Towards Explainable and Reusable Temporal Case-Based Reasoning in Predictive Maintenance: A Research Environment*

Alexander Schultheis[1,2], Maximilian Späth[1,2], Justin Weich[1,2], and Ralph Bergmann[1,2]

[1] Artificial Intelligence and Intelligent Information Systems, Trier University, 54296 Trier, Germany, http://www.wi2.uni-trier.de
{Alexander.Schultheis,s4mnspae,s4juweich,bergmann}@uni-trier.de
[2] German Research Center for Artificial Intelligence (DFKI) Branch Trier University, Behringstraße 21, 54296 Trier, Germany
{Alexander.Schultheis,Maximilian.Späth,Justin.Weich,Ralph.Bergmann}@dfki.de

Abstract *Predictive Maintenance* (PredM) relies on multivariate sensor time series to detect emerging failures and support timely interventions. *Temporal Case-Based Reasoning* (TCBR) is promising in such data-scarce maintenance settings because similar past temporal cases can support retrieval, explanation, and later reuse. Existing work, however, mainly evaluates fault classification and lacks reusable environments for investigating these directions in PredM. This paper presents a case-based research environment for PredM as a foundation for explainable and reusable TCBR. Requirements for similarity assessment are derived, and a systematic literature study is conducted to select a suitable approach and dataset. Based on this, an expert-knowledge-enhanced *Siamese Neural Network* (SiamNN) as a similarity measure and an industry-near Fischertechnik dataset are selected. For this dataset, a suitable vocabulary is investigated, and a case base is provided based on it. This case base and the SiamNN integration are implemented within the Case-Based Reasoning Framework ProCAKE. The validation shows preserved retrieval behavior and substantially accelerated retrieval of new events. This contribution and the corresponding implementation form the research foundation for explanation and case reuse through adaptation.

Keywords: Temporal Case-Based Reasoning · Predictive Maintenance · Internet of Things · Industry 4.0 Data · Time Series Similarity

1 Introduction

The analysis of machine data enables the early identification of emerging failures based on multivariate sensor time series and the estimation of their further course

*The final authenticated publication is available online at https://doi.org/10.1007/978-3-032-45236-8_17

until functional failure occurs. Such analysis is addressed in *Predictive Maintenance* (PredM), which uses condition and operational data to anticipate degradations and failures and to schedule interventions before failures occur [53, 68]. PredM has become a major application area within *Industry 4.0* (I4.0), where *Cyber-Physical Systems* (CPSs), sensors, connectivity, and data-driven services enable machines and processes to be monitored via sensor signals that capture system behavior over time [28,68]. Consequently, PredM tasks are inherently temporal and range from failure diagnosis and classification to prognostic objectives such as estimating future degradation, the *Remaining Useful Lifetime* (RUL), and the further course of time series [39, 53, 55]. In maintenance terminology, the interval between the first detectable indication of degradation (p) and functional failure (f) is commonly referred to as the p-f interval [39].

In this context, case adaptation reuses information from retrieved temporal cases for a query, for example, to predict the further course of failure-relevant sensor streams within the p-f interval. This relates to time series forecasting with established statistical and deep learning methods [23] and to PredM prognostics and RUL estimation [54]. Thus, forecasting and RUL estimation provide the application context for later case reuse through adaptation. Such reuse relies on *Case-Based Reasoning* (CBR), which represents previous situations as cases and addresses new problems by retrieving similar cases and reusing their solutions [6]. Case-based prediction likewise infers outcomes from similarity to previously observed cases [4]. For temporal data, this falls into *Temporal Case-Based Reasoning* (TCBR) [24,37], which is attractive for PredM because industrial datasets often contain many failure-free recordings but few labeled failure cases and are strongly shaped by application-specific contexts [19, 30]. However, reviewed case-based and similarity-based approaches remain largely centered on diagnosis or classification. They typically provide a label and similarity value without making the supporting temporal case information traceable [46] or providing a reusable foundation for forecasting the subsequent development of multivariate sensor time series within the p-f interval.

This paper takes a first step toward addressing this gap by presenting a case-based research environment for PredM. From a Design Science Research perspective [20], the work identifies and prepares existing artifacts from the knowledge base for reuse in subsequent design cycles for explanation and case adaptation. To this end, requirements for similarity assessment in PredM are derived, existing approaches are analyzed through a systematic literature study [44], and a suitable similarity assessment approach together with a corresponding dataset is selected. On this basis, the case-based approach is conceptualized following the knowledge containers [43] and implemented based on the ProCAKE framework [7, 8] as an open-source application for follow-up research.

The remainder of this paper is structured as follows. Sect. 2 introduces the foundations and related work, Sect. 3 derives the requirements, and Sect. 4 presents the literature study and approach selection. Sect. 5 presents the research environment concept, Sect. 6 its implementation and validation, and Sect. 7 the conclusion and future work.

2 Foundations and Related Work

This section introduces the foundations required for the proposed research environment. Sect. 2.1 presents PredM and related CBR work, Sect. 2.2 introduces TCBR, and Sect. 2.3 summarizes *Deep Learning* (DL) and *Siamese Neural Network*s (SiamNNs) for similarity assessment.

2.1 Predictive Maintenance

I4.0 denotes the paradigm of digitally interconnected industrial production, where CPSs, sensors, connectivity, and data-driven services enable flexible and intelligent manufacturing [28,68]. In this context, PredM uses continuously acquired machine and production data for condition monitoring, failure diagnosis, failure prediction, decision support, and maintenance activities [28, 35, 68]. Advances in *Internet of Things* (IoT) and data analytics enable continuous monitoring via interconnected sensors that generate time series data [35]. A common failure-progression model in PredM, the p-f curve, separates the first detectable *potential failure* (p) from *functional failure* (f). The p-f interval spans these points, whereas estimating the remaining time from a current point to f constitutes RUL estimation [35, 39].

These developments are closely linked to CPSs, which integrate physical components with digital sensing, control, and communication capabilities, and to Cyber-Physical Production Systems (CPPSs) as their industrial realization for real-time data collection, processing, and decision-making in production environments [32, 35]. Compared to corrective, time-based, and condition-based maintenance, PredM supports diagnosis and anticipates future degradations, which may include predicting the further development of time series and estimating the RUL [35]. PredM approaches are commonly grouped into model-based, data-driven, and hybrid approaches [18]. Knowledge-based methods are interpretable but rely on expert input, physics-based models are reliable when mechanisms are known but labor-intensive, and data-driven methods can be accurate yet require labeled data and often lack interpretability. Hybrid approaches combine these strengths, where expert knowledge can complement machine learning and knowledge-based reasoning can contribute explainability [18, 27, 30]. Such explainability is particularly relevant in maintenance settings, where diagnostic or prognostic results may influence costly interventions and require traceable justification for domain experts. In addition, PredM distinguishes diagnostics from prognostics, where RUL is estimated, often with uncertainty bounds, based on the interval between detectable conditions and functional failure [18, 35, 55].

Previous work on applying CBR in PredM mainly focuses on fault diagnosis, including gearbox systems [10], aero-engines [13], robotics [40, 41], the Tennessee Eastman Process [64], and bearing diagnosis [34]. However, these works either do not represent cases as time series or only consider features or conditional observations derived from time series. In contrast, this paper focuses on a PredM research environment in which multivariate time series form the basis for CBR. While PredM more broadly includes prognostic tasks such as anticipating future

degradations and estimating the RUL [18, 35, 53, 55], the case-based work focuses on diagnosis and fault classification [10, 13, 34, 40, 41, 64].

2.2 Temporal Case-Based Reasoning

CBR is an experience-based problem-solving paradigm in which new problems are addressed by retrieving and reusing similar previously observed cases [6]. The CBR cycle is commonly structured into the steps retrieve, reuse, revise, and retain, where retrieved solutions are adapted, evaluated, and stored as new experience when appropriate. Since retrieval determines which prior experiences become available for reuse, case representation and similarity assessment are central elements of a CBR system [6]. The representation and processing of time series within this cycle are considered in the area of TCBR [24, 37]. In many *Industrial Internet of Things* (IIoT) settings, temporal case knowledge is captured through multivariate sensor streams. Accordingly, time series constitute the most established form of case representation [38]. A time series is an ordered sequence of measured values associated with specific points in time, for instance originating from manufacturing sensors [38, 49]. While further representations such as episodes or graphs exist, time series allow a direct use of sensor data and can be embedded as local attributes within a higher-level case description [38].

Retrieval in TCBR depends on time series similarity measures, which determine which previously observed temporal cases are relevant for a new problem [38]. Following the local-global principle [6, pp. 106–107], local similarities between time points or local structures are aggregated to global similarities on the time series level and, if multiple series exist, further combined to an overall case similarity. Prominent measures for sequences of differing lengths include *Dynamic Time Warping* (DTW), which additionally compensates for stretching and compression effects [38]. Many widely used measures are knowledge-poor in the sense that they compare syntactic patterns without incorporating domain semantics, whereas knowledge-intensive, semantic measures have been proposed in general but are rarely used in TCBR practice [38, 49].

2.3 Deep Learning and Siamese Neural Networks

For similarity assessment on complex multivariate time series in PredM, promising approaches come from the area of DL. DL is a sub-area of *Machine Learning* (ML) that uses multi-layer *Neural Networks* (NNs) to learn task-relevant representations from data through successive transformations [5, 9]. NNs learn non-linear mappings from input to output by adjusting internal parameters based on training data, typically through gradient-based optimization such as backpropagation [5,9]. They are also used for representation learning by mapping complex objects to embeddings in vector spaces [21]. Similarities between these embeddings can then be computed using measures such as Euclidean distance or cosine similarity. For learning such similarity functions, SiamNNs provide a dedicated architecture in which two networks with identical structure and shared weights process a pair in parallel to produce embeddings whose distance is then used to calculate a

similarity value [21, 66]. Given labeled pairs, SiamNNs learn discriminative representations by minimizing distances for similar pairs and maximizing distances for dissimilar pairs [22, 30]. This makes Siamese architectures particularly suitable for similarity assessment on complex inputs, including multivariate time series data [22].

3 Requirements

The case-based research environment requires a suitable similarity assessment approach and corresponding dataset. The following requirements specify the properties needed to support subsequent research on explanation and case reuse through adaptation. They are derived from the PredM domain, the foundations in Sect. 2, and the goal of reproducible follow-up research:

Req. 1 Applicable Similarity Measure: To identify the most similar cases in PredM, the approach must be able to calculate a similarity value between an incoming event serving as the query, represented by measured sensor time series, and a stored failure-free case or failure case. This requirement follows directly from the scope of the research environment.

Req. 2 Similarity Based on Sensor Streams: Since PredM in I4.0 relies on continuously acquired multivariate sensor streams, the approach must support events represented by varying-length multivariate time series from multiple sensors and machines [28, 35, 68].

Req. 3 Domain-Agnostic Applicability: To support PredM across heterogeneous I4.0 environments, the approach must not be tailored to a single machine type or narrowly defined factory setup, but remain applicable in principle across different factories, machines, operating regimes, and sensor configurations [28, 68].

Req. 4 Suitable for Little Failure Data: In PredM deployments, datasets are typically imbalanced, with abundant failure-free data but only limited failure data available [19, 28, 47]. Accordingly, the approach must remain effective even for only a few labeled failure examples.

Req. 5 Expert Knowledge Integration: The approach must allow modeled domain expert knowledge about the factory environment, especially relevant sensors or attributes for specific failure types, to guide similarity assessment in a traceable and domain-informed way and to provide a structured basis for later explanation-oriented analyses.

Req. 6 Available Industry-Near Dataset: A suitable dataset must be openly accessible and industry-near, providing multivariate sensor time series with failure-free and failure-related events that can be prepared as a reusable case base for follow-up research.

Req. 7 Open-Source Availability: The approach must be open-source and sufficiently documented so that it can be integrated into the proposed CBR research environment and extended in subsequent research.

4 Similarity Assessment in Predictive Maintenance

The literature study structures similarity assessment approaches for failure diagnosis and classification in PredM. Sect. 4.1 summarizes the results. Sect. 4.2 evaluates them against the requirements and selects an approach and dataset.

4.1 Systematic Literature Review

The systematic literature study [44], conducted in June 2025, categorizes the identified approaches, following Xing et al. [61], into *distance-based* and *feature-based* similarity assessment. Distance-based approaches compare raw time series using lockstep or elastic measures, whereas feature-based approaches extract discriminative fixed-size feature vectors before computing similarity. Due to their prominence, SiamNN-based approaches are reported separately as *learned feature-based*. The identified approaches are summarized in Table 1.

Table 1. Identified Approaches for Similarity Assessment in PredM, Sorted by Publication Year Within the Categories.

Distance-Based

Reference	Similarity Assessment	Problem Domain	Dataset
Bregón et al. [11]	Euclidean distance, Manhattan, DTW	Laboratory plant	Own simulated
Mai & Chevalier [12]	Euclidean distance, DTW	Nuclear power plant	Own simulated
Sun et al. [58]	Euclidean distance	High-voltage LED lamps	–

Feature-Based

Reference	Similarity Assessment	Problem Domain	Dataset
Deng et al. [15]	Serial Principal Component Analysis	Nonlinear system / chemical process	Tennessee Eastman Process[3]
Ribeiro [42]	Manhattan or Euclidean distance (on extracted features)	Rotating machinery	Machinery Fault Database[4] & Bearing Data Center[5]
Attoui [2, 3]	Statistical similarity measure	Rotating machinery	Bearing Data Center[5] & Unit of Research in Advanced Materials (URMA) Dataset[4]
Aburakhia et al. [1]	Cosine, Euclidean distance, structural similarity measure	Rotating machinery	Bearing Data Center[5]

Learned Feature-Based

Reference	Similarity Assessment	Problem Domain	Dataset
Zhang et al. [63]	Siamese wide first-layer kernel *Convolutional Neural Network* (CNN)	Rotating machinery	Bearing Data Center[5]
Klein et al. [30]	Siamese two-dimensional CNN considering expert restrictions	Cyber-Physical Production System	Fischertechnik Smart Factory Dataset [29]
Yang et al. [62]	Siamese two-dimensional CNN	Rotating machinery	Bearing Data Center[5]
Fang & Wu [17]	Anti-noise Siamese network (ANS-net)	Rotating machinery	Bearing Data Center[5]
Li et al. [33]	Siamese hybrid NN (SHNN)	Fixed-wing UAVs	Real-Flight Fault Detection[6]
Zhao et al. [65]	Improved SiamNN	Rotating machinery	Own simulated
Chen et al. [14]	SiamNN based on long-short-term networks	HVAC systems	AHU[4]
Lee & Jeong [31]	Lightweight wide first-layer kernel CNN	Rotating machinery	Bearing Data Center[5]
Ding et al. [16]	Channel Attention Siamese Network (CASN)	Rotating machinery	Bearing Data Center[5]
Zheng et al. [67]	Siamese one-dimensional CNN	Rotating machinery	Bearing Data Center[5]

Overall, feature-based similarity assessment dominates in PredM, with learned feature-based, predominantly Siamese, approaches forming the largest group. Distance-based approaches apply classical measures such as Euclidean distance and DTW directly to raw time series [11, 12, 58]. Most works evaluate rotating machinery and reuse the Bearing Data Center dataset, while few address broader production environments.

4.2 Requirements Analysis and Selected Approach

Based on the literature study, the identified approaches are evaluated[7] against the requirements derived in Sect. 3 to identify a suitable similarity assessment approach and dataset for the proposed research environment. For each requirement, the analysis distinguishes whether it is already satisfied, is only partially satisfied but can be fulfilled through adaptations of the approach or its accompanying artifacts, or is not supported by the approach design. Since the literature study is scoped to similarity-based failure diagnosis in PredM, all identified approaches satisfy Req. 1 by construction. For the remaining requirements, the distance-based approaches are demonstrated only in narrowly scoped scenarios, providing limited evidence for handling varying-length multivariate sensor streams in complex I4.0 environments (Req. 2) and for domain-agnostic applicability (Req. 3). Moreover, they rely on simulated and partly non-public datasets, which limits their suitability for reproducible follow-up research (Req. 6). Therefore, these approaches are not selected as the basis of the proposed research environment. Feature-based and learned feature-based approaches are dominated by rotating-machinery benchmarks, primarily the *Bearing Data Center*[5]. This supports comparability but concentrates the available empirical evidence on a single domain and largely on one benchmark dataset, providing only limited empirical evidence for domain-agnostic transfer (Req. 3). In addition, most of these approaches do not support the explicit integration of expert knowledge (Req. 5). Taken together, this narrows the selection to approaches that combine evidence from an industry-near PredM setting with support for expert knowledge and few failure examples. Overall, none of the identified approaches fully satisfies all requirements without adaptation. Among the reviewed approaches, the selected approach by Klein et al. [30] provides the best overall fit with the derived requirements.

Therefore, the selected approach together with its dataset is chosen as the most suitable basis for preparing the case base and integrating the similarity measure within the proposed research environment. In general terms, the approach addresses similarity-based failure diagnosis for smart manufacturing data and is

[3] `https://github.com/YKatser/CPDE/tree/master/TEP_data`

[4] Not publicly available.

[5] `https://engineering.case.edu/bearingdatacenter`

[6] `https://github.com/mrtbrnz/fault_detection`

[7] The detailed analysis of all requirements is available at `https://gitlab.rlp.net/procake/publications/procake-predictive-maintenance/-/blob/main/publicationContent/Similarity-Assessment-Requirements-Analysis.pdf`

based on an expert-knowledge-enhanced SiamNN for similarity assessment [30]. Its central idea is to incorporate modeled expert knowledge about which sensor and actuator streams are relevant for specific failure modes, thereby restricting similarity assessment to the most relevant attributes [30]. The approach satisfies Req. 2 by normalizing varying-length time series to 1,000 values. Shorter streams are padded by repeating the last observed value, whereas longer streams are truncated after the first 1,000 values. This preprocessing has limitations, as the SiamNN cannot distinguish padded and recorded values and failure-relevant behavior occurring after the first 1,000 values is not considered. The approach also supports scenarios with only a few failure examples (Req. 4) and enables expert knowledge integration (Req. 5). However, Req. 3 is only partially satisfied because a model instance expects a fixed input dimensionality. Different sensor configurations, therefore, require adaptation of the model input and training for the respective dataset. Thus, further adaptation is required to fully satisfy Req. 3 within the intended research environment. Moreover, it is accompanied by a carefully modeled, industry-near dataset from the IoT Lab Trier[8] recorded on a physical Fischertechnik factory model with multiple production modules and multivariate sensor and actuator streams [27, 28], which provides the basis for the case base in the proposed research environment. Further, the physical factory setup remains available for additional data generation and can also be replicated using an equivalent factory model. However, Req. 6 is only partially satisfied because, although the underlying data artifacts are publicly available, the relevant information is distributed across multiple artifacts and requires preprocessing and consolidation before it can be used as a reusable case base. Likewise, Req. 7 is only partially satisfied because the available implementation is not provided as an encapsulated similarity component and lacks sufficient documentation for straightforward integration into the proposed research environment. The expert-knowledge-enhanced similarity assessment was evaluated in the original work [30] and outperformed baseline similarity measures and Siamese variants without such restrictions.

5 Concept of the Case-Based Research Environment for Predictive Maintenance

The case-based research environment integrates the selected dataset and expert-knowledge-enhanced SiamNN by Klein et al. [27, 28, 30] into the vocabulary, case base, and similarity knowledge containers [43]. The current concept assumes at most one failure mode per event and therefore does not identify simultaneous failures. Following the CBR procedure model [38], the three containers are presented in turn.

[8] https://iot.uni-trier.de/

5.1 Vocabulary

The vocabulary determines how experiences are represented as cases and thus which information can later be stored in the case base, retrieved, examined, and reused [6, 43]. In the PredM setting, an experience corresponds to an event captured by multivariate sensor and actuator streams originating from a smart factory CPPS. Fig. 1 summarizes the vocabulary, which is based on the Fischertechnik data model [29, 30]. The case representation is derived from the selected dataset and centers on multivariate time series, the assigned failure mode label, modeled expert knowledge about relevant attributes, and optional precomputed embeddings. At the case level, the representation captures the temporal scope of an event and metadata for dataset handling and traceability [30]. The potential failure point p and functional failure point f can be represented as timestamps in *MaintenanceSpecificKnowledge*, with their distance defining the p-f interval. Since RUL depends on the considered point in time, it is not represented as a fixed case attribute. At the stream level, sensor and actuator signals are represented with lightweight metadata, and links to ontology concepts[9] preserve the smart-factory context [29, 30]. At the time series level, timestamped numeric or Boolean values represent heterogeneous CPPS monitoring data. For failure events, the vocabulary further includes the failure mode and distinguishes direct and contextual features to support expert-knowledge-based restrictions in similarity assessment.

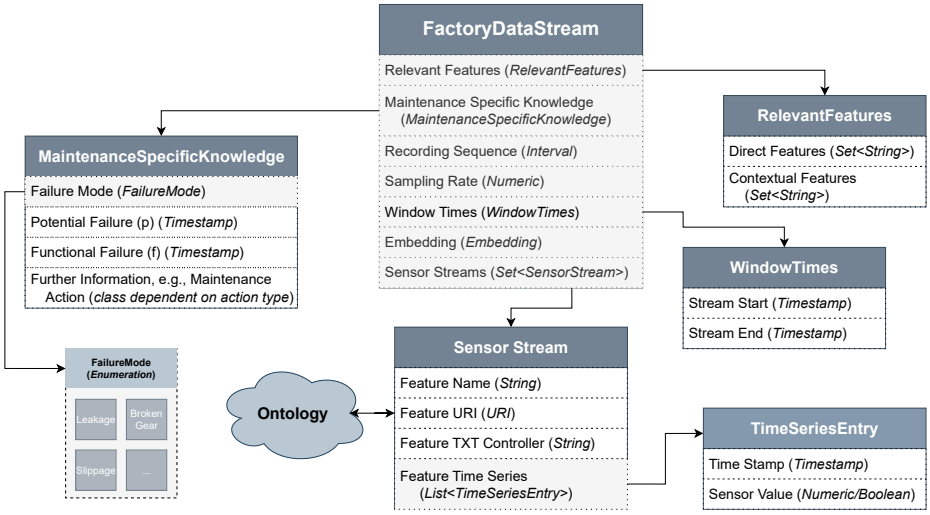


Fig. 1. The Vocabulary for Representing Data Streams from a Smart Factory.

[9] https://gitlab.rlp.net/iot-lab-uni-trier/ftonto

5.2 Case Base

In the case base, all recorded events of the selected dataset are stored as cases following Fig. 1. For the Fischertechnik dataset of Klein et al. [28,30], 28 different failure modes and numerous failure-free events are available, resulting in 27,073 cases. Each case represents multiple sensor and actuator time series, usually comprising around 1,000 sampled values per stream at a sampling interval of 4 ms, together with metadata and ontology links preserving their factory context, which leads to a substantial data volume characteristic of PredM settings. The time series are stored compactly by keeping only timestamped value changes, while unchanged intermediate values can be reconstructed losslessly based on the known recording sequence [38]. This reconstruction is required for similarity assessment and embedding creation by the SiamNN and preserves the basis for follow-up research on explanation and adaptation. Due to the overall size of the case base, especially the large share of failure-free events, the case base creates substantial storage and loading effort, corresponding to the Big Data challenge of volume [25]. Therefore, ingestion into the CBR system is parallelized so that loading and repeated evaluation runs remain feasible.

5.3 Similarity Measures

In the similarity measures container, the selected expert-knowledge-enhanced SiamNN is integrated for retrieval [30]. Similarity is computed between a query event and a stored case based on their multivariate sensor and actuator streams. For each query-case comparison, the expert knowledge regarding failure-relevant attributes is determined by the known failure mode of the stored case. The motivation for using this learned similarity measure is twofold: first, the approach is designed for high-dimensional smart-factory data streams, and second, it incorporates modeled expert knowledge about attribute relevance through direct and contextual features. For integration into the research environment, the original SiamNN is decoupled from its standalone implementation and invoked on cases represented according to Sect. 5.1.

During retrieval, the SiamNN receives the query event and the stored case, transforms their streams and relevant features into the model input, creates encodings and embeddings, and computes the similarity value. For the research environment, the model input can be adapted to a different number of streams. Computed embeddings are stored with the cases once generated and reused in later calculations. Batched retrieval processes multiple comparisons jointly instead of invoking the model for each query-case pair, reducing overhead and improving scalability.

6 Implementation and Validation of the Research Environment

This section presents the implementation of the research environment in Sect. 6.1 and its functional validation and runtime comparison in Sect. 6.2.

6.1 Implementation of the Research Environment

For the implementation, ProCAKE [7, 8] was selected as the underlying CBR framework. This choice is motivated by its support for TCBR [51] and its successful integration of ML approaches in existing prototypes [21, 51, 59]. Based on the defined vocabulary, the selected dataset [27, 28, 30], originally distributed across several files, was collected, prepared, and transformed into ProCAKE-specific XML files. The corresponding embeddings were stored in a database and linked to the case base, while the relevant ontology was integrated as background knowledge. To enhance reusability, the prepared case base has been published openly in ProCAKE format via Zenodo [57]. Furthermore, the selected SiamNN was encapsulated as a Docker component and integrated as the similarity measure. In this way, the implementation encapsulates the selected similarity assessment approach and makes both the data and retrieval functionality reusable in a consolidated form for subsequent research on explanation and adaptation. All components are documented and published for the research community[10].

6.2 Functional Validation and Runtime Comparison

For functional validation and runtime comparison, the implemented research environment was compared with the original implementation of Klein et al. [30]. Functional equivalence was tested by comparing all pairwise similarity computations of the integrated SiamNN against the original implementation, considering values as equal within a minor tolerance to account for floating-point differences (threshold of 10^{-6}). Efficiency was measured once for loading the case base and executing the complete retrieval of 3,389 test queries against 23,303 training cases with a batch size of 128 on a server with 32 CPU cores, 2,940 MHz, and 200 gigabytes of RAM.

Both versions produced identical aggregated confusion matrices, with 2,956 true positives, 104,626 true negatives, 433 false positives, and 433 false negatives, and all pairwise similarity differences remained below the accepted tolerance threshold. Loading the XML-based case base takes longer than loading the original dataset representation of Klein et al. [28, 30] (16 minutes compared to 5.8 seconds), mainly because the XML-based representation is larger than the original NPY and CSV files. However, retrieval is accelerated from 174 hours and 23 minutes to 3 hours by storing and reusing embeddings and by batch-wise SiamNN execution. Total runtime is reduced by approximately a factor of 53. A retrieval query over 23,303 cases takes 3.19 seconds on average. Thus, the retrieval gain comes at the cost of substantially longer initial loading. Following Seiger et al. [52], the similarity value may also serve as a heuristic indication of confidence in the retrieved case. Thus, the environment provides efficient and functionally equivalent retrieval over temporal cases in PredM and a basis for follow-up research on explanation and adaptation.

[10] `https://gitlab.rlp.net/procake/publications/procake-predictive-maintenance`

7 Conclusion and Future Work

This paper presents a case-based research environment for TCBR in PredM that provides a reusable and reproducible foundation for follow-up research on explanation and adaptation. A suitable similarity assessment approach and an industry-near dataset were selected, integrated into the relevant knowledge containers, and implemented as an open-source prototype based on ProCAKE. The evaluation shows that the implementation preserves the retrieval behavior of the selected similarity assessment and improves retrieval efficiency through batching and embedding reuse. Overall, the contribution comprises the development of a suitable case representation and the consolidation of the previously distributed dataset into an openly available case base, together with the encapsulation, integration, and documentation of the SiamNN as a reusable similarity component for TCBR in PredM.

Future work will address adaptation knowledge for queries for which retrieval and classification indicate a failure mode and the corresponding expert knowledge indicates that functional failure f has not yet been reached. For such queries within the p-f interval, adaptation could be used to predict the further course of failure-relevant time series and thereby support RUL estimation. For such multivariate temporal data, hybrid approaches combining CBR with DL methods for time series forecasting appear promising. A neural adaptation component could use the query and similar retrieved cases to predict the subsequent development until functional failure, for example, with Transformer-based models [60] or time series foundation models, and compare the results with non-case-based forecasting baselines. As another topic, an extension to support simultaneous failures should be investigated. Since explicitly representing combinations of rare failure modes would likely provide insufficient case coverage, alternative retrieval strategies that can identify multiple relevant failure modes, for example, based on several retrieved cases and similarity thresholds as detection boundaries, can be investigated. Furthermore, additional datasets from the extended Fischertechnik factory, including further production workflows, will be provided and used to evaluate classification accuracy and transferability across workflows. Future work should investigate alternative preprocessing and temporal abstraction strategies that explicitly distinguish padded values and preserve failure-relevant information occurring after the first 1,000 values. Furthermore, alternative similarity assessment approaches that may more fully satisfy the derived requirements should be investigated and compared with the selected approach of Klein et al. [30].

Another important research direction concerns explicit explanation. The retrieval result already offers an inherent form of explanation by presenting a similar case from the past [56]. However, the similarity value produced by the SiamNN remains opaque and constitutes a challenge in *Explainable Case-Based Reasoning* (XCBR). Possible directions include explainable artificial intelligence methods [36], such as surrogate models, and twin-system approaches [26] in which a transparent TCBR similarity measure approximates the behavior of the SiamNN. Furthermore, visualization of the query and the most similar case [45,50] can help explain the similarity value, also in combination with a twin-

system approach. Once adaptation mechanisms have been developed, explanation approaches can also be investigated for the predicted further course of the time series.

Acknowledgments. This work is funded by the Federal Ministry of Research, Technology and Space under grant No. 01MD22002C *EASY* [48].

References

1. Aburakhia, S.A., Tayeh, T., Myers, R., Shami, A.: Similarity-Based Predictive Maintenance Framework for Rotating Machinery. In: 5th ICCSPA Proceedings. pp. 1–6. IEEE (2022). https://doi.org/10.1109/ICCSPA55860.2022.10019121
2. Attoui, I.: Novel Fast and Automatic Condition Monitoring Strategy Based on Small Amount of Labeled Data. IEEE TSMC **52**(2), 1266–1275 (2022). https://doi.org/10.1109/TSMC.2020.3018102
3. Attoui, I., Boutasseta, N., Fergani, N.: Novel Machinery Monitoring Strategy Based on Time-Frequency Domain Similarity Measurement With Limited Labeled Data. IEEE TIM **70**, 1–8 (2021). https://doi.org/10.1109/TIM.2020.3011874
4. Badra, F., Lesot, M.: Case-based prediction - A survey. Int. J. Approx. Reason. **158**, 108920 (2023). https://doi.org/10.1016/J.IJAR.2023.108920
5. Başţanlar, Y., Özuysal, M.: Introduction to Machine Learning, pp. 105–128. Humana Press, Totowa, NJ (2014). https://doi.org/10.1007/978-1-62703-748-8_7
6. Bergmann, R.: Experience Management: Foundations, Development Methodology, and Internet-Based Applications, LNCS, vol. 2432. Springer (2003). https://doi.org/10.1007/3-540-45759-3
7. Bergmann, R., Grewenig, L., Hoffmann, M., Jung, L., Malburg, L., Schultheis, A., Zeyen, C.: ProCAKE Framework (8.0.0) (2026). https://doi.org/10.5281/zenodo.20524277
8. Bergmann, R., Grumbach, L., Malburg, L., Zeyen, C.: ProCAKE: A Process-Oriented Case-Based Reasoning Framework. In: 27th ICCBR Workshop Proc. vol. 2567, pp. 156–161. CEUR-WS.org (2019)
9. Bishop, C.M.: Pattern recognition and machine learning, 5th Edition. Information science and statistics, Springer (2007)
10. Boral, S., Chaturvedi, S.K., Naikan, V.N.A.: A case-based reasoning system for fault detection and isolation: a case study on complex gearboxes. JQME **25**(2), 213–235 (2019). https://doi.org/10.1108/JQME-05-2018-0039
11. Bregón, A., Hurtado, M.A.S., Rodríguez, J.J., Alonso, C.J., Junquera, B.P., Moro, Q.I.: Early Fault Classification in Dynamic Systems Using Case-Based Reasoning. In: 11th CAEPIA Proc. LNCS, vol. 4177, pp. 211–220. Springer (2005). https://doi.org/10.1007/11881216_23
12. Carole, M., Chevalier, R.: Equipment diagnostics based on comparison of past abnormal behaviors using a big data platform. In: 3rd PHME Proc. vol. 3 (2016). https://doi.org/10.36001/phme.2016.v3i1.1602
13. Chen, M., Qu, R., Fang, W.: Case-based reasoning system for fault diagnosis of aero-engines. Expert Systems with Applications **202**, 117350 (2022). https://doi.org/10.1016/j.eswa.2022.117350
14. Chen, Z., Xiao, F., Guo, F.: Similarity learning-based fault detection and diagnosis in building HVAC systems with limited labeled data. RSER **185**, 113612 (2023). https://doi.org/10.1016/j.rser.2023.113612

15. Deng, X., Tian, X., Chen, S., Harris, C.J.: Nonlinear Process Fault Diagnosis Based on Serial Principal Component Analysis. IEEE TNNLS **29**(3), 560–572 (2018). https://doi.org/10.1109/TNNLS.2016.2635111
16. Ding, P., Xu, Y., Qin, P., Sun, X.M.: A novel deep learning approach for intelligent bearing fault diagnosis under extremely small samples. Applied Intelligence **54**(7), 5306–5316 (2024). https://doi.org/10.1007/s10489-024-05429-7
17. Fang, Q., Wu, D.: ANS-net: anti-noise Siamese network for bearing fault diagnosis with a few data. Nonlinear Dynamics **104**(3), 2497–2514 (2021). https://doi.org/10.1007/s11071-021-06393-4
18. Gupta, S., Kumar, A., Maiti, J.: A critical review on system architecture, techniques, trends and challenges in intelligent predictive maintenance. Safety Science **177**, 106590 (2024). https://doi.org/10.1016/j.ssci.2024.106590
19. Hakami, A.: Strategies for overcoming data scarcity, imbalance, and feature selection challenges in machine learning models for predictive maintenance. Scientific Reports **14**(1), 9645 (2024). https://doi.org/10.1038/s41598-024-59958-9
20. Hevner, A.R.: A Three Cycle View of Design Science Research. SJIS **19**(2), 4 (2007)
21. Hoffmann, M., Bergmann, R.: Using Graph Embedding Techniques in Process-Oriented Case-Based Reasoning. Algorithms **15**(2), 27 (2022). https://doi.org/10.3390/a15020027
22. Hou, L., Jin, X., Zhao, Z.: Time Series Similarity Measure via Siamese Convolutional Neural Network. In: 12th CISP-BMEI Proceedings. pp. 1–6. IEEE (2019). https://doi.org/10.1109/CISP-BMEI48845.2019.8966048
23. Hyndman, R.J., Athanasopoulos, G.: Forecasting: principles and practice. OTexts (2021)
24. Jære, M.D., Aamodt, A., Skalle, P.: Representing Temporal Knowledge for Case-Based Prediction. In: 6th ECCBR Proc. LNCS, vol. 2416, pp. 174–188. Springer (2002). https://doi.org/10.1007/3-540-46119-1_14
25. Kaisler, S.H., Armour, F., Espinosa, J.A., Money, W.H.: Big Data: Issues and Challenges Moving Forward. In: 46th HICSS Proc. pp. 995–1004. IEEE Computer Society (2013). https://doi.org/10.1109/HICSS.2013.645
26. Kenny, E.M., Keane, M.T.: Twin-Systems to Explain Artificial Neural Networks Using Case-Based Reasoning: Comparative Tests of Feature-Weighting Methods in ANN-CBR Twins for XAI. In: 28th IJCAI. pp. 2708–2715 (2019). https://doi.org/10.24963/ijcai.2019/376
27. Klein, P.: Combining Expert Knowledge and Deep Learning with Case-Based Reasoning for Predictive Maintenance. Springer Vieweg, Wiesbaden, 1 edn. (2025). https://doi.org/10.1007/978-3-658-46986-3
28. Klein, P., Bergmann, R.: Generation of Complex Data for AI-based Predictive Maintenance Research with a Physical Factory Model. In: 16th ICINCO Proc. pp. 40–50. SciTePress (2019). https://doi.org/10.5220/0007830700400050
29. Klein, P., Malburg, L., Bergmann, R.: FTOnto: A Domain Ontology for a Fischertechnik Simulation Production Factory by Reusing Existing Ontologies. In: LWDA 2019 Workshop Proc. vol. 2454, pp. 253–264. CEUR-WS.org (2019)
30. Klein, P., Weingarz, N., Bergmann, R.: Enhancing Siamese Neural Networks Through Expert Knowledge for Predictive Maintenance. CCIS, vol. 1325, pp. 77–92. Springer (2020). https://doi.org/10.1007/978-3-030-66770-2_6
31. Lee, D., Jeong, J.: Few-Shot Learning-Based Light-Weight WDCNN Model for Bearing Fault Diagnosis in Siamese Network. Sensors **23**(14), 6587 (2023). https://doi.org/10.3390/s23146587

32. Lezoche, M., Panetto, H.: Cyber-Physical Systems, a new formal paradigm to model redundancy and resiliency. EIS **14**(8), 1150–1171 (2020). https://doi.org/10.1080/17517575.2018.1536807
33. Li, C., Li, S., Zhang, A., Yang, L., Zio, E., Pecht, M., Gryllias, K.: A Siamese hybrid neural network framework for few-shot fault diagnosis of fixed-wing unmanned aerial vehicles. JCDE **9**(4), 1511–1524 (2022). https://doi.org/10.1093/jcde/qwac070
34. Li, J., Guo, Y., Dou, Y., Wang, J., Qiu, B., Liu, X.: An approach to bearing fault diagnosis based on ensemble learning and case-based reasoning. In: JPCS. vol. 2787, p. 012042. IOP Publishing (2024). https://doi.org/10.1088/1742-6596/2787/1/012042
35. Liu, M., Li, L., Yan, F.: Intelligent Predictive Maintenance. Springer (2024). https://doi.org/10.1007/978-981-97-2677-6
36. Longo, L., Goebel, R., Lécué, F., Kieseberg, P., Holzinger, A.: Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions. In: 4th IFIP TC Proc. LNCS, vol. 12279, pp. 1–16. Springer (2020). https://doi.org/10.1007/978-3-030-57321-8_1
37. López, B.: Case-Based Reasoning: A Concise Introduction. SLAIML, Springer (2013). https://doi.org/10.1007/978-3-031-01562-5
38. Malburg, L., Schultheis, A., Bergmann, R.: Modeling and Using Complex IoT Time Series Data in Case-Based Reasoning: From Application Scenarios to Implementations. In: 31st ICCBR Workshop Proc. vol. 3438, pp. 81–96 (2023)
39. Nowlan, F., Heap, H.: Reliability-Centered Maintenance. Tech. Rep. AD/A066-579, National Technical Information Service, US Department of Commerce, Springfield, Virginia (1978)
40. Olsson, E., Funk, P., Xiong, N.: Fault diagnosis in industry using sensor readings and case-based reasoning. Journal of Intelligent & Fuzzy Systems **15**(1), 41–46 (2004). https://doi.org/10.3233/IFS-2004-00223
41. Olsson, T., Funk, P.: Case-based reasoning combined with statistics for diagnostics and prognosis. In: Journal of Physics: Conference Series. vol. 364, p. 012061 (2012). https://doi.org/10.1088/1742-6596/364/1/012061
42. Ribeiro, F.M.L.: Similarity-based Methods for Machine Diagnosis. Ph.D. thesis, Universidade Federal do Rio de Janeiro (UFRJ) (2018), pantheon UFRJ Institutional Repository
43. Richter, M.M.: Knowledge Containers. In: Readings in Case-Based Reasoning. Morgan Kaufmann Publishers (2003)
44. Ridley, D.: The Literature Review. SAGE (2012)
45. Schander, R., Schultheis, A., Bergmann, R.: Explanation of Similarities in Temporal Case-Based Reasoning by Visualization. In: 34th ICCBR Proc. LNCS, vol. 16780, pp. 169–186. Springer (2026). https://doi.org/10.1007/978-3-032-33865-5_11
46. Schoenborn, J.M., Weber, R.O., Aha, D.W., Cassens, J., Althoff, K.D.: Explainable Case-Based Reasoning: A Survey. In: AAAI-21 Workshop Proc. (2021)
47. Schultheis, A.: Exploring a Hybrid Case-Based Reasoning Approach for Time Series Adaptation in Predictive Maintenance. In: 32nd ICCBR Workshop Proc. vol. 3708, pp. 230–235. CEUR-WS.org (2024)
48. Schultheis, A., Alt, B., Bast, S., Guldner, A., Jilg, D., Katic, D., Mundorf, J., Schlagenhauf, T., Weber, S., Bergmann, R., Bergweiler, S., Creutz, L., Dartmann, G., Malburg, L., Naumann, S., Rezapour, M., Ruskowski, M.: EASY: Energy-Efficient Analysis and Control Processes in the Dynamic Edge-Cloud Continuum for Industrial Manufacturing. Künstliche Intelligenz **39**, 161–166 (2024). https://doi.org/10.1007/s13218-024-00868-3

49. Schultheis, A., Bertrand, Y., Grüger, J., Malburg, L., Bergmann, R., Serral Asensio, E.: Case-based Data Quality Management for IoT Logs: A Case Study Focusing on Detection of Data Quality Issues. IoT **6**(4) (2025). https://doi.org/10.3390/iot6040063
50. Schultheis, A., Hoffmann, M., Malburg, L., Bergmann, R.: Explanation of Similarities in Process-Oriented Case-Based Reasoning by Visualization. In: 31st ICCBR Proc. LNCS, vol. 14141, pp. 53–68. Springer (2023). https://doi.org/10.1007/978-3-031-40177-0_4
51. Schultheis, A., Zeyen, C., Bergmann, R.: An Overview and Comparison of Case-Based Reasoning Frameworks. In: 31st ICCBR Proc. LNCS, vol. 14141, pp. 327–343. Springer (2023). https://doi.org/10.1007/978-3-031-40177-0_21
52. Seiger, R., Schultheis, A., Bergmann, R.: Case-Based Activity Detection from Segmented Internet of Things Data. In: 33rd ICCBR Proc. LNCS, vol. 15662, pp. 438–453. Springer (2025). https://doi.org/10.1007/978-3-031-96559-3_29
53. Serradilla, O., Zugasti, E., Rodriguez, J., Zurutuza, U.: Deep learning models for predictive maintenance: a survey, comparison, challenges and prospects. Appl. Intell. **52**(10), 10934–10964 (2022). https://doi.org/10.1007/s10489-021-03004-y
54. Si, X., Wang, W., Hu, C., Zhou, D.: Remaining useful life estimation - A review on the statistical data driven approaches. Eur. J. Oper. Res. **213**(1), 1–14 (2011). https://doi.org/10.1016/J.EJOR.2010.11.018
55. Sikorska, J.Z., Hodkiewicz, M., Ma, L.: Prognostic modelling options for remaining useful life estimation by industry. Mech. Syst. Signal Process. **25**(5), 1803–1836 (2011). https://doi.org/10.1016/j.ymssp.2010.11.018
56. Sørmo, F., Cassens, J., Aamodt, A.: Explanation in Case-Based Reasoning – Perspectives and Goals. Artif. Intell. Rev. **24**(2), 109–143 (2005). https://doi.org/10.1007/s10462-005-4607-7
57. Späth, M., Schultheis, A.: Research Dataset from a Smart Factory Model for Case-Based Predictive Maintenance (Mar 2026). https://doi.org/10.5281/zenodo.19242538
58. Sun, F., Zhu, S., Wang, Y.: Illumination Variation Similarity Based Fault Diagnosis for HV-LED Lamp Driven by Segmented Linear Driver. In: ICSCIB 2018 Proceedings. AISC, vol. 890, pp. 547–555. Springer (2018). https://doi.org/10.1007/978-981-13-6733-5_51
59. Weich, J., Schultheis, A., Hoffmann, M., Bergmann, R.: Integration of Time Series Embedding for Efficient Retrieval in Case-Based Reasoning. In: 33rd ICCBR Proc. LNCS, vol. 15662, pp. 328–344. Springer (2025). https://doi.org/10.1007/978-3-031-96559-3_22
60. Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., Sun, L.: Transformers in Time Series: A Survey. In: 32nd IJCAI Proc. pp. 6778–6786. IJCAI.org (2023). https://doi.org/10.48550/arXiv.2202.07125
61. Xing, Z., Pei, J., Keogh, E.: A brief survey on sequence classification. ACM Sigkdd Explorations Newsletter **12**(1), 40–48 (2010). https://doi.org/10.1145/1882471.1882478
62. Yang, Y., Wang, H., Liu, Z., Yang, Z.: Few-shot Learning for Rolling Bearing Fault Diagnosis Via Siamese Two-dimensional Convolutional Neural Network. In: 11th PHM Proc. pp. 373–378. IEEE (2020). https://doi.org/10.1109/PHM-Jinan48558.2020.00073
63. Zhang, A., Li, S., Cui, Y., Yang, W., Dong, R., Hu, J.: Limited Data Rolling Bearing Fault Diagnosis With Few-Shot Learning. IEEE Access **7**, 110895–110904 (2019). https://doi.org/10.1109/ACCESS.2019.2934233

64. Zhao, H., Liu, J., Dong, W., Sun, X., Ji, Y.: An improved case-based reasoning method and its application on fault diagnosis of Tennessee Eastman process. Neurocomputing **249**, 266–276 (2017). https://doi.org/10.1016/j.neucom.2017.04.022
65. Zhao, X., Ma, M., Shao, F.: Bearing fault diagnosis method based on improved Siamese neural network with small sample. Journal of Cloud Computing **11**(1), 79 (2022). https://doi.org/10.1186/s13677-022-00350-1
66. Zheng, W., Yang, L., Genco, R.J., Wactawski-Wende, J., Buck, M., Sun, Y.: SENSE: Siamese neural network for sequence embedding and alignment-free comparison. Bioinform. **35**(11), 1820–1828 (2019). https://doi.org/10.1093/bioinformatics/bty887
67. Zheng, X., Feng, Z., Lei, Z., Chen, L.: Few-shot learning fault diagnosis of rolling bearings based on siamese network. Measurement Science and Technology **35**(9), 095018 (2024). https://doi.org/10.1088/1361-6501/ad57d9
68. Zonta, T., Da Costa, C.A., da Rosa Righi, R., de Lima, M.J., da Trindade, E.S., Li, G.P.: Predictive maintenance in the Industry 4.0: A systematic literature review. Comput. Ind. Eng. **150**, 106889 (2020). https://doi.org/10.1016/j.cie.2020.106889