Event-Based Data in Prognostics and Health Management: A Systematic Review of Models, Challenges, and Applications

Thi-Tinh Le[1], Phuc Do[2], Benoit Iung[1] and Marcos Orchard [3]

[1]*Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France*
{thi-tinh.le, benoit.iung}@univ-lorraine.fr

[2]*SyCoIA, IMT Mines Ales, Ales, France*
phuc.do@mines-ales.fr

[3]*Universidad de Chile, Santiago, Chile*
morchard@ing.uchile.cl

ABSTRACT

In the modern industrial context, Prognostics and Health Management (PHM) systems based on data-driven approaches have been widely and effectively developed to reduce maintenance costs. However, continuous data requires large memory capacity and high costs. Therefore, in recent years, the use of event-based data for PHM models has become prominent and increasingly attracts attention due to its cost-efficiency and effectiveness. This surge in data availability has opened new avenues for developing data-driven methods that leverage event patterns to enhance diagnostic, prognostic, and predictive maintenance capabilities. Building meaningful and interpretable patterns from raw event data is crucial for understanding system behavior, detecting faults early, forecasting future failures, and accurately estimating the Remaining Useful Life (RUL) of critical components. This review paper systematically surveys the state-of-the-art methodologies and frameworks for extracting, modeling, and utilizing event-based patterns in the context of diagnostic and prognostic applications. Furthermore, we analyze challenges related to event data heterogeneity, scalability, and interpretability, as well as the need for robust pattern extraction methods that can adapt to dynamic operating environments. The review further explores how these event-based patterns contribute to building reliable diagnostic models, enabling early fault detection, and supporting maintenance decision-making through precise prognostics. Finally, this paper identifies key research gaps and outlines future directions, emphasizing the need for explainable, adaptive, and scalable pattern mining approaches that effectively translate raw event data into actionable maintenance intelligence. To tackle these challenges, we propose a unified pattern construction framework general event pattern for diagnostic, prognostic, and RUL prediction tasks, designed to be adaptable across diverse industrial systems. This comprehensive survey aims to serve as a foundational reference for researchers and practitioners committed to leveraging event data for enhanced system reliability and the development of optimized, intelligent maintenance strategies.

1. INTRODUCTION

Industrial systems are increasingly complex, and their reliable operation is vital to ensure safety, minimize downtime, and reduce maintenance costs. In recent years, Prognostics and Health Management (PHM) has emerged as a transformative approach that leverages data analytics and machine learning to assess system health, early detect faults, and predict the Remaining Useful Life (RUL) of critical components (Lee et al., 2014). PHM plays a central role in enabling predictive maintenance, helping organizations shift from reactive and time-based maintenance to proactive strategies that reduce unexpected failures and optimize maintenance schedules (Jardine, Lin, & Banjevic, 2006).

Traditional PHM systems have predominantly relied on continuous time-series data gathered from sensors monitoring physical parameters such as temperature, pressure, vibration, and flow rate (Si, Wang, Hu, & Zhou, 2011). While effective, such data acquisition entails significant costs in terms of data storage, processing power, and sensor infrastructure, especially when scaled to large industrial facilities. Furthermore, continuous monitoring generates massive volumes of data, much of which may be redundant or uninformative for early fault detection (Nor, Pedapati, Muhammad, & Leiva, 2021).

To address these challenges, recent research has turned toward event-based data, which captures system behavior through discrete events such as alarms, status changes, operational triggers, offer valuable insights for understanding equipment behavior and predicting future failures without the overhead of handling full continuous streams. These sparse and meaningful representations not only reduce data volume but also offer cost-effective and scalable alternatives for PHM modeling (Saxena, Celaya, Saha, Saha, & Goebel, 2010). Event-based PHM has shown promise in enhancing both diagnostic and prognostic tasks by focusing on patterns in event sequences rather than raw time-series streams (Fronza, Sillitti, Succi, & Vlasenko, 2011). However, despite their potential, challenges remain in effectively extracting, modeling, and applying event patterns across diverse industrial systems. Current literature lacks a comprehensive review addressing these issues, including pattern construction, heterogeneity, and scalability.

Despite the growing interest, research on event-based PHM remains scattered, and few review studies comprehensively assess its methodologies, challenges, and applications. Existing reviews on PHM primarily emphasize condition monitoring, fault diagnosis, and deep learning models based on continuous signals (Fink et al., 2020; Zhang et al., 2019; Zhao et al., 2019). For instance, (Jia, Huang, Feng, Cai, & Lee, 2018) surveyed data-driven PHM using deep learning but did not focus on event-level representations. Similarly, major reviews by (Tsui, Chen, Zhou, Hai, & Wang, 2015) and (Atamuradov, Medjaher, Dersin, Lamoureux, & Zerhouni, 2017) concentrate on condition monitoring, fault diagnosis, and RUL estimation from sensor-based continuous signals without addressing the potential of event-driven techniques. This lack of focused analysis motivates the need for a systematic review consolidating developments in event-based pattern mining and modeling for PHM.

This paper presents a comprehensive systematic review of event-based PHM methods, aiming to (i)-highlight the evolution of diagnostic and prognostic approaches using event sequences, (ii)-evaluate the scalability, interpretability, and robustness of existing models, (iii)-identify open challenges in real-world deployments, and (iv) explore the design of a generalizable event-pattern framework applicable to multiple PHM tasks and diverse industrial contexts. The review also considers hybrid frameworks that combine expert knowledge models with data-driven insights, and emphasizes the need for explainable and adaptive algorithms capable of operating in dynamic environments.

maintenance actions, and threshold violations. Event-based data, which reflect transitions in system states or specific.

Current literature lacks a comprehensive review addressing these issues, including pattern construction, heterogeneity, and scalability. This review aims to fill this gap by systematically analyzing the latest developments in event-based PHM methodologies. We focus on three key research questions: RQ1: What are the main event data types and pattern extraction techniques used in PHM? RQ2: How do event patterns contribute to PHM applications? RQ3: What are the existing challenges and solutions for applying event-based methods in industrial PHM? RQ4: How can a single, flexible framework for pattern construction be designed to support multiple PHM tasks and adapt to diverse industrial systems?

The remainder of this paper is organized as follows: Section 2 describes the review methodology, including search strategy, inclusion criteria, and research questions Section 3 summarizes key methods in event-based PHM, highlighting diverse pattern mining techniques and their applications in diagnostics, prognostics, and monitoring. It emphasizes the strengths and limitations of models and provides a framework to address the remaining limitations. Section 4 concludes the paper with final observations and recommendations.

2. METHODOLOGY OF SYSTEMATIC REVIEW

2.1. Literature Search Strategy

This review employs a systematic literature review methodology (Tranfield, Denyer, & Smart, 2003) to ensure a rigorous and unbiased synthesis of research on event-based PHM approaches. The key steps of this methodology are as follows: (i)-defining research scope and objectives which were established to guide the review with a focus on event data types, pattern extraction techniques, and the challenges of applying event-based PHM; (ii)-conducting a database search; (iii)-screening and selecting relevant studies; and (iv) extracting and synthesizing the data.

To conduct a comprehensive and systematic literature review, an initial combination of keywords was constructed using logical operators (AND, OR) to accurately capture relevant studies on predictive maintenance within event-driven data applications in PHM. Following extensive screening, the finalized keyword set (Figure 1) guided concurrent searches on Web of Science, Springer Link and IEEE xplore databases to ensure completeness and accuracy.

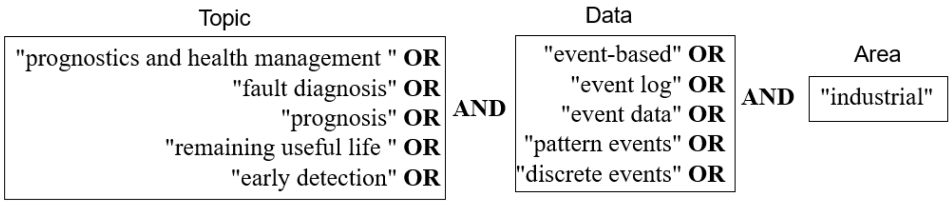

Figure 1 Keyword search strategy for literature selection for Event-Based PHM

2.2. Study Selection and Screening Process

The study selection process for this systematic review was carefully structured to ensure that only relevant and high-quality studies on event-based PHM were included. An initial search across scientific databases yielded 414 records. The screening process began with the title screening of 414 records, which led to the exclusion of 250 articles due to irrelevance or duplication.

Subsequent stages involved abstract screening, where 164 abstracts were reviewed in detail. This step resulted in the exclusion of 112 records for failing to meet the inclusion criteria, which specified that papers must focus on diagnostic or prognostic models using event-based data or methodologies that extract or model event patterns for health assessment or maintenance planning. Studies that did not contribute methodology (e.g., editorial notes, opinion pieces) or were biomedical PHM or focused on power transmission diagnostics were excluded. The eligibility phase consisted of full-text assessment for 52 publications, with 10 excluded for reasons such as inappropriate research questions (n=12) or unavailability of the full text (n=4).

Ultimately, 36 publications were included in the qualitative synthesis, including 5 identified through an expert network. Therefore, 41 publications related to discrete event data were selected for study. This comprehensive and transparent process is illustrated in the accompanying PRISMA flowchart (Figure 2), which clearly outlines the stages of the study selection process.

2.3. Keyword Co-occurrence Analysis

Figure 3 illustrates the annual distribution of publications on event-based Prognostics and Health Management (PHM) from 2013 to 2025. The data show a relatively low number of publications in the early years, followed by a significant increase starting around 2018, indicating growing research interest in event-driven PHM methods. Before 2013, event-based PHM approaches were either absent or rarely documented in academic literature, with most research focusing on continuous sensor data or physics-based modeling. Most of the publications from 2013 onward are sourced from major academic platforms such as Web of Science, IEEE Xplore, and SpringerLink, with a variety of other publishers contributing smaller shares. This reflects a broad and multidisciplinary dissemination of research in the event-based PHM domain.

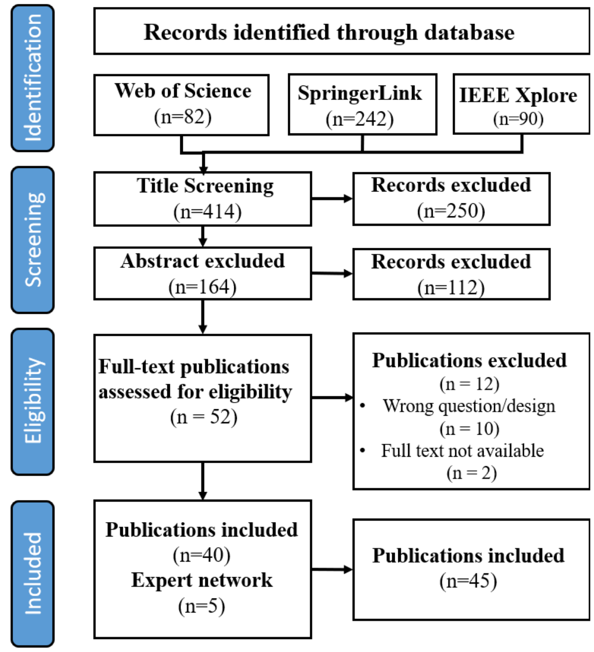

Figure 2. PRISMA flow diagram for literature selection process.

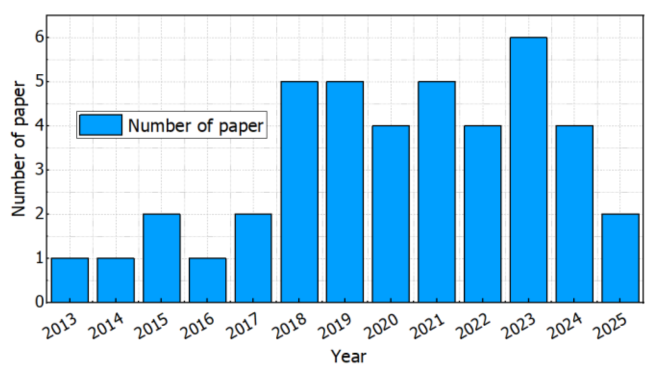

Figure 3. Number of Event-Based PHM papers published by year

The keyword co-occurrence analysis was conducted using VOSviewer software (Figure 4), and the resulting network graph is presented in the figure. Central to the network is "discrete-event systems," which is strongly linked with "diagnosability," "failure diagnosis," and "verification," emphasizing the importance of fault detection, system monitoring, and diagnostic verification in complex event-driven environments. This cluster highlights efforts to ensure that faults are not only identified but also verified in a timely and reliable manner.

Another significant cluster connects “fault diagnosis” with “design,” “optimization,” and “prognostics,” indicating a focus on developing optimized diagnostic strategies and predictive models to anticipate failures. These associations underscore the growing emphasis on early detection of anomalies and degradations, which is crucial for minimizing downtime and preventing cascading failures.Modeling terms like “petri nets” indicate the use of formal models in diagnostic processes.

Overall, this analysis reveals a close interplay between system modeling, fault diagnosis, and verification. It also highlights the increasing role of predictive and early fault detection mechanisms in enhancing system reliability, enabling proactive maintenance, and supporting decision-making in discrete-event systems.

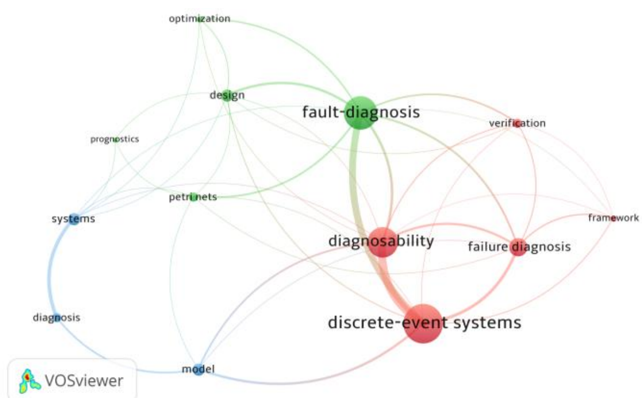

Figure 4. Keyword Co-Occurrence network for Event-Based PHM literature.

3. RESULTS DISCUSSION

3.1. Event Data and Pattern Mining Approaches in PHM

In recent years, the use of event-based data has gained increasing attention in the PHM domain due to its high interpretability, lower data acquisition costs, and ability to directly represent system transitions. Event data typically consist of discrete occurrences such as alarm activations, operational mode changes, maintenance logs, and threshold violations that mark observable changes in system behavior(Ariamuthu Venkidasalapathy & Kravaris, 2021; Benatia, Louis, & Baudry, 2020; Bezerra et al., 2019; Del Moral, Nowaczyk, & Pashami, 2022; Guillaume, Vrain, & Wael, 2020; Gutschi, Furian, Suschnigg, Neubacher, & Voessner, 2019; Petsinis, Naskos, & Gounaris, 2021). Unlike continuous sensor streams, event logs provide a sparse yet semantically meaningful abstraction of system dynamics, which can reduce data redundancy and facilitate more interpretable pattern discovery for degradation analysis (He et al., 2021; Liu et al., 2018). Event data are collected from multiple sources, including programmable logic controllers (PLCs), maintenance logs, and sensor-driven threshold alerts. Table 1 presents the major types of event data utilized in PHM. Despite their advantages, the heterogeneity of industrial systems poses challenges to the effective use of event data in PHM.

Variability in data formats, semantic interpretations, sensor granularity, and logging policies complicates the direct application of generalized modeling techniques. Consequently, there is a growing need for scalable, robust, and domain-adaptive pattern extraction methods capable of handling irregular, noisy, and non-uniform event sequences (Wuest, Weimer, Irgens, & Thoben, 2016). To address these issues, numerous pattern mining approaches have been developed to uncover diagnostic and prognostic insights from event logs. Initial efforts focused on classical data mining methods such as frequent item set mining and sequential pattern discovery, which aim to identify recurring event combinations or fault-related subsequences (Liu et al., 2018),(İfraz, Ersöz, Aktepe, & Çetinyokuş, 2024), (Kawabata, Matsubara, & Sakurai, 2019; Yan, Cao, Madden, & Rundensteiner, 2018).

Table 1 Classification of Event Data Types and Sources

Event Data	Typical Sources	References
Alarm Events	System alarms, sensor alerts	(Ariamuthu Venkidasalapathy & Kravaris, 2021), (Bezerra et al., 2019), (Benatia et al., 2020)
Status Changes	Operational logs, PLC status	(Gutschi et al., 2019), (Guillaume et al., 2020), (Marin-Castro & Tello-Leal, 2021), (Dakic, Stefanovic, Vuckovic, Zizakov, & Stevanov, 2023)
Maintenance Actions	Maintenance logs, work order databases	(Atamuradov et al., 2017), (Del Moral et al., 2022), (Inyang, Petrunin, & Jennions, 2023), (Rakesh, Shruti, Thippeswamy, Nithya, & Dheeraj, 2024)
Threshold Violations	Sensor alerts, monitoring systems	(Petsinis et al., 2021), (Giannoulidis, Gounaris, Nikolaidis, Naskos, & Caljouw, 2022), (Trilla, Mijatovic, & Vilasis-Cardona, 2023)

These approaches are computationally efficient and easy to interpret but often fail to capture complex temporal dependencies or rare but critical failure patterns. To overcome these limitations, more advanced models such as Hidden Markov Models (HMMs) and Bayesian networks were introduced to model stochastic event transitions while accounting for temporal uncertainty and incomplete data (Cartella, Lemeire, Dimiccoli, & Sahli, 2015). Tree-based classifiers like decision trees and ensemble learning methods have also proven effective in balancing predictive accuracy with model interpretability in real-world industrial settings (Wuest et al., 2016). With the rise of deep learning, recent developments have introduced neural network-based approaches such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs)—that can automatically learn nonlinear temporal features and long-range dependencies from event sequences. These models have achieved impressive predictive performance, especially in complex systems, but often require large volumes of labeled data and lack transparency (Guillaume et al., 2020),(Yuan, Zhou, Sievenpiper, Mannar, & Zheng, 2011), (Trilla et al., 2023), (Zhou et al., 2023). The progression of these techniques highlights an important trade-off between model complexity and interpretability. While classical pattern mining techniques remain valuable for their transparency and simplicity, modern deep learning methods excel in capturing intricate patterns at the cost of explainability and scalability. Table 2 provides a comparative summary of these pattern mining approaches, evaluating their strengths, limitations, scalability, interpretability, robustness, and representative studies.

Table 2 Pattern Extraction Techniques: Strengths and Limitations

Technique	Strengths	Limitations	References
Frequent Pattern Mining	Simple, interpretable patterns	May miss rare but critical event patterns	(He et al., 2021; Liu et al., 2018), (Long, Ho, Cong, Dinh-Duc, & Ngoc, 2024),
Sequential Pattern Mining	Captures event order and timing	High computational complexity for long sequences	(İfraz et al., 2024), (Kawabata et al., 2019; Yan et al., 2018), (Borah & Nath)
Probabilistic Models	Handles uncertainty and noisy data	Requires careful parameter tuning	(Cartella et al., 2015), (Ariamuthu Venkidasalapathy & Kravaris, 2021), (Trilla et al., 2023)
Deep Learning (RNN/CNN)	Learns complex temporal dependencies automatically	Less interpretable, needs large data	(Guillaume et al., 2020), (Yuan et al., 2011), (Trilla et al., 2023), (Zhou et al., 2023),

3.2. Contributions of Event-Based Patterns in PHM

Event-based pattern analysis plays a pivotal role in enhancing the diagnostic and prognostic capabilities of modern PHM systems. By abstracting system behavior into discrete and semantically meaningful events, such as fault activations or operational state changes, these patterns enable more focused and interpretable reasoning than raw sensor data. Rather than processing continuous, noisy signals, PHM models can leverage structured event sequences to detect anomalies, characterize degradation, and estimate component health trajectories (Zhou et al., 2023),(Trilla et al., 2023).

In diagnostic tasks, event patterns support the early identification of incipient faults by capturing structured relationships among abnormal behaviors. Classical methods such as association rule mining and sequential pattern mining are commonly employed to discover frequently co-occurring or temporally dependent events from historical logs (Liu et al., 2018), (İfraz et al., 2024). These approaches are both interpretable and computationally efficient, making them suitable for rule-based fault identification frameworks. In more complex industrial environments, decision trees and ensemble learning methods have demonstrated strong performance in learning discriminative features from event logs while preserving model transparency (Guillaume et al., 2020). For example, recent studies have applied these techniques to monitor rotating machinery, production lines, and energy systems achieving accurate fault classification and facilitating operator understanding without relying on black-box models (Inyang et al., 2023).

For prognostics and Remaining Useful Life (RUL) estimation, event sequences serve as temporal abstractions that reflect the system's health evolution. Unlike continuous signal trends, they can capture latent degradation stages through symbolic transitions. Survival analysis methods, particularly Cox proportional hazards models, have been employed to estimate failure probabilities from timestamped events, improving robustness in the face of censored or incomplete data (Fronza et al., 2011), (Yuan et al., 2011). Some approaches further enhance prediction by incorporating event-sequence features and domain knowledge into hybrid survival frameworks. Probabilistic models such as HMMs have also been applied to model transitions between hidden health states under uncertainty (Ariamuthu Venkidasalapathy & Kravaris, 2021). More recently, deep learning models like RNNs and LSTM networks have shown strong capability in learning long-range dependencies and degradation patterns directly from event logs (Zhou et al., 2023). In practice, these methods are often combined with statistical or physical models to form hybrid frameworks, improving both generalization and interpretability in RUL estimation.

In condition monitoring, event patterns are used to continuously track system status and identify early signs of abnormal behavior. These patterns are often structured into Petri Net models to represent causal relationships and fault propagation paths, enabling transparent monitoring in complex systems (Blancke et al., 2018). Beyond Petri Nets, other models such as finite state machines (FSMs), temporal Bayesian networks are also employed to capture sequential dependencies and system dynamics under uncertainty (Arroyo-Figueroa & Sucar, 2013). In recent studies, graph-based approaches, including dynamic dependency graphs and temporal event networks, have shown promise in modeling interactions across components in distributed systems (Narapureddy). To support real-time and scalable monitoring, especially in high-throughput industrial settings, researchers have proposed streaming pattern mining, incremental learning, and online graph updates, allowing the system to evolve adaptively with new event inputs (Patnaik, Ramakrishnan, Laxman, & Chandramouli, 2012), (Decker, Leite, Giommi, & Bonacorsi, 2020). These techniques reduce the need for full retraining, improve responsiveness, and are increasingly integrated into modern PHM pipelines for condition-aware maintenance decision-making.

Overall, event-based patterns offer a powerful and flexible foundation for PHM systems by enabling modular, interpretable, and scalable analysis across diagnostic, prognostic, and monitoring tasks. Table 3 summarizes typical methods used in each application area, highlighting the diversity and adaptability of event-based approaches for predictive maintenance

Table 3 Typical Event-Based Pattern Mining Methods Used

Application Area	Typical Methods	References
Diagnostics	Association rule mining, sequential pattern mining, decision trees	(Liu et al., 2018), (İfraz et al., 2024), (Guillaume et al., 2020), (Inyang et al., 2023)
Prognostics (RUL)	Survival analysis, HMMs, RNNs (LSTM), hybrid statistical models	(Fronza et al., 2011), (Yuan et al., 2011),(Zhou et al., 2023),(Ureta, 2022)
Condition Monitoring	Petri Nets, adaptive graphs, streaming pattern mining	(Arroyo-Figueroa & Sucar, 2013), (Narapureddy), (Patnaik et al., 2012), (Decker et al., 2020)

3.3. Challenges and Limitations in Event-Based PHM

Despite the growing interest in event-based approaches for PHM, several technical and practical challenges hinder their full deployment in real-world industrial settings. These challenges span from data-related limitations to model scalability, adaptability, and explainability, calling for both methodological advancements and practical innovations.

Industrial event logs often suffer from heterogeneity in structure, semantics, and granularity across vendors, hindering data integration and cross-system generalization (Marin-Castro & Tello-Leal, 2021). Additionally, data quality issues such as noise, missing events, and inconsistencies from sensor faults challenge reliable pattern extraction (Marin-Castro & Tello-Leal, 2021),(Dakic et al., 2023). The lack of publicly available, labeled event datasets also restricts reproducibility and benchmarking, slowing progress in developing and validating robust PHM models (Su & Lee, 2023).

From a modeling perspective, most existing pattern mining techniques emphasize frequent patterns, potentially neglecting rare but critical event sequences that precede major failures (Long et al., 2024),(Shyalika, Wickramarachchi, & Sheth, 2024). These rare events are vital for early fault detection but are difficult to discover without tailored algorithms or domain feedback. In addition, capturing complex temporal dependencies in long and noisy event streams demands sophisticated, efficient pattern mining techniques that can scale computationally (Kawabata et al., 2019; Yan et al., 2018).

Model transparency and explainability also remain key limitations. Complex models, particularly deep learning-based architectures like RNNs or hybrid neural-statistical systems, often operate as black boxes. This lack of interpretability hinders trust and adoption by practitioners (Kundu & Hoque, 2023),(Nor et al., 2021). Incorporating Explainable AI (XAI) tools such as attention visualization or surrogate models can help elucidate model behavior and build user confidence (Rakesh et al., 2024).Despite their potential, event-based PHM systems face several key challenges. Integrating event data with continuous sensor signals remains limited, and hybrid models that fuse symbolic and numerical data are still under development (Tsallis, Papageorgas, Piromalis, & Munteanu, 2025).

Real-time applicability is another concern, as many algorithms struggle with high-frequency, large-scale event streams and lack online inference capabilities (Patnaik et al., 2012),(Mayer, Mayer, & Abdo, 2017). Although streaming pattern mining and online learning show promise (Patnaik et al., 2012) their scalability in dynamic environments requires improvement. Most models lack generalizability and cannot adapt well to new systems or changing conditions. While techniques like adaptive Petri Nets and transfer learning show potential, they remain immature and are not yet widely applicable (Chiachío, Chiachío, Prescott, & Andrews, 2018).

Finally, human-in-the-loop integration is essential to ensure interpretability, trust, and continuous model refinement through expert feedback.

3.4. Toward a Generalizable Pattern Framework

Despite significant advancements in event-based PHM research, current methods continue to face key challenges that limit their practical applicability. As outlined in Section 3.3 and summarized in Table 4, these challenges do not always occur together in every scenario but represent recurring issues commonly encountered when extracting meaningful patterns from industrial event logs. To address these challenges, we propose a generalizable framework for event-pattern construction, with the ultimate goal of supporting downstream applications such as fault detection, diagnostics, and RUL prediction. The patterns extracted serve as compact, interpretable, and transferable representations of system behavior, enabling effective use of event data for predictive early faults. It also serves as a foundation for predictive maintenance and decision-making models.

The full process is illustrated in Figure 5, which provides an end-to-end overview of the pipeline: from raw symbolic logs to multi-source fusion, pattern discovery, and explainable outputs. The architecture is modular and adaptable, allowing it to scale across systems and data conditions.

Table 4 Typical Event-Based Pattern Mining Methods Used

Gap	Description	References
Gap 1	Data is not standardized, making generalization difficult	(Marin-Castro & Tello-Leal, 2021)
Gap 2	Lack of labels and context, event meaning is unclear	(Dakic et al., 2023)
Gap 3	Pattern mining often overlooks important events	(Su & Lee, 2023), (Borah & Nath)
Gap 4	Simple patterns fail to capture relational or temporal dependencies	(Kawabata et al., 2019), (Borah & Nath)
Gap 5	Patterns are difficult to apply across different systems	(Kawabata et al., 2019; Yan et al., 2018)
Gap 6	Difficult to handle real-time or streaming data	(Patnaik et al., 2012),(Mayer et al., 2017)
Gap 7	Patterns are hard to interpret, impractical to use, and lack explain ability	(Kundu & Hoque, 2023),(Nor et al., 2021)
Gap 8	Lack of integrated multi-source modeling: combining continuous and discrete data is difficult	(Tsallis et al., 2025)

The process begins with raw symbolic event logs, which often vary widely in format, terminology, and semantic granularity. To address this heterogeneity (Gap 1), event tokens are first mapped to a unified ontology or taxonomy. This standardization is guided by expert knowledge, which helps clarify the meaning of domain-specific terms and align them across different datasets or systems. By combining automated parsing with expert-driven mapping, the framework ensures semantic consistency and facilitates downstream processing. Subsequently, semantic embedding techniques such as Word2Vec or Event2Vec etc are applied to encode symbolic sequences into continuous vector representations. These sequences are segmented into time windows or sessions, preserving temporal locality and system context.

Since many datasets lack annotated fault data (Gap 2), the framework adopts unsupervised learning methods—such as clustering (e.g., KMeans, DBSCAN), LSTM autoencoders, or isolation forests—to identify structural anomalies in event sequences. These anomalies are assigned pseudo-labels, enabling the creation of surrogate target variables for downstream supervised tasks. To further mitigate data imbalance caused by rare faults (Gap 4), data augmentation techniques, including SMOTE and Generative Adversarial Networks (GANs), are employed. However, when datasets are already balanced, this step can be omitted.

In the next stage, the framework performs pattern mining to identify high-utility sub-patterns within the labeled sequences. Methods such as n-gram mining, PrefixSpan, and SPADE, along with attention mechanisms in Transformer or LSTM architectures, are utilized to extract patterns that reveal meaningful temporal dependencies. This process helps capture long-range interactions and complex transitions (Gap 3), which are often overlooked in simple rule-based systems.

To improve cross-system generalization (Gap 5), learned embeddings and sub-patterns are used in transfer learning schemes. These embeddings serve as shared representations that can be fine-tuned or reused across different machines, product lines, or operational environments. The framework thus supports flexible adaptation with minimal retraining effort.

To handle real-time or streaming scenarios (Gap 6), the framework incorporates online anomaly detection through a sliding-window approach. Incoming event logs are continuously segmented into short time-based windows, which are encoded using pre-trained embeddings enriched with temporal features (e.g., delta time). These segments are then passed through lightweight models such as LSTM Autoencoders or online classifiers that estimate reconstruction errors on the fly. If the anomaly score exceeds a threshold, a warning is triggered in real time. Additionally, online learning tools (e.g., Incremental Isolation Forest) are integrated to support adaptive behavior and enable fast response to system drift without full retraining. This allows the framework to operate effectively in dynamic industrial environments where timely fault detection is critical.

To address the limited semantic interpretability of discovered event patterns (Gap 7), the framework incorporates a dedicated explanation layer that combines Explainable AI techniques with expert knowledge. This module goes beyond structural analysis and focuses on interpreting the meaning of each pattern specifically, whether it represents a normal behavior, a sign of system degradation, or an actual fault. While domain experts contribute contextual knowledge to label and validate these behaviors. This human-in-the-loop process enables the translation of abstract sequences into semantically meaningful insights, enhancing trust, facilitating corrective actions, and aligning pattern-based predictions with real-world operational understanding.

To address the lack of integration between existing event logs and continuous sensor signals in industrial systems (Gap 8), the framework introduces a multi-modal fusion module.

Although both data sources are often available, they are typically analyzed separately, limiting their combined potential for predictive maintenance. By synchronizing event sequences with sensor measurements via timestamps, the framework creates a unified representation that captures both discrete system transitions and continuous physical dynamics. This integration enhances model performance by providing richer context for state estimation, anomaly detection, and fault diagnosis.

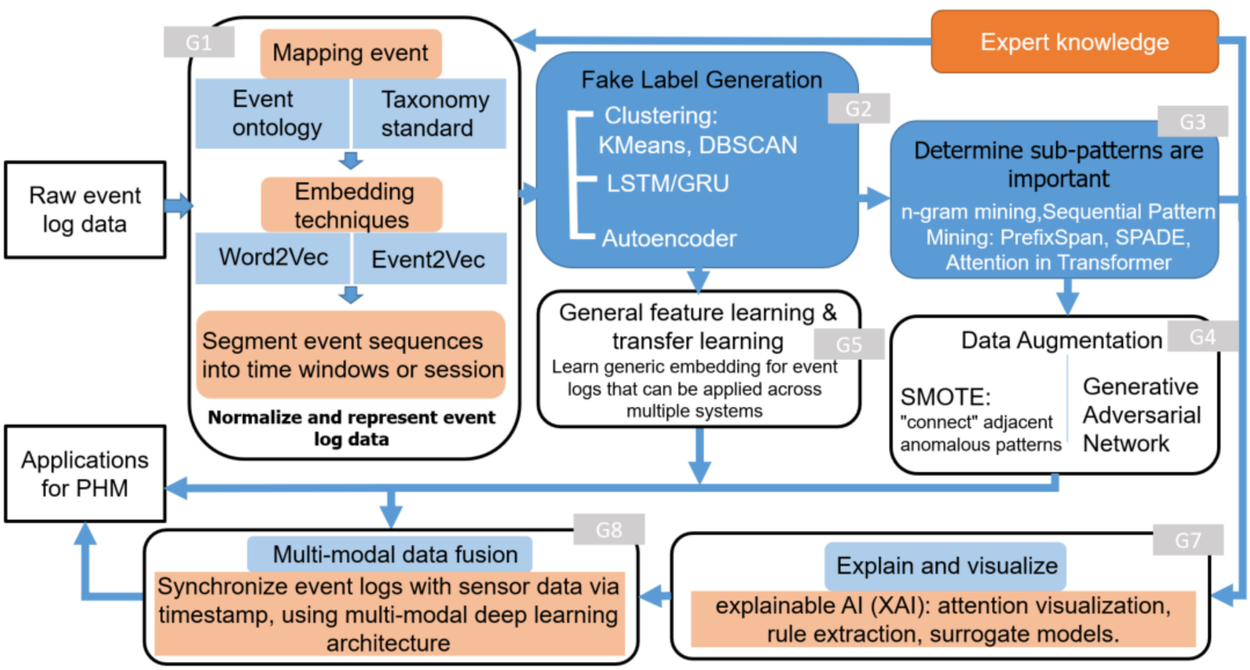

Figure 5 Proposing a generalized Event Pattern building framework for PHM

4. CONCLUSION

This paper presented a systematic review of event-based approaches in Prognostics and Health Management (PHM), emphasizing their potential to improve diagnostics, prognostics, and condition monitoring through interpretable and cost-efficient event patterns. We categorized event data types, compared various pattern extraction techniques, and analyzed their application domains. Despite notable progress, challenges remain regarding data heterogeneity, limited labels, pattern generalizability, and real-time implementation in dynamic environments. To bridge these gaps, we proposed a generalizable event-pattern framework that integrates unsupervised learning, anomaly detection, temporal modeling, expert feedback, and multi-source data fusion. This framework supports scalable PHM solutions and encourages reuse across diverse industrial systems. Future work should focus on benchmarking with real-world datasets, validating hybrid approaches, and developing practical tools that enable seamless integration into existing maintenance workflows. The proposed framework lays the foundation for a more intelligent, transparent, and sustainable maintenance ecosystem.

ACKNOWLEDGEMENT

The acknowledgement section is optional. Please list any acknowledgment here using a single paragraph.

NOMENCLATURE

PHM	Prognostics and Health Management
RUL	Remaining Useful Life
XAI	Explainable AI
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
RNN	Recurrent Neural Network
GAN	Generative Adversarial Network
AE	Autoencoder
HMM	Hidden Markov Models

REFERENCES

Ariamuthu Venkidasalapathy, J., & Kravaris, C. (2021). Hidden Markov model based approach for diagnosing cause of alarm signals. *AIChE Journal, 67*(10), e17297.

Arroyo-Figueroa, G., & Sucar, L. E. (2013). A temporal Bayesian network for diagnosis and prediction. *arXiv preprint arXiv:1301.6675*.

Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., & Zerhouni, N. (2017). Prognostics and health management for maintenance practitioners-Review, implementation and tools evaluation. *International Journal of Prognostics and health management, 8*(3), 1-31.

Benatia, M. A., Louis, A., & Baudry, D. (2020). Alarm correlation to improve industrial fault management. *IFAC-PapersOnLine, 53*(2), 10485-10492.

Bezerra, A., Silva, I., Guedes, L. A., Silva, D., Leitão, G., & Saito, K. (2019). Extracting value from industrial alarms and events: A data-driven approach based on exploratory data analysis. *Sensors, 19*(12), 2772.

Blancke, O., Combette, A., Amyot, N., Komljenovic, D., Lévesque, M., Hudon, C., & Zerhouni, N. (2018). *A predictive maintenance approach for complex equipment based on petri net failure mechanism propagation model.* Paper presented at the Proc Eur Conf PHM Soc.

Borah, A., & Nath, B. Rare pattern mining: challenges and future perspectives. Complex Intell. Syst. 5 (1), 1–23 (2018). In.

Cartella, F., Lemeire, J., Dimiccoli, L., & Sahli, H. (2015). Hidden Semi‐Markov Models for Predictive Maintenance. *Mathematical Problems in Engineering, 2015*(1), 278120.

Chiachío, J., Chiachío, M., Prescott, D., & Andrews, J. (2018). *Modelling adaptive systems using plausible Petri nets.* Paper presented at the 8th International Workshop on Reliable Computing:"Computing with Confidence".

Dakic, D., Stefanovic, D., Vuckovic, T., Zizakov, M., & Stevanov, B. (2023). Event Log Data Quality Issues and Solutions. *Mathematics, 11*(13), 2858.

Decker, L., Leite, D., Giommi, L., & Bonacorsi, D. (2020). *Real-time anomaly detection in data centers for log-based predictive maintenance using an evolving fuzzy-rule-based approach.* Paper presented at the 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE).

Del Moral, P., Nowaczyk, S., & Pashami, S. (2022). *Filtering Misleading Repair Log Labels to Improve Predictive Maintenance Models.* Paper presented at the PHM Society European Conference.

Fink, O., Wang, Q., Svensen, M., Dersin, P., Lee, W.-J., & Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence, 92*, 103678.

Fronza, I., Sillitti, A., Succi, G., & Vlasenko, J. (2011). *Failure prediction based on log files using the cox proportional hazard model.* Paper presented at the The 23rd International Conference on Software Engineering & Knowledge Engineering, SEKE 2011: technical program; July 7-9, 2011, Eden Rock Renaissance Miami Beach, Florida, USA.

Giannoulidis, A., Gounaris, A., Nikolaidis, N., Naskos, A., & Caljouw, D. (2022). Investigating thresholding techniques in a real predictive maintenance scenario. *ACM SIGKDD Explorations Newsletter, 24*(2), 86-95.

Guillaume, A., Vrain, C., & Wael, E. (2020). Predictive maintenance on event logs: Application on an ATM fleet. *arXiv preprint arXiv:2011.10996*.

Gutschi, C., Furian, N., Suschnigg, J., Neubacher, D., & Voessner, S. (2019). Log-based predictive maintenance in discrete parts manufacturing. *Procedia CIRP, 79*, 528-533.

He, K., Liu, B., Xie, M., Do, P., Iung, B., & Kuo, W. (2021). Reliability analysis of systems with discrete event data using association rules. *Quality and Reliability Engineering International, 37*(8), 3693-3712.

İfraz, M., Ersöz, S., Aktepe, A., & Çetinyokuş, T. (2024). Sequential predictive maintenance and spare parts management with data mining methods: a case study in bus fleet. *The Journal of Supercomputing, 80*(15), 22099-22123.

Inyang, U. I., Petrunin, I., & Jennions, I. (2023). Diagnosis of multiple faults in rotating machinery using ensemble learning. *Sensors, 23*(2), 1005.

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing, 20*(7), 1483-1510.

Jia, X., Huang, B., Feng, J., Cai, H., & Lee, J. (2018). *A review of PHM Data Competitions from 2008 to*

2017: Methodologies and Analytics. Paper presented at the Proceedings of the Annual Conference of the Prognostics and Health Management Society.

Kawabata, K., Matsubara, Y., & Sakurai, Y. (2019). *Automatic sequential pattern mining in data streams.* Paper presented at the Proceedings of the 28th ACM International Conference on Information and Knowledge Management.

Kundu, R. K., & Hoque, K. A. (2023). *Explainable predictive maintenance is not enough: quantifying trust in remaining useful life estimation.* Paper presented at the Annual conference of the phm society.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., & Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications. *Mechanical systems and signal processing, 42*(1-2), 314-334.

Liu, B., Do, P., Iung, B., Xie, M., Peysson, F., & Jha, M. (2018). *A study on the use of discrete event data for prognostics and health management: discovery of association rules.* Paper presented at the PHM Society European Conference.

Long, V. H., Ho, N., Cong, T. L., Dinh-Duc, A.-V., & Ngoc, T. N. (2024). Efficient Rare Temporal Pattern Mining in Time Series. *arXiv preprint arXiv:2409.05042*.

Marin-Castro, H. M., & Tello-Leal, E. (2021). Event log preprocessing for process mining: a review. *Applied Sciences, 11*(22), 10556.

Mayer, C., Mayer, R., & Abdo, M. (2017). *Streamlearner: Distributed incremental machine learning on event streams: Grand challenge.* Paper presented at the Proceedings of the 11th ACM International Conference on Distributed and Event-based Systems.

Narapureddy, A. R. AI-DRIVEN DYNAMIC DEPENDENCY GRAPH GENERATION FOR PREDICTIVE OBSERVABILITY IN DISTRIBUTED SYSTEMS.

Nor, A. K. M., Pedapati, S. R., Muhammad, M., & Leiva, V. (2021). Overview of explainable artificial intelligence for prognostic and health management of industrial assets based on preferred reporting items for systematic reviews and meta-analyses. *Sensors, 21*(23), 8020.

Patnaik, D., Ramakrishnan, N., Laxman, S., & Chandramouli, B. (2012). Streaming algorithms for pattern discovery over dynamically changing event sequences. *arXiv preprint arXiv:1205.4477*.

Petsinis, P., Naskos, A., & Gounaris, A. (2021). Analysis of key flavors of event-driven predictive maintenance using logs of phenomena described by Weibull distributions. *arXiv preprint arXiv:2101.07033*.

Rakesh, P., Shruti, J., Thippeswamy, I., Nithya, B., & Dheeraj, V. (2024). *A Surrogate Approach to Explainable AI for Predictive Maintenance: Techniques and Applications.* Paper presented at the 2024 5th International Conference on Circuits, Control, Communication and Computing (I4C).

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and health management, 1*(1), 4-23.

Shyalika, C., Wickramarachchi, R., & Sheth, A. P. (2024). A comprehensive survey on rare event prediction. *ACM Computing Surveys, 57*(3), 1-39.

Si, X.-S., Wang, W., Hu, C.-H., & Zhou, D.-H. (2011). Remaining useful life estimation–a review on the statistical data driven approaches. *European journal of operational research, 213*(1), 1-14.

Su, H., & Lee, J. (2023). Machine learning approaches for diagnostics and prognostics of industrial systems using open source data from PHM data challenges: a review. *arXiv preprint arXiv:2312.16810*.

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence ‐ informed management knowledge by means of systematic review. *British journal of management, 14*(3), 207-222.

Trilla, A., Mijatovic, N., & Vilasis-Cardona, X. (2023). Unsupervised probabilistic anomaly detection over nominal subsystem events through a hierarchical variational autoencoder. *International Journal of Prognostics and health management, 14*(1).

Tsallis, C., Papageorgas, P., Piromalis, D., & Munteanu, R. A. (2025). Application-wise review of Machine Learning-based predictive maintenance: Trends, challenges, and future directions. *Applied Sciences, 15*(9), 4898.

Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., & Wang, W. (2015). Prognostics and health management: A review on data driven approaches. *Mathematical Problems in Engineering, 2015*(1), 793161.

Ureta, D. (2022). *Gentle Correctness Verification of the Theory of Uncertain Event Prognosis to Compute Failure Time Probability.* Paper presented at the Annual Conference of the PHM Society.

Wuest, T., Weimer, D., Irgens, C., & Thoben, K.-D. (2016). Machine learning in manufacturing: advantages, challenges, and applications. *Production & Manufacturing Research, 4*(1), 23-45.

Yan, Y., Cao, L., Madden, S., & Rundensteiner, E. A. (2018). Swift: Mining representative patterns from large event streams.

Yuan, Y., Zhou, S., Sievenpiper, C., Mannar, K., & Zheng, Y. (2011). Event log modeling and analysis for system failure prediction. *IIE Transactions, 43*(9), 647-660.

Zhang, L., Lin, J., Liu, B., Zhang, Z., Yan, X., & Wei, M. (2019). A review on deep learning applications in prognostics and health management. *Ieee Access, 7*, 162415-162438.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical systems and signal processing, 115*, 213-237.

Zhou, Y., Wang, Y., Song, H., Meharizghi, T., Al Jazaery, M., Xu, P., . . . Abeyakoon, A. (2023). Deep sequence modeling for event log-based predictive maintenance.