A Systematic Review of Anomaly and Fault Detection Using Machine Learning for Industrial Machinery

ZAIDI, Haseeb <http://orcid.org/0009-0003-7357-5200>, SHENFIELD, Alex <http://orcid.org/0000-0002-2931-8077>, ZHANG, Hongwei <http://orcid.org/0000-0002-7718-021X> and IKPEHAI, Augustine <http://orcid.org/0000-0002-5254-8188>

algorithms

MDPI

Review

A Systematic Review of Anomaly and Fault Detection Using Machine Learning for Industrial Machinery

Syed Haseeb Haider Zaidi [1,*], Alex Shenfield [1,*], Hongwei Zhang [1] and Augustine Ikpehai [2]

[1] Advanced Food Innovation Centre (AFIC), Sheffield Hallam University, Sheffield S9 2AA, UK; h.zhang@shu.ac.uk
[2] School of Engineering and Built Environment, Sheffield Hallam University, Sheffield S1 1WB, UK; a.ikpehai@shu.ac.uk
* Correspondence: h.zaidi@shu.ac.uk (S.H.H.Z.); a.shenfield@shu.ac.uk (A.S.)

Abstract

Unplanned downtime in industrial machinery remains a major challenge, causing substantial economic losses and safety risks across sectors such as manufacturing, food processing, oil and gas, and transportation. This systematic review investigates the application of machine learning (ML) techniques for anomaly and fault detection within the broader context of predictive maintenance. Following a hybrid review methodology, relevant studies published between 2010 and 2025 were collected from major databases including IEEE Xplore, ScienceDirect, SpringerLink, Scopus, Web of Science, and arXiv. The review categorizes approaches into supervised, unsupervised, and hybrid paradigms, analyzing their pipelines from data collection and preprocessing to model deployment. Findings highlight the effectiveness of deep learning architectures such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders, and hybrid frameworks in detecting faults from time series and multimodal sensor data. At the same time, key limitations persist, including data scarcity, class imbalance, limited generalizability across equipment types, and a lack of interpretability in deep models. This review concludes that while ML-based predictive maintenance systems are enabling a transition from reactive to proactive strategies, future progress requires improved hybrid architectures, Explainable AI, and scalable real-time deployment.

Keywords: anomaly detection; fault detection; machine learning; industrial machinery; predictive maintenance

Academic Editors: Chih-Lung Lin, Bor-Jiunn Hwang, Shaou-Gang Miaou and Chi-Hung Chuang

Received: 3 December 2025
Revised: 22 January 2026
Accepted: 23 January 2026
Published: 1 February 2026

1. Introduction

Unplanned downtime, or unscheduled production halts from equipment failure, poses a persistent and costly challenge in modern manufacturing. Studies consistently show that downtime consumes between 5% and 20% of total production time across various sectors [1,2], with one global survey finding that 82% of companies experienced at least one significant downtime incident in the past three years [3]. The economic implications are significant; Deloitte estimates that unplanned downtime costs industrial manufacturers approximately USD 50 billion annually on a global scale [4]. For large scale automotive manufacturing plants, unplanned downtime can result in extremely high losses, hourly downtime costs can exceed USD 2 million, accounting for lost output, overtime, and logistical penalties [5]. Beyond direct financial losses, downtime can lead to wasted raw materials, missed deadlines, and supply chain disruptions, particularly in just-in-time

(JIT) systems [6]. Furthermore, equipment failures often precede industrial accidents, as tragically illustrated by the Deepwater Horizon disaster where maintenance failures contributed to an explosion that resulted in 11 deaths [7].

Traditional maintenance strategies have proven inadequate in addressing these multifaceted challenges. Reactive maintenance, which involves repairing equipment only after it fails, is simple but leads to unpredictable downtime and high long-term costs [8]. Preventive maintenance, performed on a fixed schedule, mitigates some risks but can result in unnecessary part replacements or missed failures that occur between inspections [9]; meanwhile, an improved approach, condition-based maintenance (CBM), still depends heavily on sophisticated sensors and monitoring infrastructure, meaning that maintenance actions are often still triggered only after monitored indicators cross predefined thresholds, making the approach partly reactive in practice [1]. These limitations have created a compelling need for a more intelligent, proactive approach.

The solution lies in the paradigm shift toward predictive maintenance (PdM), a strategy enabled by the fourth industrial revolution (Industry 4.0). By leveraging technologies like the Internet of Things (IoT), artificial intelligence (AI), and big data analytics, PdM transforms maintenance from a reactive or scheduled task into a data-driven process of anticipating failure [10]. Sensors collect real-time data on key parameters such as vibration, temperature, and pressure. Sophisticated machine learning (ML) algorithms then analyze this data to detect subtle signs of degradation, predict a machine's remaining useful life (RUL), and enable maintenance to be performed precisely when needed, minimizing disruption. This approach offers significant benefits, with a study from the U.S. Department of Energy suggesting a transition to PdM can reduce downtime by up to 35% and lower maintenance costs by 8 to 12% compared to preventive methods [11,12].

A core technological component of PdM is fault detection and diagnosis (FDD), which focuses on identifying and classifying known fault conditions. Model-based FDD relies on detailed process knowledge, while data-driven FDD increasingly uses machine learning to interpret sensor data [13]. However, industrial systems also encounter previously unseen or weakly labeled faults, which traditional FDD cannot capture. This is where anomaly detection methods are essential, as they identify deviations from normal behavior without requiring prior fault definitions. Online and unsupervised learning supports this by adapting to continuously evolving industrial data streams. A major challenge in both areas is interpretability. Many deep learning models operate as "black boxes," making it difficult for engineers to understand why a fault or anomaly is flagged. As a result, PdM research increasingly incorporates Explainable AI (XAI) to produce human-interpretable insights for root-cause analysis [14]. PdM also influences business models by enabling "maintenance-as-a-service," helping manufacturers reduce downtime, strengthen customer relationships, and improve competitiveness across sectors such as energy and transport [15,16].

This systematic review aims to provide a comprehensive analysis of fault and anomaly detection in the context of predictive maintenance. The primary objective is to address key knowledge gaps by answering the following research questions:

- To what extent have machine learning and deep learning methodologies supplanted traditional signal processing for fault and anomaly detection?
- What are the prevailing applications, challenges, and successful implementations of these techniques within the manufacturing sector?
- What are the key research gaps and future directions that will define the next generation of intelligent maintenance systems?

Section 2 outlines the methodology, including search strategy, study selection, data extraction, and synthesis employed in this study. Section 3 introduces conventional fault and anomaly detection approaches, covering industrial machinery types, common faults,

https://doi.org/10.3390/a19020108

and signal processing techniques. Section 4 reviews machine-learning-based approaches, focusing on supervised, unsupervised, and hybrid methods. Section 5 examines the applications of these techniques across different industries, including manufacturing, food and beverage, oil and gas, and transportation. Finally, Section 6 concludes the review by summarizing the main findings, discussing methodological limitations, and identifying directions for future research.

2. Methodology

This review adopts a hybrid methodological framework, combining elements of systematic mapping studies and scoping reviews. This approach is specifically tailored to the interdisciplinary domain of machine learning (ML) applications in fault and anomaly detection for industrial machinery. The chosen hybrid approach uniquely supports both structured exploration and thematic synthesis, which is especially appropriate for fast-evolving technical fields like applied machine learning, where diverse publication formats, experimental settings, and evaluation standards make rigid review frameworks (such as PRISMA) overly restrictive. The highly heterogeneous nature of study designs and reporting standards in this research area precludes strict adherence to frameworks primarily designed for quantitative synthesis, necessitating a more flexible yet rigorous approach. Drawing on the foundational scoping review framework proposed by Arksey and O'Malley [17], further refined by Levac et al. [18] and incorporating methodological insights from researchers, this review adopts a hybrid approach that emphasizes transparency, adaptability, and comprehensiveness.

2.1. Search Strategy

The literature search was designed to be comprehensive, iterative, and multi-source. To capture both peer-reviewed and emerging work, the following databases and repositories were systematically searched: IEEE Xplore, ScienceDirect, SpringerLink, Scopus, Web of Science, Google Scholar, and arXiv. These databases were selected for their extensive coverage of engineering, computer science, and industrial applications, ensuring a broad capture of both theoretical and applied research. Additionally, relevant technical reports and industry white papers were included when issued by reputable institutions or professional bodies, identified through targeted searches on key industrial organization websites and professional society publications. Search terms were formulated using Boolean logic and refined through a systematic, iterative process. Initial pilot searches were conducted to identify key terms and adjust Boolean operators, with search strings iteratively refined based on the relevance of preliminary results from each major database until saturation was observed. Key phrases included combinations of "machine learning", "fault detection", "anomaly detection", "predictive maintenance", "industrial machinery", and "manufacturing systems". A representative example of a search query used was: "machine learning" AND ("fault detection" OR "anomaly detection") AND ("industrial machinery" OR "manufacturing equipment"). Filters were applied where possible to narrow results to relevant fields such as engineering, computer science, and industrial technology. The search included publications from 1 January 2010, to 31 May 2025, capturing both foundational research and the latest developments.

2.2. Study Selection Criteria

A set of explicit inclusion and exclusion criteria was defined to guide the filtering process and ensure that only relevant and sufficiently robust studies were considered. To be included, studies were required to apply machine learning methods—whether supervised, unsupervised, deep learning, or hybrid models—for the purpose of detecting faults or anomalies in industrial settings. This scope specifically encompassed machinery used in food and beverage manufacturing as a subset of the broader industrial category, while studies targeting non-industrial domains such as finance, healthcare, or purely theoretical signal processing without industrial application were excluded. Methodologically, eligible studies had to employ machine learning as the primary approach for fault or anomaly detection; those relying solely on traditional rule-based methods, statistical process control (SPC), or classical signal processing without a significant ML component were excluded. Nonetheless, these conventional approaches are briefly reviewed to contextualize the field and illustrate the progression toward ML-based fault and anomaly detection. Furthermore, studies needed to provide empirical evidence through real-world deployment, simulated experiments based on industrially relevant scenarios, or validation against publicly available or private datasets with clear methodological descriptions. Purely conceptual papers, architectural proposals without implementation, or review articles lacking novel empirical contributions were not considered. With respect to publication type, only peer-reviewed journal articles, conference papers, pre-prints from reputable repositories such as arXiv, and technical reports or white papers issued by recognized institutions or professional bodies were accepted. Finally, only publications written in English were included in the review.

The selection process was conducted in three phases to systematically filter the identified literature:

- Initial screening (titles and abstracts): Screened all identified titles and abstracts against the preliminary inclusion/exclusion criteria. Studies that were clearly irrelevant were discarded.
- Full-text review: Articles that passed the initial screening, or whose relevance could not be determined from the abstract alone, underwent a full-text review. During this phase, a more detailed assessment against all inclusion criteria was performed.
- Conflict resolution and final selection: Any discrepancies during both the initial screening and full-text review stages were resolved.

Throughout the selection process, the primary reason for exclusion was systematically logged for articles rejected during the full text review stage to ensure transparency and reproducibility, while a formal PRISMA flowchart was not utilized due to the hybrid nature of this review and the diversity of study types, the entire selection process was meticulously documented to provide an audit trail and support future replication. This rigorous approach, incorporating independent review, was employed to minimize selection bias and enhance the reliability of the chosen studies. The database search yielded a total of 1142 records after duplicate removal. Following title and abstract screening, 812 records were excluded due to clear irrelevance to industrial fault or anomaly detection using machine learning. The remaining 330 articles underwent full text assessment, of which 252 were excluded for reasons including lack of an industrial context, absence of a substantive machine learning component, or insufficient empirical validation. Ultimately, 78 studies met all inclusion criteria and were retained for data extraction and synthesis. Table 1 summarizes the screening and selection process.

Table 1. Summary of the literature screening and study selection process.

Screening Stage	Description	Number of Records
Records identified	Records retrieved from databases and gray literature sources after duplicate removal	1142
Title and abstract screening	Records screened against preliminary inclusion and exclusion criteria	1142
Records excluded	Clearly irrelevant studies removed during title and abstract screening	812

https://doi.org/10.3390/a19020108

Table 1. *Cont.*

Screening Stage	Description	Number of Records
Full-text articles assessed	Articles reviewed in full for eligibility against all inclusion criteria	330
Full-text articles excluded	Articles excluded with documented reasons (e.g., non-industrial focus, lack of a substantive machine learning component, insufficient empirical validation)	252
Final studies included	Studies retained for data extraction and synthesis	78

2.3. Data Extraction and Categorization

The data extraction process focused not only on technical aspects of the ML models but also on contextual factors such as dataset type, application environment, and reported limitations. The following Table 2 summarizes the core data categories extracted from each included study.

Table 2. Data extraction categories and their descriptions used in the systematic review.

Category	Description
Publication details	Author(s), publication year, and venue (journal, conference, etc.).
ML techniques used	Type of model(s) used: supervised, unsupervised, deep learning, and hybrid methods.
Industrial context	Domain of application (e.g., packaging systems, robotics, food processing).
Fault/anomaly types	Specific failure types detected (e.g., wear, misalignment, sensor faults).
Dataset characteristics	Dataset origin (real-world or simulated), public/private status, size, and diversity.
Feature engineering	Sensor inputs, process variables, control parameters, and preprocessing techniques.
Evaluation metrics	Performance measures (e.g., accuracy, F1-score, precision, AUC).
Results and findings	Summary of model outcomes and reported effectiveness.
Challenges identified	Limitations discussed (e.g., data imbalance, lack of interpretability).
Future directions	Suggestions for improvement (e.g., Explainable AI, edge deployment).

2.4. Synthesis and Analysis

Data synthesis was carried out in two stages. First, a quantitative mapping of the extracted data was conducted to identify macro-level patterns in algorithm usage, dataset types, industrial domains, and evaluation practices. This allowed the identification of dominant research themes—for example, the frequent use of deep learning models in fault prediction for complex machinery, or the reliance on synthetic datasets in simulation heavy studies. Second, a thematic analysis was performed to extract deeper insights. Themes were identified inductively from the extracted data, categorizing recurring patterns in challenges, trends, and application contexts. Recurring challenges were grouped under themes such as data scarcity, real-time implementation complexity, generalizability across machinery types, and the opacity of model decisions. The review also highlighted methodological trends such as increased adoption of hybrid models combining data-driven and physics-based components, as well as rising interest in model explainability, particularly in safety-critical applications. Where applicable, comparative analysis was used to contrast approaches across sectors. For instance, ML techniques applied in discrete manufacturing (e.g., electronics assembly) tended to rely on structured sensor networks, while process industries (such as chemical or food production) often required time series models and multimodal data fusion.

2.5. Methodological Limitations

While this review was designed to be rigorous and inclusive, several limitations are acknowledged. First, by restricting the review to English-language sources, some relevant non-English studies may have been excluded; however, initial scoping did not reveal a significant body of non-English literature that would fundamentally alter the core findings. Second, while broad in scope, the review may have missed relevant publications indexed in databases not included in the search protocol. Third, the inclusion of preprints and gray literature, while valuable for capturing cutting-edge research, means that some findings may not have undergone formal peer review. Another acknowledged limitation is the absence of a formal quality assessment tool. Unlike traditional systematic reviews in healthcare, where bias and evidence levels are formally rated, the diverse nature of ML applications and evaluation settings often lack standardized benchmarks or control groups common in clinical trials. This precluded the direct application of traditional risk of bias tools like Cochrane's. Instead, this review evaluated methodological robustness contextually based on the clarity of reporting, presence of empirical evaluation, and transparency in results, which introduces some subjectivity. Despite these constraints, the methodology employed in this review offers a comprehensive and adaptable framework suitable for synthesizing current research in this dynamic field. It balances methodological structure with exploratory depth, providing both a high-level overview of the state of research and insights into technical and practical challenges.

The methodological framework presented in this section provides the foundation for analyzing the wide spectrum of approaches used in fault and anomaly detection. Before examining modern machine learning techniques, it is important to first understand the conventional signal processing and diagnostic methods upon which many data-driven models are built. Section 3 therefore introduces the traditional fault and anomaly detection approaches that have shaped the evolution of predictive maintenance and continue to serve as essential baselines in industrial practice.

3. Conventional Fault and Anomaly Detection

This section defines the foundational concepts, technological enablers, and analytical frameworks that underpin the field of intelligent maintenance, focusing on the conventional approaches to fault and anomaly detection. A key driver for this field is the need for proactive failure anticipation, which is powered by these analytical capabilities.

3.1. Industrial Machinery Overview and Types of Faults

This sub-section provides a focused overview of common industrial machines and the categorization of their faults and anomalies. A deep understanding of the equipment and its potential failure modes is crucial for the effective implementation of any data-driven system [19]. Predictive maintenance is applicable to a wide range of industrial machinery, and the principles of fault and anomaly detection are versatile enough to be adapted across different machine types. Among the most common examples are pumps, motors, and turbines, which serve as foundational components in many industrial processes.

Pumps, essential for transporting fluids, are susceptible to several failure mechanisms such as cavitation, bearing wear, impeller damage, and seal leakage [20]. These issues can be detected through sensor-based monitoring that measures pressure, flow rate, temperature, and vibration. Motors, often described as the workhorses of industry, also present a variety of potential faults. Mechanical issues include bearing or rotor failures, while electrical

https://doi.org/10.3390/a19020108

problems may arise from insulation breakdown, winding shorts, or phase imbalances [21]. Effective condition monitoring of motors typically involves vibration analysis, thermal measurements, and motor current signature analysis (MCSA) [22]. Turbines, which play critical roles in power generation and propulsion, are complex systems where faults can lead to catastrophic outcomes. Common turbine faults include blade erosion, imbalance, and bearing degradation. These are usually monitored through extensive sensor networks that track vibration, temperature, pressure, and acoustic emissions [23].

At the core of any predictive maintenance system are the analytical methods that distinguish between faults and anomalies. Fault detection is defined as the process of identifying a specific, known type of malfunction [24], whereas anomaly detection refers to identifying data points or patterns that deviate from expected normal behavior, often highlighting novel or unforeseen problems [25]. Industrial faults are generally divided into two broad categories: mechanical and electrical. Mechanical faults are physical in nature, typically associated with moving parts. For instance, bearing wear generates high-frequency impacts that appear as spikes in vibration signals [26], while imbalance and misalignment in rotating machinery produce strong vibrations at the fundamental rotational frequency ($1\times$ RPM) and its harmonics. Gear tooth damage, including wear or cracking, can be identified by the appearance of sidebands around the gear mesh frequency in vibration spectra. Electrical faults, on the other hand, occur within the electrical subsystems of machinery, particularly in motors. Examples include rotor bar cracks, stator winding failures, and anomalies in voltage or current, which are often detected using MCSA or thermal sensors [22,27].

Anomalies themselves can also be further classified based on their nature [28]. Point anomalies refer to individual data points that significantly differ from the remainder of the dataset. Contextual anomalies are only anomalous in relation to a specific context, such as temperature readings that may be normal in one operating condition but abnormal in another. Collective anomalies describe groups of related data instances that, when considered together, deviate from expected patterns. Finally, temporal anomalies represent events that occur at incorrect or unexpected times, a particularly relevant category for time series data obtained from industrial sensors [28].

3.2. Signal Processing Techniques

In predictive maintenance, fault and anomaly detection typically rely on a structured pipeline consisting of data collection, preprocessing and feature extraction, model application, and fault identification [29]. Signal processing plays a crucial role in the preprocessing and feature extraction stage, as raw sensor measurements such as vibration, acoustic, or ultrasonic signals are often noisy and not directly interpretable by machine learning algorithms [30,31]. By applying signal processing methods—such as Fourier transforms, wavelets, or graph-based spectral techniques—raw data can be transformed into discriminative features that capture key characteristics of machine condition, thereby improving the performance of fault detection and diagnosis models [32].

Time series signals can be analyzed primarily in two domains: the time domain and the frequency domain. Time-domain analysis evaluates the raw signal as it evolves over time, extracting statistical features such as root mean square (RMS), standard deviation, and kurtosis to capture the signal's energy and impulsive characteristics [33,34]. Although this approach is straightforward and computationally efficient, it is often insufficient on its own to discriminate between different types of faults. Frequency-domain analysis, on the other hand, transforms the signal to reveal its constituent frequency components, making it particularly effective for detecting cyclic phenomena commonly associated with rotating machinery, such as imbalance, misalignment, and gear wear [35]. Fast Fourier Transform

(FFT) is the primary tool used in this context, as it decomposes a time-domain signal into its frequency spectrum, thereby highlighting dominant frequencies and their amplitudes [34].

A variety of sophisticated techniques are used for signal processing to extract features for fault detection. These include the following:

- Envelope analysis: This is a specialized demodulation technique used to detect low-frequency impulses embedded in high-frequency carrier signals. It is primarily used for identifying faults in rolling-element bearings and gearboxes [26]. The process involves band-pass filtering the raw vibration signal, rectifying it, and then applying a low-pass filter to extract the envelope. A Fast Fourier Transform (FFT) is then performed on this envelope signal to reveal fault-related frequencies that would otherwise be hidden [36].
- Spectral analysis: This is a foundational technique in machinery diagnostics. Spectral analysis uses Fast Fourier Transform (FFT) to convert a time-domain signal into the frequency domain. The resulting plot, a spectrum, shows the amplitude of each frequency component. This is essential for identifying faults that occur at specific, predictable frequencies, such as imbalances ($1\times$ RPM), misalignment ($2\times$ RPM), and gear mesh frequencies [37].
- Wavelets analysis: This offers an advanced method that addresses a major limitation of FFT—its inability to provide time-localized frequency information. Wavelets provide high resolution in both the time and frequency domains, making them ideal for analyzing non-stationary signals and detecting transient faults, such as those from crack propagation or intermittent impacts [38]. Discrete Wavelet Transform (DWT) is commonly used to decompose a signal into different frequency sub-bands, with each band retaining its temporal information. This allows for the precise localization of a fault in both time and frequency, providing more detailed diagnostic information than traditional spectral analysis [37,39].

Figure 1 presents a set of signal processing analysis applied to the same vibration signal. Subplot (a) shows the vibration waveform in the time domain, illustrating the raw amplitude fluctuations. Subplots (b) and (c) display the frequency domain representation using Fast Fourier Transform (FFT): (b) shows the full spectrum, while (c) provides a zoomed view of the lower frequency region to highlight dominant components such as rotational and fault related frequencies. Subplot (d) shows the envelope of the signal obtained using the Hilbert transform, revealing the amplitude modulation pattern over time. Subplot (e) presents the envelope spectrum, which is the frequency domain representation of the envelope and is used to identify modulation and characteristic fault frequencies. Finally, subplot (f) shows the Continuous Wavelet Transform (CWT) scalogram, providing a time frequency representation suitable for detecting transient and non-stationary features in the vibration signal.

Although conventional signal processing methods remain fundamental for feature extraction, interpretability, and linking machine behavior to physical principles, they rely heavily on expert knowledge and predefined assumptions. They can also struggle to scale to complex, nonlinear systems or to adapt under changing operating conditions. Machine learning addresses these limitations by enabling automated feature learning and data-driven pattern recognition across diverse machinery and environments. Importantly, ML does not replace signal processing entirely; instead, many predictive maintenance pipelines integrate signal processing as preprocessing or feature engineering to improve robustness and interpretability. Building on this foundation, Section 4 examines supervised, unsupervised, and hybrid machine learning methods for fault and anomaly detection.

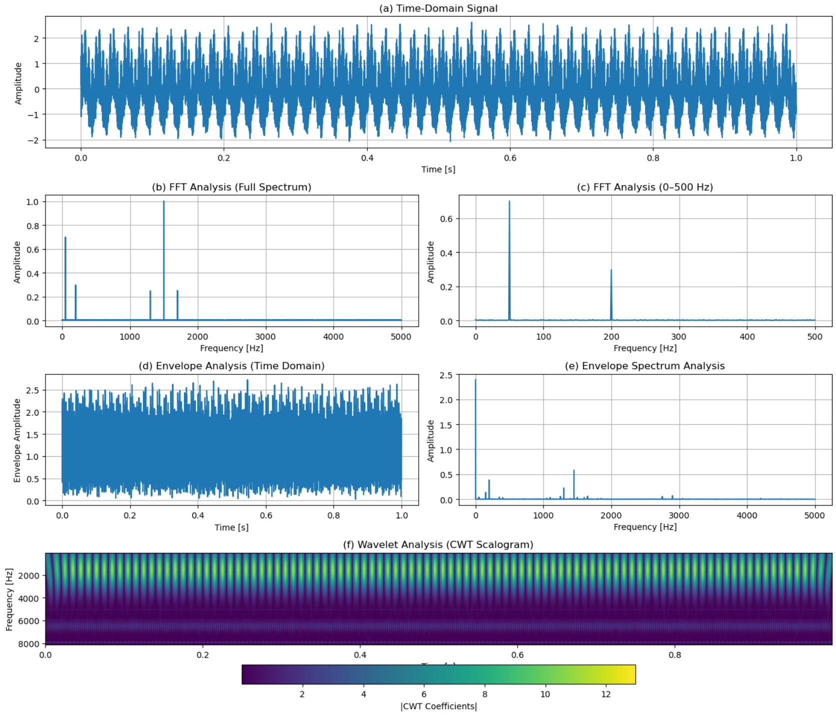

Figure 1. Vibration-signal analysis: (**a**) time-domain waveform; (**b**) full FFT spectrum; (**c**) zoomed FFT (0–500 Hz); (**d**) envelope signal; (**e**) envelope spectrum; (**f**) CWT scalogram.

4. Machine Learning Approaches

The increasing complexity and interconnectedness of modern industrial machinery have accelerated the transition from rule-based and signal-driven diagnostics toward data-driven predictive maintenance systems, while conventional signal processing provides physically meaningful representations of machine behavior, machine learning techniques extend these capabilities by learning discriminative patterns directly from data, either using engineered features or end-to-end architectures. Accordingly, this section reviews supervised, unsupervised, and hybrid machine learning paradigms for fault and anomaly detection, highlighting how they complement and extend traditional diagnostic pipelines. Figure 2 presents a schematic overview of these machine learning approaches.

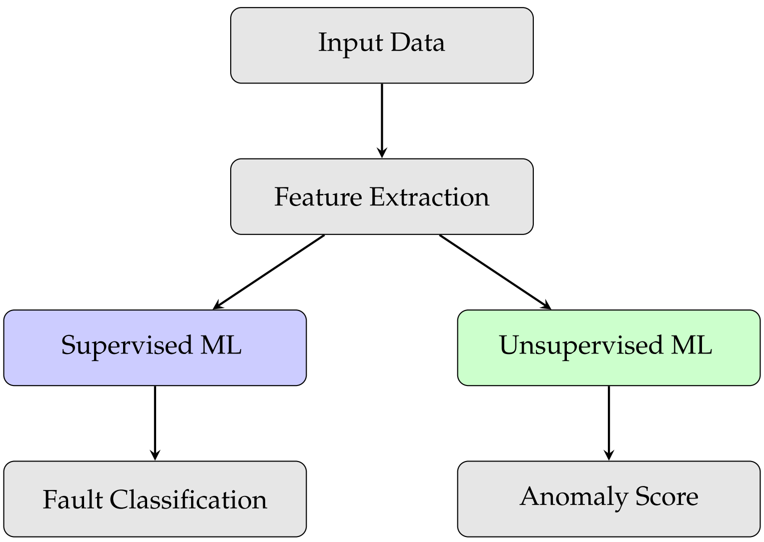

Figure 2. Schematic of machine learning approaches for industrial machines. Supervised models classify faults, while unsupervised models detect anomalies without labeled data.

Before discussing the individual learning paradigms, it is important to consider the practical data characteristics and constraints that influence model selection and performance in industrial environments.

4.1. Data Requirements and Practical Considerations

The performance of fault and anomaly detection methods in industrial environments is strongly influenced by the characteristics and quality of the available data. In practice, higher data volume and diversity generally lead to more reliable models, particularly for data-hungry machine learning approaches such as deep neural networks and ensemble methods. For complex machinery operating under variable load and environmental conditions, large and representative datasets are often required to capture normal variability and avoid false alarms [40,41].

Sensor type and sampling frequency play a critical role in determining diagnostic capability. High-frequency vibration and acoustic sensors are typically required to detect early-stage mechanical faults, such as bearing wear or gear damage; in contrast, lower-frequency measurements such as temperature, pressure, or electrical current are often sufficient for monitoring gradual degradation or process level anomalies. Inadequate sampling rates or low sensor resolution can obscure fault signatures and limit model effectiveness, regardless of the algorithm used [33–35].

Table 3 summarizes typical sensor configurations, sampling frequencies, frequency ranges, and data processing considerations commonly encountered in industrial fault and anomaly detection.

Table 3. Typical sensor configurations, data characteristics, and processing requirements for industrial fault and anomaly detection.

Signal Type	**Typical Sensors**	**Sampling Frequency**	**Key Frequency Range of Interest**	**Typical Faults Detected**	**Common Preprocessing and Feature Extraction**	**Data Volume and Storage Notes**
Vibration	Accelerometers (1–3 axis)	5–50 kHz	10 Hz–20 kHz	Bearing defects, gear wear, imbalance, misalignment	Band-pass filtering, FFT, envelope analysis, wavelet transform, RMS, kurtosis	High data rate (GB/day per machine); often processed at the edge and stored as features
Acoustic/AE	Microphones, acoustic emission sensors	20–100 kHz	5 kHz–50 kHz	Early bearing faults, lubrication issues, crack initiation	Filtering, time–frequency analysis, spectral features	Very high data rate; raw data stored selectively or event-triggered
Motor current	Current transformers, Hall-effect sensors	1–10 kHz	0–2 kHz	Rotor bar cracks, stator faults, load anomalies	Motor current signature analysis (MCSA), FFT, harmonic analysis	Moderate data size; often synchronized with vibration data
Temperature	Thermocouples, RTDs, infrared sensors	0.1–1 Hz	Very low frequency	Overheating, friction, insulation degradation	Smoothing, trend analysis, thresholding	Low data volume; long-term storage feasible
Pressure/Flow	Pressure transducers, flow meters	1–10 Hz	Low frequency	Leakage, blockage, cavitation (pumps)	Statistical features, trend deviation detection	Low–moderate data volume
Speed/Position	Encoders, tachometers	10–1000 Hz	Machine-dependent	Slip, misalignment, control faults	Resampling, synchronization, derivative features	Often fused with vibration and current data

The values reported summarize typical ranges commonly used in industrial practice and the literature; actual sensor configurations and data requirements depend on machine type, fault mechanisms, sensor specifications, and operational constraints (e.g., [19,22,26,33]).

Table 3 provides representative sensor configurations and data characteristics commonly encountered in industrial fault and anomaly detection systems. High-frequency signals such as vibration and acoustic emissions are essential for detecting early-stage mechanical faults but generate large data volumes, motivating edge-based preprocessing and feature extraction. In contrast, low-frequency process variables such as temperature, pressure, and flow support long-term condition monitoring and trend-based anomaly detection with minimal storage requirements.

Data quality is another key consideration. Industrial datasets frequently contain noise, missing values, sensor drift, and synchronization issues across multi-sensor systems, while traditional signal processing techniques can mitigate some of these effects, machine learning models particularly supervised approaches remain sensitive to biased or poorly labeled data. Larger datasets generally improve robustness, especially for ensemble models and deep learning architectures, but they also increase computational and storage requirements [9,42].

Label availability further constrains method selection. Supervised fault detection typically requires substantial labeled data for each fault type, which is often impractical in real industrial settings where failures are rare or costly to reproduce. As a result, unsupervised and hybrid approaches are commonly preferred during early deployment stages, with supervised models introduced incrementally as labeled fault data accumulate over time [19,25]. These practical data considerations largely determine which fault and anomaly detection strategies are feasible and effective in operational predictive maintenance systems.

Recent research has also advanced data-driven condition monitoring through approaches that reduce reliance on extensive labeled fault datasets. Blind diagnostic indicators aim to extract health relevant features directly from raw sensor signals without predefined fault templates, enabling robust monitoring under limited prior knowledge and variable operating conditions. In parallel, contrastive and self-supervised representation learning methods have gained traction for defect diagnosis by leveraging large volumes of unlabeled data to learn discriminative representations via augmentations, temporal consistency, or similarity based objectives, thereby improving robustness to label scarcity and domain shifts. For prognostics, remaining useful life (RUL) prediction is increasingly supported by spatial temporal multi sensor information fusion combined with prior knowledge embedding (e.g., physical constraints or degradation priors), which enhances prediction stability and generalization across operating regimes [43]. Collectively, these directions highlight a shift toward data efficient and knowledge aware learning strategies that complement the supervised, unsupervised, and hybrid paradigms reviewed in the following subsections.

4.2. Supervised Learning for Fault Detection

Supervised fault detection relies on labeled datasets in which operating conditions are explicitly annotated as normal or faulty. Data preprocessing remains a critical step, particularly for addressing missing sensor readings, noise, and synchronization issues in industrial time series data. Techniques such as imputation, normalization, and dimensionality reduction introduced earlier in this review are commonly applied to ensure robust model training [42].

Another key challenge in supervised learning is class imbalance, as normal operating data typically dominate fault samples. Without appropriate mitigation, models may fail to detect rare but critical failure events. The Synthetic Minority Over Sampling Technique (SMOTE) is widely used to generate synthetic fault samples, while class-weighted loss functions and ensemble resampling strategies further improve sensitivity to minority classes [44].

A wide range of supervised models have been applied to industrial fault detection. Logistic regression provides a simple, interpretable baseline, while Decision Trees and Random Forests offer greater robustness and accuracy. This is particularly beneficial in SCADA and hardware-in-the-loop (HIL) environments, where data are often nonlinear, noisy, and heterogeneous, and faults tend to manifest through threshold based patterns that tree based methods capture effectively [45]. KNN remains useful for small datasets but is computationally expensive at scale. Neural networks (ANNs) capture nonlinear relationships, and recurrent architectures such as LSTMs and GRUs are widely used for sequential data in degradation modeling and remaining useful life (RUL) estimation [46]. Convolutional neural networks (CNNs) are effective for raw signals and image-encoded time series, with architectures like WDCNN showing robustness to noise and strong results on benchmark datasets such as the Case Western Reserve University (CWRU)-bearing database [47]. Hybrid CNN–LSTM networks further combine spatial feature extraction with temporal modeling [48]. More recently, transformer-based models such as the Temporal Fusion Transformer have gained traction, capturing long-range dependencies while offering interpretability for industrial applications [46].

Figure 3 shows the workflow of supervised learning for industrial fault detection, from data acquisition to model deployment and monitoring. The pipeline covers data preparation, train/validation/test splitting, model training and evaluation, and long-term maintenance in operation.

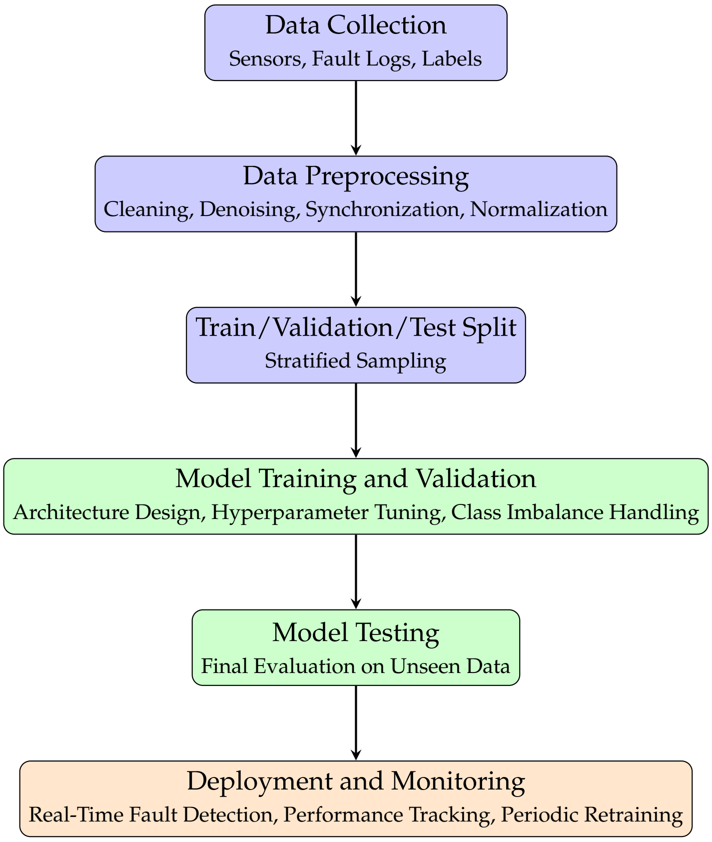

Figure 3. Workflow of supervised learning for industrial fault detection, from data preparation through model training, testing, deployment, and monitoring.

4.3. Unsupervised Learning for Anomaly Detection

In industrial systems where labeled fault data are scarce or entirely absent, unsupervised learning methods play a critical role in anomaly detection by modeling normal system behavior and identifying deviations as potential anomalies. Among the most widely adopted approaches is the Isolation Forest, which isolates anomalies through random partitioning of the feature space. Because anomalous observations tend to be separated in fewer splits, Isolation Forest provides an efficient and scalable solution that performs well on high-dimensional sensor datasets commonly encountered in industrial environments [49].

Clustering based methods such as K-Means and DBSCAN are also widely used for anomaly detection. In these approaches, normal operating conditions are expected to form dense clusters, while observations that lie far from cluster centroids or fail to belong to any cluster are flagged as anomalous. These techniques are conceptually simple, computationally efficient, and offer a degree of interpretability, making them effective for detecting unusual operational states without requiring labeled fault data [50].

Neural network based autoencoders further extend unsupervised anomaly detection capabilities by learning compact representations of normal system behavior. Trained to reconstruct healthy operating data, autoencoders produce higher reconstruction errors when presented with anomalous or degraded signals. Variants such as standard autoencoders and variational autoencoders have been widely applied to large-scale industrial sensor arrays, demonstrating strong performance in detecting subtle and previously unseen anomalies [51].

Another established unsupervised technique is principal component analysis (PCA), which assumes that normal operating data reside within a low-dimensional subspace. Anomalous observations project poorly into this space and therefore yield large residuals that can be used as anomaly indicators. PCA remains attractive for industrial applications due to its simplicity, computational efficiency, and high interpretability, particularly for correlated multivariate sensor measurements [52].

4.4. Hybrid and Integrated Approaches

Hybrid and integrated approaches combine different paradigms—such as machine learning, fuzzy inference, signal processing, and physics-based models—to build more robust and interpretable fault detection systems. They are especially effective when labeled data are scarce, operating conditions vary, or real-time deployment is required.

Recent studies highlight the effectiveness of hybrid approaches for industrial fault detection. A neuro-fuzzy framework for bearing fault prediction achieved 99.4% accuracy, outperforming a standalone neural network at 94% [53]. Another study proposed a hybrid edge–cloud architecture for bearing monitoring, combining unsupervised detection at the edge with supervised learning in the cloud and achieving an AUC of 0.96 [54]. For Cross-Linked Polyethylene (XLPE) cables, a semi-supervised ensemble method leveraging both labeled and unlabeled data reached 98% accuracy [55]. Additionally, hybrid techniques that integrate feature engineering with fuzzy systems have shown strong performance in bearing fault diagnosis [56]. Similar gains have been reported for remaining useful life (RUL) prediction using ANFIS models enhanced with wavelet-based features [57].

These studies confirm that hybrid approaches generally outperform single-method models by exploiting complementary strengths, although challenges remain in fusion strategies, computational complexity, and maintaining interpretability.

4.5. Implementation and Deployment Considerations

Beyond algorithmic performance, the practical value of fault and anomaly detection methods depends on their feasibility in real industrial deployment. Implementation constraints such as computational resources, latency requirements, and system integration strongly influence model selection. In many industrial settings, fault detection must operate in near real-time to enable timely intervention, which limits the applicability of computationally intensive models unless sufficient edge or cloud infrastructure is available [58,59]. Edge–cloud hybrid architectures are increasingly adopted to balance these constraints. Lightweight models or anomaly detection components are often deployed at the edge to enable low-latency monitoring and immediate alerts, while more complex machine learning models are executed in the cloud for model retraining, performance analysis, and long-term optimization. This distribution allows scalable deployment across fleets of machines while maintaining responsiveness at the equipment level [60,61].

Integration with existing industrial systems is another critical consideration. Effective predictive maintenance solutions must interface with supervisory control and data acquisition (SCADA) systems, industrial Internet of Things (IIoT) platforms, and computerized maintenance management systems (CMMS) to support automated alerts, maintenance scheduling, and decision making. Models that are difficult to interpret or maintain can hinder adoption, even when detection accuracy is high [9,62]. From an operational perspective, machine learning models require continuous monitoring and periodic retraining to remain effective under changing operating conditions, sensor drift, and equipment aging. Studies consistently report that predictive maintenance systems deliver tangible benefits such as reduced unplanned downtime, improved fault detection reliability, and optimized maintenance intervals when these deployment and lifecycle management challenges are addressed alongside model development. Consequently, successful industrial adoption depends not only on algorithm selection, but on the holistic design of data pipelines, deployment strategies, and maintenance workflows.

While Section 4 has outlined the key machine learning paradigms used for fault and anomaly detection, understanding their practical relevance requires examining how these models perform in real industrial environments. Machine learning techniques are increasingly embedded into manufacturing systems, food and beverage operations, energy infrastructures, and transportation networks, each presenting unique challenges and opportunities for deployment. Section 5 therefore explores the application of these approaches across different industrial sectors, highlighting practical benefits, domain-specific considerations, and emerging trends in real-world predictive maintenance systems.

5. Applications of Machine Learning in Predictive Maintenance

5.1. Manufacturing and Production

In manufacturing environments, machine learning plays a vital role in preventing unplanned downtime by enabling early detection of degradation in machines, production robots, and rotating equipment [41]. Multi-sensor data fusion combining vibration, temperature, acoustic, current, and speed signals significantly enhances diagnostic accuracy, with studies demonstrating that expanding beyond vibration only measurements raises fault classification performance from 84% to over 99% when using models such as kNN, SVM, Random Forest, and Decision Trees [63]. Manufacturing systems also benefit from ML-driven real-time thresholding, where Decision Tree models automatically derive dynamic operational limits for parameters like motor current and vibration, enabling IIoT platforms to trigger immediate alerts when machines deviate from healthy conditions. Deep learning models, particularly LSTM and GRU networks, are widely used for remaining useful life (RUL) prediction of bearings, spindles, compressors, and cutting tools, outperforming classical ML by capturing long-term degradation patterns in industrial time series data. Ensemble algorithms such as Random Forest and XGBoost further support predictive maintenance by effectively handling noisy, high dimensional sensor datasets and delivering robust fault prediction for turbines, gearboxes, pumps, and conveyor systems. Modern factories are increasingly integrating ML models into digital twin platforms, enabling virtual simulation of equipment health, early prediction of degradation, and optimization of maintenance schedules in steel mills and semiconductor manufacturing environments [64].

Roller chain systems, widely used in manufacturing conveyors, packaging lines, and material handling equipment, have also benefited from machine-learning-based condition monitoring. Typical failure modes such as chain elongation, sprocket wear, misalignment, and lubrication degradation generate characteristic vibration and acoustic signatures that can be captured using low-cost sensors. Recent studies demonstrate that data-driven approaches, including feature-based machine learning and deep learning models applied to vibration and motor current signals, enable early detection of chain degradation and support remaining useful life estimation [65,66]. These developments highlight the applicability of machine learning techniques to discrete mechanical transmission systems beyond traditional bearings and gearboxes.

To address the challenge of scarce failure samples, techniques such as SMOTE and cost-sensitive tree models are employed to improve detection of rare but critical faults in motors, hydraulic presses, and robotic systems [64]. Together, these machine learning advancements support highly reliable, data-driven predictive maintenance strategies that enhance equipment availability and operational efficiency across manufacturing and production settings.

5.2. Food and Beverage

Machine learning is becoming a central technology in food and beverage manufacturing, enabling real-time monitoring, quality control, and process optimization across production environments. A key driver of this adoption is the increasing availability of intelligent sensors that generate large volumes of data suitable for machine learning analysis. Research shows that food and drink manufacturers benefit significantly from combining online sensors with machine learning models to improve the efficiency and reliability of core processes. By integrating sensing technologies such as near infrared spectroscopy, ultrasonic and microwave sensors, hyperspectral imaging, and optical systems, machine learning algorithms can predict key quality parameters, detect deviations, and identify early signs of process faults [67]. In several industrial case studies including mixing, cleaning, and fermentation machine learning models achieved accuracies between 95 and 100%, demonstrating their effectiveness for classification, regression, and anomaly detection tasks in high throughput food operations [67].

These predictive capabilities allow manufacturers to reduce waste, prevent resource losses, and ensure consistent product quality. Machine learning also enhances operational decision making in food and beverage manufacturing. Machine-learning-driven demand forecasting significantly improves operational optimization by aligning production schedules with real market needs, reducing unnecessary machine use and avoiding overproduction. Similarly, ML-based inventory optimization models were shown to reduce waste by ensuring that raw materials and finished products are managed more efficiently [68]. These applications indirectly contribute to equipment longevity by minimizing erratic production cycles that can increase mechanical stress.

Beyond process and operational optimization, AI-based automation also supports quality control and inspection tasks. Machine learning methods including image processing, pattern recognition, and deep learning improve sorting, grading, and detection of product defects. These systems replace error prone manual inspection and ensure more consistent handling of products and equipment, which helps reduce operational disruptions and maintain smoother production flow [69]. Together, these studies demonstrate that machine learning provides a versatile set of tools that enhance monitoring, prediction, and decision making across food and beverage manufacturing. Whether applied through intelligent sensor networks, data-driven forecasting models, or automated inspection systems, machine learning contributes to higher efficiency, reduced waste, and more sustainable production in the sector.

5.3. Oil and Gas Industry

Machine learning (ML) has moved from experimental to operational in many parts of the oil and gas value chain, especially in subsurface characterization, drilling, and production optimization. In exploration and reservoir description, supervised and deep learning models now routinely classify lithology and facies from seismic and well logs, reducing turnaround time for interpretation and enabling more consistent mapping of complex reservoirs. Recent work, for example, integrates attribute selection workflows with classical classifiers to identify the most informative seismic attributes for facies classification in Malaysian offshore basins, improving both accuracy and computational efficiency [70]. Other studies use deep neural networks to predict lithology from well logs in unconventional and geologically complex reservoirs, explicitly encoding geological context or multi scale information to improve generalization beyond the training wells [71]. Together, these approaches show that ML can systematically exploit large integrated datasets (seismic, logs, cores) to refine facies models and reduce interpretation bias.

On the operations side, ML is increasingly embedded in real-time decision support for drilling and field management. Reinforcement-learning-based agents have been proposed to continuously update drilling parameters such as weight on bit and rotary speed, formulating drilling optimization as a multi objective control problem that balances rate of penetration, vibration control, and tool integrity; field-scale frameworks now couple symbolic regression, time series models, and Markov decision processes to achieve real-time optimization from streaming data. In reservoir and production engineering, physics-informed ML workflows combine reduced order flow models with neural networks to perform rapid history matching and long-term production forecasting in fractured unconventional reservoirs, offering orders of magnitude speed ups over high-fidelity simulation while preserving essential physics [72]. More recently, similar ideas have been extended to field development planning and process control, where deep reinforcement learning agents optimize well placement, drilling sequence, and plant operating conditions subject to economic and operational constraints [73]. Overall, these primary studies demonstrate that ML in oil and gas has progressed from proof-of-concept models toward workflows that are tightly integrated with physics, domain knowledge, and real-time data streams.

5.4. Transportation and Logistics

Machine learning has become an essential tool for advancing efficiency, sustainability, and decision making within transport and logistics systems. Recent studies demonstrate that ML techniques provide powerful capabilities for monitoring and optimizing transportation operations. For example, models such as Independent Component Analysis (ICA) and Gradient Descent K-Nearest Neighbors (GD-KNN) have been used to analyze real-time transport data, predict short term demand, assess accessibility, and evaluate system efficiency across parameters such as speed, cost, and fuel economy [74]. Beyond operational monitoring, machine learning also contributes to sustainability assessment in road freight transport. Supervised learning algorithms, including Support Vector Machines (SVMs), discriminant analysis, KNN, and Decision Trees, enable the classification of companies into sustainability performance levels using environmental, economic, and social indicators; this means they can be used to help organizations identify improvement areas and develop greener logistics strategies with high predictive accuracy, particularly through optimized SVM models [75]. In addition, ML plays a critical role in freight transport demand forecasting, where Artificial Neural Networks (ANNs), NARX, and NAR models have been shown to outperform traditional statistical approaches like ARIMA by capturing nonlinear patterns in historical Transportation Management System (TMS) data. These models support more accurate vehicle allocation, routing decisions, and resource planning, ultimately enhancing logistics performance and reducing uncertainty in dynamic supply chain environments [76]. Collectively, these studies highlight the broad and impactful applications of machine learning in creating more intelligent, efficient, and sustainable transport and logistics systems.

6. Evaluation and Emerging Trends

Traditional signal processing and modern machine learning approaches should not be viewed as competing paradigms, but as complementary components within predictive maintenance systems. Signal processing methods offer strong interpretability and physical grounding, while machine learning enables scalability, adaptability, and improved performance under complex and nonlinear operating conditions. Hybrid approaches increasingly bridge this gap by embedding domain knowledge into data-driven models, thereby combining transparency with predictive accuracy.

While the preceding sections have examined conventional approaches, machine learning methods, and sector specific applications, a broader evaluation helps clarify how these techniques complement one another and where the field is heading. Traditional signal processing methods remain valuable due to their interpretability and strong link to physical principles, yet they rely heavily on expert driven feature engineering and are difficult to generalize across diverse machines and environments. Machine learning methods mitigate these limitations by automatically extracting features and modeling complex nonlinear relationships, although their effectiveness is highly dependent on data availability, quality, and representativeness.

Supervised learning approaches typically achieve the highest accuracy when labeled fault data are abundant, yet they remain sensitive to class imbalance and can struggle to generalize to unseen equipment or operating regimes. Unsupervised learning approaches, including clustering, PCA-based methods, and autoencoders, are more suitable for detecting rare or novel faults but may be affected by noise, sensor drift, and changes in baseline machine behavior. Hybrid approaches integrating supervised and unsupervised learning, physics-based models, or signal processing features consistently outperform individual paradigms by combining interpretability, robustness, and adaptability. Table 4 provides a qualitative comparison of traditional signal processing, supervised learning, unsupervised learning, and hybrid approaches with respect to key industrial criteria. The comparison highlights trade-offs between interpretability, data and label requirements, and practical maturity, while signal-processing methods remain highly interpretable and well established in industrial practice, machine learning approaches offer greater adaptability at the cost of increased data dependence and deployment complexity. Hybrid approaches balance these characteristics by combining domain knowledge with data-driven learning, enabling improved robustness and generalization under varying operating conditions.

Table 4. Qualitative comparison of fault and anomaly detection paradigms across industrial criteria.

Approach	Interp.	Label Need	Data Demand	Maturity
Signal processing	High	Low	Low–Med.	High
Supervised ML	Med.	High	Med.–High	Med.
Unsupervised ML	Med.	Low	Med.	Med.
Hybrid approaches	High	Med.	Med.–High	Med.

Several emerging trends are reshaping the future of predictive maintenance. Transformer-based architectures provide superior modeling of long-range temporal dependencies and offer improved interpretability through attention mechanisms. Explainable AI (XAI) has become essential for enhancing transparency and operator trust, particularly in safety-critical environments. Physics-informed and hybrid models that blend engineering knowledge with data-driven learning help address data scarcity and improve generalization. Additionally, edge–cloud hybrid deployment frameworks support real-time detection at the edge while enabling scalable model retraining in the cloud.

https://doi.org/10.3390/a19020108

While these approaches demonstrate significant potential, many remain at a research or pilot-deployment stage. Challenges related to data availability, validation across operating regimes, certification, and integration with safety-critical industrial systems continue to limit widespread adoption.

Despite the comprehensive scope of this review, several limitations should be acknowledged. First, although a wide range of industrial sectors and machine learning paradigms are covered, the analysis is primarily qualitative and does not include a standardized quantitative benchmarking of methods, which is challenging due to heterogeneous datasets, evaluation metrics, and experimental settings reported in the literature. Second, while efforts were made to include recent advances through 2025, the rapid pace of development in machine learning means that emerging architectures and techniques may not yet be fully represented. Third, the review relies on reported results from existing studies, many of which are based on laboratory datasets or pilot-scale deployments, potentially limiting direct generalization to large-scale industrial environments. These limitations highlight the need for future work focusing on standardized evaluation frameworks, large-scale industrial validation, and longitudinal performance assessment.

A key contribution of this review, compared with existing surveys on predictive maintenance and anomaly detection, is its integrated cross sector perspective and its synthesis of machine learning paradigms within a unified evaluation framework. Whereas most prior reviews focus on a single industry, a single algorithm family, or a limited set of benchmark datasets, this work provides a comprehensive analysis covering supervised, unsupervised, and hybrid approaches across manufacturing, food and beverage, oil and gas, and transportation. Furthermore, by extending the review period through 2025 and incorporating emerging concepts such as transformers, hybrid physics-informed models, and Explainable AI, this review offers an updated and consolidated resource that captures recent advances not yet reflected in earlier surveys.

7. Conclusions

This systematic review has examined the evolution of fault and anomaly detection from conventional signal-based techniques to modern machine learning approaches, highlighting the growing maturity and practical relevance of data-driven predictive maintenance. By analyzing supervised, unsupervised, and hybrid models alongside diverse industrial applications, this review provides an integrated understanding of how machine learning is transforming maintenance practices across sectors such as manufacturing, food processing, energy, and transportation.

A key insight from this work is that the effectiveness of predictive maintenance depends not on a single model type but on the orchestration of the entire pipeline from data acquisition and preprocessing to model training, deployment, and continuous monitoring, while deep learning models, particularly CNNs, RNNs, autoencoders, and hybrid frameworks, have demonstrated strong performance in handling complex sensor data, persistent challenges remain. These include limited labeled fault datasets, class imbalance, generalization across different machines, and limited interpretability of deep models.

Unlike earlier surveys that typically concentrate on isolated algorithmic techniques or specific industrial domains, this review contributes a broader and more up-to-date synthesis that spans multiple sectors and integrates conventional, machine learning, and hybrid approaches within a unified evaluation framework. This broader scope enables practitioners and researchers to understand how different techniques compare, where they succeed or fail, and how emerging paradigms can be translated into practical industrial deployments.

The evaluation presented in this review highlights several emerging directions that are likely to shape the next generation of predictive maintenance systems. Transformer-based and attention driven architectures offer improved modeling of long-range temporal dependencies, while physics-informed and hybrid learning approaches help overcome issues related to data scarcity and domain shift. Explainable AI techniques are becoming essential for improving transparency, enabling engineers to interpret model outputs, and increasing trust in high-stakes industrial environments.

Overall, this review contributes a comprehensive and up-to-date synthesis of machine learning techniques for industrial fault and anomaly detection through 2025. Future progress will depend on integrating advanced machine learning architectures with domain knowledge, scalable edge-cloud deployment, and interpretability frameworks, ensuring that predictive maintenance solutions are not only accurate and robust but also trustworthy and practical for industrial adoption.

Author Contributions: Conceptualization, S.H.H.Z. and A.S.; methodology, S.H.H.Z.; writing—original draft, S.H.H.Z.; writing—review and editing, S.H.H.Z. and A.S.; reviewing, A.S., H.Z., and A.I.; resources, S.H.H.Z. and A.S.; supervision, A.S., H.Z., and A.I. All authors have read and agreed to the published version of the manuscript.

Funding: This study received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors gratefully acknowledge support from the Advanced Food Innovation Centre (AFIC) and Sheffield Hallam University. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version of this paper arising from this submission.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Muchiri, P.; Pintelon, L. Performance measurement using overall equipment effectiveness (OEE): Literature review and practical application discussion. *Int. J. Prod. Res.* **2008**, *46*, 3517–3535. [CrossRef]
2. Prajapati, A.; Bechtel, J.; Ganesan, S. Condition based maintenance: A survey. *J. Qual. Maint. Eng.* **2012**, *18*, 384–400. [CrossRef]
3. Vanson Bourne Global. Human Error Is More Common Cause of Unplanned Downtime in Manufacturing Than Any Other Sector, According to New Research. ServiceMax from GE Digital; Businesswire. 2017. Available online: https://www.businesswire.com/news/home/20171106006370/en/Human-Error-Common-Unplanned-Downtime-Manufacturing-Sector (accessed on 5 December 2025).
4. Deloitte. Predictive Maintenance: Optimize Maintenance with Industry 4.0 Technologies and Advanced Analytics. 2024. Available online: https://www.deloitte.com/content/dam/assets-zone3/us/en/docs/services/consulting/2024/us-smart-manufacturing-predictive-maintenance-infographic.pdf (accessed on 5 December 2025).
5. Siemens, A.G. Senseye Predictive Maintenance—The True Cost of Downtime 2022. 2023. Available online: https://assets.new.siemens.com/siemens/assets/api/uuid:3d606495-dbe0-43e4-80b1-d04e27ada920/dics-b10153-00-7600truecostofdowntime2022-144.pdf (accessed on 5 December 2025).
6. Benhanifia, A.; Cheikh, Z.B.; Oliveira, P.M.; Valente, A.; Lima, J. Systematic review of predictive maintenance practices in the manufacturing sector. *Intell. Syst. Appl.* **2025**, *26*, 200501. [CrossRef]
7. Cordes, E.E.; Jones, D.O.; Schlacher, T.A.; Amon, D.J.; Bernardino, A.F.; Brooke, S.; Carney, R.; DeLeo, D.M.; Dunlop, K.M.; Escobar-Briones, E.G.; et al. Environmental impacts of the deep-water oil and gas industry: A review to guide management strategies. *Front. Environ. Sci.* **2016**, *4*, 58. [CrossRef]
8. Liedberg, J. From Reactive Maintenance Towards Increased Proactiveness Through Digitalisation: Vital Factors to Achieve Momentum Towards Preventive Maintenance Using Digital Tools. Master's Thesis, Jönköping University, Jönköping, Sweden, 2021.
9. Tiddens, W.; Braaksma, J.; Tinga, T. Exploring predictive maintenance applications in industry. *J. Qual. Maint. Eng.* **2022**, *28*, 68–85. [CrossRef]
10. Shukla, K.; Nefti-Meziani, S.; Davis, S. A heuristic approach on predictive maintenance techniques: Limitations and scope. *Sensors* **2021**, *21*, 1184. [CrossRef]
11. Swanson, L. Linking maintenance strategies to performance. *Int. J. Prod. Econ.* **2001**, *70*, 237–244. [CrossRef]
12. Franceschini, L.; Midali, A. Industrial IoT: A Cost-Benefit Analysis of Predictive Maintenance Service. Master's Thesis, Politecnico di Milano, Milano, Italy, 2020.

13. Santos, M.R.; Costa, B.S.J.; Bezerra, C.G.; Andonovski, G.; Guedes, L.A. An evolving approach for fault diagnosis of dynamic systems. *Expert Syst. Appl.* **2022**, *189*, 115983. [CrossRef]
14. Ribeiro, R.P.; Mastelini, S.; Davari, N.; Veloso, B. From fault detection to anomaly explanation: A case study on predictive maintenance. *Web Semant. Sci. Serv. Agents World Wide Web* **2024**, *81*, 100821.
15. Li, J.; Milisavljevic-Syed, J.; Salonitis, K. Predictive Maintenance Servitisation Pathways. *IFAC PapersOnLine* **2024**, *58*, 329–334. [CrossRef]
16. Mallioris, P.; Aivazidou, E.; Bechtsis, D. Predictive maintenance in Industry 4.0: A systematic multi-sector mapping. *CIRP J. Manuf. Sci. Technol.* **2024**, *50*, 80–103. [CrossRef]
17. Arksey, H.; O'malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [CrossRef]
18. Levac, D.; Colquhoun, H.; O'brien, K.K. Scoping studies: Advancing the methodology. *Implement. Sci.* **2010**, *5*, 69. [CrossRef]
19. Zhang, W.; Yang, D.; Wang, H. Data-driven methods for predictive maintenance of industrial equipment: A survey. *IEEE Syst. J.* **2019**, *13*, 2213–2227. [CrossRef]
20. Pokharel, N.; Ghimire, A.; Thapa, B.; Thapa, B.S. Wear in centrifugal pumps with causes, effects and remedies: A Review. In *Proceedings of the IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2022; Volume 1037, p. 012042.
21. Manjare, A.A.; Patil, B.G. A review: Condition based techniques and predictive maintenance for motor. In Proceedings of the 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 25–27 March 2021; pp. 807–813.
22. Thomson, W.T.; Culbert, I. *Current Signature Analysis for Condition Monitoring of Cage Induction Motors: Industrial Application and Case Histories*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
23. Shanbhag, V.V.; Kandukuri, S.T.; Olimstad, G.; Schlanbusch, R. Predictive maintenance of critical components in hydroelectric turbines: A review. *IEEE Sens. J.* **2025**, *25*, 31959–31979. [CrossRef]
24. Zacksenhouse, M. Neural networks, diagnostic applications. In *Encyclopedia of Vibration*; Academic Press (Elsevier): San Diego, CA, USA, 2001.
25. Shaukat, K.; Alam, T.M.; Luo, S.; Shabbir, S.; Hameed, I.A.; Li, J.; Abbas, S.K.; Javed, U. A review of time-series anomaly detection techniques: A step to future perspectives. In *Proceedings of the Future of Information and Communication Conference*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 865–877.
26. Alonso-González, M.; Díaz, V.G.; Pérez, B.L.; G-Bustelo, B.C.P.; Anzola, J.P. Bearing fault diagnosis with envelope analysis and machine learning approaches using CWRU dataset. *IEEE Access* **2023**, *11*, 57796–57805. [CrossRef]
27. Garcia-Calva, T.; Morinigo-Sotelo, D.; Fernandez-Cavero, V.; Romero-Troncoso, R. Early detection of faults in induction motors—A review. *Energies* **2022**, *15*, 7855. [CrossRef]
28. Görnitz, N. *One-Class Classification in the Presence of Point, Collective, and Contextual Anomalies*; Technische Universitaet Berlin (Germany): Berlin, Germany, 2019.
29. Lee, X.Y.; Kumar, A.; Vidyaratne, L.; Rao, A.R.; Farahat, A.; Gupta, C. An ensemble of convolution-based methods for fault detection using vibration signals. In Proceedings of the 2023 IEEE International Conference on Prognostics and Health Management (ICPHM), Montreal, QC, Canada, 5–7 June 2023; pp. 172–179.
30. Moshrefi, A.; Nabki, F. Advanced industrial fault detection: A comparative analysis of ultrasonic signal processing and ensemble machine learning techniques. *Appl. Sci.* **2024**, *14*, 6397. [CrossRef]
31. Lai, S.; Cheung, T.H.; Zhao, J.; Xue, K.; Fung, K.C.; Lam, K.M. Residual Attention Single-Head Vision Transformer Network for Rolling Bearing Fault Diagnosis in Noisy Environments. In Proceedings of the 2024 6th International Conference on Video, Signal and Image Processing, Ningbo, China, 22–24 November 2024; pp. 136–150.
32. Bentrad, M.B.; Ghoggal, A.; Bahi, T.; Bahi, A. GNN-ASE: Graph-Based Anomaly Detection and Severity Estimation in Three-Phase Induction Machines. *arXiv* **2025**, arXiv:2508.00879.
33. Yang, H.; Mathew, J.; Ma, L. Vibration feature extraction techniques for fault diagnosis of rotating machinery: A literature survey. In Proceedings of the Asia-Pacific Vibration Conference, Gold Coast, Australia, 12–14 November 2003; pp. 801–807.
34. Helmi, H.; Forouzantabar, A. Rolling bearing fault detection of electric motor using time domain and frequency domain features extraction and ANFIS. *IET Electr. Power Appl.* **2019**, *13*, 662–669. [CrossRef]
35. Romanssini, M.; de Aguirre, P.C.C.; Compassi-Severo, L.; Girardi, A.G. A review on vibration monitoring techniques for predictive maintenance of rotating machinery. *Eng* **2023**, *4*, 1797–1817. [CrossRef]
36. Bao, W.; Tu, X.; Hu, Y.; Li, F. Envelope spectrum L-Kurtosis and its application for fault detection of rolling element bearings. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1993–2002. [CrossRef]
37. Li, T.; Sun, C.; Fink, O.; Yang, Y.; Chen, X.; Yan, R. Filter-informed spectral graph wavelet networks for multiscale feature extraction and intelligent fault diagnosis. *IEEE Trans. Cybern.* **2023**, *54*, 506–518. [CrossRef] [PubMed]
38. Rhif, M.; Ben Abbes, A.; Farah, I.R.; Martínez, B.; Sang, Y. Wavelet transform application for/in non-stationary time-series analysis: A review. *Appl. Sci.* **2019**, *9*, 1345. [CrossRef]

39. Shalby, E.M.; Abdelaziz, A.Y.; Ahmed, E.S.; Abd-Elhamed Rashad, B. A comprehensive guide to selecting suitable wavelet decomposition level and functions in discrete wavelet transform for fault detection in distribution networks. *Sci. Rep.* **2025**, *15*, 1160. [CrossRef]
40. Ma, L.; Juefei-Xu, F.; Zhang, F.; Sun, J.; Xue, M.; Li, B.; Chen, C.; Su, T.; Li, L.; Liu, Y.; et al. Deepgauge: Multi-granularity testing criteria for deep learning systems. In Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering, Montpellier, France, 3–7 September 2018; pp. 120–131.
41. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72*, 303–315. [CrossRef]
42. Gao, X.; Liu, Z.; Xu, L.; Ma, F.; Wu, C.; Zhang, K. Sensor Data Imputation for Industry Reactor Based on Temporal Decomposition. *Processes* **2025**, *13*, 1526. [CrossRef]
43. Ali, U.; Zia, A.; Ali, W.; Ramzan, U.; Rehman, A.; Chaudhry, M.T.; Xiang, W. Hypergraph Contrastive Sensor Fusion for Multimodal Fault Diagnosis in Induction Motors. *arXiv* **2025**, arXiv:2510.15547. [CrossRef]
44. Jang, J.; Kim, C.O. Unstructured borderline self-organizing map: Learning highly imbalanced, high-dimensional datasets for fault detection. *Expert Syst. Appl.* **2022**, *188*, 116028. [CrossRef]
45. Ovacıklı, A.K.; Yagcioglu, M.; Demircioglu, S.; Kocatekin, T.; Birtane, S. Supervised Learning-Based Fault Classification in Industrial Rotating Equipment Using Multi-Sensor Data. *Appl. Sci.* **2025**, *15*, 7580. [CrossRef]
46. Saeed, A.; A. Khan, M.; Akram, U.; J. Obidallah, W.; Jawed, S.; Ahmad, A. Deep learning based approaches for intelligent industrial machinery health management and fault diagnosis in resource-constrained environments. *Sci. Rep.* **2025**, *15*, 1114. [CrossRef]
47. Alghtus, K.; Gannan, A.; Alhajri, K.M.; Al Jubouri, A.L.; Al-Janahi, H.A. AI-Powered Machine Learning Approaches for Fault Diagnosis in Industrial Pumps. *arXiv* **2025**, arXiv:2508.15550. [CrossRef]
48. Kai, T.C.Y.; Saptoro, A.; Putra, Z.A.; Lim, K.H.; Yeo, W.S.; Sunarso, J. Supervised deep learning algorithms for process fault detection and diagnosis under different temporal subsequence length of process data. *Appl. Intell.* **2025**, *55*, 883. [CrossRef]
49. Liu, F.T.; Ting, K.M.; Zhou, Z.H. Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 413–422.
50. Rashid, U.; Saleem, M.F.; Rasool, S.; Abdullah, A.; Mustafa, H.; Iqbal, A. Anomaly detection using clustering (K-means with DBSCAN) and SMO. *J. Comput. Biomed. Inform.* **2024**, *7*, 1–10.
51. Aslam, M.M.; Tufail, A.; De Silva, L.C.; Haji Mohd Apong, R.A.A.; Namoun, A. An improved autoencoder-based approach for anomaly detection in industrial control systems. *Syst. Sci. Control Eng.* **2024**, *12*, 2334303. [CrossRef]
52. Liang, A.; Hu, Y.; Li, G. The impact of improved PCA method based on anomaly detection on chiller sensor fault detection. *Int. J. Refrig.* **2023**, *155*, 184–194. [CrossRef]
53. Kudelina, K.; Raja, H.A. Neuro-Fuzzy Framework for Fault Prediction in Electrical Machines via Vibration Analysis. *Energies* **2024**, *17*, 2818. [CrossRef]
54. Leroux, S.; Simoens, P. Hybrid Edge–Cloud Models for Bearing Failure Detection in a Fleet of Machines. *Electronics* **2024**, *13*, 5034. [CrossRef]
55. De Xu, N. Semi-supervised hybrid ensemble learning for fault detection in 20 kV XLPE cables. *Informatica* **2025**, *49*, 1–12.
56. Zhou, J.; Yang, X.; Liu, L.; Wang, Y.; Wang, J.; Hou, G. Fuzzy broad learning system combined with feature-engineering-based fault diagnosis for bearings. *Machines* **2022**, *10*, 1229. [CrossRef]
57. Wahhab Lourari, A.; Benkedjouh, T.; El Yousfi, B.; Soualhi, A. An ANFIS-based framework for the prediction of bearing remaining useful life. *Int. J. Progn. Health Manag.* **2024**, *15*, 1–13.
58. Xu, X.; Mo, R.; Yin, X.; Khosravi, M.R.; Aghaei, F.; Chang, V.; Li, G. PDM: Privacy-aware deployment of machine-learning applications for industrial cyber–physical cloud systems. *IEEE Trans. Ind. Inform.* **2020**, *17*, 5819–5828.
59. Zonta, T.; Da Costa, C.A.; da Rosa Righi, R.; de Lima, M.J.; Da Trindade, E.S.; Li, G.P. Predictive maintenance in the Industry 4.0: A systematic literature review. *Comput. Ind. Eng.* **2020**, *150*, 106889. [CrossRef]
60. Wang, T.; Wang, P.; Cai, S.; Ma, Y.; Liu, A.; Xie, M. A unified trustworthy environment establishment based on edge computing in industrial IoT. *IEEE Trans. Ind. Inform.* **2019**, *16*, 6083–6091. [CrossRef]
61. Shi, W.; Cao, J.; Zhang, Q.; Li, Y.; Xu, L. Edge computing: Vision and challenges. *IEEE Internet Things J.* **2016**, *3*, 637–646. [CrossRef]
62. Carvalho, T.P.; Soares, F.A.; Vita, R.; Francisco, R.d.P.; Basto, J.P.; Alcalá, S.G. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [CrossRef]
63. Kullu, O.; Cinar, E. A deep-learning-based multi-modal sensor fusion approach for detection of equipment faults. *Machines* **2022**, *10*, 1105. [CrossRef]
64. Liu, H.; Cheng, F.; Li, Y. Deep Learning Prediction for Bearing Remaining Useful Life. In *Prognostics and Health Management for Intelligent Electromechanical Systems*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 73–90.

65. Wu, M.; Goh, K.W.; Chaw, K.H.; Koh, Y.S.; Dares, M.; Yeong, C.F.; Su, E.L.; William, H.; Zhang, Y. An intelligent predictive maintenance system based on random forest for addressing industrial conveyor belt challenges. *Front. Mech. Eng.* **2024**, *10*, 1383202. [CrossRef]
66. Radicioni, L.; Bono, F.M.; Cinquemani, S. On the use of vibrations and temperatures for the monitoring of plastic chain conveyor systems. *Mech. Syst. Signal Process.* **2025**, *223*, 111935.
67. Watson, N.J.; Bowler, A.L.; Rady, A.; Fisher, O.J.; Simeone, A.; Escrig, J.; Woolley, E.; Adedeji, A.A. Intelligent sensors for sustainable food and drink manufacturing. *Front. Sustain. Food Syst.* **2021**, *5*, 642786. [CrossRef]
68. Idoko, N.A.; Umoh, V.M.; Onah, A.O. Machine learning and cost reduction in food and beverage manufacturing firms in enugu state. *Int. J. Allied Res. Econ. Bus. Manag.* **2025**, *13*, 34–51.
69. Kumar, I.; Rawat, J.; Mohd, N.; Husain, S. Opportunities of artificial intelligence and machine learning in the food industry. *J. Food Qual.* **2021**, *2021*, 4535567. [CrossRef]
70. Babikir, I.; Abdul Latiff, A.H.; Elsaadany, M.; Pratama, H.; Sajid, M.; Mad Sahad, S.; Ishak, M.A.; Laudon, C. Enhancing machine learning-based seismic facies classification through attribute selection: Application to 3D seismic data from the Malay and Sabah Basins, offshore Malaysia. *Geomech. Geophys. Geo-Energy Geo-Resour.* **2024**, *10*, 138. [CrossRef]
71. Khan, S.Q.; Kirmani, F.U.D. Applicability of deep neural networks for lithofacies classification from conventional well logs: An integrated approach. *Pet. Res.* **2024**, *9*, 393–408. [CrossRef]
72. Srinivasan, S.; O'Malley, D.; Mudunuru, M.K.; Sweeney, M.R.; Hyman, J.D.; Karra, S.; Frash, L.; Carey, J.W.; Gross, M.R.; Guthrie, G.D.; et al. A machine learning framework for rapid forecasting and history matching in unconventional reservoirs. *Sci. Rep.* **2021**, *11*, 21730. [CrossRef]
73. Nasir, Y.; He, J.; Hu, C.; Tanaka, S.; Wang, K.; Wen, X. Deep reinforcement learning for constrained field development optimization in subsurface two-phase flow. *Front. Appl. Math. Stat.* **2021**, *7*, 689934. [CrossRef]
74. Yadav, M.S.; Shukla, R.; Parthasarathy, C.; Chikati, D.; Chandan, R.R.; Gupta, K.K.; Gupta, S.K. Original research article transportation logistics monitoring for transportation systems using the machine learning. *J. Auton. Intell.* **2024**, *7*, 1–13.
75. Castaneda, J.; Cardona, J.F.; Martins, L.d.C.; Juan, A.A. Supervised machine learning algorithms for measuring and promoting sustainable transportation and green logistics. *Transp. Res. Procedia* **2021**, *58*, 455–462. [CrossRef]
76. Salais-Fierro, T.E.; Martínez, J.A.S. Demand forecasting for freight transport applying machine learning into the logistic distribution. *Mob. Netw. Appl.* **2022**, *27*, 2172–2181. [CrossRef]