

Anomaly and Fault Detection and Predictive Maintenance: a Review on Industrial Applications

Davide Frizzo¹, Valentina Zaccaria¹, David Petrovic¹, Francesco Borsatti¹, Davide Sartor¹, Felice Russo²,
Genoveffa Marcario², and Gian Antonio Susto¹

Abstract—Diagnostics and prognostics are crucial components of Smart Factories. Their continuous development is driven by the Industry 4.0 and 5.0 revolution, which relies on Artificial Intelligence (AI), Internet of Things (IoT) and Industrial Internet of Things (IIoT). This survey provides a comprehensive review of AI-enhanced Anomaly Detection (AD), Fault Detection (FD), and Predictive Maintenance (PdM) frameworks, considering both state-of-the-art methodologies and practical applications across diverse industrial domains. Key aspects covered include benchmark datasets, the integration of IoT/IIoT infrastructures, and the economic implications of diagnostics and prognostics in smart manufacturing. Moreover, the emphasis on the human-centric vision of Industry 5.0, the significance of Explainable Artificial Intelligence (XAI) and Uncertainty Quantification (UQ) for fostering transparent and trustworthy industrial AI systems is highlighted. Finally, research gaps and future research directions are considered, notably in model transferability, computational efficiency, and human-centric AI development for real-world deployment.

Index Terms—Anomaly Detection, Fault Detection, Predictive Maintenance, Internet of Things, Industrial Internet of Things, Industry 4.0, Industry 5.0, Machine Learning, Deep Learning

I. INTRODUCTION

In recent years the phenomena of Industry 4.0 and 5.0 have been continuously gaining popularity. In particular Industry 4.0 [1], [2] driven by the complementarity of various emerging technologies (i.e. nanotechnology, biotechnology, AI, and others) gave birth to the concept of *Smart Factory*. This refers to an industrial system that learns as it works, undergoing a process of continuous adaptation and optimization. The Fourth Industrial revolution became possible due to multiple technological advancements with the most important certainly being the development of IoT and IIoT based technologies. These innovations are revolutionizing the connectivity between hardware and software by building Cyber Physical Systems (CPS), an example of which is Digital Twin (DT) [3].

As an example the 2021 report by Deloitte [4] considers IoT as the most impactful Industry 4.0 technology, followed by AI and Cloud Computing. The result has been made possible by the development of cheaper and more reliable sensors. Additionally, the availability of cloud infrastructures enables the storage, processing and sharing of the data collected by these sensors in real time without the need of a centralized data center, further automatizing production and maintenance processes.

The proliferation of Big Data and IoT has revolutionized industrial processes by facilitating the collection and processing of vast amounts of data. The data-rich environment arising from the application of these emerging technologies provides a fertile ground for integrating AI algorithms, enabling the automation of crucial manufacturing tasks such as diagnostics and prognostics [5].

As a consequence, this technology is being widely adopted in the industrial domain for the deployment of diagnostics and prognostics methodologies in order to provide an efficient and cost-effective way to reactively and proactively monitor the health of industrial systems. IIoT infrastructures are essentially the main source of data used to produce the diagnostic and prognostic decisions with the main goals of identifying anomalous behavior, faults and their underlying causes, reducing downtime, maintenance costs and extending the lifespan of the industrial assets.

This study focuses on providing practitioners and researchers a comprehensive review on the state-of-the art and applications of AI-enhanced diagnostic and prognostic frameworks which are fastly becoming the standard in smart manufacturing: AD, FD and PdM.

AD and FD, which will be introduced in Sections II and III, are diagnostics approaches. Their aim is to first identify the presence of anomalous behaviors or faults in the manufacturing process (i.e. AD and Fault Detection). Successively it will be analyzed if the type, root cause and location of such malfunctioning can be identified (i.e. Fault Classification).

On the other hand, PdM, presented in IV, is a prognostic strategy. It comes later in the production pipeline and its focus is on forecasting the future state, the Remaining Useful Life (RUL) or the likelihood of failure of an equipment where a fault or a malfunctioning process has been diagnosed. Even though AD, FD and PdM are defined as three separated tasks, they are closely interconnected: diagnostic results produced by AD or FD are the inputs of a prognostic system, such as one that performs RUL estimation.

The main characteristics of the three diagnostic and prognostic frameworks surveyed in this chapter are presented in Table I.

While the goals of Industry 4.0 focus on economical sustainability and operational efficiency, Industry 5.0 shifts the vision towards the central role of humans and societal needs [7], [8], [9], [10]. The focus has shifted toward to the development of new technologies that prioritize human well-being and interaction within industrial processes. To address this paradigm shift, particularly in the context of AD, FD,

¹Department of Information Engineering, University of Padova, Via Giovanni Gradenigo, 6b, Padova, Italy.

²LFoundry SRL, Via Antonio Pacinotti 7, Avezzano, Italy.

TABLE I
MAIN CHARACTERISTICS OF AD, FD AND PDM FRAMEWORKS

Framework	Field	Paradigm	Task	Data modalities
Anomaly Detection	Diagnostic	Un- and Semi-supervised	Classification	Tabular, Image
Fault Detection	Diagnostic	Supervised and Semi-supervised	Multi-Class Classification	Tabular, Multivariate Time Series
Predictive Maintenance	Prognostic	Supervised and Semi-supervised	Regression, Classification [6]	Multivariate Time Series

and PdM, this paper also surveys XAI (Section IV-C) and UQ (Section IV-B2) approaches, which are crucial for fostering human-centric and transparent industrial AI systems.

A. Benchmark Datasets

One of the main challenges when it comes to the integration of AD, FD and PdM in the industrial domain is the availability of high quality datasets. These are needed to train and validate the AI models used to perform diagnostics or prognostics tasks. In industrial settings, in fact, historical data of faults or machine breakages are usually rare since companies are willing to avoid the high-costly and catastrophic situations generated by such events. Moreover also privacy reasons have to be considered. Furthermore, in the past, humans were primarily responsible for constructing maintenance and fault reports on machines. This often resulted in data that is not well-structured or standardized, making it challenging for AI models to effectively process and learn from. Consequently, there is a significant need to build comprehensive benchmark datasets to enable researchers to compare and validate different proposed methods on a common and consistent ground.

Because of the data collection challenges described above these datasets are usually obtained from simulated data, aiming to mimic the degradation process which leads to the machine failure or specific faults.

Even though AD, FD and PdM are quite different tasks they are naturally connected in the manufacturing pipeline and thus usually these datasets can be easily manipulated to be adapted to any of these problems.

In particular:

- For the AD labels contained in the dataset can be ignored because of the unsupervised nature of the task.
- In FD problems usually labels characterizing different faults are available.
- In PdM workflows a degradation model, such as the ones presented in IV-B2, or an Health Index (HI) can be exploited to construct the labels.

Here is a list of the benchmarks datasets most commonly used in the literature:

- **Commercial Modular Aero-Propulsion System Simulation:** This is probably the most common and well documented benchmark dataset in the literature for the tasks of FD and PdM. It is produced by the Commercial Modular Aero-Propulsion System Simulation [11]. It contains sensor measurements from turbofan engines used in passenger aircraft vehicles. In the simulations each

engine starts from random health states which, after a short period of time, start to degrade towards engine failure. The dataset is already divided into a training a training set (sensor measurements available until the end of life (i.e. RUL = 0)) and test set (sensor measurements stop some time before failure). Each dataset is formed by 24 features, 3 being the operations conditions and the remaining 21 being sensor measurements. The dataset is further divided into four sub datasets (i.e. FD001, FD002, FD003, FD004) with different numbers of faults and operating conditions making the RUL prediction task more challenging. Additional challenges posed by this dataset are the heterogeneous lengths of the run to failure life cycles of different engines and the different characteristics of the engines due to manufacturing variations. Moreover multiple degradation modes can develop on different engines even if starting from the same RUL value making it extremely challenging for data-driven models to produce accurate predictions [12].

- **New Commercial Modular Aero-Propulsion System Simulation: The New Commercial Modular Aero-Propulsion System Simulation (N-CMAPSS)** [13], [14] (also referred to as C-MAPSS-2) is an improved version of the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) dataset, introduced in [14]. The improvements include the fact that several common maneuvers performed by commercial flights are now recorded in the data, the equipment degradation onset is related to the operational history and the dataset now also include labels regarding faults. As a consequence, this dataset can be used to benchmark models designed for the FD task.
- **PHM18:** This dataset was introduced for the PHM Data Challenge 2018 [15] and it contains data about Ion Milling etching tools and the aim of the challenge was the prediction of faults on these kind of equipments. Ion Milling is a crucial step in the process of manufacturing integrated circuits, in particular in the production of wafers. This operation, refer to [15] for a detailed description, is quite complex and multiple failures can occur. It is thus beneficial to predict when and where the failures occur to strategically schedule equipment's downtime to intervene with maintenance operations. Thus this dataset can be employed both for the FD and the PdM tasks. The dataset contains time series data regarding the execution of 20 ion milling machines described with 3 failure causes and 22 features. Labels for the three failure

types are provided while the time corresponding to the machine's end of life (needed for the RUL estimation task) is provided indirectly as the time when the machine is switched off by the operator to perform a maintenance intervention. This poses a challenge since the actual time when the failure start may occur much earlier than the failure time and it is not provided. Moreover failure times are not available in case the equipment did not fail within a specified time horizon. As for C-MAPSS the data are splitted into a training and a test set and a specific score function is designed to have an objective evaluation of the solutions proposed by the competition's participants.

- **Lithium-ion batteries:** Another common field of application for PdM methodologies is the one of lithium-ion batteries (i.e. Section IV-D1), used to power electric vehicles. A common benchmark dataset used in this field is contained in the NASA Prognostics Center of Excellence [16] which is provided by National Aeronautics and Space Administration (NASA) and contains charge and discharge data of lithium-ion batteries. Three batteries are considered, all at a capacity of 2 Ah. Charging and discharging experiments are performed on the batteries. Experiments stopped when the battery capacity dropped below a predefined threshold (Acceptable Performance Threshold (APT) of 70%). The battery capacity is recorded at the end of each discharge cycle.
- **Rolling element bearing:** Rolling element bearings constitute one of the main causes of rotating machines failures. There are several benchmark dataset which can be used to benchmark PdM methodologies in this field. One of the most common is the IEEE PHM Challenge 2012 bearing dataset produced through the PRONOSTIA [17] experimental platform, whose aim is to provide real experimental data on rolling bearings degradation. This platform provides degradation data for all possible type of defects and can produce bearing degradation data under different operating conditions. A rotating speed sensor and force sensor are used to monitor the operating conditions while health monitoring is conducted through the collection of temperature and vibration signals through accelerometers. The sampling frequency of these sensors is set in order to capture the entirety of the frequency spectrum the bearing during the whole degradation period. Another benchmark dataset commonly used in the application of FD and PdM approaches on bearings is the Intelligence Maintenance Systems (IMS) dataset [18]. The experimental setting consists of four bearings installed on a shaft. Each bearing is equipped with two Sensitivity Quartz ICP accelerometers measuring the x and y coordinates of the acceleration to which each bearing is subjected. Three datasets are available: both accelerometer data are considered in dataset 1, while only one accelerometer is considered for datasets 2 and 3. This dataset is characterized by a very high sampling frequency of 20 kHz. The three provided datasets correspond to three different experiments which ended with different defects on one or multiple bearings which are used to

construct the failure labels. Other popular benchmark dataset for rolling element bearings are Case Western Reserve University (CWRU) and [19].

- **Tennessee Eastman Process (TEP)**
Dataset: The TEP [20] dataset is typically used as a benchmark for FD. It collects data from Tennessee Eastman (TE) [21], a crucial process for the production of fibers, chemicals and advanced materials for every day purpose. This is a synthetic dataset producing normal and faulty runs. TEP also comes with ground truth knowledge of the main root causes of failures, making it a commonly used benchmark dataset for XAI models [22], [23].

For a more comprehensive list of available benchmark datasets refer to [24].

B. Scopes and Topics of interest

The research fields of AD, Fault Detection and Diagnosis (FDD) and PdM are very wide, containing applications using multiple learning frameworks based on different data modalities. In order to provide a general overview of the field and its application on the industrial world we decided to focus on a portion of topics which we think are more related and interesting for Industry 4.0 and 5.0.

Before the emergence of data-driven approaches, several methods were developed for diagnostics and prognostics, typically relying on domain expertise or mathematical models representing the underlying physical processes. While these methods can provide more than acceptable results they require to communicate with domain expert and to develop overly complex physical models which dramatically increase the computational requirements of the final system. Data-driven solutions are more efficient since, being based just on the relationships between the input and output variables, do not require the design of complex representations of the equipment.

For the above mentioned reasons, we decide to focus mainly on data-driven approaches (i.e. Machine Learning (ML) and Deep Learning (DL) based) since we believe in their superiority in terms of efficiency and predictive accuracy with the respect to less recent methodologies based on domain expertise and non data-driven AI solutions which are leaving the floor for learning-based approaches as the new standard for diagnostics and prognostics.

For what concerns the data modalities for all the three frameworks we restrict our study to tabular or sequential data excluding domains such as images, video, and other high-dimensional perceptual data, we refer the interested readers to [1] [2], which have been published jointly with this review.

The survey is organized as follows: Sections II, III, IV are dedicated to the three frameworks: AD, FDD and PdM, respectively. Each one of these sections provides an initial theoretical background on the specific task (i.e. Sections II-A1, III-A1, IV-A), followed by an in depth description of the state of the art methods (i.e. Sections II-A2, III-A2, IV-B) used to solve them. In Sections II-B, III-B, IV-D an overview of the applications in

industrial environments is outlined. Remaining in the realm of industrial applications, Section V describes some use cases in the context of the Artificial Intelligence in Manufacturing leading to Sustainability and Industry 5.0 (AIMS5.0) project providing a real world example of challenges in integrating the three presented frameworks in a sustainable industrial process. Section VI is dedicated to the identification of research gaps. Finally, the study concludes in Section VII.

II. ANOMALY DETECTION

In this section theoretical foundations and state of the art methods for the AD framework will be presented.

A. AD Methods and Theoretical Foundations

1) *AD Definitions*: It is necessary to formalize the notion of anomaly detection by defining the task of identifying such instances within data.

Definition 1. Anomaly detection, in the context of machine learning, is generally an unsupervised task, though allowing weakly or semi-supervised variants. Its primary objective is to infer binary labels (*normal* versus *anomalous*) or to compute continuous anomaly scores that quantify the degree of deviation from normal behavior.

In general, given the absence of extensive ground-truth annotations, the detection task is treated as unsupervised with respect to the anomalous class, implying that the notion of anomaly must be inferred directly from the data. Consequently, the problem is inherently ill-posed, as the absence of clearly defined criteria and ground-truth labels introduces ambiguity in both the formulation and evaluation of the task.

Anomaly: Numerous studies on anomaly detection provide their own definitions of what constitutes an anomaly. For example, D. H. Hawkins defines an anomaly as “an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism” [25]. Or put more simply, anomalies are “cases that are in some way unusual and do not appear to fit the general patterns of the dataset” [26].

The definition of anomaly adopted in this work, based on Definition 1, is stated as follows.

Definition 2. An anomaly is a sample identified by both its rarity and significant deviation from the normal data points.

To formalize this definition, it is necessary to clarify the notion of rarity, to specify what constitutes a significant deviation, and to define which data points represent normality.

In Definition 2, rarity is identified as one of the defining characteristics of anomalous samples. This property is typically interpreted as the anomaly residing in a low-density region relative to the overall data distribution. However, as noted in the comprehensive review of anomaly types by R. Foorthuis [27], a more nuanced interpretation allows for anomalies to occur in high-density regions, for instance, when the majority of normal observations themselves occupy low-density areas.

In an unsupervised setting, anomalies are identified by their deviation from the majority of the data. In this review, two main scenarios observed in the literature are presented, each describing a different relationship between normal and anomalous data.

- 1) *Outlier*: An outlier is an anomaly present within the data set used for modeling. This implies the available data is contaminated with anomalies, and the detection algorithm must be able to distinguish them from the majority of the data, which is considered normal [28].
- 2) *Novelty*: Given an initial dataset assumed to be free of anomalies, a novelty refers to a previously unseen observation that significantly deviates from the established normal distribution and is therefore identified as anomalous by the detection algorithm [28]. Since the modeling of the data is performed exclusively on normal instances, *novelty detection* can be regarded as a semi-supervised learning task [29], [30], where the model learns the characteristics of normal behavior and subsequently distinguishes deviations from this learned representation as novel or anomalous patterns.

The concepts of outlier and novelty are also discussed in the survey by V. Hodge *et al.* [29], which further introduces a third scenario in which both anomalous and normal data are labeled. This last case, however, falls outside the scope of the present definition of anomaly detection and is instead referred to as fault detection when label information is available.

Given the two scenarios—outlier and novelty—which describe how data is presented to the algorithm, the literature further introduces several types and categories of anomalies. These classifications are discussed in the following section.

Types of Anomaly: The survey by Chandola *et al.* [28] classifies anomalies into three principal categories: point, contextual, and collective anomalies. A point anomaly corresponds to an individual observation that significantly deviates from the expected data distribution. Contextual anomalies typically occur in time series or sequential data, where an observation may appear normal in isolation but is anomalous when considered within its contextual attributes. Collective anomalies, on the other hand, refer to groups of observations that may individually appear normal but collectively exhibit abnormal behavior, often due to relational dependencies or patterns among them. In contrast to contextual anomalies—where the context is derived from intrinsic data attributes—collective anomalies are defined by the relationships between multiple data instances.

The work by Foorthuis [27], [26] provides a comprehensive and systematic characterization of anomaly types, emphasizing their diverse nature and the multiple dimensions along which they can be distinguished. Anomalies may exhibit different characteristics, which can be categorized along several axes, as described below.

Two anomaly types can be identified from a structural point of view:

- *Atomic* (also called *point anomalies*): individual data instances that deviate from the normal distribution.
- *Aggregate* (also called *collective anomalies*): anomalies that arise from groups of related samples, where the

collection as a whole deviates from normal behavior, even if individual instances appear normal.

Two anomaly types can be defined on their dimensionality:

- *Univariate*: anomalies identifiable from a single feature or variable.
- *Multivariate*: anomalies that emerge from specific combinations of multiple features, where each individual feature may appear normal, but their joint occurrence is atypical.

Anomaly groups can be also dependent on the data type:

- *Numeric*: anomalies defined over continuous or discrete numerical attributes.
- *Categorical*: anomalies occurring within categorical or symbolic data.
- *Mixed*: anomalies described by both numerical and categorical features.

In essence, anomalies can be described by any combination of these characteristics, for example, a single faulty sensor reading in an industrial process (atomic, univariate, numeric) or a coordinated network intrusion involving multiple correlated events across hosts (aggregate, multivariate, mixed).

2) *AD Methods*: Anomaly detection methods based on ML can be categorized according to their learning paradigm or model architecture. Following the taxonomy proposed by L. Ruff *et al.* [31], a primary distinction is drawn between *shallow* and *deep* learning approaches. Furthermore, the same authors classify existing methods into four fundamental categories based on their underlying principles: *classification-based*, *probability-based*, *reconstruction-based*, and *distance-based* approaches, which are better described in III.

Shallow machine learning is often referred to as traditional machine learning and typically relies on explicit feature engineering, where domain knowledge is used to design relevant input representations. In contrast, *deep learning* approaches employ end-to-end models that jointly learn feature representations and perform the detection task within a unified architecture.

Given that anomalies are inherently rare and unpredictable, extensive manual feature engineering often implies the use of labeled or domain-specific knowledge. Such cases align more closely with supervised *fault detection* and are therefore discussed separately in Section III-A.

Since extensive feature engineering is not generally applicable within the anomaly detection framework, shallow methods are instead characterized by the use of models that do not rely on deep neural architectures. These methods operate directly on the available feature space and typically include approaches such as support vector machines (SVM), k-nearest neighbors (k-NN), and ensemble-based techniques like Isolation Forests. Such algorithms focus on identifying deviations in data distributions or local density structures without the hierarchical representation learning typical of deep models.

Taxonomy of AD methods: Anomaly detection methods are often specialized depending on the data modality. This survey focuses on **tabular** and **time series** data, which present distinct structural and modeling challenges, summarized in Table II.

Other data modalities such as visual, textual, graph, or 3D point cloud data introduce different challenges and modeling paradigms, but are beyond the scope of this survey.

Table III shows the main categories of anomaly detection methods and their core principles.

In subsequent sections, each category will be further divided into **shallow** and **deep** methods, reflecting differences in model capacity, feature learning, and scalability.

3) *Density-Based AD*: Density-based anomaly detection methods operate on the principle that normal instances occur in high-density regions of the data distribution, while anomalies reside in sparse or isolated areas. These approaches estimate the probability density function of the data, either explicitly or implicitly, and identify samples located in low-density regions as anomalous. The core assumption underlying these methods is that the data-generating process produces normal observations with higher probability than anomalous ones.

Shallow Density-Based Approaches: Shallow density-based methods can be broadly categorized into three families: local density estimation approaches, histogram-based methods, and probabilistic mixture models.

Shallow density-based methods span a spectrum from local density approaches that capture fine-grained neighborhood structure to histogram-based methods that sacrifice multivariate dependencies for scalability. This trade-off is reflected in computational complexity: local methods such as LOF and LDF typically require $O(n^2)$ operations for pairwise distance computations, though approximations and indexing structures can reduce this in practice. Histogram-based approaches achieve linear or near-linear complexity, making them suitable for large-scale applications. LOCI additionally introduces an approximate algorithm with linear complexity through box-counting methods.

Local methods excel at detecting contextual anomalies in datasets with varying density clusters but require careful parameter tuning. Histogram-based approaches provide substantial speedups suitable for large-scale or real-time detection but are generally limited to global outliers unless combined with spatial reasoning. Copula and mixture model approaches offer principled statistical foundations with varying degrees of distributional assumptions. The choice among these methods depends on computational constraints, the nature of expected anomalies (global versus local), and the validity of underlying distributional assumptions.

Local Density Estimation. A foundational approach in density-based anomaly detection is the Local Outlier Factor (LOF) [32], which introduced the concept of quantifying the degree to which an observation deviates from its local neighborhood rather than treating outlier detection as a binary classification problem. LOF computes a local reachability density for each point and compares this density to those of neighboring points, addressing datasets with clusters of varying densities where global methods fail.

Building upon LOF, the Local Correlation Integral (LOCI) method [33] introduced the Multi-Granularity Deviation Factor (MDEF), providing automatic, statistically-principled outlier determination without user-specified cutoffs through standard deviation-based flagging across multiple scales. The Local

TABLE II
DATA MODALITIES CONSIDERED IN THIS SURVEY AND THEIR ASSOCIATED CHALLENGES.

Modality	Description	Challenges
Tabular	Fixed-length vector of features, without inherent order or temporal dependency.	Mixed-type features, high dimensionality, missing data.
Time series	Ordered sequence of observations over time, with temporal dependencies.	Temporal correlations, seasonality, irregular sampling, concept drift.

TABLE III
CATEGORIES OF ANOMALY DETECTION METHODS AND THEIR CORE PRINCIPLES

Category	Core Principle
Density	Estimate the probability density function of the data and identify samples located in low-density regions as anomalies. These methods assume that normal data points occur in dense regions, while anomalies lie in sparse or isolated areas. Approaches range from classical statistical estimations to neural or flow density models.
Distance	Compute dissimilarities between samples using a distance metric and label points that deviate significantly from their neighbors as anomalies. Such methods rely on the assumption that anomalies are distant from most other data points in input space, feature space, or along temporal trajectories.
Ensemble	Combine the outputs or decision functions of multiple detectors to improve robustness and reduce model bias. Ensembles exploit model diversity through random subsampling, feature bagging, or heterogeneous learners to capture different anomaly characteristics and achieve better generalization.
Reconstruction	Learn a model capable of reconstructing normal data patterns and detect anomalies as samples with large reconstruction errors. The underlying hypothesis is that models trained on normal data cannot accurately reproduce unseen or irregular patterns. Both linear (e.g., PCA) and nonlinear (e.g., autoencoders, GANs) reconstruction frameworks are common.
Clustering	Group data into clusters and identify anomalies as points that do not belong to any cluster or are far from cluster centroids. The intuition is that normal data form compact clusters, whereas anomalies are isolated or weakly associated with any group. Methods vary from classical k-Means and density clustering to deep clustering variants.
One-Class	Learn a boundary or manifold that encloses normal data and treats points outside this region as anomalies. These methods directly model the support of the normal distribution, often via kernel machines or deep representation learning, and are particularly suited when only normal samples are available for training.
Forecasting	Model temporal dependencies to predict future observations and identify anomalies as deviations between predicted and actual values. These methods assume that normal behavior follows predictable temporal patterns, while anomalies correspond to unpredictable deviations. Although forecasting detection is conceptually similar to reconstruction as both involve reproducing parts of the input it differs fundamentally in that time series data are inherently ordered, and prediction exploits the temporal dependency between successive observations.

Density Factor (LDF) [34] further refined this direction by providing a statistically grounded framework through variable-width kernel density estimation, handling multivariate distributions where normal points in sparse clusters may exhibit lower absolute density than outliers in dense regions.

Further developments include methods employing specialized kernels and weighted neighborhood formulations for improved robustness to parameter selection [35], frameworks that explicitly decouple density estimation from outlier scoring phases [36], and approaches that unify kernel density estimation with k -nearest neighbor methods through observation-specific kernels adapted to local covariance structure [37].

Histogram-Based Methods. Histogram-based approaches trade modeling sophistication for computational efficiency. The Histogram-Based Outlier Score (HBOS) [38] constructs independent univariate histograms for each feature and combines their density estimates to compute anomaly scores. By assuming feature independence, HBOS achieves substantial speedups over local density methods, though this assumption precludes detection of local outliers dependent on feature correlations.

To retain local outlier detection capability while maintaining efficiency, multivariate histogram approaches have been developed [39]. These methods identify and discard dense, uniform

regions where outliers are unlikely to exist, reducing the computational burden of subsequent local density calculations. The Lightweight On-line Detector of Anomalies (Loda) [40] demonstrates that an ensemble of weak detectors—one-dimensional histograms on random sparse projections—can achieve competitive performance with complex methods, while additionally supporting streaming data and providing anomaly explanations.

Copula and Mixture Models. COPOD [41] leverages copula theory to detect outliers by computing tail probabilities in multivariate distributions. The method constructs an empirical copula to separate marginal distributions from dependency structure, identifying outliers as observations with extremely small tail probabilities. COPOD is parameter-free, deterministic, and provides interpretability through dimensional outlier graphs quantifying each feature's contribution to abnormality.

Gaussian Mixture Models (GMM) offer a parametric alternative [42], fitting a mixture of Gaussian components and identifying outliers as points falling outside high-probability regions. While GMMs capture multi-modal distributions and provide probabilistic interpretability, they assume the data follows a Gaussian mixture distribution, which may not hold for complex real-world datasets.

Deep Density-Based Approaches: Deep density-based methods leverage neural networks to model complex, high-dimensional probability distributions that are intractable for shallow approaches. These methods can be organized into three main families: normalizing flow-based approaches, hybrid autoencoder-density models, and diffusion-based methods. Normalizing flows provide exact likelihood computation through invertible transformations, with variations including conditional flows for structured data, likelihood ratio construction, and pseudo-anomaly generation for classifier augmentation. Hybrid approaches such as DAGMM demonstrate that combining autoencoders with explicit density models through joint optimization outperforms decoupled alternatives. Diffusion models offer stable training dynamics and effective handling of contaminated training data through rejection schemes. Computational considerations vary across families: normalizing flows require invertible architectures with tractable Jacobian computation; DAGMM adds mixture model estimation overhead but avoids autoencoder pre-training; diffusion models incur costs from iterative denoising during both training and inference.

Normalizing Flow-Based Methods. Normalizing flows provide a framework for explicit density estimation through invertible transformations that map complex data distributions to simple base distributions (typically Gaussian). The change of variables formula enables exact likelihood computation, making flows particularly suitable for density-based anomaly detection where anomalies are identified as samples with low probability under the learned distribution.

Conditional normalizing flows extend this framework by factorizing joint probabilities along relevant dimensions [43]. For multivariate time series, factorization along temporal and attribute dimensions transforms complex joint density evaluation into tractable conditional density estimation, with attention mechanisms parameterizing the conditional information

to capture long-range dependencies and dynamic inter-variable correlations.

Normalizing flows also enable construction of fully data-driven likelihood ratios by estimating conditional densities in different data regions [44]. By computing the ratio of densities learned from potentially anomalous regions versus background regions, an optimal test statistic for anomaly presence can be derived, providing both classification capability and direct background estimation.

Beyond direct density evaluation, normalizing flows have been applied to generate pseudo-anomalous samples for training augmentation [45]. By sampling from low-density regions of a flow trained on normal data, synthetic anomalies can be generated to transform one-class classification into binary classification. This model-agnostic approach bridges one-class and two-class methods, enabling effective utilization of any available labeled anomalies while maintaining robustness to unknown anomaly types.

Emerging research explores quantum implementations of normalizing flows [46], leveraging the natural invertibility of unitary quantum gates. Quantum architecture search optimizes gate sequences to map data distributions to target distributions, with anomaly scores computed via divergence measures between quantum measurement distributions and the learned normal distribution.

Hybrid Autoencoder-Density Models. A limitation of using autoencoders alone for anomaly detection is that reconstruction error provides only an implicit density proxy. Hybrid approaches address this by combining dimensionality reduction with explicit density estimation in a unified framework.

The Deep Autoencoding Gaussian Mixture Model (DAGMM) [47] jointly optimizes a deep autoencoder with a Gaussian mixture model through end-to-end training. The key innovation is constructing a low-dimensional representation that concatenates encoded features with reconstruction error features, capturing both how feature correlations differ for anomalies and the reconstruction difficulty. An estimation network predicts mixture membership probabilities, replacing the traditional Expectation-Maximization algorithm and enabling differentiable training. This joint optimization ensures that dimensionality reduction directly serves the density estimation objective, avoiding the information loss inherent in decoupled two-stage approaches.

Diffusion-Based Methods. Diffusion models learn data distributions by training a neural network to reverse a gradual noising process. For anomaly detection, the denoising capability provides a natural density proxy: normal samples are easier to denoise than anomalous ones.

For tabular data, diffusion-based approaches address the challenge of training on contaminated (unlabeled) data through iterative anomaly rejection schemes [48]. By progressively identifying and removing likely anomalies based on denoising difficulty, the learned density increasingly focuses on the normal data distribution. Anomaly scores are computed by aggregating reconstruction losses across multiple diffusion timesteps, providing multi-scale detection that captures anomalies at different noise levels. This approach demonstrates

particular effectiveness on high-dimensional data where classical methods suffer from the curse of dimensionality.

4) *Distance-Based AD*: Distance-based anomaly detection methods operate on the principle that anomalies are significantly distant from most other data points in the feature space. These approaches compute dissimilarities between samples using a distance metric and identify points that deviate substantially from their neighbors as anomalous. The core assumption is that normal instances form cohesive structures where points are relatively close to one another, while anomalies reside in isolated regions far from the majority of observations.

Shallow Distance-Based Approaches: Shallow distance-based methods can be organized into two main families: nearest neighbor approaches and alternative dissimilarity measures. Nearest neighbor methods rely on distances to neighboring points as anomaly indicators, with various strategies for computing and aggregating these distances. Alternative dissimilarity approaches address fundamental limitations of geometric distance in the original feature space, either by introducing data-dependent measures that adapt to local density variations, by exploiting directional information rather than distance magnitudes, or by computing distances in robust subspaces resistant to contamination. The computational complexity of these methods ranges from $O(n^2)$ for classical k -NN to $O(n)$ for ensemble-based density estimators, with the choice depending on dataset scale and dimensionality characteristics.

Nearest Neighbor Methods.

The k -nearest neighbor (k -NN) framework forms the foundation of distance-based anomaly detection, where anomaly scores are derived from distances to neighboring points. Classical approaches compute the distance to the k -th nearest neighbor or aggregate distances across multiple neighbors, with larger values indicating potential anomalies.

A key challenge in nearest neighbor methods is parameter sensitivity, particularly the choice of distance threshold or neighborhood size. Hybrid approaches address this through adaptive, data-driven neighborhood definitions [49]. By computing the average pairwise distance across the dataset and using it as an automatic neighborhood radius, such methods eliminate the need for prior knowledge of data distribution. Recursive partitioning based on neighbor counts progressively concentrates candidate outliers, followed by k -NN distance refinement to extract true anomalies.

For multivariate time series, the choice of local distance metric within Dynamic Time Warping (DTW) significantly impacts detection performance. Mahalanobis distance-based DTW learns a distance matrix that captures feature correlations and assigns appropriate weights to different variables [50]. This approach addresses the limitation that standard DTW with Euclidean distance treats all variables equally, ignoring heterogeneous feature relevance and inter-variable correlations present in multivariate data.

Computational efficiency represents a significant concern for nearest neighbor methods, which traditionally require $O(n^2)$ distance computations. The LiNearN algorithm [51] achieves $O(n)$ time complexity by fundamentally reconsidering the density estimation problem: rather than computing nearest neighbors for every instance, it estimates density by

averaging over multiple local regions derived from small subsamples. This ensemble approach yields adaptive local regions that automatically adjust to varying density distributions without requiring global parameters.

Alternative Dissimilarity Measures.

A fundamental limitation of geometric distance is its inability to capture the data-dependent characteristic where two instances in a dense region should be judged more dissimilar than two instances at the same geometric distance in a sparse region. Mass-based dissimilarity addresses this by replacing geometric distance with probability mass [52]. The dissimilarity between two points is defined as the probability mass of the smallest region covering both instances under a hierarchical partitioning model, typically implemented using isolation trees. A critical departure from traditional distance measures is that mass-based dissimilarity exhibits non-constant self-dissimilarity: points in high-density regions have larger self-dissimilarity than points in sparse regions. This property enables detection of local anomalies in clusters of varying densities, where fringe points of dense and sparse clusters receive similar anomaly scores despite having vastly different absolute densities.

In high-dimensional spaces, distance-based methods suffer from the curse of dimensionality: the relative contrast between nearest and farthest neighbors converges to zero as dimensionality increases, regardless of data distribution. Angle-Based Outlier Detection (ABOD) addresses this by using the variance of angles between difference vectors rather than distance magnitudes [53]. The key insight is that angular relationships remain meaningful discriminators in high dimensions where distance comparisons deteriorate. Points deep within clusters have difference vectors pointing in all directions, yielding high angle variance, while outliers have most points clustered in limited directions, producing low angle variance. ABOD computes the variance of angles between all pairs of difference vectors from a query point, weighted inversely by the squared distances to reduce the influence of distant points. A significant advantage is the parameter-free nature of the method, directly producing outlier rankings without requiring neighborhood size or distance threshold specification.

An alternative strategy computes distances in transformed spaces designed to be robust to outlier contamination. Standard subspace projection techniques such as PCA are vulnerable because anomalous points distort the covariance matrix, corrupting the principal components. CoMadOut [54] addresses this by employing comedian PCA, a robust variant that replaces covariance with the comedian matrix based on median absolute deviation. Since median-based statistics have high breakdown points, the resulting principal components maintain correct orientation even under substantial contamination. Orthogonal distances to subspace axes then serve as anomaly indicators, with distribution-aware scoring through kurtosis weighting adapting decision boundaries to handle varying tail behaviors across different principal components.

Deep Distance-Based Approaches: Deep distance-based methods leverage neural networks to learn representations where distance computations become more meaningful for anomaly detection. These approaches address limitations of

computing distances in raw feature spaces by projecting data into learned latent spaces where normal samples cluster near a reference distribution or manifold, and anomalies manifest as distant points. The computational overhead of deep architectures is offset by the ability to handle high-dimensional data and capture complex patterns that shallow methods cannot model effectively.

Bidirectional GAN architectures enable learning low-dimensional latent representations where distances to a reference distribution serve as anomaly scores [55]. By training an encoder jointly with a generator through adversarial objectives and reconstruction losses, normal data is mapped to approximately follow a standard normal distribution in latent space. The Mahalanobis distance to this reference distribution simplifies to the Euclidean norm of encoded features, providing a computationally efficient anomaly score that requires only a single encoder forward pass at inference time.

For data-center-scale applications involving millions of machines with multivariate time series, learned representations enable tractable clustering through distribution-based distances [56]. Rather than computing distances between raw high-dimensional time series, RNN-VAE encoders compress data to low-dimensional latent representations, and machines are clustered based on Wasserstein distances between their latent representation distributions. This approach reduces the problem from training millions of individual models to training hundreds of cluster-specific models through transfer learning, achieving near-optimal accuracy with orders-of-magnitude speedup.

Self-supervised contrastive learning provides an alternative framework where distances in embedding space reflect transformation consistency [57]. Rather than relying on hand-crafted transformations appropriate only for images, learnable neural network transformations are jointly optimized with an encoder to produce diverse yet semantically consistent views of normal data. The training loss, which measures similarity between transformed and original samples in the learned embedding space, directly serves as the anomaly score, with anomalies exhibiting high scores due to inconsistency with learned transformation patterns.

5) Ensemble-Based AD: Ensemble-based anomaly detection methods combine multiple detectors to improve robustness and reduce model bias. These approaches exploit diversity through random subsampling, feature bagging, or heterogeneous learners to capture different anomaly characteristics. The core principle is that aggregating multiple weak detectors yields more reliable scores than any single model.

Shallow Ensemble-Based Approaches: Shallow ensemble methods comprise two main families: isolation-based approaches that build ensembles of partitioning structures to explicitly isolate anomalies, and selective combination methods that intelligently aggregate outputs from multiple base detectors based on estimated competence. Isolation methods achieve linear complexity through subsampling, while selective methods add competence evaluation overhead but avoid degradation from poorly performing members.

Isolation-Based Methods.

The isolation paradigm [58] exploits the observation that anomalies, being few and different, require fewer partitions to separate from normal data. Random recursive partitioning isolates anomalies closer to the root, with scores derived from average path length across an ensemble. The partitioning can use axis-parallel splits as in the original formulation or random hyperplanes [59] to handle non-axis-aligned structures. However, infinite hyperplane extension causes score inconsistencies as splits in dense regions affect distant sparse areas; finite hyperspheres [60] constrain each split to its local neighborhood, restoring proper locality.

Tree-based partitioning produces global scores that may miss local anomalies. Hypersphere-based isolation [61] centers regions on instances with radii from nearest-neighbor distances, automatically adapting to local density. Multi-scale approaches [62] combine global and local isolation by partitioning space into subregions with separate forests, capturing anomalies at different spatial scales.

For streaming data, static structures cannot adapt to distribution drift. Dynamic histogram-based approaches [63] replace fixed trees with structures that split and merge bins as data arrives and expires, enabling continuous adaptation. The randomness in isolation can also be guided by data statistics: kurtosis-weighted feature selection [64] directs splits toward heavy-tailed dimensions, while single-neighbor similarity in small subsamples [65] demonstrates that extreme simplicity suffices when aggregated across an ensemble.

Selective Combination Methods.

Generic ensemble combination assumes equal detector contribution, but competence varies across feature space and poor detectors can neutralize good ones. Selective methods address this by evaluating which detectors to trust for which instances.

Dynamic per-instance selection [66] evaluates competence within local neighborhoods, selecting detectors whose scores correlate with pseudo ground truth from ensemble statistics. When seeking a global subset, graph-based approaches [67] map rankings to similarity graphs where good detectors form dense communities of mutual agreement, enabling selection through community detection without pseudo ground truth. Competence estimation can be further refined near decision boundaries where detector disagreement matters most, using adaptive neighborhoods stretched along boundary directions [68] to improve validation in regions where class probabilities change rapidly.

Deep Ensemble-Based Approaches: Deep ensemble methods extend the isolation and combination principles to learned representation spaces, addressing limitations of shallow approaches that operate directly on raw features. The core insight is that anomalies difficult to isolate or detect in the original feature space may become separable in appropriately transformed spaces, with neural networks providing the non-linear mappings required for complex data structures.

Deep Isolation Forest [69] addresses two fundamental limitations of shallow isolation: the inability to isolate non-linearly separable anomalies and systematic score biases from axis-parallel partitioning artifacts. Rather than improving the splitting geometry as in extended or generalized variants, the method transforms the data space itself using randomly initial-

ized neural networks, making simple axis-parallel cuts in the transformed space equivalent to non-linear cuts in the original space. Critically, these networks require no training—casual initialization provides sufficient diversity across ensemble members, maintaining the computational efficiency that makes isolation methods practical. The approach generalizes across data types by simply changing the random architecture: fully connected networks for tabular data, recurrent layers for time series, and graph neural network layers for graph-structured data.

When labeled anomalies are available, even in small quantities, supervised deep ensembles can refine decision boundaries beyond what unsupervised methods achieve. Stacked autoencoders combined with ensembles of probabilistic neural networks [70] use deep feature extraction to project data into more separable representations, then employ multiple classifiers with varying smoothing parameters to balance the trade-off between false alarms and missed detections. The ensemble across hyperparameters provides diversity complementary to data-based diversity, and the framework naturally handles multiple outlier types through multi-class probabilistic outputs.

For time series data, hybrid approaches combining statistical preprocessing with deep architectures [71] demonstrate that feature selection maintaining interpretability can enhance deep models including LSTM-autoencoders, GANs, and transformers. Ensembling multiple deep architectures captures different aspects of temporal anomalies, with reconstruction-based and forecasting-based paradigms converging on similar effectiveness when properly combined with domain-informed preprocessing.

6) Reconstruction-Based AD: Reconstruction-based anomaly detection methods learn models capable of reconstructing normal data patterns and identify anomalies as samples with large reconstruction errors. The underlying hypothesis is that models trained on normal data cannot accurately reproduce unseen or irregular patterns, causing anomalies to exhibit systematically higher reconstruction loss. This principle applies across representation frameworks from linear subspace methods to sparse coding and deep autoencoders, with deep methods enabling non-linear mappings through neural architectures and introducing adversarial training or probabilistic latent spaces to improve reconstruction quality and anomaly discrimination.

Shallow Reconstruction-Based Approaches: Shallow reconstruction methods primarily build on Principal Component Analysis (PCA) and sparse representation frameworks, with reconstruction error measuring deviation from learned normal structure.

PCA-based approaches project data onto principal subspaces capturing dominant variance directions. Classical linear PCA computes reconstruction error in the original feature space but fails to capture complex non-linear manifolds. Kernel PCA [72] addresses this by performing PCA in an implicit infinite-dimensional feature space via the kernel trick, where reconstruction error becomes the squared distance to the principal subspace in feature space. For RBF kernels, decision boundaries follow principal subspace geometry with heterogeneous variance modeling, providing tighter enclosure than

uniform boundaries from one-class SVM or SVDD. However, both linear and kernel PCA face scalability challenges on large or streaming datasets where eigenvalue decomposition and covariance matrix storage become prohibitive. Online variants [73] address this through incremental updating without explicit eigen-decomposition, using an oversampling strategy that duplicates target instances to amplify outlier effects on principal directions—necessary because single anomalies produce negligible eigenvector variation in large datasets.

Sparse representation methods offer an alternative framework where normal data admit sparse decomposition over dictionaries. Rather than measuring reconstruction magnitude, joint sparse coding and anomaly detection [74] explicitly models outliers as structured deviations through a unified optimization that separates sparse normal components from a column-sparse outlier matrix. Non-zero columns directly indicate anomalous samples, treating anomalies as signals unexplainable by the sparse dictionary model.

Deep Reconstruction-Based Approaches: Deep reconstruction methods leverage neural network architectures to learn non-linear mappings between input and latent spaces, enabling modeling of complex data distributions beyond the capacity of linear methods. These approaches can be organized into two families: autoencoder-based methods that focus on deterministic or probabilistic encoding-decoding pipelines, and adversarial methods that incorporate discriminator networks to improve reconstruction quality or anomaly discrimination.

Autoencoder and Variational Methods.

Standard autoencoders learn compressed representations through encoder-decoder architectures, with reconstruction error serving as the anomaly score. However, the choice of reconstruction metric significantly impacts detection performance. While most methods use Mean Squared Error treating dimensions independently, Mahalanobis distance [75] incorporates covariance structure between variables, providing scale-invariant measurement particularly suitable for correlated multivariate data where sensor measurements exhibit strong interdependencies.

Variational Autoencoders (VAE) extend deterministic autoencoders with probabilistic latent spaces, learning distributions rather than point encodings. For time series data, temporal structure in both observation and latent spaces becomes critical. OmniAnomaly [76] addresses this by combining VAEs with GRUs and introducing stochastic variable connections that establish temporal dependence among latent variables across time steps, enabling the model to capture patterns missed by methods treating latent representations independently. Planar normalizing flows further enable learning non-Gaussian posterior distributions, while Peaks-Over-Threshold from Extreme Value Theory provides principled automatic threshold selection.

Architectural innovations for multivariate time series include explicit modeling of inter-variable dependencies alongside temporal patterns. GTAD [77] combines Temporal Convolutional Networks for capturing long-range temporal dependencies with Graph Attention Networks that learn input features correlation structures without requiring domain expertise. The hybrid prediction-reconstruction objective and

per-dimension loss optimization enable fine-grained anomaly localization while the parallel architecture provides computational efficiency over recurrent alternatives.

Adversarial Methods.

Adversarial training addresses limitations of standard autoencoders, which minimize reconstruction error universally and may fail to distinguish subtle anomalies from normal data. USAD [78] employs two autoencoders sharing an encoder, where the second autoencoder learns to distinguish real data from reconstructions of the first through adversarial objectives. This dual-path architecture amplifies reconstruction errors for anomalous patterns while maintaining training stability absent in pure GANs, with configurable sensitivity through weighted combination of both reconstruction paths at inference.

Dual-discriminator architectures provide complementary regularization at multiple points. DAEMON [79] uses separate discriminators for latent space regularization (guiding posterior distributions toward the prior) and reconstruction space regularization (preventing overfitting by ensuring reconstructed distributions match real data statistics). This addresses both generalization failures from irregular latent encodings and overfitting to training-specific patterns, while per-dimension reconstruction errors enable root cause analysis.

GAN-based methods offer alternatives where generators learn normal data distributions and anomalies are identified through various comparison mechanisms. AnoGAN principles adapted for tabular data [80] use per-sample latent vector optimization to find representations best reconstructing each test instance, with reconstruction error indicating distributional deviation. Rather than relying solely on reconstruction, GASN [81] exploits the implicit bias of GANs toward majority distributions when trained on contaminated data. Since normal samples dominate statistically, generated synthetic data serves as a normal reference benchmark, with anomalies identified through neighborhood analysis comparing original samples to synthetic neighbors via Euclidean distance.

7) *Clustering-Based AD*: Clustering-based anomaly detection operates on the principle that normal data points naturally form cohesive groups, while anomalies appear as isolated instances or points weakly associated with any cluster. Anomalies are identified either as noise points that cannot be assigned to clusters under density or distance constraints, or as points exhibiting significant deviation from cluster centroids. This approach provides unsupervised structure discovery without requiring labeled anomalies and naturally handles multimodal distributions through multiple clusters, though performance depends critically on the appropriateness of the clustering algorithm for the data distribution and careful parameter selection.

Shallow Clustering-Based Approaches: Shallow clustering methods apply classical algorithms directly to the feature space, identifying anomalies through distance to centroids, membership in sparse clusters, or classification as noise points. A fundamental challenge across these methods is parameter sensitivity, determining cluster counts, density thresholds, and neighborhood sizes typically requires domain expertise or expensive tuning.

Density-based approaches address this challenge through various parameter automation strategies. Both [82] and [83] enhance classical density-based clustering by incorporating domain-specific structure rather than relying on manual parameter tuning. The former introduces automated parameter selection for OPTICS through geometric covering space models that identify optimal neighborhood sizes based on pairwise distance distributions, while the latter adapts DBSCAN to seasonal time series by transforming temporal data into circular coordinate spaces that capture periodicity. This geometric feature engineering enables detection of local anomalies within seasonal contexts, deviations that appear normal globally but are anomalous relative to their temporal position. Taking a different approach, [84] replaces density thresholds entirely with statistical modeling, using multivariate normal distributions and Mahalanobis distance for membership determination in streaming data. The method performs single-pass clustering and outlier detection simultaneously, constructing virtual statistical distributions around outliers to enable consistent distance calculations and tracking population evolution through hierarchical sub-clustering.

Hybrid methods combine clustering with supervised classification to leverage complementary strengths. [85] and [86] both employ cascaded architectures where clustering pre-processes data for subsequent supervised learning. The earlier work uses k-Medoids with SVM in a two-stage framework, exploiting k-Medoids' robustness to outliers, medoid-based representatives are less biased by anomalies than mean-based centroids, critical when anomalies are the detection target. The recent work extends this to a three-stage cascade: K-Means for efficient initial partitioning, DBSCAN for density refinement, and SVM for final classification. This progression reflects recognition that no single algorithm excels across all data distributions, with each stage compensating for limitations of previous components. However, multi-stage pipelines increase complexity and propagate errors between stages.

[87] addresses multivariate time series through weighted Fuzzy C-Means, learning variable importance weights via Particle Swarm Optimization to balance contributions from time series components with different scales and relevance. The method unifies amplitude and shape anomaly detection by applying clustering to both original subsequences and their autocorrelation representations, the latter removing temporal shifts to enable pattern-based comparison.

These approaches demonstrate enhancement strategies for classical algorithms: geometric feature engineering (circular coordinates, covering space models), statistical modeling (Mahalanobis distance, virtual distributions), architectural composition (cascaded clustering-classification), and adaptive weighting (PSO-optimized distance metrics). While computationally efficient and interpretable, shallow methods operate on fixed feature representations and struggle with high-dimensional data or complex patterns without appropriate preprocessing.

Deep Clustering-Based Approaches: Deep clustering methods jointly learn feature representations and cluster structure through neural architectures, operating in learned latent spaces rather than hand-crafted features. This paradigm en-

ables automatic discovery of representations optimized for clustering separability while addressing high-dimensional and complex temporal data where shallow methods struggle.

A central question in deep clustering design is how to combine representation learning with clustering objectives. [88] demonstrates that simple architectural constraints can outperform complex joint optimization, showing that L2 normalization of autoencoder representations, constraining encoded vectors to the unit hypersphere, yields superior clustering with standard k-means compared to sophisticated methods like DEC and IDEC. The geometric insight is that normalization makes Euclidean distances equivalent to angular distances, inherently more suitable for clustering algorithms. For anomaly detection, this approach dramatically outperforms reconstruction error-based methods by measuring deviation from cluster structure rather than reconstruction quality, revealing a critical flaw: reconstruction-based detection fails when anomalies are structurally simple despite being semantically anomalous.

Recognizing that single representations cannot optimally serve multiple objectives, [89] splits the autoencoder latent space into specialized subsets: clustering-friendly variables optimized for discrimination and reconstruction-friendly variables for input fidelity. A clustering loss applies only to the top-K most discriminative features, selected dynamically during training based on inter-cluster distances. This meta-learning principle, the network learns which representational components specialize in which tasks, distinguishes the approach from static feature allocation. Validation on industrial time series demonstrates substantial improvements over standard autoencoders and shallow methods.

Beyond pure accuracy, [90] addresses algorithmic fairness by integrating adversarial training with deep clustering. The framework jointly optimizes reconstruction, clustering, and fairness through a dynamic weighting scheme that de-emphasizes outliers during training while using adversarial losses to hide sensitive attribute information in learned representations. This demonstrates that deep clustering's representational flexibility enables multi-objective optimization, accuracy, fairness, discrimination, simultaneously, with fairness constraints potentially improving detection by forcing more robust features.

[91] eliminates the requirement to specify cluster count by replacing k-means with BIRCH clustering, which dynamically determines cluster assignments through hierarchical tree structures. The method combines hyperparameter-optimized autoencoders with BIRCH's single-scan processing capability, suitable for streaming network traffic where cluster counts evolve over time. Experimental results show that this component-based approach, carefully optimized autoencoder plus adaptive clustering algorithm, can outperform integrated joint optimization methods.

The evolution of deep clustering reflects three design philosophies: architectural simplicity through normalization constraints versus specialized objectives through latent space splitting, integrated joint optimization versus optimized component composition, and pure accuracy versus multi-objective optimization including fairness and automatic adaptation. The field has progressed from establishing feasibility to addressing

practical deployment requirements, streaming data, unknown cluster counts, fairness constraints, with methods representing different trade-offs between performance, interpretability, and computational cost.

8) *One-Class-Based AD*: One-class methods explicitly learn decision boundaries or manifolds that enclose normal data. Observations outside this region are treated as anomalies. While many anomaly detection approaches train exclusively on normal samples, this category is distinguished by its direct optimization of boundary compactness or support estimation. This contrasts with reconstruction-based methods that measure deviation through reconstruction error, or distance-based methods that rely on proximity metrics. The explicit boundary formulation enables principled threshold determination and theoretical interpretability. However, it introduces challenges in preventing degenerate solutions where boundaries collapse to trivial representations.

The one-class paradigm was established through Support Vector Data Description and OCSVM in the early 2000s. These classical methods remain relevant for contextualizing modern deep learning extensions. Contemporary research addresses three main challenges: representation collapse in neural architectures, contamination robustness for real-world data, and modality-specific adaptations for temporal and tabular domains. The progression from foundational collapse-prevention (2020) through contamination robustness (2024) to decomposition and disentanglement paradigms (2025) reflects maturing research addressing practical deployment challenges.

Shallow One-Class-Based Approaches: Classical kernel-based methods established theoretical foundations through OCSVM extensions that address two key limitations: adapting to sequential data and handling contaminated training sets.

For temporal data, phase space embeddings [92] bridge one-class SVMs to time series by converting sequential data into vector representations via time-delay windows. The core innovation addresses low-frequency signal bias through projected phase space. This orthogonal projection implements adaptive high-pass filtering without manual tuning. Multi-scale detection across varying embedding dimensions ensures robustness by requiring consistent anomalousness. This exemplifies shallow learning characteristic approach: explicit feature engineering precedes established algorithms.

For contaminated training, adaptive sample weighting [93] combines local density with global positioning to identify suspicious samples. Standard OCSVM equally penalizes all training samples, allowing outliers to distort boundaries. The adaptive approach downweights suspicious samples through combined local-global criteria, preventing contamination from corrupting the decision boundary. Experiments demonstrate significant improvements, validating that intelligent sample weighting enhances boundary robustness without requiring clean training data.

Deep One-Class-Based Approaches: Deep one-class learning confronts representation collapse as its fundamental challenge. This occurs when networks learn trivial mappings that satisfy objectives without capturing meaningful structure. Solutions diverge into three areas: preventing collapse through

adversarial training or architectural design, handling contaminated training data, and adapting to specific modalities.

Preventing Representation Collapse. Early deep one-class methods, particularly DeepSVDD, suffered from collapse where all points map to identical representations. This trivially satisfies hypersphere objectives but captures no meaningful patterns. Two complementary strategies emerged to address this challenge: adversarial training and hierarchical architectural design.

The adversarial approach [94] leverages the manifold hypothesis: normal data resides on low-dimensional manifolds, enabling synthetic anomaly generation through off-manifold perturbations. Training alternates between creating hard negatives via gradient ascent around normal points and optimizing classification between originals and perturbations. Collapse becomes impossible because identical mappings eliminate discriminability. This manifold-based strategy generalizes across modalities without requiring domain-specific augmentations, unlike transformation-based self-supervision. The DROCC-LF variant learns Mahalanobis distances when limited negatives are available.

In contrast, the hierarchical approach [95] prevents collapse through architectural constraints rather than adversarial pressure. THOC introduces multi-hypersphere structures for time series, replacing DeepSVDD single global hypersphere with hierarchical layers that capture multi-scale temporal dynamics. Dilated RNNs with exponentially increasing skip connections enable lower layers to model short-term patterns while upper layers capture long-term dependencies. Each layer learns multiple hypersphere centers organized via differentiable soft clustering. Orthogonal regularization prevents center redundancy while temporal self-supervision enhances representation learning. Ablations validate that both hierarchical structure and multiple hyperspheres per layer are essential, removing either substantially degrades performance. These complementary approaches establish that collapse prevention requires either discriminative pressure or explicit structural constraints.

Contamination-Robust Time-Series Methods. Real-world training data frequently contains unlabeled anomalies, violating clean-data assumptions. Recent work addresses this through three distinct philosophies: soft uncertainty-based masking, hard input decomposition, and learned feature masking with pseudo-labeling.

Uncertainty modeling [96] treats contamination as a confidence problem. COUTA produces dual distance predictions whose disagreement signals potential contamination. High uncertainty triggers exponential masking coefficient reduction while penalizing uncertainty itself to ensure confident predictions on legitimate normal samples. This soft masking preserves informative boundary samples that hard filtering would discard. Domain-specific synthetic anomalies, point, contextual, and collective perturbations, provide anomaly-aware guidance through an auxiliary classification head. The dual calibration (uncertainty modeling and synthetic anomalies) contributes synergistically.

Hard decomposition [97] takes a fundamentally different approach inspired by Robust Principal Component Analysis. Rather than weighting samples, ROC explicitly separates

inputs into low-rank clean structure and sparse anomalous components. Only clean data feeds the autoencoder while sparsity constraints apply to anomalous parts. Neural networks enable nonlinear temporal decomposition through alternating optimization, extending linear Robust PCA to complex temporal patterns. This hard separation proves particularly effective when contamination exhibits structured sparsity. The method also addresses single-hypothesis limitations by combining reconstruction and one-class objectives.

Learned masking with pseudo-labeling [98] addresses both contamination and the identical shortcut problem simultaneously. AMFormer network-driven masking breaks identity shortcuts by forcing reconstruction from partial features, while active MSE loss assigns exponentially decreasing weights based on reconstruction errors. Unlike COUTA uncertainty-based soft masking, this exponential pseudo-label weighting provides more aggressive downweighting. Unlike ROC hard decomposition, the soft re-weighting preserves boundary sample information. The unified single-stage optimization jointly learns which features to mask and how to weight samples, avoiding the misaligned objectives of two-stage methods.

These approaches represent a spectrum from soft to hard contamination handling, each suited to different contamination characteristics.

Tabular-Specific Adaptations. Tabular data presents unique challenges: unknown structure, mixed types, heterogeneous features, low correlation, and sensitivity to single-feature changes. Image-focused methods fail to address these characteristics. Two recent approaches demonstrate that specialized architectures are essential, though they pursue fundamentally different design philosophies.

The GAN-based approach [99] adapts generative adversarial training for tabular heterogeneity through dual encoders and 1D-convolutional layers. Where image GANs use 2D convolution, 1D convolution captures tabular feature correlations and single-feature sensitivity. The dual encoder architecture addresses the non-intersecting manifold problem during training: the real encoder learns the generator inverse for detection while the fake encoder co-trains with the generator for learning guidance. This separation differs from single encoder approaches or integrated dual encoders by providing separated training objectives that prevent mode collapse. Ablations validate that 1D convolution contributes feature correlation modeling while dual encoders provide training stability.

The disentanglement approach [100] addresses a fundamental limitation of prior masking methods where learnable masks produce uniform soft values that essentially mask nothing, leading to identity mappings. Disent-AD employs two-head self-attention to extract non-overlapping attribute subsets with orthogonality constraints ensuring genuine separation. Successful reconstruction from either subset independently validates learned correlation rather than memorization. Ablations confirm that removing disentangling loss causes dramatic drops, the model fails to extract non-overlapping subsets without orthogonality constraints. Analysis reveals two subsets are optimal; more subsets become harder to train and may exceed actual correlation structure.

Both approaches validate that tabular characteristics demand

specialized architectures rather than adapted image methods. The GAN-based strategy focuses on robust training through dual encoders, while the disentanglement strategy focuses on explicit correlation modeling through subset decomposition.

9) *Forecasting-Based AD*: Forecasting-based anomaly detection operates on the principle that normal behavior exhibits predictable temporal patterns. These methods model expected future values and identify anomalies as significant deviations between predictions and observations. The paradigm assumes that models trained on normal data cannot accurately forecast anomalous events, manifesting as elevated prediction errors or violations of learned uncertainty bounds.

Shallow Forecasting-Based Approaches: Classical statistical time series methods form the foundation of forecasting-based anomaly detection. [101] demonstrates this approach through ARIMA and a modified Holt-Winters method applied to network traffic monitoring. Both techniques construct confidence bands around forecasts, flagging observations that simultaneously violate thresholds on multiple metrics as anomalies. The ARIMA model captures linear dependencies through autoregressive and moving average components, while the modified Holt-Winters method explicitly models trend and seasonality through exponential smoothing. A critical design choice emerges in confidence band construction: symmetric intervals based on variance-to-mean ratios versus asymmetric intervals adapted to error distribution characteristics, with the latter proving more effective at reducing false positives.

The 2001 work by [102], while focused on outlier detection within ARIMA model fitting rather than operational anomaly detection, provides important historical context. This earlier research addressed a fundamental challenge: classical maximum likelihood estimation fails when training data contains outliers, creating a bootstrapping problem where contaminated data yields biased models. The solution of robust parameter estimation combined with iterative detection and cleaning influenced subsequent developments in making forecasting methods resilient to noisy real-world conditions.

Modern shallow methods have evolved toward online operation and novel anomaly indicators. [103] introduces online gradient descent for continuous ARIMA parameter adaptation, enabling true streaming detection without batch retraining. Beyond computational efficiency, this approach proposes monitoring discrete differences in model weight coefficients rather than prediction residuals. The rationale is that sudden shifts in optimal parameters may signal system state changes more robustly than residual spikes, though this introduces heuristic complexity through multiple aggregation metrics and threshold selection strategies.

Deep Forecasting-Based Approaches: The evolution from statistical to neural forecasting methods began with recurrent architectures capable of learning long-term dependencies. [104] established stacked LSTM networks as the foundation for this paradigm. By modeling prediction errors as multivariate Gaussian distributions estimated from normal training data, the method provides anomaly likelihoods rather than binary classifications. A key insight is multi-step prediction: forecasting several timesteps ahead ensures the network cap-

tures meaningful temporal structure, generating richer error signatures than single-step prediction.

Alternative neural architectures soon emerged. DeepAnT [105] demonstrated that convolutional networks can effectively replace recurrent models for time series forecasting, offering dramatic computational advantages through parallel processing while maintaining competitive accuracy. Meanwhile, HTM [106] explored neuromorphic principles, using sparse distributed representations and continuous online learning to handle concept drift. A critical innovation in HTM is the two-layer detection framework: the sequence memory predicts future representations, then an anomaly likelihood layer models the distribution of prediction errors to reduce false positives in noisy streams. This separation of prediction from anomaly scoring, rather than direct residual thresholding, proves essential for robust performance.

The shift to multivariate time series required explicitly modeling inter-sensor relationships. Graph Deviation Network [107] addresses this through learned graph structures where attention mechanisms weight the importance of different sensor dependencies for forecasting. Each sensor receives an embedding that captures its behavioral characteristics, and the graph structure itself emerges from similarity in embedding space. Anomalies manifest as deviations from graph-constrained predictions, with attention weights providing interpretability about which relationships were violated. A related idea is applied to in-vehicle CAN traffic in [108], where correlated signals are forecast jointly and an anomaly is raised when their combined prediction error exceeds a threshold.

Transformer architectures brought significant advances through their ability to process entire sequences in parallel and model long-range dependencies. TranAD [109] combines transformers with adversarial training and self-conditioning: reconstruction errors from an initial inference phase become focus scores that guide attention in a second phase, amplifying subtle anomalies. Two decoders engage in an evolutionary adversarial game where one minimizes reconstruction error while the other maximizes it, with training weights shifting from reconstruction-focused to adversarial-focused over epochs. The GAT-Informer hybrid [110] extends this by combining graph attention for sensor relationships with Informer architecture for efficient long-sequence processing, using both short-term and long-term forecasting to validate predictions against future trends.

A distinct line of research focuses on uncertainty quantification rather than point predictions. [111] constructs prediction intervals using Monte Carlo dropout and quantile bootstrap methods across multiple architectures including cascaded networks, echo state networks, and LSTMs. Anomalies are identified when observations fall outside learned uncertainty bounds, with dynamic thresholding based on Chebyshev's inequality adapting to data characteristics. This approach makes explicit what remains implicit in residual-based methods: the confidence in predictions varies across the time series, and anomaly detection should account for this uncertainty.

The comprehensive benchmark study by [71] systematically compares seven deep learning paradigms on the same problem, revealing that hybrid architectures combining LSTM

temporal modeling with autoencoder reconstruction principles significantly outperform either approach alone. The study demonstrates that pure reconstruction methods without temporal modeling produce excessive false positives, while classification-based approaches require labeled data. Ensemble methods and data augmentation via GANs prove particularly effective for handling the severe class imbalance inherent to anomaly detection.

Several design principles emerge across these methods. The shift from univariate to multivariate modeling requires explicit mechanisms for capturing inter-dimensional dependencies, whether through graph structures, attention mechanisms, or joint reconstruction. Finally, interpretability increasingly comes not just from detecting that an anomaly occurred, but from explaining which features or relationships caused the deviation, a capability that graph-based and attention-based methods provide through their learned structures.

10) Summary of AD Methods: Table IV provides a comprehensive summary of the AD methods discussed in this section, organized by category and approach, along with their key references.

11) XAI in AD: In this section, we explore the application of XAI approaches in AD. In many real-world industrial scenarios, explicit fault or anomaly labels are scarce, unreliable, or entirely unavailable, making unsupervised learning a natural and often unavoidable choice. While this paradigm enables the identification of deviations from normal operating conditions without prior knowledge of failure modes, it further increases the interpretability challenges of data-driven models. In particular, it is often unclear why a specific observation is flagged as anomalous, which process variables, signal features, or temporal dynamics contribute most to the detected deviation, or whether the anomaly reflects a meaningful physical phenomenon rather than noise or data artifacts. This lack of transparency limits trust, hampers root-cause analysis, and complicates the deployment of AD systems in safety-critical industrial environments. In this context, XAI techniques become essential to uncover the internal reasoning of AD models, providing both global insights into learned normality patterns and local explanations of individual anomalies, in support of reliable, human-centered monitoring solutions aligned with the Industry 5.0 vision. As will be discussed in the following section, the majority of application-oriented studies on AD in the industrial domain rely on a limited set of deep learning architectures that have demonstrated high detection performance. However, despite their effectiveness, deep anomaly detection models are often criticized for their limited interpretability. For this reason, the majority of XAI methods adopted in this context are post-hoc, with several works proposing their use to improve the transparency of the underlying Neural Networks (NNs) [112]. Importantly, the availability and suitability of XAI methods strongly depend on the specific model architecture employed. In particular, Recurrent Neural Network (RNN) such as Long Short Term Memory Network (LSTM) and Gated Recurrent Unit (GRU), together with reconstruction- and generation-based approaches including Auto Encoders (AE), Variational Auto Encoder (VAE), and Generative Adversarial Network (GAN), are among the

most commonly adopted solutions. Consequently, the following discussion focuses on XAI methods that are applicable to these model architectures, with the aim of demonstrating how interpretability can be enabled in real-world, industrial AD scenarios. AEs are among the most widely adopted architectures for AD, based on the assumption that anomalous samples are more difficult to reconstruct than normal ones. However, their encoder-decoder structure inherently limits interpretability, as anomalies are typically identified through aggregate reconstruction errors. To address this issue, most explainability approaches for AE-based models rely on post-hoc XAI techniques. Common strategies include feature attribution methods, such as Shapley Additive Values (SHAP) [113] and gradient-based explanations, which quantify the contribution of individual input variables to the reconstruction error, as well as model-specific analyses inspecting feature-wise or temporal reconstruction residuals. In addition, visualization techniques, particularly in image and video applications, are often employed to localize anomalous regions by highlighting discrepancies between input and reconstructed data. Despite these efforts, explanations for AE-based AD models are generally local and instance-specific, providing limited insight into the global behavior of the learned normality model [112], as observed in industrial applications such as the AE-based approach proposed by Jang et al. [114]. RNNs, in particular LSTM and GRU architectures, are widely used for AD in sequential and time-series data due to their ability to model temporal dependencies and predict normal system dynamics. However, the temporal recurrence and internal gating mechanisms of RNNs make their decision process difficult to interpret. As a result, explainability for RNN-based AD models is predominantly addressed through post-hoc XAI techniques. Common approaches include feature attribution methods such as SHAP [113] and gradient-based explanations to identify the most influential variables or time steps contributing to an anomalous prediction, as well as surrogate models like Local Interpretable Model-Agnostic Explanations (LIME) [115] to approximate local model behavior. Additionally, attention mechanisms and relevance propagation techniques are sometimes exploited to provide insights into temporal importance patterns. Despite these efforts, explanations for RNN-based AD models are typically local and instance-specific, often computationally demanding, and strongly dependent on the specific RNN architecture and input sequence length, offering limited insight into the global temporal behavior learned by the model, as observed in industrial applications such as the LSTM-based approach proposed by Nor et al. [116].

B. AD Industrial Applications

This is the section where several industrial case studies are considered to showcase how the methodologies presented in II-A2 are applied in real-world scenarios.

Unlike II-A2, where the focus was on the methodological aspects, in this section we mainly treat case studies highlighting the specific choices that are made to ensure a successful integration of such AI-based technologies inside complex industrial scenarios.

TABLE IV
SUMMARY OF AD METHODS REVIEWED IN SECTION II-A, ORGANIZED BY CATEGORY, LEARNING PARADIGM (SHALLOW/DEEP), AND APPROACH.

Category	Type	Approach	Key Methods	References
Density-Based	Shallow	Local density estimation Histogram-based Copula and mixture models	LOF, LOCI, LDF HBOS, Loda COPOD, GMM	[32], [33], [34], [35], [36], [37] [38], [39], [40] [41], [42]
	Deep	Normalizing flows Hybrid AE-density Diffusion-based	Conditional flows, likelihood ratio DAGMM TabADM	[43], [44], [45], [46] [47] [48]
Distance-Based	Shallow	Nearest neighbor	k -NN, adaptive neighborhood, DTW variants	[49], [50], [51]
		Mass-based dissimilarity	Probability mass distance	[52]
		Angle-based Robust subspace	ABOD CoMadOut	[53] [54]
	Deep	Learned representations	BiGAN-Mahalanobis, CTF, contrastive learning	[55], [56], [57]
Ensemble-Based	Shallow	Isolation-based	Isolation Forest, Extended IF, Hypersphere IF	[58], [59], [60], [61], [62], [63], [64], [65]
		Selective combination	LSCP, graph-based selection	[66], [67], [68]
	Deep	Deep isolation Hybrid ensembles	Deep Isolation Forest Stacked AE + PNN, LSTM-AE ensembles	[69] [70], [71]
Reconstruction-Based	Shallow	PCA-based Sparse representation	Linear PCA, Kernel PCA, Online PCA Joint sparse coding	[72], [73] [74]
	Deep	Autoencoder variants Adversarial methods GAN-based	AE, VAE, OmniAnomaly, GTAD USAD, DAEMON AnoGAN variants, GASN	[75], [76], [77] [78], [79] [80], [81]
Clustering-Based	Shallow	Density-based clustering Hybrid clustering-classification	OPTICS, DBSCAN variants k-Medoids + SVM, K-Means + DBSCAN + SVM	[82], [83], [84] [85], [86]
		Fuzzy clustering	Weighted FCM with PSO	[87]
	Deep	Deep clustering	Normalized AE + k-means, latent space splitting, fair clustering	[88], [89], [90], [91]
One-Class-Based	Shallow	Kernel-based	OCSVM, phase space embedding, adaptive weighting	[92], [93]
	Deep	Collapse prevention Contamination-robust Tabular-specific	DROCC, THOC COUTA, ROC, AMFormer Dual encoder GAN, Disent-AD	[94], [95] [96], [97], [98] [99], [100]
Forecasting-Based	Shallow	Statistical	ARIMA, Holt-Winters, Online ARIMA	[101], [102], [103]
	Deep	RNN-based	Stacked LSTM, HTM	[104], [106]
		CNN-based	DeepAnT	[105]
		Graph-based	Graph Deviation Network, GAT-Informer	[107], [110]
		Transformer-based Uncertainty quantification	TranAD MC dropout, quantile bootstrap	[109] [111], [71]

Before diving into specific approaches and applications, it is useful to provide an overview of the main industrial contexts in which AD, FD, and PdM techniques are deployed. Industrial systems typically involve multiple interacting layers — mechanical, electrical, cyber-physical, and human — each of which can be a potential source of faults or abnormal behaviors. Identifying and classifying such faults is therefore essential to understand how AI-based diagnostic and prognostic frameworks are tailored to different operational environments.

Tables V and VIII present a taxonomy of representative industrial fault categories and subcategories that form the conceptual foundation for this overview. The classification highlights the domains where data-driven and hybrid diagnostic systems are most relevant, including:

- Equipment / Mechanical faults, such as wear, imbalance, or component degradation;
- Process / Quality deviations, including process drifts, raw material inconsistencies, or quality defects;
- Sensor / Data anomalies, arising from calibration errors, data loss, or malicious signal manipulation;
- Control System / Cyber faults, including PLC malfunctions, control loop instabilities, and cybersecurity threats.

These fault categories capture the complex and interconnected nature of industrial operations, illustrating how anomalies and failures may emerge and propagate across physical and digital layers. Each group also corresponds to a set of challenges and research efforts in developing AI-driven models capable of ensuring robust, interpretable, and human-centric decision-making.

Building on this taxonomy, the following subsections provide a structured examination of the three key paradigms in

industrial diagnostics and prognostics — AD, FD, and PdM — highlighting their methodological principles, application domains, and relevance within the evolving landscape of Industry 4.0 and 5.0.

As summarized in Table V, AD techniques are utilized across multiple fault categories — from mechanical degradation and process deviations to sensor and cyber-physical anomalies — enabling early identification of abnormal patterns and improving overall system resilience.

1) Mechanical/equipment faults: Mechanical and equipment-related faults represent one of the most extensively studied application domains for AD, encompassing a wide range of phenomena such as rotating machinery degradation, tool wear, bearing defects, and electromechanical component failures. Within this broad category, rotating machinery has attracted significant attention due to the availability of rich vibration data and well-established condition monitoring practices, motivating the development of numerous AD approaches. For instance, [117] employs a hybrid bidirectional LSTM–GRU autoencoder trained on healthy vibration data; reconstruction error is used to identify abnormal machine behavior and predict maintenance needs. [118] also adopts an AE framework but directly processes raw vibration signals without manual feature extraction, reaching an F1-score of 99.6% on real-world datasets. [119] proposes a modular methodology combining AD with XAI tools (SHAP and Local-DIFFI), enabling both automated fault diagnosis and interpretability of model decisions. Recent studies on tool wear detection emphasize AD approaches to overcome the limitations of traditional FD, which often requires extensive labeling. [120] proposes two low-label or label-free strategies for predicting the end of life of drilling tools: a classification-based AD method using XGBoost, and an unsupervised AE model. Both approaches demonstrate high accuracy across different cutting parameters, with the XGBoost model providing an interpretable wear indicator and the autoencoder's reconstruction error effectively correlating with tool degradation. [121] presents an unsupervised deep learning–K-means hybrid method for tool health monitoring, where vibration signals undergo unsupervised pre-training to automatically extract relevant fault features, exploiting an AE model. This integration allows adaptive feature learning and accurate clustering of tool conditions in large-scale, multi-tool datasets, eliminating the dependence on expert knowledge or manual feature design. AD methods have shown strong potential for identifying faults in induction motors and electrical drives, where abnormal data are often scarce. [122] introduces a GAN-based model (GANomaly) that uses audio signals transformed into mel-spectrograms from multiple microphones to detect bearing faults in induction motors. The method achieves high performance, with an AUROC of 0.9822 and 92.52% accuracy, effectively distinguishing between normal and faulty conditions without prior exposure to fault data. [123] proposes a VAE-based approach for anomaly detection in electric drives, relying on reconstruction error as the fault indicator. Trained and tested on data from a 200 W servo-motor testbed, the model demonstrates reliable detection performance and is successfully implemented on a

DSP platform for real-time evaluation, confirming its practical applicability to industrial systems.

2) Process/Quality faults: Within the industrial domain, AD methods have also been applied to monitor process stability and product quality, addressing both process drift and manufacturing defects.

For process drift and anomalies, [124] presents an AD framework for complex industrial machinery such as gas turbines. The proposed method combines AE-based architectures with sliding windows to detect early signs of abnormal behavior directly from operational data, without requiring previously recorded fault instances. Validated on real industrial environments, the approach provides real-time monitoring capabilities and generates valuable information for predictive maintenance. Its adaptable design allows application to other variable-duration industrial processes where labeled data are limited or unavailable.

Regarding product quality defects, [125] investigates vacuum testing in refrigerator production, introducing a Bayesian Neural Network (BNN) for DL-based AD on time-series pressure profiles. The model produces an anomaly score for each unit, supporting operators in inspection and quality control decisions. By explicitly accounting for model uncertainty, the BNN enhances interpretability and reliability, reducing dependence on manual evaluation. Integrated within a decision support system, this approach mitigates human bias, optimizes inspection efforts, and minimizes production costs associated with false positives and false negatives.

3) Sensor/Data faults: In modern industrial environments, the integration of cyber-physical systems (CPS) and Industrial Internet of Things (IIoT) infrastructures has increased exposure to security threats such as false data injection (FDI) and other signal manipulation attacks. AD methods have been proposed to address these challenges by learning normal system behavior directly from data, without requiring labeled attack instances.

[126] introduces an unsupervised RNN-based approach for detecting cyber-attacks in CPS, using a time-series prediction model combined with the Cumulative Sum (CUSUM) technique to identify anomalies. Applied to the Secure Water Treatment Testbed (SWaT) dataset, the method effectively detects a wide range of attack scenarios while maintaining a low false positive rate, and can also pinpoint the specific sensor under attack, enhancing interpretability and response accuracy.

[127] focuses on false data injection detection within IIoT systems, leveraging AEs to exploit both temporal and spatial correlations among sensor data. The approach not only detects falsified measurements but also employs Denoising Autoencoders (DAEs) to recover clean data from compromised signals. Experimental results demonstrate that this method outperforms Support Vector Machine (SVM)-based techniques, offering a more robust and self-correcting mechanism for safeguarding industrial sensor networks.

4) Control/Cyber systems faults: AD techniques are increasingly employed in control and CPS to ensure operational reliability, real-time responsiveness, and resilience against both process-related faults and cyber intrusions.

TABLE V
TAXONOMY OF INDUSTRIAL FAULT TYPES, REPRESENTATIVE SUBCATEGORIES, AND CORRESPONDING USED METHODS.

Fault Type	Category	References
Equipment / Mechanical	Rotating machinery faults (bearings, shafts, gears, imbalance)	[117], [118], [119]
	Tool wear and breakage (cutting tools, milling, drilling)	[120], [121]
	Induction motor / electrical drive faults (stator winding, rotor bars)	[122], [123]
Process / Quality	Process drift and anomalies (chemical, thermal, flow processes)	[124]
	Product quality defects (surface cracks, mis-shapes, etc.)	[125]
Sensor / Data	Cyber-physical signal injection (spoofed/manipulated IoT data)	[126], [127]
Control / Cyber systems	ICS faults (PLC malfunction, loop instability)	[128], [129], [130]
	Cybersecurity faults (intrusion, denial-of-service, adversarial manipulation)	[131]

For Industrial Control System (ICS) faults, [128] demonstrates the use of a composite AE model trained on HIL-augmented ICS data to identify anomalous behavior in ICS without relying on labeled datasets. The approach effectively captures both temporal and value-based deviations, enabling robust real-time monitoring and enhanced process reliability. [129] investigates AD in ICS communication networks under simulated lateral cyberattacks, comparing several machine learning models such as One Dimensional Convolutional Neural Networks (1DCNNs), LSTMs, SVMs, and Isolation Forests (IFs). The 1DCNN achieved the best trade-off between accuracy (0.92) and processing time, making it well-suited for real-time intrusion detection in critical infrastructure. [130] introduces a decentralized and real-time AD framework for cyber-physical production systems, employing multiple 1D Convolutional Autoencoders (1DCAEs) in a sliding-window configuration. By decentralizing model execution across individual subsystems, the approach reduces communication overhead and eliminates the need for expert configuration, achieving reliable detection of local process anomalies in real industrial settings.

Regarding Cybersecurity faults, [131] explores unsupervised clustering-based AD applied to heterogeneous cybersecurity data streams. The study evaluates several clustering algorithms using metrics such as WSSSE and Silhouette scores, identifying K-Means as the most effective method for detecting anomalies in real-time streaming data. The proposed framework demonstrates high detection accuracy while maintaining computational efficiency, supporting timely decision-making in dynamic cybersecurity environments.

III. FAULT DETECTION AND DIAGNOSIS

A. FDD Methods and Theoretical Foundations

In this section the theoretical foundations and the state of the art methodological approaches for FDD will be presented.

1) **FDD Definitions:** FDD is a key function in industrial automation, aimed at identifying malfunctions in monitored systems. As part of condition monitoring, it enables timely recognition of faults, helping to maintain reliability, safety and efficiency.

FDD is usually divided into two main tasks:

- 1) **Fault Detection:** deciding whether the system is operating normally or if a fault is present, producing a binary outcome.
- 2) **Fault Diagnosis:** determining the specific type, cause and/or location of the fault, namely fault classification.

Methods for FDD are commonly grouped into three categories: *knowledge-based*, *model-based*, and *data-driven*. Knowledge-based methods rely on expert knowledge, heuristics, or rule-based reasoning to identify faults.

Model-based approaches, on the other hand, employ mathematical or physical representations of a system nominal behavior. Faults are detected by comparing outputs predicted with such representations against measured data. Abnormal conditions are identified when significant deviations are obtained in the latter comparison.

Both categories perform well when the process is thoroughly understood and accurate models or expert rules can be established. In many complex industrial systems, however, developing such detailed knowledge or precise models is impractical or prohibitively costly, which limits their applicability in such cases. Data-driven approaches, by contrast, overcome these limitations by extracting patterns that characterize normal and faulty system behaviors by large volumes of data [132].

Within data-driven techniques, statistical-based approaches, such as Statistical Process Control (SPC) and Principal Component Analysis (PCA), are widely used for FDD [133]. These methods detect deviations from normal operations by analyzing the underlying data structure, for example through hypothesis testing or dimensionality reduction. Since they do not learn predictive mappings, they are not considered ML methods and are therefore not reviewed here.

In the context of machine learning, FDD is typically formulated as a *supervised learning* problem. Labeled examples of both normal operation and known faults are used to train models, which can then detect and/or classify incoming data according to these predefined categories. This limits FDD to faults seen during training, so novel faults cannot be reliably classified. In contrast, AD is typically unsupervised, detecting any deviation from a learned normal behavior. While this distinction is not universally shared in the literature, we adopt it in this work, forming the basis for the following discussion. Finally, FDD differs from PdM in terms of the task objective.

FDD is diagnostic, assessing the current state of the system, whereas PdM is prognostic, aiming to predict future system degradation and failures (see Section IV-B).

In ML-based FDD, different types of data are collected and used. For machinery fault diagnosis, vibrational data are the most utilized [24], as they are informative for a wide range of machinery faults. Acoustic emissions are also employed, particularly for early fault detection in rotating components and induction motors [134], [135], as they are cheaper and less invasive than vibration-based analysis. Motor current signals are commonly used in Motor Current Signature Analysis (MCSA) for the diagnosis of electrical faults, such as broken rotor bars or shorted stator windings [136]. Other data, such as wear debris analysis, oil analysis or thermal imaging are also employed [24]. For discrete manufacturing machines, common types of logs are cycle time, throughput, utilization, or acceleration/deceleration rates [132]. In addition, for industrial processes, control variables are typically used, such as temperature, pressure, flow rates. For IIoT and CPSs faults may include not only mechanical or process ones, but also cyber-attacks (e.g., in SCADA systems), communication failures or firmware faults. In such context, data for FDD can include network traffic logs or protocol in addition to sensor and control data [137], [138].

2) FDD Methods:

Traditional ML: FDD using traditional machine learning relies heavily on feature engineering, a critical pre-process for the successful deployment of the models. Raw sensor measurements or system logs must be transformed into informative descriptors that capture the characteristic of normal and faulty operation.

The specific types of designed features depend on the domain and available measurements. For *rotating machinery and equipment* standard transformations such as Fast Fourier Transform (FFT), Wavelet Transform (WT), Wavelet Packet Transform (WPT), Empirical Mode Decomposition (EMD), Hilbert-Huang Transform (HHT), Variational Mode Decomposition (VMD), or Short Time Fourier Transform (STFT), are commonly used to expose frequency- and time-frequency patterns that indicate mechanical or electrical faults [24]. Features are typically grouped into:

- **Time-domain features:** statistical descriptors such as Root Mean Square (RMS), kurtosis, skewness, peak-to-peak value, and crest factors. It is also a common practice to compute cumulative versions of these features in order to obtain monotonically increasing variables to better represent the advancement of equipment degradation.
- **Frequency-domain features:** spectral peaks, fault-characteristic frequencies (e.g. Ball Pass Frequency Outer Race), band energy ratios or sideband patterns in induction motors (via MCSA, indicative of rotor or eccentricity faults) [139]. In particular, fault-characteristic frequencies are particularly useful in rotating machinery equipment (i.e. bearings, gears, ...) [140] where the frequency spectrum is characterized by the presence of harmonics associated with the rotational speed of the equipment.
- **Time-frequency features:** These kind of features are used in presence of non-stationary or highly oscillatory

signals, where time and frequency domain features are not able to provide HI capturing the true health state of the equipment. Typically used time-frequency domain features are: STFT, Wigner-Ville Distribution (WVD) and WT.

For *Discrete Manufacturing Machines* (i.e., machines that operate in distinct cycles to produce individual items [132]), additional features such as cycle time, throughput, utilization, and other discrete operational KPIs are often used [132], [141]. Discrete events such as actuator commands and sensor triggers can also serve as informative indicators of faulty behavior [142]. Despite their relevance FDD approaches for discrete manufacturing machine are relatively underexplored in the literature. For *IIoT-specific faults*, including attacks, intrusions or communication failures, features typically describe network communication patterns (e.g., traffic flows, protocol usage), enabling the detection of cyber-physical faults.

Once features are extracted, standard supervised models are trained. Widely used classifiers include Random Forest (RF) [143], [144], [145], [146], [147], gradient-boosted trees (e.g., XGBoost, LightGBM) [148], [145], [147], SVM [149], [150], [151], K-nearest neighbors (k-NN) [152], shallow NN [153], [150] and ensemble models [154]. These models have also been shown to perform effectively in multi-fault scenarios, where a single model must identify and classify multiple simultaneous faults, a more realistic and robust setting [155].

In addition to feature engineering, feature selection is important for improving model accuracy and generalization and mitigate overfitting, for both machinery faults and cyber-attacks. Common approaches include *recursive feature elimination* [148], filtering based on their importance (e.g., derived from RF or XGBoost) [156], correlation-based selection [157], [145], or more advanced optimization-based methods [158].

Hybrid architectures combining deep learning for automatic feature extraction with shallow classifiers for decision making have also been proposed. These approaches often outperform both purely traditional and purely deep models, especially when labeled datasets are limited or fault signatures are complex. For example, [159] transformed vibration signals into 2D Continuous Wavelet Transform (CWT) images, extracted multi-level features with a pretrained LeNet-5 Convolutional Neural Network (CNN), and classified them using an ensemble of RF. Similarly, [160] demonstrated that CNN features combined with RF or SVM improved diagnostic performance while maintaining computational efficiency. On the contrary, [156] first applied RF-based feature selection and then trained a deep network, surpassing both shallow and deep baselines for arc-series fault detection. Hybrid strategies are also applied in multi-fault and explainable settings [155].

Although shallow ML is no longer the state-of-the-art for industrial FDD, it remains effective when distinctive fault signatures can be explicitly extracted, labeled data are limited [143], physical interpretability is important, or computational resources are constrained. In contrast, DL dominates in scenarios with high-dimensional, large-scale, and heterogeneous data or when engineered features are not sufficient [161], [162].

The hybrid approaches described above provide a practical compromise.

Deep Learning: Industrial IIoT systems generate large-scale, high-dimensional and heterogeneous data from diverse sources, including physical sensors and/or cyber sources. These data streams often exhibit nonlinear, noisy and nonstationary behavior, making manual feature engineering for traditional ML methods challenging and potentially insufficient for capturing complex fault patterns. DL methods address these challenges by automatically learning hierarchical discriminative features directly from raw data, enabling end-to-end modeling. Architectures such as CNNs, RNNs, Transformers (TFs) or Graph Neural Networks (GNNs) effectively extract spatial, temporal and contextual features across diverse data modalities, achieving state-of-the-art performance.

Despite the end-to-end feature learning ability, studies have demonstrated that data preprocessing remains a critical step also in DL. For physical sensors, for example, techniques such as filtering, normalization, time-frequency transformations or empirical mode decomposition can enhance significantly signal quality and highlight fault-relevant structures, resulting in improved performance and robustness [163], [164].

A trend is multiscale modeling, with the idea of capturing fault signatures that manifest at different temporal, spatial and spectral resolutions. Mechanical degradation process often produce both short-lived transient impulses and long-term effects, which cannot be fully represented at a single analysis scale. Multiscale architectures address this by constructing feature extractors (convolutional or wavelet-based) with multiple receptive fields sizes or frequency bandwidths, followed by adaptive fusion mechanisms such as attention weights or hierarchical pooling. In practice, such architectures tend to have better generalization under varying speeds, loads and noise conditions [165], [166].

In the following an overview of the most used architecture is described. The distinction lies in feature extraction, while the final decision-making is performed by a classification head that maps these learned features to fault labels.

Convolutional Neural Networks: CNNs are the most adopted models in FDD literature [167] due to their ability to learn local structured features from raw sensory inputs, multivariate signal streams or transformed time–frequency representations.

Two principal structural forms are prevalent:

- **1D-CNNs:** operate on ordered 1D sequences and perform convolution along that axis. They are parameter-efficient, preserve the natural ordering of the data and are appropriate when the discriminative information is one-dimensional (for example temporal) with strong local correlations.
- **2D-CNNs:** operate on two-dimensional representations, typically time–frequency images (e.g., spectrograms, scalograms) or spatially arranged multi-sensor maps, and capture joint spatial, time–frequency or cross-sensor patterns via 2D receptive fields.

Beyond these structural forms, numerous model architectures have been adopted, ranging from ResNet, DenseNet, CapsNet and CNNs with dilated convolutions [168], [24].

The 1D-CNNs are the preferred choice for univariate time series, either in the time or frequency domain (e.g., after applying FFT). For example, [169] presents a gearbox fault diagnosis framework under variable operating conditions that uses vibration data and 1D-CNN. To mitigate the effect of variable operating conditions, which increase variability in the learned features of the same fault types and reduce separability between different fault types, the network applies intra-class and inter-class constraints during training. This ensures that features of the same fault type cluster tightly, while features of different fault types remain well separated, enhancing discriminability and improving fault diagnosis robustness. The study in [170] addresses bearing fault diagnosis using motor-current signals. It applies time-shifting, to remove dominant current components, wavelet-based denoising and a lightweight 1D causal-convolutional network that preserves temporal causality while enabling efficient parallel training. The results show higher accuracy and faster training relative to 2D-CNNs and RNNs. In [171], features extracted from multi-modal sensor signals (vibration and acoustic) are fused and fed into a 1D-CNN: empirical evaluation on a dedicated test platform shows that sensor fusion improves diagnostic performance over single-sensor approaches.

To apply 2D-CNNs, one typically need to convert the time series into a 2D input. Some common approaches are:

- **image transformation:** a univariate time series is divided into equal-length segments and stack them as rows of a 2D matrix, with the pixel/cell values being normalized amplitudes of the signal, resulting in a grayscale image [172]. However, the spatial correlations captured by the network may be artificial, since the 2D structure does not necessarily reflect inherent temporal or frequency relationships in the original signal.
- **Multivariate arrangement:** for multivariate time series, each row may correspond to a different sensor channel, and grayscale values represent amplitude, or multiple channels in the image may encode different sensors [173], [174].
- **Time-frequency transforms:** such as wavelet transform [175], [176] or Short Time Fourier Transform [177], [178], [179] to convert mechanical/process time series into 2D time-frequency representations. These advanced techniques often results in better performance [164]. Some recent research embed these transforms directly in the first convolutional layer so that the model can learn the transformation parameters jointly with the classification task (see Sec. III-A2).

Recently, attention mechanisms have demonstrated to be a powerful enhancement to CNNs in fault diagnosis, allowing network to selectively emphasize the most informative parts of the input while suppressing irrelevant information. Depending on the dimension along which adaptive weighting is applied, attention modules in CNNs can be broadly categorized into channel attention, spatial (local) attention and nonlocal or self-attention. Channel attention learn the relative importance of each feature map, enabling the network to emphasize discriminative sensor channels or frequency bands. For example

in [180], a two-branch CNN-based architecture is used for jointly learn bearing fault diagnosis and signal denoising. Channel-wise attention modules in both branches adaptively reweight feature maps according to task relevance, improving robustness under noisy conditions. In [181], channel attention is combined with nonlocal attention in a residual CNN for multi-task diagnosis from multi-sensor vibration data. In [168] channel attention is used in conjunction with multi-feature parallel fusion.

Spatial attention focuses on intra-feature spatial or temporal local dependencies, highlighting the most informative regions or time intervals within feature map. Nonlocal and self-attention extend this principle to a global scale, capturing long-range temporal or contextual correlations beyond the convolutional receptive field [182]. For instance, in [164], raw bearing vibration signals are decomposed using empirical wavelet transform into low-, medium-, and high-frequency components, which are then stacked as channels and processed by a 1D self-attention-enhanced CNN. The attention mechanism adaptively reweights the frequency channels, emphasizing fault-relevant temporal and spectral features. Other studies effectively combine different types of attention, capturing richer feature interactions and improving generalization in many domains, including rotating machinery, industrial processes and industrial robotics [182], [183], [184], [185].

Recurrent Neural Networks.: RNNs and enhanced variants, LSTM, GRU and bi-directional architectures (BiRNNs) remain conceptually appealing for time-series-based fault diagnosis because they implement a principled mechanism for modeling temporal dependencies. In supervised industrial FDD, however, they increasingly appear as components of larger hybrid architectures rather than as standalone predictors. The reason is structural: while recurrent units capture temporal dynamics, they do not represent spatial correlations or multiscale patterns. Hybrid designs fill these gaps and empirical studies shows that they surpass pure recurrent models. An example is the Bi-LSTM-Quadratic NN (QNN) framework in [186]. The Bi-LSTM provides long-range temporal modeling, while the QNN excels at nonlinear feature extraction and offers improved interpretability. Evaluated on both the CWRU dataset and a custom experimental platform for rolling bearing fault diagnosis, the approach outperforms transformer-based and residual architectures, even if computational complexity remain a challenge for deployment in resource-constrained industrial environments.

Hybrid CNN-RNN architectures are widely spread in the literature. Convolutions extract localized and multiscale structure and recurrent units integrate these features across time, surpassing both architectures when used in isolation. In [162], a modified LSTM performs multi-step autoregression, and a 1D-CNN with channel decay handles fault classification on industrial process data with injected faults. A similar approach is presented in the Bayesian-optimized CNN-LSTM model for hydro-turbine FDD in [187]. In there, the CNN constructs a compressed denoised feature basis before the LSTM processes temporal evolution, while Bayesian optimization tunes the hyperparameters. Similar approaches are specifically developed for industrial robots [188] or rotating machinery [189]. More

elaborated designs extend this principle to multi-sensor and multiscale settings. The integrated parallel CNN-deep residual network-LSTM architecture in [190] fuses heterogeneous sensor streams through parallel convolutional branches, followed by a deep residual network and an LSTM module, improving robustness on both cutting-tool monitoring and bearing-fault diagnosis. Extensions combining CNNs, recurrent units and attention mechanisms further enhance the accuracy, as demonstrated in [165], [191].

Overall, these works converge on the same insight: recurrent networks remain valuable, but their strength is highlighted within hybrid models.

Transformers.: Transformers have recently gained traction in time-series modeling and FDD due to their ability to capture long-range dependencies without relying on recurrence. Unlike RNNs, which propagate information sequentially through hidden states, TFs use self-attention to model global relationships across the entire input. This allows scalable parallelization, making TFs particularly effective for large or high-resolution sequences. However, they typically necessitate more training data than recurrent architectures, which can be limited in the data-scarce and imbalanced scenarios often encountered in industrial FDD.

Transformer-based FDD architectures typically use only the encoder, consisting of stacked self-attention and feedforward layers, followed by a classification head. The self-attention mechanism computes contextual relationships across all input positions, whether in the time, frequency, or time-frequency domain, enabling global interaction throughout the representation space.

In mechanical fault diagnosis, early studies applied TFs directly to raw time-domain signals, such as in [192] and [193]. The latter introduces a hybrid CNN-TF architecture to extract multiscale temporal features, where convolutional layers downsample redundant local information to generate refined key-value representations for self-attention. This approach enhances noise robustness and outperforms baseline models, including CNN, CNN-LSTM, RNN, LSTM and GRU on both CWRU dataset and a rotating machinery experimental platform. In the same direction, [194] extends the hybrid CNN-TF framework into a multitask learning setting, enabling simultaneous optimization of fault type and severity while preserving robustness under variable operating conditions and noise.

TFs can also be successfully adapted to the frequency domain. For example, [195] first transforms raw signals into spectral representations via FFT, which are then processed through parallel transformer encoder branches to capture local, medium-scale and global spectral features. Similarly, [196] introduces a global frequency encoding layer within the TF, performing learnable amplitude and phase filtering to facilitate multiscale feature fusion, achieving robust fault detection under high noise levels.

Time-frequency representations are a dominant input modality for TF-based FDD in rotating machinery. In [197], vibration signals are processed using synchrosqueezed wavelet transforms before attention-based feature extraction. In [198], the vibration signal is decomposed into frequency-band sub-

signals via discrete wavelet transform, converted into time-frequency maps using continuous wavelet transform and fed into a multiple Vision Transformers (ViTs), with final classification achieved through soft voting. Similarly [199] combines CNN and TF modules on vibration data, surpassing purely CNN-based methods on a experimental testbed. While most of these studies focus on vibrational data, TFs have also been applied to other signal modalities, such as acoustics [200].

Efforts to reduce TF-based FDD complexity while retaining global dependency modeling include [201], which reformulate self-attention as depthwise convolution. This lightweight design reduces parameters and computational cost by over 60% while maintaining superior performance on bearing and gearbox datasets.

Although TFs generally outperform CNN-based baselines when sufficient data are available, their generalization suffers under limited or imbalanced datasets, a common challenge in FDD [202]. Beyond conventional data augmentation strategies such as SMOTE or GANs [203], [204], tailored solutions have been explored. Notably, [202] treat attention weights as random variables learned via Bayesian variational inference, enhancing robustness to data scarcity and noise. Experiments demonstrate that this Bayesian transformer surpasses ResNet, Bidirectional Long Short Term Memory (BiLSTM), and vision transformer baselines on small-scale datasets, maintains performance under noisy conditions and achieves good cross-domain generalization.

Graph Neural Networks. In many industrial systems, sensors and process components exhibit structured interconnections that can be naturally represented as graphs. Such data are inherently non-Euclidean, making conventional architectures such as CNNs suboptimal. In contrast, GNNs leverage relational structure to learn fault-sensitive representations that capture both local and global dependencies among process variables.

Formally, a system at a given time (or over a time window) can be represented as a graph

$$G = (\mathbf{X}, \mathbf{A}),$$

where $\mathbf{X} \in \mathbb{R}^{N \times F}$ is the node feature matrix (each row encodes F features for one of the N nodes) and $\mathbf{A} \in \mathbb{R}^{N \times N}$ is the adjacency matrix defining pairwise relationships among nodes. Under normal operating conditions, the system graph is denoted as $G = (\mathbf{X}, \mathbf{A})$, while under faulty conditions it becomes $G_f = (\mathbf{X}_f, \mathbf{A}_f)$, since both node features and topological relations may vary due to faults. Hence, a GNN can, in principle, detect and localize faults whose signatures manifest either in node attributes or in the graph topology [205].

To apply GNNs, two main steps are typically required:

- 1) **Graph construction:** define the set of nodes, their feature vectors, and the adjacency matrix encoding pairwise relationships.
- 2) **Model selection, training, and inference:** choose a GNN architecture and train it to propagate information through message passing, producing node or graph embeddings used for fault classification.

Regarding graph construction, the definition of nodes and edges depend on the data source and, for time series, in the

univariate or multivariate nature. For univariate time series, two main modeling approaches are commonly adopted [206]:

- **Series-as-graph:** each time series is segmented into windows, and each window is treated as a node. Features extracted from each segment form the node attributes, while the adjacency matrix encodes relationships among segments. The task is graph classification: extracted node embeddings are aggregated into a graph-level representation, on which the final classifier operates to predict the fault class for the entire time series [207], [208], [209].
- **Series-as-node:** each complete time series in a dataset is assigned to a single node, with statistical or learned features forming the node attributes. The task is node classification: the classifier outputs a label (or probability) for each node. This configuration enables modeling of inter-sample relationships through the adjacency matrix, for instance by encoding similarities between healthy and faulty conditions [210].

For multivariate (multi-sensor) time series, inter-variable dependencies can be modeled explicitly. Each sensor or process variable corresponds to a node, and its measurements or extracted features (statistical, temporal, or spectral) form the node attributes. The adjacency matrix encodes relationships among sensors, based on physical connectivity (e.g., from process topology [211]), statistical dependencies or learned during training. Single graphs are usually defined per time window to represent the system state. Some strategies do not explicitly model temporal continuity between windows/graphs [212], [213], [214], treating each snapshot independently. Other studies, instead, jointly capture both spatial and temporal dependencies by constructing spatio-temporal graphs. In this case, a sequence of graph instances:

$$\{G_t = (\mathbf{X}_t, \mathbf{A}_t)\}_{t=1}^T$$

is defined over consecutive time intervals, where both $\mathbf{X}_t$ and $\mathbf{A}_t$ may evolve with time. The GNN-based model used to deal with such representations, called Spatio-Temporal Graph Neural Network (STGNN), comprises two interacting modules: a spatial module (GNN, Graph Convolutional Network (GCN), ...) operating on each G_t to capture inter-sensor dependencies at each time step, and a temporal module (e.g., RNN-, Temporal Convolutional Networks (TCN)- or attention-based) to model temporal dynamics, either in time or frequency domain [215], [216], [217]. Alternative formulations also exist. For instance, [218] divide multivariate time series into non-overlapping windows, each containing data from all sensors. Each window is treated as a node in the graph, reformulating the problem as node classification.

Regarding edges/adjacency matrix construction, the literature distinguishes between heuristic-based and learning-based approaches. Heuristic methods rely on pairwise similarity, where edges are created between nodes with similar features (e.g., using cosine similarity [207], [219] or Pearson correlation [210], [216]). To maintain sparsity, only the top-k most similar nodes are typically retained, forming a k-NN graph [220], [218]. Other heuristics capture functional dependencies, where edges represent causal relationships (e.g.,

inferred from Granger causality tests [221]) or reflect known physical process topologies [212], [222]. An increasingly adopted alternative is learning-based adjacency construction, where the matrix $\mathbf{A}$ is optimized jointly with the downstream task and defined as a function of trainable parameters and input features [223]. A common formulation expresses $\mathbf{A}$ via attention mechanisms, as in Graph Attention Network (GAT) [224], which assign data-dependent edge weights based on node relevance. Attention can be static, when learned once and fixed during inference, or dynamic, when recomputed adaptively [215], [213], [217].

In the input layer of the GNN, each node can be associated with raw or transformed signals, or with extracted feature vectors. In the latter case, the features can be obtained manually, based on expert-designed statistical, temporal, or spectral descriptors [221], or automatically, for example using CNN [223], in an end-to-end manner. In particular, in [223], the authors transform the CNN features into node embedding through a dynamic graph embedding module. Then, the constructed graph is fully connected, with an edge-dropout mechanism, enabling the diagnosis of rolling bearing faults under variable working condition.

Once the graph is defined, different GNN-based architectures can be used:

- GCN [225]: performs graph convolutions to aggregate feature information from neighboring nodes according to the adjacency matrix $\mathbf{A}$. Two main formulations exist: *spectral* and *spatial*. Spectral GCNs define convolution operations in the graph Fourier domain, using the graph Laplacian eigenbasis, and are effective when $\mathbf{A}$ reflects global statistical relationships or spectral similarity among features (e.g., in vibration or frequency-domain graphs). Spatial GCNs, instead, aggregate messages directly over local neighborhoods and are more efficient and interpretable for process or sensor networks with explicit topological connections. Both are particularly suitable for systems with relatively static process structures, where physical or empirical couplings among sensors remain approximately constant over time.
- GAT [224]: generalizes GCN by introducing learnable, data-dependent attention coefficients. Instead of uniformly weighting neighbors, each edge e_{ij} receives a learned attention score reflecting the relative importance of node j to node i . This mechanism enables the model to adaptively focus on fault-relevant relationships, for example when correlation strengths between sensors change under fault conditions. Multi-head attention further stabilizes training and enhances feature discrimination.
- GraphSage [226]: addresses inductive learning—generalization to unseen nodes or graphs—by learning an aggregation function rather than relying on the full adjacency matrix. Instead of computing convolutions over the entire graph, GraphSage samples a fixed-size neighborhood and aggregates its features via mean, max pooling, or LSTM-based operators. This sampling-based strategy allows generalization to unseen nodes or new operating conditions, advantageous in dynamic IIoT environments.

- STGNN: designed for dynamic processes in which fault evolution exhibits both spatial and temporal dependencies. The spatial module (often a GCN or GAT) captures inter-sensor correlations at each time step, while the temporal module (e.g., RNN, TCN, or attention mechanism) models temporal evolution. This joint modeling enables learning spatio-temporal fault signatures.

Beyond conventional GNN-based architectures, recent research has explored more flexible frameworks. Some works explore more discriminative architectures, for example Graph Isomorphism Networks [227]. A major direction is hierarchical or multi-scale modeling, where nodes or subgraphs are progressively coarsened into higher-level abstractions to capture fault patterns across different structural or temporal resolutions [166]. For instance, [228] introduces a hierarchical supergraph framework for rotating machinery fault diagnosis. Segments from a vibration signal are first represented as spatio-temporal graphs, whose nodes correspond to frequency bins from the STFT and edges encode spectral proximity. Each graph is embedded into a low-dimensional representation and it is then treated as a node in a higher-level supergraph. This design reformulates the original graph-classification task as meta-level node classification. Similarly, [216] combines local subsystem graphs and a global graph to jointly model intra- and inter-unit interactions within industrial systems.

Large Language Models.: In the last couple of years, the proliferation of Large Language Models (LLMs) beyond conventional text generation tasks has also prompted interest in their use for FDD. Emerging research on integrating LLMs into industrial FDD can be broadly grouped based on the role the LLM plays in the diagnostic pipeline: (i) as the primary fault detector, or (ii) as supporting component.

In the first paradigm, the LLM serves as the primary fault detector. Some approaches leverage in-context reasoning using knowledge graphs (KGs). For example, [229] incorporates a causal knowledge graph via prompt-tuning to guide quality defect analysis in aerospace manufacturing, while [230] exploits the LLM to construct a Dynamic Master Logic (DML) knowledge graph from unstructured documentation for nuclear power plant subsystems. In this latter framework, the LLM effectively extracts DML logic, links functional models to natural-language reasoning, interprets user queries, selects reasoning tools to propagate faults through the KG, and retrieves relevant KG segments to generate human-readable explanations. Another major class fine-tuned the LLM on sensor data. The LLM is fine-tuned on paired datasets, where the input is a processes representation of the sensor signal, eventually augmented with context-information, and the label is the corresponding fault information, often with explanatory text. For example, [231], [232] convert vibration features into semantic text enriched with bearing geometry, operating conditions, and characteristic fault frequencies; these texts are paired with fault type and severity labels for fine-tuning. Similarly, [233] constructs a multimodal dataset of envelope spectral images paired with textual descriptions of the fault class and underlying mechanisms; during inference, new spectral images are converted to the same representation, and the fine-tuned LLM predicts the fault class while generating explanations. More

complex pipelines, such as [234], first embed raw time-series signals via an LSTM and align them with textual embeddings using contrastive learning; fuzzy semantic embeddings handle ambiguity across similar fault patterns, prompt learning injects domain knowledge, and then a fine-tuning adapts the model for inference. The authors in [235] conduct an empirical study on fine-tuning LLM for fault diagnosis in complex systems, using both a simulated chemical process dataset (TEP) and a real high-speed train braking system dataset. The study demonstrates high diagnostic accuracy and strong few-shot learning performance, with the models performing well even when data are limited or imperfect. The effectiveness of LLMs is further enhanced by incorporating contextual information, while smaller models offer greater efficiency with minimal accuracy loss. However, despite their predictive strength, the LLMs exhibit limited explainability, often producing reasoning that is unreliable or inconsistent.

In the second paradigm, the LLM functions as a supporting component rather than directly predicting faults. LLMs help to construct or refine knowledge graphs [236], manage human-in-the-loop workflows by overseeing data preprocessing, model selection, and result explanation [237]. For example, [236] leverages an LLM to build detailed ontologies for robotic fault diagnosis and to act as a “teacher” that refines synthetically generated training data, improving model reliability in low-data environments. Similarly, [237] implements a human-in-the-loop pipeline where the LLM manages preprocessing and visualization, corrects causal relations obtained from standard algorithms using domain knowledge, controls model selection and updates, and explains diagnostic results in plain language, while humans provide feedback to ensure accuracy and consistency. An exhaustive review is provided in [238].

Physics-Informed Machine Learning: Previous approaches to FDD often face challenges like scarcity of labeled fault and the inherent “black-box” nature of many models. Physics-Informed Machine Learning (PIML) mitigates these issues by embedding physical domain knowledge into the diagnostic framework, shifting FDD from a purely statistical pattern-recognition task to a more physics-consistent inference task. This enhances model interpretability, diagnostic accuracy and robustness under varying operational conditions.

While physical models can be employed either to generate synthetic faults for data augmentation [239], [240] or to introduce new features based on system equations, enhancing discriminativity [241], PIML goes further. It incorporates physical laws directly into the model structure or the learning process. Following [242], two categories can be identified:

- 1) Physics-informed regularization, and
- 2) Physics-guided architectures.

In physics-informed regularization, physical laws appear explicitly in the training loss function:

$$\mathcal{L} = \mathcal{L}_{\text{accuracy}} + \lambda \mathcal{L}_{\text{physics}}. \quad (1)$$

where $\mathcal{L}_{\text{accuracy}}$ measures prediction error with respect to labeled data, $\mathcal{L}_{\text{physics}}$ penalizes violations of known physical constraints. A widely adopted example of this idea are Physics-Informed Neural Networks (PINNs), which incorporate differential equations describing system dynamics into the loss

function. In FDD, these losses have been used, for example, to estimate physical parameters which are interpretable fault indicators. For instance, [243] trained a PINN to model pressure dynamics in axial piston pumps, an indicator of wear severity, embedding conservation law of mass and fluid compressibility, while [244] applied multistage PINN to estimate valve flow characteristic under normal conditions, ensuring physical consistency via valve sizing equations. The residuals between predictions and observed values are analyzed to detect anomalies. Authors of [245] use a PINN to estimate five parameters that characterize the health state of rotating shafts (i.e., the radial and angular position of the disk-induced static unbalance, the stiffness along the principal elastic axes, and the damping coefficient in the non-rotating reference frame). Alternative approaches enforce energy conservation in rotating machinery under normal and faulty operation regimes and using the resulting estimates as diagnostic features for downstream classifiers [246], constrain decisions to agree with coarse physics-based threshold models to stabilize critical classes [247], or align the model to approximate physics-based labels [248]. Recent extensions apply partial derivative equations residuals together with boundary conditions to hydraulic pipelines and pump flow-ripple prediction [249].

On the other hand, physics-guided architectures embed physics directly into the network structure. A representative family of methods replaces or initializes early convolutional kernels with physically motivated filters. [250] designed a modular CNN, for bearing fault diagnosis, with one specialized feature extractor module for each type of fault. The first convolutional layer of each module is replaced with kernels corresponding to analytically derived fault characteristic frequencies. The features are then combined to make the final classification. Similarly, [251], [252], [253], [254] initialize convolutional filters with wavelet functions (e.g., Morlet or Laplace) providing a physically interpretable basis for vibration analysis while still allowing the filters to be fine-tuned during training. In [255], a continuous wavelet layer replaced the first layer of a CNN, where only the scale and translation parameters are learned. Deterministic modal-property layers explicitly extract modal features that are invariant to operating speed and excitation, improving cross-condition robustness [256]. In [257], a parametric spectral-weighting layer learn a weighting function over the envelope/order spectrum that emphasizes analytically derived characteristic frequencies. The weighting parameters are trained jointly with the classifier to maximize discriminativity while remaining physically interpretable. Finally, physics-informed priors is also used to guide hyperparameter selection [258].

Deep Reinforcement Learning: The architectures discussed so far have been trained under the supervised learning paradigm, relying on labeled pairs of examples $(x_i, y_i)_{i=1}^N$. In contrast, Reinforcement Learning (RL) follows a fundamentally different trial-and-error approach, in which an agent learns to select *actions* based on a *reward signal* that indicates the effectiveness of those actions, particularly suitable for sequential decision making. Deep RL has also been applied to diagnostic tasks.

In a large portion of studies, FDD is formulated as a

single-step decision making problem or "guessing game" problem [24]. The RL agent receives a portion of a sensor signal (vibrational, acoustic, etc.) as its *state*/input and selects an action corresponding to a diagnostic decision. In a K -class classification problem, the action space contains K elements (one for each fault type plus the absence of the fault). The reward signal replaces traditional labels, providing positive feedback for correct diagnoses and negative feedback otherwise. These formulations have demonstrated effectiveness in a range of FDD applications, including rotating machinery, hydraulic pumps, gearboxes and nuclear power plant equipment [259], [260], [261], [262]. Differences consist in various learning algorithms (e.g., Q-learning, REINFORCE, etc.), underlying architectures that constitute the feature extractor parts of the agent (e.g., CNNs, attention-based models), preprocessing strategies and datasets used for evaluation. Since RL does not solve the problem of class imbalance, [263] introduces a data-augmentation method based on a generative model that combines a Variational Autoencoder with a Wasserstein Generative Adversarial Network and Gradient Penalty, before training an RL agent. This approach enhances training data diversity and has demonstrated effectiveness in both laboratory experiments and electric locomotive applications.

While these guessing-game formulations are advantageous in scenarios with nonstationary or evolving data, where online RL allows the agent to adapt to changing patterns, the benefit for static datasets, where all labels are already available, is less clear. In such cases conventional supervised learning may be generally straightforward and efficient, as RL may require extensive data interactions.

Beyond direct fault classification, RL has been employed to optimize diagnostic model components. Neural architecture search is one such application, where RL agents explore, for example, candidate CNN architectures and receive rewards based on resulting accuracy [264]. Similarly, RL has been applied to hyperparameter optimization, as adaptive learning-rate scheduling [265].

Other approaches use RL to mitigate class imbalance typical of FDD. In [266], an agent learns adaptive sample weights to scale a cross-entropy loss of a downstream classifier. The rewards are derived from the classifier performance, especially on minority classes. An alternative approach treats imbalance as a sample-selection problem, modeled as a multi-armed bandit [267]. Here, the agent selects a subsample of points for training of a separate classifier, with rewards based on its validation performance. The method outperforms conventional oversampling and undersampling techniques. A similar methodology has been successfully applied to HVAC systems [268]. While effective, these approaches require iterative retraining, which can introduce substantial computational overhead.

Challenges:

Scarce and imbalanced datasets: Since faults cannot be freely generated in industrial environments, real-world datasets for FDD are typically small and highly imbalanced. In standard supervised learning, this imbalance leads to poor classification performance on minority fault classes, degrading overall diagnostic accuracy. Data augmentation techniques aim

to mitigate this issue by generating synthetic samples from existing data. Oversampling approaches are the most common, with SMOTE [269] being a widely adopted example [24]. Transformations in the time or frequency domains are also possible. Generative models such as VAE or GAN have been explored for fault data synthesis [270], [271], [272]. In a real manufacturing plant, [273] embeds GAN-based synthetic data generation into an automated MLOps pipeline [274] to compensate for the scarcity of defective-component data in fault detection. Although these methods can model complex data distributions, they are less straightforward to apply and require careful training generating realistic and diverse samples.

A particularly promising approach is Transfer Learning (TL), which enables the transfer of knowledge from a source domain, typically consisting of abundant and well-labeled data, to a target domain characterized by limited, imbalanced or unlabeled samples. This capability enables improved model generalization under challenging data conditions. For example, TL allows models training on a large public dataset (like CWRU) or simulated and laboratory-controlled data (source domain) to be fine-tuned for deployment in specific industrial environments with distinct operating conditions (target domain). This setting is often referred to as cross-domain fault diagnosis and conventional DL models may fail because shifts in data distributions. Parameter-based and feature-based TL using deep architectures have proven effective [275], as they can extract invariant feature representations across domains [276], [24].

3) *XAI in FDD:* Most FDD approaches, whether based on traditional ML or DL models, function as "black boxes," providing little insight into how fault-related decisions are made. For example, it is often unclear which process variables, signal features, or temporal patterns contribute to a given fault detection or classification outcome. While Physics-Informed Machine Learning can naturally enforce physically consistent reasoning and improve interpretability, it is not always feasible due to incomplete knowledge of system dynamics, unmodeled interactions, or complex process behaviors. In such cases, Explainable Artificial Intelligence techniques become essential, not only for validating model reasoning against domain knowledge (with *global* explanations), but also for supporting effective root-cause analysis, enhancing the actionability of diagnostic results, and enabling operators to make informed, trustworthy decisions (with *local* explanations), in line with the Industry 5.0 paradigm.

The majority of the FDD literature applies post-hoc XAI methods to explain pretrained models, rather than relying on inherently interpretable architectures. This preference arises from the typically superior predictive performance of complex, opaque models. Among post-hoc techniques, feature relevance scores (also referred to as feature importance) are the most common modality [277]. In FDD tasks, which are generally formulated as classification problems, these approaches explain the predicted class probability rather than the final discrete decision. The model-agnostic SHAP [278] is the most widely adopted method due to its solid theoretical foundation, ease of use, and ability to provide both global and local explanations [279], [280], [113], [281], [282], [283].

TABLE VI
SUMMARY OF FDD METHODS IN MANUFACTURING AND INDUSTRIAL IIOT.

Method	Description	Notes
Shallow ML	Feature engineering (time-, frequency- or time-frequency-domain) → traditional ML classifier (e.g., SVM, RF).	Suitable when data size is moderate and features are well understood. Works well for interpretable, low-latency and edge deployable FDD. Good for proof-of-concept with low compute cost. Less effective for high-dimensional or time-dependent signals. Relies heavily on domain expertise for feature extraction.
CNNs	Automatically learn local spatial or temporal features from raw or transformed sensor data (e.g., 1D vibration or 2D spectrograms).	Appropriate when local fault signatures dominate (e.g., impact transient) and sufficient labeled data are available. Less effective for long-range temporal patterns or strongly non-stationary processes.
RNNs/LSTMs/GRUs	Capture temporal dependencies and degradation dynamics.	Appropriate when fault behavior evolves over time (wear, drift, progressive failures) and sequential relationships are key. Handles long time windows better than CNNs but higher computational cost, but impractical for very long sequences.
CNNs+RNNs (Hybrid)	Combine convolutional feature extraction with recurrent temporal modeling.	Best suited when short-term and long-term signal patterns are relevant. Strong diagnostic accuracy. Demands larger, well-labeled dataset and more computational power.
Transformers	Model global temporal and cross-sensor dependencies using attention mechanism.	Preferred for high-dimensional, multi-sensor or multi-modal setup. Enable transfer learning. Requires large datasets, high compute and careful regularization.
GNNs	Represent sensors, process units as graph nodes with physical or statistical interconnections.	Capture structural and non-euclidean correlations that CNNs and RNNs ignore. Need proper graph definition (either learned or manually defined). Powerful for multi-machine or networked IIoT monitoring.
PIML	Integrate process physics, conservation laws, analytical models into learning to guide or constrain predictions.	Valuable when labeled fault data are scarce but physical understanding is strong. Physically consistent and explainable outcomes. Flexibility is limited, performance may drop for unmodeled failure modes.
RL	Learn optimal diagnostic policies via interaction with the environment or digital twin.	Require a reliable simulator/safe experimentation, training is resource-intensive.

LIME [115], which offers local explanations, is also used, though less frequently [155]. In [284], when detecting faults on a rotating electric motor testbed using RF, the authors state that LIME is especially beneficial for understanding the reasons behind single predictions, particularly with LIME rules provided for RF. Other model-agnostic, feature-importance-based techniques, such as Permutation Feature Importance and Partial Dependence Plots, are less frequently applied in recent studies [285], [277].

When dealing with particular types of classifiers, model-specific XAI methods can be applied to exploit intrinsic model properties and produce more faithful or computationally efficient explanations. For instance, ensembles of decision trees, such as RF, provide built-in feature importance measures, which have been used to provide explanations for rotating machines using multi-sensor fusion [155].

Gradient-weighted Class Activation Mapping [286] is a widely used visualization technique for CNNs that highlights the regions of the input most influential in a specific model

prediction. It is particularly popular in FDD, where it is employed to generate heatmaps over 2D images [287], as well as over 1D time-series data, both in the time domain [209], [288] and in the frequency domain [209], [289], [290], [291], [292], where it identifies the frequency bands most responsible for a particular decision. Building upon this, [293] proposed applying Gradient-weighted Class Activation Mapping (GradCAM) to CNNs trained on fuzzy-based energy pattern images rather than raw sensor signals, enhancing interpretability. The approach was validated on a robotic post-welding application. Many variations of GradCAM have been proposed. The study in [294] introduced a multi-layer GradCAM variant that considers gradient flow across multiple CNN layers to improve explainability in vibration-based fault diagnosis for industrial equipment. Another adaptation, specifically designed for HVAC fault diagnosis, was presented in [295]. In parallel, a few studies have explored Layer-Wise Relevance Propagations (LRPs) [296], a technique that

decomposes the model output backward through the network layers to assign a relevance score to each input feature. This allows identification of the parts of the input that contribute most to the decision. LRP has been applied, for example, in gearbox fault diagnosis with CNNs [297], in chemical processes [298] or HVAC [299]. Moreover, other studies have explored integrated gradients [300], [301]. Beyond standard XAI methods, novel explanation strategies for FDD have been proposed. For instance, [291] introduced a frequency-domain perturbation-based method to visualize and assess the decision boundaries of vibration-based classifiers. After selecting the most relevant frequencies using GradCAM, the authors perturb the power spectrum of a given sample around these dominant frequencies, varying bandwidth and intensity, to evaluate and visualize the robustness of the classifier. XAI, and particularly feature-importance analysis, is not only used for model interpretation but also for feature selection and model refinement. Studies such as [285], [298], [302], [301] demonstrate how interpretability can guide the identification of the most discriminative features, improving the performance of fault classifiers.

Even if other XAI modalities exist, such as counterfactual, example-based and concept-based explanations, their application remains very limited in FDD [277].

While post-hoc XAI methods dominates the FDD literature, an increasingly important direction focuses on models that are inherently interpretable by design. One branch of this research aims to improve interpretability by encouraging sparsity in the model parameters [254], [303], [304]. Beyond its interpretability benefit, sparsity is also advantageous in fault diagnosis for rotating machinery. As highlighted in [305], vibration signals from faulty components often exhibit sparse spectral signatures. Sparse representation therefore provide a natural way to isolate these fault-related patterns and can facilitate downstream tasks, including data generation [306].

Other work uses logical inference to obtain symbolic expressions that describe systems properties directly from data, an approach consistent with neurosymbolic AI. In this setting, the diagnostic task is formulated as the generation of expression in a formal language, for example temporal-logic formulas, and these expressions act as explicit classifiers. The interpretability follows immediately from the fact that the learned rules have a fixed syntax and well-defined semantics. Recent studies to interpretable FDD effectively use temporal-logic-based languages, learning the parameters of temporal operators [307], [308], [309], [310], [311]. To enhance deep learning models inherent interpretability for FDD, *algorithm unrolling* is an emerging alternative [305]. It transforms iterative optimization or signal-processing algorithms into neural network architectures. Each iteration of the original algorithm corresponds to a layer in the network and algorithm parameters are made trainable, allowing end-to-end optimization while retaining the interpretability of the original procedure. This strategy has been applied to sparse coding, where signals are represented as linear combinations of few dictionary elements. For example, convolutional sparse coding solved via iterative shrinkage-threshold algorithm (ISTA) was unrolled into CISTA-Net [312]. Extensions further incorporated multi-

layer sparse coding formulations that preserving theoretical transparency [313], [314].

Attention mechanisms are also considered to enhance interpretability. Some studies go beyond pure attention, for example in [315], a Variational Attention-based Transformer Network is proposed. Unlike standard transformers, whose attention weights may lack physical meaning, their model introduces probabilistic and sparsity constraints through variational inference with Dirichlet priors. This forces the attention mechanism to focus on a few fault-relevant impulsive segments, aligning model behavior with the physical characteristics of vibration signals, offering both high diagnostic accuracy and causally interpretable attention maps.

Finally, PIML, which is thoroughly discussed in Section III-A2, is another form of inherent model interpretability [305].

4) *Summary of FDD Methods:* Table VII provides a comprehensive summary of the FDD methods discussed in this section, organized by category and approach, along with their key references.

B. FDD Industrial Applications

In this section, the most common applications of FDD in the industrial domain are analyzed. We follow the same organization as in Section II-B, so that the reader can directly compare unsupervised anomaly detection approaches with supervised fault detection methods applied to the same use cases.

A non-exhaustive overview of specific papers, divided by application domain, is reported in Table VIII, while more general considerations are made in the subsections.

1) Mechanical/equipment faults:

Rotating Machinery: Rotating machinery is the most extensively studied application domain in the FDD literature for manufacturing systems. Commonly investigated fault types include bearing defects (e.g., inner-race, outer-race and rolling-element faults), shaft-related faults (such as rotor imbalance and misalignment) and gear/gearboxes faults (e.g., missing, broken or chipped teeth, surface wear, spalling and root cracks). Scenarios involving multiple simultaneous faults are also considered [318].

Vibration signals are by far the most commonly used data source, typically acquired via accelerometers and sampled at rates sufficient to capture characteristic fault frequencies [223] and suitable for early FD. Additional measurements, such as temperature and torque are sometimes used [24].

From a methodological perspective, the literature shows a clear evolution from feature-engineered pipelines, based on frequency- or time-frequency features [143], [148], toward deep learning models operating directly on time-frequency inputs. In this domain, CNNs are the dominant architecture [24], [271], [178], [175], even if hybrid LSTM-CNN models are also frequently investigated. More recent work explores newer architectures, such as TF [193], [202], often in combination with synthetic data generation strategies [272], attention mechanisms [164], [185], [168], [184], or graph-based models [135]. Due to its scale, this application is the

TABLE VII
SUMMARY OF FDD METHODS REVIEWED IN SECTION III-A, ORGANIZED BY CATEGORY AND APPROACH.

Category	Approach	Key Methods/Techniques	References
Traditional ML	Feature engineering	FFT, WT, EMD, STFT, time/frequency/time-frequency features	[24], [139], [140], [132], [141], [142]
	Classifiers	RF, XGBoost, LightGBM, SVM, k-NN, shallow NN, ensembles	[143], [144], [145], [146], [147], [148], [149], [150], [151], [152], [153], [154], [161]
	Feature selection	Recursive elimination, importance-based, correlation, optimization	[148], [156], [157], [145], [158]
	Hybrid DL-ML	CNN features + RF/SVM, DL extraction + shallow classifier	[159], [160], [156], [155], [162]
CNNs	1D-CNN	Time/frequency domain convolutions	[169], [170], [171], [167]
	2D-CNN	Image transformation, multivariate arrangement, time-frequency transforms	[172], [173], [174], [175], [176], [177], [178], [179], [164], [163]
	Attention-enhanced	Channel attention, spatial attention, self-attention	[180], [181], [168], [182], [164], [183], [184], [185]
RNNs	Enhanced variants	Bi-LSTM, QNN hybrids	[186]
	CNN-RNN hybrids	CNN + LSTM, parallel multi-sensor fusion	[162], [187], [188], [189], [190], [165], [191]
Transformers	Time domain	Raw signal transformers, CNN-TF hybrids	[192], [193], [194]
	Frequency domain	FFT-based, spectral encoding	[195], [196]
	Time-frequency	Wavelet transforms + ViT, synchrosqueezed WT	[197], [198], [199], [200]
GNNs	Lightweight/robust	Depthwise convolution, Bayesian attention	[201], [202], [203], [204], [316]
	Univariate graphs	Series-as-graph, series-as-node	[207], [208], [209], [210], [206]
	Multivariate/STGNN	Spatio-temporal graphs, inter-sensor modeling	[211], [212], [213], [214], [215], [216], [217], [218], [205]
	Adjacency learning GNN architectures Advanced architectures	Similarity-based, attention-based, causal GCN, GAT, GraphSAGE Hierarchical, supergraph, multiscale, uncertainty-aware	[207], [219], [210], [221], [222], [223], [220] [225], [224], [226] [227], [166], [228], [317]
LLMs	Primary detector	Knowledge graphs, fine-tuned on sensor data	[229], [230], [231], [232], [233], [234], [235]
	Supporting component	KG construction, human-in-the-loop	[236], [237], [238]
PIML	Data augmentation/features	Synthetic fault generation, physics-based features	[239], [240], [241], [242]
	Physics-informed regularization	PINNs, conservation laws, physics losses	[243], [244], [245], [246], [247], [248], [249]
	Physics-guided architectures	Wavelet kernels, modal layers, spectral weighting	[250], [251], [252], [253], [254], [255], [256], [257], [258], [166]
Deep RL	Direct classification	Q-learning, REINFORCE for FDD	[259], [260], [261], [262], [263]
	Model optimization	Neural architecture search, hyperparameter tuning	[264], [265]
	Imbalance handling	Adaptive sample weighting, multi-armed bandit	[266], [267], [268]
Data Challenges	Data augmentation	SMOTE, VAE, GAN synthesis	[269], [270], [271], [272]
	Transfer learning	Cross-domain adaptation, feature-based TL	[275], [276]

only one in which niche approaches like RL and LLMs have been studied [262], [231], [232].

Practical considerations. TL is often employed to mitigate the gap between models trained on simulated/lab-controlled datasets and deployed in real industrial plants [275], the gap across different machines or operating conditions. When computational resources or labeled data are limited, feature-engineered representations combined with traditional ML may be preferable. Conversely, when richer datasets are available, CNN-based models operating on spectrograms are typically adopted.

As highlighted in [24], FDD for rotating machinery presents several open challenges, including complex motion dynamics and frequent occurrence of multiple interacting faults, which complicate the identification of characteristic fault frequencies, external noise, non-stationary fault behaviors and strong

dependence of diagnostic performance on sensor placement.

Tool wear and breakage: Tool wear monitoring is a key task in manufacturing, as undetected wear can degrade machining quality. In milling, turning and drilling, tool wear monitoring is typically formulated as an estimation problem, with the objective of quantifying the tool's current state. The most widely used metric is *flank wear*, which directly reflects tool degradation.

ML-based monitoring is generally performed online using indirect sensor measurements, enabling data collection and estimation without stopping the machine. Cutting forces are the most used signals [319], followed by vibration and acoustic emissions [320].

The SVM and NN models dominate traditional approaches, using time, frequency and time-frequency features. SVM often shows superior performance even with respect to DL on small

sample sizes [149], [319]. However, especially in milling applications, CNN-based models, sometimes combined with LSTM, represent the current state of the art [321], [322], [323], [324], [325], [326], [327]. Emerging methods have been proposed to reduce the required number of training samples, such as GNNs-based one [323]

Hydraulic and pneumatic systems: Hydraulic and pneumatic systems are widely employed for actuation, clamping and material handling in manufacturing and process industries.

The most commonly investigated fault types include leakage, valve-related faults (e.g., sticking, clogging, spool wear, solenoid failures), pump and compressor faults (wear, cavitation, seal degradation), actuator faults (e.g., cylinder seal wear) and sensor malfunctions. In hydraulic subsystems, faults are frequently reported in pumps, valves, cylinders, and hydraulic motors, with axial and radial piston pumps being among the most extensively studied components.

A wide range of sensing modalities is employed. Pressure signals directly reflect energy losses, flow restrictions and system dynamics. Additional measurements include vibration, acoustic emission, flow rate, temperature and current, often collected in multi-sensor configurations to improve fault observability. However, unlike rotating machinery, vibration transmission in fluid power systems is strongly affected by fluid resistance and pressure losses, leading to more attenuated and indirect fault signatures.

From the methodological perspective, early studies rely on feature-engineered representations combined with traditional ML classifiers, suitable for small-sample or low-computational settings. An example is given in [328], where an SVM pipeline using variational mode decomposition and evolutionary optimization strategy is used for hydraulic pump diagnosis. Classical pipelines are also employed in more structured fault-detection workflows, such as the multi-step electromagnetic hydraulic valve group diagnosis method in [329], which combines time-frequency analysis, similarity-based health indicators, statistical change detection, and SVM-based fault classification.

However, also in this application, recent work adopt DL architectures. CNNs, RNNs and hybrid CNN-RNN models dominate the literature, often enhanced with attention mechanisms to improve feature discrimination under varying operating conditions. For example, [187] proposes a Bayesian-optimized CNN-LSTM architecture for hydro-turbine fault diagnosis using acoustic vibration signals, demonstrating improved robustness over standard deep models. Multi-channel and multi-modal fusion strategies are also widely investigated, aiming to exploit heterogeneous sensor data and different sampling frequencies. Representative examples include attention-based multi-sensor fusion networks [330], adaptive multichannel time-series fusion frameworks coupled with attention-integrated residual networks [331], and transformer-based models designed to capture multiscale interactions in hydraulic pump signals [332]. A distinctive trend in this domain is the increasing integration of data-driven learning with physics-based knowledge. Digital twin technologies are used to generate labeled fault data under controlled conditions, as in [333], where domain-adaptation Wasserstein GANs are employed to

bridge the distribution gap between simulated and real data before training a CNN-GRU-attention model for directional valve diagnosis. Physics-informed neural networks further extend this paradigm by embedding hydraulic principles directly into the learning process. Examples include multistage PINN frameworks for regulating valve fault detection in nuclear power plant hydraulic systems [244], and inverse PINN-based approaches that reconstruct pump flow ripple from pressure measurements to diagnose axial piston pump faults without direct flow sensing [249]. Transfer learning and hybrid residual-based strategies are also explored to address data scarcity and domain shift. For instance, [334] combines residual multi-head deep models with transfer learning for hydraulic pump diagnosis using vibration signals, improving generalization across operating conditions. Siamese neural networks have been proposed for piston pump wear detection under small-sample regimes, addressing the limited availability of labeled industrial fault data.

Electrical drives, induction motors: Electrical drives are another major application domain in ML-based FDD literature. Among them, induction motors and permanent magnet synchronous motors (PMSMs) are of particular interest.

Reported fault classes span electrical, mechanical, magnetic and drive-related categories. Electrical faults include stator winding short circuits, broken rotor bars and cracked end rings. Mechanical faults encompass bearing defects, air-gap eccentricity, rotor unbalance and shaft misalignment. Drive-related faults affecting auxiliary subsystems, such as cooling system failures, are also investigated [136]. In PMSMs, additional magnetic faults are considered, e.g., partial or irreversible demagnetization [335].

Current signals are among the most common data source for diagnosis electrical faults [336], [337], [338], [170], owing to their non-intrusive acquisition. Voltage measurements are sometimes combined with current signals to enhance model performance [161]. For faults with a dominant mechanical or thermal nature, vibration, temperature and acoustic emission signals are frequently employed, either individually [339], [340] or within multi-sensor configurations [341], [342]. To train a model to detect both electrical and mechanical faults, vibration signals can be used [343].

Methodologically, a considerable amount of the literature adopts feature-engineered pipelines based on spectral and time-frequency representations, including Fast Fourier Transform, Short-Time Fourier Transform, continuous and discrete wavelet transforms, empirical mode decomposition, Hilbert and Park transforms [344]. These representations are typically combined with classical machine learning classifiers such as support vector machines (SVM), k-nearest neighbors (kNN), random forests (RF), multilayer perceptrons (MLP), and tree-based ensemble methods [335], [148], [339]. In parallel, deep learning approaches have gained increasing attention. CNNs are the dominant architecture, often operating on transformed signal representations to enable automated feature extraction [179], [170]. For example, [337] employs bispectrum-based preprocessing of current signals as input to a pretrained CNN for PMSM stator winding fault diagnosis. Vibration-based CNN models have also been explored, including multi-

head 1D-CNN architectures for detecting multiple fault types using multi-directional accelerometer data [341]. Multi-input CNNs fusing vibration and acoustic features derived from time-frequency spectra have been proposed for induction motor diagnosis [342]. Attention-enhanced CNNs operating on continuous wavelet transform representations of vibration signals further improve discriminative capability under varying operating conditions [340]. Beyond standard CNNs, hybrid and advanced architectures have been investigated. Hierarchical CNN-SVM models have been used to jointly detect incipient faults and estimate fault severity from current signals [338]. Hybrid RNN-CNN with attention mechanisms have been proved effective to diagnose induction motor faults directly from raw voltage and current signals [345]. Graph-based deep learning has also emerged, which combines signal-to-image transformation with superpixel-based graph construction and graph attention networks to achieve robust induction motor fault diagnosis under steady and transient operating regimes [343].

2) *Process faults*: Process fault diagnosis concerns the monitoring of complex industrial processes in which failures do not originate from a single component, but emerge from deviations in system-level behavior. Typical application scenarios include chemical production units, semiconductor manufacturing lines, metallurgical plants, and large-scale continuous processes.

In chemical and process industries, faults often manifest as slow drifts in temperature, flow, or level measurements, abnormal reaction dynamics, or inconsistencies between correlated variables. These behaviors are difficult to detect with local thresholds and motivate predictive monitoring strategies. For example, multi-step fault prediction has been investigated in chemical processes subject to false-data injection and slow-developing faults, where temporal models are used to anticipate abnormal evolution before operational limits are violated [162].

In semiconductor manufacturing, an applicative example is the detection of abnormal wafer produced multi-stage fabrication processes such as thin-film chemical vapor deposition. Here, faults appear as unexpected changes in sensor readings, deviations in processing time steps, or abnormal temporal patterns across recipe phases. Data-driven fault diagnosis methods are applied directly to real fab data to flag defective wafers before downstream processing, reducing yield loss and rework costs [154]. Related studies on benchmark semiconductor datasets confirm that ensemble-based classifiers remain effective for this application, even when synthetic data augmentation is required to compensate for the rarity of faulty samples [346]. CNNs are also used, for example in [347], where an approach to detect faulty wafers is implemented via a multiple time-series CNN using internal equipment sensor variables (SVIDs), and in [282], which uses a CNN-autoencoder. Since these last two studies use real data and real faults, they are better described at the end of this subsection.

In heavy industries such as iron-making, process fault diagnosis is applied to large-scale systems like blast furnaces, where faults emerge under strongly nonlinear and non-stationary operating conditions. The application challenge here

lies in separating fault-related behavior from normal process variability caused by changing raw materials, loads, and operating regimes. Hybrid pipelines that remove non-stationary components before classification have been shown to improve fault detection reliability when applied to real blast furnace data [348]. In metal processing industries, FDD is also applied to rolling mills, where faults directly affect product thickness, surface quality and throughput. The paper [141] diagnoses faults in the Finishing Mill Area using a modified recurrent unit, using as input process variables and KPIs.

Accounting for interactions between sensor and process units is fundamental in industrial process FDD. In large interconnected facilities, such as three-phase flow installations, power systems and nuclear plants, faults propagate across subsystems. Graph-based modeling approaches have been adopted to explicitly represent dependencies among components [213], [214].

Interestingly, almost all the reported studies in this section use real world data collected under controlled laboratory conditions for dedicated test rigs, where faults are intentionally introduced and operating conditions carefully regulated. While this is fundamental to validate and benchmark models, it only partially reflects the complexity of real industrial environments. Nevertheless, there are two recent real-case studies that use actual operational data and naturally occurring faults for process FDD. The first, [144], addresses the problem of alarm floods in industrial plants, situations where a large number of alarm trigger simultaneously, overwhelming operators and delaying critical responses. Using real operational data from an Algerian gas plant (Hassi R'Mel), the authors develop and compare several supervised machine-learning models to classify alarms and predict whether a new alarm is true or false. Among all tested algorithms, Random Forest achieves the best performance, reaching the highest precision, recall, and F1-score, making it the most reliable model for industrial alarm management. Additionally, the study in [282] is based on real multivariate time-series data from an 8-inch semiconductor ashing tool, including genuine failure events associated with actual part replacements. Using this industrial dataset, the authors construct a framework that transforms sensor streams into correlation-based signature matrices and detects anomalies with a convolutional autoencoder. Residuals from the autoencoder are then used to classify which equipment part failed, and SHAP analysis identifies the sensor-pair correlations driving each prediction. The approach demonstrates strong performance in both failure detection and root-cause identification under real manufacturing conditions. Always in the semiconductor domain, [347] use real wafer-level data from a chemical vapor deposition process. The empirical study is based on actual production wafers exhibiting real process faults.

3) *Cyber attacks*: Cyber-attack detection for IIoT is a major application domain that can be framed within FDD. The detection is typically addressed through intrusion detection systems that monitor network traffic and system behavior to identify abnormal patterns. The literature considers a broad spectrum of attack types, including malware, denial-of-service, man-in-the-middle, phishing, supply-chain attacks, botnets and

false data injection [349].

The primary data sources are network- and system-level measurements, such as packet-level features, flow statistics, protocol attributes and SCADA communication variables, collected either from real IIoT deployments or realistic testbeds. These data are often high-dimensional and highly imbalanced.

While approaches based on traditional ML (e.g., on random forest and tree-based ensembles) still appear in the literature [350], [147], [145], state-of-the-art performance is consistently achieved by DL models. Recurrent architectures, including GRU, LSTM, Bi-LSTM, are widely adopted to capture temporal dependencies [351], [352]. Feedforward NNs and CNNs are also employed [353], [354]. Hybrid architectures that jointly model spatial and temporal patterns, such as CNN-GRU combinations, have proven especially effective for detecting complex attacks like botnets [349]. In [162], the authors propose a model that integrates a modified LSTM with multi-lookback inputs to enable multi-step autoregressive prediction, and couples it with a 1D-CNN, enabling the detection of false data injection attacks on networked industrial control systems. Hybrid and ensemble models are frequently explored to improve robustness and coverage of diverse attack patterns. Examples include multi-modal deep ensembles models, combining RNN, Bi-LSTM and Deep Belief Network (DBN) for cyber-physical systems [355], as well as architectures that integrate deep feature extractors with lightweight classifiers. A representative approach is the two-layer denoising autoencoder framework in [356], where successive layers suppress neutral features and extract discriminative representations, which are then classified using an SVM. For SCADA-based IIoT networks, hybrid models combining advanced recurrent units with decision-tree aggregation have been proposed [357].

TL is sometimes adopted to mitigate the scarcity of labeled attack data and to improve generalization, as demonstrated by the ResNet-based transfer-learning intrusion detector in [353]. Data imbalance is typically addressed through augmentation strategies such as SMOTE [351], while there exists also recent works that explore generative diffusion models to synthesize realistic attack traffic [352]. Since modern attacks often affect multiple system layers, while most existing approaches rely on single-modality time-series data, the design of robust multi-modal models and effective fusion strategies remains an open challenge. This issue is further compounded by common limitations such as the scarcity of publicly available datasets, limited explainability of AI-based methods, increasing system complexity and scalability requirements and the need for privacy-preserving solutions [358].

4) *Human-machine interaction*: Human-machine collaboration is a pillar of Industry 5.0. Although human-related errors cannot be strictly classified as faults in the traditional sense, they can still be framed within the FDD as deviations from expected operational behavior. Consequently, recent works investigate data-driven methods to detect human-related anomalies, primarily relying on visual input such as images or video streams. To give some examples, [359] introduces an end-to-end spatio-temporal learning framework for real-time monitoring of assembly operations. The method predicts future human actions from video using a convolu-

tional encoder and ConvLSTM-based sequence modeling and detects human faults such as unfixed parts, incorrect assembly order and missing components, by comparing predicted and observed action sequences. In [360], a two-stage deep learning pipeline is introduced to identify unsafe ergonomic behaviors. Monocular video is first used to reconstruct 3D human body poses via the VIBE architecture; the resulting decimated body meshes are then classified using MeshCNN to recognize ergonomically hazardous actions across diverse industrial tasks. Finally, [361] presents a multimodal framework for safety monitoring in human-robot collaboration. The approach combines scene-graph-based visual understanding with a BERT-Bi-LSTM-conditional random field model to interpret scene descriptions and verify compliance with robot-specific personal protective equipment (PPE) requirements, enabling the detection of missing or improperly used PPE.

5) *HVAC faults*: HVAC systems are important in manufacturing plants for maintaining thermal and humidity conditions critical to processes such as semiconductor fabrication, precision machining, pharmaceuticals and clean-room-assembly. Faults can affect both energy efficiency and process reliability, with typical issues including sensor drift, heat exchanger fouling, refrigerant leaks, valve or damper malfunctions, chiller degradation and air-handling-unit (AHU) control faults. HVAC FDD relies on operational data (temperatures, pressures, flows, power and control signals). A key challenge is separating true fault symptoms from variations caused by production schedules or environmental conditions. Causal discovery and inference methods have been shown to identify real fault drivers and build interpretable models with high accuracy and low computational cost [362]. Chiller faults are particularly complex due to thermal coupling and compound failure modes. Knowledge-embedded models that integrate expert rules into data-driven networks (e.g., deep belief networks) improve generalization and enable detection of multi-fault scenarios, validated on real experimental data and deployed in cloud-based monitoring platforms [363]. Data scarcity is another practical challenge: some studies use physically induced faults under controlled conditions to select informative features for chiller diagnostics [158], while others generate synthetic fault samples with GANs and reinforcement learning to enhance AHU fault detection under extreme data limitations [268]. For a comprehensive overview of methods and applications, see [364].

IV. PREDICTIVE MAINTENANCE

This section is devoted to introducing the field of PdM which differs from the previously presented AD (Section II) and FDD (Section III) because of its prognostic nature. In fact, PdM is a component of the broader field of Prognostics and Health Management (PHM) which encompasses the combination of several technologies including advanced monitoring, diagnostics, prognostics and prognostics management. These are characterized by the common goal of minimizing the occurrences of failures or downtime in industrial manufacturing processes [367].

TABLE VIII
REVIEW OF INDUSTRIAL APPLICATIONS OF ML-BASED FDD

Fault Type	Category	References
Mechanical/ Equipment	Rotating machinery (bearings, shafts, gears, imbalance).	ML: [143], [148], [152], [153], [150], [157], [151], [146], DL: [271], [272], [135], [365], [172], [175], [176], [178], [177], [169], [171], [163], [180], [164], [185], [181], [168], [183], [182], [186], [197], [193], [202], [195], [223], [258], Hybrid: [160], [159], [366], [184], [191].
	Tool wear and breakage (cutting tools, milling, drilling).	ML: [319], [149], [320], DL: [326], [321], [322], [323], [325].
	Hydraulic and pneumatic.	ML: [328], [329], DL: [187], [330], [331], [332], [333], [244], [249], [334].
	Induction motor, electrical drive faults (stator winding, rotor bars).	ML: [344], [335], [148], [339], DL: [179], [170], [337], [341], [342], [340], [338], [343], Hybrid: [345].
Process	Three-phase flow facility.	DL: [213], [317], [214].
	Chemical plants.	DL: [162].
	Nuclear power systems.	DL: [214].
	Metallurgic process.	DL: [348], [141].
	Semiconductor manufacturing.	ML: [154], DL: [347], [282].
	Gas plant.	ML: [144].
Cyber attacks	Product quality.	LLM: [229].
	Intrusion, denial-of-service, man-in-the-middle, supply-chain attacks, botnets and false data injection.	ML: [147], [145], [350], DL: [351], [352], [353], [354], [349], [162], [355], [356], [357], [353], [351], [352].
Human-machine interaction	Operator error/unsafe action detection.	DL: [359], [360], [361].
HVAC	HVAC faults.	ML: [362], [158], DL: [363], RL: [268].

A. PdM Theoretical Foundations

In this section we start by reporting the definitions of some basics concepts in the field of PHM, PdM and RUL estimation which will be used throughout the rest of the survey.

Definition 3. Predictive Maintenance (also referred to as **Condition-based Maintenance**) is an advanced maintenance strategy which employs data analytics, ML and IoT technologies to predict equipment failure before their occurrence with the aim of reducing maintenance costs and minimize unexpected breaks and downtime in industrial applications.

The pipeline of maintenance policies adopted in the industrial field is usually divided into five main steps [368]:

- **Remote Monitoring:** This step involves monitoring the equipment under analysis by embedding sensors or cameras to enable real-time data collection regarding its health state. The acquired data may than be transmitted on site or to external devices through cloud infrastructures. Nowadays, this step is handled by IIoT infrastructures which are becoming more and more advanced and able to handle huge amounts of data produced by sensors, more details of which will be presented in IV-D3. Monitoring data represents the main resource for the predictive models used in later stages of the pipeline. Some pre-processing steps (i.e. data cleaning, data normalization and handling of missing or corrupted data) are traditionally executed before feeding the data to the models for the prediction.
- **Failure Prediction:** Pre-processed data are analysed through ML or DL based models to produce predictions. In this stage we may be interested both in prognostics predictions (i.e. when the asset is most likely the fail) or in diagnostic ones (i.e. identify the most vulnerable

components of the system). This achieved through RUL or AD estimation, respectively. This highlights the strong interconnection between these two fields.

- **Task Scheduling:** Based on the predictions produced by prognostic systems and the domain knowledge of experts, maintenance operations are scheduled in order to prevent the equipment from failing. In order to solve this complex optimization problem Reinforcement Learning based approaches have recently been proposed, more details of which will be provided in IV-B4.
- **Communication and Integration:** To translate the decisions made in the previous steps into actionable measures, communicating with maintenance personell is crucial. This step is traditionally carried out through the usage of Enterprise Resource Planning (ERP) and Asset Management software tools.
- **User Interface and Reporting:** The User Interface should guarantee a clear understanding of the model decision to the machine operators. This is normally carried out through interactive dashboards, reports and alerts. In this context the trustworthiness of the predictions produced by the predictive models, achieved through XAI (presented in IV-C), is also a required guarantee.

PdM is employed in the second stage of the maintenance pipeline (i.e. *Failure Prediction*) with the aim of outperforming two other traditional prognostic approaches, typically referred to as **Reactive Maintenance** and **Preventive Maintenance**.

Definition 4. Maintenance interventions based on **Reactive Maintenance** (Run-to-Failure) wait for the machine failure to be performed. While this strategy fully exploits a machine's operational lifespan, it incurs significant repair costs and risks, such as cascading failures (i.e. subsequent failures in dependent components or interconnected systems). As the

name suggests, this strategy represents a reactive approach with no way of controlling or monitoring the system's health state during its lifespan.

Definition 5. Preventive Maintenance occurs at fixed time intervals, decided based on historical failure times of the machine. Although more systematic and proactive than Run-to-Failure approaches, this strategy often results in premature maintenance cycles, underutilization of component lifespan, inefficient resource allocation, and excessive spare part consumption.

To address the downsides of both the reactive and preventive maintenance approaches, in PdM maintenance is dynamically scheduled based on predicted time-to-failure estimates, enabling near-optimal utilization of component lifespan while mitigating failure risks. Differently from Preventive Maintenance, where the time interval between two consecutive maintenance interventions is decided a priori based on the time-to-failure experienced in the past, in PdM the maintenance time is determined based on the current condition of the machine which is monitored in real-time.

In order to ensure a continuous monitoring of the equipment's health state sensor networks, often incorporating signal processing techniques, are employed. Several benefits arise from this approach including the prevention of catastrophic failures, identification of undetected failure modes, minimization of downtime, reduction of maintenance costs, optimized maintenance scheduling and also valuable insights for equipment design improvements.

This survey will mainly focus on the RUL estimation task, an alternative formulation of the PdM task, defined as follows:

Definition 6. Remaining Useful Life is defined as the amount of time remaining in the productive life of a manufacturing equipment. The RUL is typically represented as a time varying signal characterized by a monotonically decreasing trend. It can be expressed in terms of number of cycles, seconds, minutes or hours, depending on the specific application.

PdM gained popularity thanks to the rapid development of IoT-based systems and it's one of the main components of Industry 4.0. In fact, sensors composing IIoT infrastructures represent a fast and reliable approach to collect data (i.e. vibration, temperature, humidity) strictly related to the equipment's health which are used to design and train the predictive models used for the estimation of the RUL signal.

In particular, the huge amount of data collectable through IoT sensors is the main source of information for ML and DL based algorithms. Differently from older approaches where the machine's life degradation is modelled exactly through complex physical or mathematical equations, these approach base their predictions just on the relationship found between the input sensor readings and the RUL signal. This provides a more computationally efficient method to solve this complex task.

This is the main feature of data-driven PdM, introduced in IV-B2, which will be the main focus of the PdM survey.

The application of PdM technologies has brought several benefits in industry. According to [369], the integration of

PdM strategies in manufacturing plants led to a 20% reduction in unplanned downtime and decrease in maintenance cost of about 15%.

By predicting the RUL the objective is to perform maintenance interventions at the right moment: not too late (as it happens in reactive maintenance) or too early as in preventive maintenance. In this way both costly machine repairment and long downtime intervals are avoided.

B. PdM Methods

This section is devoted to the presentation of the current state of the art in the field of PdM. The operational efficacy of PdM systems fundamentally depends on accurate RUL estimation, considered as the most complex yet most beneficial PHM task [370]. This problem serves as the foundation for maintenance decision-making processes and will be the main focus of this section.

PdM has become the predominant strategy in the field of prognostics and health monitoring in industrial applications in recent years [371], [372]. PdM is applied in several industrial fields, such as: turbofan engines [373], wind turbines [374], rotating machinery [375], high-speed trains [376], aviation systems [377], railway infrastructures [378], [379], [380], pharmaceuticals production [381], lithium-ion batteries [382], vehicle lifts [383] and also operating systems [384]. Moreover PdM can also be integrated as one of the services offered by DT technologies [371], [385], predominant in the field of Industry 4.0 and 5.0.

The RUL estimation research field is very broad and comprises several different industrial domains which will be covered in IV-B, IV-D. For the sake of clarity the different branches of RUL-based PdM are graphically represented in Figure 1.

Three main macro areas of PdM estimation can be extrapolated from 1: *model-based*, *data-driven* and *hybrid* RUL estimation whose main characteristics and differences are highlighted in IV-B1, IV-B2 and IV-B3, respectively.

The methods presented in this section and in IV-D are grouped in Tables IX and X divided in hybrid and data-driven methods. The few model-based approaches presented in IV-B1 are not considered in this comparison because they are not the main target of the survey which focuses on the usage of AI-enhanced models for the automation of prognostic decisions.

1) *Model based RUL Estimation*: Model based methodologies for RUL estimation exploit prior physical and mechanical knowledge on the equipment with the aim to design a physical model of the degradation process which represents how the machine transitions from an initial state of full health to the end of its life (i.e. machine breakage). The main advantage of these methods is that if an accurate model of the machine is built it will fully characterize the degradation process and provide accurate estimates of the RUL. However crafting these kind of models is highly challenging because of the complexity of the physical and mechanical processes typically involved in real world equipments. Physical processes are in fact well understood just for critical or very simple components [490], [491], [492], [493].

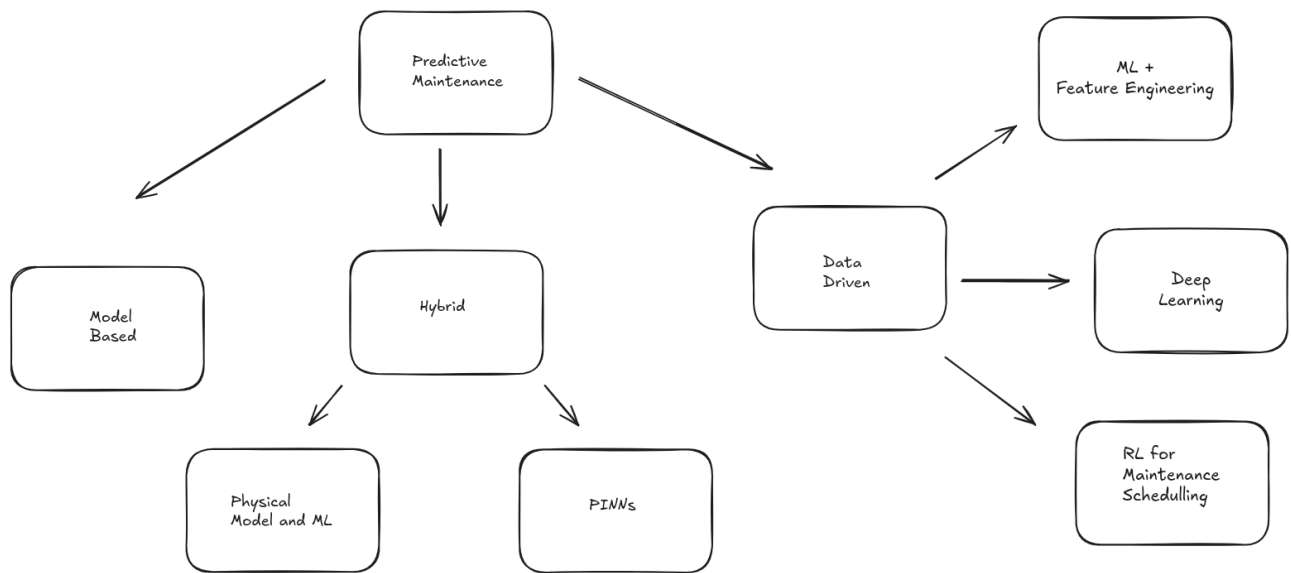


Fig. 1. The scheme above subdivides the topic of PdM into different sub topics which will be covered in the next sections

TABLE IX
HYBRID PDM APPROACHES

Category	Techniques	References
Hybrid PdM Approaches	RUL estimation approaches where both model-based and data-driven approaches are used	[386], [387], [388], [389], [390]
PINN	PINN models embed physical principles as a regularization term in the loss function	[391], [392], [393], [394], [395]

Additionally these models are usually equipment specific and may not be easily transferable to completely different or similar machines working with different operating or manufacturing conditions. For these reasons, the employment of these methodologies is becoming less and less common in the literature in favor of the more flexible and general purpose data-driven methodologies.

2) *Data driven RUL Estimation:* Data driven methods do not require domain knowledge of the system under analysis but will instead rely on the availability of historical data, usually in the form of raw sensor measurements, acquired over the entire life cycle of the machine. These methods raised in popularity thanks to the developments of AI related technologies (i.e. ML and DL) and IoT networks for data collection. These models are capable of estimating the degradation process of the machine directly from the historical data and, based on that, will provide estimates of the RUL.

Once enough historical data are available, accurate RUL predictions can be obtained without the need of designing overly complicated physical models. In particular, models are usually designed in order to easily generalize to assets similar to the ones producing the training data. AI powered models may be adapted to work on different equipments through the employment of transfer learning [494], [12] or domain adaptation [445] techniques. In general, the machine degradation processes estimated by such models are less accurate with respect to the one produced by a model based approach. In

fact, the outcomes generated by these models are not based on a physical model of the system but rather on statistical patterns identified in the historical data. However, even if not perfect, these estimates are in general good enough to trigger maintenance operations at the right time and come at a significantly lower cost in terms of computational and human resources. As a result, nowadays data driven models are overtaking model based approaches in the industrial field [402]. For this reason, data driven models will be the main focus of this review.

With the rapid development of data-driven prognostics and the high amount of data required the issue of data privacy arises. As a consequence in the literature this challenge is addressed through Federated Learning (FL), a paradigm in which the privacy of user's data is ensured by the fact that each user trains the model locally with private data and shares with other users just the model's updates. Examples of FL based RUL estimation are [495], [496], [479].

RUL Estimation using Feature Extraction: The main source of information for data driven models employed in the task of RUL estimation are the raw data produced by sensors monitoring key physical quantities of the equipment under analysis (i.e. temperature, flow, pressure, acceleration, humidity, ...). These signals are typically distorted by noise and may interfere with other signals. This means that the signs of the machine's useful life degradation, which are the key component for a precise estimate of the RUL, are often not directly observable in the raw data. Consequently, some

TABLE X
DATA-DRIVEN PDM APPROACHES

Category	Techniques	References
Data Driven PDM Approaches	Generic RUL estimation approaches where the RUL signal is estimated through data (i.e. sensor readings) rather than with the usage of complex physical model of the equipment	[396], [397], [398], [399], [400], [401], [402], [140], [403], [404], [405], [406], [407], [405], [408], [409], [410], [411], [412], [413], [414], [415], [416], [417]
Feature Extraction + ML	These methods exploit ML algorithms for the RUL estimation, thus an initial feature extraction step is required to handcraft the features needed by the model to accurately estimate the equipment's end of life.	[396], [397], [398], [399], [400], [401], [402], [140], [403], [404], [405], [406], [407], [405], [408], [409], [410], [411], [412], [413], [414], [415], [416], [417]
Lithium-ion batteries	RUL estimation is a very important task in the context of lithium-ion batteries whose degradation process can lead to catastrophic consequences as unpredictable burn of electric vehicles.	[418], [419], [406], [396], [391], [382], [420], [421], [422], [423]
Turbofan engines C-MAPSS	The C-MAPSS dataset is the most commonly used benchmark dataset in the context of RUL estimation. The sequential nature of the data promotes the usage of sequential models such as LSTM, CNN and combinations of those.	[370], [424], [425], [426], [427], [428], [429], [430], [431], [432], [433], [434], [399], [435], [436], [437], [438], [439], [440]
Turbofan engines N-CMAPSS Bearings	N-CMAPSS is a more complex and recent version of C-MAPSS. The approaches adopted on this benchmark are similar to the ones used on its predecessor. Rolling element bearings are one of the most common reasons of machine breakdown in the industry. The periodical nature of the data enables the definition of notable fault-frequencies that facilitate the detection of faults in the context of FDD and inherent interpretability in PdM.	[441], [12], [442], [443], [444], [433], [445]
Transformer	Attention based models are widely used in the field of NLP but have also extensive applications in time series modelling. In particular, in the context of RUL prediction the inclusion of attention modules in the model architecture facilitates comprehension over the root causes of the machine degradation.	[446], [403], [447], [448], [374], [400], [449]
Uncertainty	UQ is a fundamental requirement in PdM considering the elevated economical costs associated with underestimation or overestimation of the target RUL signal. The most common approaches are based on BNN and Quantile Regression.	[450], [451], [452], [433], [453], [444], [454], [455], [445], [456]
XAI	UQ is a fundamental requirement in PdM considering the elevated economical costs associated with underestimation or overestimation of the target RUL signal. The most common approaches are based on BNN and Quantile Regression.	[457], [395], [458], [459], [460], [461], [462], [463], [441], [464], [465], [466], [467]
IoT	In the context of the fifth industrial revolution, where the role of the human being is gaining more and more relevance, being able to provide insights on the model decisions is of paramount importance. Traditionally XAI approaches exploit variants of the SHAP explanation algorithm, which may also be included on the AD module of a complete PdM pipeline that performs both diagnostic and prognostic activities.	[468], [469], [470], [433], [471], [446], [472], [473], [474], [475]
Industrial Case Studies	IoT infrastructures are one of the key elements of the latest industrial revolution and their interconnection with health monitoring and AI is considered in order to enhance performance of PdM based applications. In these settings usually cloud computing infrastructures are employed to better store and manage the data collected from IoT sensors which are later used by ML or DL algorithms to perform end of life predictions.	[476], [477], [478], [479], [376], [480], [481]
	In this category several case studies of PdM applications in the context of Industry 4.0 and 5.0 are presented. One of the adaptations employed to embed RUL prediction models in high-stakes industrial contexts is the usage of business metrics that shape the model's predictions to account for the business costs associated with imprecise estimations, preferring a pessimistic approach rather than an optimistic one.	[6], [482], [483], [484], [485], [486], [487], [488], [489]

pre processing steps are required before feeding the sensor measurement to the ML models. These involve the creation of handcrafted features, extracted through transformations directly applied on the raw data. The resulting variables are often referred to as Health Indexes (HI) or State of Health (SOH) [396] since they provide a complete representation of the health state of the monitored equipment throughout its entire lifespan. This pre processing step is often denominated as **Feature Engineering**.

Industrial equipments are characterized by a monotonic deterioration of their properties (i.e. without sudden regenerations). In order to effectively capture their health condition over time, features showing a monotonic trend should be extracted [397]. Coble et al. defined in [398] a set of measures to score the importance of a given feature as a prognostic parameter: *Monotonicity*, *Prognosability* and *Trendability*.

These are employed in [399], [400] where they are merged together into a weighted sum to measure the relevance of sensor measurements for the RUL estimation task, and in [401] where *Monotonicity* is used as a feature selection criteria. An interesting study is the one reported in [497] where SHAP values are computed on models employed on the C-MAPSS dataset showing an high correlation with these three prognostics measures.

Typical features extracted in this stage are presented in III-A2.

In addition to the construction of handcrafted features other typically used methods are: dimensionality reduction (i.e. PCA) ([443], [403], [404]), EMD [396], clustering to discover additional features, noise filtering (i.e. Gaussian, Kalman or Exponential filters) or even physics-inspired feature transformations [406].

The extracted features or HI are successively used as inputs to the ML models: Regression models trained to directly predict the RUL values. An alternative approach consists in defining a threshold on the values of the HI to trigger maintenance operations [403], [498] once that value is reached.

There are also unsupervised methods of HI estimation used where no predefined degradation process information on the equipment are given but are automatically extracted through signal preprocessing or signal entropy [407]. Also DL models can be exploited to extract hidden features from sensor measurements, in particular AE and VAE based architectures [403], [405], [408], [409].

Several works in the literature exploit HI for their RUL prediction using different strategies to extract the equipment State of Health (SOH): graph modelling [410], weighted correlation of known fault frequencies over cycles in bearings [411], a quasi-orthogonal sparse projection algorithm [412] to trace different degradation trends for different fault types, information entropy [413], frequency energy shift exploiting the signal produced by a microphone sensor [414], CWT [415] and Gaussian Process (GP) Regression for SOH estimation [416].

Another example of the usage of feature extraction for RUL estimation is represented by sensor fusion based approaches [404], [417].

In recent times Feature Engineering based methods became less popular because of the development of DL algorithms. Deep models are known for stacking multiple hidden layers sequentially building a hierarchical representation of the features embedded in the data. This makes it possible to directly extract key degradation features from raw data without the need of an handcrafting feature extraction process as the ones explained above.

Moreover relying on handcrafted features makes it challenging to adapt to different prognostic applications [435]. However, the combination of feature engineering with ML based prediction model can still be considered as a valid alternative in case end-to-end DL solutions become too computationally complex for the implementation on a real-world IoT based industrial system.

A peculiar case is represented by [482], [453], [443] where handcrafted features are combined to raw data using feature fusion to enrich the training data used to fit a deep learning model combining feature engineering with the most recent DL approaches.

RUL Estimation for Turbofan Engines: As mentioned in paragraph I-A the C-MAPSS dataset is very commonly used to evaluate the performances of RUL estimation methods, as highlighted in [370], [424]. The main challenge posed by this dataset are the different operating conditions and degradation patterns of the turbofan engines due to manufacturing variations. In order to solve these issues the majority of the works in the literature employs DL algorithms. In particular, sequence learning DL architectures are employed [425], [426], [427], [428], [429]. A common example is the combination of the LSTM [430], [431], [432], [433] and CNN [499] architectures. In fact, CNNs are good in merging different sensor dimensions such as vibration and acoustic data. On the other hand, LSTM networks are specialized in capturing sequential dependencies

in time-series data. As a consequence, the combination of these two architectures is usually the go-to solution for a task like RUL estimation where input data are multi-sensor time-series in the majority of the cases. In some cases the addition of an attention mechanisms, to weight the importance of different sensors and time stamps, is also considered a valid solution.

In [434] and [399] the machine degradation process is not assumed to start from the beginning of the production phase but after the first incipient fault. Before this moment the machine is considered to be in a full health state, thus the RUL signal is constant at the initial value and the degradation towards failure begins only after the incipient fault. This model for RUL degradation is typically named **piecewise linear degradation** model while the simpler approach of considering the RUL as a monotonically decreasing signal is called **linear degradation** model. The piecewise linear degradation model represents a more realistic scenario but adds an additional step in the estimation pipeline: the detection of the point in which the machine transitions from a full health to a degradation state. This is achieved in [434] by using a pair of CNNs architectures: the first one to estimate the incipient fault and the second one devoted to the RUL estimation. Moreover, a reliability index is also used in order to distinguish the RUL of two different components. Another application of the piecewise linear degradation model is presented in [399] where the so called *elbow point* (i.e. the point when degradation is first observable) is detected through a Z-test and then only sensor data from the *elbow point* onwards are employed for prognostics, which is realized through an Echo State Network (ESN) optimized through a Differential Evolution (DE) optimization algorithm.

In [435] the capacity of DBN to extract hierarchical features is employed as an alternative to the commonly used feature engineering combined with ML models approaches. To increase model efficiency an ensemble of DBN is considered and it is trained through a multi-objective evolutionary algorithm needed to tune the optimal hyperparameters of the base learners (i.e. number of hidden neurons in each hidden layer). Another example of ensemble learning is employed in [436] where Stacking Ensemble Learning is used to exploit the predictions of different base learners (i.e. RF, Decision Tree, Gradient Boosting) and combine them through a meta-learner (i.e. Linear Regression).

To take into account for the different working conditions characterizing the C-MAPSS dataset a Mixture of Encoders based model is presented in [437] in order to specialize and Autoencoder model on the specific features of each working condition. Finally, features extracted in this way are fed in input to a Transformer to obtain the final RUL prediction.

Together with CNNs and LSTM based architectures another effective approach to learn the spatio-temporal relationships underlying the RUL prediction is to employ TCN. TCN is an optimized version of 1DCNN to model both long and short temporal dependencies through the usage of dilated convolutions, causal convolutions (i.e. the output of layer t depends solely on the convolutional outputs at layers $t-1, t-2, \dots$) and residual connections. In [438] TCN layers are used together with LSTM layers and a data augmentation strategy

is employed to enrich the training and testing trajectories contained in C-MAPSS. Other works employing TCN for the feature extraction part are [439], [440].

Another commonly used dataset to evaluate PdM models in the field of turbofan engines is the N-CMAPSS dataset and several works have used it as the benchmark dataset: [441], [12], [442], [443], [444], [433], [445].

RUL Estimation employing Transformer based architectures: The Transformer architecture, introduced in [500], has had a lot of success in NLP but several applications and adaptations of attention-based models on time series data have also been proposed, as an alternative to the classic CNN and LSTM architectures, presented in IV-B2. These models employ the attention mechanism to discover correlations between different time stamps in order to identify the most important ones. Some of these Transformer's variants have also been used in prognostics: [450] uses a Temporal Fusion Transformer (TFT) [451] and [452] a Temporal Flow Transformer which employs ProbSparse self-attention, a modification of the traditional self-attention mechanism tailored for time series processing, introduced in [501].

The employment of the attention mechanism, which assigns weights to the time stamps composing the input sequence based on their importance, provides also some explainability since it focuses on the most important features composing the input sequence [433], [453].

It also helps in the feature extraction process by identifying the most relevant spatio-temporal features for the task. As an example in [373] a Dual-Attention mechanism is used to enhance RUL prediction: the combination of CNN with a Self Attention layer is used to extract the most relevant spatial features while the temporal ones are retrieved through the combination between a LSTM and a Self Attention layer. In [444] a dual channel is used to separately treat sensor data and time-varying operating conditions which are firstly modelled through a 1DCNN to extract spatial features that successively are fed into two parallel LSTM layers to generate temporal features which are passed through a soft attention layer to filter out invalid information. Finally, the RUL prediction is produced through a Regression Head represented by two Fully Connected (FC) layers.

Attention based feature extraction is also employed in [454] together with a gated convolutional unit specialized in discovering local features. The combination of CNNs and Transformer architectures is also used in [455] where a Convolutional Transformer model is proposed.

In [445] the focus is not solely on the RUL estimation task but also on the Domain Adaptation one which has a lot of beneficial aspects in prognostics: for example, it can achieve transferability of RUL prediction capabilities across different machines. The main contribution brought by this work is the focus on the adaptation not only at the feature level but also at the semantic one. In particular, an adversarial approach is employed for the Domain Adaptation task while the Transformer Encoder is used as a feature extractor. In order to consider both feature-level and semantic-level adaptation two discriminator networks are used: feature-level source identification is obtained from the extracted features

while semantic-level identification is done based on the output RUL curves. Essentially the semantic discriminator acts as a regularization term during training, punishing inaccurate and unreasonable outputs that can be produced in the target domain if Domain Adaptation is only applied at the feature level.

The Transformer architecture was originally introduced to solve the machine translation task in the field of NLP and was later applied to other common NLP tasks such as sentence and emotion classification with the Deep Bidirectional Transformers for Language Understanding (BERT) [502] model. In [456] this model is adapted to bearing RUL estimation. Time features are extracted from the raw vibration signals and passed as features blocks to the BERT input layers. The training strategy involves a pre-training phase on a large dataset in order to equip the model with a general knowledge on the RUL estimation task and then fine-tuning is used to specialize the model on a specific scenario.

Uncertainty Quantification in RUL estimation: One of the main advantages of PdM is the reduction of downtime and machine breakages costs, guaranteed by a precise estimation of an equipment's health state and RUL. In this context it is of paramount importance to take into account the uncertainty in the model predictions which can cause an *unexpected break* if too optimistic (i.e. RUL overestimation) or *unexploited lifetime* conditions if too cautious (i.e. RUL underestimation). This estimate is successively used to guide maintenance operators in interventions making uncertainty quantification equally important as prediction accuracy [395]. From an economical perspective designing a more pessimistic model is the preferable choice, although in specific cases when an intense use of the machine is required (e.g. in agriculture where the time to collect crops is limited and highly subjected to weather conditions) an aggressive use of the machine may instead be considered. For this reason, another crucial aspect to take into account is not only the point estimation of the RUL value but also the uncertainty associated with that. UQ is a very important task to produce trustworthy models and it is widely adopted in DL [458] and, consequently, in PdM and RUL [457] estimation.

A natural solution to introduce uncertainty estimation in RUL prediction, or in any other regression task, is the usage of BNN, a class of NN with weights modelled by a probability distribution that naturally provide confidence intervals in their estimates. This approach is applied to different data-driven architectures: DBN [435], traditional NN [459], RNN [460], GRU [390], PINN [395] and CNN [461]. Traditionally training BNN is achieved through Bayesian Variational Inference [459] or Markov Chain Monte Carlo (MCMC) derived methods such as Hamiltonian Monte Carlo (HMC), employed in [463]. [462] applies BNN to RUL prediction and uses the Stein Variational Gradient method for the efficient training of such models. Moreover in [462], [463] the the RUL prediction is enhanced by its uncertainty which works as a corrective factor in order to reduce it in case a too high value is produced. Moreover BNN models are inherently regularized and this provides a natural improvement of the predictive accuracy with the respect to their non-bayesian counterparts. [441], instead, introduces a Bayesian Gated-Transformer including a

customized encoder block with gated connections to improve feature extraction capabilities and both epistemic and aleatoric uncertainty provided by the probabilistic nature of the model weights and the noise channel, respectively.

In [464] a GNN model is used to perform RUL prediction on a complex mechanical equipment. In particular, GCN and GAT modules are used for spatiotemporal feature extraction which is succeeded by a Bidirectional Long Short Term Memory (Bi-LSTM) module to also produce temporal features. Finally Maximum Likelihood Estimation (MLE) is incorporated in the loss function to model the output space as a Gaussian distribution and enrich the model with an estimation of the uncertainty on the RUL prediction, which is a fundamental ingredient in prognostics to improve system reliability and operational efficiency.

Another approach for uncertainty estimation is Quantile Regression where different quantiles of the posterior predictive distribution are estimated. Applications of Quantile Regression in the field of RUL estimation can be found in [465], [466]. In [465] a State Space Model (SSM) architecture is proposed for RUL prediction and it is enhanced by the Simultaneous Quantile Regression (SQR) [503] framework for both training and evaluation. On the other hand, [466] uses eXtreme Gradient Boosting (XGBoost), a non DL architecture, for feature selection and extraction and Multi-Quantile Recursive Neural Networks (MQ-RNN) [467] for a Quantile Regression based RUL estimation. The advantage of both SQR and MQ-RNN is that, being used for multi-quantile regression there are no prior requirements on the output distribution.

3) *Hybrid Approaches for RUL Estimation:* Data-driven models are nowadays the standard for RUL estimation but they still have some limitations. First of all, in the field of PdM it's difficult to collect high-quality data because of the rarity of faults and of the unwillingness to reproduce faults on purpose just for the sake of training such models, as explained in section I-A. Additionally, environmental conditions may have a huge impact on the physics of the system and it's difficult to collect data representing all the possible environmental conditions encounterable. A possible solution is represented by data augmentation techniques [504], [438] but still researchers and practitioners have to deal with model complexity which is strictly correlated to the predictive accuracy: the more accurate the model the higher its parameter count needs to be hardening the deployment stage. Finally, a common characteristic of data-driven models is their lack of transparency and explainability which is instead a key requirement in Industry 5.0, as explained in section I.

For the reasons mentioned above, model based approaches can still be useful in the modern era of PdM in the development of hybrid systems where domain knowledge is merged with data-driven architectures leveraging the advantages of both worlds. The inaccuracy of physics-based models can be enhanced using ML or DL algorithms to model more complex phenomena through data and reduce the amount of training data for data-driven solutions acting as data simulators and producing physics-enhanced input features for the models.

In [386] model-based approaches are used to produce data but also to enrich the feature space of the training set by

discovering some additional parameters which can improve the predictive power of DL models. Specifically, the training set of the data-driven model (i.e. a CNN architecture), which is normally composed of condition monitoring data produced by sensor readings, is enriched with scenario-descriptor operating conditions, virtual sensors and unobservable model parameters obtained as the result of the calibration of a physics-based model solved using an Unscented Kalman Filter (UKF) [505]. The obtained hybrid model shows improved performances with the respect to a solely data-driven model and it also requires less training samples thanks to its physics-enhanced input features.

The typical methodology employed in hybrid RUL estimation involves using a physics-based model to describe easy phenomenon in a very precise way, while more complex behaviors are explained through data using DL models. For example in [387] a PINN based RNN model is composed of physics-informed layers to model well known components of airplane fleets (i.e. fatigue crack growth) while data-driven layers are used to model more complex aspects like the internal loads. Similarly in [389], where the task is to predict corrosion-fatigue of an AI 2024-T3 alloy employed in fuselage panels of aircraft wings, physics-based layers are used to model crack growth through the Paris Law while challenging effects such as damage bias due to corrosion are modelled by data-driven layers. Finally, in RUL estimation of main bearings in wind turbines [388] relatively well known physical effects (i.e. L10 fatigue life) are modelled through physics-informed layers while degradation due to poor greasing conditions are left for data-driven layers.

Another very common example of hybrid PdM is represented by PINN [506], NN enhanced or integrated with physics rules and concepts. PINN combine the strengths of physics-based models, which are highly accurate in estimating machines' HI and degradation patterns, with those of purely data-driven models, which offer computational efficiency.

There are two kinds of PINN: *sequential integration models* in which the two components (i.e. the physics-based model and the data-driven model) act independently and *hybridized physical models* where the equations governing the physics-based models are integrated in the data-driven model. An example of this approach is contained in [391] where a CNN model for RUL estimation of lithium-ion batteries is integrated with Open Circuit Voltage (OCV) curves reconstruction methods to fully characterize the batteries degradation pattern and exploit those information in the training process.

A traditional application of PINN is presented in [392] where physical rules are included as regularization terms in the definition of the loss function used to train an RNN model which is used for the RUL estimation of discrete power electronic devices. In particular, the loss function is enriched with a term to control the monotonicity of the predicted RUL signal and to make sure that the prediction falls inside a specific bound. Similarly in [393] domain knowledge is incorporated in the loss function of an LSTM network for RUL prediction on supercapacitors. Moreover, Bayesian Optimization (BO) Optimization is used to tune the balance between the two components of the loss function (i.e. the training loss and the

physics-inspired regularization term).

In [390] a hybrid method for RUL estimation of rolling element bearings is proposed. In this case the degree of bearing damage is estimated by a combination of the Adaptive Spectrum Mode Extraction (ASME) theory with Fault Feature Gain (FFG) to adaptively monitor bearings faults frequencies. After the detection of incipient faults a DL models combining GRU and BNN is used to obtain the final RUL prediction.

In [394] a central role is played by the self-attention mechanism, typical of the transformer model [500], which is firstly used to map the high-dimensional sensor measurement data into a low-dimensional space by selecting only the key features. An additional attention-based network is used to learn the physics of the system as the partial differential equations between the hidden states and the predicted RUL.

4) *Reinforcement Learning Approaches for RUL Estimation*: Maintenance Scheduling is the third and final step of the maintenance pipeline, presented in IV-A. It is usually a component of Decision-Making Modules triggered after the ML or DL models failure predictions are produced [507]. It is a challenging task because it may involve several machines with different failure types and it requires to coordinate the maintenance operations among the team of experts that will act on the equipments necessity reparations. The optimization of this process is crucial in the industrial production chain since it may lead to a significant reduction of the machines downtime under maintenance and to a more efficient use of human resources. This task requires the ability to perform sequential decisions in a stochastic environment, which is the main reason why RL approaches are becoming more and more popular in Maintenance Scheduling. As RL is an automatic learning paradigm that learns through interaction with the environment it can solve some of the typical challenges of PdM like the scarcity of labeled data, documented failures and the presence of multiple operating and working conditions of equipments. RL agents are in fact able to learn the optimal policy to follow in order to minimize the downtime of the machines and to maximize the efficiency of the maintenance operations.

It is worth noting how the application of RL in PdM is quite challenging. In fact, differently from traditional RL formalizations in which the environment is fully observable, stationary and states fully measurable, in industrial applications there are several factors that can quickly modify the state transition probabilities such as production load variations, performed or missed maintenance and external weather [367]. Moreover, in Multi Agent RL applications in PdM an action taken by an agent can change the environment for the other agent/s. As a consequence Hidden-Mode MDP (HM-MDP)s are used to formalize the problem. This is the most complex type of Markov Decision Process (MDP) since no model of the environment is available and states are not fully measurable because determined by noisy sensor recordings. Moreover there is also the necessity of considering continuous and variable length time intervals for action duration, in fact using discrete time formulations could lead to the miss of a maintenance event. This additional requirement defines Semi HM-MDP.

The typical formulation of an RL agent in the field of PdM can thus be summarized as follows:

- **Environment**: The environment is represented by the equipment to monitor or even by the entire plant. Producing a detailed model of the environment can turn into an extremely complex task, thus model-free approaches are preferred.
- **States**: States represent different equipment health stages and are determined by condition monitoring performed through sensors measuring various quantities such as vibration, temperature, toll wear and similar.
- **Actions**: Typically three actions can be performed by the PdM scheduler agent: in case the machine's health is acceptable no action is performed, a single component can be replaced or a complete maintenance operation can be scheduled.

Several works can be found in the literature, according to [367] the most commonly used RL approaches are the basic Q-learning and its variants [508], Actor Critic based methods [509], policy-gradient methods (i.e. Proximal Policy Optimization (PPO) [510] and Deep Deterministic Policy Gradient (DDPG) [511]).

Some works in the literature are worth to be described in more details: for instance, [368] proposed a Multi Agent RL algorithm. The agents work in parallel on different copies of the same machine with the aim of maximizing the time in which the machine is in the *working* state. Maintenance is scheduled when needed assigning the task to a technician.

[512] focuses on a serial production system and uses a RL algorithm based on value function estimation to avoid Machine Stoppage Bottlenecks. In [513] a PPO-LSTM model is proposed in the IIoT framework. As already discussed in previous sections the LSTM architecture is chosen thanks to its excellent capabilities of extracting temporal features from sensor measurements. To speed up training of Deep Reinforcement Learning (DRL) based architectures [514] proposed a knowledge transfer method called Transfer Learning from Demonstrations (TLD) where expert-demonstration data are stored in an Experience Replay (ER) buffer to improve learning efficiency when the system is adapted to a new environment. [515] combined supervised learning for RUL prediction with DRL for maintenance scheduling. In particular a CNN network equipped with Monte Carlo Dropout (MCD) is used to estimate the probability distribution of the RUL signal, which is later used to generate the states of the DRL algorithm (i.e. Soft Actor Critic) which is used to schedule maintenance action based on the trend of the estimated RUL. A final work worth citing is [516] which focuses on bringing interpretability on the table through a Hierarchical DRL framework where, to facilitate the RL task, a probabilistic model (i.e. Input-Output Hidden Markov Model (IOHMM)) is used to filter the large amount of condition monitoring data to include just recordings of equipment's critical states. When these states are reached the DRL model comes in the picture. Moreover to limit the RL agent training time Behavioral Cloning (BC) is used to initialize a baseline policy.

C. XAI in PdM

One of the main drawbacks associated to the usage of AI related technologies in PdM is the black-box nature of the used models and the lack of transparency and explainability on their outputs, which can cause untrustworthiness in the users of these systems, who are generally not familiar with these technologies. Consequently, one of the key requirements when applying PdM methodologies in industrial settings is the ability to provide explanations on the model predictions to increase their trustworthiness towards the domain experts.

Traditionally the field of XAI modelling is characterized by two stages: model prediction and prediction explanation [468]. This is particularly true in case a model agnostic explanation algorithm is used. These algorithms, of which SHAP [278] and LIME [115] are the two most common examples, are agnostic in the sense that are independent on the model architecture but base their interpretation just on the relationship between model's inputs and outputs. A more efficient solution is represented by model specific algorithms which exploit the internal structure of a specific model to provide insights on its reasoning. As a consequence these XAI approaches come with the limitation of being usable just on the specific models for which they are designed. LRP is an example of model specific XAI since it is tailored for NN architectures.

Model agnostic XAI approaches find applications also in PHM: [468] benchmarks different regression models (i.e. from Random Forests to NN) on end of life prediction and successively interprets their outcomes using SHAP and LIME.

A peculiar ad-hoc interpretation method is introduced in [469] which consists in inserting a so called *feature selection layer* immediately after the input layer of a Deep Neural Network (DNN) based model which will contain the weight (i.e. importance) of each input feature which is learned during model training. The model is trained using an L1 loss function in order to produce sparse solutions and thus eliminate irrelevant features, indirectly performing a feature selection operation. Moreover a Ranking Quality Score (RQS) is introduced in order to provide a quantitative evaluation of the goodness of different feature subsets and also to compare different models on the RUL estimation task.

As described in previous sections CNNs architectures are very common in the field of RUL estimation. While being very accurate and efficient in the extraction of local spatial features, crucial for the production of acceptable RUL estimates, the architecture of these models is highly complex making interpretability an hard task. A possible approach to increase its trustworthiness is contained in [470] where the LRP XAI algorithm is employed on time-frequency spectral images of vibration signals produced by an induction motor. With LRP it's possible to identify regions of the input highly contributing to the final prediction by analyzing the pixel-level representation returned by the algorithm. Similarly [406] also applies LRP as an explanation algorithm on a Multi Scale Deep CNN used for the RUL prediction. Additionally, in the pre processing steps raw data are clusterized into different operational modes and physics-informed features are extracted to better capture the nature of degradation. Another DL based

model widely used in PdM is LSTM which is explained through an hybrid attention mechanism in [433].

Another utilization of XAI tools for PdM is their insertion in a complete PdM pipeline comprehending AD, RUL estimation and maintenance planning. In this context, explanations are performed in the AD stage with the aim of detecting the most relevant feature describing the detected anomaly or failure mode, acting as a feature selection mechanism to equip RUL estimation model with the most essential descriptors needed for end of life estimation [471], [446], [119].

For a more detailed description of how XAI is used in PdM the reader is referred to the following surveys on the topic: [472], [473], [474], [475].

D. PdM Industrial Applications

In this section an overview of the most common applications of PdM in the industrial domain is provided, with a focus on how these methodologies are applied and deployed and how they can enhance the performance of industrial systems.

When considering the application of PdM in the industrial setting several challenges have to be taken into account. This task in fact requires failure data of the equipment, a rarity in the industrial domain, where the practice of sharing fault records is avoided for economical and privacy reasons. Moreover multiple failures are usually responsible of a machine's life degradation, making it challenging to correctly label each one of them. As a result researchers and practitioners lack of standard guidelines on the implementation of such prognostics systems. Moreover, the lack of publicly available datasets decreases the chances of being able to carry out a fair comparison between different RUL estimation methods making most of the RUL prediction approaches to be case-dependent [482].

1) *RUL Estimation for Lithium-Ion Batteries*: A common field of application of PdM methodologies is the one of lithium-ion batteries that is becoming increasingly popular in the recent years thanks to the development of electric vehicles. The development of prognostic applications in this field is of crucial importance because of the risks associated to the usage of these devices outside of their operating limits: fire, toxic gas release and electric shock [419].

One of the main challenges for the application of RUL estimation on these kind of devices is that the degradation process may also be inverted, with batteries capable of restoring some of their capacity. This phenomenon is called Capacity Regeneration Phenomena (CRP) [406] and dramatically increases the difficulty of the task since most of the RUL estimation approaches are based on the assumption that the degradation phenomena leading an equipment to the end of its life is a monotonic process.

Considering the application of RUL based prognostic methods in the field of lithium-ion batteries it is possible to see the usage of all the different RUL estimation approaches listed in the previous sections.

An hybrid based approach is employed in [406] to deal with CRP. In particular, the battery capacity degradation signal is decomposed into a seasonal, residual and trend component.

The latter component is the one mostly correlated with the RUL and thus estimated through a Gaussian model. On the other hand the local features (i.e. seasonal and residual components), which represent the Capacity Regeneration Phenomena (CPR) phenomena, are estimated through a Bidirectional GRU data-driven model.

A similar approach is employed in [396] where the global deterioration trend and the local capacity regeneration are decoupled using EMD and then estimated through an Auto Regressive Integrated Moving Average (ARIMA) based model. As already described in IV-B3, another hybrid approach for RUL estimation consists in enriching the knowledge of data-driven models with domain knowledge on the batteries degradation process [391].

In [382] a purely data-driven approach is considered. In fact, the hierarchical spatial characteristics together with the temporal dependencies typical of the prognostics scenario are learned through the so called *Convo-LSTM* network which combines the 1DCNN (i.e. 1 dimensional CNN) architecture with the LSTM one. In particular, an LSTM cell is used to model the temporal patterns while convolutional operations are used instead of the classical multiplication to exploit their weight-sharing strategy and locally focused representation abilities. The combination of CNNs and LSTM networks is also exploited in [420] with the addition of an Autoencoder that deals with the lack of data by extracting additional features from the input space. Moreover a post prediction smoothing filter is applied on the RUL signal to solve the discontinuity problem in the prediction result. In [421] BO is used in the hyperparameter tuning of these hybrid models in order to maximize their predictive accuracy.

LSTM model is employed also in [418] where the focus is on the estimation of the SOH of the battery. Inaccurate SOH estimation is known to lead to spontaneous combustion accidents so it's crucial to be able to provide an estimate of this quantity which is defined as the ratio of a fully charged battery in the current state on the capacity of a fully charged on as it is when initially bought (i.e Nominal Capacity):

$$SOH \approx \frac{\text{Current actual capacity (Ah)}}{\text{Nominal Capacity (Ah)}} \times 100\%$$

Conventionally the end of life is determined when a device capacity is lower than the APT defined as 70-80 % of the rated capacity. This quantity is correlated to other features in batteries data such as the number of discharging cycles, the time required to reach a certain voltage level or the time required to reach a certain temperature. This fact is exploited in [418] where Support Vector Regression (SVR) and LSTM models where trained on those three extracted features to predict the SOH.

Another commonly used architecture in data-driven temporal modelling is the Transformer [500] which is however also known for its quadratic time complexity that makes its use unpractical when dealing with long degradation life patterns. This challenge is dealt in [422] with the introduction of an Exponential Transformer where an exponential decaying measure is employed in the attention mechanism to focus

more on nearby data points. Moreover, similarly to the works cited above, the input data are decoupled into their trend and fluctuation components to produce highly interpretable results. A Transformer based architecture is also used in [423] where a Patch-wise Attention Network (DPAN) correlates patches of the signal to capture both global features (i.e. intra-patch attention) and local ones (i.e. inter-patch attention). Successively the relationships among these features are modelled through a Feature-wise Attention Network (FAN) employing the traditional self-attention.

2) *RUL Estimation for Rolling Element Bearings*: Rolling bearings have an huge impact in the performances of rotating machinery equipment and have been recognized as one of the main sources of failures. For this reason the RUL estimation task for rolling element bearings is crucial in the field of machine prognostics to avoid unexpected failure and to enhance efficiency and safety.

Feature extraction techniques, as the ones outlined in IV-B2, are highly beneficial in the field of bearing diagnostics and prognostics, in particular if frequency domain features are extracted from vibration signals. In fact, as outlined in [517], ground truth knowledge is available since some specific harmonic frequency are directly related to specific faults in bearings: Ball Pass Frequency Outer (BPFO), Ball Pass Frequency Inner (BPFI), Ball Spin Frequency (BFS).

These features are exploited in [446] where a complete PdM pipeline is implemented: firstly incipient bearing faults are detected through the IF AD algorithm, successively Un-supervised Fault Classification is performed thanks to the usage of the XAI SHAP algorithm which ranks the input features in terms of importance in the AD task and the most important fault related feature is used to detect the fault affecting the equipment. Finally RUL prediction is conducted by estimating the Weibull Distribution on the most important variable according to SHAP. Weibull-based failure probability modelling is also studied on its own in [518], for both individual machines and entire systems and across different dataset sizes.

A more complex method for the extraction of fault related features is explored in [390] called ASME which decomposes the original vibration signal into multiple simple Intrinsic Mode Functions (IMF)s. Successively the FFG method is used to adaptively calculate the amplitude of each fault frequency which is a crucial components for the estimation of RUL in bearings. Finally, the study considers the combination of a GRU architecture for temporal modelling and a BNN architecture to enrich the RUL prediction with a quantification of its uncertainty, which is a crucial factor in the field of PHM as detailed in IV-B2.

Feature extraction is addressed in an original way in [403] where a Dynamic Complexity Characteristic Entropy (DCCE) method is used to fully characterize the degradation trend of rolling bearing's useful life. Successively the HI is constructed using PCA and VAE. Finally, features extracted in this way are fed into a Deep Quantum Neural Network (DQNN) to obtain the final RUL prediction.

Even though feature extraction based approaches are naturally more suitable for rolling elements PdM, end-to-end

DL solutions are also explored. For example, in [447] the Recurrent Convolutional Network with Wide First Layer Kernels and Residual Shrinkage Building Unit (MSWR-LRCN) architecture is presented. In this model soft thresholding and attention are incorporated in a CNN model to form an Residual Shrinkage Building Unit (RSBU) unit whose aim is to remove noisy and redundant feature from the input and provide the model just with the features relevant for the task at hand. This is the first path of the feature extraction pipeline which focuses on local features. In parallel an Attention LSTM block is used to extract information spanning large number of time steps. Finally a BiLSTM block is used for the final RUL prediction.

The combination of soft thresholding and attention to eliminate redundant information is also employed in [448] in order to improve the prediction accuracy of a TCN based model.

A fundamental requirement for a task so deeply connected to industrial and real-world scenarios is the ability of producing solutions able to transfer knowledge between different scenarios (from the test rig to the real equipment) and working conditions. Transfer Learning techniques represent a viable solution in this direction. In [374] the Multisource Domain Meta Transfer Learning (MD-MTL) transfer learning method is used to migrate the estimation capabilities of a model composed of a Convolutional Encoder and ProbSparse Multihead Attention Decoder (CE-PSAD) from the test rig to real wind turbine bearings.

Another example of the utility of transfer learning is represented by [400]. In this work firstly a detailed construction of HI indexes is performed. Next, a dual-task (i.e. health assessment and RUL prediction) feature extractor is used to extract useful features exploiting a Convolutional Block Attention Module (CBAM) and Bidirectional Gated Recurrent Unit (BiGRU) architecture. Successively a Model Migration module is responsible to perform Domain Adaptation to move from the training (i.e. test rig) to the target domain (i.e. real-world scenario). Moreover the Weibull Proportional Hazards Model (WPHM) physics-based model is integrated into the data-driven model loss function, thus creating a PINN model as the ones presented in section IV-D1

3) *IoT Papers*: In this sections we consider works whose main focus is posed on the IoT architecture used to provide the monitoring data to the prognostic system. As presented in I, IoT infrastructure are nowadays the standard for PdM systems because of the large amount of benefits associated with them. Other than potential benefits, several challenges have also to be considered, as outlined in [476]. These include integrating diverse sensor systems, handling high volumes of data, ensuring data security and addressing costs and scalability of the implementation of such systems. Moreover, it is also crucial to consider that the transition from traditional maintenance practices to IoT-based ones will come with technical and organizational difficulties.

The typical pipeline of an IoT-based PdM system involves deploying multiple sensors on industrial machines to monitor critical variables, such as vibration or temperature, transmit these data through a cloud infrastructure where data are pre processed to deal with noisy or missing recordings and to perform predictive analysis through ML or DL models.

An example of such pipeline is provided in [477] where a complete PdM infrastructure is presented: temperature, humidity and vibration data are acquired with an IoT system composed of an Arduino Uno and ESP-32 sensors. Successively a cloud infrastructure is employed for data storage, data preprocessing and for identification of incipient faults through the usage of a RF model for Feature Selection and an initial classification of the machine's status as normal or fault. Successively an LSTM model is used to detect anomalous and faulty patterns in the data by providing alarms to maintenance personell through SMS in case a fault is detected.

A similar system is implemented in [478] where communication protocols like Message Queuing Telemetry Transport (MQTT), Hypertext Transfer Protocol (HTTP), OPC Unified Architecture (OPC-UA) are employed for the transmission of data from sensors to the cloud infrastructure. Differently from [477] in this case notification of potential issues in the monitored equipment are communicated through a web-based dashboard.

[479] deals with two crucial aspects in the application of PdM to the industry: data privacy which is handled by employing a FL framework for distributed and secure model training and XAI to equip model predictions with interpretable insights for the maintenance operators.

[376] proposed an Artificial Intelligence of Things (AIoT) system for PdM of the wheel and axle systems of high-speed trains. Differently from typical approaches employing IoT in this case quite complex DL algorithms are considered for the predictive part: Multi Scale Transformer (MST) and Deep Wavelet Transform (DWT). Considering the complexity and computational power required by these models some edge computing solutions were adopted for the deployment of this solution on a railway infrastructure. The predictions produced by these models are used by a Maintenance Decision-Making Layer which can generate maintenance schedules based on RUL predictions, raise alarms in case anomalies are detected and it is also integrated with the train management system to synchronize maintenance actions with train schedules to avoid disruptions of the transportation system.

[480] proposed a ML-based PdM system for a circular knitting machine. Machine speed and stops data are collected through an IoT infrastructure which uses a Printed Circuit Board (PCB), a ESP32 microcontroller and several sensors whose recordings are collected and managed through the *mongoBD* database management system. For what concerns the ML component, the Ada Boost model is used to classify different classes of machine's faults in advance in order to provide target interventions.

As it can be noted from the examples above in IoT-based PdM, predictions are not done on the RUL of the equipment but sensor recordings are directly analysed to detect failures in advance and consequently perform maintenance operations before taking place.

The advent of 5G cellular technologies has an impact on IoT-based systems which can be furtherly enhanced improving real-time data collection. [481] studies the application of this technology in the case study of a factory floor where Automated Guided Vehicles (AGV) are used to transport hazardous

liquid that, if leaked, can endanger workers. Thus, exploiting a 5G based IIoT network for data collection a DL model is trained for RUL prediction. RUL is predicted on the server side and an alert is sent to the AGV in case a failure is forecasted to trigger maintenance intervention.

4) *Business related metrics for the application of PdM in the industry:* As detailed in IV-B2 an aspect of paramount importance in PdM is the trade-off between the two extreme conditions of *unexpected breaks* and *unexploited lifetime* which assume a great importance when PdM systems have to be deployed in an industrial environment where premature breakages or maintenance actions can translate in significant economical losses. For this reason there are several applications of PdM in which the economical aspect is embedded in the RUL estimation model in the form of so called **business metrics**. In [6] Susto et al. proposed a business oriented metric in which, assuming that underestimating the RUL is the preferred error type from an economical point of view, a more cautious approach is embedded in the model by subtracting a constant quantity m (labeled *maintenance window*) from the model output: $\hat{RUL} = \bar{RUL} - m$ where $\bar{RUL}$ is the original model prediction and $\hat{RUL}$ is the adjusted value after the application of the *maintenance window* m . The hyperparameter m needs to be tuned in order to find the correct balance between *unexpected breaks* and *unexploited lifetime*. The following loss function is proposed:

$$J = \rho_{UB} C_{UB} + \rho_{UL} C_{UL} \quad (2)$$

where ρ_{UB}, ρ_{UL} represent the percentage of *unexpected breaks/unexploited lifetime* encountered in the test set and C_{UB}, C_{UL} are the costs associated to those situations, which have to be defined by the users based on the application needs. By optimizing 2 the optimal m window can be defined.

5) *PdM Industrial Case Studies:* In this section some case studies on the integration of PdM in smart manufacturing are presented with the aim of showing the benefits and challenges that this maintenance approach provides.

[483] proposes a hybrid framework for the integration of PdM in manufacturing. As usual the starting point is represented by data collection, performed by IoT-enabled sensors. Successively CNNs networks are employed for time-series feature extraction, followed by the combination of LSTM for sequence modelling and ARIMA model for forecasting. AD algorithms, such as IF and One-Class SVM, can also be considered for an early detection of potentially anomalous conditions. Finally, after a fault is detected, maintenance is scheduled following and Adaptive Maintenance Scheduling approach realized through the usage of an RL agent whose rewards coincide with the operators feedback.

The work presented in [484] is a great example of the computational challenges in the deployment of complex DL models in an industrial scenario. The case study focuses on the implementation of a web application for RUL prediction to guide maintenance interventions for cutting tool machines in milling industries. Due to the computational constraints usually faced in the deployment stage the RUL prediction

methods are all ML-based (i.e. DT, RF, Multi Layer Perceptron (MLP), ...) with MLP being the better performing one.

[498] presents a PdM strategy to substitute the conventional three-stage maintenance strategy employed in nuclear power plants, where maintenance plays a crucial role considering the devastating consequences that an incident in such a plant can cause to the environment. The developed prognostics system is implemented on the circulating water pump of the power plant. The system is composed of different stages: first a sensor selection strategy is used to reduce the number of input features to the RUL prediction models and thus guarantee higher efficiency. Because of the hardness in the collection of degradation data, pseudo-labels are created to train a *LassoNet* model to discard redundant sensors and select solely the ones relevant to for the RUL estimation task. Successively degradation-related sensors are used to construct HI which are used to establish a degradation trend prediction model to perform a short-term prognosis. Considering the fact that multiple degradation sensors are considered for the construction of the HI, an ensemble approach is used for trend prediction. It is worth noting how among the base models for the ensemble Structured State Space Sequence model (S4) is considered. This model is part of the family of SSM models, which are rarely used [465], [519] in the field of RUL estimation. Finally, a long-term RUL prediction is obtained through a model-based approach which is selected based on the degradation trend prediction performed at the previous stage. The RUL prediction is defined as the time when a specific threshold value (determined) is overcome by the HI. For what concerns the industrial deployment, an IoT infrastructure is considered composed by an end device (i.e. the sensors), an edge device represented by a computer which processes real-time monitoring data and perform data cleaning, AD and HI construction. Successively the elaborated data are sent to the cloud server whose task is to train the models for sensor selection, trend degradation and RUL prediction.

Another application of RUL estimation in a nuclear power plant is introduced in [485] where the focus is on Class II components (i.e. critical piping systems that can experience the Flow Accelerated Corrosion (FAC) fault). Due to the scarcity of real data the proposed models for RUL prediction were validated on a synthetic dataset produced by a Finite Element Model (FEM) model of the components under analysis. The synthetic dataset reproduces the operating conditions experienced in a catastrophic pipe rupture incident (caused by the FAC fault) in Surry Unit 2 in December 1986. For what concerns the methodology employed for RUL estimation an AE was used for feature selection and the reconstruction error was used to represent the machine's health status: the smaller the reconstruction error the healthier the machine. Finally an ARIMA model was considered to perform time-series forecasting on the learned HI to predict equipment's degradation trend and produce a RUL estimation. This case study represents an example on how challenging data collection can be in the field of PdM, in particular in cases in which such complicated physical degradation phenomena are considered. In these cases there is the need to exploit model-based approaches (FEM analysis in this case) to simulate

the FAC fault which normally expands at a rate of a few millimeters per year, making it impractical to be monitored through sensing technologies.

A novel PdM approach is presented in [486] for prognostics on marine engines. An LSTM based model is used for failure prediction and it is trained in an online fashion by using data collected in a three-week period to predict failure in the subsequent week and by continuously discarding old data as new ones are available, making it possible for the model to adapt to new environmental conditions that a marine engine can face during the year. Another important contribution is represented by an accurate selection of relevant Key Performance Indicators (KPIs) for model training. Since historical data are not considered in this online-based approach the selection of KPI is performed by analysing common faults in previously designed marine engines and by simulating them to confirm the correlation with the selected KPI.

Another case study is described in [487] where a PdM system is implemented by performing a regression analysis on the sensor data produced by the Alarm Monitoring System (AMS) of large cargo boats. Data Augmentation is applied to produce additional abnormal conditions for model training using a 1D simulation model of the monitored engine. The PdM algorithm is based on the method proposed in [488] where the Revision Generator Engine Condition Criterion (RGCCV) index is introduced to take the maintenance decisions. This index is based on the analysis of the exhaust gas temperature, one of the variables measured with the IoT sensors which is highly correlated with engine failures.

Traditionally, PdM is solved through RUL estimation as a regression task. However there are some case studies in which PdM is posed as a classification problem [6], [489]. In particular in [489] the aim is to predict whether a distribution transformer will fail within a given time horizon. The authors of this case study specify several challenges encountered in dealing with real world data on distribution transformers: missing data, high feature dimensionality, presence of both numerical and categorical features. Finally, when considering PdM as a classification problem dataset imbalance is guaranteed considering that equipment failures are rare events. In order to deal with these challenges, that are typical of real-world datasets, the authors used a tree based method (i.e. RF) to deal with mixed feature types and employed random under-sampling with boosting to create a more balanced dataset. Moreover the authors developed a specific metric called *match in top N* (MITN) in order to avoid a high number of false positives predictions. In fact the costs of premature replacements of transformers may exceed the repairment cost for a sudden failure. This final consideration highlights another crucial factor to take into account when deploying a PdM solution in the industrial world: the economical consequences of prediction errors. Depending on the specific applications a pessimistic or optimistic prediction may have different economical impacts, creating a trade-off of difficult solution. A possible approach to deal with challenge, presented for the first time in [6], is represented by the usage of business-related metrics (IV-D4).

V. INDUSTRIAL USE CASES FROM THE AIMS5.0 PROJECT

This section focuses on the AIMS5.0 European project by introducing its scopes and missions and reporting its current state on some of its Use Case (UC) strictly related to the topics of this chapter.

A. Overview of the AIMS5.0 Project

The AIMS5.0 project is a European research and innovation initiative aimed at accelerating the transition from Industry 4.0 to Industry 5.0 by enabling the systematic integration of AI into industrial processes. The project addresses key technological, organizational, and human-centric challenges to foster sustainable, resilient, and trustworthy manufacturing ecosystems.

The final aim of the project is to develop and validate a comprehensive framework for AI-enabled industrial transformation, supporting the full lifecycle of AI adoption — from use case specification and data management to algorithm development, deployment, validation, and standardization. Through the development of AI methods, architectures, and reference solutions, AIMS5.0 seeks to empower industrial stakeholders to deploy scalable, explainable, and human-centric AI systems that enhance productivity, sustainability, and decision-making in alignment with the Industry 5.0 paradigm.

AIMS5.0 integrates twenty industrial use cases (UCs) from globally renowned companies across sectors such as semiconductor manufacturing, machine tools, robotics, logistics, and production systems. These UCs serve as real-world validation environments to demonstrate how AI-driven solutions can improve operational efficiency, sustainability KPIs, human-machine collaboration, and resilience across complex industrial value chains.

Among the twenty UCs, four are directly aligned with the frameworks presented in this chapter, namely AD, PdM, and FDD:

- Anomaly Detection on Production Processes;
- Edge Anomaly Detection and Predictive Maintenance;
- Tool Breakage Detection (Fault Detection);
- Industrial Production-Ready Anomaly Detection.

The following subsections describe the objectives and operational principles of these four UCs, highlighting their relevance to diagnostic and prognostic AI frameworks in Smart Manufacturing.

B. AI-related Challenges in AIMS5.0 Use Cases

Before delving into a detailed description of the four UCs listed above, this subsection describes the AI-related challenges faced in AIMS5.0 UCs. The most commonly faced challenges can be divided into the following groups:

- **Data Collection** is the most frequent obstacle when dealing with industrial AI integration. In particular issues regard the presence of unbalanced or low-quality data (i.e. defects are rare), difficult connection between ML or DL based tools when data are stored in Manufacturing Execution Systems (MES) or ERP systems, privacy issues due to datasets confidential or owned by third-parties and

data distribution issues as a result of the post COVID pandemic era.

- **Algorithmic problems** regard the lack of best practices to train and validate AI models. Interestingly in some UC the so-called Fear of Missing Out (FOMO) and *fast time to market* pressure can hinder AI adoption. Other challenges regard the lack of computing power, available AI and ML platforms.
- **Human-centric problems** regard the lack of explainability and trustworthiness of some AI solution.

C. Anomaly Detection on Production Processes

This use case addresses anomaly detection within industrial production processes. The primary objective is to identify abnormal behaviours or deviations from expected operational patterns that may affect product quality or process efficiency.

The UC focuses on the development of ML models capable of detecting anomalous conditions in manufacturing workflows. These anomalies may correspond to unusual production states, deviations in process parameters, or unexpected variations in output characteristics. The approach involves collecting process data, defining relevant operational indicators, and designing models that distinguish between normal and abnormal operating regimes.

At its current stage, this use case is characterized by an exploratory development phase. The main efforts are directed toward defining appropriate problem formulations, identifying meaningful data sources, and selecting suitable anomaly detection techniques. The challenge lies not only in model development but also in clearly specifying validation metrics and performance indicators that reflect industrial requirements.

From a methodological perspective, this use case exemplifies the early stages of industrial anomaly detection deployment, where problem definition, data strategy, and KPI specification are critical enablers for successful AI integration.

D. Edge Anomaly Detection and Predictive Maintenance

This use case focuses on PdM and AD executed directly at the edge level. The objective is to enable real-time monitoring of industrial assets without relying on continuous cloud connectivity, thereby addressing latency, bandwidth, and data privacy constraints.

The UC leverages lightweight and interpretable ML models deployed on industrial edge devices such as Programmable Logic Controllers (PLCs), embedded systems, or gateways. Typical algorithms include decision trees, rule-based models, and TinyML neural networks optimized for constrained hardware environments. Gradient-boosted decision trees are used as core models, fine-tuned on self-collected industrial datasets.

A key design choice in this use case is the development of specialized models tailored to specific operational contexts rather than a single generalized solution. This specialization allows the models to capture equipment-specific behaviour and degradation patterns more accurately. The system architecture supports incremental updates and continuous refinement as new operational data becomes available.

This UC represents a concrete implementation of PdM in distributed industrial environments, where computational constraints, model interpretability, and deployment robustness are central design requirements.

E. Tool Breakage Detection in Machine Tool Operations

This use case addresses FDD in precision machining processes, with a specific focus on automatic tool breakage detection. The objective is to detect tool failures during machining operations using high-frequency vibration signals and other sensor data, enabling timely intervention and reducing scrap and downtime.

The approach combines signal processing and supervised ML. First, frequency-domain features are extracted using FFT analysis to capture harmonic patterns associated with tool condition. These features are then used as inputs to binary classification models, including RF, SVM, k-NN, and Logistic Regression, to determine whether the tool is in a *broken* or *not broken* state.

In addition to supervised learning, unsupervised techniques such as motif discovery and change-point detection are used to segment repetitive machining cycles automatically. This segmentation step is crucial for isolating comparable operational patterns and improving classification reliability.

A central challenge in this UC is the transition from experimental datasets to real-world production data, which are often unlabelled and more heterogeneous. Therefore, the use case highlights common industrial issues such as limited labelled data, variability across machining conditions, and the need for robust model generalization.

This UC constitutes a representative example of data-driven FDD in manufacturing systems and illustrates the practical steps required to industrialize supervised diagnostic models.

F. Industrial Production-Ready Anomaly Detection

This use case represents a mature implementation of anomaly detection deployed in a production environment. The objective is to develop and operationalize an application-specific ML model capable of identifying abnormal production conditions at scale.

In contrast to earlier-stage UCs, this use case is characterized by fully specified targets, validation metrics, and KPIs. The developed model has been deployed within the production infrastructure, and a pilot phase is being conducted to evaluate its performance under real operational conditions.

The system follows a complete industrial AI lifecycle: customization to the specific production context, model training on industrial data, deployment within the production environment, and continuous monitoring during operation. Particular attention is given to scalability, robustness, and integration with existing manufacturing systems.

This use case demonstrates the transition from research-oriented anomaly detection methods to production-grade AI solutions. It exemplifies how well-defined objectives, structured validation procedures, and close collaboration between AI developers and industrial stakeholders enable successful deployment in real manufacturing settings.

Overall, these four UCs collectively illustrate different maturity stages and application contexts of AD, PdM, and FDD within Smart Manufacturing. They provide practical insights into how AI-based diagnostic and prognostic techniques are designed, validated, and integrated into industrial environments under the Industry 5.0 paradigm.

VI. COMPARISON, DISCUSSION AND RESEARCH GAP

This section is devoted to perform a comparison between the different methods and applications surveyed in Sections II, III, IV. The comparison will not be targeting the quantitative performances of different diagnostic and prognostic approaches but will rather discuss the main difference and similarities between those and potentially highlight some common research trends in the field and identify potential gaps.

Among data-driven models there is a division in different domains of application of the RUL estimation task. Turbofan engine is the most popular domain, probably because of the popularity of the C-MAPSS based datasets which contain high-quality data and represent a standard for the benchmarking of RUL estimation methods. Given the sequential nature of the C-MAPSS data most of the proposed methods employ the LSTM, CNNs architectures to exploit their excellent temporal and spatial feature extraction capabilities. A lot of solutions decide to merge together both these qualities by combining the LSTM and CNNs architectures. Additionally, the inclusion of attention based architectures for an importance weighting of the time samples is also explored. The high-quality of the C-MAPSS data pushes researchers to design solutions that may be excessively complicated and too focused on the performances showed on this dataset. In this context, it is important to note that the data contained in C-MAPSS and N-CMAPSS are the result of a simulation that, albeit precise, does not guarantee to be a good enough representation of real-world conditions. Consequently, comparable performances once these solutions are applied to real cases are not guaranteed. Moreover the elevated computational complexity associated to the DL architectures employed requires extra work to develop edge computing solutions, as in [376], to make the deployment of such architectures sustainable.

Another negative aspect of approaches benchmarked on the C-MAPSS dataset is reported in [370]: the evaluation criteria used to validate the different approaches presented differs from work to work making impractical to conduct a full comparison in terms of quantitative metrics.

The elevated computational complexity associated to DL based models makes the solution of enriching simpler ML models with engineered input features still a viable choice. This approach is in fact the one with the highest amount of review articles in Table X.

In this context the topic of feature engineering becomes of crucial importance. As explained in IV-B2 pre-defined statistical features in time or frequency domain are typically used. These features usually are able to extract the degradation process of the machine assisting the predictive models in the prediction of the RUL signal. However in some cases this

is not enough and domain knowledge should be employed by pairing data modelling with physical knowledge on the machine deterioration. This is where hybrid RUL methods, grouped in Table IX, should be used. Among these approaches a valid class of models are PINN, in which prior knowledge on the equipment can be used to craft highly informative input features or as a regularization term in the loss function which guides the training process of the underlying NN model.

Inside the manufacturing process PHM constitutes the final step where the fundamental decision on how to face the incoming failure on the machine is taken. Thus is crucial to provide accurate results that should be able to deal with the multiple operating conditions and different environments and domains where the same machine may be employed. As described in previous sections numerous contributions were presented in this direction but this is a challenging tasks and some research gaps have still to be faced. One of them is the transferability of RUL prediction models from simulated environments (i.e. C-MAPSS) to real industrial processes. Developing overly complex models on a synthetic datasets like C-MAPSS, designing specific architectural choices to deal with a very unique and peculiar setting may produce unsatisfiable results when applying said architectures to completely different scenarios. For this reason it's important to employ transfer learning or domain adaptation solutions such as the ones presented in [445], [374], [494], [12], [514], [449] which represent a small fraction of the current state of the art on RUL estimation.

Other than the difficulty on transferring the knowledge between different domains complex DL approaches, normally associated with demanding computational requirements, are not satisfying the requirements of operational efficiency and transparency required by Industry 4.0 and 5.0 requiring the need to develop more efficient and trustworthy solutions. A solution in this direction can be represented by hybrid-based RUL models where the computational load of DL model is reduced by the presence of prior physical knowledge that can help in modelling some less complex components/features of the degradation process leaving to the DL component to explain what cannot be efficiently modelled by model-based approaches to refine the predictions.

Finally, to deal with the human-inspired principles required by Industry 5.0 additional focus should be addressed towards the integration of XAI approaches and UQ methodologies in order to equip the final model prediction with auxiliary information such as a list of the features more closely related to machine degradation and confidence intervals on the RUL estimate. These information can promote informed decision and increase trust in maintenance personell.

VII. CONCLUSION

This survey has comprehensively reviewed AD, FD, and PdM within the context of Industry 4.0 and 5.0. The critical role of AI, IoT, and IIoT in enabling Smart Factories was established, alongside the advancements in diagnostic and prognostic capabilities. State-of-the-art methodologies, encompassing model-based, data-driven, and hybrid approaches for

RUL estimation, were thoroughly examined. The increasing importance of UQ and XAI for building trustworthy and transparent systems, aligning with Industry 5.0's human-centric principles, was also highlighted.

Industrial applications across various sectors demonstrated tangible benefits, though significant real-world challenges persist. These include data scarcity, computational demands of complex AI models, integration complexities and legal or privacy issues related to the data collection process.

To fully realize the potential of these frameworks, several research gaps require addressing: enhance model transferability to real world applications and different domains through transfer learning and domain adaptation approaches, employ edge computing solutions to enable a sustainable application of complex DL based architectures, and increase the application of XAI and UQ methodologies to form human-friendly predictions in order to fulfill the requirements of the Industry 4.0 and 5.0 revolutions. Another key challenge to address is the one related to data privacy, mainly caused by private companies concerned about protecting their resources, which can be faced with FL based solutions.

In conclusion, AD, FD, and PdM are pivotal for intelligent industrial operations. Addressing these identified gaps will enable the industrial sector to fully leverage the transformative capabilities of Industry 4.0 and 5.0, leading to enhanced efficiency, reduced downtime and sustainability.

ACKNOWLEDGMENT

The research leading to these results was conducted in the AIMS5.0 project, which received funding from the Chips Joint Undertaking under grant agreement No. 101112089 and from the partners' national funding programmes.

REFERENCES

- [1] P. Adolphs, H. Bedenbender, D. Dirzus, M. Ehlich, U. Epple, M. Hankel, R. Heidel, M. Hoffmeister, H. Huhle, B. Kärcher, H. Koziol, R. Pichler, S. Pollmeier, F. Schewe, A. Walter, B. Waser, and M. Wollschlaeger, "Reference architecture model industrie 4.0 (RAMI4.0)," VDI/VDE Society for Measurement and Automatic Control (GMA) and ZVEI, Status Report, Jul. 2015.
- [2] H. Lasi, P. Fietke *et al.*, "Industry 4.0," *Business & Inf. systems Eng.*, vol. 6, pp. 239–242, 2014.
- [3] A. Keshar, "Advancing industrial iot and industry 4.0 through digital twin technologies: A comprehensive framework for intelligent manufacturing, real-time analytics and predictive maintenance," *World Journal of Advanced Engineering Technology and Sciences*, vol. 14, pp. 228–240, 01 2025.
- [4] Deloitte, "Internet of things (iot) the rise of the connected world," 2021. [Online]. Available: https://www.deloitte.com/content/dam/Deloitte/in/Documents/technology-media-telecommunications/int-tmt-IoT_Theriseoftheconnectedworld-28aug-noexp.pdf
- [5] A. Frankó, G. Hollósi, D. Ficzer, and P. Varga, "Applied machine learning for IIoT and smart production—methods to improve production quality, safety and sustainability," *Sensors*, vol. 22, no. 23, p. 9148, 2022.
- [6] G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 812–820, 2015.
- [7] J. Leng, W. Sha, B. Wang, P. Zheng, C. Zhuang, Q. Liu, T. Wuest, D. Mourtzis, and L. Wang, "Industry 5.0: Prospect and retrospect," *Journal of Manufacturing Systems*, vol. 65, pp. 279–295, 2022. [Online]. Available: <https://www.sciencedirect.com/journal/journal-of-manufacturing-systems/article/pii/S0278612522000917>
- [8] E. Valette, H. B. El-Haouzi, and G. Demesure, "Industry 5.0 and its technologies: A systematic literature review upon the human place into iot and cps-based industrial systems," *Computers & Ind. Eng.*, vol. 184, p. 109 426, 2023.
- [9] M. C. Zizic, M. Mladineo *et al.*, "From industry 4.0 towards industry 5.0: A review and analysis of paradigm shift for the people, organization and technology," *Energies*, vol. 15, no. 14, p. 5221, 2022.
- [10] G. Dimitrakopoulos, P. Varga, T. Gutt, G. Schneider, H. Ehm, A. Hoess *et al.*, "Industry 5.0: Research areas and challenges with artificial intelligence and human acceptance," *IEEE Industrial Electronics Magazine*, vol. 18, no. 4, pp. 43–54, 2024.
- [11] A. Saxena and K. Goebel, "Turbopfan engine degradation simulation data set," NASA Ames Prognostics Data Repository, 2008. [Online]. Available: <https://c3.ndc.nasa.gov/dashlink/resources/139/>
- [12] X. Du, J. Chen, J. Wang, H. Zhang, and J. Wen, "Fusion-based dual-task architecture for predicting the remaining useful life of an aeroengine," *Journal of Aerospace Engineering*, vol. 38, no. 3, p. 04 025 004, 2025. [Online]. Available: <https://ascelibrary.org/doi/abs/10.1061/JAEEZ.ASENG-5887>
- [13] S. Chatterjee and A. Keprate, "Exploratory data analysis of the ncmapps dataset for prognostics," in *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2021, pp. 1114–1121.
- [14] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics," *Data*, vol. 6, no. 1, 2021. [Online]. Available: <https://www.mdpi.com/2306-5729/6/1/5>
- [15] A. C. J. Bonatakis and N. Propes, "Phm data challenge 2018," PHM Society, 2018. [Online]. Available: <https://c3.ndc.nasa.gov/dashlink/resources/139/>
- [16] B. Saha and K. Goebel, "Battery data set," NASA Ames Prognostics Data Repository, 2007. [Online]. Available: <https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/>
- [17] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Chebel-Morello, N. Zerhouni, and C. Varnier, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Conference on Prognostics and Health Management*, vol. sur CD ROM. Denver, Colorado, United States: IEEE Catalog Number: CPE12PHM-CDR, Jun. 2012, pp. 1–8. [Online]. Available: <https://hal.science/hal-00719503>
- [18] NASA, "Ims bearings — dataset," <https://data.nasa.gov/dataset/ims-bearings>.
- [19] C. W. R. U. Case School of Engineering, "Bearing data center," <https://engineering.case.edu/bearingdatacenter>.
- [20] C. A. Rieth, B. D. Amsel, R. Tran, and M. B. Cook, "Additional tennessee eastman process simulation data for anomaly detection evaluation," *Harvard Dataverse*, vol. 1, p. 2017, 2017.
- [21] J. Downs and E. Vogel, "A plant-wide industrial process control problem," *Computers & Chemical Eng.*, vol. 17, no. 3, pp. 245–255, 1993, industrial challenge problems in process control.
- [22] V. Zaccaria, C. Masiero, D. Dandolo, and G. A. Susto, "Enabling efficient and flexible interpretability of data-driven anomaly detection in industrial processes with acme-ad," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2024, pp. 1375–1380.
- [23] D. Frizzo, F. Borsatti, A. Arcudi, A. De Moliner, R. Oboe, and G. A. Susto, "Interpretable data-driven anomaly detection in industrial processes with exiffi," in *2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*. IEEE, 2024, pp. 595–600.
- [24] D. Neupane, M. R. Bouadjenek, R. Dazeley, and S. Aryal, "Data-driven machinery fault diagnosis: A comprehensive review," *Neurocomputing*, vol. 627, p. 129 588, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S09252321225002607>
- [25] D. M. Hawkins, *Identification of outliers*. Springer, 1980, vol. 11.
- [26] R. Foorthuis, "A typology of data anomalies," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations*, J. Medina, M. Ojeda-Aciego, J. L. Verdegay, D. A. Pelta, I. P. Cabrera, B. Bouchon-Meunier, and R. R. Yager, Eds. Cham: Springer International Publishing, 2018, pp. 26–38.
- [27] —, "On the nature and types of anomalies: a review of deviations in data," *International Journal of Data Science and Analytics*, vol. 12, no. 4, pp. 297–331, Oct. 2021. [Online]. Available: <https://doi.org/10.1007/s41060-021-00265-1>

- [28] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, Jul. 2009. [Online]. Available: <https://doi.org/10.1145/1541880.1541882>
- [29] V. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, Oct. 2004. [Online]. Available: <https://doi.org/10.1023/B:AIRE.0000045502.10941.a9>
- [30] M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko, "A review of novelty detection," *Signal Processing*, vol. 99, pp. 215–249, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S016516841300515X>
- [31] L. Ruff, J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, and K.-R. Müller, "A unifying review of deep and shallow anomaly detection," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 756–795, 2021.
- [32] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 93–104.
- [33] S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos, "Loc: Fast outlier detection using the local correlation integral," in *Proceedings 19th international conference on data engineering (Cat. No. 03CH37405)*. IEEE, 2003, pp. 315–326.
- [34] L. J. Latecki, A. Lazarevic, and D. Pokrajac, "Outlier detection with kernel density functions," in *International workshop on machine learning and data mining in pattern recognition*. Springer, 2007, pp. 61–75.
- [35] W. Hu, J. Gao, B. Li, O. Wu, J. Du, and S. Maybank, "Anomaly detection using local kernel density estimation and context-based regression," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 2, pp. 218–233, 2018.
- [36] E. Schubert, A. Zimek, and H.-P. Kriegel, "Generalized outlier detection with flexible kernel density estimates," in *Proceedings of the 2014 SIAM international conference on data mining*. SIAM, 2014, pp. 542–550.
- [37] K. Olsen, R. M. H. Lindrup, and M. Mørup, "Think global, adapt local: Learning locally adaptive k-nearest neighbor kernel density estimators," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2024, pp. 4114–4122.
- [38] M. Goldstein and A. Dengel, "Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm," *KI-2012: poster and demo track*, vol. 1, pp. 59–63, 2012.
- [39] M. Gebski and R. K. Wong, "An efficient histogram method for outlier detection," in *International Conference on Database Systems for Advanced Applications*. Springer, 2007, pp. 176–187.
- [40] T. Pevn' y, "Loda: Lightweight on-line detector of anomalies," *Machine Learning*, vol. 102, no. 2, pp. 275–304, 2016.
- [41] Z. Li, Y. Zhao, N. Botta, C. Ionescu, and X. Hu, "Copod: copulabased outlier detection," in *2020 IEEE international conference on data mining (ICDM)*. IEEE, 2020, pp. 1118–1123.
- [42] W. Liu, D. Cui, Z. Peng, and J. Zhong, "Outlier detection algorithm based on gaussian mixture model," in *2019 IEEE international conference on power, intelligent computing and systems (ICPICS)*. IEEE, 2019, pp. 488–492.
- [43] S. Guan, Z. He, S. Ma, and M. Gao, "Conditional normalizing flow for multivariate time series anomaly detection," *ISA transactions*, vol. 143, pp. 231–243, 2023.
- [44] B. Nachman and D. Shih, "Anomaly detection with density estimation," *Physical Review D*, vol. 101, no. 7, p. 075 042, 2020.
- [45] A. Ryzhikov, M. Borisyak, A. Ustyuzhanin, and D. Derkach, "Nfad: fixing anomaly detection using normalizing flows," *PeerJ Computer Science*, vol. 7, p. e757, 2021.
- [46] B. Rosenhahn and C. Hirche, "Quantum normalizing flows for anomaly detection," *Physical Review A*, vol. 110, no. 2, p. 022 443, 2024.
- [47] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *International conference on learning representations*, 2018.
- [48] G. Zamberg, M. Salhov, O. Lindenbaum, and A. Averbuch, "Tabadm: Unsupervised tabular anomaly detection with diffusion models," *arXiv preprint arXiv:2307.12336*, 2023.
- [49] Y. Huang, Z. Zhang, M. Liao, Y. Tan, and S. Zhou, "A hybrid distanced-based outlier detection approach," in *2012 International Conference on Systems and Informatics (ICSAI2012)*. IEEE, 2012, pp. 2212–2216.
- [50] J. Mei, M. Liu, Y.-F. Wang, and H. Gao, "Learning a mahalanobis distance-based dynamic time warping measure for multivariate time series classification," *IEEE transactions on Cybernetics*, vol. 46, no. 6, pp. 1363–1374, 2015.
- [51] J. R. Wells, K. M. Ting, and T. Washio, "Linearn: A new approach to nearest neighbour density estimator," *Pattern Recognition*, vol. 47, no. 8, pp. 2702–2720, 2014.
- [52] K. M. Ting, Y. Zhu, M. Carman, Y. Zhu, and Z.-H. Zhou, "Overcoming key weaknesses of distance-based neighbourhood methods using a data dependent dissimilarity measure," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 1205–1214.
- [53] H.-P. Kriegel, M. Schubert, and A. Zimek, "Angle-based outlier detection in high-dimensional data," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2008, pp. 444–452.
- [54] A. Lohrer, D. Kazempour, M. Hünemörder, and P. Kröger, "Comadout—a robust outlier detection algorithm based on comad," *Machine Learning*, vol. 113, no. 10, pp. 8061–8135, 2024.
- [55] Y. Hou, Z. Chen, M. Wu, C.-S. Foo, X. Li, and R. M. Shubair, "Mahalanobis distance based adversarial network for anomaly detection," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 3192–3196.
- [56] M. Sun, Y. Su, S. Zhang, Y. Cao, Y. Liu, D. Pei, W. Wu, Y. Zhang, X. Liu, and J. Tang, "Ctf: Anomaly detection in high-dimensional time series with coarse-to-fine model transfer," in *IEEE INFOCOM 2021-IEEE conference on computer communications*. IEEE, 2021, pp. 1–10.
- [57] C. Qiu, T. Pfommer, M. Kloft, S. Mandt, and M. Rudolph, "Neural transformation learning for deep anomaly detection beyond images," in *International conference on machine learning*. PMLR, 2021, pp. 8703–8714.
- [58] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth IEEE international conference on data mining*. IEEE, 2008, pp. 413–422.
- [59] J. Lesouple, C. Baudoin, M. Spigai, and J.-Y. Tournet, "Generalized isolation forest for anomaly detection," *Pattern Recognition Letters*, vol. 149, pp. 109–119, 2021.
- [60] J. Choudhury, P. Ky, Y. Ren, and C. Shi, "Hypersphere for branching node for the family of isolation forest algorithms," in *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, 2021, pp. 418–423.
- [61] T. R. Bandaragoda, K. M. Ting, D. Albrecht, F. T. Liu, Y. Zhu, and J. R. Wells, "Isolation-based anomaly detection using nearest-neighbor ensembles," *Computational Intelligence*, vol. 34, no. 4, pp. 968–998, 2018.
- [62] T. Liu, Z.-Y. Zhou, and L. Yang, "Layered isolation forest: A multi-level subspace algorithm for improving isolation forest," *Neurocomputing*, vol. 581, p. 127 525, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:268334315>
- [63] F. Leveni, G. W. Cassales, B. Pfahringer, A. Bifet, and G. Boracchi, "Online isolation forest," *ArXiv*, vol. abs/2505.09593, 2025. [Online]. Available: <https://api.semanticscholar.org/CorpusID:278602380>
- [64] A. Putina, M. Sozio, D. Rossi, and J. M. Navarro, "Random histogram forest for unsupervised anomaly detection," *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 1226–1231, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:228195076>
- [65] G. Pang, K. M. Ting, and D. W. Albrecht, "Lesinn: Detecting anomalies by identifying least similar nearest neighbours," *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 623–630, 2015. [Online]. Available: <https://api.semanticscholar.org/CorpusID:18178448>
- [66] Y. Zhao, M. K. Hryniewicki, Z. Nasrullah, and Z. Li, "Lscp: Locally selective combination in parallel outlier ensembles," *ArXiv*, vol. abs/1812.01528, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:54445467>
- [67] H. Sarvari, C. Domeniconi, and G. Stilo, "Graph-based selective outlier ensembles," *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, 2018. [Online]. Available: <https://api.semanticscholar.org/CorpusID:4903507>
- [68] B. Wang and Z. Mao, "A dynamic ensemble outlier detection model based on an adaptive k-nearest neighbor rule," *Inf. Fusion*, vol. 63, pp. 30–40, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:219499716>

- [69] H. Xu, G. Pang, Y. Wang, and Y. Wang, "Deep isolation forest for anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, pp. 12 591–12 604, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:249642184>
- [70] D. Chakraborty, V. Narayanan, and A. Ghosh, "Integration of deep feature extraction and ensemble learning for outlier detection," *Pattern Recognit.*, vol. 89, pp. 161–171, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:126415389>
- [71] A. Iqbal and R. Amin, "Time series forecasting and anomaly detection using deep learning," *Computers & Chemical Engineering*, vol. 182, p. 108 560, 2023.
- [72] H. Hoffmann, "Kernel pca for novelty detection," *Pattern Recognition*, vol. 40, no. 3, pp. 863–874, 2007. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320306003414>
- [73] Y.-J. Lee, Y.-R. Yeh, and Y.-C. F. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, pp. 1460–1470, 2013.
- [74] A. Adler, M. Elad, Y. Hel-Or, and E. Rivlin, "Sparse coding with anomaly detection," *Journal of Signal Processing Systems*, vol. 79, no. 2, pp. 179–188, 2015.
- [75] L. Gjorgiev and S. Gievska, "Time series anomaly detection with variational autoencoder using mahalanobis distance," in *ICT Innovations 2020. Machine Learning and Applications*, V. Dimitrova and I. Dimitrovski, Eds. Cham: Springer International Publishing, 2020, pp. 42–55.
- [76] Y. Su, Y. Zhao, C. Niu, R. Liu, W. Sun, and D. Pei, "Robust anomaly detection for multivariate time series through stochastic recurrent neural network," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ser. KDD '19*. New York, NY, USA: Association for Computing Machinery, 2019, p. 2828–2837. [Online]. Available: <https://doi.org/10.1145/3292500.33030672>
- [77] S. Guan, B. Zhao, Z. Dong, M. Gao, and Z. He, "Gtad: Graph and temporal neural network for multivariate time series anomaly detection," *Entropy*, vol. 24, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:249166131>
- [78] J. Audibert, P. Michiardi, F. Guyard, S. Marti, and M. A. Zuluaga, "Usad: Unsupervised anomaly detection on multivariate time series," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ser. KDD '20*. New York, NY, USA: Association for Computing Machinery, 2020, p. 3395–3404. [Online]. Available: <https://doi.org/10.1145/3394486.3403392>
- [79] X. Chen, L. Deng, F. Huang, C. Zhang, Z. Zhang, Y. Zhao, and K. Zheng, "Daemon: Unsupervised anomaly detection and interpretation for multivariate time series," in *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, 2021, pp. 2225–2230.
- [80] P. Reddy and A. Singh, "Anogan for tabular data: A novel approach to anomaly detection," in *HCI International 2024 – Late Breaking Posters*, C. Stephanidis, M. Antona, S. Ntoa, and G. Salvendy, Eds. Cham: Springer Nature Switzerland, 2025, pp. 319–330.
- [81] L. Chen, H. Jiang, L. Wang, J. Li, M. Yu, Y. Shen, and X. Du, "Generative adversarial synthetic neighbors-based unsupervised anomaly detection," *Scientific Reports*, vol. 15, no. 1, p. 16, 2025.
- [82] Y. F. Wang, Y. Jiong, G. P. Su, and Y. R. Qian, "A new outlier detection method based on optics," *Sustainable Cities and Society*, vol. 45, pp. 197–212, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2210670718318328>
- [83] P. K. Jain, M. S. Bajpai, and R. Pamula, "A modified dbscan algorithm for anomaly detection in time-series data with seasonality," *Int. Arab J. Inf. Technol.*, vol. 19, no. 1, pp. 23–28, 2022.
- [84] D. Krleža, B. Vrdoljak, and M. Brčić, "Statistical hierarchical clustering algorithm for outlier detection in evolving data streams," *Machine Learning*, vol. 110, no. 1, pp. 139–184, 2021.
- [85] R. Chitrakar and H. Chuanhe, "Anomaly detection using support vector machine classification with k-medoids clustering," in *2012 Third Asian Himalayas International Conference on Internet*, 2012, pp. 1–5.
- [86] U. Rashid, M. F. Saleem, S. Rasool, A. Abdullah, H. Mustafa, and A. Iqbal, "Anomaly detection using clustering (k-means with dbscan) and smo," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 02, 2024.
- [87] J. Li, H. Izakian, W. Pedrycz, and I. Jamal, "Clustering-based anomaly detection in multivariate time series data," *Applied Soft Computing*, vol. 100, p. 106919, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494620308577>
- [88] C. Aytekin, X. Ni, F. Cricri, and E. Aksu, "Clustering and unsupervised anomaly detection with l2 normalized deep auto-encoder representations," in *2018 International Joint Conference on Neural Networks (IJCNN)*, 2018, pp. 1–6.
- [89] G. S. Chadha, I. Islam, A. Schwung, and S. X. Ding, "Deep convolutional clustering-based time series anomaly detection," *Sensors*, vol. 21, no. 16, p. 5488, 2021.
- [90] H. Song, P. Li, and H. Liu, "Deep clustering based fair outlier detection," in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, ser. KDD '21*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1481–1489. [Online]. Available: <https://doi.org/10.1145/3447548.3467225>
- [91] A. G. Roselin, P. Nanda, S. Nepal, and X. He, "Intelligent anomaly detection for large network traffic with optimized deep clustering (odc) algorithm," *IEEE Access*, vol. 9, pp. 47 243–47 251, 2021.
- [92] J. Ma and S. Perkins, "Time-series novelty detection using one-class support vector machines," in *Proceedings of the International Joint Conference on Neural Networks*, 2003., vol. 3, 2003, pp. 1741–1745 vol. 3.
- [93] M. Ji and H.-J. Xing, "Adaptive-weighted one-class support vector machine for outlier detection," in *2017 29th Chinese Control And Decision Conference (CCDC)*, 2017, pp. 1766–1771.
- [94] S. Goyal, A. Raghunathan, M. Jain, H. V. Simhadri, and P. Jain, "DROCC: Deep robust one-class classification," in *Proceedings of the 37th International Conference on Machine Learning, ser. Proceedings of Machine Learning Research*, H. D. III and A. Singh, Eds., vol. 119. PMLR, 13–18 Jul 2020, pp. 3711–3721. [Online]. Available: <https://proceedings.mlr.press/v119/goyal20c.html>
- [95] L. Shen, Z. Li, and J. Kwok, "Timeseries anomaly detection using temporal hierarchical one-class network," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33. Curran Associates, Inc., 2020, pp. 13 016–13 026. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/97e401a02082021fd24957f852e0e475-Paper.pdf
- [96] H. Xu, Y. Wang, S. Jian, Q. Liao, Y. Wang, and G. Pang, "Calibrated one-class classification for unsupervised time series anomaly detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 5723–5736, 2024.
- [97] Z. Yang, X. Wang, Y. Chen, H. Dou, and H. Sang, "Robust deep one-class classification time series anomaly detection," *Computers, Materials & Continua*, vol. 83, no. 3, pp. 5181–5197, 2025. [Online]. Available: <http://www.techscience.com/cmc/v83n3/60982>
- [98] G. Zhong, F. Liu, J. Jiang, B. Wang, and C. P. Chen, "Refining one-class representation: A unified transformer for unsupervised time-series anomaly detection," *Information Sciences*, vol. 656, p. 119 914, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0020025523014998>
- [99] X. Yang and X. Li, "Atdad: One-class adversarial learning for tabular data anomaly detection," *Computers & Security*, vol. 134, p. 103 449, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167404823003590>
- [100] J. Ye, Z. Tan, Y. Hu, X. Yang, G. Cheng, and K. Huang, "Disentangling tabular data towards better one-class anomaly detection," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 12, pp. 13 061–13 068, Apr. 2025. [Online]. Available: <https://ojs.aaai.org/index.php/AAAI/article/view/33425>
- [101] E. H. M. Pena, M. V. O. de Assis, and M. L. Proença, "Anomaly detection using forecasting methods arima and hwsd," in *2013 32nd International Conference of the Chilean Computer Science Society (SCCC)*, 2013, pp. 63–66.
- [102] A. M. Bianco, M. García Ben, E. J. Martínez, and V. J. Yohai, "Outlier detection in regression models with arima errors using robust estimates," *Journal of Forecasting*, vol. 20, no. 8, pp. 565–579, 2001. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.768>
- [103] V. Kozitsin, I. Katser, and D. Lakontsev, "Online forecasting and anomaly detection based on the arima model," *Applied Sciences*, vol. 11, no. 7, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/7/3194>
- [104] P. Malhotra, L. Vig, G. M. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *The European Symposium on Artificial Neural Networks*, 2015. [Online]. Available: <https://api.semanticscholar.org/CorpusID:43680425>

- [105] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "Deepant: A deep learning approach for unsupervised anomaly detection in time series," *IEEE Access*, vol. 7, pp. 1991–2005, 2019.
- [106] S. Ahmad, A. Lavin, S. Purdy, and Z. Agha, "Unsupervised real-time anomaly detection for streaming data," *Neurocomputing*, vol. 262, pp. 134–147, 2017, online Real-Time Learning Strategies for Data Streams. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231217309864>
- [107] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, 2021, pp. 4027–4035.
- [108] B. Koltai, A. Gazdag, and G. Ács, "Improving CAN anomaly detection with correlation-based signal clustering," *Infocommunications Journal*, vol. 15, no. 4, pp. 17–25, 2023.
- [109] S. Tuli, G. Casale, and N. R. Jennings, "Tranad: Deep transformer networks for anomaly detection in multivariate time series data," *Proc. VLDB Endow.*, vol. 15, pp. 1201–1214, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:246035402>
- [110] M. Zhao, H. Peng, L. Li, and Y. Ren, "Graph attention network and informer for multivariate time series anomaly detection," *Sensors*, vol. 24, no. 5, 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/5/1522>
- [111] T. Mathonsi and T. L. v. Zyl, "Multivariate anomaly detection based on prediction intervals constructed using deep learning," *Neural Computing and Applications*, vol. 37, no. 2, pp. 707–721, 2025.
- [112] Z. Li, Y. Zhu, and M. Van Leeuwen, "A survey on explainable anomaly detection," *ACM Trans. Knowl. Discov. Data*, vol. 18, no. 1, Sep. 2023. [Online]. Available: <https://doi.org/10.1145/3609333>
- [113] E. Brusa, L. Cibrario, C. Delprete, and L. G. Di Maggio, "Explainable ai for machine fault diagnosis: understanding features' contribution in machine learning models for industrial condition monitoring," *Applied Sciences*, vol. 13, no. 4, p. 2038, 2023.
- [114] K. Jang, K. E. S. Pilario, N. Lee, I. Moon, and J. Na, "Explainable artificial intelligence for fault diagnosis of industrial processes," *IEEE Transactions on Industrial Informatics*, vol. 21, no. 1, pp. 4–11, 2025.
- [115] M. T. Ribeiro, S. Singh, and C. Guestrin, "why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1135–1144. [Online]. Available: <https://doi.org/10.1145/2939672.2939778>
- [116] A. K. M. Nor, S. R. Pedapati, M. Muhammad, and V. Leiva, "Abnormality detection and failure prediction using explainable bayesian deep learning: Methodology and case study with industrial data," *Mathematics*, vol. 10, no. 4, 2022. [Online]. Available: <https://www.mdpi.com/2227-7390/10/4/554>
- [117] K. Patra, R. N. Sethi, and D. K. Behera, "Anomaly detection in rotating machinery using autoencoders based on bidirectional lstm and gru neural networks," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 30, no. 4, pp. 1637–1653, 2022.
- [118] S. Ahmad, K. Styp-Rekowski, S. Nedelkoski, and O. Kao, "Autoencoder-based condition monitoring and anomaly detection method for rotating machines," in *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 2020, pp. 4093–4102.
- [119] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. Duarte, "An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery," *Mechanical Systems and Signal Processing*, vol. 163, p. 108 105, 2022.
- [120] T. Reeber, J. Henninger, N. Weingarz, P. M. Simon, M. Berndt, M. Glatt, B. Kirsch, R. Eisseler, J. C. Aurich, and H.-C. Möhring, "Tool condition monitoring in drilling processes using anomaly detection approaches based on control internal data," *Procedia Cirp*, vol. 121, pp. 216–221, 2024.
- [121] C. Shi, B. Luo, H. Li, B. Li, X. Mao, and F. Peng, "Anomaly detection via unsupervised learning for tool breakage monitoring," *International Journal of Machine Learning and Computing*, vol. 6, no. 5, p. 256, 2016.
- [122] J. Shim, T. Joung, S. Lee, and J.-I. Ha, "Audio data-driven anomaly detection for induction motor based on generative adversarial networks," in *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, 2022, pp. 1–5.
- [123] J. Shim, G. C. Lim, and J.-I. Ha, "Unsupervised anomaly detection for electric drives based on variational auto-encoder," in *2022 IEEE applied power electronics conference and exposition (appec)*. IEEE, 2022, pp. 1703–1708.
- [124] J. M. Barrera, A. Reina, A. Mate, and J. C. Trujillo, "Fault detection and diagnosis for industrial processes based on clustering and autoencoders: a case of gas turbines," *International Journal of Machine Learning and Cybernetics*, vol. 13, no. 10, pp. 3113–3129, 2022.
- [125] M. Carletti, C. Masiero, A. Beghi, and G. A. Susto, "A deep learning approach for anomaly detection with industrial time series data: a refrigerators manufacturing case study," *Procedia Manufacturing*, vol. 38, pp. 233–240, 2019.
- [126] J. Goh, S. Adepu, M. Tan, and Z. S. Lee, "Anomaly detection in cyber physical systems using recurrent neural networks," in *2017 IEEE 18th international symposium on high assurance systems engineering (HASE)*. IEEE, 2017, pp. 140–145.
- [127] M. M. Aboelwafa, K. G. Seddik, M. H. Eldefrawy, Y. Gadallah, and M. Gidlund, "A machine-learning-based technique for false data injection attacks detection in industrial iot," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8462–8471, 2020.
- [128] W.-H. Choi and J. Kim, "Unsupervised learning approach for anomaly detection in industrial control systems," *Applied System Innovation*, vol. 7, no. 2, p. 18, 2024.
- [129] D. Benka, D. Horváth, L. Špendla, G. Gašpar, and M. Strémy, "Machine learning-based detection of anomalies, intrusions and threats in industrial control systems," *IEEE Access*, 2025.
- [130] C. Goetz and B. Humm, "Decentralized real-time anomaly detection in cyber-physical production systems under industry constraints," *Sensors*, vol. 23, no. 9, p. 4207, 2023.
- [131] C. Sánchez-Zas, X. Larriva-Novo, V. A. Villagrà, M. S. Rodrigo, and J. I. Moreno, "Design and evaluation of unsupervised machine learning models for anomaly detection in streaming cybersecurity logs," *Mathematics*, vol. 10, no. 21, p. 4043, 2022.
- [132] D. Leite, E. Andrade, D. Rativa, and A. M. Maciel, "Fault detection and diagnosis in industry 4.0: a review on challenges and opportunities," *Sensors (Basel, Switzerland)*, vol. 25, no. 1, p. 60, 2024.
- [133] H. Li, M. Jia, and Z. Mao, "Dynamic reconstruction principal component analysis for process monitoring and fault detection in the cold rolling industry," *Journal of Process Control*, vol. 128, p. 103 010, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0959152423000951>
- [134] O. AlShorman, F. Alkahatni, M. Masadeh, M. Irfan, A. Glowacz, F. Althobiani, J. Kozik, and W. Glowacz, "Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study," *Advances in Mechanical Engineering*, vol. 13, no. 2, p. 1687814021996915, 2021.
- [135] D. Zhang, E. Stewart, M. Entezami, C. Roberts, and D. Yu, "Intelligent acoustic-based fault diagnosis of roller bearings using a deep graph convolutional network," *Measurement*, vol. 156, p. 107585, 2020.
- [136] G. Niu, X. Dong, and Y. Chen, "Motor fault diagnostics based on current signatures: A review," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–19, 2023.
- [137] D. Ding, Q.-L. Han, Y. Xiang, X. Ge, and X.-M. Zhang, "A survey on security control and attack detection for industrial cyber-physical systems," *Neurocomputing*, vol. 275, pp. 1674–1683, 2018.
- [138] A. Garcés-Jiménez, A. Rodrigues, J. M. Gomez-Pulido, D. Raposo, J. A. Gomez-Pulido, J. S. Silva, and F. Boavida, "Industrial internet of things embedded devices fault detection and classification. a case study," *Internet of Things*, vol. 25, p. 101 042, 2024.
- [139] M. E. H. Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE transactions on industrial electronics*, vol. 47, no. 5, pp. 984–993, 2002.
- [140] J. Sim, S. Kim, H. J. Park, and J.-H. Choi, "A tutorial for feature engineering in the prognostics and health management of gears and bearings," *Applied Sciences*, vol. 10, no. 16, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/16/5639>
- [141] C. Zhang, W. Yin, C. Zhang, K. Peng, and X. Zhang, "A kpi-related fault diagnosis method for multimode manufacturing processes based on supervised minimal gated unit and sparse broad learning system," *Applied Soft Computing*, p. 113 679, 2025.
- [142] D. Leite, A. Martins Jr, D. Rativa, J. F. De Oliveira, and A. M. Maciel, "An automated machine learning approach for real-time fault detection and diagnosis," *Sensors*, vol. 22, no. 16, p. 6138, 2022.
- [143] J. Yao, C. Liu, K. Song, C. Feng, and D. Jiang, "Fault diagnosis of planetary gearbox based on acoustic signals," *Applied Acoustics*, vol. 181, p. 108 151, 2021.
- [144] S. Sekiou, A. Behloul, R. Nait-Said, and Z. Chiremsel, "Alarms prediction and classification in industrial processes using supervised machine learning techniques: A case study in an algerian gas plant," *Computers & Chemical Engineering*, p. 109 378, 2025.

- [145] L. Arya and G. P. Gupta, "Ensemble filter-based feature selection model for cyber attack detection in industrial internet of things," in *2023 9th International conference on advanced computing and communication systems (ICACCS)*, vol. 1. IEEE, 2023, pp. 834–840.
- [146] L. Wan, K. Gong, G. Zhang, X. Yuan, C. Li, and X. Deng, "An efficient rolling bearing fault diagnosis method based on spark and improved random forest algorithm," *Ieee Access*, vol. 9, pp. 37 866–37 882, 2021.
- [147] W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "Cyber-attacks detection in industrial systems using artificial intelligence-driven methods," *International journal of critical infrastructure protection*, vol. 38, p. 100542, 2022.
- [148] A. N. Saberi, A. Belahcen, J. Sobra, and T. Vaimann, "Lightgbmbased fault diagnosis of rotating machinery under changing working conditions using modified recursive feature elimination," *IEEE Access*, vol. 10, pp. 81 910–81 925, 2022.
- [149] M. C. Gomes, L. C. Brito, M. B. da Silva, and M. A. V. Duarte, "Tool wear monitoring in micromilling using support vector machine with vibration and sound sensors," *Precision Engineering*, vol. 67, pp. 137–151, 2021.
- [150] Y. E. Karabacak, N. G. Özmen, and L. Gümüşel, "Intelligent worm gearbox fault diagnosis under various working conditions using vibration, sound and thermal features," *Applied Acoustics*, vol. 186, p. 108 463, 2022.
- [151] J. Wei, H. Huang, L. Yao, Y. Hu, Q. Fan, and D. Huang, "New imbalanced bearing fault diagnosis method based on sample-characteristic oversampling technique (scote) and multi-class ls-svm," *Applied Soft Computing*, vol. 101, p. 107 043, 2021.
- [152] K. Ravikumar, C. Madhusudana, H. Kumar, and K. Gangadharan, "Classification of gear faults in internal combustion (ic) engine gearbox using discrete wavelet transform features and k star algorithm," *Engineering Science and Technology, an International Journal*, vol. 30, p. 101 048, 2022.
- [153] G. Rayjade, A. Bhagure, P. B. Kushare, R. Bhandare, V. Matsagar, and A. Chaudhari, "Performance evaluation of machine learning algorithms and impact of activation functions in artificial neural network classifier for bearing fault diagnosis," *Journal of Vibration and Control*, vol. 31, no. 9–10, pp. 1859–1873, 2025.
- [154] S.-K. S. Fan, C.-Y. Hsu, D. ming Tsai, F. He, and C.-C. Cheng, "Data-driven approach for fault detection and diagnostic in semiconductor manufacturing," *IEEE Transactions on Automation Science and Engineering*, vol. 17, pp. 1925–1936, 2020. [Online]. Available: <https://api.semanticscholar.org/CorpusID:218776407>
- [155] S. Gawde, S. Patil, S. Kumar, P. Kamat, and K. Kotecha, "An explainable predictive maintenance strategy for multi-fault diagnosis of rotating machines using multi-sensor data fusion," *Decision Analytics Journal*, vol. 10, p. 100 425, 2024.
- [156] J. Jiang, W. Li, Z. Wen, Y. Bie, H. Schwarz, and C. Zhang, "Series arc fault detection based on random forest and deep neural network," *IEEE Sensors Journal*, vol. 21, no. 15, pp. 17 171–17 179, 2021.
- [157] M. Hosseinpour-Zarnaq, M. Omid, and E. Biabani-Aghdam, "Fault diagnosis of tractor auxiliary gearbox using vibration analysis and random forest classifier," *Information Processing in Agriculture*, vol. 9, no. 1, pp. 60–67, 2022.
- [158] Z. Wang, J. Guo, P. Xia, L. Wang, C. Zhang, Q. Leng, and K. Zheng, "Feature selection for chillers fault diagnosis from the perspectives of machine learning and field application," *Energy and Buildings*, vol. 307, p. 113 937, 2024.
- [159] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, 2019.
- [160] W. Xie, Z. Li, Y. Xu, P. Gardoni, and W. Li, "Evaluation of different bearing fault classifiers in utilizing cnn feature extraction ability," *Sensors*, vol. 22, no. 9, 2022. [Online]. Available: <https://www.mdpi.com/1424-8220/22/9/3314>
- [161] S. Esakimuthu Pandarakone, Y. Mizuno, and H. Nakamura, "A comparative study between machine learning algorithm and artificial intelligence neural network in detecting minor bearing fault of induction motors," *Energies*, vol. 12, no. 11, p. 2105, 2019.
- [162] H. Ruan, B. Dorneanu, H. Arellano-Garcia, P. Xiao, and L. Zhang, "Deep learning-based fault prediction in wireless sensor network embedded cyber-physical systems for industrial processes," *IEEE Access*, vol. 10, pp. 10 867–10 879, 2022.
- [163] Z. Jin, D. Chen, D. He, Y. Sun, and X. Yin, "Bearing fault diagnosis based on vmd and improved cnn," *Journal of Failure Analysis and Prevention*, vol. 23, no. 1, pp. 165–175, 2023.
- [164] Z. Dong, D. Zhao, and L. Cui, "An intelligent bearing fault diagnosis framework: One-dimensional improved self-attention-enhanced cnn and empirical wavelet transform," *Nonlinear Dynamics*, vol. 112, no. 8, pp. 6439–6459, 2024.
- [165] S. Zhao, Y. Duan, N. Roy, and B. Zhang, "A deep learning methodology based on adaptive multiscale cnn and enhanced highway lstm for industrial process fault diagnosis," *Reliability engineering & system safety*, vol. 249, p. 110 208, 2024.
- [166] T. Li, C. Sun, O. Fink, Y. Yang, X. Chen, and R. Yan, "Filter-informed spectral graph wavelet networks for multiscale feature extraction and intelligent fault diagnosis," *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 506–518, 2023.
- [167] B. A. Tama, M. Vania, S. Lee, and S. Lim, "Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals," *Artificial Intelligence Review*, vol. 56, no. 5, pp. 4667–4709, 2023.
- [168] Y. Hou, H. Huang, Z. Chen, X. Cai, Y. Han, and X. Zhai, "A novel multi-scale adaptive feature fusion framework for accurate and efficient bearing fault diagnosis," *Signal, Image and Video Processing*, vol. 19, no. 15, pp. 1–12, 2025.
- [169] X. Zhao, J. Yao, W. Deng, P. Ding, Y. Ding, M. Jia, and Z. Liu, "Intelligent fault diagnosis of gearbox under variable working conditions with adaptive intra-class and inter-class convolutional neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 6339–6353, 2022.
- [170] B. Guan, X. Bao, H. Qiu, and D. Yang, "Enhancing bearing fault diagnosis using motor current signals: A novel approach combining time shifting and causal convnets," *Measurement*, vol. 226, p. 114 049, 2024.
- [171] X. Wang, D. Mao, and X. Li, "Bearing fault diagnosis based on vibroacoustic data fusion and 1d-cnn network," *Measurement*, vol. 173, p. 108 518, 2021.
- [172] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognitive Systems Research*, vol. 53, pp. 42–50, 2019.
- [173] C.-F. Chien, W.-T. Hung, and E. T.-Y. Liao, "Redefining monitoring rules for intelligent fault detection and classification via cnn transfer learning for smart manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 35, no. 2, pp. 158–165, 2022.
- [174] J. Jiao, M. Zhao, J. Lin, and K. Liang, "A comprehensive review on convolutional neural network in machine fault diagnosis," *Neurocomputing*, vol. 417, pp. 36–63, 2020.
- [175] Y. Cheng, M. Lin, J. Wu, H. Zhu, and X. Shao, "Intelligent fault diagnosis of rotating machinery based on continuous wavelet transform-local binary convolutional neural network," *Knowledge-Based Systems*, vol. 216, p. 106 796, 2021.
- [176] Y.-M. Hsueh, V. R. Ittangihai, W.-B. Wu, H.-C. Chang, and C.-C. Kuo, "Fault diagnosis system for induction motors by cnn using empirical wavelet transform," *Symmetry*, vol. 11, no. 10, p. 1212, 2019.
- [177] B. Peng, H. Xia, X. Lv, M. Annor-Nyarko, S. Zhu, Y. Liu, and J. Zhang, "An intelligent fault diagnosis method for rotating machinery based on data fusion and deep residual neural network," *Applied Intelligence*, vol. 52, no. 3, pp. 3051–3065, 2022.
- [178] Q. Zhang and L. Deng, "An intelligent fault diagnosis method of rolling bearings based on short-time fourier transform and convolutional neural network," *Journal of Failure Analysis and Prevention*, vol. 23, no. 2, pp. 795–811, 2023.
- [179] A. Mohammad-Alikhani, E. Jamshidpour, S. Dhale, M. Akrami, S. Pardhan, and B. Nahid-Mobarakeh, "Fault diagnosis of electric motors by a channel-wise regulated cnn and differential of stft," *IEEE Transactions on Industry Applications*, 2025.
- [180] H. Wang, Z. Liu, D. Peng, and Z. Cheng, "Attention-guided joint learning cnn with noise robustness for bearing fault diagnosis and vibration signal denoising," *ISA transactions*, vol. 128, pp. 470–484, 2022.
- [181] G. Niu, E. Liu, X. Wang, P. Ziehl, and B. Zhang, "Enhanced discriminate feature learning deep residual cnn for multitask bearing fault diagnosis with information fusion," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 762–770, 2022.
- [182] J. Miao, C. Deng, H. Zhang, and Q. Miao, "Interactive channel attention for rotating component fault detection with strong noise and limited data," *Applied Soft Computing*, vol. 138, p. 110 171, 2023.
- [183] B. Hu, J. Liu, and Y. Xu, "A novel multi-scale convolutional neural network incorporating multiple attention mechanisms for bearing fault diagnosis," *Measurement*, vol. 242, p. 115 927, 2025.

- [184] Y. Chen and R. Zhang, "Hybrid dual-channel attention cnn and extreme gradient boosting for industrial process model development and fault diagnosis," *IEEE Internet of Things Journal*, 2025.
- [185] K. Lu, C. Chen, T. Wang, L. Cheng, and J. Qin, "Fault diagnosis of industrial robot based on dual-module attention convolutional neural network," *Autonomous Intelligent Systems*, vol. 2, no. 1, p. 12, 2022.
- [186] Y. Keshun, W. Puzhou, and G. Yingkui, "Toward efficient and interpretative rolling bearing fault diagnosis via quadratic neural network with bi-lstm," *IEEE Internet of Things Journal*, vol. 11, no. 13, pp. 23 002–23 019, 2024.
- [187] F. Dao, Y. Zeng, and J. Qian, "Fault diagnosis of hydro-turbine via the incorporation of bayesian algorithm optimized cnn-lstm neural network," *Energy*, vol. 290, p. 130326, 2024.
- [188] H. Bilal, M. S. Obaidat, M. S. Aslam, J. Zhang, B. Yin, and K. Mahmood, "Online fault diagnosis of industrial robot using iort and hybrid deep learning techniques: An experimental approach," *IEEE Internet of Things Journal*, vol. 11, no. 19, pp. 31 422–31 437, 2024.
- [189] G. Fu, Q. Wei, Y. Yang, and C. Li, "Bearing fault diagnosis based on cnn-bilstm and residual module," *Measurement Science and Technology*, vol. 34, no. 12, p. 125050, 2023.
- [190] X. Xu, Z. Tao, W. Ming, Q. An, and M. Chen, "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, p. 108 086, 2020.
- [191] J. Li, Y. Liu, and Q. Li, "Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method," *Measurement*, vol. 189, p. 110 500, 2022.
- [192] Y. Jin, L. Hou, and Y. Chen, "A time series transformer based method for the rotating machinery fault diagnosis," *Neurocomputing*, vol. 494, pp. 379–395, 2022.
- [193] S. Han, H. Shao, J. Cheng, X. Yang, and B. Cai, "Convformer-nse: A novel end-to-end gearbox fault diagnosis framework under heavy noise using joint global and local information," *IEEE/ASME Transactions on Mechatronics*, vol. 28, no. 1, pp. 340–349, 2022.
- [194] Y. Han, F. Zhang, Z. Li, Q. Wang, C. Li, P. Lai, T. Li, F. Teng, and Z. Jin, "Mt-convformer: a multi-task bearing fault diagnosis method using a combination of cnn and transformer," *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [195] Y. Hou, J. Wang, Z. Chen, J. Ma, and T. Li, "Diagnosisformer: An efficient rolling bearing fault diagnosis method based on improved transformer," *Engineering Applications of Artificial Intelligence*, vol. 124, p. 106 507, 2023.
- [196] X. Luo, H. Wang, T. Han, and Y. Zhang, "Fft-trans: Enhancing robustness in mechanical fault diagnosis with fourier transform-based transformer under noisy conditions," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–12, 2024.
- [197] Y. Ding, M. Jia, Q. Miao, and Y. Cao, "A novel time-frequency transformer based on self-attention mechanism and its application in fault diagnosis of rolling bearings," *Mechanical Systems and Signal Processing*, vol. 168, p. 108 616, 2022.
- [198] X. Tang, Z. Xu, and Z. Wang, "A novel fault diagnosis method of rolling bearing based on integrated vision transformer model," *Sensors*, vol. 22, no. 10, p. 3878, 2022.
- [199] C. Li, J. Chen, C. Yang, J. Yang, Z. Liu, and P. Davari, "Convolutional neural network-based transformer fault diagnosis using vibration signals," *Sensors*, vol. 23, no. 10, p. 4781, 2023.
- [200] S. Wang, Q. Xu, S. Zhu, and B. Wang, "Making transformer hear better: Adaptive feature enhancement based multi-level supervised acoustic signal fault diagnosis," *Expert Systems with Applications*, vol. 264, p. 125 736, 2025.
- [201] W. Sun, R. Yan, R. Jin, J. Xu, Y. Yang, and Z. Chen, "Liteformer: a lightweight and efficient transformer for rotating machine fault diagnosis," *IEEE Transactions on Reliability*, vol. 73, no. 2, pp. 1258–1269, 2023.
- [202] Y. Xiao, H. Shao, J. Wang, S. Yan, and B. Liu, "Bayesian variational transformer: A generalizable model for rotating machinery fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 207, p. 110 936, 2024.
- [203] L.-z. Wang, J.-f. Chi, Y.-q. Ding, H.-y. Yao, Q. Guo, and H.-q. Yang, "Transformer fault diagnosis method based on smote and ngo-gbdt," *Scientific Reports*, vol. 14, no. 1, p. 7179, 2024.
- [204] H. Chen, J. Wei, H. Huang, L. Wen, Y. Yuan, and J. Wu, "Novel imbalanced fault diagnosis method based on generative adversarial networks with balancing serial cnn and transformer (bctgan)," *Expert Systems with Applications*, vol. 258, p. 125 171, 2024.
- [205] Z. Chen, J. Xu, C. Alippi, S. X. Ding, Y. Shardt, T. Peng, and C. Yang, "Graph neural network-based fault diagnosis: a review," *arXiv preprint arXiv:2111.08185*, 2021.
- [206] M. Jin, H. Y. Koh, Q. Wen, D. Zambon, C. Alippi, G. I. Webb, I. King, and S. Pan, "A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [207] T. Li, Z. Zhao, C. Sun, R. Yan, and X. Chen, "Multireceptive field graph convolutional networks for machine fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 12, pp. 12 739–12 749, 2020.
- [208] C. Li, L. Mo, and R. Yan, "Rolling bearing fault diagnosis based on horizontal visibility graph and graph neural networks," in *2020 international conference on sensing, measurement & data analytics in the era of artificial intelligence (ICSMD)*. IEEE, 2020, pp. 275–279.
- [209] K. Wen, R. Huang, D. Li, Z. Chen, and W. Li, "Gradient-based interpretable graph convolutional network for bearing fault diagnosis," in *2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2023, pp. 1–6.
- [210] W. Lu, W. Wang, X. Qin, and Z. Cai, "Enhancing fault diagnosis in mechanical systems with graph neural networks addressing class imbalance," *Mathematics*, vol. 12, no. 13, p. 2064, 2024.
- [211] M. Jia, J. Hu, Y. Liu, Z. Gao, and Y. Yao, "Topology-guided graph learning for process fault diagnosis," *Industrial & Engineering Chemistry Research*, vol. 62, no. 7, pp. 3238–3248, 2023.
- [212] D. Wu and J. Zhao, "Process topology convolutional network model for chemical process fault diagnosis," *Process Safety and Environmental Protection*, vol. 150, pp. 93–109, 2021.
- [213] D. Chen, R. Liu, Q. Hu, and S. X. Ding, "Interaction-aware graph neural networks for fault diagnosis of complex industrial processes," *IEEE Transactions on neural networks and learning systems*, vol. 34, no. 9, pp. 6015–6028, 2021.
- [214] R. Liu, Q. Zhang, D. Lin, W. Zhang, and S. X. Ding, "Causal intervention graph neural network for fault diagnosis of complex industrial processes," *Reliability Engineering & System Safety*, vol. 251, p. 110 328, 2024.
- [215] Y. Xu, C. Zhang, Y. Luo, W. Ke, Q.-X. Zhu, Y.-L. He, Y. Zhang, and M.-Q. Zhang, "Multi-attention spatio-temporal fusion graph neural network for chemical process fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, 2025.
- [216] G. Jiang, K. Shen, X. Liu, X. Cheng, and P. Xie, "Hierarchical spatiotemporal graph network for fault diagnosis of industrial processes," *IEEE Internet of Things Journal*, 2024.
- [217] Z. Wang, Z. Wu, X. Li, H. Shao, T. Han, and M. Xie, "Attention-aware temporal-spatial graph neural network with multi-sensor information fusion for fault diagnosis," *Knowledge-Based Systems*, vol. 278, p. 110 891, 2023.
- [218] J. Xu, S. Mo, Z. Jiang, Z. Chen, W. Gui, and H. Wang, "A novel positive-negative graph convolutional network-based fault diagnosis method with application to complex systems," *Neurocomputing*, vol. 600, p. 128 145, 2024.
- [219] W. Cai, D. Zhao, and T. Wang, "Multi-scale dynamic graph mutual information network for planet bearing health monitoring under imbalanced data," *Advanced Engineering Informatics*, vol. 64, p. 103 096, 2025.
- [220] T. Li, Z. Zhou, S. Li, C. Sun, R. Yan, and X. Chen, "The emerging graph neural networks for intelligent fault diagnostics and prognostics: A guideline and a benchmark study," *Mechanical Systems and Signal Processing*, vol. 168, p. 108 653, 2022.
- [221] Z. Zhang and L. Wu, "Graph neural network-based bearing fault diagnosis using granger causality test," *Expert Systems with Applications*, vol. 242, p. 122 827, 2024.
- [222] Z. Chen, J. Xu, T. Peng, and C. Yang, "Graph convolutional networkbased method for fault diagnosis using a hybrid of measurement and prior knowledge," *IEEE transactions on cybernetics*, vol. 52, no. 9, pp. 9157–9169, 2021.
- [223] Z. Yu, C. Zhang, and C. Deng, "An improved gnn using dynamic graph embedding mechanism: A novel end-to-end framework for rolling bearing fault diagnosis under variable working conditions," *Mechanical Systems and Signal Processing*, vol. 200, p. 110 534, 2023.
- [224] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.
- [225] T. Kipf, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

Anomaly and Fault Detection and Predictive Maintenance: a Review on Industrial Applications

- [226] W. Hamilton, Z. Ying, and J. Leskovec, "Inductive representation learning on large graphs," *Advances in neural information processing systems*, vol. 30, 2017.
- [227] C. Li, L. Mo, and R. Yan, "Fault diagnosis of rolling bearing based on whvg and gcn," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [228] C. Yang, K. Zhou, and J. Liu, "Supergraph: Spatial-temporal graph-based feature extraction for rotating machinery diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 4, pp. 4167–4176, 2021.
- [229] B. Zhou, X. Li, T. Liu, K. Xu, W. Liu, and J. Bao, "Causalkgpt: Industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing," *Advanced Engineering Informatics*, vol. 59, p. 102 333, 2024.
- [230] S. Marandi, Y.-S. Hu, and M. Modarres, "Complex system diagnostics using a knowledge graph-informed and large language model-enhanced framework," *Applied Sciences*, vol. 15, no. 17, p. 9428, 2025.
- [231] C. Xiao, X. Liu, A. Wulamu, and D. Zhang, "Kg-sr-llm: Knowledge-guided semantic representation and large language model framework for cross-domain bearing fault diagnosis," *Sensors*, vol. 25, no. 18, p. 5758, 2025.
- [232] L. Tao, H. Liu, G. Ning, W. Cao, B. Huang, and C. Lu, "Llm-based framework for bearing fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 224, p. 112 127, 2025.
- [233] J. Wang, T. Li, Y. Yang, S. Chen, and W. Zhai, "Diagllm: multimodal reasoning with large language model for explainable bearing fault diagnosis," *Science China Information Sciences*, vol. 68, no. 6, p. 160 103, 2025.
- [234] L. Lin, S. Zhang, S. Fu, and Y. Liu, "Fd-llm: Large language model for fault diagnosis of complex equipment," *Advanced Engineering Informatics*, vol. 65, p. 103 208, 2025.
- [235] S. Zheng, K. Pan, J. Liu, and Y. Chen, "Empirical study on finetuning pre-trained large language models for fault diagnosis of complex systems," *Reliability Engineering & System Safety*, vol. 252, p. 110 382, 2024.
- [236] X. Liao, C. Chen, Z. Wang, Y. Liu, T. Wang, and L. Cheng, "Large language model assisted fine-grained knowledge graph construction for robotic fault diagnosis," *Advanced Engineering Informatics*, vol. 65, p. 103 134, 2025.
- [237] Q. Zhang, C. Xu, J. Li, Y. Sun, J. Bao, and D. Zhang, "Llm-tsfd: An industrial time series human-in-the-loop fault diagnosis method based on a large language model," *Expert Systems with Applications*, vol. 264, p. 125 861, 2025.
- [238] X. Chen, Y. Lei, Y. Li, S. Parkinson, X. Li, J. Liu, F. Lu, H. Wang, Z. Wang, B. Yang et al., "Large models for machine monitoring and fault diagnostics: Opportunities, challenges, and future direction," *Journal of Dynamics, Monitoring and Diagnostics*, vol. 4, no. 2, pp. 76–90, 2025.
- [239] M. W. Hopwood, J. S. Stein, J. L. Braid, and H. P. Seigneur, "Physics-based method for generating fully synthetic iv curve training datasets for machine learning classification of pv failures," *Energies*, vol. 15, no. 14, p. 5085, 2022.
- [240] P. Seventekidis, D. Giagopoulos, A. Arailopoulos, and O. Markogiannaki, "Structural health monitoring using deep learning with optimal finite element model generated data," *Mechanical Systems and Signal Processing*, vol. 145, p. 106972, 2020.
- [241] W. Hao, L. Tan, X. Yang, D. Shi, M. Wang, G. Miao, and Y. Fan, "A physics-informed machine learning approach for notch fatigue evaluation of alloys used in aerospace," *International Journal of Fatigue*, vol. 170, p. 107 536, 2023.
- [242] Y. Wu, B. Sicard, and S. A. Gadsden, "Physics-informed machine learning: A comprehensive review on applications in anomaly detection and condition monitoring," *Expert Systems with Applications*, vol. 255, p. 124 678, 2024.
- [243] Z. Wang, Z. Zhou, W. Xu, C. Sun, and R. Yan, "Physics informed neural networks for fault severity identification of axial piston pumps," *Journal of Manufacturing Systems*, vol. 71, pp. 421–437, 2023.
- [244] C. Lai, I. Ahmed, E. Zio, W. Li, Y. Zhang, W. Yao, and J. Chen, "A multistage physics-informed neural network for fault detection in regulating valves of nuclear power plants," *Energies*, vol. 17, no. 11, p. 2647, 2024.
- [245] M. Parziale, L. Lomazzi, M. Giglio, and F. Cadini, "Physics-informed neural networks for the condition monitoring of rotating shafts," *Sensors*, vol. 24, no. 1, p. 207, 2023.
- [246] J. Shen, J. Chowdhury, S. Banerjee, and G. Terejanu, "Machine fault classification using hamiltonian neural networks," *arXiv preprint arXiv:2301.02243*, 2023.
- [247] S. Shen, H. Lu, M. Sadoughi, C. Hu, V. Nemani, A. Thelen, K. Webster, M. Darr, J. Sidon, and S. Kenny, "A physics-informed deep learning approach for bearing fault detection," *Engineering Applications of Artificial Intelligence*, vol. 103, p. 104 295, 2021.
- [248] Z. Xu, K. Zhao, J. Wang, and M. Bashir, "Physics-informed probabilistic deep network with interpretable mechanism for trustworthy mechanical fault diagnosis," *Adv. Eng. Informatics*, vol. 62, p. 102 806, 2024. [Online]. Available: <https://api.semanticscholar.org/CorpusID:272638356>
- [249] C. Dong, J. Tao, H. Sun, Q. Wei, H. Tan, and C. Liu, "Innovative fault diagnosis for axial piston pumps: A physics-informed neural network framework predicting pump flow ripple," *Mechanical Systems and Signal Processing*, vol. 225, p. 112 274, 2025.
- [250] C. Gao, Z. Wang, Y. Guo, H. Wang, and H. Yi, "Mpinet: multiscale physics-informed network for bearing fault diagnosis with small samples," *IEEE Transactions on Industrial Informatics*, 2024.
- [251] C. He, H. Shi, J. Si, and J. Li, "Physics-informed interpretable wavelet weight initialization and balanced dynamic adaptive threshold for intelligent fault diagnosis of rolling bearings," *Journal of Manufacturing Systems*, vol. 70, pp. 579–592, 2023.
- [252] Z. Shang, Z. Zhao, and R. Yan, "Denoising fault-aware wavelet network: A signal processing informed neural network for fault diagnosis," *Chinese Journal of Mechanical Engineering*, vol. 36, no. 1, p. 9, 2023.
- [253] C. He, H. Shi, X. Liu, and J. Li, "Interpretable physics-informed domain adaptation paradigm for cross-machine transfer diagnosis," *Knowledge-Based Systems*, vol. 288, p. 111 499, 2024.
- [254] T. Gao, H. Zhu, J. Wu, Q. Yuan, and Z. Li, "A fully interpretable convolutional neural network for intelligent fault diagnosis of rotating machinery," *Applied Soft Computing*, vol. 180, p. 113 472, 2025.
- [255] T. Li, Z. Zhao, C. Sun, L. Cheng, X. Chen, R. Yan, and R. X. Gao, "Waveletkernelnet: An interpretable deep neural network for industrial intelligent diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 4, pp. 2302–2312, 2021.
- [256] Q. Ni, J. Ji, B. Halkon, K. Feng, and A. K. Nandi, "Physics-informed residual network (piresnet) for rolling element bearing fault diagnostics," *Mechanical Systems and Signal Processing*, vol. 200, p. 110544, 2023.
- [257] H. Lu, V. P. Nemani, V. Barzegar, C. Allen, C. Hu, S. Laflamme, S. Sarkar, and A. T. Zimmerman, "A physics-informed feature weighting method for bearing fault diagnostics," *Mechanical Systems and Signal Processing*, vol. 191, p. 110 171, 2023.
- [258] Y. Chen, M. Rao, K. Feng, and M. J. Zuo, "Physics-informed lstm hyperparameters selection for gearbox fault detection," *Mechanical Systems and Signal Processing*, vol. 171, p. 108 907, 2022.
- [259] Y. Ding, L. Ma, J. Ma, M. Suo, L. Tao, Y. Cheng, and C. Lu, "Intelligent fault diagnosis for rotating machinery using deep q-network based health state classification: A deep reinforcement learning approach," *Advanced Engineering Informatics*, vol. 42, p. 100 977, 2019.
- [260] H. Wang, J. Xu, C. Sun, R. Yan, and X. Chen, "Intelligent fault diagnosis for planetary gearbox using time-frequency representation and deep reinforcement learning," *IEEE/ASME Transactions On Mechatronics*, vol. 27, no. 2, pp. 985–998, 2021.
- [261] G. Qian and J. Liu, "Development of deep reinforcement learning-based fault diagnosis method for rotating machinery in nuclear power plants," *Progress in Nuclear Energy*, vol. 152, p. 104 401, 2022.
- [262] R. Wang, H. Jiang, K. Zhu, Y. Wang, and C. Liu, "A deep feature-enhanced reinforcement learning method for rolling bearing fault diagnosis," *Advanced Engineering Informatics*, vol. 54, p. 101 750, 2022.
- [263] Z. Li, H. Jiang, and X. Wang, "A novel reinforcement learning agent for rotating machinery fault diagnosis with data augmentation," *Reliability Engineering & System Safety*, vol. 253, p. 110 570, 2025.
- [264] R. Wang, H. Jiang, X. Li, and S. Liu, "A reinforcement neural architecture search method for rolling bearing fault diagnosis," *Measurement*, vol. 154, p. 107 417, 2020.
- [265] L. Wen, X. Li, and L. Gao, "A new reinforcement learning based learning rate scheduler for convolutional neural network in fault classification," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 12, pp. 12 890–12 900, 2020.
- [266] X. Zhang, S. Fan, and Z. Song, "Reinforcement learning-based cost-sensitive classifier for imbalanced fault classification," *Science China Information Sciences*, vol. 66, no. 11, p. 212 201, 2023.
- [267] S. Fan, X. Zhang, and Z. Song, "Imbalanced sample selection with deep reinforcement learning for fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 4, pp. 2518–2527, 2021.

- [268] K. Yan, C. Lu, X. Ma, Z. Ji, and J. Huang, "Intelligent fault diagnosis for air handing units based on improved generative adversarial network and deep reinforcement learning," *Expert Systems with Applications*, vol. 240, p. 122 545, 2024.
- [269] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [270] J. Tian, Y. Jiang, J. Zhang, H. Luo, and S. Yin, "A novel data augmentation approach to fault diagnosis with class-imbalance problem," *Reliability Engineering & System Safety*, vol. 243, p. 109 832, 2024.
- [271] Y. Liu, H. Jiang, C. Liu, W. Yang, and W. Sun, "Data-augmented wavelet capsule generative adversarial network for rolling bearing fault diagnosis," *Knowledge-Based Systems*, vol. 252, p. 109 439, 2022.
- [272] H. Gao, X. Zhang, X. Gao, F. Li, and H. Han, "Icot-gan: Integrated convolutional transformer gan for rolling bearings fault diagnosis under limited data condition," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–14, 2023.
- [273] S. Dahdal, L. Colombi, M. Brina, A. Gilli, M. Tortonesi, M. Vignoli, and C. Stefanelli, "An MLOps framework for GAN-based fault detection in Bonfiglioli's EVO plant," *Infocommunications Journal*, vol. 16, no. 2, pp. 2–10, 2024.
- [274] C. Hegedűs and P. Varga, "Tailoring MLOps techniques for industry 5.0 needs," in *2023 19th International Conference on Network and Service Management (CNSM)*, 2023, pp. 1–7.
- [275] I. Misbah, C. K. Lee, and K. L. Keung, "Fault diagnosis in rotating machines based on transfer learning: Literature review," *Knowledge-Based Systems*, vol. 283, p. 111 158, 2024.
- [276] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on resnet-50," *Neural Computing and Applications*, vol. 32, no. 10, pp. 6111–6124, 2020.
- [277] J. Caçao, J. Santos, and M. Antunes, "Explainable ai for industrial fault diagnosis: A systematic review," *Journal of Industrial Information Integration*, p. 100 905, 2025.
- [278] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf
- [279] F. Kakavandi, P. Han, R. D. Reus, P. G. Larsen, and H. Zhang, "Interpretable fault detection approach with deep neural networks to industrial applications," *2023 International Conference on Control, Automation and Diagnosis (ICCAD)*, pp. 1–7, 2023.
- [280] H. Tang, H. Wang, and C. Li, "Explainable fault diagnosis method based on statistical features for gearboxes," *Eng. Appl. Artif. Intell.*, vol. 148, p. 110 503, 2025.
- [281] M. J. Hasan, M. Sohaib, and J.-M. Kim, "An explainable ai-based fault diagnosis model for bearings," *Sensors*, vol. 21, no. 12, p. 4070, 2021.
- [282] M. Baek and S. B. Kim, "Failure detection and primary cause identification of multivariate time series data in semiconductor equipment," *IEEE Access*, vol. 11, pp. 54 363–54 372, 2023.
- [283] J. Hu, Y. Zhang, W. Li, X. Zheng, and Z. Tian, "Trustworthy artificial intelligence based on an explicable temporal feature network for industrial fault diagnosis," *Cognitive Computation*, vol. 16, no. 2, pp. 534–545, 2024.
- [284] A. N. Gummadi, J. C. Napier, and M. Abdallah, "Xai-iot: an explainable ai framework for enhancing anomaly detection in iot systems," *IEEE Access*, vol. 12, pp. 71 024–71 054, 2024.
- [285] M. Keshk, N. Koroniotis, N. Pham, N. Moustafa, B. Turnbull, and A. Y. Zomaya, "An explainable deep learning-enabled intrusion detection framework in iot networks," *Information Sciences*, vol. 639, p. 119 000, 2023.
- [286] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.
- [287] H.-Y. Chen and C.-H. Lee, "Vibration signals analysis by explainable artificial intelligence (xai) approach: Application on bearing faults diagnosis," *IEEE Access*, vol. 8, pp. 134 246–134 256, 2020.
- [288] L. Guo, X. Gu, Y. Yu, A. Duan, and H. Gao, "An analysis method for interpretability of convolutional neural network in bearing fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–12, 2024.
- [289] M. S. Kim, J. P. Yun, and P. Park, "An explainable convolutional neural network for fault diagnosis in linear motion guide," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 4036–4045, 2020.
- [290] L. C. Brito, G. A. Susto, J. N. Brito, and M. A. V. Duarte, "Fault diagnosis using explainable ai: A transfer learning-based approach for rotating machinery exploiting augmented synthetic data," *Expert Systems with Applications*, vol. 232, p. 120 860, 2023.
- [291] M. Gwak, M. S. Kim, J. P. Yun, and P. Park, "Robust and explainable fault diagnosis with power-perturbation-based decision boundary analysis of deep learning models," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 6982–6992, 2022.
- [292] M. S. Kim, J. P. Yun, and P. Park, "Deep learning-based explainable fault diagnosis model with an individually grouped 1-d convolution for three-axis vibration signals," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 12, pp. 8807–8817, 2022.
- [293] J. Lee, I. Noh, J. Lee, and S. W. Lee, "Development of an explainable fault diagnosis framework based on sensor data imagification: A case study of the robotic spot-welding process," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 10, pp. 6895–6904, 2021.
- [294] S. Li, T. Li, C. Sun, R. Yan, and X. Chen, "Multilayer grad-cam: An effective tool towards explainable deep neural networks for intelligent fault diagnosis," *Journal of manufacturing systems*, vol. 69, pp. 20–30, 2023.
- [295] G. Li, Q. Yao, C. Fan, C. Zhou, G. Wu, Z. Zhou, and X. Fang, "An explainable one-dimensional convolutional neural networks based fault diagnosis method for building heating, ventilation and air conditioning systems," *Building and Environment*, vol. 203, p. 108 057, 2021.
- [296] A. Binder, G. Montavon, S. Lapuschkin, K.-R. Müller, and W. Samek, "Layer-wise relevance propagation for neural networks with local renormalization layers," in *International conference on artificial neural networks*. Springer, 2016, pp. 63–71.
- [297] J. Grezmak, P. Wang, C. Sun, and R. X. Gao, "Explainable convolutional neural network for gearbox fault diagnosis," *Procedia CIRP*, vol. 80, pp. 476–481, 2019.
- [298] P. Agarwal, M. Tamer, and H. Budman, "Explainability: Relevance based dynamic deep learning algorithm for fault detection and diagnosis in chemical processes," *Computers & Chemical Engineering*, vol. 154, p. 107 467, 2021.
- [299] Y. Gao, S. Miyata, and Y. Akashi, "How to improve the application potential of deep learning model in hvac fault diagnosis: Based on pruning and interpretable deep learning method," *Applied Energy*, vol. 348, p. 121591, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261923009558>
- [300] P. Peng, Y. Zhang, H. Wang, and H. Zhang, "Towards robust and understandable fault detection and diagnosis using denoising sparse autoencoder and smooth integrated gradients," *ISA transactions*, vol. 125, pp. 371–383, 2022.
- [301] J. Du, X. Li, Y. Gao, and L. Gao, "Integrated gradient-based continuous wavelet transform for bearing fault diagnosis," *Sensors*, vol. 22, no. 22, p. 8760, 2022.
- [302] M. R. Santos, A. Guedes, and I. Sanchez-Gendriz, "Shapley additive explanations (shap) for efficient feature selection in rolling bearing fault diagnosis," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 316–341, 2024.
- [303] H. Pu, K. Zhang, and Y. An, "Restricted sparse networks for rolling bearing fault diagnosis," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 11, pp. 11 139–11 149, 2023.
- [304] T. Yan, D. Wang, and Y. Wang, "Discrimination-and sparsity-driven weight-oriented optimization model for interpretable initial fault detection and fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2023.
- [305] G. Chen, J. Yuan, Y. Zhang, H. Zhu, R. Huang, F. Wang, and W. Li, "Enhancing reliability through interpretability: A comprehensive survey of interpretable intelligent fault diagnosis in rotating machinery," *IEEE Access*, 2024.
- [306] L. Ma, Y. Ding, Z. Wang, C. Wang, J. Ma, and C. Lu, "An interpretable data augmentation scheme for machine fault diagnosis based on a sparsity-constrained generative adversarial network," *Expert Systems with Applications*, vol. 182, p. 115 234, 2021.
- [307] R. Tian, M. Cui, and G. Chen, "A neural-symbolic network for interpretable fault diagnosis of rolling element bearings based on temporal logic," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–14, 2024.

- [308] G. Chen, M. Liu, and J. Chen, "Frequency-temporal-logic-based bearing fault diagnosis and fault interpretation using bayesian optimization with bayesian neural networks," *Mechanical Systems and Signal Processing*, vol. 145, p. 106 951, 2020.
- [309] G. Chen, P. Wei, H. Jiang, and M. Liu, "Formal language generation for fault diagnosis with spectral logic via adversarial training," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 119–129, 2020.
- [310] G. Chen, P. Lu, and Y. Tang, "Interpretable fault diagnosis of rolling element bearings with temporal logic neural network," in *2024 18th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 2024, pp. 290–296.
- [311] G. Chen, Y. Lu, and R. Su, "Interpretable fault diagnosis with shapelet temporal logic: Theory and application," *Automatica*, vol. 142, p. 110 350, 2022.
- [312] F. Rao, M. Zeng, and Y. Cheng, "A novel interpretable model via algorithm unrolling for intelligent fault diagnosis of machinery," *IEEE Sensors Journal*, vol. 24, no. 1, pp. 495–505, 2023.
- [313] Z. Zhao, T. Li, B. An, S. Wang, B. Ding, R. Yan, and X. Chen, "Modeldriven deep unrolling: Towards interpretable deep learning against noise attacks for intelligent fault diagnosis," *ISA transactions*, vol. 129, pp. 644–662, 2022.
- [314] B. An, S. Wang, Z. Zhao, F. Qin, R. Yan, and X. Chen, "Interpretable neural network via algorithm unrolling for mechanical fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.
- [315] Y. Li, Z. Zhou, C. Sun, X. Chen, and R. Yan, "Variational attentionbased interpretable transformer network for rotary machine fault diagnosis," *IEEE transactions on neural networks and learning systems*, vol. 35, no. 5, pp. 6180–6193, 2022.
- [316] D. Liu, L. Cui, G. Wang, and W. Cheng, "Interpretable domain adaptation transformer: a transfer learning method for fault diagnosis of rotating machinery," *Structural Health Monitoring*, vol. 24, no. 2, pp. 1187–1200, 2025.
- [317] D. Chen, Z. Xie, R. Liu, W. Yu, Q. Hu, X. Li, and S. X. Ding, "Bayesian hierarchical graph neural networks with uncertainty feedback for trustworthy fault diagnosis of industrial processes," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.
- [318] R. Mishra, A. Choudhary, S. Fatima, A. Mohanty, and B. Panigrahi, "A systematic review on advancement and challenges in multi-fault diagnosis of rotating machines," *Engineering Applications of Artificial Intelligence*, vol. 156, p. 111 306, 2025.
- [319] Y. Zhou and W. Sun, "Tool wear condition monitoring in milling process based on current sensors," *IEEE access*, vol. 8, pp. 95 491–95 502, 2020.
- [320] P. Twardowski, M. Tabaszewski, M. Wiciak-Pikuła, and A. Felusiak-Czyrca, "Identification of tool wear using acoustic emission signal and machine learning methods," *Precision Engineering*, vol. 72, pp. 738–744, 2021.
- [321] H. Miao, Z. Zhao, C. Sun, B. Li, and R. Yan, "A u-net-based approach for tool wear area detection and identification," *IEEE Transactions on instrumentation and measurement*, vol. 70, pp. 1–10, 2020.
- [322] C.-W. Hung, C.-H. Lee, C.-C. Kuo, and S.-X. Zeng, "Soc-based early failure detection system using deep learning for tool wear," *Ieee Access*, vol. 10, pp. 70 491–70 501, 2022.
- [323] Y. Zhou, G. Zhi, W. Chen, Q. Qian, D. He, B. Sun, and W. Sun, "A new tool wear condition monitoring method based on deep learning under small samples," *Measurement*, vol. 189, p. 110 622, 2022.
- [324] R. Munaro, A. Attanasio, and A. Del Prete, "Tool wear monitoring with artificial intelligence methods: a review," *Journal of Manufacturing and Materials Processing*, vol. 7, no. 4, p. 129, 2023.
- [325] M. Cheng, L. Jiao, P. Yan, H. Jiang, R. Wang, T. Qiu, and X. Wang, "Intelligent tool wear monitoring and multi-step prediction based on deep learning model," *Journal of Manufacturing Systems*, vol. 62, pp. 286–300, 2022.
- [326] Z. Huang, J. Zhu, J. Lei, X. Li, and F. Tian, "Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations," *Journal of intelligent manufacturing*, vol. 31, no. 4, pp. 953–966, 2020.
- [327] Q. Wang, H. Wang, L. Hou, and S. Yi, "Overview of tool wear monitoring methods based on convolutional neural network," *Applied Sciences*, vol. 11, no. 24, p. 12041, 2021.
- [328] Y. Zhao, A. Jiang, W. Jiang, Y. Zhang, and S. Zhang, "Hydraulic pump fault diagnosis by modified slime mold algorithm optimized support vector machine," *Measurement*, vol. 254, p. 117 878, 2025.
- [329] D. Ma, Z. Liu, Q. Gao, and Y. Ding, "Fault diagnosis of multi-step electromagnetic hydraulic valve group based on localized current signal cs-svm," *Measurement*, vol. 245, p. 116 632, 2025.
- [330] N. Liang, F. Zhang, Z. Yuan, H. Wang, J. Zhang, Z. Fan, and X. Yu, "Fault diagnosis in hydraulic systems via multi-channel multi-modal fusion," *Measurement Science and Technology*, vol. 36, no. 5, p. 055 023, 2025.
- [331] R. Jiang, Z. Yuan, H. Wang, Z. Fan, Y. Zhang, N. Liang, and X. Yu, "Mcfpred: A novel multichannel signal adaptive fusion framework for fault diagnosis in hydraulic systems," *IEEE Transactions on Instrumentation and Measurement*, 2025.
- [332] S. Liu, J. Yin, Y. Zhang, and P. Liang, "Mlift: Multi-scale linear interaction fusion transformer for fault diagnosis of hydraulic pumps," *Measurement*, p. 117892, 2025.
- [333] G. Zhang, K. Zhang, W. Shen, G. Xu, and Y.-J. Lin, "Intelligent fault diagnosis of multi-way directional valves in hydraulic systems using digital twin and deep learning approaches," *Mechanical Systems and Signal Processing*, vol. 230, p. 112 579, 2025.
- [334] S. Tang, Z. Yan, Y. Xue, K. M. Lim, and Y. Zhu, "Deep transferable model with multi-head attention for small sample fault diagnosis of hydraulic pumps," *Engineering Applications of Artificial Intelligence*, vol. 160, p. 111 808, 2025.
- [335] K.-J. Shih, M.-F. Hsieh, B.-J. Chen, and S.-F. Huang, "Machine learning for inter-turn short-circuit fault diagnosis in permanent magnet synchronous motors," *IEEE Transactions on Magnetics*, vol. 58, no. 8, pp. 1–7, 2022.
- [336] C. Zachariades and V. Xavier, "A review of artificial intelligence techniques in fault diagnosis of electric machines," *Sensors (Basel, Switzerland)*, vol. 25, no. 16, p. 5128, 2025.
- [337] P. Pietrzak, M. Wolkiewicz, and T. Orłowska-Kowalska, "Pmsm stator winding fault detection and classification based on bispectrum analysis and convolutional neural network," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 5, pp. 5192–5202, 2022.
- [338] F. Husari and J. Seshadrinath, "Incipient interturn fault detection and severity evaluation in electric drive system using hybrid hcnn-svm based model," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1823–1832, 2021.
- [339] M.-C. Kim, J.-H. Lee, D.-H. Wang, and I.-S. Lee, "Induction motor fault diagnosis using support vector machine, neural networks, and boosting methods," *Sensors*, vol. 23, no. 5, p. 2585, 2023.
- [340] M.-Q. Tran, M.-K. Liu, Q.-V. Tran, and T.-K. Nguyen, "Effective fault diagnosis based on wavelet and convolutional attention neural network for induction motors," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2021.
- [341] R. F. R. Junior, I. A. dos Santos Areias, M. M. Campos, C. E. Teixeira, L. E. B. da Silva, and G. F. Gomes, "Fault detection and diagnosis in electric motors using 1d convolutional neural networks with multi-channel vibration signals," *Measurement*, vol. 190, p. 110 759, 2022.
- [342] A. Choudhary, R. K. Mishra, S. Fatima, and B. K. Panigrahi, "Multiinput cnn based vibro-acoustic fusion for accurate fault diagnosis of induction motor," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105 872, 2023.
- [343] Y. Tang, X. Zhang, G. Qin, Z. Long, S. Huang, D. Song, and H. Shao, "Graph cardinality preserved attention network for fault diagnosis of induction motor under varying speed and load condition," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3702–3712, 2021.
- [344] W. Lang, Y. Hu, C. Gong, X. Zhang, H. Xu, and J. Deng, "Artificial intelligence-based technique for fault detection and diagnosis of ev motors: A review," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 384–406, 2021.
- [345] T.-T. Vo, M.-K. Liu, and M.-Q. Tran, "Harnessing attention mechanisms in a comprehensive deep learning approach for induction motor fault diagnosis using raw electrical signals," *Engineering Applications of Artificial Intelligence*, vol. 129, p. 107643, 2024.
- [346] A. A. Nuhu, Q. Zeeshan, B. Safaei, and M. A. Shahzad, "Machine learning-based techniques for fault diagnosis in the semiconductor manufacturing process: a comparative study," *The Journal of Supercomputing*, vol. 79, no. 2, pp. 2031–2081, 2023.
- [347] C.-Y. Hsu and W.-C. Liu, "Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing," *Journal of Intelligent Manufacturing*, vol. 32, no. 3, pp. 823–836, 2021.
- [348] S. Lou, C. Yang, P. Wu, L. Kong, and Y. Xu, "Fault diagnosis of blast furnace iron-making process with a novel deep stationary kernel learning support vector machine approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

- [349] H. Nandanwar and R. Katarya, "Deep learning enabled intrusion detection system for industrial iot environment," *Expert Systems with Applications*, vol. 249, p. 123 808, 2024.
- [350] M. Mohy-Eddine, A. Guezaz, S. Benkirane, M. Azrou, and Y. Farhaoui, "An ensemble learning based intrusion detection model for industrial iot security," *Big Data Mining and Analytics*, vol. 6, no. 3, pp. 273–287, 2023.
- [351] S. Soliman, W. Oudah, and A. Aljuhani, "Deep learning-based intrusion detection approach for securing industrial internet of things," *Alexandria Engineering Journal*, vol. 81, pp. 371–383, 2023.
- [352] B. Tang, Y. Lu, Q. Li, Y. Bai, J. Yu, and X. Yu, "A diffusion model based on network intrusion detection method for industrial cyberphysical systems," *Sensors*, vol. 23, no. 3, p. 1141, 2023.
- [353] S. T. Mehedi, A. Anwar, Z. Rahman, K. Ahmed, and R. Islam, "Dependable intrusion detection system for iot: A deep transfer learning based approach," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 1006–1017, 2022.
- [354] M.-Q. Tran, M. Amer, A. Y. Abdelaziz, H.-J. Dai, M.-K. Liu, M. Elsis, et al., "Robust fault recognition and correction scheme for induction motors using an effective iot with deep learning approach," *Measurement*, vol. 207, p. 112 398, 2023.
- [355] M. A. Alohal, F. N. Al-Wesabi, A. M. Hilal, S. Goel, D. Gupta, and A. Khanna, "Artificial intelligence enabled intrusion detection systems for cognitive cyber-physical systems in industry 4.0 environment," *Cognitive Neurodynamics*, vol. 16, no. 5, pp. 1045–1057, 2022.
- [356] A. Abusitta, G. H. de Carvalho, O. A. Wahab, T. Halabi, B. C. Fung, and S. Al Mamoori, "Deep learning-enabled anomaly detection for iot systems," *Internet of Things*, vol. 21, p. 100 656, 2023.
- [357] F. Khan, R. Alturki, M. A. Rahman, S. Mastorakis, I. Razzak, and S. T. Shah, "Trustworthy and reliable deep-learning-based cyberattack detection in industrial iot," *IEEE transactions on industrial informatics*, vol. 19, no. 1, pp. 1030–1038, 2022.
- [358] R. Canonico and G. Sperli, "Industrial cyber-physical systems protection: A methodological review," *Computers & Security*, vol. 135, p. 103 531, 2023.
- [359] Z. Zhang, G. Peng, W. Wang, and Y. Chen, "Real-time human fault detection in assembly tasks, based on human action prediction using a spatio-temporal learning model," *Sustainability*, vol. 14, no. 15, p. 9027, 2022.
- [360] A. M. Vukicevic, M. N. Petrovic, N. M. Knezevic, and K. M. Jovanovic, "Deep learning-based recognition of unsafe acts in manufacturing industry," *IEEE Access*, vol. 11, pp. 103 406–103 418, 2023.
- [361] L. Wang, J. Yuan, S. Zhou, and M. König, "Automatic detection of unsafe behaviors resulting from the improper use of ppe during human–robot collaboration," *Journal of Computing in Civil Engineering*, vol. 40, no. 2, p. 04 025 147, 2026.
- [362] C. Zhang, Y. Zhao, Y. Zhao, T. Li, and X. Zhang, "Causal discovery and inference-based fault detection and diagnosis method for heating, ventilation and air conditioning systems," *Building and Environment*, vol. 212, p. 108 760, 2022.
- [363] Z. Du, S. Chen, P. Li, K. Chen, X. Liang, X. Zhu, and X. Jin, "Knowledge-extracted deep learning diagnosis and its cloud-based management for multiple faults of chiller," *Building and Environment*, vol. 235, p. 110 228, 2023.
- [364] Z. Chen, Z. O'Neill, J. Wen, O. Pradhan, T. Yang, X. Lu, G. Lin, S. Miyata, S. Lee, C. Shen et al., "A review of data-driven fault detection and diagnostics for building hvac systems," *Applied Energy*, vol. 339, p. 121 030, 2023.
- [365] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1d cnn classifier," *Journal of signal processing systems*, vol. 91, no. 2, pp. 179–189, 2019.
- [366] Q. Zhang and Z. Ju, "Rolling bearing fault diagnosis based on 2d cnn and hybrid kernel fuzzy svm," *Advanced Theory and Simulations*, p. 2 400 793, 2025.
- [367] K. S. Siraskar R. and P. S., "Reinforcement learning for predictive maintenance: A systematic technical review," *Artif Intell Rev*, vol. 56, p. 12885–12947, 2023. [Online]. Available: <https://doi.org/10.1007/s10462-023-10468-6>
- [368] M. L. Ruiz Rodríguez, S. Kubler, A. de Giorgio, M. Cordy, J. Robert, and Y. Le Traon, "Multi-agent deep reinforcement learning based predictive maintenance on parallel machines," *Robotics and Computer-Integrated Manufacturing*, vol. 78, p. 102 406, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0736585422000928>
- [369] J. Kairo, "Machine learning algorithms for predictive maintenance in manufacturing," *Journal of Technology and Systems*, vol. 6, pp. 66–79, 08 2024.
- [370] S. Vollert and A. Theissler, "Challenges of machine learning-based rul prognosis: A review on nasa's c-mapss data set," in *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2021, pp. 1–8.
- [371] S. S. Sharma, V. V., and A. Malviya, "Ai-enhanced predictive maintenance in intelligent systems for industries," in *2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET)*, 2024, pp. 1–6.
- [372] A. E. Frankó and P. Varga, "A survey on machine learning based smart maintenance and quality control solutions," *Infocommunications Journal*, vol. 13, no. 4, pp. 28–35, 2021.
- [373] F. Wang, A. Liu, C. Qu, R. Xiong, and L. Chen, "A deep-learning method for remaining useful life prediction of power machinery via dual-attention mechanism," *Sensors*, vol. 25, no. 2, 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/2/497>
- [374] L. Cao, X. Wang, H. Zhang, Z. Meng, J. Li, and M. Liu, "A novel crossscenario transferable rul prediction network with multisource domain meta transfer learning for wind turbine bearings," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–1, 2025.
- [375] X. Liu, Z. Zhang, Z. Li, J. Wang, Y. Zhu, and H. Ma, "Advancements in bearing health monitoring and remaining useful life prediction: Techniques, challenges, and future directions," *Measurement Science and Technology*, 2025. [Online]. Available: <http://iopscience.iop.org/article/10.1088/1361-6501/adafcb>
- [376] D. Li, Y.-F. Li, and H. Wang, "Next-gen aiot-driven predictive maintenance and health management system for high-speed train wheel and axle systems," in *2024 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, 2024, pp. 1–6.
- [377] P. Suebsuwong, "Integrating horizons: A holistic view of predictive maintenance in aviation maintenance practices," *Journal of Science and Technology Kasetsart University*, vol. 13, no. 3, pp. 38–60, 2024. [Online]. Available: <https://ph03.tci-thaijo.org/index.php/JSTKU/article/view/3495>
- [378] G. Bianchi, C. Fanelli, F. Freddi, F. Giuliani, and A. L. Placa, "Systematic review railway infrastructure monitoring: From classic techniques to predictive maintenance," *Advances in Mechanical Engineering*, vol. 17, no. 1, p. 16878132241285631, 2025. [Online]. Available: <https://doi.org/10.1177/16878132241285631>
- [379] G. Papa, Š. Poklukar, A. Frankó, A. Sillitti, A. Kancilija, M. Šterk, C. Hegedűs et al., "Improving the maintenance of railway switches through proactive approach," *Electronics*, vol. 9, no. 8, p. 1260, 2020.
- [380] C. Hegedűs, P. Ciancarini, A. Frankó, A. Kancilija, I. Moldován, G. Papa et al., "Proactive maintenance of railway switches," in *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2018.
- [381] S. R. Katta, "Optimizing pharmaceutical manufacturing through ai-powered calibration and maintenance workflow management," *Journal of Artificial Intelligence, Machine Learning and Data Science*, vol. 1, no. 4, pp. 2175–2181, 2023.
- [382] S. Gupta, J. Maiti, and A. Kumar, "Prognostic modeling of li-ion battery using convolutional long short term memory (convolstm)," *International Journal of System Assurance Engineering and Management*, 2025. [Online]. Available: <https://doi.org/10.1007/s13198-025-02706-7>
- [383] J. Heredia and E. Ayala, "IoT and AI-Based Predictive Maintenance System Design for Express Auto Repair Shops," *Revista Técnica energía*, vol. 21, pp. 81–86, 06 2025. [Online]. Available: http://scielo.senecyt.gob.ec/scielo.php?script=sci_arttext&pid=S2602-84922025000100081&nrm=iso
- [384] F. Battisti, A. V. B. da Silva, L. F. A. Pereira, T. B. A. de Carvalho, J. C. T. de Araujo, E. Choi, T. A. Nguyen, and D. Min, "hLSTMAging: A hybrid LSTM model for software aging forecast," *Applied Sciences*, vol. 12, no. 13, p. 6412, 2022.
- [385] C. Chen, H. Fu, Y. Zheng, F. Tao, and Y. Liu, "The advance of digital twin for predictive maintenance: The role and function of machine learning," *Journal of Manufacturing Systems*, vol. 71, pp. 581–594, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S027861252300211X>
- [386] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Fusing physicsbased and deep learning models for prognostics," *Reliability Engineering & System Safety*, vol. 217, p. 107 961, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832021004725>

- [387] R. G. Nascimento and F. A. C. Viana, "Fleet prognosis with physics-informed recurrent neural networks," *ArXiv*, vol. abs/1901.05512, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:58014195>
- [388] Y. Yucesan and F. Viana, "Wind turbine main bearing fatigue life estimation with physics-informed neural networks," *Annual Conference of the PHM Society*, vol. 11, 09 2019.
- [389] A. Dourado and F. Viana, "Physics-informed neural networks for corrosion-fatigue prognosis," *Annual Conference of the PHM Society*, vol. 11, 09 2019.
- [390] N. Yang, W. Zhang, J. Zhang, K. Wang, Y. Su, and Y. Liu, "A method for remaining useful life prediction and uncertainty quantification of rolling bearings based on fault feature gain," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–1, 2025.
- [391] R. Xiong, Y. He, Y. Sun, Y. Jia, and W. Shen, "Enhanced electrode-level diagnostics for lithium-ion battery degradation using physics-informed neural networks," *Journal of Energy Chemistry*, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2095495625000609>
- [392] Z. Lu, C. Guo, M. Liu, and R. Shi, "Remaining useful lifetime estimation for discrete power electronic devices using physics-informed neural network," *Scientific Reports*, vol. 13, no. 1, p. 101 67, jun 2023. [Online]. Available: <https://doi.org/10.1038/s41598-023-37154-5>
- [393] L. E. J. Wang, R. Yang, C. Wang, H. Li, and R. Xiong, "A physics-informed neural network-based method for predicting degradation trajectories and remaining useful life of supercapacitors," *Green Energy and Intelligent Transportation*, vol. 4, no. 3, p. 100 291, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2773153725000416>
- [394] X. Liao, S. Chen, P. Wen, and S. Zhao, "Remaining useful life with self-attention assisted physics-informed neural network," *Advanced Engineering Informatics*, vol. 58, p. 102 195, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1474034623003233>
- [395] S. Xie, W. Cheng, Z. Nie, Q. Huang, J. Xing, X. Chen, R. Zhang, and Y. Yang, "Bayesian physics-informed neural networks with iterative ensemble kalman inversion for rul prediction and uncertainty quantification," *Advanced Engineering Informatics*, vol. 69, p. 103 907, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1474034625008006>
- [396] Y. Zhou and M. Huang, "Lithium-ion batteries remaining useful life prediction based on a mixture of empirical mode decomposition and arima model," *Microelectronics Reliability*, vol. 65, pp. 265–273, 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0026271416303018>
- [397] K. Javed, R. Gouriveau, N. Zerhouni, and P. Nectoux, "A feature extraction procedure based on trigonometric functions and cumulative descriptors to enhance prognostics modeling," in *2013 IEEE Conference on Prognostics and Health Management (PHM)*, 2013, pp. 1–7.
- [398] J. Coble and J. Baalis, "Merging data sources to predict remaining useful life – an automated method to identify prognostic parameters." PhD Diss, University of Tennessee, 2010. [Online]. Available: https://trace.tennessee.edu/utk_graddiss/683
- [399] M. Rigamonti, P. Baraldi, E. Zio, I. Roychoudhury, K. Goebel, and S. Poll, "Echo state network for the remaining useful life prediction of a turbofan engine," *PHM Society European Conference*, 2016. [Online]. Available: <https://doi.org/10.36001/phme.2016.v3i1.1623>
- [400] X. Zhu, P. Yu, and S. Su, "A deep transfer network based on dual-task learning for predicting the remaining useful life of rolling bearings," *Measurement Science and Technology*, vol. 36, no. 3, p. 036108, feb 2025. [Online]. Available: <https://dx.doi.org/10.1088/1361-6501/adafd2>
- [401] X. W. Pandit R, "Data-driven models for predicting remaining useful life of high-speed shaft bearings in wind turbines using vibration signal analysis and sparrow search algorithm," *Energy Science & Engineering*, 2023. [Online]. Available: <https://misclibrary.wiley.com/doi/10.1002/ese3.1075>
- [402] W. Zhang, D. Yang, and H. Wang, "Data-driven methods for predictive maintenance of industrial equipment: A survey," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2213–2227, 2019.
- [403] Y. Liu, Y. Wang, Y. Wang, S. Xue, Z. Wang, and Z. Gao, "Method for predicting remaining useful life of rolling bearings based on dynamic complexity characteristic entropy and quantum neural networks," *Engineering Failure Analysis*, vol. 170, p. 109 315, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1350630725000561>
- [404] Y. Su, Z. Lei, G. Wen, and X. Chen, "A novel multi-sensor data fusion enabled health indicator construction and remaining useful life prediction of aero-engine," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 0, no. 0, p. 09544054241310485, 0. [Online]. Available: <https://doi.org/10.1177/09544054241310485>
- [405] Y. Qin, J. Zhou, and D. Chen, "Unsupervised health indicator construction by a novel degradation-trend-constrained variational autoencoder and its applications," *IEEE/ASME Transactions on Mechatronics*, vol. PP, pp. 1–1, 07 2021.
- [406] M. Kim, S. Yoo, S. Son, S. Y. Chang, and K.-Y. Oh, "Physics-informed deep learning framework for explainable remaining useful life prediction," *Engineering Applications of Artificial Intelligence*, vol. 143, p. 110072, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0952197625000727>
- [407] L. Wen, G. Yang, L. Hu, C. Yang, and K. Feng, "A new unsupervised health index estimation method for bearings early fault detection based on gaussian mixture model," *Engineering Applications of Artificial Intelligence*, vol. 128, p. 107562, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0952197623017463>
- [408] T. Sun, C. Yin, H. Zheng, and Y. Dong, "An unsupervised framework for dynamic health indicator construction and its application in rolling bearing prognostics," 2025. [Online]. Available: <https://arxiv.org/abs/2506.05438>
- [409] P. Malhotra, V. TV, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "Multi-sensor prognostics using an unsupervised health index based on lstm encoder-decoder," 2016. [Online]. Available: <https://arxiv.org/abs/1608.06154>
- [410] X. Ye, G. Li, L. Meng, and G. Lu, "Dynamic health index extraction for incipient bearing degradation detection," *ISA Transactions*, vol. 128, pp. 535–549, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0019057821006005>
- [411] C. Lim, S. Kim, Y.-H. Seo, and J.-H. Choi, "Feature extraction for bearing prognostics using weighted correlation of fault frequencies over cycles," *Structural Health Monitoring*, vol. 19, no. 6, pp. 1808–1820, 2020. [Online]. Available: <https://doi.org/10.1177/1475921719900917>
- [412] X. Zhang, P. Tengyi, S. Sun, and Y. Zhou, "New multifeature information health index (mihi) based on a quasi-orthogonal sparse algorithm for bearing degradation monitoring," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–14, 08 2021.
- [413] D. An, J.-H. Choi, and N. H. Kim, "Remaining useful life prediction of rolling element bearings using degradation feature based on amplitude decrease at specific frequencies," *Structural Health Monitoring*, vol. 17, no. 5, pp. 1095–1109, 2018. [Online]. Available: <https://doi.org/10.1177/1475921717736226>
- [414] J. Park, S. Kim, J.-H. Choi, and S. H. Lee, "Frequency energy shift method for bearing fault prognosis using microphone sensor," *Mechanical Systems and Signal Processing*, vol. 147, p. 107 068, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327020304544>
- [415] X. Du, W. Jia, P. Yu, Y. Shi, and S. Cheng, "A remaining useful life prediction method based on time-frequency images of the mechanical vibration signals," *Measurement*, vol. 202, p. 111 782, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S026322412200985X>
- [416] X. Luo, Y. Song, W. Bu, H. Liang, and M. Zheng, "A joint prediction of the state of health and remaining useful life of lithium-ion batteries based on gaussian process regression and long short-term memory," *Processes*, vol. 13, no. 1, 2025. [Online]. Available: <https://www.mdpi.com/2227-9717/13/1/239>
- [417] M. Assafo and P. Langendoerfer, "Tool remaining useful life prediction using feature extraction and machine learning-based sensor fusion," *Results in Engineering*, vol. 28, p. 107 297, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2590123025033523>
- [418] I. Obisakin and E. C. Vanessa, "State of health estimation of lithium-ion batteries using support vector regression and long short-term memory," *Open Journal of Applied Sciences*, vol. 12, pp. 1366–1382, 2022.
- [419] M. Yenioğlu, E. Ayçiçek, and O. Erdinç, "Remaining useful life estimation of lithium-ion batteries," in *2024 15th National Conference on Electrical and Electronics Engineering (ELECO)*, 2024, pp. 1–5.
- [420] L. Ren, J. Dong, X. Wang, Z. Meng, L. Zhao, and M. J. Deen, "A data-driven auto-cnn-lstm prediction model for lithium-ion battery remaining useful life," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3478–3487, 2021.

- [421] A. Rastegarpanah, M. E. Asif, and R. Stolkin, "Hybrid neural networks for enhanced predictions of remaining useful life in lithium-ion batteries," *Batteries*, vol. 10, no. 3, 2024. [Online]. Available: <https://www.mdpi.com/2313-0105/10/3/106>
- [422] C. Wang, Z. Bao, H. Lin, Z. He, and M. Gao, "An exponential transformer for learning interpretable temporal information in remaining useful life prediction of lithium-ion battery," *IEEE Transactions on Transportation Electrification*, pp. 1–1, 2025.
- [423] L. Liu, J. Huang, H. Zhao, T. Li, and B. Li, "Patchformer: A novel patch-based transformer for accurate remaining useful life prediction of lithium-ion batteries," *Journal of Power Sources*, vol. 631, p. 236 187, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378775325000230>
- [424] R. E and S. A., "Performance benchmarking and analysis of prognostic methods for cmappss datasets," *International Journal of Prognostics and Health Management*, 2020. [Online]. Available: <https://doi.org/10.36001/ijphm.2014.v5i2.2236>
- [425] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–6.
- [426] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *2017 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2017, pp. 88–95.
- [427] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Reliability Engineering and System Safety*, vol. 172, pp. 1–11, 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832017307779>
- [428] H. Li, W. Zhao, Y. Zhang, and E. Zio, "Remaining useful life prediction using multi-scale deep convolutional neural network," *Applied Soft Computing*, vol. 89, p. 106113, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1568494620300533>
- [429] R. Jin, M. Wu, K. Wu, K. Gao, Z. Chen, and X. Li, "Position encoding based convolutional neural networks for machine remaining useful life prediction," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 8, pp. 1427–1439, 2022.
- [430] J. Wang, G. Wen, S. Yang, and Y. Liu, "Remaining useful life estimation in prognostics using deep bidirectional lstm neural network," in *2018 Prognostics and System Health Management Conference (PHMChongqing)*, 2018, pp. 1037–1042.
- [431] S. Deng and J. Zhou, "Prediction of remaining useful life of aeroengines based on CNN-LSTM-Attention," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, 2024.
- [432] Y. He, C. Wen, and W. Xu, "Residual life prediction of sacnn-bilstm aero-engine based on a multichannel hybrid network," *Applied Sciences*, vol. 15, no. 2, 2025. [Online]. Available: <https://www.mdpi.com/2076-3417/15/2/966>
- [433] J. Gao, Y. Wang, and Z. Sun, "An interpretable rul prediction method of aircraft engines under complex operating conditions using spatio-temporal features," *Measurement Science and Technology*, vol. 35, no. 7, p. 076003, apr 2024. [Online]. Available: <https://dx.doi.org/10.1088/1361-6501/ad3b2c>
- [434] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9521–9530, 2019.
- [435] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2306–2318, 2017.
- [436] A. Z. Ture, B. A. Akbulut, H. Ahmet et al., "Stacking-based ensemble learning for remaining useful life estimation," *Soft Computing*, vol. 28, no. 3, pp. 1337–1349, 2024.
- [437] Y. Liu, B. Xu, and Y.-a. Geng, "Multi-condition remaining useful life prediction based on mixture of encoders," *Entropy*, vol. 27, no. 1, 2025. [Online]. Available: <https://www.mdpi.com/1099-4300/27/1/79>
- [438] L. Jayasinghe, T. Samarasinghe, C. Yuen, J. C. N. Low, and S. S. Ge, "Temporal convolutional memory networks for remaining useful life estimation of industrial machinery," 2018. [Online]. Available: <https://arxiv.org/abs/1810.05644>
- [439] R. Li, Z. Chu, W. Jin, Y. Wang, and X. Hu, "Temporal convolutional network based regression approach for estimation of remaining useful life," in *2021 IEEE International Conference on Prognostics and Health Management (ICPHM)*, 2021, pp. 1–10.
- [440] J. Chen, D. Chen, and G. Liu, "Using temporal convolution network for remaining useful lifetime prediction," *Engineering Reports*, vol. 3, no. 3, p. e12305, 2021. [Online]. Available: <https://misclibrary.wiley.com/doi/abs/10.1002/eng2.12305>
- [441] F. Xiang, Y. Zhang, S. Zhang, Z. Wang, L. Qiu, and J.-H. Choi, "Bayesian gated-transformer model for risk-aware prediction of aeroengine remaining useful life," *Expert Systems with Applications*, vol. 238, p. 121859, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0957417423023618>
- [442] N. DeVol, C. Saldana, and K. Fu, "Inception based deep convolutional neural network for remaining useful life estimation of turbofan engines," *Annual Conference of the PHM Society*, 2021. [Online]. Available: <https://doi.org/10.36001/phmconf.2021.v13i1.3109>
- [443] T. Berghout, M.-D. Mouss, L.-H. Mouss, and M. Benbouzid, "Prognnet: A transferable deep network for aircraft engine damage propagation prognosis under real flight conditions," *Aerospace*, vol. 10, no. 1, 2023. [Online]. Available: <https://www.mdpi.com/2226-4310/10/1/10>
- [444] Z. Zhao and P. Zou, "Attention-based dual-channel deep neural network for aero-engine rul prediction under time-varying operating conditions," *Journal of Physics: Conference Series*, vol. 2386, no. 1, p. 012 027, dec 2022. [Online]. Available: <https://dx.doi.org/10.1088/1742-6596/2386/1/012027>
- [445] X. Li, J. Li, L. Zuo, L. Zhu, and H. T. Shen, "Domain adaptive remaining useful life prediction with transformer," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.
- [446] L. C. Brito, L. L. C. Araújo, G. A. Susto, J. N. Brito, and M. A. V. Duarte, "Rotating machinery health monitoring: Unsupervised anomaly detection and remaining useful life (RUL) prediction using explainable artificial intelligence (XAI) and statistical distribution," *IFACPapersOnLine*, vol. 59, no. 26, pp. 97–102, 2025.
- [447] Y. Chen, D. Zhang, and W. an Zhang, "Mswr-lrcn: A new deep learning approach to remaining useful life estimation of bearings," *Control Engineering Practice*, vol. 118, p. 104 969, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S096706612100246X>
- [448] K. Xia, Q. Li, L. Han, Z. Ren, and H. Luo, "Prediction of residual life of rolling bearings based on multi-scale improved temporal convolutional network (mitcn) model," *Machines*, vol. 13, no. 2, 2025. [Online]. Available: <https://www.mdpi.com/2075-1702/13/2/137>
- [449] T. Liu, Y. Li, Z. Hu, C. Fu, and G. Song, "Remaining useful life prediction of bearings based on small-sample enhanced and interpretable transfer learning," *Advanced Engineering Informatics*, vol. 69, p. 103 938, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1474034625008316>
- [450] A. Fentaye and K. Kyprianidis, "Gas turbine prognostics via temporal fusion transformer," *The Aeronautical Journal*, vol. 128, no. 1325, 2024.
- [451] B. Lim, S. O. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169207021000637>
- [452] Y. Chang, F. Li, J. Chen, Y. Liu, and Z. Li, "Efficient temporal flow transformer accompanied with multi-head probparse self-attention mechanism for remaining useful life prognostics," *Reliability Engineering & System Safety*, vol. 226, p. 108 701, 10 2022.
- [453] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, "Machine remaining useful life prediction via an attention-based deep learning approach," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 3, pp. 2521–2531, 2021.
- [454] Y. Mo, Q. Wu, X. Li, Y. Zhang, Y. Zhang, and Y. Zhang, "Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit," *Journal of Intelligent Manufacturing*, vol. 32, no. 8, pp. 1997–2006, 2021. [Online]. Available: <https://doi.org/10.1007/s10845-021-01750-x>
- [455] Y. Ding and M. Jia, "A convolutional transformer architecture for remaining useful life estimation," in *2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing)*, 2021, pp. 1–7.
- [456] J. Zhu, Z. Long, X. Ma, and F. Luan, "Bearing remaining useful life prediction based on bert fine-tuning," in *2023 Global Reliability and Prognostics and Health Management Conference (PHM-Hangzhou)*, 2023, pp. 1–6.

- [457] S. Sankararaman, "Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction," *Mechanical Systems and Signal Processing*, vol. 52–53, pp. 228–247, 2015. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0888327014002052>
- [458] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makaremkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information Fusion*, vol. 76, p. 243–297, Dec. 2021. [Online]. Available: <http://dx.doi.org/10.1016/j.inffus.2021.05.008>
- [459] M. Kraus and S. Feuerriegel, "Forecasting remaining useful life: Interpretable deep learning approach via variational bayesian inferences," *Decision Support Systems*, vol. 125, p. 113 100, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0167923619301290>
- [460] B. Wang, Y. Lei, T. Yan, N. Li, and L. Guo, "Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery," *Neurocomputing*, vol. 379, pp. 117–129, 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231219315024>
- [461] Z. Zhao, J. Wu, D. Wong, C. Sun, and R. Yan, "Probabilistic remaining useful life prediction based on deep convolutional neural network," in *9th International Conference on Through-life Engineering Service*, ser. SSRN Electronic Journal. Social Science Research Network, Oct. 2020.
- [462] L. D. Libera, J. Andreoli, D. D. Pezze, M. Ravanelli, and G. A. Susto, "Bayesian deep learning for remaining useful life estimation via stein variational gradient descent," 2024. [Online]. Available: <https://arxiv.org/abs/2402.01098>
- [463] M. Benker, L. Furtner, T. Semm, and M. F. Zaeh, "Utilizing uncertainty information in remaining useful life estimation via bayesian neural networks and hamiltonian monte carlo," *Journal of Manufacturing Systems*, vol. 61, pp. 799–807, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0278612520301928>
- [464] Liu, S., C. Lv, F. Song *et al.*, "Remaining useful life prediction integrating working conditions and uncertainty quantification based on multilayer graph neural networks," *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 2025. [Online]. Available: <https://doi.org/10.1007/s40430-025-05400-8>
- [465] D. Frizzo, F. Borsatti, and G. A. Susto, "A quantile regression approach for remaining useful life estimation with state space models," 2025. [Online]. Available: <https://arxiv.org/abs/2506.17018>
- [466] Y. Yin, J. Zhao, X. Zhang, and Y. Liu, "A novel remaining useful life prediction approach combined extreme gradient boosting and multiquantile recurrent neural network," *IEEE Access*, vol. 12, pp. 44 648–44 658, 2024.
- [467] R. Wen, K. Torkkola, and B. Narayanaswamy, "A multi-horizon quantile recurrent forecaster," 11 2017.
- [468] T. Khan, K. Ahmad, J. Khan, I. Khan, and N. Ahmad, "An explainable regression framework for predicting remaining useful life of machines," in *2022 27th International Conference on Automation and Computing (ICAC)*, 2022, pp. 1–6.
- [469] J. Figueroa Barraza, E. Drogue, and M. Ramos Martins, "Towards interpretable deep learning: A feature selection framework for prognostics and health management using deep neural networks," *Sensors*, vol. 21, p. 5888, 09 2021.
- [470] J. Grezma, J. Zhang, P. Wang, K. Loparo, and R. Gao, "Interpretable convolutional neural network through layer-wise relevance propagation for machine fault diagnosis," *IEEE Sensors Journal*, vol. PP, pp. 1–1, 12 2019.
- [471] L. Rajaoarisoa, R. Randrianandraina, G. J. Nalepa, and J. a. Gama, "Decision-making systems improvement based on explainable artificial intelligence approaches for predictive maintenance," *Eng. Appl. Artif. Intell.*, vol. 139, no. PB, Feb. 2025. [Online]. Available: <https://doi.org/10.1016/j.engappai.2024.109601>
- [472] S. Vollert, M. Atzmueller, and A. Theissler, "Interpretable machine learning: A brief survey from the predictive maintenance perspective," in *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2021, pp. 01–08.
- [473] D. Solís-Martín, J. Galán-Páez, and J. Borrego-Díaz, "On the soundness of xai in prognostics and health management (phm)," 2023. [Online]. Available: <https://arxiv.org/abs/2303.05517>
- [474] R. K. Kundu and K. A. Hoque, "Explainable predictive maintenance is not enough: Quantifying trust in remaining useful life estimation," *Annual Conference of the PHM Society*, vol. 15, no. 1, 2023. [Online]. Available: <https://doi.org/10.36001/phmconf.2023.v15i1.3472>
- [475] L. Cummins, A. Sommers, S. B. Ramezani, S. Mittal, J. Jabour, M. Seale, and S. Rahimi, "Explainable predictive maintenance: A survey of current methods, challenges and opportunities," 2024. [Online]. Available: <https://arxiv.org/abs/2401.07871>
- [476] M. Achouch, M. Dimitrova, K. Ziane, S. Sattarpanah Karganroudi, R. Dhouib, H. Ibrahim, and M. Adda, "On predictive maintenance in Industry 4.0: Overview, models, and challenges," *Applied Sciences*, vol. 12, no. 16, p. 8081, 2022.
- [477] S. Arulnithika, D. R. Marur, and P. Taranalakshmi, "Iot based predictive maintenance in manufacturing," in *2025 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI)*, 2025, pp. 1–5.
- [478] H. Singh, "Iot based predictive maintenance in industrial machinery," *International Journal for Modern Trends in Science and Technology*, vol. 11, no. 1, pp. 98–101, 07 2025. [Online]. Available: <https://feplb.com/index.php/csitrj/article/view/716>
- [479] D. Dhinakaran, S. Edwin Raja, R. Velselvi, S. Senthil Kumar, and S. Senthil Kumar, "Intelligent iot-driven advanced predictive maintenance system for industrial applications," *SN COMPUT. SCI.*, vol. 6, p. 151, 12 2025.
- [480] S. Elkateb, A. Métwalli, A. Shendy, and A. E. Abu-Elanien, "Machine learning and iot – based predictive maintenance approach for industrial applications," *Alexandria Engineering Journal*, vol. 88, pp. 298–309, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1110016823011572>
- [481] N. Longhi, L. M. Amorosa, S. Cavallero, E. Buracchini, and R. Verdone, "5g architectures enabling remaining useful life estimation for industrial iot: A holistic study," *IEEE Open Journal of the Communications Society*, pp. 1–1, 2025.
- [482] L. Lorenti, D. D. Pezze, J. Andreoli, C. Masiero, N. Gentner, Y. Yang, and G. A. Susto, "Predictive maintenance in the industry: A comparative study on deep learning-based remaining useful life estimation," in *2023 IEEE 21st International Conference on Industrial Informatics (INDIN)*, 2023, pp. 1–9.
- [483] N. Alshahrani, "Machine learning approaches for predictive maintenance in industrial operations," in *2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2024, pp. 365–372.
- [484] A. Al-Rafea, M. Al-atrash, and N. Lepkova, "Prediction of the remaining useful life of a milling machine using machine learning," *MethodsX*, vol. 14, p. 103 195, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2215016125000433>
- [485] S. Cancemi, M. Angelucci, A. Chierici, S. Paci, and R. Lo Frano, "Hybrid neural network and statistical forecasting methodology for predictive monitoring and residual useful life estimation in nuclear power plant components," *Nuclear Engineering and Design*, vol. 433, p. 113 900, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0029549325000779>
- [486] F. Maione, P. Lino, G. Maione, and G. Giannino, "Predictive maintenance of marine engines using a long short-term memory algorithm," in *2024 IEEE 8th Forum on Research and Technologies for Society and Industry Innovation (RTSI)*, 2024, pp. 578–583.
- [487] S. Lim, J. Oh, and J. Park, "Maintenance time prediction for predictive maintenance of ship engines," *Applied Sciences*, vol. 15, p. 4764, 04 2025.
- [488] J. Park and J. Oh, "A machine learning based predictive maintenance algorithm for ship generator engines using engine simulations and collected ship data," *Energy*, vol. 285, no. C, p. None, None 2023. [Online]. Available: <https://ideas.repec.org/a/eee/energy/v285y2023ics0360544223026634.html>
- [489] F. Kabir, B. Foggo, and N. Yu, "Data driven predictive maintenance of distribution transformers," in *2018 China International Conference on Electricity Distribution (CICED)*, 2018, pp. 312–316.
- [490] N. Bolander, H. Qiu, N. Eklund, E. Hindle, and T. Rosenfeld, "Physicsbased remaining useful life prediction for aircraft engine bearing prognosis," *Annual Conference of the Prognostics and Health Management Society*, 2009.
- [491] M. J. Daigle and K. Goebel, "A model-based prognostics approach applied to pneumatic valves," *Int J Progn Health Manag*, vol. 2, pp. 1–16, 2011.

- [492] — —, “Model-based prognostics with concurrent damage progression processes,” *IEEE Trans Syst Man Cybern A*, vol. 43, no. 3, pp. 535–546, 2013.
- [493] D. Zhang, S. Dey, H. E. Perez, and S. J. Moura, “Remaining useful life estimation of lithium-ion batteries based on thermal dynamics,” in *2017 American Control Conference (ACC)*, 2017, pp. 4042–4047.
- [494] M. M. Özcan and U. Kutbay, “Improving predictive maintenance performance with limited data under different operating conditions with transfer learning,” in *2024 15th National Conference on Electrical and Electronics Engineering (ELECO)*, 2024, pp. 1–5.
- [495] X. Chen, Z. Chen, M. Zhang, Z. Wang, M. Liu, M. Fu, and P. Wang, “A remaining useful life estimation method based on long short-term memory and federated learning for electric vehicles in smart cities,” *PeerJ Computer Science*, 2023. [Online]. Available: <https://peerj.com/articles/cs-1652/>
- [496] L. Chen, H. Gao, L. Guo, J. Liang, and L. Peng, “Remaining useful life prediction of machinery using federated public feature representation in edge-cloud collaboration architecture,” *Engineering Applications of Artificial Intelligence*, vol. 143, p. 110 059, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0952197625000594>
- [497] M. L. Baptista, K. Goebel, and E. M. Henriques, “Relation between prognostics predictor evaluation metrics and local interpretability shap values,” *Artificial Intelligence*, vol. 306, p. 103 667, 2022. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0004370222000078>
- [498] X. Liu, W. Cheng, J. Xing, X. Chen, L. Li, Y. Guan, B. Ding, Z. Nie, R. Zhang, and Y. Zhi, “Predictive maintenance system for high-end equipment in nuclear power plant under limited degradation knowledge,” *Advanced Engineering Informatics*, vol. 61, p. 102 506, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S147403462400154X>
- [499] G. Sateesh Babu, P. Zhao, and X.-L. Li, “Deep convolutional neural network based regression approach for estimation of remaining useful life,” in *Database Systems for Advanced Applications*, S. B. Navathe, W. Wu, S. Shekhar, X. Du, X. S. Wang, and H. Xiong, Eds. Cham: Springer International Publishing, 2016, pp. 214–228.
- [500] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” 2023. [Online]. Available: <https://arxiv.org/abs/1706.03762>
- [501] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, “Informer: Beyond efficient transformer for long sequence time-series forecasting,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 11 106–11 115, 05 2021.
- [502] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *North American Chapter of the Association for Computational Linguistics*, 10 2018.
- [503] N. Tagasovska and D. Lopez-Paz, “Single-model uncertainties for deep learning,” 2019. [Online]. Available: <https://arxiv.org/abs/1811.00908>
- [504] J. Shang, D. Xu, H. Qiu, L. Gao, C. Jiang, and P. Yi, “A novel data augmentation framework for remaining useful life estimation with dense convolutional regression network,” *Journal of Manufacturing Systems*, vol. 74, pp. 30–40, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0278612524000347>
- [505] S. J. Julier and J. K. Uhlmann, “New extension of the Kalman filter to nmiscar systems,” in *Signal Processing, Sensor Fusion, and Target Recognition VI, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, I. Kadar, Ed., vol. 3068, Jul. 1997, pp. 182–193.
- [506] M. Raissi, P. Perdikaris, and G. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *Journal of Computational Physics*, vol. 378, pp. 686–707, 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0021999118307125>
- [507] A. Ucar, M. Karakose, and N. Kırımça, “Artificial intelligence for predictive maintenance applications: Key components, trustworthiness, and future trends,” *Applied Sciences*, vol. 14, no. 2, 2024. [Online]. Available: <https://www.mdpi.com/2076-3417/14/2/898>
- [508] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,” 2013. [Online]. Available: <https://arxiv.org/abs/1312.5602>
- [509] V. Konda and J. Tsitsiklis, “Actor-critic algorithms,” in *Advances in Neural Information Processing Systems*, S.olla, T. Leen, and K. Müller, Eds., vol. 12. MIT Press, 1999. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/1999/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf
- [510] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal policy optimization algorithms,” 2017. [Online]. Available: <https://arxiv.org/abs/1707.06347>
- [511] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, “Deterministic policy gradient algorithms,” *31st International Conference on Machine Learning, ICML 2014*, vol. 1, 06 2014.
- [512] M. Feng and Y. Li, “Predictive maintenance decision making based on reinforcement learning in multistage production systems,” *IEEE Access*, vol. 10, pp. 18 910–18 921, 2022.
- [513] K. S. H. Ong, W. Wang, D. Niyato, and T. Friedrichs, “Deepreinforcement-learning-based predictive maintenance model for effective resource management in industrial iot,” *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5173–5188, 2022.
- [514] K. S. H. Ong, W. Wang, N. Q. Hieu, D. Niyato, and T. Friedrichs, “Predictive maintenance model for iiot-based manufacturing: A transferable deep reinforcement learning approach,” *IEEE Internet of Things Journal*, vol. 9, no. 17, pp. 15 725–15 741, 2022.
- [515] J. Lee and M. Mitici, “Deep reinforcement learning for predictive aircraft maintenance using probabilistic remaininguseful-life prognostics,” *Reliability Engineering & System Safety*, vol. 230, p. 108908, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0951832022005233>
- [516] A. N. Abbas, G. C. Chasparis, and J. D. Kelleher, “Hierarchical framework for interpretable and specialized deep reinforcement learning-based predictive maintenance,” *Data & Knowledge Engineering*, vol. 149, p. 102240, 2024. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0169023X23001003>
- [517] L. Brito, G. A. Susto, J. Brito, and M. Duarte, “Fault detection of bearing: An unsupervised machine learning approach exploiting feature extraction and dimensionality reduction,” *Informatics*, vol. 8, p. 85, 11 2021.
- [518] A. Agárdi and K. Nehéz, “Failure prediction with Weibull distribution,” *Infocommunications Journal*, vol. 17, no. 5, pp. 33–41, 2025.
- [519] H. K. Wang, X. Dai, Q. Ran, and et al., “Lithium-ion battery remaining useful life prediction based on dynamic filter frequency mixing learner and dual-stream mamba,” *Ionics*, 2025. [Online]. Available: <https://doi.org/10.1007/s11581-025-06715-1>