# Monitoring cyberthreats in railway systems: A hybrid framework for detecting stealthy data tampering attacks

Sara Abdellaoui, Emil Dumitrescu, Cédric Escudero, Eric Zamai

# Monitoring Cyberthreats in Railway Systems: A Hybrid Framework for Detecting Stealthy Data Tampering Attacks

Sara Abdellaoui*, Emil Dumitrescu, Cédric Escudero, Eric Zamai

*INSA Lyon, Universite Claude Bernard Lyon 1, Ecole Centrale de Lyon, CNRS, Ampère, UMR5005, , Villeurbanne, 69621, , France*

**Abstract**

Railway cybersecurity has become a critical concern as the integration of advanced monitoring systems increases reliance on technology. Cyberattacks targeting railway systems can disrupt operations, compromise data integrity, and mislead maintenance decisions, jeopardizing safety and efficiency. Despite these risks, existing detection methods often struggle to address stealthy data tampering attacks designed to either mask failures or trigger unnecessary maintenance. To remedy this gap, this article proposes a novel framework combining Turnout Lifecycle Analysis (TLA) and Expected Behavior Analysis (EBA), complemented by a weighted, modified Dempster-Shafer theory to integrate threat estimations from both approaches. The proposed framework supports the detection of stealthy cyberattacks and the diagnosis of turnout faults, while enabling resilient decision-making under uncertainty. The framework is validated on simulated cyberattack scenarios, successfully identifying six out of seven attacks while reducing false positives. The results highlight the potential of this framework to give railway maintenance operators more accurate insights, help improve decision-making, and help enhance the safety and resilience of railway operations against cyberthreats.

*Keywords:* Cyberattacks, Stealthy Attacks, Cybersecurity, Data Tampering, Railway Systems, Turnouts, Maintenance

## 1. Introduction

### *1.1. Growing cybersecurity concerns in the railway sector*

The advent of Industry 4.0 promotes hyper-connectivity as a vector for operational efficiency. While the integration of cyber physical interactions enhances efficiency and improves performance in the railway sector, it also increases potential attack surfaces for cyberattacks, making it more vulnerable to intrusion and manipulation. Among others, the transport industry has witnessed significant impacts [1, 2], making cybersecurity one of the most pressing concerns in recent years, particularly following the Russian invasion of Ukraine in 2022 [3].

*Corresponding author

*Email addresses:* sara.abdellaoui@insa-lyon.fr (Sara Abdellaoui, emil.dumitrescu@insa-lyon.fr (Emil Dumitrescu), cedric.escudero@insa-lyon.fr (Cédric Escudero), eric.zamai@insa-lyon.fr (Eric Zamai)

This ongoing threat is profoundly affecting the railway sector, by drastically altering railway operations and maintenance, impacting availability, safety and security, more specifically cybersecurity of railway systems [4]. Cyber-malware can have severe consequences for human life and equipment [5, 6], as well as cause financial losses due to disruptions in railway services and equipment availability [7, 8]. According to the 2023 report by the European Union Agency for Cybersecurity (ENISA) [3], 21% of all observed cyber-incidents targeting transport sectors were against the railway sector. This report highlights that the primary objectives of these cyberattacks were disrupting operations and gaining financial benefits. Earlier, in 2021, the ENISA published a report on good practices in railway cyber-risk management [9]. These guidelines are based on practices and standards performed by European railway undertakings and infrastructure managers to assess and mitigate cyber-risks. The ENISA states that Operational Technology (OT) systems are more exposed to cyberattacks than Information Technology (IT) systems, due to a lack of cybersecurity awareness when OT systems were designed [10].

### *1.2. Attacks against railways: overview, consequences and considered types*

Among various cyberattack typologies, railways primarily encounter three main types of cyberattacks [11]:

- *Data theft*: compromises confidentiality by collecting stored data exchanged by railway infrastructure components;
- *Access denial*: compromises availability by preventing operators from having access to the railway system to operates legitimate queries;
- *Data tampering*: compromises integrity by altering data to interfere with railway system performance and security.

In this article, we consider cyberattacks targeting OT in railway systems, with a particular focus on scenarios where IT security measures have been compromised by attackers in an attempt to manipulate data. This emphasis is derived from the fact that even with robust IT security, breaches can still occur, potentially disrupting OT in railway systems. A notable example is the 2015 BlackEnergy malware attack on the Ukrainian power grid [12], which resulted in railway operation disruptions. Although IT security measures aimed at preventing intrusion were in place, OT systems remained vulnerable once the grid was compromised. More recently, in 2020, a cyberattack targeting Israel's railways [13] demonstrated that bypassing IT defenses can lead to direct attempts on OT systems. These examples highlight the critical need for dedicated OT security measures.

The railway Remote Monitoring System (RMS) is an advanced monitoring system that gathers information from sensors distributed across the infrastructure. Its primary goal is to provide maintenance operators with real-time data reflecting the health state of railway components, including rails, turnouts, crossings, sensors, and communication links. The RMS detects faults and helps anticipate failures. However, its dependency on distributed data transmission and processing makes it particularly vulnerable to cyberattacks that could compromise the integrity of the observed data.

Securing the RMS can be approached from two perspectives:

- **Securing against intrusions:** This refers to preventive measures designed to block unauthorized access and detect attempted breaches before they succeed. Intrusion detection systems and strong encryption protocols can help prevent attackers from accessing or tampering with the RMS data. Despite these efforts, vulnerabilities may persist. For instance, in [14], attackers exploit a zero-day vulnerability in the data transmission protocol in order to bypass RMS security measures and inject false data into the system, thereby misleading maintenance operators;

- **Securing in case IT security barriers are breached:** This focuses on maintaining the resilience and integrity of the RMS even after a successful intrusion into the IT layers of the railway infrastructure [15]. In such cases, OT components require additional safeguards to detect and mitigate anomalies. For example, an attacker who breaches IT defenses might manipulate sensor data to conceal an impending turnout failure [5]. If the RMS lacks OT-level safeguards, this manipulation could lead to undetected faults and eventually cause operational incidents resulting in accidents and derailments [16, 17].

Hence, while securing the RMS against intrusions is essential, ensuring resilience after IT barriers are breached is equally critical to guaranteeing the safety and security of railway systems in the face of evolving cyberthreats.

One of railway's key components is the turnout [18]. As illustrated in Fig. 1, turnouts consist of a pair of rails positioned to direct trains in a direct or deviated direction. Hence, they must meet strict security and reliability specifications [19]. To ensure the turnout remains operational, railway maintenance operators must assess its health condition based on data provided by the RMS.

The integrity of the data delivered by the RMS must be ensured in order to accurately assess the state of a turnout.

Figure 1: Turnout. 

This article focuses especially on data tampering attacks targeting turnout data monitored by the RMS, which aim to alter turnout maintenance decisions. As maintenance operators rely heavily on digital platforms for maintenance scheduling and system monitoring, such as the RMS, successful attacks on these platforms may disrupt services and compromise operational safety [20]. The cyberattacks considered in this work are inspired by the Stuxnet attack [14] against Iran's nuclear centrifuges. Its particularity was its design to conceal its modifications by feeding false feedback to operators, misleading them about

what was really happening. The objective intended in this work is to detect stealthy attacks that mislead maintenance operators about turnout's health condition:

- *Fault Masking (FM) attacks*: Turnout data reflecting problems in switch operations are replaced with normal, fault-free switch operations data or other operations data that do not need maintenance. The objective is to avoid triggering necessary maintenance actions.

- *Fault Injecting (FI) attacks*: Turnout data reflecting normal switch operations are replaced with data corresponding to failures. The objective is to trigger unnecessary maintenance actions.

### *1.3. Related research*

The following paragraphs will first present the state of the art on various cyberattack detection methods employed in Cyber Physical Systems (CPS). Subsequently, cyberattack detection methods specific for railway systems are analyzed.

The detection of cyberattacks targeting CPS is possible by relying on two main Intrusion Detection Systems (IDSs) [21, 22]: Signature-based IDS (SIDS) and Anomaly-based IDS (AIDS). SIDS rely on databases of known, previously observed intrusion types, which limits their ability to identify new, unknown threats [23]. Meanwhile, AIDS characterizes a reference system behavior and compares it with current behaviors to identify deviations. Considering the nature of the cyberattacks studied in this article, an AIDS that takes into account the behavior of remotely monitored railway turnouts seems appropriate, unlike SIDS that cannot address the new, stealthy and targeted nature of these attacks.

According to [24], AIDS are classified into: statistical-based techniques, expert knowledge-based techniques and Machine Learning (ML)-based techniques. Statistical approaches, such as Hidden Markov Models (HMM) [25] and Bayesian Networks [26], model dependencies among data and variables to detect deviations in the probability distributions. These methods are effective for identifying outliers, strange behaviors and known attacks but often struggle to detect previously unseen or novel attacks. Expert knowledge techniques [27, 28] rely on experience-based rules to flag suspicious behaviors [15]. While they can identify predefined abnormal patterns, they are generally ineffective against new or unknown attacks. ML techniques have been widely applied in intrusion detection, including Isolation Forest [29, 30], Random Forest [31, 32], Support Vector Machine [33, 34, 35], K-Nearest Neighbor (KNN) [36], Convolutional Neural Network (CNN) [37]. Despite their popularity, these methods have limitations. They rely solely on modeling the system's normal behavior and detecting deviations, making them ineffective at distinguishing between normal/failure behaviors and cyberattacks that mimic normal operations. Furthermore, some approaches are validated based on assumptions of secure communications between low-levels (sensors) and higher-levels (monitoring systems) or the availability of labeled datasets.

Overall, the above mentioned techniques are not well suited to the type of cyberattacks considered in this study, as they cannot reliably differentiate between stealthy attacks and normal or failure system behavior, often depend on secure data channels, and frequently require labeled data, which may not be available in real-world scenarios.

In the context of railways and cyber-risk assessment, several authors have explored cybersecurity concerns and highlighted the risks of the integration of cyber and physical security

systems. Authors in [38, 39] argue that traditional safety measures are insufficient for addressing modern cyberattacks. Rekik et al. [40] concentrate on the risks in train control and monitoring system, focusing on cyber physical vulnerabilities. Additionally, in [41], the authors underscore the critical need for ethical cybersecurity standards within the railway sector. Together, these studies advocate for increased cyber-awareness and stronger defenses to protect railway systems from evolving threats. However, the survey of Soderi et al. [42] concludes that the analyzed articles rarely consider the overall environment in which a system is situated, and thus they fail to acknowledge that railways are composed of interconnected critical systems. Therefore, there is a limited amount of literature that deals with the detection of cyberattacks targeting railway systems from a CPS standpoint. Among them, authors in [43] propose a method to detect attacks on railway sensors. The method involves comparing predicted sensors disturbances, which are predefined by topographic information, with the actual estimated sensors disturbances. In [44], a detection system against false data injection attacks targeting the sensors of trains' voltage, current and position was proposed by setting thresholds on sensor data. Kour et al. [45] introduce a method called Railway Defender Kill Chain (RDKC), which is designed to predict, prevent, detect and respond to known cyberattacks. The RDKC is based on an adapted Open System Architecture for Condition-Based Maintenance (OSA-CBM) framework for the railway cybersecurity, with the aim of providing cybersecurity information from a technological perspective along an extended Cyber Kill Chain (CKC) [46]. Authors in [47] provide a systematic review and outline cybersecurity emerging trends and approaches, concluding that there is a lack of focus on some important railway assets such as infrastructure and Supervisory Control and Data Acquisition (SCADA) systems.

### *1.4. Research motivation and contribution*

The field of railway cybersecurity remains both complex and challenging due to several critical issues. Among the major challenges are integrating cyber and physical components and detecting stealthy attacks. Hybridizing these components is particularly challenging due to the lack of integrated models that can capture both, physical dynamics of the railway infrastructure and interactions with IT systems. Additionally, developing robust anomaly detection algorithms that integrate machine learning techniques with physical models remains an ongoing challenge. A key issue is distinguishing between natural failures and malicious data tampering, which complicates maintenance decisions. The need for intelligent diagnosis methods combining both cyber and physical indicators has become increasingly evident, aiming to reduce false positives and negatives. Stealthy attacks pose an even greater challenge, as they can perfectly mimic normal behavior, making them difficult to detect. These attacks often involve subtle, undetectable changes that align with legitimate system operation, posing a significant detection challenge [48]. To address this, algorithms capable of identifying small deviations in system behavior without triggering excessive false alerts are necessary. Furthermore, the lack of realistic databases of stealthy attack scenarios further complicates model training. These research gaps present significant barriers for the railway sector, particularly maintenance operators, as they limit the ability to make precise maintenance decisions that are crucial for ensuring the availability and safety of railway systems.

This research contributes to the reliability and safety of railway systems by ensuring the integrity of maintenance decisions in the presence of cyberthreats. This is achieved by

exposing potential cyberthreats that compromise the authenticity of monitored turnout data and ensuring more informed and resilient maintenance decisions. Unlike approaches that focus primarily on identifying disruptive anomalies, i.e. non-stealthy attacks, this article proposes a data-driven framework for detecting stealthy attacks. The proposed framework addresses data tampering scenarios by relying on collected turnout data, without interfering with the turnout infrastructure or requiring additional security layers. This article makes the following key contributions:

- *Hybrid cyber-physical threat detection for critical infrastructure*: Develop a hybrid detection framework that integrates physical lifecycle representation with temporal behavior analysis to identify cyberattacks affecting railways' turnout data.
- *Enhanced failure-attack differentiation*: Improve the ability to distinguish between natural failures and malicious tampering by leveraging both cyber and physical indicators. This supports more accurate fault diagnosis and reduces false alarms that could lead to unnecessary maintenance or unsafe decisions.
- *Validation through adversarial scenarios*: Evaluate the proposed detection framework under diverse simulated cyberattack scenarios that emulate stealthy data tampering, addressing the common challenge of limited access to real-world datasets on cyberthreats targeting railway infrastructure. This enables a robust assessment of detection reliability under uncertainty.

To achieve this, two existing approaches - focusing respectively on turnout lifecycle modeling [49] and temporal behaviors analysis [50] - are combined, using the Dempster-Shafer theory of evidence [51, 52], to propose an enhanced cyberattack detection framework that involves:

1. Estimating cyberattack likelihoods related to each observed turnout behavior using two approaches:
   - The first approach classifies and analyzes the turnout behavior with regard to its lifecycle using aging criteria [49];
   - The second approach analyzes the temporal turnout behavior evolution in order to contextualize observed turnout behavior with regard to expected behavior based on previous behaviors' observations [50].
2. Combining the cyberattack likelihoods obtained using an additional fusion layer based on the Dempster-Shafer evidence theory. As detailed in Section 3.3.2, the Dempster-Shafer theory is applied to mitigate subjectivity in the evaluation of sources and to aggregate assessments originating from heterogeneous perspectives [53].

The need for a combined approach arises from each approach's limitations when used individually. The main strategy involves leveraging the exploitation of the turnout's lifecycle represented by the *bathtub* curve, as illustrated in Fig. 7. However, in certain lifecycle phases as described by the bathtub curve, accurately assessing cyberattack risks based solely on this analysis becomes difficult. Relying solely on the analysis of temporal behavior evolution can be imprecise when addressing specific turnout behaviors. Therefore, combining

both approaches aims to enhance cyberattack detection and the decision-making process by leveraging the strengths and compensating for the limitations of each. This article is structured into five main sections. Section 1 introduces the concept of cybersecurity in railways and cyberattack types, and presents cyberattack detection methods. Section 2 provides an overview of the railway RMS architecture, outlines its vulnerabilities, and presents the research hypothesis and problem statement. Section 3 describes the methodological steps of the proposed framework in detail. Section 4 presents the application of the proposed framework, followed by a discussion of the findings and suggestions for future improvements. The article concludes in Section 5 with a summary of the key results and insights.

## 2. The Cybersecurity problem in railway monitoring


Figure 2: Railway RMS. © Vossloh 2024. All rights reserved.

### *2.1. Railway RMS architecture*

The railway RMS, illustrated in Fig. 2, is composed of sensors distributed throughout the railway infrastructure. Data from these sensors are combined and merged in data hubs before being transmitted to the Central Monitoring System (CMS) to facilitate maintenance and operational decision-making. The RMS's vulnerabilities lie in its wide geographical distribution, providing many opportunities for attackers to breach the system at various points: overriding physical sensors' data or intercepting communications between sensors and the CMS.

This article focuses on turnouts; therefore, the emphasis will be on switching sensor data. During each switching operation, these sensors measure the current and voltage of the electrical motor moving the switch rail and then compute the power consumption. Henceforth,

this data will be referred to as current, voltage and power data. Based on [54], to accurately assess the operational condition of turnouts and the deterioration of their mechanical parts, attention should be directed towards analyzing power data. Consequently, power curves reflecting turnout switch operations that are transmitted to the CMS are studied in the sequel.

### *2.2. Hypothesis & problem definition*

As mentioned beforehand in Section 1, two stealthy cyberattacks corresponding to two data manipulation scenarios are considered: the FM attack and the FI attack. Thus, the cyberattacker's profile is assumed to match the required skills to perform such attacks:

1. Have an extensive knowledge about the railway infrastructure, except for information concerning the turnout lifecycle;
2. Have the means to manipulate only data reflecting the health of turnouts;
3. Have a high level of expertise to ensure that compromised data consistently reflects realistic turnout behaviors, depicting either normal or failure scenarios.

Several studies employ the analysis of switch operation curve shapes and patterns to diagnose turnout health and detect failures [18, 55, 56, 57, 58]. The authors in these studies do not question the trustworthiness of the observed switch operation curves, implicitly assuming that they faithfully reflect the actual turnout behavior. Therefore, the reliability of such failure detection methods is grounded on the assumption that the curves accurately represent the turnout's true state. Various factors influence the shape of the collected turnout data: external ones, such as obstacles, humidity, ambient temperature, and cyberattacks, and internal ones, such as friction forces and lubrication [59]. This article tackles the challenge of cyberattack detection by analyzing observed field curves in a highly exposed railway environment. The key challenge is differentiating data that accurately reflects the true state of the turnout from data that has been strategically manipulated by an attacker to mimic legitimate behavior and evade traditional anomaly detection mechanisms.

## 3. Methodology: from data modeling to cyberattacks detection

Previous sections have outlined vulnerabilities in railway systems, particularly in railway RMS, along with the types of cyberattacks considered. Domain expertise of railway operators is considered central in the cyberattack detection framework developed here. This framework leverages a collection of advanced technical tools and operator expertise. Thus, as raw field data generally does not provide clear interpretations on its own, it must be modeled and cross-examined with human expertise to derive meaningful domain insights into each curve. Subsequently, each collected piece of field data is interpreted and assessed with respect to cyberthreat likelihood. This framework is highlighted in Fig. 3. It consists of two phases: a development phase conducted offline with raw data, and an operation phase where cyberattack detections are performed after each switch operation. This section provides a detailed explanation of each phase.

The development phase utilizes an uncompromised turnout dataset to identify specific behavioral reference patterns. During this phase, classification and forecasting models are developed to interpret each curve and anticipate expected turnout behavior by analyzing

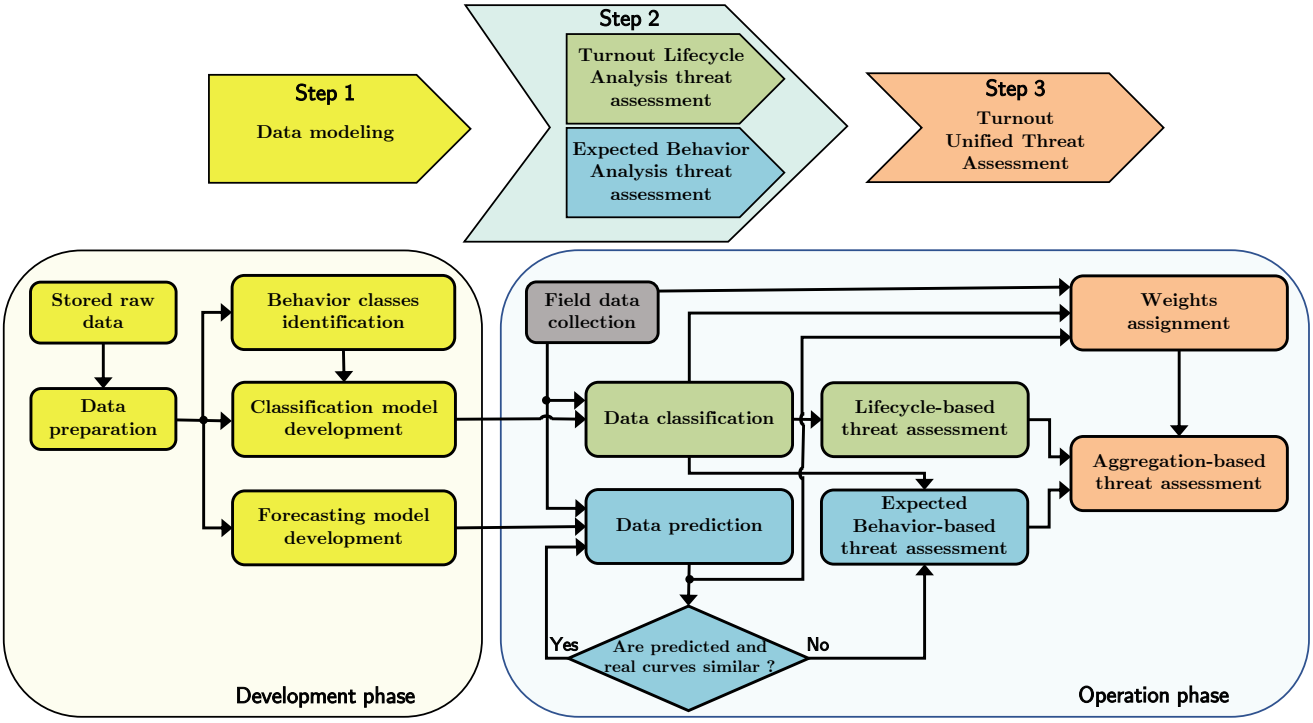



Figure 3: Methodology of the proposed framework.

its temporal evolution. The operational phase establishes a robust framework capable of accurately distinguishing between legitimate turnout behaviors and potential cyberattacks. This is achieved through a double assessment process that cross-analyzes both the expected system response and real-time behavioral deviations, enabling the detection of subtle manipulations that could otherwise appear normal to conventional detection methods. Two cyberattack detection approaches run in parallel, each providing a cyberthreat estimation by contextualizing the observed behavior with respect to either the *turnout lifecycle* or the *expected behavior*:

- *Turnout Lifecycle Analysis* (TLA) is based on the premise that the meaning of a field curve changes throughout a turnout's lifecycle. Indeed, a curve reflecting perfect operation can be considered normal in the early stages of a turnout's useful life, but increasingly suspicious as the turnout ages. Hence, accurate knowledge about the turnout lifecycle is fundamental. Yet, this reasoning has its limitations: at mid-lifecycle, wear is barely noticeable and the failure rate increases. In such situations, it is no longer possible to take a firm stance regarding the cyberthreat likelihood;

- *Expected Behavior Analysis* (EBA) addresses the shortcomings of the TLA. It provides a refined assessment of each field curve, put into perspective with respect to a sequence of curves previously observed.

The cyberthreat detection is then enhanced by combining the TLA and EBA estimations into a single indicator.

### *3.1. Step 1: Data modeling*

This step is illustrated in Fig. 4. It involves extracting insights from a large amount of turnout field data representing power curves of switch operations to determine how best to use this data. This is accomplished by first identifying behavior classes through a semantic analysis step, followed by building two predictive models, a classification model supporting the TLA and EBA and a forecasting model supporting the EBA.

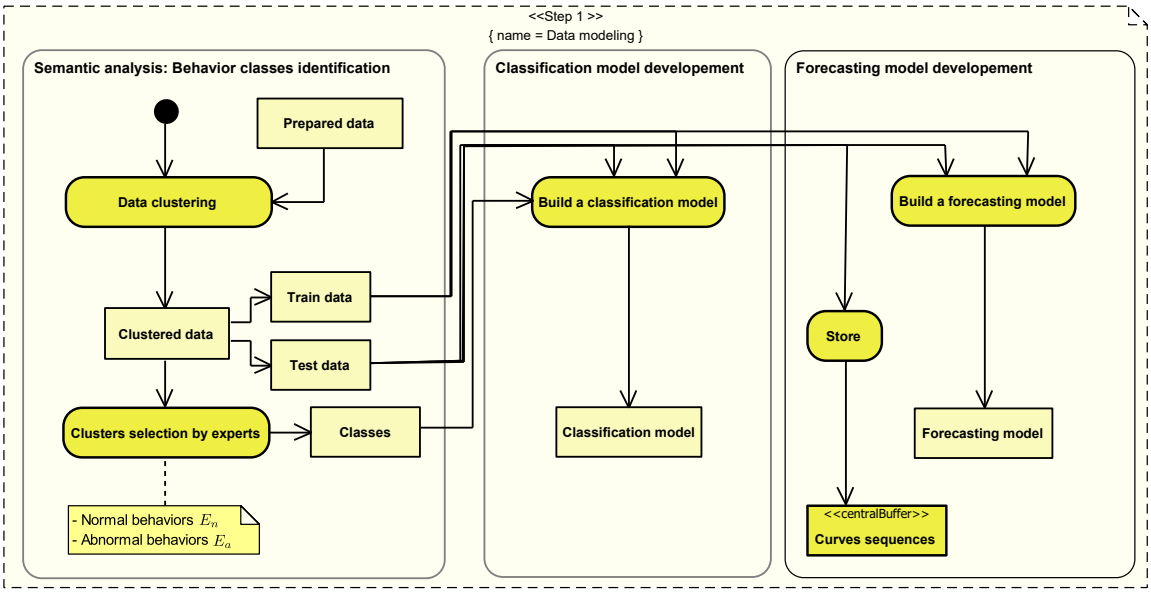



Figure 4: Data processing and predictive modeling.

#### *3.1.1. Semantic analysis: Behavior classes identification*

Data clustering is a straightforward technique and convenient way of identifying behaviors allowing to automatically organize data into clusters using a predetermined similarity measure. Railway expert knowledge is crucial in determining an appropriate meaning of each cluster.

A relevant similarity criterion must consider that a turnout power curve is a sequenced set of data. This specificity shows that Time Series (TS) based methods are necessary. Among other methods [60], a shape-based method for time-series clustering is employed because the shape of the turnout time-series data is a critical in identifying and distinguishing different types of behaviors. To measure similarities between TS curves in both time and shape, two common metrics are used: the Euclidean distance and the Dynamic Time Warping (DTW). The Euclidean distance involves one-to-one point comparisons, making it straightforward but also sensitive to small changes along the time axis [61]. This sensitivity can lead to misleading measures in case of slight time shifts. Conversely, DTW "elastic" measure allows for one-to-many point comparisons, making it flexible for aligning curves that may be out of phase. For clustering turnout data, DTW is chosen as a metric for the clustering algorithm because it can handle variations in data speed or temporal shifts, aligning with the goal of defining behavioral classes in the dataset.

Moreover, when comparing the two widely used clustering algorithms, k-Means [62] and hierarchical clustering [63], the turnout dataset's size favors k-Means. K-Means is faster as it only requires one pass through the data for cluster assignment, while hierarchical clustering demands more iterations and calculations, leading to a higher time complexity. However, the k-Means algorithm requires specifying the number of clusters $k$, which can be determined using methods such as the elbow method or the silhouette score method [64].

By combining railway domain expertise with findings from related works [65, 66, 67, 18, 58, 68], this phase produces a set of clusters and corresponding labels that systematically link each cluster to a distinct turnout behavior class, highlighting the generality and robustness of the approach. Without any loss in overall scope, for this study, the focus is narrowed down to two broad categories, normal and abnormal curves, along with their respective subcategories. Let $E_n$ and $E_a$ represent the sets of labels for these two categories, where $E_n = \{e_{n_e}, e_{n_w}\}$ are the labels associated with normal curves, and $E_a = \{e_{a_p}, e_{a_m}, e_{a_s}\}$ are the labels associated with abnormal curves. The complete set of labels is given by $E = E_n \cup E_a$ . Table 1 provides a short description of each considered curve type.

Table 1: Turnout behavioral classes.

| Label | Description |
|---|---|
| $e_{n_e}$ | Curves reflecting normal, fault-free turnout behavior |
| $e_{n_w}$ | Curves reflecting normal aging behavior of the turnout |
| $e_{a_p}$ | Curves reflecting progressive pre-faults degradations |
| $e_{a_m}$ | Curves reflecting minor faults with no need for maintenance |
| $e_{a_s}$ | Curves reflecting sudden failure behavior |

As described in the next paragraph, these labels contribute to the supervised construction of a predictive model capable of identifying any field power curve.

*3.1.2. Classification model development*

In the semantic analysis, meaning is assigned to each curve. The results of this process need to be generalized to assign meaning to newly acquired field curves, which is fundamental for implementing the TLA. This is achieved through the development of a *classification model* based on a supervised learning approach, wherein the classes generated in the preceding stage serve as labels throughout the training and validation phases.

To develop the classification model, the K-Nearest Neighbor (KNN) technique [69] is used, as it is a straightforward yet powerful algorithm commonly used for classifying time series data.

The KNN learning algorithm can use the same distance metrics, as discussed earlier, to compute similarities among data. Findings from a set of experiments on time series data in [70] confirm that the DTW metric is more effective than the Euclidean distance metric in the KNN classification for TS data. This supports the use of DTW in constructing the classification model in this study.

*3.1.3. Forecasting model development*

The goal of developing a *forecasting model* is to be able to predict the next expected behavior of the turnout based on observed behaviors history. This is fundamental for imple-

menting the EBA.

**Forecasting model selection** Two of the most widely used deep learning tools for sequence-to-sequence prediction are examined: Convolutional Neural Network (CNN) [71] and Long Short-Term Memory (LSTM) networks [72]. CNNs have proven effective in analyzing visual data for applications such as pattern and image recognition. They can automatically learn hierarchical representations from data, allowing them to efficiently capture complex features and patterns [73]. LSTMs are a type of Recurrent Neural Networks (RNN) designed to address the RNN vanishing gradient issue by managing the information flow in the network. Unlike CNN, LSTM can capture long-term dependencies and handle inputs and outputs of varying lengths. After analyzing the performance of CNN and LSTM applied to turnout operation data, and considering research contributions comparing various prediction methods [74, 75, 76], the LSTM method is selected to develop the turnout field curve forecasting method.

Table 2: LSTM & CNN performance.

| Method | Root Mean Squared Error (RMSE) | Mean Absolute Error (MAE) | Time to train |
|---|---|---|---|
| CNN | 0.065 | 0.028 | $170s$ |
| LSTM | 0.043 | 0.015 | $50s$ |

Indeed, according to the figures presented in Table 2, LSTM provides both better training performance and accuracy.

The LSTM learning and prediction process requires a particular preparation step, as illustrated in Fig. 5. This involves using a set of $N$ curves, called window size, in a sliding sequence to predict the $N + 1^{th}$ curve. With a total of $M$ switch operation curves in the training dataset, the size of the training set becomes $M - N$. The validation dataset goes through the same data preparation process. The trained and validated model serves as the forecasting model used in the EBA threat assessment described in Section 3.2.2.

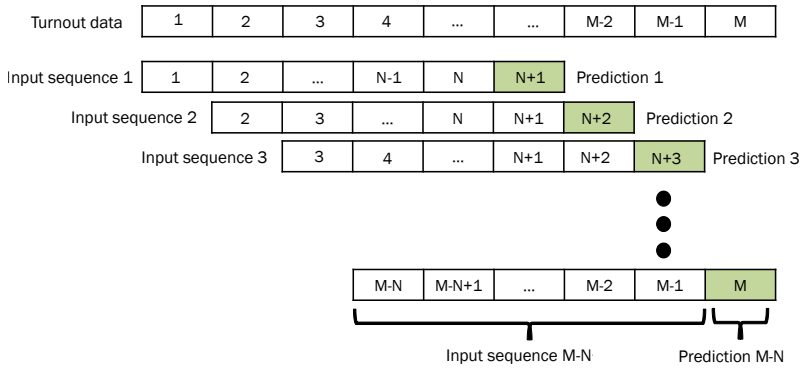



Figure 5: LSTM learning data structure.

### *3.2. Step 2: Cyberthreat assessment*

#### *3.2.1. Turnout Lifecycle Analysis (TLA) threat assessment*

The TLA threat assessment process is described in Fig. 6. As mentioned in Section 3.1.2, turnout power data are classified as either normal $E_n$ or abnormal $E_a$. From this stage onward in the threat assessment approach, the focus shifts from the actual shape of the power curve to its assigned label by the classification model, as illustrated in Fig. 6.

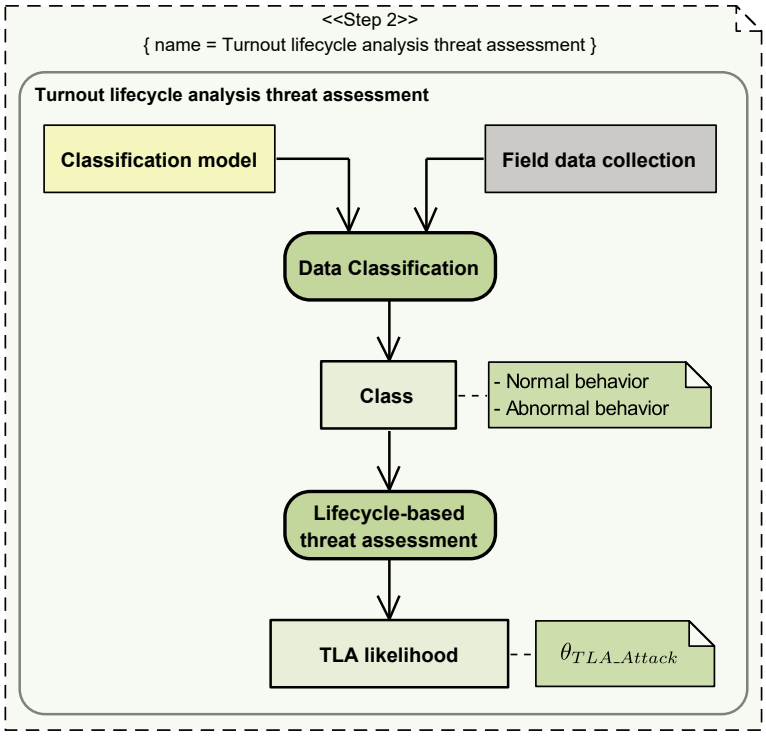



Figure 6: Turnout Lifecycle Analysis Threat Assessment.

The TLA approach produces an indicator, $\theta_{TLA\_Attack}$, representing the likelihood of the existence of a cyberattack. This likelihood is estimated by analyzing each label occurrence according to its context within the turnout lifecycle according to the following principles:

- Every occurrence of $E_n$ or $E_a$ results in an estimation of $\theta_{TLA\_Attack}$;
- It is considered that $\theta_{TLA\_Attack}$ is within the range $\in ]0, 1[$, indicating that the threats cannot be dismissed as impossible nor regarded as entirely certain;
- The presence of $e_{n_e}$ causes less suspicion after maintenance operation and high suspicion pre-scheduled maintenance;
- The presence of $e_{n_w}$ or $E_a$ is considered suspicious after maintenance operation, and less suspicious before scheduled maintenance.

It is assumed that after maintenance operations, turnouts are restored to an *as-good-as-new* state, and behave as if newly installed. This allows to represent the turnout lifetime

failure rate using the Weibull distribution [77, 78, 79, 80]. As turnouts combine both electronic and mechanical components, exhibiting wear, fatigue and corrosion, their malfunctioning behavior is categorized according to a *bathtub* shaped curve, as illustrated in Fig. 7. While other failure rate profiles exist in railway systems [81], the choice of the bathtub shape is guided by railway domain-specific studies [82, 83, 84], and validated by expert feedback from the industrial partner who confirmed its relevance in a practical maintenance context. Therefore the turnout malfunctioning behavior typically progresses through three stages:

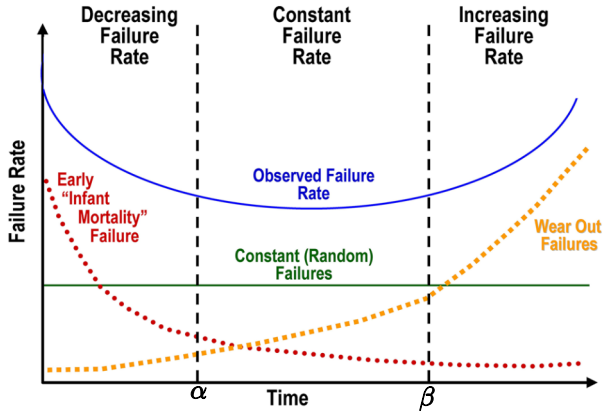



Figure 7: The Bathtub curve [85].

- Early infant mortality failures (burn-in) period: The first stage following installation or maintenance of the system, marked by high failure rates due to material defects and installation issues;
- Constant failures (useful life) period: This phase denotes a stable period in the turnout lifecycle where failure rates remain consistently low;
- Wear-out period: The final stage in the turnout's lifecycle, after prolonged operations, where failure rates start to increase because of factors like aging, material deterioration and other factors.

Consequently, detecting FM attacks and FI attacks relies on monitoring the turnout's aging time since being installed or maintained. It is considered that the turnout operates as well after maintenance operations as it does after installation.

Therefore, the TLA approach monitors incoming events online, focusing on the turnout wear, and on the timing $t$ of switch operations and utilizing two specific milestones known as $\alpha$ and $\beta$. Railway operators specify these milestones, which correspond to different phases of the turnout's useful life, where "natural" anomalies are infrequent. During the burn-in period, the threat level associated with acquired power curves cannot be reliably determined due to the high risk of failure during this phase, while normal behavior is also anticipated. To reflect this uncertainty, the corresponding curves are assigned a threat estimation value of 0.5, indicating that, at this stage, both genuine and cyberattack scenarios are considered equally probable.

The threat monitoring is achieved by the *Time monitor* block modeled in Fig. 8. The states $\{0, 1, 2\}$ represent the three intervals of the *bathtub* curve. Transitions between these states are triggered by the labels $E = E_n \cup E_a$, associated to a field curve $e$, as well as by $e_m$, which represents maintenance actions.

The *Time monitor* $Tm$ is defined as $Tm = \langle Q, q^0, E, e_m, \gamma, \theta_{TLA\_Attack} \rangle$ where $Q$ and $E$ represent the sets of states and labels, respectively, $q^0$ is the initial state, and $\gamma$ is the transition function of $Tm$. Let $\epsilon$ be an infinitely small positive real number. The $Tm$'s state evolves according to an internal clock, $clk$, which is reset upon each maintenance action $e_m$.

The output function $\theta_{TLA\_Attack} : Q \times E \times \mathbb{R}^+ \rightarrow ]0, 1[$ of $Tm$, which estimates the likelihood of a cyberattack, is defined as:

$$\theta_{TLA\_Attack}(q, e, t) = \begin{cases} 0.5 & \text{if } q = 0 \\ \frac{\beta - t}{\beta - \alpha} - \epsilon & \text{if } q = 1 \ \& \ e \in E_a \cup \{e_{n_w}\} \\ \frac{t - \alpha}{\beta - \alpha} + \epsilon & \text{if } q = 1 \ \& \ e = e_{n_e} \\ \epsilon & \text{if } q = 2 \ \& \ e \in E_a \cup \{e_{n_w}\} \\ 1 - \epsilon & \text{if } q = 2 \ \& \ e = e_{n_e} \end{cases} \tag{1}$$

*3.2.2. Expected Behavior Analysis (EBA) threat assessment*

The EBA threat assessment process is based on contextualizing each field curve, with respect to a sequence of immediately preceding curves, according to the following principles:

- The turnout cannot heal by itself: whenever a sequence of abnormal curves is observed, a new abnormal curve is expected as long as no maintenance action has been taken;
- Failures documented as being progressive cannot happen suddenly.

As illustrated in Fig. 9, the assessment involves applying the forecasting model from Section 3.1.3 to predict the upcoming power curve by considering previously observed curve sequences. To perform a prediction, the forecasting model requires a sequence of $N$ curves. Subsequently, this sequence evolves over time by incorporating new, non-corrupt field data, while discarding older data from the prediction sequence.

To detect a potential cyberattack, a systematic comparison is performed between the predicted and actual field curves. This comparison aims to identify discrepancies between

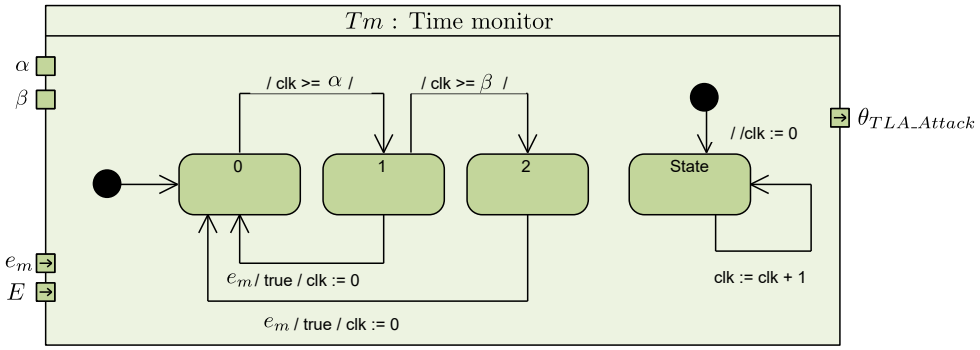



Figure 8: Time monitor.

expected and observed behavior, requiring an appropriate similarity metric. The two distance metrics, DTW and Euclidean distance, are considered. This dual-metric approach ensures a robust and comprehensive comparison. To formalize the comparison process, it is essential to establish thresholds for both DTW and Euclidean distance. These thresholds serve as references for determining whether two curves should be considered similar or different. As detailed in Section 4.3.2, these values are derived from an analysis of the training dataset.

Based on these thresholds, observed turnout behavior is classified as either similar to or different from the expected behavior. As shown in Fig. 9, if the collected curve is found to be similar to the predicted curve, no cyberthreat is ruled out. This curve is then stored and used in the next input sequence to maintain temporal continuity in the forecasting model. However, if the collected curve differs from the predicted curve, further investigation based on their respective types is required to assess the cyberthreat likelihood.

The EBA evaluation generates a cyberthreat indicator $\theta_{EBA\_Attack}$. The threat assessment relies on the classification model developed in Section 3.1.2 to categorize both the collected and predicted curves. Let $f$ be the predicted curve label and $F$ be the label set of predicted curves.

Given that sudden failures cannot be deduced by just analyzing the temporal evolution of turnout behavior, these scenarios are excluded from the training data. Hence, $F = E \setminus \{e_{a_s}\}$ and $\theta_{EBA\_Attack} : E \times F \rightarrow ]0, 1[$ according to the following rules:

- **No-cyberattack suspected** ($\epsilon$):

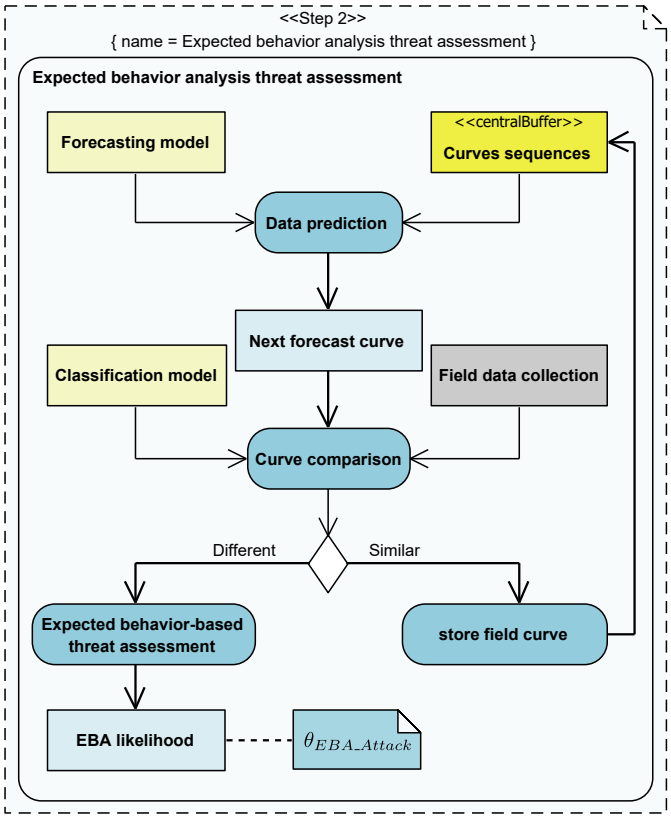



Figure 9: Turnout Expected Behavior Analysis Threat Assessment.

  - The field curve $e$ matches the EBA predicted curve $f$, indicating that the observed behavior is as expected;
  - The field curve shows a minor abnormality $e_{a_m}$, while the EBA predicted curve reflects a normal behavior $E_n$. The cyberattack risk is negligible. Indeed, it seems unlikely that cyberattackers would try to raise such alerts, as minor abnormalities do not require maintenance actions.

- **Undetermined situation** $(0, 5)$:
  - The field curve represents a sudden failure $e_{a_s}$, regardless of the behavior $(F)$ reflected by the predicted curve. The cyberattack risk cannot be accurately estimated, as sudden failures are always possible and cannot be distinguished from similar malicious curves. In this case, assigning a threat estimation value of 0.5 reflects the uncertainty in determining whether the anomaly is caused by a natural failure or a malicious cyberattack. This shortcoming is sometimes compensated by TLA.

- **High cyberattack risk** $(1 - \epsilon)$:
  - The field curve represents a normal behavior $e \in E_n$ or a minor abnormality $e_{a_m}$, while the EBA predicted curve is related to a progressive abnormality $e_{a_p}$. The cyberattack risk is considered high, as an evolving abnormal situation cannot heal spontaneously;
  - The field curve indicates a progressive abnormality $e_{a_p}$, while the predicted curve shows normal behavior $E_n$. The cyberattack risk is considered high, as the progressive abnormality appears suddenly, without any prior indication in the previously observed sequence;
  - The field curve reflects a normal fault-free $e_{n_e}$ behavior, while the EBA predicted curve indicates wear $e_{n_w}$ behavior. The opposite situation is also significant: a wear field curve $e_{n_w}$ opposed to a fault-free $e_{n_e}$ EBA prediction. The cyberattack risk is high, as the observed behavior has no prior indication in the sequence on which the prediction is based.

The rules above are expressed formally as follows:

$$\theta_{EBA\_Attack}(e, f) = \begin{cases} \epsilon & \text{if } e = f \text{ or } (e = e_{a_m} \ \& \ f \in E_n) \\ 0.5 & \text{if } e = e_{a_s} \ , \forall f \in F \\ 1 - \epsilon & \text{if } (e = e_{a_m} \ \& \ f = e_{a_p}) \\ & \quad \text{or } (e = e_{a_p} \ \& \ f \in E_n) \\ & \quad \text{or } (e = e_{n_w} \ \& \ f \in \{e_{n_e}, e_{a_p}\}) \\ & \quad \text{or } (e = e_{n_e} \ \& \ f \in \{e_{n_w}, e_{a_p}\}) \end{cases} \tag{2}$$

### *3.3. Step 3: Turnout Unified Threat Assessment (UTA)*

Despite the threat assessment results provided by the TLA and EBA approaches, in certain instances, as outlined in the subsequent section, they fail to individually provide a clear indication of cyberthreat to support decisive actions. This points to the necessity of adopting an approach that merges the estimations derived from these approaches to enhance the detection process.

This section assesses the TLA and EBA approaches to identify situations where clear-cut decisions cannot be made. This assessment enhances the global accuracy of the cyberthreat assessment by proposing a Unified Threat Assessment process that combines the two approaches' results based on assigned trust levels.

#### *3.3.1. Evaluating the TLA and EBA threat assessment approaches*

The TLA approach is based on the *bathtub* curve, which is directly related to turnout failure rates and maintenance operations. The primary blind zone is the burn-in period ($[t_0, \alpha[$), which begins immediately after maintenance or installation and marks the start of the turnout lifecycle at $t_0$. During this period, the TLA threat estimation is 0.5, this value reflects indecision due to the variety of possible turnout behaviors. Furthermore, field knowledge considers that when $\theta_{TLA\_Attack} \in ]0.3, 0.7[$, in other words when $t \in ]0.3\beta + 0.7\alpha, 0.7\beta + 0.3\alpha[$, the TLA cyberattack estimation is not conclusive and clear-cut decisions cannot be made. Consequently, the TLA threat estimations in these zones are inconclusive.

Meanwhile, the EBA approach relies on comparing a field curve with the expected behavior predicted by the forecasting model. Yet, by construction, the forecasting model does not account for sudden turnout failures. Consequently, if the field curve represents a sudden failure $e_{a_s}$, the EBA threat estimation remains inconclusive.

To address the inconclusiveness, trust levels are assigned to each approach based on the moment $t$ during which the switch curve is collected, the type of the field curve $e$, and the type of expected behavior $f$. These trust levels are fixed by considering intervals where each approach results are not conclusive. Let $T = [t_0, t_f]$ represent the turnout life interval. Hence, the trust levels $\omega_{TLA}$ and $\omega_{EBA}$ are defined as $\omega_{TLA}, \omega_{EBA} : \mathbb{R}^+ \times E \times F \rightarrow [0, 1]$, as illustrated in Table 3.

The trust levels presented in Table 3 will be used in next sections to achieve a combined cyberattack estimation. In cases where only one approach is considered (trust level equals to 1), the overall cyberattack estimation corresponds to the one given by that approach. However, when non-zero trusts are assigned to each approach, the aggregation of their results into one single cyberthreat estimation is not straightforward, necessitating a combination solution.

To this end, given the variety of data sources (such as sensors, expert knowledge, predictive models, etc.) and the incomplete nature of information regarding all circumstances concerning switch operations, in addition to the potential conflict between the two approaches, conventional probabilistic approaches like Bayesian inference necessitate a complete comprehensive probabilistic model. They often struggle to handle situations where the information is ambiguous or incomplete. While Fuzzy Logic effectively addresses uncertainty, it lacks a strong framework for combining knowledge from different sources when faced with conflicting information. Weighted Voting, on the other hand, offers a straightforward averaging method that favors the source with the highest weight. Although this simplicity can be beneficial,

Table 3: Trust Levels associated to the TLA and EBA approaches.

| $t$ | $e$ | $f$ | $\omega_{TLA}$ | $\omega_{EBA}$ | Description |
|---|---|---|---|---|---|
| $[t_0, \alpha[$ | $E_n$ | $E_n \cup \{e_{a_p}\}$ | 0 | 1 | TLA approach is inconclusive during the burn-in period |
| | $e_{a_p}$ | $E_n$ | 0 | 1 | TLA approach is inconclusive during the burn-in period |
| | $e_{a_m}$ | $F$ | 0 | 1 | TLA approach is inconclusive during the burn-in period |
| | $e_{a_s}$ | $F$ | 0.5 | 0.5 | Both, TLA and EBA approaches are equally inconclusive when observing a sudden failure during the burn-in period |
| $[\alpha, 0.3\beta + 0.7\alpha]$ *or* $[0.7\beta + 0.3\alpha, \beta[$ | $e_{a_s}$ | $F$ | 1 | 0 | EBA approach is inconclusive when observing a sudden failure |
| | $e_{a_m}$ | $F$ | 0.1 | 0.9 | EBA approach is more accurate, as it allows to differentiate between abnormal behaviors |
| | $else$ | $F$ | 0.4 | 0.6 | EBA approach is slightly more accurate than TLA approach, as it provides better contextualization for each curve $e$ |
| $]0.3\beta + 0.7\alpha, 0.7\beta + 0.3\alpha[$ | $e_{a_s}$ | $F$ | 1 | 0 | EBA approach is inconclusive when observing a sudden failure |
| | $else$ | $F$ | 0.1 | 0.9 | EBA approach is more accurate is the zone where $\theta_{TLA\_Attack} \in ]0.3, 0.7[$ |
| $[\beta, t_f[$ | $e_{n_e}$ | $e_{n_w}$ | 0.9 | 0.1 | TLA approach is more accurate, as it enables more direct reasoning when analyzing fault-free curves |
| | $e_{a_s}$ | $F$ | 1 | 0 | EBA approach is inconclusive when observing a sudden failure |
| | $else$ | $F$ | 0.1 | 0.9 | EBA approach is more accurate, as it enables a more comprehensive analysis of the context |

it may overlook the complexities of conflicting evidence (leading to overconfidence in one outcome over another). In contrast, The Dempster-Shafer (D-S) theory of evidence provides a balanced, conflict-resilient method for combining different sources of evidence that can effectively manage uncertainty and lack of information [86, 87].

*3.3.2. The evidence theory: basic concepts and notations*

The evidence theory, also referred to as the Dempster-Shafer (D-S) theory, is a technique introduced by Dempster [51] and later refined by Shafer [52], that allows reasoning with uncertain, imprecise information without requiring additional assumptions to represent that information [88]. D-S theory is known to provide well-structured rules for combining information, enabling the aggregation of data from various independent sources [53] to support decision-making processes [89, 90].

The universal set of observed situations, also known as the Frame Of Discernment (FOD), is denoted by $S$ and contains N elements.

$$S = \{s_1, s_2, \ldots, s_N\} \tag{3}$$

The power set of $S$, denoted $2^S$, is the set of all subsets of $S$. It includes $2^N$ elements, which are listed as follows:

$$\begin{aligned} 2^S = \{\varnothing, \{s_1\}, \{s_2\}, \ldots, \{s_N\}, \{s_1, s_2\}, \\ \ldots, \{s_1, s_N\}, \ldots, \{s_1, s_2, \ldots, s_i\}, \ldots, S\} \end{aligned} \tag{4}$$

The mass function $m : 2^S \rightarrow [0, 1]$, representing the Basic Belief Assignment (BBA), must satisfy the following conditions:

$$m(\varnothing) = 0 \tag{5}$$

$$\sum_{A \in 2^S} m(A) = 1 \tag{6}$$

Where $A$ is a subset of $2^S$ referred to as a focal element, and $m(A)$ denotes the amount of knowledge associated with every subset $A$ [88].

Let $m_1$ and $m_2$ be two independent BBAs from the same FOD ($S$). According to the D-S theory, these BBAs can be combined using Dempster's combination rule as follows:

$$m_{12}(A) = \begin{cases} \dfrac{\sum\limits_{B \cap C = A \neq \varnothing} m_1(B) m_2(C)}{1 - k}, & A \neq \varnothing \\ 0, & A = \varnothing \end{cases} \tag{7}$$

where $k$ denotes the conflict coefficient, which indicates the level of conflict between $m_1$ and $m_2$ [88]:

$$k = \sum_{B \cap C = \varnothing} m_1(B) m_2(C) \tag{8}$$

When the conflict coefficient $k$ approaches to 1, it indicates a high conflict level between evidences. This issue, initially highlighted by Zadeh in [91], arises when multiple sources contradict each other, leading to counterintuitive results if the Dempster's combination rule in Eq. 7 is applied. In [92], the authors propose an advanced modified D-S combination approach that effectively addresses the high conflict level when compared to various other combination rules.

The modified combination rule utilizes distance or similarity between pieces of evidence, as described in Eq. 9, to compute their "Credibility" weight. This weight represents the relative importance of each piece of evidence [53, 93, 94].

$$sim(m_1, m_2) = \sum_{B \cap C \neq \varnothing} \frac{m_1(B) m_2(C)}{\sqrt{(\sum (m_1(B))^2).(\sum (m_2(C))^2)}} \tag{9}$$

The resulting similarity matrix $s$ between $p$ sources is constructed as follows:

$$s = \begin{bmatrix} 1 & sim(m_1, m_2) & \dots & sim(m_1, m_p) \\ sim(m_2, m_1) & 1 & \dots & sim(m_2, m_p) \\ \dots & \dots & \dots & \dots \\ sim(m_p, m_1) & sim(m_p, m_2) & \dots & 1 \end{bmatrix} \tag{10}$$

A high similarity value between two sources indicates a low distance between them, suggesting they support each other [92]. The support degree of $m_i$, reflecting its reliability, is defined as:

$$sup(m_i) = \sum_{j=1, j \neq i}^{p} sim(m_i, m_j) \quad (i = 1, 2, \dots, p) \tag{11}$$

The credibility weight of each evidence is calculated then by normalizing the support degree, as shown in Eq. 12. Subsequently, the weighted average of the belief provided by each source is computed using Eq. 13:

$$Crd(m_i) = \frac{sup(m_i)}{\sum_{i=1}^{p} sup(m_i)} \quad (i = 1, 2, \ldots, p) \tag{12}$$

$$m_c(A) = \sum_{i=1}^{p} m_i(A).Crd(m_i) \quad (A \in 2^S) \tag{13}$$

Finally, by normalizing the resulting weighted average using Eq. 14, according to [53, 93], the aggregated value $m(A)$ is obtained

$$m(A) = \frac{m_c^2(A)}{\sum_{A \in 2^S} m_c^2(A)} \quad (A \in 2^S) \tag{14}$$

*3.3.3. Combined TLA and EBA threat assessment likelihoods: UTA*

When assessing cyberthreats in the collected turnout data, each approach provides an estimation of its belief regarding the existence of a cyberattack. The D-S theory is applied to combine the cyberthreat assessments from each approach when their assigned trust levels are neither 0 nor 1, as discussed in Section 3.3.1. The two approaches estimations can be represented as $M = \{m_i \mid i = 1, 2\}$ where $m_1$ refers to the estimations of the TLA approach, and $m_2$ refers to the estimations of the EBA approach. Consequently, the FOD $S = \{S_1, S_2\}$ consists of the two possible observed scenarios (situations): cyberattack, denoted by $S_1$ or no-cyberattack, denoted by $S_2$. It is important to note that the cyberattack estimations (BBA) provided by each approach (source) are considered independent, as the degrees of belief generated by each source do not influence one another. Indeed, the two estimations are based on different analysis approaches.

Let $2^S = \{\varnothing, \{S_1\}, \{S_2\}, \{S_1, S_2\}\}$ represent the set of focal elements within the FOD, obtained from the two approaches $m_1$ and $m_2$. These pieces of evidence satisfy the conditions given in Eqs. 5 and 6: $m_1(\varnothing) = m_2(\varnothing) = 0$ and $\sum_{A \in 2^S} m_1(A) = 1$; $\sum_{A \in 2^S} m_2(A) = 1$, where $m_1(A)$ and $m_2(A)$ correspond to the likelihood of the scenario $A$ given by the TLA and EBA approaches, respectively. The focal element $\{S_1, S_2\}$ indicates uncertainty, and the mass of this uncertainty, $m_i(S_1, S_2)$, is considered to be zero. This is because the likelihoods of the two focal elements $S_1$ and $S_2$ are complementary and account for the uncertainty of each approach, as in the case where $m_1(S_1) = m_1(S_2) = 0.5$ which indicates that the two situations $S_1$ and $S_2$ are equally possible. This guarantees that the total belief assigned by each approach to all possible scenarios sums to 1, as required by the D-S theory.

To combine results from the two threat assessment approaches, it is essential to integrate the trust levels associated to each approach, since their trustworthiness varies based on the field curve's type $e$, its appearance time $t$ and the predicted curve's type $f$, as detailed in Table 3. To address this variability, a modified BBA that accounts for varying trust levels can be employed, as suggested by [95, 96]. The modified BBAs, denoted as $\bar{m}_1$ and $\bar{m}_2$, are formulated as follows:

$$\bar{M} = \{\bar{m}_i \mid i = 1, 2\}, \quad \text{where } \bar{m}_i = \omega_i(t, e, f) \times m_i \tag{15}$$

Here, $\omega_1$ and $\omega_2$ represent the trust levels $\omega_{TLA}$, $\omega_{EBA}$ associated with each approach, satisfying the condition $\omega_1 + \omega_2 = 1$.

In the specific context of this article, where only two approaches are involved, Eqs. 11 and 12 become:

$$sup(\bar{m}_1) = sup(\bar{m}_2) = sim(\bar{m}_1, \bar{m}_2) \tag{16}$$

$$Crd(\bar{m}_1) = Crd(\bar{m}_2) = \frac{sup(\bar{m}_1)}{sup(\bar{m}_1) + sup(\bar{m}_2)} = \frac{1}{2} \tag{17}$$

Consequently, the resulting combination approach outlined in Eq. 13 becomes:

$$\bar{m}_c(A) = \frac{1}{2} \sum_{i=1}^{2} \bar{m}_i(A) \quad (A \in 2^S) \tag{18}$$

The final normalized combined belief for $A$ is obtained by:

$$\bar{m}(A) = \frac{\bar{m}_c^2(A)}{\sum_{A \in 2^S} \bar{m}_c^2(A)} \quad (A \in 2^S) \tag{19}$$

From now on, the values of $m_1(S_1)$ and $m_1(S_2)$ are given by $\theta_{TLA\_Attack}$ and $\theta_{TLA\_NoAttack}$, respectively, where $\theta_{TLA\_NoAttack} = 1 - \theta_{TLA\_Attack}$. Similarly, $m_2(S_1)$ and $m_2(S_2)$ are given by $\theta_{EBA\_Attack}$ and $\theta_{EBA\_NoAttack}$ respectively, where $\theta_{EBA\_NoAttack} = 1 - \theta_{EBA\_Attack}$. Tables 4 & 5 summarize likelihoods given by each approach.

Table 4: TLA likelihoods.

| $T$ | $e$ | $\theta_{TLA\_Attack}$ | $\theta_{TLA\_NoAttack}$ |
|---|---|---|---|
| $[\alpha, \beta[$ | $E_a \cup \{e_{n_w}\}$ | $\frac{\beta - t}{\beta - \alpha} - \epsilon$ | $\frac{t - \alpha}{\beta - \alpha} + \epsilon$ |
| $[\alpha, \beta[$ | $e_{n_e}$ | $\frac{t - \alpha}{\beta - \alpha} + \epsilon$ | $\frac{\beta - t}{\beta - \alpha} - \epsilon$ |
| $[\beta, t_f[$ | $E_a \cup \{e_{n_w}\}$ | $\epsilon$ | $1 - \epsilon$ |
| $[\beta, t_f[$ | $e_{n_e}$ | $1 - \epsilon$ | $\epsilon$ |

Table 5: EBA likelihoods.

| $e$ | $f$ | $\theta_{EBA\_Attack}$ | $\theta_{EBA\_NoAttack}$ |
|---|---|---|---|
| $e$ | $e$ | $\epsilon$ | $1 - \epsilon$ |
| $e_{a_m}$ | $E_n$ | $\epsilon$ | $1 - \epsilon$ |
| $e_{a_m}$ | $e_{a_p}$ | $1 - \epsilon$ | $\epsilon$ |
| $e_{a_p}$ | $E_n$ | $1 - \epsilon$ | $\epsilon$ |
| $e_{n_w}$ | $\{e_{n_e}, e_{a_p}\}$ | $1 - \epsilon$ | $\epsilon$ |
| $e_{n_e}$ | $\{e_{n_w}, e_{a_p}\}$ | $1 - \epsilon$ | $\epsilon$ |

This will allow calculating, $\Theta$, the combined cyberthreat evaluation, that is given by $\bar{m}$ in Eq. 19. Where $\Theta_{Attack}$ and $\Theta_{NoAttack}$ are given by $\bar{m}(S_1)$ and $\bar{m}(S_2)$, respectively. Specifically, $\Theta_{Attack}$ and $\Theta_{NoAttack}$ are defined as $\Theta_{Attack}, \Theta_{NoAttack} : Q \times E \times F \times \mathbb{R}^+ \rightarrow ]0, 1[$.

## 4. Application: results & discussion

In this section, the proposed cyberattack detection framework is applied to a turnout dataset, and the results are discussed. The dataset used in this study was provided by our industrial partner. The labeling process is performed in the development phase (see Section 4.2) by the classification model and consulting the industrial partner. This process results in categorizing each curve as normal, i.e. without faults, or abnormal, i.e. with faults. These labeled curves do not reveal if their behaviors is resulted from an attack of a real turnout behavior. Thus the available dataset does not contain documented cyberattacks that are aligned with FM and FI attacks. In the absence of publicly available datasets for this specific context, cyberattack scenarios are simulated by replacing existing genuine normal/abnormal turnout behavior curves with alternative abnormal (FI) or normal (FM) ones. This corresponds to an attacker who had perfect knowledge which curve is normal and which is abnormal as stated in Hypothesis 1. These scenarios were designed to produce a range of detection outcomes allowing accurate evaluation of the individual and combined performance of the proposed detection approaches.

### *4.1. Overview*

The diagram in Fig. 10 provides a scattered view of the cyberattack detection system in its physical environment. The development phase has resulted in the production of the Data Processor (DP) block, which contains the classification and forecasting models. The TLA and EBA functions compute each a cyberthreat likelihoods from each field curve. Subsequently, the UTA function combines these likelihoods into a single global likelihood, which is communicated to maintenance operators via CMS to assist them in making maintenance decisions.

This work is illustrated using a raw dataset consisting of 3879 switch operations, each represented by a power consumption curve.

The raw dataset is split into two subsets: a training dataset consisting of 2480 curves, and a validation dataset consisting of 1399 curves. Each curve contains 252 points. The training dataset is used to learn curves, while the validation dataset is used to assess the model's accuracy. This is achieved by relying on the basic learning/testing approach supported by the Scikit-learn library.

In the following, the process of developing the classification and forecasting models is outlined. Subsequently, the results obtained from applying the proposed cyberattack detection framework to the validation dataset are presented and discussed.

### *4.2. The Development Phase*

The training dataset is analyzed to understand the turnout's behavior. It is assumed that this dataset is uncompromised and was securely obtained.

**Classification model development.** This model is at the core of the TLA and EBA approaches. The first goal is to identify different behavior types through a semantic analysis step, where dataset similarities were explored to differentiate various shapes of power consumption curves. By applying k-Means clustering within the Scikit-learn environment [97] and consulting railway experts, five distinct clusters are identified: $e_{n_e}$, $e_{n_w}$, $e_{a_p}$, $e_{a_m}$ and $e_{a_s}$,

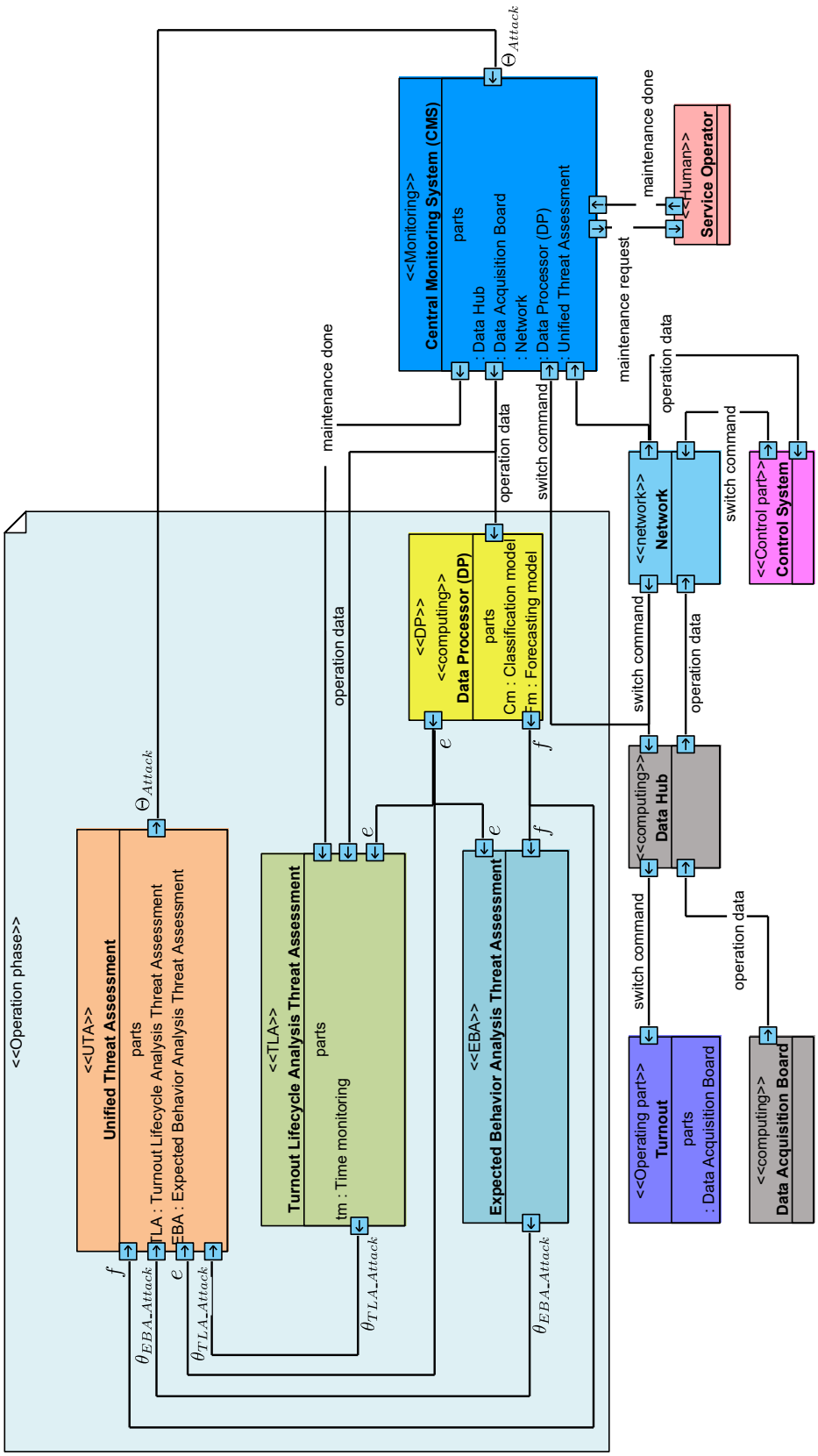



Figure 10: Switch operations monitoring process.

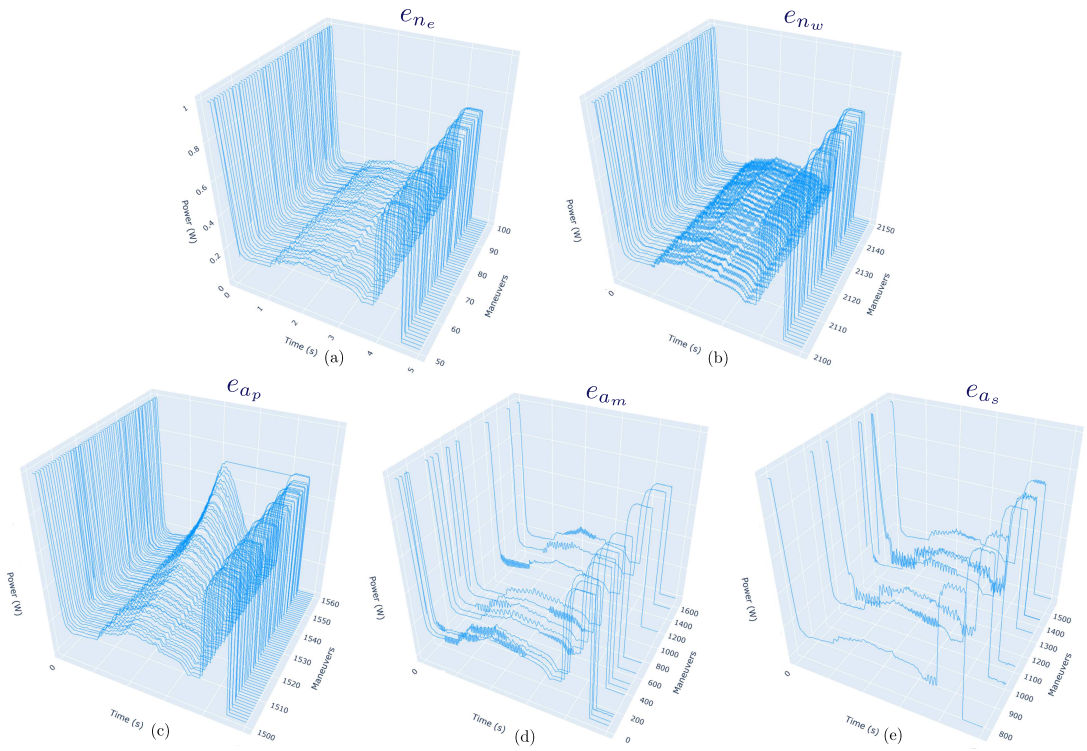



Figure 11: Turnout behaviors classes.

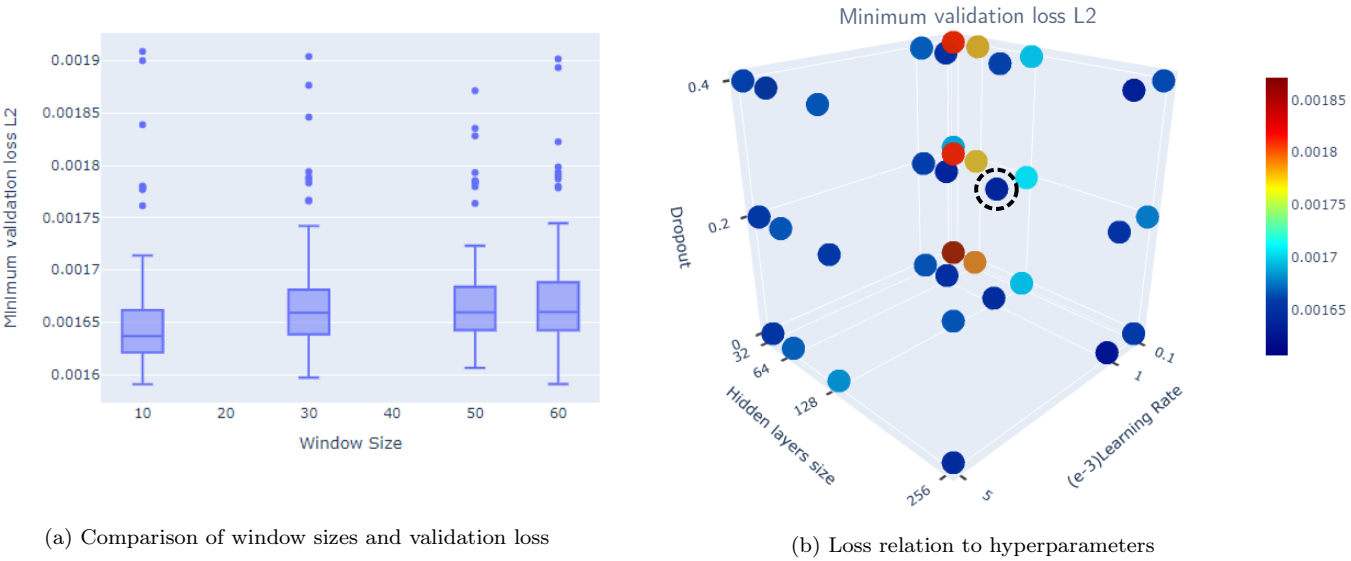
as listed in Table 1 and illustrated in Fig. 11. The clustering results were employed to train a KNN classification model which labels all switch operation power curves.

**LSTM model development.** This model is at the core of the EBA approach. It aims to analyze turnout behavior sequences and predict its potential future behaviors. For that, it is assumed that the used training dataset is free from sudden failures $e_{a_s}$, therefore it contains normal behaviors $E_n$, progressive abnormalities $e_{a_p}$ and minor abnormalities $e_{a_m}$. This exclusion is essential because the forecasting model is trained using field data, which is at the moment unable to exhibit sufficient scenarios to learn and predict sudden failures.

The performance of the obtained LSTM forecasting model is directly related to hyper-parameters choice listed in Table 6. The results of a grid search procedure are summarized below.

Table 6: LSTM model hyperparameters.

| Hyperparameter | Range | Choice |
|---|---|---|
| Window size | $[10, 30, 50, 60]$ | 50 |
| Hidden states | $[32, 64, 128, 256]$ | 128 |
| Dropout | $[0.0, 0.2, 0.4]$ | 0.2 |
| Learning rate | $[0.0001, 0.001, 0.005]$ | 0.001 |
| Batch | $[16, 32, 64]$ | 32 |




(a) Comparison of window sizes and validation loss

(b) Loss relation to hyperparameters

Figure 12: LSTM hyperparameters choice.

The grid search first step focuses on selecting the optimal window size through an evaluation of the model's accuracy by varying all the identified hyperparameters and computing the corresponding validation loss L2 (Mean Square Error MSE) over 100 epochs. Fig. 12a presents the model's accuracy computed as the minimum validation loss observed across different window sizes. While a window size of 10 field curves results in the lowest median validation loss, a window size of 50 curves shows less dispersion in the performance figures with fewer outliers. Performance with a window size of 30 curves exhibits high variability, despite a median loss comparable to other window sizes. Similarly, performance with a window size of 60 curves is comparable to the 50-curve window but exhibits slightly more variability. This suggests that, despite not having the absolute lowest validation loss, a window size of 50 curves is better because it yields a narrower L2 interval and better captures longer-term dependencies in power curves. This allows the model to account for more historical information, potentially leading to more accurate predictions.

After setting the LSTM window size to 50, various combinations of LSTM hidden state sizes, dropout rates and learning rates were evaluated. The resulting minimum validation losses are illustrated in Fig. 12b. With 256 hidden states, more variability in performance is observed, with some points indicating higher losses (lighter colors), potentially due to the size being larger than necessary. Conversely, hidden state sizes of 32 and 64 may be too small, failing to adequately capture the data variation. The LSTM model with 128 hidden states consistently results in lower validation loss (darker blue points), indicating a good balance in the model's performance. Regarding dropout rates, a rate of 0.2 yields better results, as shown by the concentration of darker blue points. For the learning rate, the standard value of 0.001 consistently produces good performance, suggesting that the model converges well with this rate, unlike other learning rates, which tend to result in higher losses. Therefore, the optimal configuration appears to be 128 hidden states, a dropout rate of 0.2, and a learning rate of 0.001.

The KNN and LSTM models obtained are both instrumental for implementing the oper-

ation phase. It is worth noting that given the data set at hand, KNN did not raise particular accuracy issues, unlike LSTM. The upstream study and improvement of these models' accuracy allows for a significant choice: prevent the introduction of additional uncertainty into the aggregation process through the Dempster-Shafer method.

### *4.3. The Operation Phase*

The raw data available does not feature documented cyberattack scenarios. For the current study, specifically for FM and FI attacks, cyberattack scenarios have been handcrafted, using previously seen turnout behaviors, and added to the global dataset to assess the validity of the proposed framework. These scenarios are illustrated in Fig. 13, where $A_1$, $A_2$, $A_3$ and $A_4$ correspond to FI attacks, while $A_5$, $A_6$ and $A_7$ represent FM attacks. Descriptions of these scenarios are listed in Table 7, where $t_0$ stands for the installation/maintenance date of the turnout.

Table 7: Cyberattacks descriptions.

| Label | Date (days) | Description | Attack's type |
|---|---|---|---|
| $A_1$ | $t_0 + 8d$ | Sudden failure behavior | Fault Injecting attack |
| $A_2$ | $t_0 + 18d$ | Aging behavior | Fault Injecting attack |
| $A_3$ | $t_0 + 24d$ | Progressive pre-fault degradation behavior | Fault Injecting attack |
| $A_4$ | $t_0 + 56d$ | Sudden failure behavior | Fault Injecting attack |
| $A_5$ | $t_0 + 90d$ | Minor fault behavior | Fault Masking attack |
| $A_6$ | $t_0 + 94d$ | Normal, fault-free behavior | Fault Masking attack |
| $A_7$ | $t_0 + 149d$ | Normal, fault-free behavior | Fault Masking attack |

#### *4.3.1. The TLA threat assessment*

The TLA threat assessment is derived from the direct application of Eq. 1, which requires information about the turnout's lifecycle and each observed field curve. Key inputs include $\alpha$ and $\beta$, which identify the time interval $T$ related to each period from the turnout's *bathtub* curve, as well as specific details on each switch operation: the observed behavior $e$ and its corresponding timing $t$. Due to the reduced scope of the available dataset, all lifecycle intervals have been scaled down. Due to confidentiality constraints, $\alpha$ and $\beta$ are not given as absolute dates. They are hence defined as $t_0 + 6d$ and $t_0 + 123d$, respectively. This has a limited impact on the genericity of this study, because the nature of the observations does not change from one time interval to another: only the detection logic evolves.

Table 8 presents a selection of threat estimations provided by the TLA approach. The results shown in the table have been selected to illustrate the TLA approach operation and offer a representative overview of its outcomes based on different 18 situations (field inputs).

The TLA approach identifies 7 turnout behaviors as potential cyberattacks on $t_0 + 8d$, $t_0 + 18d$, $t_0 + 24d$, $t_0 + 44d$, $t_0 + 94d$, $t_0 + 114d$ and $t_0 + 149d$, as shown in Table 8. However, no decisions can be made for 3 curves observed on $t_0 + 4d$, $t_0 + 5d$ and $t_0 + 56d$ since they could represent either a cyberattack or a legitimate behavior. The remaining behaviors appear to be legitimate as they exhibit consistency with the turnout expected lifecycle. Further evaluations are discussed in Section 4.3.4.

Table 8: TLA threat assessment results with $\epsilon = 0.1$.

| $T$ | $t$ | $e$ | $\theta_{TLA\_Attack}$ | $\theta_{TLA\_NoAttack}$ | Description |
|---|---|---|---|---|---|
| $[t_0, t_0 + 6d[$ | $t_0 + 4d$ | $e_{a_m}$ | 0.5 | 0.5 | TLA approach is inconclusive during the burn-in period |
| | $t_0 + 5d$ | $e_{n_e}$ | 0.5 | 0.5 | TLA approach is inconclusive during the burn-in period |
| $[t_0 + 6d, t_0 + 123d[$ | $\boldsymbol{t_0 + 8d}$ | $e_{a_s}$ | **0.9** | 0.1 | Cyberattack is detected due to the observation of a sudden failure in the beginning of the turnout useful life |
| | $t_0 + 10d$ | $e_{n_e}$ | 0.1 | 0.9 | No suspicion is considered due to the observation of a normal, fault-free behavior |
| | $\boldsymbol{t_0 + 18d}$ | $e_{n_w}$ | **0.8** | 0.2 | Cyberattack is detected due to the observation of an aging behavior in early stages of the turnout useful life |
| | $\boldsymbol{t_0 + 24d}$ | $e_{a_p}$ | **0.7** | 0.3 | Cyberattack is detected due to the observation of a progressive pre-fault degradation behavior in early stages of the turnout useful life |
| | $\boldsymbol{t_0 + 44d}$ | $e_{a_m}$ | **0.6** | 0.4 | Cyberattack is detected due to the observation of a minor fault in early stages of the turnout useful life |
| | $t_0 + 56d$ | $e_{a_s}$ | 0.5 | 0.5 | TLA approach is inconclusive near the mid useful life |
| | $t_0 + 90d$ | $e_{a_m}$ | 0.2 | 0.8 | No suspicion is considered due to the observation of a minor fault in late stages of the useful life |
| | $t_0 + 92d$ | $e_{a_p}$ | 0.2 | 0.8 | No suspicion is considered due to the observation of a progressive pre-fault degradation behavior in late stages of the useful life |
| | $\boldsymbol{t_0 + 94d}$ | $e_{n_e}$ | **0.8** | 0.2 | Cyberattack is detected due to the observation of a normal, fault-free behavior in late stages of the turnout useful life |
| | $\boldsymbol{t_0 + 114d}$ | $e_{n_e}$ | **0.9** | 0.1 | Cyberattack is detected due to the observation of a normal, fault-free behavior in late stages of the turnout useful life |
| $[t_0 + 123, t_0 + 155d[$ | $t_0 + 125d$ | $e_{a_m}$ | 0.1 | 0.9 | No suspicion is considered due to the observation of a minor fault in the wear-out period |
| | $t_0 + 126d$ | $e_{a_m}$ | 0.1 | 0.9 | No suspicion is considered due to the observation of a minor fault in the wear-out period |
| | $t_0 + 143d$ | $e_{a_m}$ | 0.1 | 0.9 | No suspicion is considered due to the observation of a minor fault in the wear-out period |
| | $\boldsymbol{t_0 + 149d}$ | $e_{n_e}$ | **0.9** | 0.1 | Cyberattack is detected due to the observation of a normal, fault-free behavior in the wear-out period |
| | $t_0 + 151d$ | $e_{a_s}$ | 0.1 | 0.9 | No suspicion is considered due to the observation of a sudden failure in the wear-out period |
| | $t_0 + 154d$ | $e_{n_w}$ | 0.1 | 0.9 | No suspicion is considered due to the observation of an aging behavior in the wear-out period |

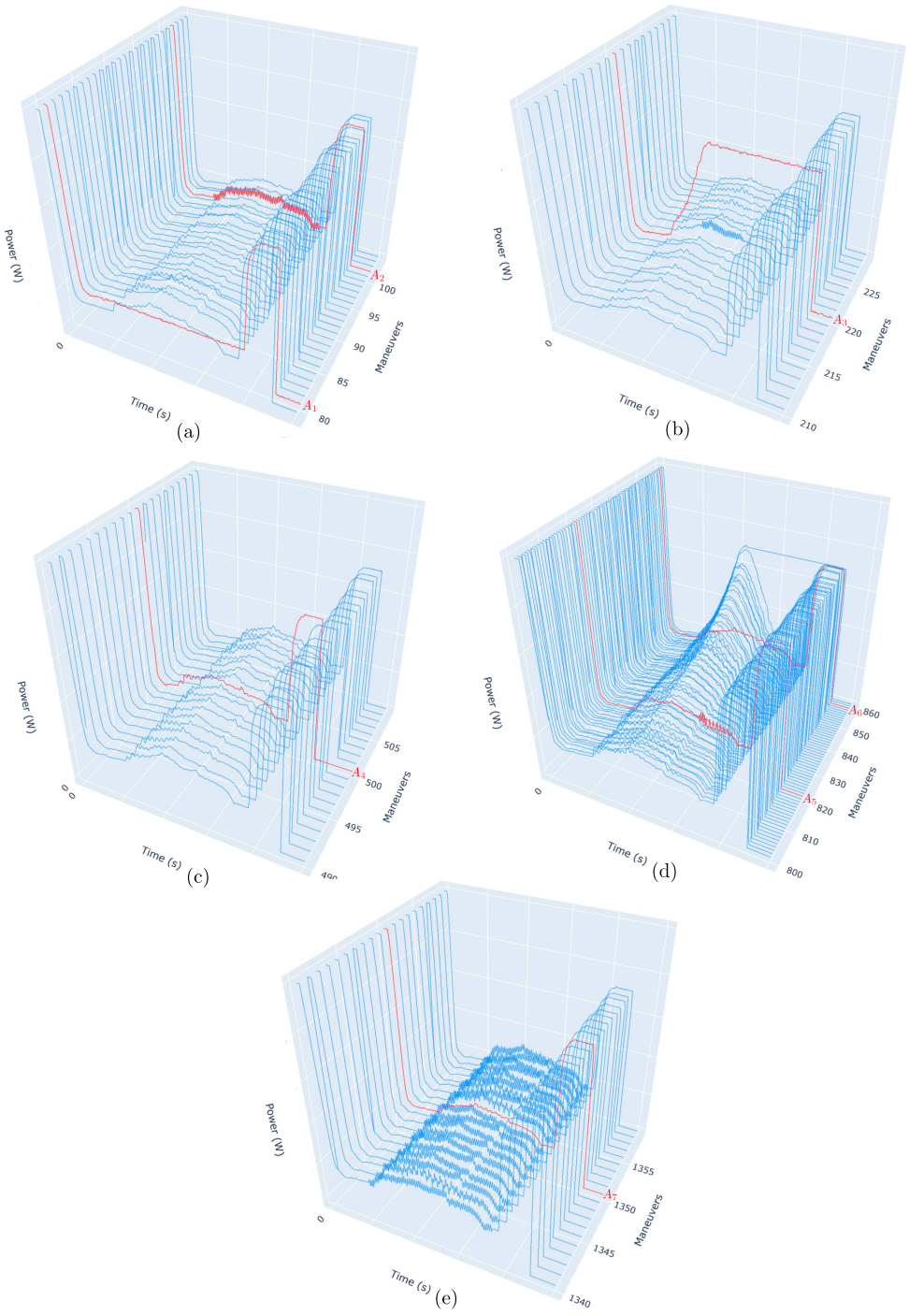



Figure 13: Cyberattack scenarios.

It is worth noting that it is difficult to work around indecision ($\theta_{TLA\_Attack} = 0.5$) in the middle of the useful life time interval. The same also goes for the early life interval. The EBA assessment provides additional insight as shown in the next section.

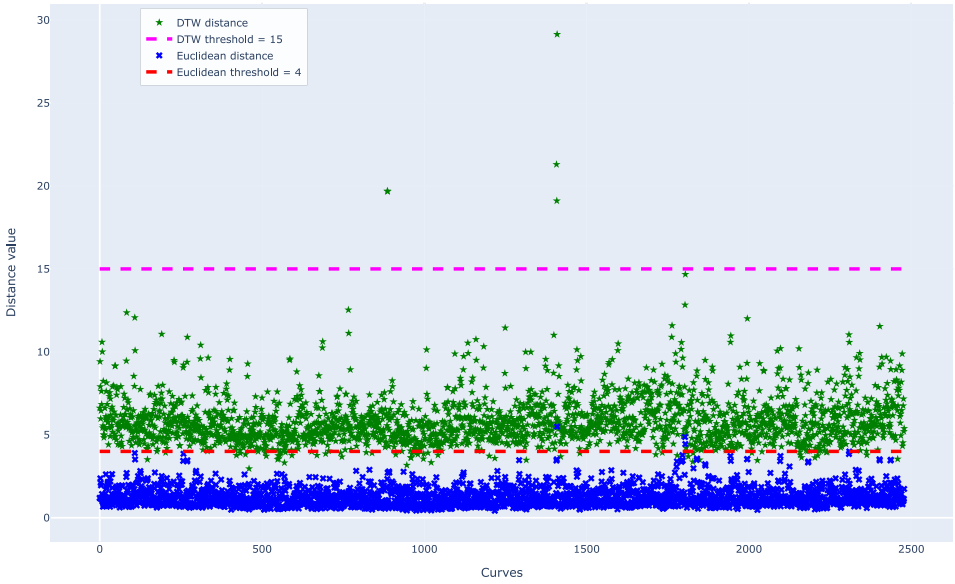



Figure 14: Euclidean and DTW distance thresholds.

*4.3.2. The EBA threat assessment*

To investigate cyberattacks based on the temporal evolution of turnout behavior, the LSTM forecasting model is used with sequences of 50 power curves each, starting with an initial sequence acquired securely to ensure the integrity of the prediction process. The predicted curve is then compared to the observed field curve using fixed thresholds for Euclidean and DTW distances. Fig. 14 illustrates the computed distances between curves from the training dataset that were used to establish these thresholds and define when two curves are considered similar or distinct. Consequently, 4 and 15 were fixed as thresholds for Euclidean and DTW distances, respectively.

The Table 9 lists the 11 turnout behaviors that were identified as deviant from expected behaviors based on a comparison using the established distance thresholds. ED and DTW correspond to the computed Euclidean and DTW distances, respectively, between expected and observed behaviors. By comparing the observed curve behavior $e$ with the expected curve $f$, and considering cases in Eq. 2, cyberattack evaluations are performed for each curve from the 11 curves, as outlined in the Description column in Table 9. Of these, 4 behaviors are found to represent cyberattacks (observed on days 18, 24, 90 and 94 after $t_0$, as highlighted in Table 9). The curves observed on days 18, 24, and 94 after $t_0$ are illustrated in Fig. 15(b), 15(c) and 15(e), respectively, along with their associated predicted curves. Three behaviors from the 11 deviated behaviors are associated with safe switch operations (observed on days 94, 102 and 126 after $t_0$ and illustrated in Fig. 16(a), 16(b) and 16(c), respectively. As for the remaining four curves (observed on days 8, 56, 127 and 151 day after $t_0$ and shown in Fig. 15(a), 15(d), Fig. 16(d) and Fig. 16(e), respectively), a clear decision of whether they represent cyberattacks could not be made. Further evaluations are discussed in Section 4.3.4.

Table 9: EBA threat assessment results with $\epsilon = 0.1$.

| $t$ | ED | DTW | $e$ | $f$ | $\theta_{EBA\_Attack}$ | $\theta_{EBA\_NoAttack}$ | Description |
|---|---|---|---|---|---|---|---|
| $t_0 + 8d$ | 2.7 | 28.98 | $e_{a_s}$ | $e_{n_e}$ | 0.5 | 0.5 | Decision cannot be made due to the observation of a sudden failure |
| $\boldsymbol{t_0 + 18d}$ | 1.85 | 16.22 | $e_{n_w}$ | $e_{n_e}$ | **0.9** | 0.1 | Cyberattack is detected since the expected behavior is a fault-free behavior, while the observed curve reflects wear of the turnout |
| $\boldsymbol{t_0 + 24d}$ | 13.21 | 126.29 | $e_{a_p}$ | $e_{n_e}$ | **0.9** | 0.1 | Cyberattack is detected since the observed curve is related to a progressive pre-faults degradation that appear without prior notice |
| $t_0 + 56d$ | 17.36 | 71.70 | $e_{a_s}$ | $e_{n_e}$ | 0.5 | 0.5 | Decision cannot be made due to the observation of a sudden failure |
| $\boldsymbol{t_0 + 90d}$ | 2.06 | 17.05 | $e_{a_m}$ | $e_{a_p}$ | **0.9** | 0.1 | Cyberattack is detected since the observed curve interrupts a sequence of a progressive pre-faults degradation |
| $\boldsymbol{t_0 + 94d}$ | 7.97 | 76.80 | $e_{n_e}$ | $e_{a_p}$ | **0.9** | 0.1 | Cyberattack is detected since the observed curve interrupts a sequence of a progressive pre-faults degradation |
| $t_0 + 94d$ | 11.84 | 123.18 | $e_{a_p}$ | $e_{a_p}$ | 0.1 | 0.9 | No suspicion is considered since the observed curve is related to the progressive pre-faults degradation |
| $t_0 + 102d$ | 3.11 | 20.62 | $e_{a_m}$ | $e_{n_e}$ | 0.1 | 0.9 | No suspicion is considered since the observed curve is related to minor fault |
| $t_0 + 126d$ | 1.10 | 15.56 | $e_{a_m}$ | $e_{n_w}$ | 0.1 | 0.9 | No suspicion is considered since the observed curve is related to minor fault |
| $t_0 + 127d$ | 5.21 | 32.50 | $e_{a_s}$ | $e_{n_w}$ | 0.5 | 0.5 | Decision cannot be made due to the observation of a sudden failure |
| $t_0 + 151d$ | 1.45 | 15.32 | $e_{a_s}$ | $e_{n_w}$ | 0.5 | 0.5 | Decision cannot be made due to the observation of a sudden failure |

*4.3.3. The Unified Threat Assessment*

To combine the threat estimations from the TLA and EBA approaches, the 18 turnout behaviors used to illustrate the TLA approach results, and the 11 behaviors considered using the EBA approach, are combined and assigned weights according to the cases discussed in Table 3. Then, the modified D-S combination approach equation in Eq. 19 is applied. The results of the combined threat estimations for each observed curve are listed in Table 10.

The UTA detects cyberattacks on days 8, 18, 24, 90, 94 and 149 after $t_0$, as highlighted in Table 10. However, an ambiguous situation is observed on day 56 after $t_0$, where $\Theta_{Attack} = \Theta_{NoAttack} = 0.5$. The remaining observed curves are considered legitimate, reflecting true turnout behavior. Further evaluations are discussed in the next section.

*4.3.4. Evaluation, discussion & future work*

The evaluation of the proposed framework is based on two criteria: its effectiveness compared to the TLA and EBA threat assessment results and its accuracy in detecting FM and FI cyberattacks.

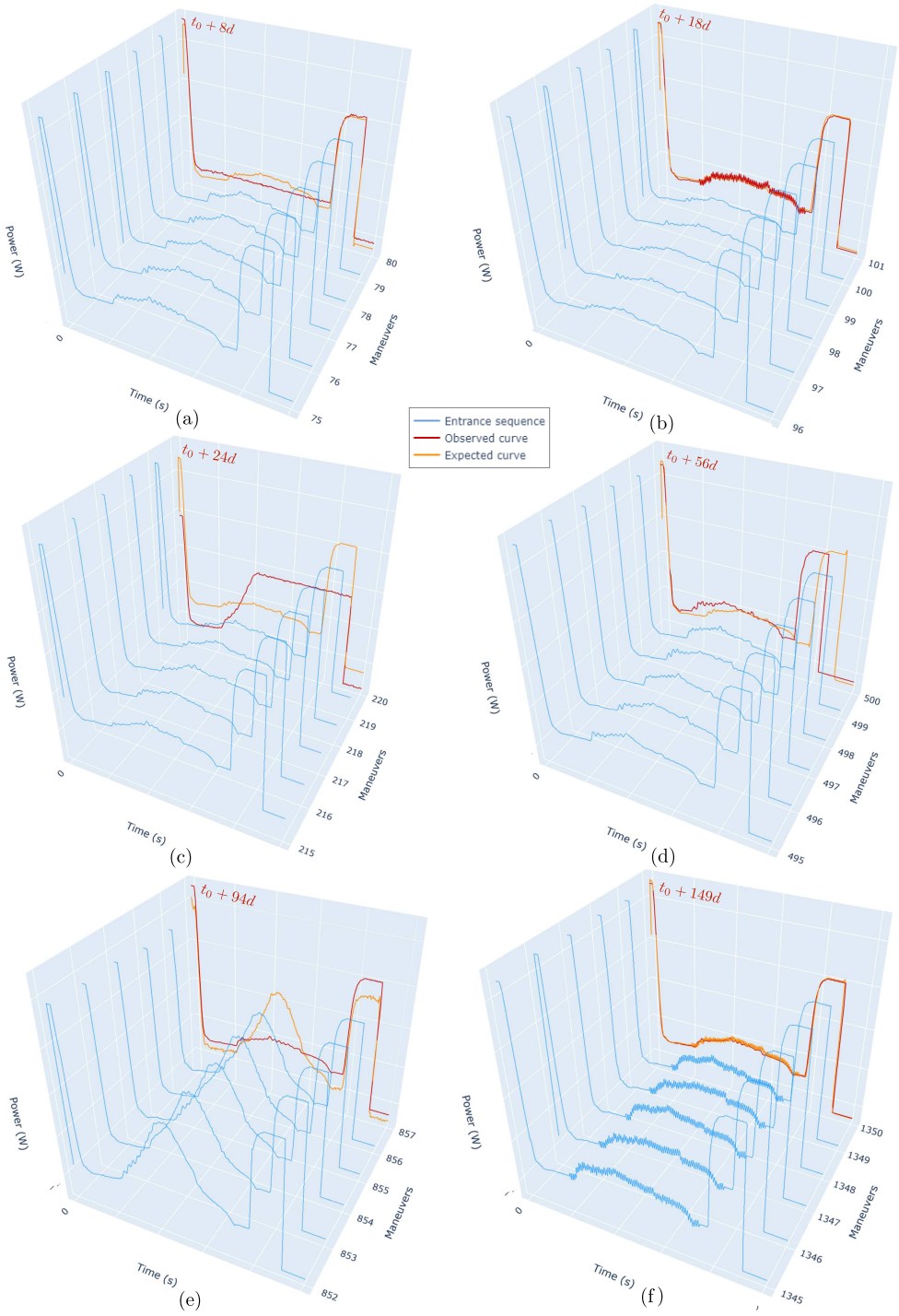



Figure 15: Cyberattack scenarios: observations vs. predictions.

As detailed in Table 11, the cyberattack results for each observed curve are presented alongside the true state to assess the reliability of the cyberattack estimations. Of the seven injected cyberattacks, the TLA approach detected five, the EBA approach detected four, while the UTA successfully detected six. In terms of false positives, only the TLA approach mistakenly identified two legitimate behaviors as cyberattacks (on days $t_0 + 44d$

and $t_0 + 114d$). The first false positive occurred during a minor fault, which was considered suspicious during the early stage of the useful life period when the TLA approach expected only fault-free behaviors. The second false positive was a normal, fault-free behavior observed during a later stage of the useful life, when the TLA approach anticipated faulty behavior instead.

The results also show that the TLA approach fails to detect the $A_5$ cyberattack, while

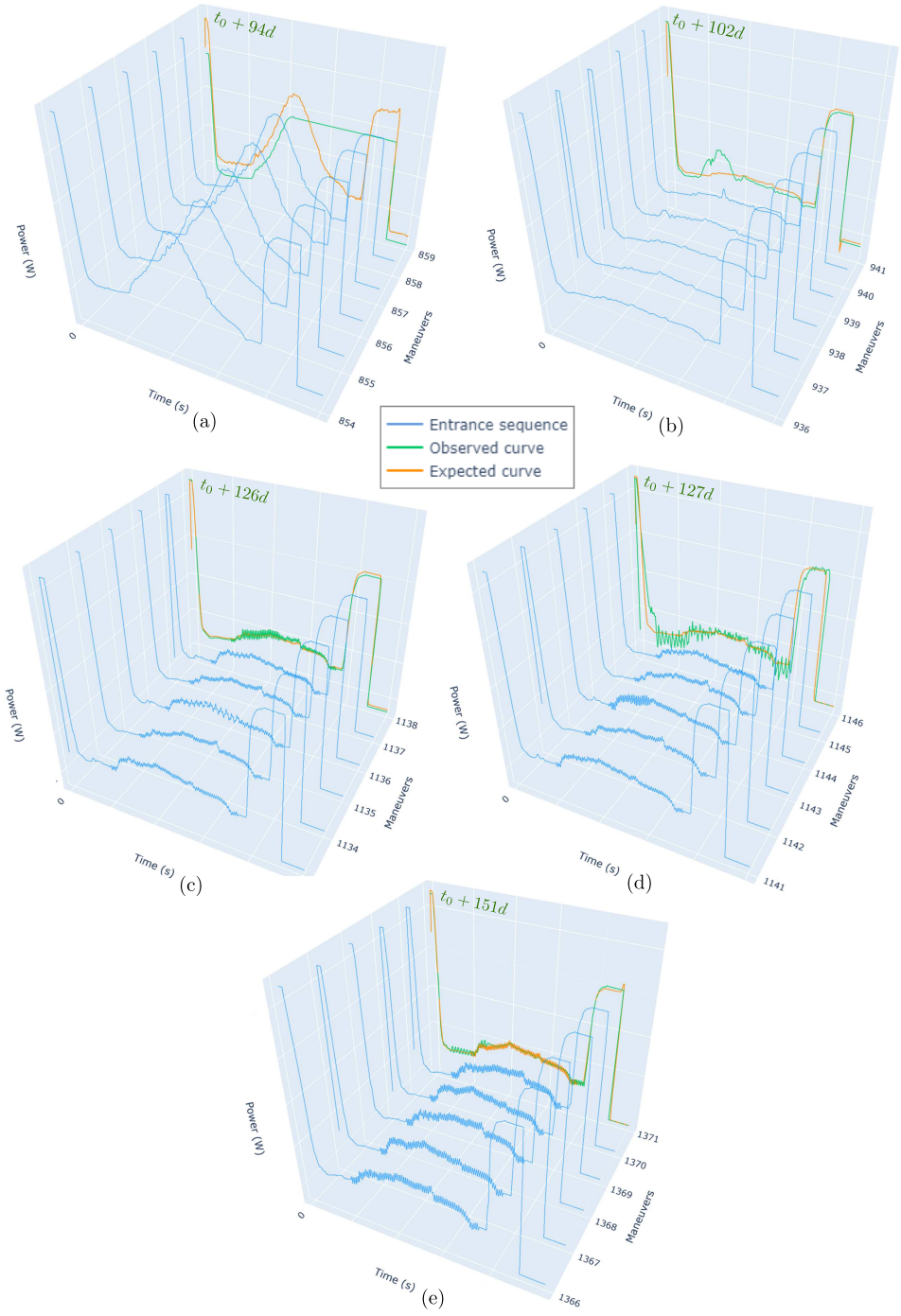



Figure 16: Legitimate observations vs. predictions.

Table 10: UTA threat assessment results.

| $t$ | $e$ | $f$ | $\omega_{TLA}$ | $\omega_{EBA}$ | Attack | | | No Attack | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | TLA | EBA | UTA | TLA | EBA | UTA |
| $t_0 + 4d$ | $e_{a_m}$ | $f$ | 0 | 1 | 0.5 | 0.1 | 0.1 | 0.5 | 0.9 | 0.9 |
| $t_0 + 5d$ | $e_{n_e}$ | $f$ | 0 | 1 | 0.5 | 0.1 | 0.1 | 0.5 | 0.9 | 0.9 |
| $\boldsymbol{t_0 + 8d}$ | $e_{a_s}$ | $e_{n_e}$ | 1 | 0 | 0.9 | 0.5 | **0.9** | 0.1 | 0.5 | 0.1 |
| $t_0 + 10d$ | $e_{n_e}$ | $f$ | 0.4 | 0.6 | 0.1 | 0.1 | 0.02 | 0.9 | 0.9 | 0.98 |
| $\boldsymbol{t_0 + 18d}$ | $e_{n_w}$ | $e_{n_e}$ | 0.4 | 0.6 | 0.8 | 0.9 | **0.97** | 0.2 | 0.1 | 0.03 |
| $\boldsymbol{t_0 + 24d}$ | $e_{a_p}$ | $e_{n_e}$ | 0.4 | 0.6 | 0.7 | 0.9 | **0.95** | 0.3 | 0.1 | 0.05 |
| $t_0 + 44d$ | $e_{a_m}$ | $f$ | 0.1 | 0.9 | 0.6 | 0.1 | 0.03 | 0.4 | 0.9 | 0.97 |
| $t_0 + 56d$ | $e_{a_s}$ | $e_{n_e}$ | 1 | 0 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 |
| $\boldsymbol{t_0 + 90d}$ | $e_{a_m}$ | $e_{a_p}$ | 0.1 | 0.9 | 0.2 | 0.9 | **0.96** | 0.8 | 0.1 | 0.04 |
| $t_0 + 92d$ | $e_{a_p}$ | $f$ | 0.4 | 0.6 | 0.2 | 0.1 | 0.03 | 0.8 | 0.9 | 0.97 |
| $\boldsymbol{t_0 + 94d}$ | $e_{n_e}$ | $e_{a_p}$ | 0.4 | 0.6 | 0.8 | 0.9 | **0.97** | 0.2 | 0.1 | 0.03 |
| | $e_{a_p}$ | $e_{a_p}$ | 0.4 | 0.6 | 0.2 | 0.1 | 0.03 | 0.8 | 0.9 | 0.97 |
| $t_0 + 102d$ | $e_{a_m}$ | $e_{n_e}$ | 0.1 | 0.9 | 0.1 | 0.1 | 0.02 | 0.9 | 0.9 | 0.98 |
| $t_0 + 114d$ | $e_{n_e}$ | $f$ | 0.4 | 0.6 | 0.9 | 0.1 | 0.34 | 0.1 | 0.9 | 0.66 |
| $t_0 + 125d$ | $e_{a_m}$ | $f$ | 0.1 | 0.9 | 0.1 | 0.1 | 0.03 | 0.9 | 0.9 | 0.97 |
| $t_0 + 126d$ | $e_{a_m}$ | $e_{n_w}$ | 0.1 | 0.9 | 0.1 | 0.1 | 0.03 | 0.9 | 0.9 | 0.97 |
| $t_0 + 127d$ | $e_{a_s}$ | $e_{n_w}$ | 1 | 0 | 0.1 | 0.5 | 0.1 | 0.9 | 0.5 | 0.9 |
| $t_0 + 143d$ | $e_{a_m}$ | $f$ | 0.1 | 0.9 | 0.1 | 0.1 | 0.03 | 0.9 | 0.9 | 0.97 |
| $\boldsymbol{t_0 + 149d}$ | $e_{n_e}$ | $e_{n_w}$ | 0.9 | 0.1 | 0.9 | 0.1 | **0.95** | 0.1 | 0.9 | 0.05 |
| $t_0 + 151d$ | $e_{a_s}$ | $e_{n_w}$ | 1 | 0 | 0.1 | 0.5 | 0.1 | 0.9 | 0.5 | 0.9 |
| $t_0 + 154d$ | $e_{n_w}$ | $f$ | 0.1 | 0.9 | 0.1 | 0.1 | 0.03 | 0.9 | 0.9 | 0.97 |

the EBA approach missed the $A_7$ cyberattack, mistakenly classifying them as legitimate behaviors. The $A_5$ was overlooked by the TLA approach because it interpreted the faulty behavior observed as the turnout approached the wear-out period as normal. However, the EBA approach was able to analyze the sequence of behaviors leading up to $A_5$ and correctly identified the corresponding curve as a cyberattack, as it interrupted a progressive pre-fault degradation. Conversely, the EBA approach missed the $A_7$ cyberattack because the computed distance between the predicted and observed curves on $t_0 + 149d$ was below the set thresholds, leading to an incorrect classification of the curve as legitimate. The UTA, however, successfully detected both cyberattacks due to the weighted D-S combination of approaches: it assigned a higher weight to the EBA approach when handling progressive pre-fault degradation, and a higher weight to the TLA approach when analyzing normal aging behavior during the wear-out period. This strategic weighting allowed UTA to make more accurate decisions in both cases.

Ambiguity was the most common issue affecting cyberattack decisions using TLA or EBA. The TLA approach reported three cases of ambiguity. Two of these curves are collected during the burn-in period, while the third curve reflects a sudden failure occurring midway through the useful life period, which led to equal evaluation for both cyberattack and no-cyberattack scenarios. The EBA approach identified four ambiguous situations after observing sudden failure behaviors. However, the UTA was unable to provide a clear decision

Table 11: Cyberattack detection results evaluation.

| $t$ | $e$ | True state status | TLA | EBA | UTA |
|---|---|---|---|---|---|
| $t_0 + 4d$ | $e_{a_m}$ | Legitimate | AM | TN | TN |
| $t_0 + 5d$ | $e_{n_e}$ | Legitimate | AM | TN | TN |
| $t_0 + 8d$ | $e_{a_s}$ | Cyberattack $A_1$ | TP | AM | TP |
| $t_0 + 10d$ | $e_{n_e}$ | Legitimate | TN | TN | TN |
| $t_0 + 18d$ | $e_{n_w}$ | Cyberattack $A_2$ | TP | TP | TP |
| $t_0 + 24d$ | $e_{a_p}$ | Cyberattack $A_3$ | TP | TP | TP |
| $t_0 + 44d$ | $e_{a_m}$ | Legitimate | FP | TN | TN |
| $t_0 + 56d$ | $e_{a_s}$ | Cyberattack $A_4$ | AM | AM | AM |
| $t_0 + 90d$ | $e_{a_m}$ | Cyberattack $A_5$ | FN | TP | TP |
| $t_0 + 92d$ | $e_{a_p}$ | Legitimate | TN | TN | TN |
| $t_0 + 94d$ | $e_{n_e}$ | Cyberattack $A_6$ | TP | TP | TP |
| | $e_{a_p}$ | Legitimate | TN | TN | TN |
| $t_0 + 102d$ | $e_{a_m}$ | Legitimate | TN | TN | TN |
| $t_0 + 114d$ | $e_{n_e}$ | Legitimate | FP | TN | TN |
| $t_0 + 125d$ | $e_{a_m}$ | Legitimate | TN | TN | TN |
| $t_0 + 126d$ | $e_{a_m}$ | Legitimate | TN | TN | TN |
| $t_0 + 127d$ | $e_{a_s}$ | Legitimate | TN | AM | TN |
| $t_0 + 143d$ | $e_{a_m}$ | Legitimate | TN | TN | TN |
| $t_0 + 149d$ | $e_{n_e}$ | Cyberattack $A_7$ | TP | FN | TP |
| $t_0 + 151d$ | $e_{a_s}$ | Legitimate | TN | AM | TN |
| $t_0 + 154d$ | $e_{n_w}$ | Legitimate | TN | TN | TN |

in only one scenario: the sudden failure observed at $t_0 + 56d$. In this instance, distinguishing between a natural failure and a cyberattack is not possible. This ambiguity does not reflect a limitation of the UTA itself, but rather, arises from both the TLA and EBA approaches being unable to conclusively assess the suspicious behavior due to the anomaly occurring in the mid useful life (TLA) and the sudden nature of this anomaly (EBA).

Table 12: Performance metrics.

| | Sensitivity (%) | Specificity (%) | Precision (%) | Accuracy (%) |
|---|---|---|---|---|
| TLA threat assessment | 55 | 66 | 50 | 71 |
| EBA threat assessment | 44 | 75 | 50 | 76 |
| Unified threat assessment | 86 | 93 | 86 | 95 |

This analysis confirms that the UTA effectively integrates the strengths of both the TLA and EBA approaches, as demonstrated in Table 12, which summarizes the performance metrics of the three approaches. The TLA approach shows slightly higher sensitivity to cyberattacks than the EBA, while the EBA approach has better specificity in identifying legitimate cases compared to the TLA. Comparing these performances reveals that the UTA outperforms both the TLA and EBA approaches. It achieves high sensitivity (86%), meaning

it successfully detects most cyberattacks, and a high specificity (93%), indicating that it rarely incorrectly classifies legitimate behaviors as cyberattacks (as in $t_0 + 56d$). These results lead to high precision (86%) and an overall accuracy of 95%.

It is worth noting that other cyberattack scenarios may occur at different periods of the turnout lifecycle. The Table 13 lists some examples of other possible cyberattacks. The TLA approach often overlooks cyberattacks reflecting aging or abnormal behaviors in late useful life, while the EBA approach helps contextualize these behaviors (e.g. $A_9$ and $A_{10}$). Similarly, cyberattacks reflecting progressive pre-fault degradations may be missed by the TLA approach but are effectively detected by the EBA approach (e.g. the $A_{11}$). Conversely, cyberattacks reflecting sudden failures consistently generate ambiguity (e.g. $A_8$ and $A_{12}$), and the proposed framework cannot provide clear threat estimations in such cases, except in early useful life when the TLA approach may help in provide clear threat estimations (e.g. the $A_1$).

Table 13: Other scenarios threat assessment results.

| Label | $t$ | $e$ | $f$ | $\omega_{TLA}$ | $\omega_{EBA}$ | Attack | | | Description |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | TLA | EBA | UTA | |
| $A_8$ | $t_0 + 4d$ | $e_{a_s}$ | $e_{n_e}$ | 0.5 | 0.5 | 0.5 | 0.5 | 0.5 | Both approaches are inconclusive, the TLA in the burn-in period and the EBA when observing sudden failures |
| $A_9$ | $\boldsymbol{t_0 + 102d}$ | $e_{n_w}$ | $e_{n_e}$ | 0.4 | 0.6 | 0.1 | **0.9** | **0.67** | No suspicion is considered by the TLA due to the observation of an aging behavior in late stages of the useful life, while the EBA detects an attack since the observed curve reflects wear of the turnout that appear without prior aging indications |
| $A_{10}$ | $\boldsymbol{t_0 + 105d}$ | $e_{a_p}$ | $e_{n_e}$ | 0.4 | 0.6 | 0.1 | **0.9** | **0.67** | No suspicion is considered by the TLA due to the observation of an abnormal behavior in late stages of the useful life, while the EBA detects an attack since the observed curve is related to a progressive pre-faults degradation that appear without prior notice |
| $A_{11}$ | $\boldsymbol{t_0 + 132d}$ | $e_{a_p}$ | $e_{n_w}$ | 0.1 | 0.9 | 0.1 | **0.9** | **0.78** | No suspicion is considered by the TLA due to the observation of an abnormal behavior in the wear-out period, while the EBA detects an attack since the observed curve is related to a progressive pre-faults degradation that appear without prior notice |
| $A_{12}$ | $t_0 + 135d$ | $e_{a_s}$ | $e_{n_w}$ | 1 | 0 | 0.1 | 0.5 | 0.1 | No suspicion is considered by the TLA due to the observation of an abnormal behavior in the wear-out period, while the EBA remains inconclusive when observing sudden failures |

The proposed framework has proven effective in detecting stealthy cyberattacks designed to mislead maintenance operators. By analyzing the observed curves with respect to their types, collection times, and corresponding expected curves, weights are assigned to each threat estimation provided by the TLA and EBA approaches, significantly improving the accuracy of threat estimation for each observation. The validity of the proposed detection framework does not strictly depend on the use of the specific LSTM method, nor on the assumption of a three-phase bathtub shape. The framework remains applicable with other

forecasting methods and alternative failure rate representations regardless of their specific shapes, as long as the failure rate evolves over time. Thus, the use of the LSTM model and the bathtub curve serves primarily as a heuristic guide rather than a prescriptive statistical model. However, there are several areas that could be addressed to further increase the cyberattack detection rate:

- Predictive modeling. The EBA threat assessment approach detects cyberattacks by analyzing the temporal evolution of turnout behavior using the forecasting LSTM model. However, the LSTM has the drawback of basing its prediction only on past behavior, which may not always be sufficient for accurate prediction (prediction in Fig. 16(a)). In some cases, context from both the past and future is required to improve prediction accuracy. To address this limitation, future work could explore using a Bidirectional LSTM (Bi-LSTM) model, which incorporates both past and future behavior into its analysis. In situations requiring real-time decisions, the EBA could initially rely on the LSTM model for quick analysis, with the Bi-LSTM model used as an additional step to verify the LSTM prediction once the entire sequence is available. However, in cases where real-time decisions are not necessary, the Bi-LSTM model could be fully integrated into the predictive modeling process, allowing for more comprehensive contextualization by considering both past and future behaviors to enhance the accuracy of the predicted curve;

- Sudden failures' analysis. When observing sudden failures, it is crucial to provide accurate threat assessments, particularly during the period $]0.3\beta+0.7\alpha, 0.7\beta+0.3\alpha[$. For instance, this challenge contributed to missing the $A_4$ cyberattack due to the ambiguous nature of the corresponding curve. To counter this difficulty, the overall cyberattack estimation could be improved by introducing an additional indicator specifically related to sudden failures. This indicator could be derived from maintenance operators' expertise and predictive maintenance techniques, allowing for better assessment and contextualization of observed sudden failure curves;

- Turnout lifecycle. The TLA threat assessment approach relies on analyzing the turnout *bathtub* curve, which in this article is assumed to be known. However, to apply this approach in real-world scenarios, it would be necessary to use approaches that estimate the start of each period in the *bathtub* curve. This process should be carried out in collaboration with railway experts to ensure accurate lifecycle assessments and improve the overall effectiveness of the threat detection;

- Progressive intrusions. The detection of progressive intrusions that evolve over several days or weeks, gradually modifying parameters to avoid detection, needs further exploration. Approaches based on traceability and the analysis of behavioral footprints should be developed to detect subtle, long-term changes, such as unusual equipment usage patterns. These changes could indicate the presence of a cyberattack that might otherwise remain undetected in the short term;

- The evidence theory. The D-S theory of evidence choice was made due to its ability to combine information from different sources while managing uncertainty. In this article, the uncertainty of each approach was addressed independently of the D-S theory,

by assigning a trust level of 0 to approaches yielding inconclusive results (during the burn-in period for the TLA approach or when observing sudden failure for the EBA approach) and a trust level of 1 to the other approach, except in the only situation when both approaches are inconclusive (observing sudden failure during the burn-in period) where a trust level of 0.5 is associated to both approaches. Future research should seek to directly incorporate this uncertainty into the D-S framework by considering the uncertainty focal element $\{S_1, S_2\}$. This enhancement would provide a more explicit definition of uncertainty and enable the association of this uncertainty with an estimation, particularly in situations where the two approaches combined do not help to make a clear decision. Additionally, it is important to note that when considering only two sources, the resulting combined belief, $\bar{m}$, can effectively be reduced to a straightforward voting mechanism based on threat estimations and trust levels. This simplification arises because the support degrees of the two sources are equal when the distance between them remains constant. However, the full potential of the D-S theory becomes apparent when information from three or more sources is combined, as the varying distances between sources allow for more nuanced weighting of evidence. This article examined and used the D-S theory not only to integrate information from the two approaches, but also to lay groundwork for incorporating additional sources or methods that may be developed in future research;

- Validation on testbed: Due to the absence of real-world turnout datasets with documented cyberattacks, the proposed framework was validated using simulated attack scenarios. It is important to validate the proposed framework with a greater variety of realistic and diverse cyberattacks to enable a more comprehensive and representative evaluation and enhance its robustness. Thus, future work will involve applying the proposed detection framework on a testbed currently under development by the industrial partner. This platform will support controlled injection of realistic cyberattack scenarios in order to facilitate the assessment of the framework's robustness and effectiveness under conditions that more closely resemble real-world operational environments;

- Generalization of the cyberattack detection framework. The proposed framework is designed to detect cyberattacks by analyzing the behavior of a specific turnout, using the power switch operations curves. These curves are influenced by factors like location, temperature and weather conditions [98, 99], resulting in unique behavioral patterns across different turnouts. Thus, generalizing this cyberattack detection framework to a heterogeneous collection of turnouts is both crucial and challenging, as the learned patterns and their associated behavioral classifications may not apply uniformly across different contexts. Future research should aim to identify key discriminant parameters between different turnouts and investigate turnout behaviors in relation to these factors, allowing for broader applicability of the detection framework.

## 5. Conclusion

This article proposes a novel framework for detecting cyberattacks targeting railway systems, particularly focusing on stealthy data tampering attacks that compromise turnout monitoring either by hiding turnouts failures or triggering unnecessary maintenance actions.

By combining two complementary approaches, Turnout Lifecycle Analysis (TLA) and Expected Behavior Analysis (EBA), and integrating their outputs using the weighted modified Dempster-Shafer theory, the proposed Unified Threat Assessment (UTA) demonstrates improved detection accuracy. The UTA provides a robust mechanism for handling limitations in individual detection approaches, such as indecision during early lifecycle phases or challenges in identifying sudden failures. The methodology was validated on simulated cyberattack scenarios, successfully identifying six out of seven injected attacks while reducing false positives, thus proving its effectiveness.

For future work, the proposed framework offers a strong foundation for further enhancements in railway cybersecurity. Future research could focus on extending the framework by incorporating additional data sources, refining prediction models, and addressing limitations related to decision-making during ambiguous scenarios. Additionally, ethical and societal implications of deploying cybersecurity measures in railway infrastructure require further investigation through interdisciplinary studies. In conclusion, the proposed framework can provide maintenance operators with more reliable insights, enabling timely and accurate decision-making to ensure operational safety and resilience against cyberthreats.

## Acknowledgement

This work was supported by the French government and BPI France under the RAILMON project with the funding of the Future Investment Program (Programme d'investissements d'avenir). The authors acknowledge the partner, Vossloh, for providing the field data and their domain expertise.

## References

[1] S. Bakhtiari, M. R. Najafi, K. Goda, H. Peerhossaini, A dynamic Bayesian network approach to characterize multi-hazard risks and resilience in interconnected critical infrastructures, Reliability Engineering & System Safety 257 (2025) 110815, publisher: Elsevier BV. doi:10.1016/j.ress.2025.110815.

[2] E. Zio, Challenges in the vulnerability and risk analysis of critical infrastructures, Reliability Engineering & System Safety 152 (2016) 137–150. doi:10.1016/j.ress.2016.02.009.

[3] ENISA, ENISA threat landscape 2023: July 2022 to June 2023., Tech. rep., European Network and Information Security AgencyURL https://data.europa.eu/doi/10.2824/782573 (2023) [Accessed 2025-02-06].

[4] R. Kour, A. Patwardhan, R. Karim, P. Dersin, J. Kumari, A cybersecurity approach for improved system resilience, Proceedings of the 32nd European Safety and Reliability Conference ESREL (2022). doi:10.3850/978-981-18-5183-4_S13-03-586-cd.

[5] J. Leyden, Polish teen derails tram after hacking train network, URL https://www.theregister.com/2008/01/11/tram_hack/ (2008) [Accessed 2025-01-15].

[6] N. Tabak, Ransomware attack hits short line rail operator OmniTRAX, URL https://www.freightwaves.com/news/ransomware-attack-hits-short-line-rail-operator-omnitrax (Jan. 2021) [Accessed 2025-01-15].

[7] Reuters, Danish train standstill on Saturday caused by cyber attack, URL https://www.reuters.com/technology/danish-train-standstill-saturday-caused-by-cyber-attack-2022-11-03/ (Nov. 2022) [Accessed 2025-01-15].

[8] A. Roth, 'Cyberpartisans' hack Belarusian railway to disrupt Russian buildup, URL https://www.theguardian.com/world/2022/jan/25/cyberpartisans-hack-belarusian-railway-to-disrupt-russian-buildup (Jan. 2022) [Accessed 2025-01-15].

[9] ENISA, Railway cybersecurity: good practices in cyber risk management, Tech. rep., European Union Agency for Cybersecurity.URL https://data.europa.eu/doi/10.2824/92259 (2021) [Accessed 2025-01-15].

[10] F. Sicard, C. Escudero, E. Zamaï, J.-M. Flaus, From ICS attacks' analysis to the S.A.F.E. approach: implementation of filters based on behavioral models and critical state distance for ICS cybersecurity, 2018 2nd Cyber Security in Networking Conference (CSNet) (2018) 1–8.

[11] A. Thaduri, M. Aljumaili, R. Kour, R. Karim, Cybersecurity for eMaintenance in railway infrastructure: risks and consequences, International Journal of System Assurance Engineering and Management 10 (2) (2019) 149–159. doi:10.1007/s13198-019-00778-w.

[12] J. Styczynski, N. Beach-Westmoreland, When the lights went out, URL https://www.boozallen.com/content/dam/boozallen/documents/2016/09/ukraine-report-when-the-lights-went-out.pdf (2019) [Accessed 2025-01-15].

[13] DarkReading, Pro-Iranian Attackers Claim to Target Israeli Railroad Network, URL https://www.darkreading.com/ics-ot-security/pro-iranian-attackers-target-israeli-railroad-network (Sep. 2023) [Accessed 2025-01-15].

[14] R. Langner, Stuxnet: dissecting a cyberwarfare weapon, IEEE Security & Privacy Magazine 9 (3) (2011) 49–51. doi:10.1109/MSP.2011.67.

[15] R. Arunthavanathan, F. Khan, Z. Sajid, M. T. Amin, K. R. Kota, S. Kumar, Are the processing facilities safe and secured against cyber threats?, Reliability Engineering & System Safety 260 (2025) 111011. doi:10.1016/j.ress.2025.111011.

[16] J. Liu, K. Chen, H. Duan, C. Li, A knowledge graph-based hazard prediction approach for preventing railway operational accidents, Reliability Engineering & System Safety 247 (2024) 110126, publisher: Elsevier BV. doi:10.1016/j.ress.2024.110126.

[17] Y. Liu, K. Li, D. Yan, Quantification analysis of potential risk in railway accidents: A new random walk based approach, Reliability Engineering & System Safety 242 (2024) 109778, publisher: Elsevier BV. doi:10.1016/j.ress.2023.109778.

[18] C. Bian, S. Yang, T. Huang, Q. Xu, J. Liu, E. Zio, Degradation state mining and identification for railway point machines, Reliability Engineering & System Safety 188 (2019) 432–443. doi:10.1016/j.ress.2019.03.044.

[19] X. Tang, X. Cai, Y. Wang, P. Wang, F. Yang, Advanced VTSDREF for vehicle-turnout system dynamic reliability analysis: Integration of hybrid deep learning and adaptive probability density evolution method, Reliability Engineering & System Safety 256 (2025) 110762, publisher: Elsevier BV. doi:10.1016/j.ress.2024.110762.

[20] M. Rodríguez-Hernández, A. Crespo-Márquez, A. Sánchez-Herguedas, V. González-Prida, Digitalization as an Enabler in Railway Maintenance: A Review from "The International Union of Railways Asset Management Framework" Perspective, Infrastructures 10 (4) (2025) 96 [Accessed 2025-09-04]. doi:10.3390/infrastructures10040096. URL https://www.mdpi.com/2412-3811/10/4/96

[21] M.-H. Monzer, K. Beydoun, A. Ghaith, J.-M. Flaus, Model-based IDS design for ICSs, Reliability Engineering & System Safety 225 (2022) 108571. doi:10.1016/j.ress.2022.108571.

[22] A. Beaudet, C. Escudero, E. Zamaï, Malicious Anomaly Detection Approaches Robustness in Manufacturing ICSs, IFAC-PapersOnLine 54 (1) (2021) 146–151. doi:10.1016/j.ifacol.2021.08.016.

[23] M. Iaiani, A. Tugnoli, S. Bonvicini, V. Cozzani, Analysis of Cybersecurity-related Incidents in the Process Industry, Reliability Engineering & System Safety 209 (2021) 107485. doi:10.1016/j.ress.2021.107485.

[24] M. Kaouk, J.-M. Flaus, M.-L. Potet, R. Groz, A Review of Intrusion Detection Systems for Industrial Control Systems, in: 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE, Paris, France, 2019, pp. 1699–1704. doi:10.1109/CoDIT.2019.8820602.

[25] L. Desgeorges, J.-P. Georges, T. Divoux, Detection of cyber-attacks in network control planes using Hidden Markov Model, IFAC-PapersOnLine 55 (28) (2022) 66–72. doi:10.1016/j.ifacol.2022.10.325.

[26] M. Iaiani, G. Fazari, A. Tugnoli, V. Cozzani, Identification of reference security scenarios from past event datasets by Bayesian Network analysis, Reliability Engineering & System Safety 254 (2025) 110615. doi:10.1016/j.ress.2024.110615.

[27] Q. Liu, V. Hagenmeyer, H. B. Keller, A Review of Rule Learning-Based Intrusion Detection Systems and Their Prospects in Smart Grids, IEEE Access 9 (2021) 57542–57564. doi:10.1109/ACCESS.2021.3071263.

[28] H. Lin, A. Slagell, C. Di Martino, Z. Kalbarczyk, R. K. Iyer, Adapting Bro into SCADA: building a specification-based intrusion detection system for the DNP3 protocol, in: Proceedings of the Eighth Annual Cyber Security and Information Intelligence Research Workshop, ACM, Oak Ridge Tennessee USA, 2013, pp. 1–4. doi:10.1145/2459976.2459982.

[29] F. T. Liu, K. M. Ting, Z.-H. Zhou, Isolation-Based Anomaly Detection, ACM Transactions on Knowledge Discovery from Data 6 (1) (2012) 1–39. doi:10.1145/2133360.2133363.

[30] M. Elnour, N. Meskin, K. Khan, R. Jain, A Dual-Isolation-Forests-Based Attack Detection Framework for Industrial Control Systems, IEEE Access 8 (2020) 36639–36651, conference Name: IEEE Access. doi:10.1109/ACCESS.2020.2975066.

[31] M. Choubisa, R. Doshi, N. Khatri, K. Kant Hiran, A Simple and Robust Approach of Random Forest for Intrusion Detection System in Cyber Security, in: 2022 International Conference on IoT and Blockchain Technology (ICIBT), IEEE, Ranchi, India, 2022, pp. 1–5. doi:10.1109/ICIBT52874.2022.9807766.

[32] S. Mokhtari, A. Abbaspour, K. K. Yen, A. Sargolzaei, A Machine Learning Approach for Anomaly Detection in Industrial Control Systems Based on Measurement Data, Electronics 10 (4) (2021) 407. doi:10.3390/electronics10040407.

[33] H. Zheng, X. Li, F. Li, Unsupervised cyberattack detection in smart grids: A novel approach integrating horizontal federated learning for the control center and substations, Reliability Engineering & System Safety 264 (2025) 111444. doi:10.1016/j.ress.2025.111444.

[34] S. Bhattacharya, N. Saqib, M. Govindarasu, ML-based Anomaly Detection System for IEC 61850 Communication in Substations, in: 2024 IEEE Power & Energy Society General Meeting (PESGM), IEEE, Seattle, WA, USA, 2024, pp. 1–5. doi:10.1109/PESGM51994.2024.10688773.

[35] R. C. Borges Hink, J. M. Beaver, M. A. Buckner, T. Morris, U. Adhikari, S. Pan, Machine learning for power system disturbance and cyber-attack discrimination, in: 2014 7th International Symposium on Resilient Control Systems (ISRCS), IEEE, Denver, CO, USA, 2014, pp. 1–8. doi:10.1109/ISRCS.2014.6900095.

[36] D. Ulybyshev, I. Yilmaz, B. Northern, V. Kholodilo, M. Rogers, Trustworthy Data Analysis and Sensor Data Protection in Cyber-Physical Systems, in: Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems, ACM, Virtual Event USA, 2021, pp. 13–22. doi:10.1145/3445969.3450432.

[37] D. Tang, Y.-P. Fang, E. Zio, Vulnerability analysis of demand-response with renewable energy integration in smart grids to cyber attacks and online detection methods, Reliability Engineering & System Safety 235 (2023) 109212. doi:10.1016/j.ress.2023.109212.

[38] S. Marrone, R. J. Rodríguez, R. Nardone, F. Flammini, V. Vittorini, On synergies of cyber and physical security modelling in vulnerability assessment of railway systems, Computers & Electrical Engineering 47 (2015) 275–285. doi:10.1016/j.compeleceng.2015.07.011.

[39] B. Chen, C. Schmittner, Z. Ma, W. G. Temple, X. Dong, D. L. Jones, W. H. Sanders, Security Analysis of Urban Railway Systems: The Need for a Cyber-Physical Perspective, in: Computer Safety, Reliability, and Security, Vol. 9338, Springer International Publishing, Cham, 2015, pp. 277–290, series Title: Lecture Notes in Computer Science. doi:10.1007/978-3-319-24249-1_24.

[40] M. Rekik, C. Gransart, M. Berbineau, Cyber-Physical Security Risk Assessment for Train Control and Monitoring Systems, in: 2018 IEEE Conference on Communications and Network Security (CNS), IEEE, Beijing, 2018, pp. 1–9. doi:10.1109/CNS.2018.8433201.

[41] L. Marassi, S. Marrone, What Would Happen if Hackers Attacked the Railways? Consideration of the Need for Ethical Codes in the Railway Transport Systems, in: A. Esposito, M. Faundez-Zanuy, F. C. Morabito, E. Pasero (Eds.), Applications of Artificial Intelligence and Neural Systems to Data Science, Vol. 360, Springer Nature Singapore, Singapore, 2023, pp. 289–296. doi:10.1007/978-981-99-3592-5_27.

[42] S. Soderi, D. Masti, Y. Z. Lun, Railway Cyber-Security in the Era of Interconnected Systems: A Survey, IEEE Transactions on Intelligent Transportation Systems 24 (7) (2023) 6764–6779 [Accessed 2023-10-18]. doi:10.1109/TITS.2023.3254442. URL https://ieeexplore.ieee.org/document/10075050/

[43] B. Yu, Y. Eun, Sensor attack detection for railway vehicles using topographic information, in: 2017 17th International Conference on Control, Automation and Systems (ICCAS), IEEE, Jeju, 2017, pp. 149–154. doi:10.23919/ICCAS.2017.8204433.

[44] S. Lakshminarayana, T. Z. Teng, R. Tan, D. K. Y. Yau, Modeling and Detecting False Data Injection Attacks against Railway Traction Power Systems, ACM Transactions on Cyber-Physical Systems 2 (4) (2018) 1–29. doi:10.1145/3226030.

[45] R. Kour, A. Thaduri, R. Karim, Railway defender kill chain to predict and detect cyber-attacks, Journal of Cyber Security and Mobility 9 (1) (2020) 47–90. doi:10.13052/JCSM2245-1439.912.

[46] M. Lockheed, Cyber Kill Chain, URL https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html (2024) [Accessed 2025-01-15].

[47] R. Kour, A. Patwardhan, A. Thaduri, R. Karim, A review on cybersecurity in railways, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit (2022) 095440972210893 [Accessed 2022-12-12]doi:10.1177/09544097221089389. URL http://journals.sagepub.com/doi/10.1177/09544097221089389

[48] C. Escudero, P. Massioni, E. Zamaï, B. Raison, Analysis, prevention, and feasibility assessment of stealthy ageing attacks on dynamical systems, IET Control Theory & Applications 16 (4) (2022) 381–397. doi:10.1049/cth2.12178.

[49] S. Abdellaoui, E. Dumitrescu, C. Escudero, E. Zamaï, Cyber threat assessment in monitoring turnout railway systems, in: 2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA), Sinaia, Romania, 2023, pp. 1–8. doi:10.1109/ETFA54631.2023.10275401.

[50] S. Abdellaoui, E. Dumitrescu, C. Escudero, E. Zamaï, Temporal Assessment of Malicious Behaviors: Application to Turnout Field Data Monitoring, in: 2024 International Conference on Control, Automation and Diagnosis (ICCAD), IEEE, Paris, France, 2024, pp. 1–6. doi:10.1109/ICCAD60883.2024.10553981.

[51] A. P. Dempster, Upper and Lower Probabilities Induced by a Multivalued Mapping, The Annals of Mathematical Statistics 38 (2) (1967) 325–339. doi:10.1214/aoms/1177698950.

[52] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976. doi:10.1515/9780691214696.

[53] E. Uflaz, S. I. Sezer, A. L. Tunçel, M. Aydin, E. Akyuz, O. Arslan, Quantifying potential cyber-attack risks in maritime transportation under Dempster–Shafer theory FMECA and rule-based Bayesian network modelling, Reliability Engineering & System Safety 243 (2024) 109825. doi:10.1016/j.ress.2023.109825.

[54] Y. Zhang, Y. Cheng, T. Xu, G. Wang, C. Chen, T. Yang, Fault Prediction of Railway Turnout Systems Based on Improved Sparse Auto Encoder and Gated Recurrent Unit Network, IEEE Transactions on Intelligent Transportation Systems 23 (8) (2022) 12711–12723. doi:10.1109/TITS.2021.3116966.

[55] F. P. García Márquez, F. Schmid, A digital filter-based approach to the remote condition monitoring of railway turnouts, Reliability Engineering & System Safety 92 (6) (2007) 830–840. doi:10.1016/j.ress.2006.02.011.

[56] O. Fink, E. Zio, U. Weidmann, Predicting component reliability and level of degradation with complex-valued neural networks, Reliability Engineering & System Safety 121 (2014) 198–206. doi:10.1016/j.ress.2013.08.004.

[57] F. Zhou, L. Xia, W. Dong, X. Sun, X. Yan, Q. Zhao, Fault diagnosis of high-speed railway turnout based on support vector machine, in: 2016 IEEE International Conference on Industrial Technology (ICIT), IEEE, Taipei, Taiwan, 2016, pp. 1539–1544. doi:10.1109/ICIT.2016.7474989.

[58] D. Ou, R. Xue, K. Cui, A Data-Driven Fault Diagnosis Method for Railway Turnouts, Transportation Research Record: Journal of the Transportation Research Board 2673 (4) (2019) 448–457. doi:10.1177/0361198119837222.

[59] F. P. G. Márquez, J. M. C. Muñoz, A pattern recognition and data analysis method for maintenance management, International Journal of Systems Science 43 (6) (2012) 1014–1028. doi:10.1080/00207720903045809.

[60] S. Aghabozorgi, A. Seyed Shirkhorshidi, T. Ying Wah, Time-series clustering – A decade review, Information Systems 53 (2015) 16–38. doi:10.1016/j.is.2015.04.007.

[61] S. Chu, E. Keogh, D. Hart, M. Pazzani, Iterative Deepening Dynamic Time Warping for Time Series, in: Proceedings of the 2002 SIAM International Conference on Data Mining, Society for Industrial and Applied Mathematics, 2002, pp. 195–212. doi:10.1137/1.9781611972726.12.

[62] J. Macqueen, Some methods for classification and analysis of multivariate observations, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1 (14) (1967) 281–297 [Accessed 2025-01-15]. URL https://api.semanticscholar.org/CorpusID:6278891

[63] F. Murtagh, P. Contreras, Algorithms for hierarchical clustering: an overview, WIREs Data Mining and Knowledge Discovery 2 (1) (2012) 86–97. doi:10.1002/widm.53.

[64] T. M. Kodinariya, D. P. R. Makwana, Review on determining number of Cluster in K-Means Clustering, International Journal 1 (6) (2013) 90–95.

[65] Z. Guo, Y. Wan, H. Ye, An Unsupervised Fault-Detection Method for Railway Turnouts, IEEE Transactions on Instrumentation and Measurement 69 (11) (2020) 8881–8901 [Accessed 2022-12-12]. doi:10.1109/TIM.2020.2998863. URL https://ieeexplore.ieee.org/document/9104763/

[66] D. Ou, Y. Ji, L. Zhang, H. Liu, An Online Classification Method for Fault Diagnosis of Railway Turnouts, Sensors 20 (16) (2020) 4627 [Accessed 2022-11-24]. doi:10.3390/s20164627. URL https://www.mdpi.com/1424-8220/20/16/4627

[67] J. Kisilowski, R. Kowalik, Railroad Turnout Wear Diagnostics, Sensors 21 (20) (2021) 6697 [Accessed 2023-02-06]. doi:10.3390/s21206697. URL https://www.mdpi.com/1424-8220/21/20/6697

[68] F. P. García Márquez, D. J. Pedregal Tercero, F. Schmid, Unobserved Component models applied to the assessment of wear in railway points: A case study, European Journal of Operational Research 176 (3) (2007) 1703–1712 [Accessed 2023-10-18]. doi:10.1016/j.ejor.2005.10.037. URL https://linkinghub.elsevier.com/retrieve/pii/S0377221705008726

[69] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1) (1967) 21–27. doi:10.1109/TIT.1967.1053964.

[70] A. Bagnall, J. Lines, An Experimental Evaluation of Nearest Neighbour Time Series Classification, (Jun. 2014). doi:arXiv:1406.4757[cs].

[71] Y. Shi, L. Zhang, Modelling long- and short-term multi-dimensional patterns in predictive maintenance with accumulative attention, Reliability Engineering & System Safety 237 (2023) 109306, publisher: Elsevier BV. doi:10.1016/j.ress.2023.109306.

[72] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780. doi:10.1162/neco.1997.9.8.1735.

[73] A. Alsaiari, M. Ilyas, Deep Learning for Smart Grid Intrusion Detection: A Hybrid CNN-LSTM-Based Model, International Journal of Artificial Intelligence & Applications 15 (3) (2024) 01–16. doi:10.5121/ijaia.2024.15301.

[74] Z. Han, J. Zhao, H. Leung, K. F. Ma, W. Wang, A Review of Deep Learning Models for Time Series Prediction, IEEE Sensors Journal 21 (2021) 7833–7848. doi:10.1109/JSEN.2019.2923982.

[75] S. Siami-Namini, N. Tavakoli, A. Siami Namin, A Comparison of ARIMA and LSTM in Forecasting Time Series, in: 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, Orlando, FL, 2018, pp. 1394–1401. doi:10.1109/ICMLA.2018.00227.

[76] S. Thapa, Z. Zhao, B. Li, L. Lu, D. Fu, X. Shi, B. Tang, H. Qi, Snowmelt-Driven Streamflow Prediction Using Machine Learning Techniques (LSTM, NARX, GPR, and SVR), Water 12 (6) (2020) 1734. doi:10.3390/w12061734.

[77] Z. Su, Z. Hua, Y. Tang, L. Wang, Q. Zhu, A method for reliability analysis of railway signal equipment at the station level based on universal generating function, Reliability Engineering & System Safety 261 (2025) 111168, publisher: Elsevier BV. doi:10.1016/j.ress.2025.111168.

[78] S. D. Bemment, R. M. Goodall, R. Dixon, C. P. Ward, Improving the reliability and availability of railway track switching by analysing historical failure data and introducing functionally redundant subsystems, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 232 (5) (2018) 1407–1424, publisher: SAGE Publications. doi:10.1177/0954409717727879.

[79] D. Rama, J. D. Andrews, A reliability analysis of railway switches, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit 227 (4) (2013) 344–363, publisher: SAGE Publications. doi:10.1177/0954409713481725.

[80] M. Rausand, A. Høyland, System reliability theory: models, statistical methods, and applications, 2nd Edition, Wiley series in probability and statistics Applied probability and statistics, Wiley-Interscience, Hoboken, NJ, 2004. doi:10.1002/9780470316900.ch12.

[81] M. J. Knoester, N. Bešinović, A. P. Afghari, R. M. Goverde, J. Van Egmond, A data-driven approach for quantifying the resilience of railway networks, Transportation Research Part A: Policy and Practice 179 (2024) 103913, publisher: Elsevier BV. doi:10.1016/j.tra.2023.103913.

[82] M. Macchi, M. Garetti, D. Centrone, L. Fumagalli, G. Piero Pavirani, Maintenance management of railway infrastructures based on reliability analysis, Reliability Engineering & System Safety 104 (2012) 71–83, publisher: Elsevier BV. doi:10.1016/j.ress.2012.03.017.

[83] N. F. Rachman, A. Ependi, Ahkwan, M. Z. Arifin, The Reliability Analysis of Railroad Switch Machine, in: Advances in Engineering Research, Atlantis Press, 2024, pp. 468–479. doi:10.2991/978-94-6463-384-9_42.

[84] L. F. P. Pinhão, Failure modes, effects and criticality of switches and crossings - Inoperability risk determination, 2017.

[85] E. Suhir, Analytical bathtub curve with application to electron device reliability, Journal of Materials Science: Materials in Electronics 26 (9) (2015) 6633–6638. doi:10.1007/s10854-015-3263-1.

[86] H. Li, L. Qi, M. Wang, J. Liu, Operational resilience modeling of cross-border freight railway systems: A study of strategies to improve proactive and reactive capabilities, Reliability Engineering & System Safety 257 (2025) 110856, publisher: Elsevier BV. doi:10.1016/j.ress.2025.110856.

[87] N. Wang, M. Wu, K. F. Yuen, Dynamic enterprise resilience assessment for port systems: A framework integrating Bayesian networks and Dempster-Shafer evidence theory, Reliability Engineering & System Safety 262 (2025) 111105, publisher: Elsevier BV. doi:10.1016/j.ress.2025.111105.

[88] J. Liu, X. Yang, Y. Yang, W. Wang, Z. Chen, F. Ding, H. Zhu, Research on fire risk quantification for extralong highway tunnels based on Wuli–Shili–Renli theory, dempster–shafer theory, and bayesian network, Reliability Engineering & System Safety 264 (2025) 111414, publisher: Elsevier BV. doi:10.1016/j.ress.2025.111414.

[89] E. L. M. González, X. Desforges, B. Archimède, Assessment method of the multicomponent systems future ability to achieve productive tasks from local prognoses, Reliability Engineering & System Safety 180 (2018) 403–415. doi:10.1016/j.ress.2018.08.005.

[90] P. Li, C. Wei, An emergency decision-making method based on D-S evidence theory for probabilistic linguistic term sets, International Journal of Disaster Risk Reduction 37 (2019) 101178. doi:10.1016/j.ijdrr.2019.101178.

[91] L. A. Zadeh, Fuzzy sets and information granularity, Advances in fuzzy set theory and applications (1979) 3–18.

[92] K. Guo, W. Li, Combination rule of D–S evidence theory based on the strategy of cross merging between evidences, Expert Systems with Applications 38 (10) (2011) 13360–13366. doi:10.1016/j.eswa.2011.04.161.

[93] S. I. Sezer, E. Akyuz, O. Arslan, An extended HEART Dempster–Shafer evidence theory approach to assess human reliability for the gas freeing process on chemical tankers, Reliability Engineering & System Safety 220 (2022) 108275. doi:10.1016/j.ress.2021.108275.

[94] L. Liang, Y. Shen, Q. Cai, Y. Gu, A reliability data fusion method based on improved D-S evidence theory, in: 2016 11th International Conference on Reliability, Maintainability

and Safety (ICRMS), IEEE, Hangzhou City, China, 2016, pp. 1–6. doi:10.1109/ICRMS.2016.8050147.

[95] J. Yang, H.-Z. Huang, L.-P. He, S.-P. Zhu, D. Wen, Risk evaluation in failure mode and effects analysis of aircraft turbine rotor blades using Dempster–Shafer evidence theory under uncertainty, Engineering Failure Analysis 18 (8) (2011) 2084–2092. doi:10.1016/j.engfailanal.2011.06.014.

[96] X. Su, Y. Deng, S. Mahadevan, Q. Bao, An improved method for risk evaluation in failure modes and effects analysis of aircraft engine rotor blades, Engineering Failure Analysis 26 (2012) 164–174. doi:10.1016/j.engfailanal.2012.07.009.

[97] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research 12 (85) (2011) 2825–2830 [Accessed 2025-01-15]. URL http://jmlr.org/papers/v12/pedregosa11a.html

[98] Z. Zhang, C.-Y. Lin, Risk analysis of weather-related railroad accidents in the United States, Reliability Engineering & System Safety 255 (2025) 110647, publisher: Elsevier BV. doi:10.1016/j.ress.2024.110647.

[99] Y. Huang, Z. Zhang, H. Hu, Risk propagation mechanisms in railway systems under extreme weather: A knowledge graph-based unsupervised causation chain approach, Reliability Engineering & System Safety 260 (2025) 110976, publisher: Elsevier BV. doi:10.1016/j.ress.2025.110976.