AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(2026)

Contextual Embedding-Based Railway Communication Threat Detection with Transparent AI Models

Bulusu Rama[1*], Degala Sai Varun[2], Maroju Tharun Kumar[2], Suthrave Shashank[2]

[1]Associate Professor, [2]UG Student, [1,2]Department of Computer Science and Engineering (AI & ML), [1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.

*Correspondence: Bulusu Rama (bulusurama1967@gmail.com)

ABSTRACT

In modern railway networks, real-time monitoring and secure control communications are essential for ensuring operational safety, efficiency, and reliability. Critical operations such as signal control, train scheduling, and automated track switching generate large volumes of data that must be analysed instantly to prevent system failures or malicious intrusions. Traditional approaches relying on manual inspections or rule-based systems are often slow, error-prone, and inadequate for handling high-volume, dynamic communication data. To address these limitations, this research proposes a robust, automated anomaly detection framework for accurate real-time classification. Existing methods, including Decision Tree with Cost Complexity Pruning (DTCCP) and Deep Neural Decision Tree (DNDT), offer interpretability and moderate accuracy. However, DTCCP tends to overfit complex sequential data, while DNDT struggles to capture subtle contextual relationships, resulting in missed anomalies and false alarms. The proposed framework employs a RuleFit classifier (RF) that combines linear rules with decision tree logic, enhanced by semantic embeddings generated using Sentence Bi-directional Encoder Representations from Transformers (SBERT). This hybrid model effectively learns both hierarchical decision boundaries and contextual patterns within communication sequences. Performance evaluation is conducted using metrics such as accuracy, precision, recall, F1-score, confusion matrix, and Receiver Operating Characteristic (ROC) curves. Experimental results demonstrate that the RF-based approach significantly outperforms DTCCP and DNDT, achieving improved anomaly detection accuracy, reduced false positives, and reliable real-time classification of secure and insecure rail communications, thereby enhancing overall railway network safety and operational efficiency.

Keywords: Real-time monitoring, railway communication systems, anomaly detection, secure control communication, signal control, train scheduling, automated track switching, data analysis.

1. INTRODUCTION

The first steam locomotive undoubtedly heralded a transformative era, and since their inception in the early 19th century, railways have remained central to public transport. Recently, the potential of railways to alleviate road and air congestion and environmental challenges has brought them back into the spotlight. There has been a noticeable increase in rail traffic across Europe for both passenger and freight transport. Between 1990 and 2007, passenger kilometres increased by 28%, while freight ton kilometres increased by 15% in the EU-15 countries. Worldwide, rail networks carried more than 3.5 trillion passenger kilometres in 2019, with China, India, and Japan leading in passenger traffic [1].

Meanwhile, European railways recorded around 643 billion passenger kilometres in the same year. On the economic front, the global rail freight market was valued at $247.4 billion in 2020, with projections of growth to nearly $280 billion by 2026. In the railway sector, the integrity of train wheels is of paramount importance. Various defects such as wheel flats, spalling, chipping, and polygonization are common. Advances in sensor technology, driven by the integration of the Internet of Things (IoT) and Artificial Intelligence (AI), have revolutionized monitoring and diagnostics in

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

various industries, including construction, energy, healthcare, renewable energy, security, and transport [2].

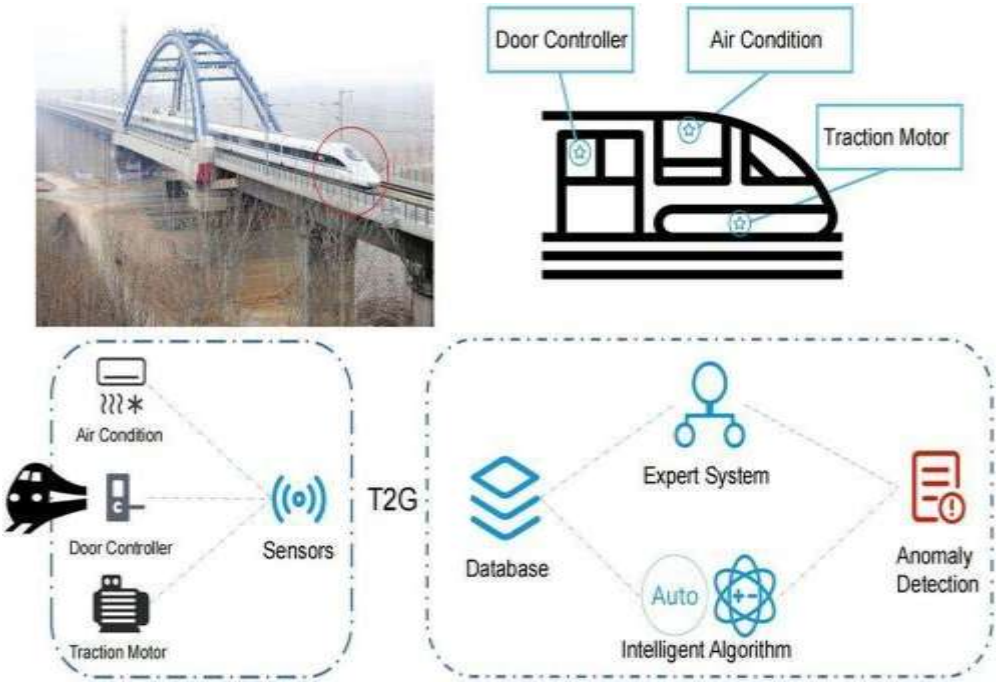

Fig. 1: Rail transit anomaly detection framework using sensor data and intelligent algorithms

Specifically in the railway context, a typical train bogie can house between 10 and 50 sensors. Among these, acoustic sensors are critical as they monitor vibrations and help in the early detection of anomalies in railway components. Such anomalies, if left unchecked, could lead to catastrophic consequences such as derailment. Traditional monitoring approaches often struggle with the complex patterns of anomalies manifested in the time-series data generated by these sensors, as shown in fig 1. However, the introduction of deep learning techniques to railway monitoring has yielded promising results, catalysing the development of models capable of processing, and interpreting vast amounts of data, in particular identifying unusual or unexpected events [3]. The plethora of anomaly detection methods still raises the need not only for their qualitative review and evaluation but also for their empirical performance evaluation. As a result, researchers are compelled to engage in comprehensive analyses that not only examine their design and current applications but, more importantly, delve into empirical evaluations of their performance [4].

This growing demand underscores the critical importance of refining and improving anomaly detection techniques to ensure their effectiveness when applied to real-world datasets. Furthermore, the dynamic nature of data landscapes requires a continuous and adaptive approach to the evaluation of anomaly detection methods, fostering continuous evolution in the field. However, to the best of their knowledge, there are few studies that have undertaken similar efforts. Thus, their study attempts to fill this gap by conducting a systematic literature review followed by experimental research [5].

2. LITERATURE SURVEY

El-Shafeiy, et al. [6] Introduced and applied a pioneering technology, Multivariate Multiple Convolutional Networks with Long Short-Term Memory (MCN-LSTM), to real-time water quality monitoring. MCN-LSTM is a cutting-edge deep learning technology designed to address the difficulty of detecting anomalies in complicated time series data, particularly in monitoring water quality in a real-world setting. The growing reliance on automated systems, the Internet of Things (IoT), and sensor networks for continuous water quality monitoring is driving the development and deployment

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

of the MCN-LSTM approach. As these technologies become more widely used, the rapid and precise identification of unexpected or aberrant data points becomes critical. Technical difficulties, inherent noise, and a high data influx pose significant hurdles to manual anomaly detection processes. The MCN-LSTM technique takes advantage of deep learning by integrating Multiple Convolutional Networks and Long Short-Term Memory networks.

Oh, et al. [7] Examined artificial intelligence applications for railway safety, mainly focused on deep learning approaches. This paper first introduces deep learning methods widely used for railway safety. Then, they investigated and classified earlier studies into four representative application areas: (1) railway infrastructure (catenary, surface, components, and geometry), (2) train body and bogie (door, wheel, suspension, bearing, etc.), (3) operation (railway detection, railroad trespassing, wind risk, train running safety, etc.), and (4) station (air quality control, accident prevention, etc.). They present fundamental problems and popular approaches for each application area. Finally, based on the literature reviews, they discuss the opportunities and challenges of artificial intelligence for railway safety.

Islam, et al. [8] Analysed The Internet of Railways (IoR) network is made up of a variety of sensors, actuators, network layers, and communication systems that work together to build a railway system. The IoR's success depends on effective communication. A network of railways uses a variety of protocols to share and transmit information amongst each other. Because of the widespread usage of wireless technology on trains, the entire system is susceptible to hacks. These hacks could lead to harmful behavior on the Internet of Railways if they spread sensitive data to an infected network or a fake user. For the previous few years, spotting IoR attacks has been incredibly challenging. To detect malicious intrusions, models based on machine learning and deep learning must still contend with the problem of selecting features. k-means clustering has been used for feature scoring and ranking because of this. To categorize attacks in two datasets, the Internet of Railways and the University of New South Wales, they employed a new neural network model, the extended neural network (ENN). Accuracy and precision were among the model's strengths. According to their proposed ENN model, the feature-scoring technique performed well. The most accurate models in dataset 1 (UNSW-NB15) were based on deep neural networks (DNNs) (92.2%), long short-term memory LSTM (90.9%), and ENN (99.7%). To categorize attacks, the second dataset (IOR dataset) yielded the highest accuracy (99.3%) for ENN, followed by CNN (87%), LSTM (89%), and DNN (82.3%).

Kim, et al. [9] Proposed the method uses a deep learning technique to train periodic data acquisition sequences, which is one of the common characteristics of IIoT. The trained model determined the sequence of packet is normal. The proposed technique can be applied without an additional analysis. The proposed method is expected to prevent security threats by proactively detecting cyberattacks. To verify the proposed method, a dataset was collected from the Korea Electric Power Control System. The model that defines normal behaviour based on the application layer exhibits an accuracy of 79.6%. The other model, defining normal behaviour based on the transport layer, has an accuracy of 80.9%. In these two models, most false positives and false negatives only occur when the abnormal packet is in a sequence.

Ahn, et al. [10] Proposed the method for detecting anomalies and characterizing failures for spacecraft attitude control systems is proposed. Herein, features are extracted from multidimensional time-series data of a simulation of the attitude control system. Then, the artificial neural network learning algorithms based on two types of generation models are applied. A Bayesian optimization algorithm with a Gaussian process is used to optimize the hyperparameters for the neural network to improve the performance. The performance is evaluated based on the reconstruction error through the algorithm using the newly generated data not used for learning as input data. Results show that the

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

detection performance depends on the operating characteristics of each sub mode in the operation scenarios and type of generation model. The diagnostic results are monitored to detect anomalies in operation modes and scenarios.

Kim, et al. [11] Suggested a hyper-parameter-tuned convolutional neural network (CNN) for multiclass unbalanced anomaly detection. A multiclass time series of anomaly data from a real-world cable-stayed bridge is used to test the 1D CNN model, and the dataset is balanced by supplementing the data as necessary. An overall accuracy of 97.6% was achieved by balancing the database using data augmentation to enlarge the dataset, as shown in the research. Song, et al. [12] Proposed a novel intrusion detection method that considers both the status of the networks and those of the equipment to identify if the abnormality is caused by cyber-attacks or by system faults. The proposed method is verified on a hardware-in-the-loop simulation platform of CBTC systems. Simulation results indicate that the proposed method has achieved 97.64% true positive rate, which can significantly improve the security protection level of CBTC systems.

Karapalidou, et al. [13] Discussed a new dataset that was made publicly available, collected from an industrial blower, is presented, analyzed and modeled using a Sequence-to-Sequence Stacked Sparse Long Short-Term Memory Autoencoder. Specifically, the right and left mounted ball bearing units were measured during several months of normal operational condition as well as during an encumbered operational state. An anomaly detection model was developed for the purpose of analysed the operational behavior of the two bearing units. A stacked sparse Long Short-Term Memory Autoencoder was successfully trained on the data obtained from the left unit under normal operating conditions, learning the underlying patterns and statistical connections of the data. The model was evaluated by means of the Mean Squared Error using data from the unit's encumbered state, as well as using data collected from the right unit. The model performed satisfactorily throughout its evaluation on all collected datasets. Also, the model proved its capability for generalization along with adaptability on assessing the behaviour of equipment like the one it was trained on.

Zhao, et al. [14] Proposed a novel solution to this problem based on measurement data. The proposed method combines a one-dimensional convolutional neural network (1DCNN) and a bidirectional long short-term memory network (BiLSTM) and uses particle swarm optimization (PSO), which is called PSO-1DCNN-BiLSTM. It enables the system to detect any abnormal activity in the system, even if the attacker tries to conceal it in the system's control layer. A supervised deep learning model was generated to classify normal and abnormal activities in an ICS to evaluate the method's performance. This model was trained and validated against the open-source simulated power system dataset from Mississippi State University. In the proposed approach, they applied several deep-learning models to the dataset, which showed remarkable performance in detecting the dataset's anomalies, especially stealthy attacks. The results show that PSO-1DCNN-BiLSTM performed better than other classifier algorithms in detecting anomalies based on measured data.

Choi, et al. [15] Described a deep learning-based anomaly detection method using time-series vibration and current data, which were obtained from endurance tests on driving modules applied in industrial robots and machine systems. Unlike traditional classification models that depend on labelled fault data for detection, acquiring sufficient fault data in real industrial environments is highly challenging due to various conditions and constraints. To address this issue, they employ a semi-supervised learning approach that relies solely on normal data to effectively detect abnormal patterns, overcoming the limitations of conventional methods. The performance of semi-supervised models was first validated using a statistical feature-based anomaly detection approach, from which the GCN-VAE model was adopted. By combining the spatial feature extraction capability of Graph

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

Convolutional Networks (GCNs) with the latent temporal feature modelling of Variational Autoencoders (VAEs), their method can effectively detect abnormal signs in the data, particularly in the lead-up to system failures. The experimental results confirmed that the proposed GCN-VAE model outperformed existing hybrid deep learning models in terms of anomaly detection performance in the pre-failure section.

3. PROPOSED SYSTEM

The proposed system is designed to automatically detect anomalies in rail control communications, distinguishing between secure and insecure messages in real time. At a high level, the system ingests raw communication logs, cleans and transforms the data into a suitable format, extracts semantic embeddings using deep sequence models and applies tree-based classifiers to identify potential security breaches. The architecture integrates preprocessing, feature extraction, classification, and visualization in a seamless workflow, as shown in fig. 2. enabling operators to monitor and act on anomalous messages efficiently.

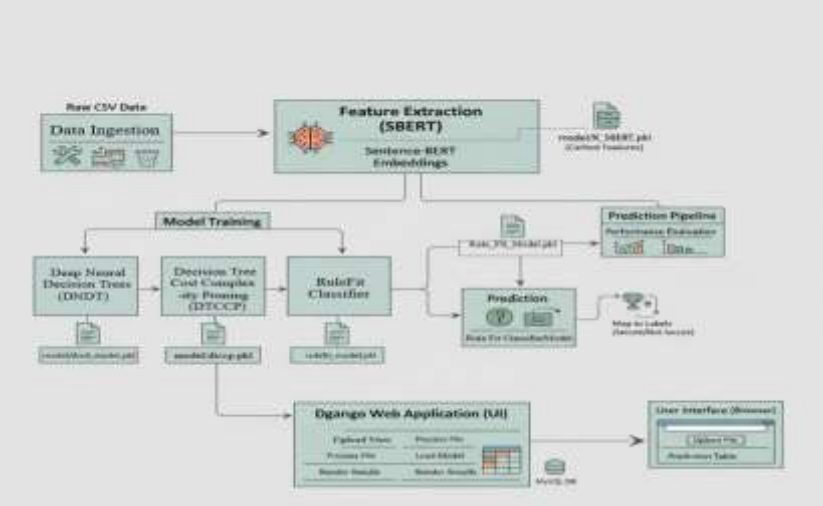

Fig. 2: Proposed system architecture of anomaly detection in rail control communications.

Data Acquisition and Loading: The system begins by collecting raw communication datasets from rail control systems. These datasets, typically in CSV or structured log formats, contain textual messages, status codes, and other metadata. The raw data is loaded into the system using a standardized upload mechanism, ensuring that it can handle multiple file formats and sizes. This initial step provides the foundation for subsequent preprocessing and modelling.

Data Preprocessing and Cleaning: Once the data is loaded, text-based features undergo cleaning and normalization. This involves:

- Lowercasing and removal of punctuation to standardize the text.
- Tokenization into individual words or terms.
- Stop word removal and lemmatization to reduce noise and focus on meaningful content. Numerical features, such as message timing or signal parameters, are preserved. The cleaned text and numerical features are then combined into a unified representation for downstream modelling. Target labels, such as “Secure” or “Not Secure,” are encoded into numeric form for classifier training.

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

Feature Extraction Using Deep Sequence Models: The cleaned textual data is transformed into dense, semantic representations using SBERT embeddings. SBERT captures contextual meaning from sequences, encoding complex patterns and relationships in communication messages. The system processes data in batches, leveraging GPU acceleration when available, and caches the resulting embeddings for faster training. This step ensures that subtle linguistic or structural anomalies in the messages can be detected effectively.

Classification with Tree-Based Models: The system employs multiple tree-based classifiers for robust anomaly detection:

1. **DNDT:** Integrate neural network representations with decision tree logic to handle high-dimensional embeddings.
2. **DTCCP:** Constructs a tree that splits features based on impurity reduction and prunes branches to prevent overfitting.
3. **RF:** Extracts interpretable rules from ensembles and linear combinations to capture non-linear relationships.

Each classifier is trained using an 80/20 train-test split, evaluated with metrics such as Accuracy, Precision, Recall, and F1-Score, and stored for future predictions. The system also calculates probability scores for risk assessment and visualization.

Model Evaluation and Visualization: Trained models are evaluated using comprehensive metrics dashboards. Visualizations such as bar charts of class-wise accuracy, confusion matrices, and performance tables allow operators to assess classifier reliability. This step provides insights into the types of anomalies detected and the strengths or limitations of each algorithm.

Prediction on New Data: When new communication logs are received, the system preprocesses the messages, generates SBERT embeddings, and applies the trained classifiers to predict security labels. Predictions are mapped to descriptive labels (“Secure” or “Not Secure”) for human interpretation. The system appends these predictions to the dataset, creating a ready-to-use report for railway operators.

Integration and Feedback Loop: Finally, the system can be integrated with a web interface (e.g., Django-based) to allow continuous monitoring. Users can upload datasets, visualize predictions, and export results. Feedback from real-world anomalies can be used to retrain models periodically, ensuring the system adapts to evolving communication patterns and maintains high detection accuracy.

4. RESULTS ANALYSIS

The results section presents the key findings of the study in a clear and organized manner. It highlights the main outcomes derived from data analysis, often using tables, charts, or summaries to support the observations. This section focuses only on what was discovered, without interpreting or explaining the reasons behind the findings. It may include patterns, trends, or significant differences identified during the research. The information is usually presented logically to help readers easily understand the outcomes. It provides a factual summary of the research results.

Fig. 3 displays 5 true negatives, 511 false positives, 483 false negatives, and 1 true positive, showing a slight improvement over DNDT in true positives but still poor Secure prediction. The matrices highlight varying model performance, with DTCCP excelling at Not Secure predictions but lacking Secure detection, while RF and DNDT show more balanced but imperfect results.

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

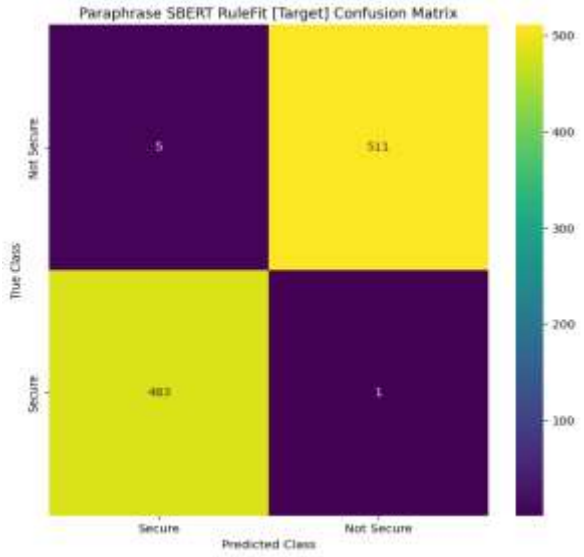

Fig. 3: Proposed Paraphrase SBERT RF confusion matrix.

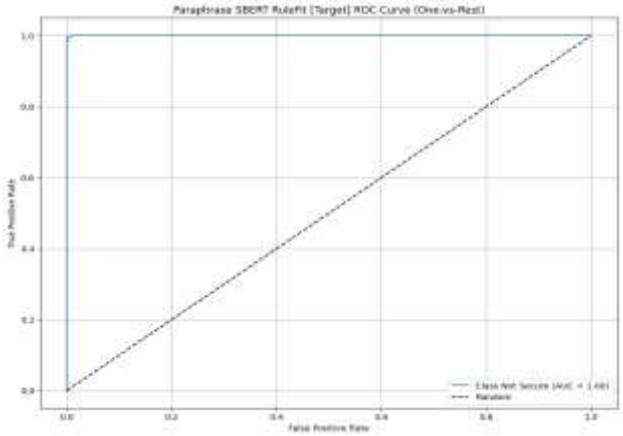

Fig. 4: Proposed Paraphrase SBERT Rule Fit ROC curve.

Fig. 4 presents ROC curves Plot for RF shows a near-perfect ROC curve hugging the top-left corner with an AUC of 1.00 for the "Not Secure" class, suggesting exceptional classification performance. The legend in each plot identifies the class ("Not Secure" or "Class Not Secure") and its AUC, with the RF model demonstrating the highest predictive accuracy among the three.

Fig. 5: Batch Prediction with uploading test data.

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

Fig. 5 illustrates the prediction interface where users can upload a test dataset in CSV format for analysis. Once the dataset is uploaded, the system processes it using the deep learning model to predict potential anomalies in communication data. The results are displayed in a detailed tabular format containing features such as transmission time, error rate, throughput, and data integrity. Each record is analysed, and the final column indicates whether the data is secure or not secure. This feature helps users efficiently evaluate the safety and reliability of railway communication systems based on model predictions.

Table 1: Performance evaluation obtained using DNDT, DTCCP and proposed RF.

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
DTCCP [Target]	48.4	24.200	50.000	32.615
DNDT [Target]	97.8	97.822	97.779	97.797
RF [Target]	99.4	99.390	99.412	99.400

Table 1 presents a comparative performance analysis of three SBERT-based anomaly detection models for railway control communications. The DTCCP model exhibited the lowest performance, with 48.4% accuracy, 24.2% precision, 50% recall, and a 32.615% F1-score, indicating limited capability in capturing complex communication patterns. The DNDT model performed significantly better, achieving 97.8% accuracy, 97.822% precision, 97.779% recall, and 97.797% F1-score, demonstrating strong effectiveness in identifying secure and insecure messages. The RF model achieved the highest performance, with 99.4% accuracy, 99.390% precision, 99.412% recall, and 99.400% F1-score, highlighting its ability to provide highly accurate and interpretable rule-based predictions for anomaly detection.

5. CONCLUSION

The proposed Paraphrase SBERT RF model represents a significant breakthrough in railway communication anomaly detection by integrating advanced semantic text embeddings with interpretable rule-based classification. By leveraging Paraphrase SBERT's deep contextual language understanding, the model captures intricate communication nuances that traditional models like the DNDT and DTCCP fail to recognize. Achieving an outstanding 99.4% accuracy, 99.39% precision, 99.41% recall, and 99.40% F1-score, the RF model demonstrates exceptional capability in accurately differentiating between "Secure" and "Not Secure" messages, ensuring reliability and safety within railway control networks. Its interpretable rule-based nature enhances transparency and accountability crucial for high-stakes, safety-critical domains while the SBERT embeddings contribute to a more profound semantic understanding of communication data. The model delivers superior performance, robustness, and interpretability, marking a pivotal step toward intelligent, secure, and context-aware railway communication systems.

REFERENCES

1. Máté, T.; Zwierczyk, P.T. Finite Element Analysis of Cracks Propagation in Railway Wheels. In Proceedings of the 33rd International ECMS Conference on Modelling and Simulation, Caserta, Italy.
2. Ministry of Land, Infrastructure and Transport; Safety Standards for Urban Railway Vehicles, Korea, 2023; Article 43, Table 3.

Received: 08-03-2026 | Accepted: 16-04-2026 | Published: 23-04-2026 | www.ajaccm.com

3. Lee, K.S.; Kim, J.W. A Study on Strategy of Condition Based Maintenance for Rolling Stock. J. Korean Soc. Railw. Korea 2024, 391–395.
4. Hyundai Rotem; Responsible for Safety and Operational Efficiency with Smart Technology, Hyundai Rotem's Health-Based Maintenance (CBM) System. Korea. 2024.
5. Kim, G.; Kim, H.; Koo, J. A Study on Cepstrum Analysis for Wheel Flat Detection in Railway Vehicles. J. Korean Soc. Saf. 2023, 31, 28–33.
6. El-Shafeiy, E.; Alsabaan, M.; Ibrahem, M.I.; Elwahsh, H. Real-Time Anomaly Detection for Water Quality Sensor Monitoring Based on Multivariate Deep Learning Technique. Sensors 2023, 23, 8613. https://doi.org/10.3390/s23208613
7. Oh, K.; Yoo, M.; Jin, N.; Ko, J.; Seo, J.; Joo, H.; Ko, M. A Review of Deep Learning Applications for Railway Safety. Appl. Sci. 2022, 12, 10572. https://doi.org/10.3390/app122010572
8. Islam, U.; Malik, R.Q.; Al-Johani, A.S.; Khan, M.R.; Daradkeh, Y.I.; Ahmad, I.; Alissa, K.A.; Abdul-Samad, Z.; Tag-Eldin, E.M. A Novel Anomaly Detection System on the Internet of Railways Using Extended Neural Networks. Electronics 2022, 11, 2813. https://doi.org/10.3390/electronics11182813
9. Kim, S.; Jo, W.; Kim, H.; Choi, S.; Jung, D.-I.; Choi, H.; Shon, T. Two-Phase Industrial Control System Anomaly Detection Using Communication Patterns and Deep Learning. Electronics 2024, 13, 1520. https://doi.org/10.3390/electronics13081520
10. Ahn, H.; Jung, D.; Choi, H.-L. Deep Generative Models-Based Anomaly Detection for Spacecraft Control Systems. Sensors 2022, 20, 1991. https://doi.org/10.3390/s20071991
11. Kim, S.-Y.; Mukhiddinov, M. Data Anomaly Detection for Structural Health Monitoring Based on a Convolutional Neural Network. Sensors 2023, 23, 8525. https://doi.org/10.3390/s23208525
12. Song, Y.; Bu, B.; Zhu, L. A Novel Intrusion Detection Model Using a Fusion of Network and Device States for Communication-Based Train Control Systems. Electronics 2022, 9, 181. https://doi.org/10.3390/electronics9010181
13. Karapalidou, E.; Alexandris, N.; Antoniou, E.; Vologiannidis, S.; Kalomiros, J.; Varsamis, D. Implementation of a Sequence-to-Sequence Stacked Sparse Long Short-Term Memory Autoencoder for Anomaly Detection on Multivariate Timeseries Data of Industrial Blower Ball Bearing Units. Sensors 2023, 23, 6502. https://doi.org/10.3390/s23146502
14. Zhao, X.; Zhang, L.; Cao, Y.; Jin, K.; Hou, Y. Anomaly Detection Approach in Industrial Control Systems Based on Measurement Data. Information 2022, 13, 450. https://doi.org/10.3390/info13100450
15. Choi, S.-H.; An, D.; Lee, I.; Lee, S. Anomaly Detection Based on Graph Convolutional Network–Variational Autoencoder Model Using Time-Series Vibration and Current Data. Mathematics 2024, 12, 3750. https://doi.org/10.3390/math12233750