Received 04 September 2025, accepted 01 May 2026, date of publication 30 September 2026
www.computingonline.net / computing@computingonline.net

*Print ISSN 1727-6209*
*Online ISSN 2312-5381*
*DOI 10.47839/ijc.25.3.4849*

# Proactive Railway Cable Infrastructure Protection Using Wireless Sensor Networks and Time Series Analytics

**MOYAHABO ROSSETT MOHLABENG[1], ERNEST MNKANDLA[1], AMINREZA KARAMOOZIAN[1,2]**

[1]Centre for Augmented Intelligence and Data Science, School of Computing, University of South Africa, Johannesburg, South Africa
[2]Institute of Artificial Intelligence, Shaoxing University, Shaoxing, Zhejiang, China

Corresponding author: Moyahabo Rossett Mohlabeng (e-mail: 41101820@mylife.unisa.ac.za).

**ABSTRACT** In recent years, protecting railway cable infrastructure has become increasingly important due to a surge in theft, vandalism, and system failures, many of which have led to costly disruptions and safety concerns. This study introduces a practical, data-driven approach that combines wireless sensor networks (WSNs) with time-series analysis to improve how we monitor and respond to such threats. The proposed method unfolds in three stages: first, a detailed analysis of sensor data collected across several stations; second, the use of clustering techniques to identify unusual patterns; and third, the application of forecasting models (ARIMA and DCS Gaussian) to anticipate potential cable sagging or theft. We tested the system using both real-world sensor deployments and publicly available datasets. Results show high reliability, with prediction errors as low as RMSE = 0.0212 and MAPE = 0.23%, and clear correlations between cable behavior across stations. Overall, the framework offers a timely and cost-effective solution for improving railway infrastructure security and contributes useful insights to the growing field of AI-based transport monitoring.

**KEYWORDS** Railway; cable; security; clustering; wireless sensor; time series; sagging.

## I. INTRODUCTION

The modern world relies heavily on sophisticated railway critical cable infrastructure technologies to execute various tasks such as energy, finance, food, transportation, health, and manufacturing. Globally, railway critical cable infrastructure has been the victim of failures, threats, and vandalism. These failures and vandalisms of railway cable infrastructure led to several accidents, damages, economic downgrades, and even human deaths. A railway cable can be defined as an essential part of the infrastructure that powers and links numerous railway systems [1].

Despite the severe failures, accidents, threats and vandalism occurring on critical cable railway infrastructure, inadequate safety and defense techniques are still available to protect it. Several countries have experienced a spike in cable theft, damage, threats, and vandalism over the last few years, negatively affecting their economies. As a result, cable theft, vandalism, and disruptions have led to many financial downgrades, accidents, and human deaths [2]. There was an increase in cable infrastructure theft and sabotage on the continent between 2013/2014 and 2014/2015, resulting in accidents and financial losses. In 2020/2021, cable thefts became more uncontrollable and affected all sizes of railway stations. As a result of these attacks, large amounts of railway critical cable infrastructure were destroyed, leading to financial downgrades and accidents [3]. Statistics of the types of cable incidents in several countries over a 10-year period is shown in Fig. 1.

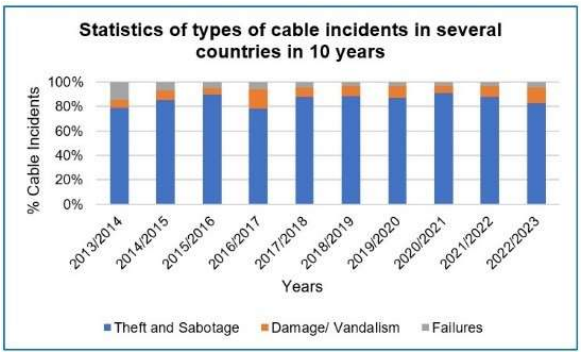



Figure 1. Statistics of the types of cable incidents in several countries over a 10-year period.

In several countries, cable theft and sabotage increased from 2016/2017 to 2018/2019, declined in 2019/2020, and increased again in 2020/2021, reaching 8727 incidents of cable

theft and sabotage, 987 incidents of damage and vandalism and 331 incidents of failures. South Africa, Kenya, Uganda, Benin, Botswana, and the UK contributed significantly to the failures, damage, vandalism, and stolen cable on the continent. The number of cable failures and vandalism decreased in 2020/2021 and gradually increased through 2022/2023. The railway cable infrastructure suffers significantly from these incidents and attacks [3], [4].

A data science-driven approach with wireless sensor network (WSN) integration is described in our work for railway cable infrastructure protection to address the shortcomings presented by current methods. Our model offers a new insight on a data science-driven approach with wireless sensor network (WSN) integration. The research helps to address the inadequacies in current techniques that result in delayed decision- making. A three-stage method is outlined below: Stage 1: time series qualitative and quantitative analysis for all stations, Stage 2: time series clustering analytics to complement Stage 1, and Stage 3: forecasting cable behavior using time-series.

Based on the above problem, the following research question is formulated: How do the proactive integration of WSNs and the design of a three-stage data analytics approach contribute to the dependability and effectiveness of securing critical cable railway infrastructure for addressing emerging security threats? The recent influx of vandalism and attacks on railway critical cable infrastructure has made it imperative for organizations to ensure their critical cable infrastructure is protected from attacks. This paper makes the following main contributions:

- Proactive data science-driven methodology integrating wireless sensor networks: The research introduces a new proactive approach to securing critical railway infrastructure, which offers a comprehensive solution to recent failures and emerging security threats. It integrates wireless sensor networks for real-time observations, which provides a robust foundation for enhanced monitoring and timely interventions in securing railway cable infrastructure.
- Three-stage analytics approach and experimental validation: The research presents a structured approach, including qualitative and quantitative analysis, time series clustering analytics, and forecasting cable behavior using time series, which address delayed decision-making gaps in existing techniques. Through the use of real-life sensors and publicly available data, the research demonstrates the dependability of the proposed framework, establishing correlations, covariances, and minimal forecasting errors as key indicators of its reliability and accuracy.
- Contributions to railway security literature as a reference guide: This study contributes significantly to the literature on securing railway infrastructure by highlighting defense gaps, bridging existing limitations, and providing valuable insights into the application of the Internet-of-Things (IoT), AI, and sustainable technologies in this domain. It serves as a beneficial reference guide to the research community and makes it a noteworthy contribution to industry practitioners.

This paper is organized in the following key elements in chronological order: Section 2 describes a theory of an overview of related infrastructure safety techniques, Section 3 explains the proposed methodology, and in Section 4, experimental evaluations are provided to simulate railway cable behavior across sensor stations and determine whether or not sagging or theft may occur. Section 5 presents the key findings of the proposed work and compares the existing techniques. Section 6 concludes with outcomes and concluding remarks.

## II. RELATED WORK

Several studies have explored the protection of railway cable infrastructure, focusing on threats such as theft, vandalism, and system failure. However, many of these methods face key limitations in scalability, responsiveness, and data interpretability. Table 1 provides a structured overview of existing infrastructure safety techniques and their shortcoming.

Andrysiak et al. [4] presented methods for addressing cable theft, vandalism, and disruptions to railway cable infrastructure, which are restricted to the specific railway environments and employ safety and defense strategies. These techniques involve a three-dimensional (3D) LiDAR-based approach, communications-based train control (CBTC) systems, and overhead contact line monitoring systems (OCMS). However, these techniques do not adequately safeguard railway infrastructure. The limitations of current methods include: vulnerability to Wi-Fi interference, inability to interpret large datasets, and inability to provide alerts quickly.

Current research on railway cable infrastructure protection is comprehensive but has several limitations that present opportunities for further exploration. The key primary drawbacks identified in existing studies are: (i) artificial intelligence/machine learning (AI/ML) methods struggle with interpreting large railway datasets but the literature suggests that future research should focus on improving the scalability and data handling capabilities of these technologies; (ii) sensor dependency highlights the need for more versatile and reliable sensor technologies that are less dependent on external factors. Future work should aim to develop more adaptable and scalable sensor networks that can accommodate varying operational conditions and requirements.

While WSNs have proven effective in domains such as healthcare, environmental monitoring, and industrial systems, their application to railway infrastructure presents distinct challenges in terms of environmental harshness, data volume, and network robustness. These insights inform the adaptation of WSN-based systems for dynamic and large-scale railway environments [12].

Aela et al. [13] noted that WSNs have some limitations, such as inefficiency and limited bandwidth, which make them vulnerable to attacks from third parties. Despite the shortfall, the benefits of WSNs over traditional observation methods include scalability and the ability to accommodate new nodes, improved real-time monitoring accuracy, and more fault tolerance. According to the literature, there is a great correlation with this study. Hence, the WSN integration aligns with the goals of this study in that real-time observations provide a robust foundation for enhanced monitoring and timely interventions in securing railway cable infrastructure.

Data science refers to the extraction of information and the production of abstract knowledge from the information set [14]. These provide an extension of the statistical capabilities of handling large amounts of data. Using data science, businesses

can forecast their future based on knowledge and prepare for unknown risks in advance. The data visualization method uses graphical elements such as charts, graphs, and maps to understand patterns, anomalies, and trends in data. Whilst statistical analysis collects, analyzes, and interprets an extensive amount of data to determine trends and develop valuable insights. As an example, these techniques have been successfully applied in economics and finance, social media, and health by supporting the detection and treatment of disease.

In contribution to the scalability, this research develops a rapid 3-stage time-series decision support system integrated into railway cable infrastructure protection for early warning to security agents. In this study, data science techniques were successfully used, such as machine learning, data visualization, statistical analysis techniques, to obtain a normal traffic model by studying historical traffic and comparing it with the normal model. Comparatively, ML detected abnormalities by establishing spatial values through training data.

## III.METHODOLOGY

This research was conducted using both qualitative and quantitative methods by collecting data from numerous sensors which were placed at critical hotspot stations to demonstrate the condition of the cables. Sensors at various hotspot stations were configured, connected, and tested at a base station. Data was then extracted from sensor devices and transmitted wirelessly to the base station. Sensor data was used in the proposed proactive data science-driven wireless sensor network integration framework.

ARIMA was selected for its proven effectiveness in time series forecasting, particularly when dealing with stationary data. Its strength lies in modelling temporal dependencies and identifying patterns in historical sensor data, such as temperature fluctuations or signal degradation over time. While, the DCS model has the ability to process real-time, distributed sensor data across multiple cable segments. It integrates sensor inputs to provide continuous monitoring of cable health, identifying localized issues such as overheating. While other models such as exponential smoothing, ARIMA, or machine learning-based approaches were considered, ARIMA and DCS were selected due to their complementary strengths.

### A. ESTABLISHING THE FRAMEWORK

This section presents a proactive data science-driven wireless sensor network (WSN) integration framework, which is shown in Fig. 2. This framework contains components with powerful capabilities, such as data collection, data preprocessing, feature selection, graphical analytics, covariance time-series, correlation time-series, and time-series clustering which is described further in the subsections.

### B. DATA GATHERING

In box 1 of Fig. 2, a diverse time series dataset from sensor stations is captured, including normal and abnormal cable readings. Sensor technologies are connected through the Internet of Things (IoT) to monitor data in real-time. The sensor data is stored at the centralized base and can be retrieved anytime and at any station [14].

### C. DATA PREPROCESSING

In box 2 of Fig. 2, data gathered is preprocessed. Data preprocessing consists of techniques such as cleaning, noise reduction, integration, transformation, drift correction, resolving redundancy, correct inconsistencies, normalization, and outlier removal or any operation performed on the data to enhance the quality, consistency, accuracy, identification, unbiased validation, completeness, and traceability of the input data [14].

### D. FEATURE SELECTION

To gain valuable insights from the sensor data in box 3 of Fig. 2, features and their associated importance are selected based on best practices [15]. This ultimately contributes to a more robust and interpretable railway CIP system. Voltage refers to identifying changes in voltage, whether low or high to prevent downtime on the railway cable, while displacement measures the position, movement, and dimensions (height, width, and thickness) of the railway cable, which can be either positive, negative, or zero. Temperature feature measures the degree of hotness, coolness, or movement of heat to improve cable quality, stability, quick response, and accuracy. Furthermore, movements of electric charge in a cable indicate a low or high current or its absence in the railway cable. A low current means that the cable has been stolen or cut; on the other hand, high current refers to the cable being intact [15].

### E. GRAPHICAL ANALYTICS

In diverse graph patterns and their associated meanings are explored, which unravel intricate relationships inherent in the cable data. It provides valuable insights into box 4 of Fig. 2 decision-making processes. In [16], cyclicality refers to the rise and fall leading to either cable theft or sagging between stations A to B., while seasonality can be defined as a regular repeating patterns at a particular time consistently between stations A to B. According to [16], trends can be defined as a long-term increase or decrease in the sensor data in stations A to B in a consistent direction. On the other hand, co-relationship indicates as positive or negative or no relationship between stations A to B overtime. Moreover, leading/lagging refers to lagging assesses the past state of stations A to Bovertime, while leading predicts the future state of Stations A to B overtime.

### F. TIME SERIES COVARIANCE

In fulfillment of box 5 in Fig. 2, the time series covariance is expressed for detecting cable sagging and theft in stations A - D using parameters such as Displacement = Disp; Voltage = Volt. Mathematically, the covariance COV (A, B) between two sensor time-series in stations A and B at different time steps $t_1$, $t_2$, ...$t_n$ is shown in Eqn. (1).

$$\mathrm{COV}(A,B)t_n = \frac{\sum(A_i - \overline{A})(B_j - \overline{B})}{N}, \quad (1)$$

where n is the number of data points in the time series, $A_i$ is the values of the A-variable in the time series, $B_j$ is the values of the B-variable in the time series, A is the mean (average) of the A-variable in the time series and B is the mean (average) of the B-variable in the time series. Based on covariance (COV) for two data series at stations A and B, the following guidelines are defined:

- Definition: For covariance adapted Principle 1: If there is a high positive COV between the data series for

stations A and B, then cable behaviors between stations A and

- Principle 2: If a correlation coefficient (Cc) = -1, then cable behaviors between the stations A and B = = perfectly, negatively correlated. This implies inconsistency and suspicion of cable sagging.
- Principle 3: If a correlation coefficient (Cc) = 0, then
- cable behaviors between the stations A and B = = independence. This implies the occurrence of a strange event.

### G. TIME SERIES CORRELATION COEFFICIENT

The time series correlation coefficient for detecting cable sagging and theft in stations A to B using parameters including Disp and Volt is accomplished in box 6 of Fig. 2. Pearson's Correlation Coefficient (Cc) between two-time series A and B is shown in Eqn. (2).

$$\text{Correlation Coefficient } (\rho) = \frac{\sum (Ai - \overline{A})(Bi - \overline{B})}{\sqrt{(\sum (Ai - \overline{A})^2)(\sum (Bi - \overline{B})^2)}}, \quad (2)$$

if $\Theta^{(t+1)}$ $\Theta^{(t)}$ $<\varepsilon$ then stop; else, return to E-Step. The Expectation maximization clustering (EMc) analysis identified two distinct data clusters across the stations.

$A_i$ and $B_i$ are the values of A and B at time point I; A is the mean (average) overtime of the A-variable and A is the mean (average) overtime of the B-variable. Here are the guidelines for deriving correlation coefficients (Cc) [17] from two data series at stations A and B:

- Definition: For correlation coefficient adapted Principle 1: If a correlation coefficient (Cc) = 1, then cable behaviors between stations A and B = = perfectly, positively correlated. This implies consistency and normalcy.
- Principle 2: If a correlation coefficient (Cc) = -1, then cable behaviors between the stations A and B = = perfectly, negatively correlated. This implies inconsistency and suspicion of cable sagging.
- Principle 3: If a correlation coefficient (Cc) = 0, then cable behaviors between the stations A and B = = independence. This implies the occurrence of a strange event.

### H. TIME SERIES CLUSTERING

This section entails expectation maximization clustering, which is referred to as an iterative approach for finding maximum-likelihood estimation of parameters overtime making use of the E-Step, M-Step and the convergence phase [18]. Since stations A and B are connected on a railway cable and coevolve over time, clusters are received simultaneously in time and are expected to be correlated.

Step 1: Initialization of EM using Eqn. (3).

$$\theta^t = (\mu_1^t, \mu_2^t, \dots, \mu_w^t) \quad (3)$$

Step 2: The E-Step provides the estimated and expected values of a hidden variable as indicated in Eqn. (4).

$$H(V_{iw}) = \frac{\exp[-\frac{(A_i - \mu_w^{(t)})^2}{2\sigma^2}]}{\sum_{j=1}^{w} \exp[-\frac{(A_i - \mu_w^{(t)})^2}{2\sigma^2}]} \quad (4)$$

Step 3: A new estimate of the parameters is computed using Eqn. (5).

$$\mu_w^{r+1} = \frac{\sum_{i=1}^{n} E(V_{iw}) A_i}{\sum_{i=1}^{n} E(V_{iw})} \quad , \quad (5)$$

where:

- $H(V_{iw})$ is the expected value of the hidden variables in the time series, the current condition distribution of B given A, $\theta^{(t)}$ is the current estimate of the parameters in the time series, σ is represented as the standard deviation overtime, $\theta^0$ is the estimate at the 0th iteration in time series.
- W is the current number of Gaussians, μ can be defined as the mean overtime, t is the number of iteration in the time series.

The convergence step indicates the stop condition checkup and if $||\theta^{(t+1)} - \theta^{(t)}|| < \varepsilon$ then stop; else, return to E-Step. Based on EM Clustering (EMc) for two data series clusters at stations A and B, definition 3.2 of section 3.7 is perfectly applicable.

### I. FORECASTING FUTURE CABLE BEHAVIOR

In box 8 of Fig. 2, a future forecast is an approach that uses historical data to formulate predictions regarding the direction of future trends using informed estimates. Forecasting provides cable infrastructure environments with the opportunity to anticipate the likelihood of future sagging or theft. Autoregressive integrated moving average (ARIMA) is a widely used and effective model that analyzes time series data to forecast future trends [18]. The model utilizes a mixture of past values, differences, and errors to determine patterns, trends, and seasonality in the data. Dynamic conditional score (DCS) Gaussian can be defined as a method for analyzing and forecasting various time series. It focuses on the heavy-tailed conditional distributions and varying location and/or scale of observations of data overtime [18].

**Table 1. An overview of related infrastructure safety techniques**

| Citations | Problem Addressed | Technique Used | Technique Description | Results | Drawbacks |
|---|---|---|---|---|---|
| [4] | Internal and external cyber security threats on the railway critical infrastructure could lead to railway accidents. | Communication-based Train Control (CBTC) system. | CBTC is an advanced railway signaling system that uses continuous, high-capacity wireless communication between trains and trackside equipment to manage train movements with precision. CBTC allows maximum capacity to be achieved and detection of traffic | The results indicate a lack of sufficient theoretical support for security issues of railway systems. | Susceptible to interference of Wi-Fi devices. |

| | | | | | |
|---|---|---|---|---|---|
| | | | demands while still maintaining safety standards. | | |
| [5] | Monitoring of critical infrastructure such as railway power lines to ensure resilience against current and emerging hazards. | LiDAR-based approach. | LiDAR is a remote sensing technology that uses laser pulses to measure distances between a sensor and surrounding objects with high precision. LiDAR is a light detection and ranging system uses a laser beam to detect objects nearby. | An accuracy of 97% results was reported for light pole extraction. | An inability to interpret large datasets. |
| [6] | Overhead contact line monitoring technology to ensure the reliability and safety of the high-speed electric railway system. | Overhead Contact Line Monitoring System (OCLMS) | An Overhead Contact Line Monitoring System (OCLMS) is a specialized diagnostic and surveillance platform designed to continuously assess the condition and performance of overhead electrification infrastructure used in railway systems. OCLMS detects wire variations throughout normal railway infrastructure operation. | The results show that OCMS can accurately measure and return in real time the data that are needed by the system. | Depends more on the heat source of an object to move the trigger sensor. |
| [7] | Field maintenance workload and safety pressure extremely affect the protection of railway operation. | Millimeter wave radar | Millimeter wave radar is a high-frequency sensing technology that operates within the electromagnetic spectrum's 30 to 300 GHz range, corresponding to wavelengths between 1 and 10 millimeters. It works by emitting short-wavelength radio waves and analyzing their reflections from surrounding objects to determine distance, speed, and angle. | As shown in the results, real-time monitoring, prediction, and early warning of natural disasters occurred on the eastern suburbs railway ring line in real-time. | Overlapping responsibilities and unclear interface of power. |
| [8] | The detection of the fallen objects in the railway infrastructure against accidents or threats. | Vision-based object detection. | Analyze image captions and object detection to determine object location, activity, and interactions between objects. | Results demonstrate that the detection of train state and object has been conducted vigorously. | Limited capability to maintain safety from the boarding platform. |
| [9] | Railway cables used for powering the locomotive are stolen which prevents the train and signaling from operating leading to accidents. | Wireless Sensor Network (WSN). | WSN refers to a network involving several sensors communicating with each other to share data, sense, compute, and communicate with one another. | The readings obtained only comprised a 1% error and greater accuracy of ultrasonic sounds. | The system is not flexible. |
| [10] | High-speed railway lines overhead catenary protection against breakages and failures. | Vibration-based diagnostics. | A diagnostic tool for detecting, monitoring, and preventing cable failures. | The results show a stage duration of t4 and an average value of the fixed acceleration of 0.5 m/s2. | Control of weak spots is even more tricky from continuous monitoring fulfillment. |
| [11] | Monitoring the safety of railway infrastructures. | Fiber Bragg Gratings (FBG) and Raman Distributed Temperature Sensing (RDTS) | Fiber Bragg Gratings (FBG) are optical sensing elements embedded within the core of an optical fiber, consisting of periodic variations in the refractive index. FBG is lightweight, easy to install and can sense multiple parameters, such as temperature, on one cable at the same time. While Raman Distributed Temperature Sensing is a fiber-optic technology that measures temperature continuously along the entire length of an optical fiber. | The results indicate that the proposed technology has been validated in railway laboratories and can detect defects. | Compatibility with maintenance procedures is unsustainable. |

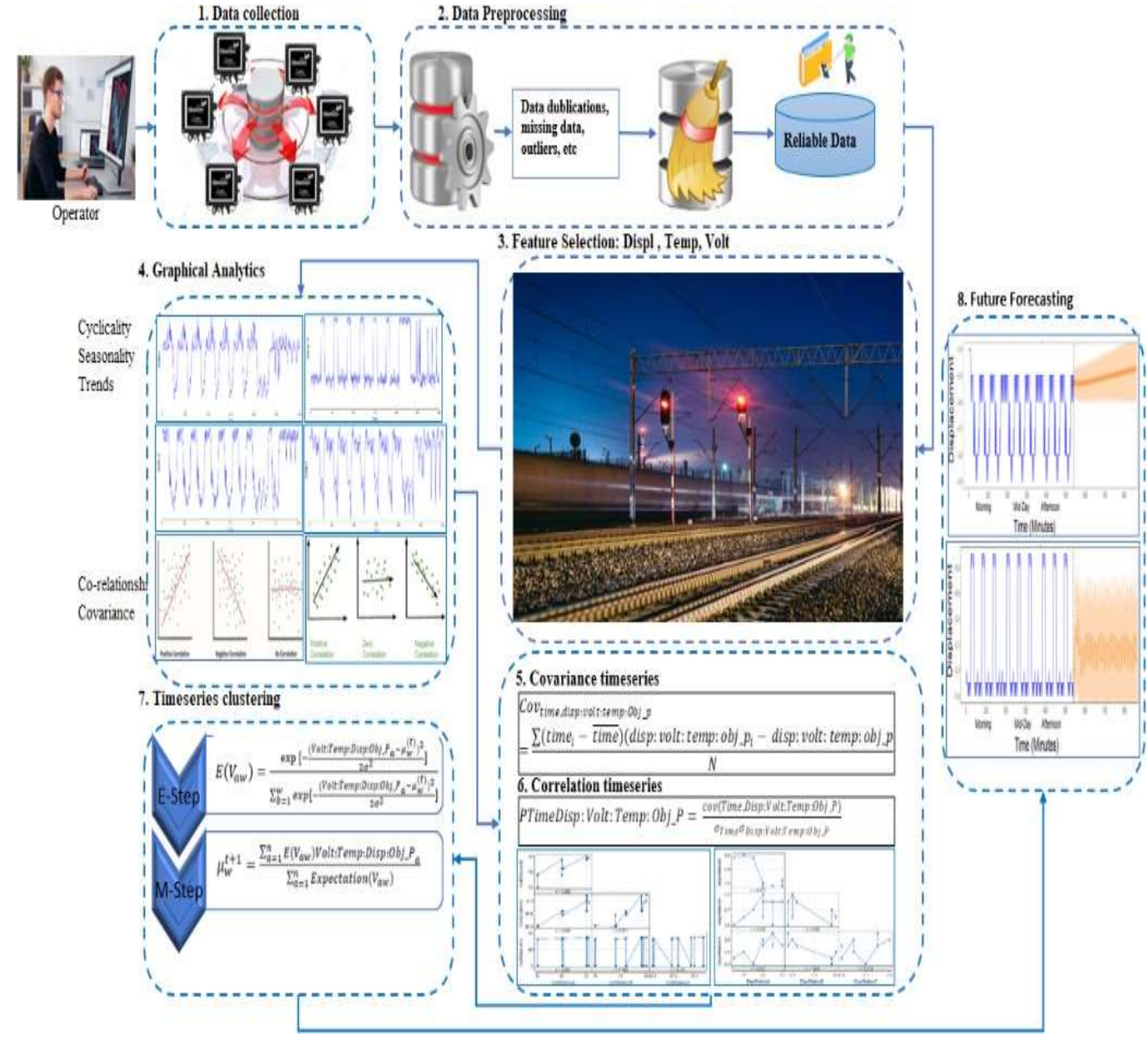



Figure 2. A proactive data science-driven wireless sensor network (WSN) integration framework.

## IV EXPERIMENTAL EVALUATIONS

This study investigates the effects of the inadequacies in current techniques that result in delayed decision-making. While earlier studies explored the impact of existing defensive techniques, they did not explicitly address its influence on delays in decision-making processes.

### *A. EXPERIMENTAL SETUP*

An experimental prototype was undertaken in a procedural manner which included four sensors deployed at various hotspot stations. These sensors were properly configured, mounted, connected, and tested to a base station at various hotspot stations. As a result, sensor devices collected data from numerous stations and transmitted it wirelessly to the base station. The sensors were mounted at various stations and labeled as sensors 1 – 4 located at different physical stations named stations A – D to demonstrate the condition or state of the cables.

**Table 2. Sensors data for data for cable stations A - D of voltages**

| Day | Periods | Times in minutes | Station A | Station B | Station C | Station D |
|---|---|---|---|---|---|---|
| 1 | Morning | 30 | 90 | 80 | 70 | 70 |
| | Mid-Day | 240 | 90 | 80 | 70 | 70 |
| | Noon | 400 | 90 | 80 | 70 | 70 |
| 2 | Morning | 30 | 89 | 80 | 70 | 70 |
| | Mid-Day | 240 | 90 | 79 | 68 | 70 |
| | Noon | 400 | 90 | 80 | 70 | 70 |
| 3 | Morning | 30 | 90 | 79 | 68 | 0 |
| | Mid-Day | 240 | 90.1 | 79 | 68 | 0 |
| | Noon | 400 | 90 | 79.8 | 68 | 0 |
| 4 | Morning | 30 | 89 | 80 | 68 | 70 |
| | Mid-Day | 240 | 85 | 80 | 79 | 69 |
| | Noon | 400 | 90 | 80 | 68 | 69 |
| 5 | Morning | 30 | 80 | 79 | 68 | 70 |
| | Mid-Day | 240 | 89 | 80 | 68.2 | 70 |
| | Noon | 400 | 90 | 80 | 68 | 70 |

Figure 3. Experimental prototype setup as railway cable infrastructure.

An experiment was conducted between 12 May and 16 May 2025, from 8:00 AM to 17:00 PM, and monitored throughout. A cable cut was intentionally introduced on Wednesday, 14 May 2025, between 8:00 AM and 17:00 PM to demonstrate cable theft for testing purposes. The experiment was conducted to fulfill the following objectives: (a) to ensure daily exposure of the cables to wind; (b) to ensure the disconnection at Station D on May 14, 2023; (c) to ensure the daily presence of objects on the cables and to observe other environmental behaviors on the railway cables. Approximately 32 Kbytes of data are captured, and the host performs quality checks. A sample of approximately 540 data instances was captured on an average day. Fig. 3 shows experimental prototype setup as railway cable infrastructure.

Various statistical tools used include SPSS, Orange data mining, WEKA, and Timeseries Lab. A sample dataset extracted from sensor devices contains attributes such as displacement and voltage, which are used to predict cable patterns. Table 2 shows sensor data for cable stations A-D of voltage and sensor data for cable stations A-D of displacement is indicated in Table 3.

Publicly available data: In Table 4, public sensor data for cable stations Darlington Melrose (DM), Rise Carr Ianson (RCI), and Rise Carr substations (RCD), while public sensor data for cable stations DM, RCI and RCD of current are shown in Table 5. As reflected in Tables 4 - 5, the days, periods, and times were transformed from publicly available data, whereby the Days 1 – 5 are referred to the day value taken from the date, i.e., 01/9/2014, 12:01:00 PM, where there are various periods linked with times such as 30 from morning until 400 in the afternoon.

### *B. EXPERIMENT I: ON REAL-LIFE RAILWAY CABLE PROTOTYPE TEST BEDS*

Here, it is important to determine whether railway cables sag or are stolen using time-series anomaly detection pattern behavior across sensor stations. Security agents would require answers to the following questions: (a) What patterns of behaviors are revealed as red flags of cable sagging or theft? (b) Is there a co-relationship behavior of one cable station to another? (c) Is there any seasonality or trend behavior? (d) What observations are suspected as abnormal? (e) From the patterns of behaviors, what is likely to happen next? A rapid 3-stage time series decision system in the next sections is integrated into the railway cable infrastructure protection to guide in answering the questions.

1) Stage 1: time series qualitative and quantitative analysis for all stations: A graphical visual analysis of stage 1 demonstrates that patterns are observed over time for the sensors' displacement, and voltage parameters. Fig. 4 shows graphical analytics of displacement and voltage for Stations A to B and Stations B to C. In Fig. 4a, one can observe high and low data movements at Stations A to B from 0 to 500 minutes of displacement that implies cyclical behavior in Stations A to B. Fig. 4b shows that there is a mixture of high, medium, and low data movements at Stations A to B from 0 to 500 minutes of voltage. This infers cyclical and seasonality behavior in Stations A to B. The Preliminary Decision for Zero Tolerance as: (a) Cable normalcy or sagging behavior is suspected in Stations A to B, (b) Security agents may need to respond to the affected site, but the intermediate DSS may be required. In Fig. 4c, there is high data and a bit of low data movement at a particular time from 0 to 500 minutes of displacement at Stations B to C. It seems there is cyclical behavior from Stations B to C.

**Table 3. Sensor data for cable stations A -D of displacement**

| Day | Periods | Times in minutes | Station A | Station B | Station C | Station D |
|---|---|---|---|---|---|---|
| 1 | Morning | 30 | 3.0 | -1.5 | -2.6 | -3 |
| | Mid-Day | 240 | 2.9 | -1.5 | -2.5 | 2.9 |
| | Noon | 400 | 3.0 | -1.4 | -2.5 | -3 |
| 2 | Morning | 30 | 3.1 | -1.5 | -2.5 | -3 |
| | Mid-Day | 240 | 3.0 | -1.5 | -2.5 | 3.1 |
| | Noon | 400 | 3.0 | -1.5 | -2.9 | -3 |
| 3 | Morning | 30 | 3.2 | -1.5 | -2.8 | -3 |
| | Mid-Day | 240 | 3.0 | -1.5 | -2.9 | 2.9 |
| | Noon | 400 | 2.9 | -1.0 | -3.0 | 2.9 |
| 4 | Morning | 30 | 2.8 | -0.5 | -2.8 | 2.9 |
| | Mid-Day | 240 | 2.8 | -0.5 | -2.8 | 2.9 |
| | Noon | 400 | 2.8 | -0.5 | -2.9 | -3 |
| 5 | Morning | 30 | 2.8 | -0.5 | -2.9 | 3.1 |
| | Mid-Day | 240 | 2.8 | -1.5 | -2.8 | 3.2 |
| | Noon | 400 | 2.7 | -0.5 | -2.8 | -3 |

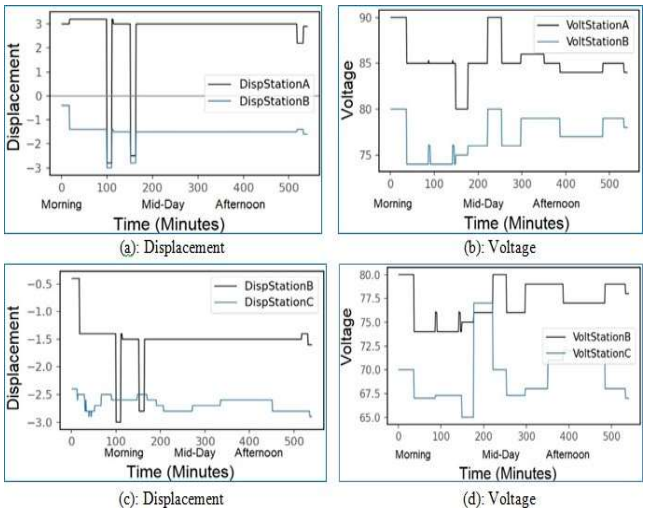



Figure 4. Graphical analytics of displacement and voltage for stations A to D.

**Table 4. Public sensor data for cable stations DM, RCI and RCD of temperature**

| Days | Periods | Times in minutes | Temp -DM | Temp -RCI | Temp -RCD |
|---|---|---|---|---|---|
| 1 | Morning | 30 | 0 | 0 | 0 |
| | Mid-day | 240 | 14.1 | 17 | 0 |
| | Noon | 400 | 14 | 17 | 14.3 |
| 2 | Morning | 30 | 0 | 0 | 0 |
| | Mid-day | 240 | 14 | 0 | 0 |
| | Noon | 400 | 14 | 17 | 14.3 |
| 3 | Morning | 30 | 0 | 17 | 0 |

| | | | | | |
|---|---|---|---|---|---|
| | Mid-day | 240 | 0 | 0 | 14.3 |
| | Noon | 400 | 14 | 17 | 14.3 |
| 4 | Morning | 30 | 0 | 0 | 0 |
| | Mid-day | 240 | 0 | 17 | 0 |
| | Noon | 400 | 14 | 17 | 14.3 |
| 5 | Morning | 30 | 14.3 | 17.1 | 0 |
| | Mid-day | 240 | 0 | 0 | 14.3 |
| | Noon | 400 | 14 | 17 | 14.3 |

**Table 5. Public sensor data for cable stations DM, RCI and RCD of current**

| Days | Periods | Times in minutes | Current -DM | Current -RCI | Current -RCD |
|---|---|---|---|---|---|
| 1 | Morning | 30 | 69 | 0 | 0 |
| | Mid-day | 240 | 67 | 81 | 0 |
| | Noon | 400 | 73.7 | 81 | 31 |
| 2 | Morning | 30 | 68.4 | 0 | 0 |
| | Mid-day | 240 | 70.4 | 0 | 30 |
| | Noon | 400 | 67.6 | 80 | 30 |
| 3 | Morning | 30 | 68.1 | 0 | 0 |
| | Mid-day | 240 | 68.7 | 0 | 0 |
| | Noon | 400 | 69.6 | 78.6 | 30 |
| 4 | Morning | 30 | 69.6 | 78.6 | 0 |
| | Mid-day | 240 | 68.6 | 0 | 0 |
| | Noon | 400 | 67.8 | 77.9 | 31 |
| 5 | Morning | 30 | 68.6 | 0 | 31 |
| | Mid-day | 240 | 68.2 | 77.8 | 32 |
| | Noon | 400 | 65.8 | 77.4 | 0 |

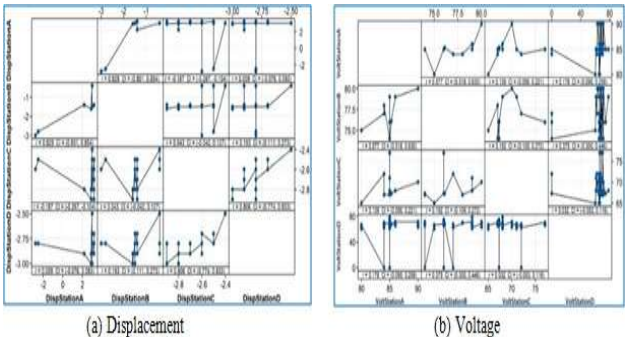
One can spot some high and low data movement from 0 to 500 minutes of voltage at Stations B to C in Fig. 4d. This implies cyclical and seasonality behavior from Stations B to C. The Preliminary Decision for Zero Tolerance is as: (a) Cable sagging is suspected in Stations B to C, (b) Security agents may respond to the affected site, however, a higher DSS may be required. For Covariance (COV) and correlation (CORR) analytics, Table 6 shows the displacement covariance (COV) matrix, while Table 7 indicates voltage covariance (COV). Fig. 5 illustrates the correlation (CORR) coefficient matrix of (a) displacement and (b) voltage.

**Table 6. Displacement covariance (COV) matrix**

| | Station A | Station B | Station C | Station D |
|---|---|---|---|---|
| Station A | 1.3295011 | | | |
| Station B | 0.3360896 | 0.1236441 | | |
| Station C | -0.0243387 | 0.0017087 | 0.0127565 | |
| Station D | 0.0011 | 0.0074614 | 0.0100011 | 0.0120762 |

**Table 7. Voltage covariance (COV)v matrix.**

| | Station A | Station B | Station C | Station D |
|---|---|---|---|---|
| Station A | 4.67317 | | | |
| Station B | 2.58194 | 4.28903 | | |
| Station C | 0.90934 | 1.19903 | 9.11157 | |
| Station D | 9.9185 | 20.08832 | 2.46806 | 667.81927 |



(a) Displacement (b) Voltage

Figure 5. Correlation (CORR) coefficient matrix of (a) displacement and (b) voltage.

In Table 6, the result of the covariance value for displacement between Stations A to B shows COV is 0.3360896. This covariance infers weak consistency and normalcy in Stations A to B. The result of the correlation coefficient value for displacement amongst Stations A to B presents CORR is 0.829 as depicted in Fig. 5a. This correlation infers stronger positive continuity in Stations A to B. In Table 7, the result of the covariance value for the voltage between Stations A to B shows COV as 2.58194. This covariance implies strong consistency and normalcy in Stations A to B. The result of the correlation coefficient value for voltage amongst Stations A to B indicates CORR is 0.577 as shown in Fig. 5b. This correlation infers strong positive continuity in Stations A to B. The Preliminary decision for zero tolerance is indicated as: (a) Strong cable normalcy is suspected in Stations A to B, Security agents may need not worry, but complementary DSS is crucial.

Moreover, in Table 6, the result of the covariance value for displacement between Stations B to C indicates that COV is 0.0017087. This covariance infers weak consistency and normalcy in Stations B to C. The result of the correlation coefficient value for displacement amongst Stations B to C shows CORR is 0.043 as shown in Fig. 5a. This correlation infers a weak positive continuity in Stations B to C. In Table 6, the result of the covariance value for the voltage between Stations B to C indicates that COV is 1.19903. This covariance implies strong consistency and normalcy in Stations B to C. The result of the correlation coefficient value for voltage amongst Stations B to C illustrates that CORR is 0.192 as shown in Fig. 5b. This correlation infers weak positive continuity in Stations B to C. So, the preliminary decision for zero tolerance is outlined as: the weak cable normalcy or sagging is suspected in Stations B to C and the security agents may need to respond to the affected sites, but complementary DSS is essential.

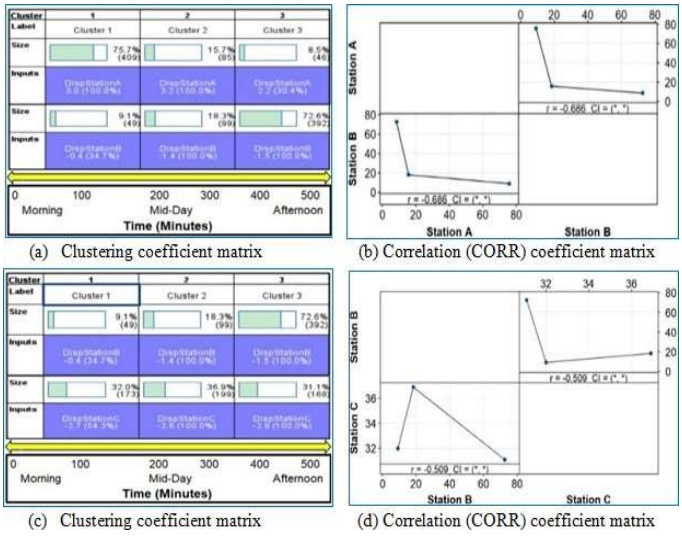
2) Stage-2: time series clustering analytics complementing stage 1: Stage 2 demonstrates the time series clustering analytics complementing the cable behavior monitoring in Stage 1. The objective here is to find any possible consistency of clustered displacement and voltage readings in sizes overtime. Having the results from various clusters allows some inferences on the railway operatives. Fig. 6 depicts the clustering and correlation (CORR) coefficient matrix of displacement for stations A to B and B to C for further discussion. In Fig. 6a, the concurrent displacement value sizes for Stations A to B are: C1 = 75.7%, C2 = 15.7% and C3 = 8.5% and C1 = 9.1%, C2 = 18.3% and C3 = 72.6% respectively. The result of the correlation coefficient value for displacement amongst Stations A to B indicates CORR is 0.686, as illustrated in Fig. 6b. This implies strong negative continuity in Stations A to B. The Intermediate decision with zero tolerance for security agents are as follow: (a) In Stations A to B, there is strong negative continuity for displacement, cable sagging is suspected [19]. Therefore, cable sagging is suspected in Stations A to B, security agents should respond to the affected site, but an advanced DSS stage is required.

In Fig. 6c, simultaneous displacement value sizes for Stations B to C are: C1 = 9.1%, C2 = 18.3% and C3 = 72.6% and C1 = 32.0%, C2 = 36.9% and C3 = 31.1% separately. The result of the correlation coefficient value for displacement amongst Stations B to C indicates CORR is 0.509, as depicted in Fig. 6d. This implies strong negative continuity in Stations B to C. So, the intermediate decision with zero tolerance for

security agents could be indicated as: (a) There is a strong negative continuity for displacement in Station B to C, cable sagging is suspected, (b) Therefore, cable sagging is suspected in Stations B to C, (c) Security agents need to respond to the affected site, but an advanced DSS stage is required.

3) Stage-3: time series forecasting cable behaviors complementing stage 2: Stage 3 focuses on forecasting cable behavior for predicting possible cable sagging, theft, and/or normal in the future using autoregressive integrated moving average (ARIMA) and dynamic conditional score (DCS) Gaussian using Timeseries Lab software. Security agents would particularly like to answer these questions (i) Is there any seasonality or trend behavior? (ii) What observations are suspected as abnormal? (iii) From the patterns of behaviors, what is likely to happen next?

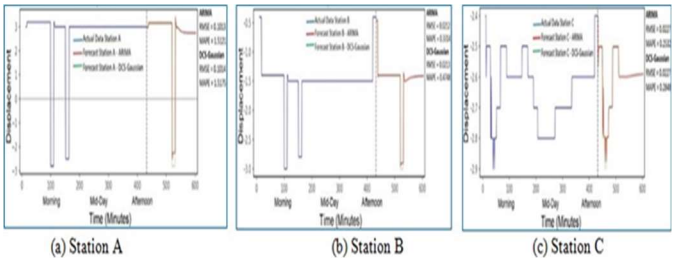
Table 8 indicates the performance of the predictive model of the displacement for stations A – D, while Fig. 8 depicts forecast for displacement of (a) station A and (b) station B. In Fig. 8c, forecast for displacement in station C is demonstrated. The training dataset used is 80%, while the test dataset is 20% of displacement for Stations A – D.




(a) Clustering coefficient matrix (b) Correlation (CORR) coefficient matrix

(c) Clustering coefficient matrix (d) Correlation (CORR) coefficient matrix

Figure 6. Clustering and correlation (CORR) coefficient matrix: (a), (b) displacement for stations A to B and (c), (d) displacement for stations B to C




(a) Station A (b) Station B (c) Station C

Figure 7. Forecast for the displacement of (a) station A (b) station B and (c) Station C.

**Table 8. Performance of predictive model of displacement for stations A – D.**

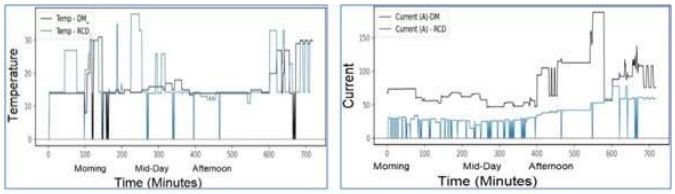
| | ARIMA | | DCS – Gaussian | |
|---|---|---|---|---|
| | RMSE | MAPE | RMSE | MAPE |
| Station A | 0.1013 | 1.5121 | 0.1014 | 1.5175 |
| Station B | 0.0212 | 0.3314 | 0.0213 | 0.4748 |
| Station C | 0.0227 | 0.2532 | 0.0227 | 0.2848 |
| Station D | 0.0309 | 0.2327 | 0.0307 | 0.2594 |

In Station A, a displacement's root mean squared error (RMSE) value in ARIMA is 0.1013 and mean absolute percentage error (MAPE) = 1.5121, whereas in DCS Gaussians, the displacement's RMSE value = 0.1014 and MAPE value = 1.5175. As shown in Table 8, Stations B to D also show ARIMA and DCS-Gaussian's RMSE and MAPE values. It appears that ARIMA has better performance than DCS- Gaussians since its error values are low.

In Fig. 7a, one can observe high and a bit low data movement which displays cyclic behavior at Station A from 0 to 420 minutes of displacement. This implies a likelihood of future cable normalcy behavior in Station A. Security agents may need not to respond in nearest time to Station A. It seems that there is a high and low data movement which indicates cyclic behavior at Station B from 0 to 420 minutes of displacement indicated in Fig. 7b. This infers a likelihood of future cable normalcy behavior in Station B. The security agents may need not to respond at the nearest time to Station B. In Fig. 7c, there is a mixture of a high and a bit of a low data movement which illustrates cyclic behavior from 0 to 420 minutes of displacement. It seems that there might be a possibility of future cable sagging behavior in Station C. The security agents may need to respond at the nearest time to Station C.

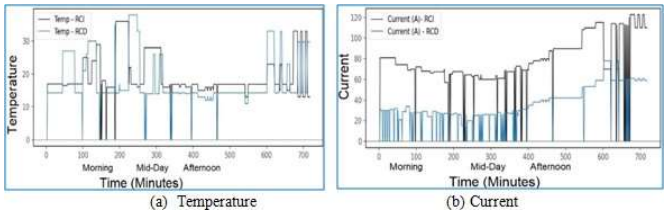
### *C. EXPERIMENT II: VALIDATING PROPOSED APPROACH ON PUBLICLY AVAILABLE CABLE STATIONS*

1) Stage 1: time series qualitative and quantitative analysis for all stations: A graphical visual analysis of stage 1 depicts that patterns are observed over time for the sensors' temperature and current for Darlington Melrose (DM), Rise Carr Ianson (RCI), and Rise Carr substations (RCD). Fig. 8 demonstrates stations DM to RCD of (a) temperature and (b) current. Stations RCD to RCI of (c) temperature and (d) current is generated using Timeseries Lab software to reveal the behavior of the railway cables over time and is depicted in Fig. 9. Fig. 8a shows a mixture of high and low data movements from 0 to 710 minutes for temperature from Stations DM to RCD. This implies cyclical and seasonality behavior in Stations DM to RCD. There seems to be a bit of high and low data movement from 0 to 710 minutes for current from Stations DM to RCD in Fig. 8b.




(a) Temperature (b) Current

Fig. 8. (a) temperature and (b) current of Stations DM to RCD.




(a) Temperature (b) Current

Figure 9. (a) temperature and (b) current of Stations RCD to RCI.

**Table 9. Temperature covariance (COV) matrix stations DM, RCI and RCD.**

| | Temp-DM | Temp-RCI | Temp-RCD |
|---|---|---|---|
| Temp-DM | 34.80443 | | |
| Temp-RCI | 10.2503 | 43.1513 | |
| Temp-RCD | 10.5373 | 2.225 | 57.4236 |

This infers cyclical and seasonality behavior in Stations DM to RCD. So, the preliminary decision for zero tolerance is considered as: Cable sagging is suspected in Stations DM to RCD and security agents may respond to the affected sites, but the intermediate DSS may be needed. In Fig. 9a, there are low and high data movements from 0 to 710 minutes of temperature for Stations RCD to RCI. This implies cyclical and seasonality behavior in Stations RCD to RCI. One can spot high and low data movements from 0 to 710 minutes for current from Stations RCD to RCI in Fig. 9b. This implies cyclical and seasonality behavior in Stations RCD to RCI. So, the preliminary decision for zero tolerance is considered as: Cable sagging is suspected in Stations RCD to RCI and the security agents may respond to the affected site, but the intermediate DSS is vital. Table 9 shows temperature covariance (COV) matrix stations DM, RCI and RCD. The result of the covariance value for temperature between Stations RCD to RCI shows that COV is 2.225. This implies weak consistency and normalcy in Stations RCD to RCI.

The result of the covariance value for temperature between Stations DM to RCD shows COV is computed as 10.5373. This covariance implies strong consistency and normalcy in Stations DM to RCD. In Fig. 10, the result of the correlation coefficient value for temperature amongst Stations DM to RCD indicates CORR is 0.236. This correlation infers a weak positive continuity between Stations DM to RCD. In Fig. 10, the result of the correlation coefficient value for temperature amongst Stations DM to RCD indicates CORR.

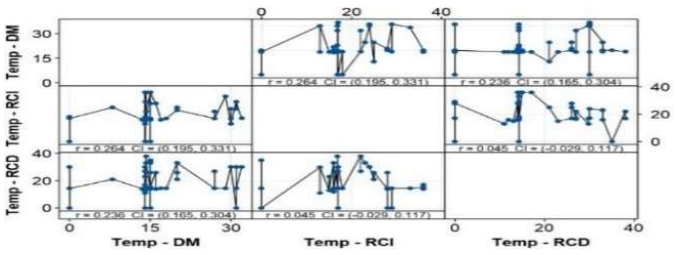
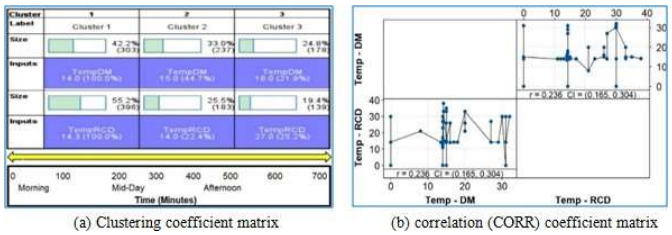



Figure 10. Temperature correlation coefficient matrix for stations DM, RCI and RCD.

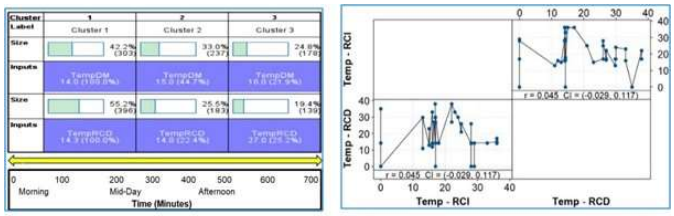



(a) Clustering coefficient matrix (b) correlation (CORR) coefficient matrix

Figure 11. Temperature for stations DM to RCD of (a) clustering and (b) correlation coefficient Matrix.

This correlation infers a weak positive continuity between Stations DM to RCD. The result of the correlation coefficient value for temperature amongst Stations RCD to RCI indicates CORR is 0.045 as depicted in Fig. 10. This implies a weak positive continuity in Stations RCD to RCI. The preliminary decision with zero tolerance could be taken as weak cable normalcy or sagging is suspected in Stations RCD to RCI and the security agents must rush to the affected sites, but complementary DSS is crucial.

2) Stage 2: time series clustering analytics complementing the cable behavior in stage 1: In Stage 2, time series clustering analytics is demonstrated to complement the cable behavior monitoring in Stage 1. Fig. 11 demonstrates temperature for stations DM to RCD of (a) clustering and (b) correlation coefficient matrix. The concurrent temperature value sizes for Stations DM to RCD are C1 = 42.2%, C2 = 33.0% and C3 = 24.8%, and C1 = 55.2%, C2 = 25.5%, and C3 = 19.4% respectively. The result of the correlation coefficient value for temperature amongst Stations DM to RCD indicates CORR is 0.236. This implies a weak positive continuity in Stations DM to RCD. The intermediate decision for security agents can be considered as weak cable normalcy is suspected in Stations DM to RCD and the security agents may respond to the affected site, but advanced DSS is vital. Fig. 12 depicts the temperature for stations RCD to RCI of (a) clustering and (b) correlation (CORR) coefficient matrix. In stations RCD to RCI, the simultaneous temperature value sizes are: C1 = 42.2%, C2 = 33.0% and C = 24.8% and C1 = 55.2%, C2 = 25.5%, and C3 = 19.4% separately as demonstrated in Fig. 12a. The result of the correlation coefficient value for temperature amongst Stations RCD to RCI presents CORR is 0.045 as shown in Fig. 12. This correlation infers weak positive continuity in Stations RCD to RCI and the intermediate decision can be drawn as weak cable normalcy is suspected in Stations RCD to RCI and the security agents may respond to the affected site, but advanced DSS is vital.

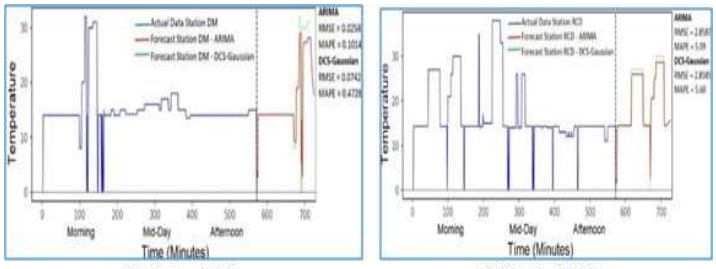



(a) Clustering coefficient matrix (b) correlation (CORR) coefficient matrix

Figure 12. Temperature for stations RCD to RCI of (a) clustering and (b) correlation (CORR) coefficient matrix.

**Table 10. Performance of predictive model - displacement**

| | ARIMA | | DCS – Gaussian | |
|---|---|---|---|---|
| | RMSE | MAPE | RMSE | MAPE |
| Station-DM | 0.0258 | 0.1014 | 0.0742 | 0.4728 |
| Station-RCD | 2.8587 | 5.09 | 2.8585 | 5.6 |
| Station-RCI | 2.328 | 4.5457 | 2.3343 | 5.06 |




(a) Station DM (b) Station RCD

Figure 13. Forecast for temperature of (a) station DM and (b) station RCD.

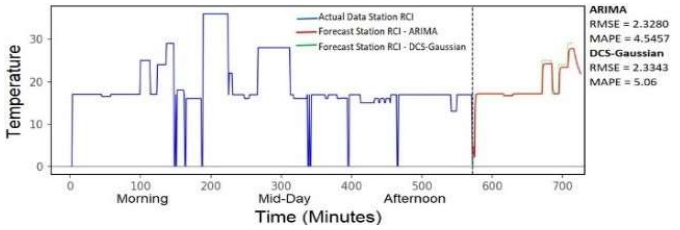



Figure 14. Forecast for temperature in station RCI.

3) Stage 3: forecasting cable behavior using time series: This section explored stage 3, focusing on forecasting cable behavior using time series. Table 10 shows performance of predictive model for displacement, whereas Fig. 13 illustrates forecast for temperature of (a) station DM and (b) station RCD. Forecast for temperature in station RCI is demonstrated in Fig. 14. In Stations DM, RCD and RCI, 80% of temperature is used in the training dataset, while 20% of temperature is used in the test dataset. In Station DM, the temperature's RMSE value in ARIMA is 0.0258 and MAPE = 0.1014, whereas in DCS Gaussian, the temperature's RMSE value = 0.0742 and MAPE value = 0.4728. Similarly, Stations RCD and RCI show ARIMA and DCS-Gaussian's RMSE and MAPE values. Due to low error values, ARIMA performs better than DCS-Gaussian.

One can observe high and low data movements which demonstrate cyclic behavior at Station DM from 0 to 570 minutes of temperature in Fig. 13a. It seems that there might be a likelihood of future cable sagging behavior in Station DM. Security agents may need to rush at the nearest time to Station DM. In Fig. 13b, there is a high, medium and low data movement which indicates cyclic and seasonality behavior from 0 to 570 minutes of temperature at Station RCD. There are huge possibilities for future cable sag behavior in Station RCD. Security agents might need to hurry to the nearest time to Station RCD. It seems that there is a high and low data movement which illustrates cyclic behavior at Stations RCI from 0 to 570 minutes of temperature as depicted in Fig. 14. This implies a likelihood of future cable normalcy or sagging behavior in Station RCI. The security agents may respond at the nearest time to Station RCI.

## V. DISCUSSION AND COMPARISON

Sensor data revealed both strong and weak correlations and covariance among different hotspot stations. Strong positive relationships typically reflect consistent cable performance and operational stability, whereas weak or fluctuating correlations may signal anomalies such as sagging, tampering, or theft. These variations suggest different degrees of interdependence across the network, influenced by environmental conditions, cable age, and localized stress factors. A critical analysis indicates that stations with high traffic density or exposure to extreme weather tend to exhibit weaker correlations, possibly due to accelerated wear or external interference.

The adoption of a three-stage time-series decision system has proven highly effective in advancing proactive security measures. By analyzing displacement and voltage data, the system classifies monitoring stations according to temporal patterns. Cyclical variations often signal potential risks that warrant immediate intervention, whereas consistent, stable readings indicate that routine oversight is sufficient. This structured approach strengthens situational awareness for security teams and enables the efficient prioritization of resources based on real-time risk indicators.

The wireless sensor network (WSN) architecture provided detailed data collection capabilities, yet several operational challenges were identified. Limited power reserves within sensor nodes constrained transmission frequency and reduced system longevity, particularly in remote or interference prone environments. In addition, the framework produced a moderate level of false positives, occasionally misinterpreting normal fluctuations as anomalies. These limitations emphasize the importance of adaptive calibration strategies and redundant sensor deployment to ensure consistent reliability and resilience.

Practical implementation of the monitoring system must account for cost, scalability, and latency. High-resolution sensors and edge computing capabilities improve responsiveness but increase capital and operational expenditure. Scalability is challenged by the need for consistent data quality across diverse geographic and infrastructural conditions. Latency in data processing—especially during peak operational hours—can hinder timely decision-making. These factors must be balanced against the strategic value of early anomaly detection and infrastructure protection.

We compare the proposed data science-driven approach on time series observations of wireless sensor network integration with existing related research. The experiments aim to determine whether railway cables sag or stolen cables using time series anomaly detection pattern behavior across sensor stations. The detection may reveal normal, cable theft and cable sagging states for publicly available and real-life data. Table 10 shows a comparison of the proposed framework with related research.

Our study suggests that higher accuracy is not associated with poor performance in the critical railway infrastructure. The proposed method may benefit from the integration of wireless sensor networks for real-time observations of without impacting the railway cable infrastructure. Our model explored a comprehensive three-stage data analytics-driven framework with the integration of wireless sensor networks (WSNs). The proposed model demonstrates that the proposed framework is more resilient and robust than other related research. Additionally, the challenge of accurately distinguishing between instances of cable sagging and those of cable theft, which has historically led to unwarranted alarm triggers, has been effectively addressed in this work, thereby ameliorating the issue of false alarms. This is evident in Sections 4.2.1 and 4.3.1, indicating probable cable sagging and theft in the proposed framework.

**Table 11. Comparing the proposed framework with related research.**

| Literature | Problem Addressed | Methodology | Results |
|---|---|---|---|
| [20] | Railway cable fault detection and localization safety. | Orthogonal multi one time domain reflectometry-based fault detection. | Results show that 70% of the cables are healthy, 25% have minor faults, and 5% have significant faults. |
| [21] | The protection wire in the railway stations. | Multi-conductor Transmission Lines (MTL). | Results of simulations indicate that the rail cable returns a higher current. |

| | | | |
|---|---|---|---|
| [22] | The defects and damage that cause railway cable failures. | Distributed temperature sensors (DTS) monitoring scheme contains online monitoring and fault location. | Results show that the heating size (0.2 m) is half of the spatial resolution of the DTS (0.41 m), maximum temperature rise along the cable is about 110C and 70C respectively. |
| [23] | The protection of railway power cables defects and faulty at stations. | MATLAB-Simulink electric circuit model. | i) Superconducting power cable with terminal resistance = 20 []. ii) Superconducting fault current limiter with critical current at 77 K and Inductance = 1.1 [mH]. |
| [24], [25] | Rail cable electrical protection against failures. | Finite element method (FEM) and Finite Difference Method (FDM). | i) Fault current response = 100 m and the DC link capacitance (Cf) is varied between 50 mF, 225 mF and 500 mF. ii) Peak temperature = filter capacitance (from 0 to 250mF) and shunt resistor (from 11m to 100m). |
| Proposed approach | Railway cables sagging problem and cable theft problem. | A proactive data science -driven wireless sensor network (WSN) integration framework and Time series datasets. | i) Average errors for a displacement's RMSE value in ARIMA is 0.1013 and MAPE = 1.5121. ii) Indication of high accuracy. Aligned correlations and covariance results in sections 4.2.1 |

## VI. CONCLUSIONS AND FUTURE WORK

This research proposed a proactive, data science-driven methodology that integrates wireless sensor networks (WSNs) to address the limitations of existing techniques for railway cable infrastructure protection, particularly in reducing delays in threat detection. The three-stage framework includes initial time-series analysis, clustering to identify abnormal behavior, and forecasting to predict potential threats such as cable sagging or theft. Its goal is to enhance situational awareness and provide early alerts to security agents through a structured, real-time decision support process.

The study yielded several quantifiable outcomes that underscore the effectiveness of the proposed framework:

- High forecasting accuracy: The application of ARIMA and DCS Gaussian models resulted in exceptionally low prediction error rates, with RMSE at 0.0212 and MAPE at 0.23%, demonstrating the reliability of the system in anticipating cable sagging and theft.
- Correlations and covariance across sensor stations: Analysis of sensor data revealed both strong and weak inter-station relationships, providing valuable insights into cable behavior patterns. Strong correlations indicated operational stability, while weaker correlations flagged potential anomalies, enabling targeted monitoring.
- Improved decision-making via three-stage analytics: The structured analytics framework—comprising data analysis, clustering, and forecasting—enhanced situational awareness and supported proactive intervention strategies. This approach enabled timely risk classification and resource prioritization across the network.

Experimental validation using both real-life and publicly available datasets demonstrated the dependability of the framework. Low forecasting errors (e.g., RMSE = 0.0212, MAPE = 0.23%) and consistent correlations and covariances between stations support its robustness and accuracy. Compared to conventional methods, this approach offers improved anomaly detection, clearer fault differentiation, and better adaptability across diverse network conditions.

The study also highlights existing defense gaps in railway cable systems and contributes to the growing body of research on IoT-enabled, AI-enhanced infrastructure security. While the results are promising, ongoing improvements are essential. Future work will explore the application of more advanced AI models, such as graph neural networks and federated learning, as well as emerging technologies like quantum sensing. Additionally, further efforts will aim to test the system under real-world deployment conditions and enhance resilience against cyber and physical disruptions.

## Acknowledgment

The authors gratefully acknowledge the financial support from the University of South Africa.

## References

[1] M. A. Rahman, H. Taheri, F. Dababneh, S. S. Karganroudi, and S. Arhamnamazi, "A review of distributed acoustic sensing applications for railroad condition monitoring," *Mech. Syst. Signal Process.*, vol. 208, pp. 1-26, 2024. https://doi.org/10.1016/j.ymssp.2023.110983.

[2] T. Hoang et al., "Autonomous end-to-end wireless monitoring system for railroad bridges," *Adv. Bridg. Eng.*, vol. 1, no. 1, pp. 1–27, 2020. https://doi.org/10.1186/s43251-020-00014-7.

[3] L. Zhu, C. Chen, H. Wang, F. R. Yu, and T. Tang, "Machine learning in urban rail transit systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 3, pp. 2182–2207, 2024. https://doi.org/10.1109/TITS.2023.3319135.

[4] T. Andrysiak, Ł. Saganowski, and W. Mazurczyk, "Network anomaly detection for railway critical infrastructure based on autoregressive fractional integrated moving average," *Eurasip J. Wirel. Commun. Netw.*, vol. 245, pp. 1-14, 2016. https://doi.org/10.1186/s13638-016-0744-8.

[5] Z. Yu, H. Wang, and F. Chen, "Security of railway control systems: A survey, research issues and challenges," *High-speed Railway*, vol. 1, no. 1, pp. 6–17, 2023, https://doi.org/10.1016/j.hspr.2022.12.001.

[6] V. J. Hodge, S. O'Keefe, M. Weeks, and A. Moulds, "Wireless sensor networks for condition monitoring in the railway industry: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 1088–1106, 2015. https://doi.org/10.1109/TITS.2014.2366512.

[7] Z. Zuo, X. Tian, and Z. Shao, "Deepening research on the comprehensive application and development of railway intelligent detection and monitoring system and key technologies," *Proceedings of the IEEE 7th Inf. Technol. Mechatronics Eng. Conf. (ITOEC'2023)*, 2023, vol. 7, pp. 425–433. https://doi.org/10.1109/ITOEC57671.2023.10291609.

[8] S. C. Oh, G. D. Kim, W. T. Jeong, and Y. T. Park, "Vision-based object detection for passenger's safety in railway platform," *Proceedings of the Int. Conf. Control. Autom. Syst. (ICCAS 2008)*, 2008, pp. 2134–2137. https://doi.org/10.1109/ICCAS.2008.4694449.

[9] P. L. Croucamp, S. Rimer, and C. Kruger, "Locomotive monitoring system using wireless sensor networks," *Proceedings of the 2014 12th IEEE Int. Conf. Ind. Informatics, (INDIN'2014)*, 2014, pp. 639–644. https://doi.org/10.1109/INDIN.2014.6945588.

[10] D. Efanov, G. Osadchy, D. Sedykh, D. Pristensky, and D. Barch, "Monitoring system of vibration impacts on the structure of overhead catenary of high-speed railway lines," *Proceedings of the 2016 IEEE East-West Des. Test Symp, (EWDTS'2016)*, 2016, pp. 1–8. https://doi.org/10.1109/EWDTS.2016.7807691.

[11] P. Velha et al., "Monitoring large railways infrastructures using hybrid optical fibers sensor systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5177–5188, 2020. https://doi.org/10.1109/TITS.2019.2949752.

[12] A. Eduard, D. Urazayev, A. Sabyrbek, D. Orel, and D. Zorbas, "Ad-hoc train-arrival notification system for railway safety in remote areas," *Internet of Things (Netherlands)*, vol. 25, pp. 1-15, 2024. https://doi.org/10.1016/j.iot.2024.101062.

[13] P. Aela, J. Cai, G. Jing, and H. L. Chi, "Vision-based monitoring of railway superstructure: A review," *Constr. Build. Mater.*, vol. 442, pp. 1-25, 2024. https://doi.org/10.1016/j.conbuildmat.2024.137385.

[14] H. Sarih, A. P. Tchangani, K. Medjaher, and E. Pere, "Data preparation and preprocessing for broadcast systems monitoring in PHM framework," *Proceedings of the 6th Int. Conf. Control. Decis. Inf. Technol. (CoDIT'2019)*, 2019, pp. 1444–1449. https://doi.org/10.1109/CoDIT.2019.8820370.

[15] D. Vyroubal and I. Lackovic, "Target temperature effect on eddy-current displacement sensing," *Proceedings of the 2015 IEEE Sensors Appl. Symp.*, 2015, pp. 5–9. https://doi.org/10.1109/SAS.2015.7133621.

[16] Z. Zhang, R. Wang, R. Ding, and Y. Gu, "Unravel anomalies: an End-to-End seasonal-trend decomposition approach for time series anomaly detection," *Proceedings of the 2024 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 2024),* 2024, pp. 5415–5419. https://doi.org/10.1109/ICASSP48485.2024.10446482.

[17] E. Vlachos and K. Berberidis, "Adaptive completion of the correlation matrix in wireless sensor networks," *Proceedings of the Eur. Signal Process. Conf.,* 2016, pp. 1403–1407. https://doi.org/10.1109/EUSIPCO.2016.7760479.

[18] T. Dogdas and S. Akyokus, "Document clustering using GIS visualizing and EM clustering method," *Proceedings of the IEEE Int. Symp. Innov. Intell. Syst. Appl. (INISTA 2013)*, 2013, pp. 1–4. https://doi.org/10.1109/INISTA.2013.6577647.

[19] A. U. Mahin, S. N. Islam, F. Ahmed, and M. F. Hossain, "Measurement and monitoring of overhead transmission line sag in smart grid: A review," *IET Gener. Transm. Distrib.,* vol. 16, no. 1, pp. 1–18, 2022. https://doi.org/10.1049/gtd2.12271.

[20] R. Vijay, G. Prasad, Y. Huo, S. Sm, and P. Tv, "Measurement-based condition monitoring of railway signaling cables," *Proceedings of the IEEE Int. Conf. Commun. Control. Comput. Technol. Smart Grids, (SmartGridComm'2021)*, 2021, pp. 33–38. https://doi.org/10.1109/SmartGridComm51999.2021.9632301.

[21] X. He, Y. Wen, and D. Zhang, "Influence of protection wire and positive feeder arrangement of traction network on the return current in high-speed railway," *Proceedings of the IEEE 2nd Int. Conf. Electron. Technol. Commun. Information, (ICETCI'2022)*, pp. 205–210, 2022. https://doi.org/10.1109/ICETCI55101.2022.9832341.

[22] K. Chen, Y. Yue, and Y. Tang, "Research on temperature monitoring method of cable on 10 Kv railway power transmission lines based on distributed temperature sensor," *Energies,* vol. 14, issue 12, pp. 1-15, 2021. https://doi.org/10.3390/en14123705.

[23] K. Qian, H. Ohsaki, and M. Tomita, "Fault current analysis of DC electric railway feeding systems using superconducting power cables," *Proceedings of the 17th Int. Conf. Electr. Mach. Syst., (ICEMS'2014)*, pp. 261–266, 2014. https://doi.org/10.1109/ICEMS.2014.7013475.

[24] S. Nolan, C. E. Jones, P. J. Norman, and S. J. Galloway, "Understanding the impact of failure modes of cables for the design of turbo-electric distributed propulsion electrical power systems," *Proceedings of the International Conference on Electrical Systems for Aircraft, Railway, Ship Propulsion and Road Vehicles & International Transportation Electrification Conference (ESARS-ITEC)*, 2016, pp. 1–6. https://doi.org/10.1109/ESARS-ITEC.2016.7841363.

[25] G. Piao, J. Li, L. Udpa, J. Qian and Y. Deng, "Finite-element study of motion-induced eddy current array method for high-speed rail defects detection," *IEEE Transactions on Magnetics*, vol. 57, no. 12, Art no. 6201010, pp. 1-10, 2021. https://doi.org/10.1109/TMAG.2021.3119267.

***Moyahabo Rossett Mohlabeng** is a PhD Student in Computer Science from University of South Africa, South Africa. He completed BTech in Information Technology in University of South Africa, South Africa in the 2011 and also completed MTech in Information Networks at Tshwane University of Technology, South Africa in the year 2014. His research areas include Information Technology, Railway, Cable infrastructure, Security, Clustering, Internet of Things (IoT), Artificial Intelligence (AI), Wireless Sensors, Data Science.*

***Ernest Mnkandla** is a Professor and Head of the Centre for Augmented Intelligence and Data Science (CAIDS) at the University of South Africa (UNISA), South Africa. He received his PhD degree in electrical engineering from the University of the Witwatersrand, South Africa. He has a track record of accomplishment in supervising students and authoring articles in accredited refereed journals, book chapters, international Conferences, and practical research interests in emerging areas of 4IR, IoT, Big Data, Data Science, Machine Learning, Cloud Computing, etc.*

***Aminreza Karamoozian** received his PhD degree in mechanical engineering from the Nanjing University of Science and Technology, Nanjing, China. He received the B.Sc. and M.Sc. (Hons.) degrees in mechanical engineering from the Amirkabir University of Technology (Tehran Polytechnic), Iran. He joined the School of Automotive and Traffic Engineering, Jiangsu University, in 2019. His research interests include stability and dynamic analysis of microstructures and cellular structures, specially lattices, vehicle dynamics, advanced driver assistance systems (ADAS) for active safety applications, and reliability analysis of mechanical systems.*