007/978–981–96–4702–6_12

[31] Youn, S.J., Lee, Y.–J., Han, H.–E., et al., 2024. A Data Analytics and Machine Learning Approach to Develop a Technology Roadmap for Next–Generation Logistics Utilizing Underground Systems. Sustainability. 16(15), 6696. DOI: https://doi.org/10.3390/su16156696

[32] Lewis, R., 2015. A semantic approach to railway data integration and decision support [Doctoral Dissertation]. University of Birmingham: Birmingham, UK. Available from: http://etheses.bham.ac.uk/id/eprint/5959 (cited 24 March 2025).

[33] Tutcher, J., 2016. Development of semantic data models to support data interoperability in the rail industry [Doctoral Dissertation]. University of Birmingham: Birmingham, UK. Available from: https://etheses.bham.ac.uk/id/eprint/6774/1/Tutcher16PhD.pdf (cited 24 March 2025).

[34] Jägare, V., 2022. A challenge–driven framework for innovations in railways [Doctoral Dissertation]. Luleå University of Technology: Luleå, Sweden. Available from: https://www.diva–portal.org/smash/get/diva2:1701543/FULLTEXT02.pdf (cited 24 March 2025).

[35] Ariyachandra, M.R.M.F., Wen, Y., 2025. Building a holistic smart technology–centric value assessment framework for rail infrastructure. University of Strathclyde Publishing: Glasgow, Scotland. DOI: https://doi.org/10.17868/STRATH.00093275

[36] Hansen, C., Daim, T., Ernst, H., et al., 2016. The future of rail automation: A scenario–based technology roadmap for the rail automation market. Technological Forecasting and Social Change. 110, 196–212. DOI: https://doi.org/10.1016/j.techfore.2015.12.017

[37] Unger, S., Heinrich, M., Scheuermann, D., et al., 2023. Securing the Future Railway System: Technology Forecast, Security Measures, and Research Demands. Vehicles. 5(4), 1254–1274. DOI: https://doi.org/10.3390/vehicles5040069

[38] Karasev, O., Beloshitskiy, A., Shitov, E., et al., 2022. Integral Assessment of the Level of Innovative Development of the Railway Industry Companies. The Open Transportation Journal. 16(1), e187444782203141. DOI: https://doi.org/10.2174/18744478–v16–e2203141