Pertanika J. Sci. & Technol. 33 (1): 149 - 177 (2025)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Review Article

Concept Drift Early Fault Detection in Wind Turbine Based on Distance Metric: A Systematic Literature Review

Dongqi Zhang[1,2], Zainura Idrus[1]* and Raseeda Hamzah[3]

[1]*College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia*
[2]*School of Big Data Science, Hebei Finance University, Baoding 071051, China*
[3]*College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Melaka Branch, 77300 Merlimau, Melaka, Malaysia*

ABSTRACT

The Supervisory Control and Data Acquisition (SCADA) system in wind turbines generates substantial data that remains underutilized in terms of wind farm operation and maintenance (O&M). Numerous fault detection methods leveraging SCADA data are being extensively researched to reduce O&M costs. The detection methods are revolutionizing wind farm O&M strategies, shifting from scheduled passive detection to predictive active detection, with the potential to significantly reduce spare parts and labor costs. This paper presents a systematic review of wind turbine fault detection methods based on concept drift and distance metrics, employing the PRISMA methodology. The selected literature is analyzed from three perspectives: fault components, modeling methods, and data sources. Additionally, this review addresses research questions related to current trends, concept drift applications, and distance metric utilization in wind turbine fault detection. Lastly, it provides valuable insights for researchers and industry practitioners in wind energy engineering to explore future research and development in fault detection techniques for enhancing the reliability and efficiency of wind turbine operations.

Keywords: Concept drift, distance metric, fault detection, wind turbines

ARTICLE INFO

Article history:
Received: 24 February 2024
Accepted: 03 September 2024
Published: 27 January 2025

DOI: https://doi.org/10.47836/pjst.33.1.07

E-mail addresses:
104465048@qq.com (Dongqi Zhang)
zainura@tmsk.uitm.edu.my (Zainura Idrus)
raseeda@uitm.edu.my (Raseeda Hamzah)
* Corresponding author

INTRODUCTION

In 2022, China's grid-connected wind power cumulative installed capacity surpassed 300 million kilowatts, with over 155,000 operational wind turbines. As these turbines age, a significant proportion will exceed the manufacturer's warranty period. Wind

ISSN: 0128-7680
e-ISSN: 2231-8526

turbines beyond the typical 5-year warranty often cannot receive maintenance from the original manufacturer, necessitating third-party companies' procurement of inspection and maintenance services. The five components with the highest fault rates in wind turbines are the pitch system, inverter, generator, control system, and electrical system. The main shaft, generator, gearbox, pitch, and hydraulic systems are responsible for the most extended downtimes. These downtimes render wind turbines inoperative and incur high maintenance costs, highlighting the urgent need for intelligent maintenance transformation in wind farms.

Wind farms are typically equipped with Supervisory Control and Data Acquisition (SCADA) systems, which monitor 30 to 150 parameters. This data is stored in the SCADA database in real-time. However, a substantial amount of valuable data remains underutilized due to the lack of effective analysis methods and tools for time series data in wind farms. Therefore, analyzing these data effectively for early fault detection in wind turbines is crucial. It is worth noting that while SCADA data encompasses both time-series data and status codes, this study focuses exclusively on the analysis of time-series data. Vibration data (Feng, Ji, Ni, et al., 2023; Ni et al., 2023), although crucial for monitoring wind turbine health, is not included in the scope of this research.

Normal data can be utilized to establish the basic performance of machines, where it serves as a benchmark or threshold, indicating the machine's normal operational parameters. The monitoring process typically involves comparing newly acquired data against this established threshold. The primary objective is to detect any abnormal state of the machine, the phenomenon often referred to as concept drift. Concept drift occurs when the machine's performance deviates significantly from the threshold, exceeding or falling below it. Such deviations are frequently indicative of system faults and warrant detection and investigation. In essence, wind turbine SCADA data is typical time series data in complex industrial systems. This data is characterized by its detailed, multi-dimensional nature, encompassing information from multiple devices and components. Through the application of concept drift and related algorithms, it is possible to monitor the operational state of wind turbine equipment. Consequently, this enables a shift in the assessment of wind turbine working conditions, transitioning from periodic passive inspections to condition-based active maintenance.

This study employs the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method to identify and analyze 65 papers pertinent to wind turbine fault detection. The analysis begins with examining the basic characteristics of these publications, including literature type, publication date, and the first author's country affiliation. Subsequently, a comprehensive investigation focuses on three key aspects: fault types, modeling methods, and data sources. Furthermore, the study addresses three critical questions: the state-of-the-art fault detection methods for wind turbines, the current status of concept drift application in wind turbine fault detection, and the advantages of distance-based concept drift approaches. The paper concludes with a synthesis of current research findings and an outlook on future research directions.

RESEARCH QUESTIONS

This review aims to concisely overview the current wind turbine fault detection research. Recent literature has examined fault detection from various perspectives, including non-destructive detection (Márquez & Chacón, 2020), tribology (Dhanola & Garg, 2020), machine learning (Fernandes et al., 2022), and condition monitoring (Badihi et al., 2022). However, a notable gap remains in synthesizing recent wind turbine fault detection methodologies advancements. This paper intends to address this gap and focus on three core research questions, summarized in Table 1.

Table 1
Description of research question

No.	Description	Why this is important
Research question 1	What is the current trend in advanced fault detection methods for wind turbines?	In recent years, algorithmic research has experienced exponential growth, yielding diverse methodologies, including classification, prediction, regression, and supervised and unsupervised learning approaches. The central focus of this review is to examine how these burgeoning algorithmic advancements can be effectively synthesized and applied to shape the emerging research trends in wind turbine fault detection.
Research question 2	What is the current research status of concept drift methods in wind turbine fault detection?	The fundamental principle of concept drift algorithms lies in their ability to detect changes in data distribution. Despite this potential, there is a notable absence of comprehensive literature reviews examining the application of concept drift algorithms in wind turbine fault detection. This paper will address this gap by concentrating on concept drift-based wind turbine fault detection developments over the past years.
Research question 3	What are the advantages of distance metrics for wind turbine fault detection and concept drift?	Distance metrics are essential in wind turbine fault detection and concept drift algorithms. These metrics find diverse applications across various aspects of the analytical process, including data similarity measurement, fundamental distance functions within models, and evaluation indices. This paper will specifically examine the application of distance metric algorithms in concept drift-based wind turbine fault detection research, analyzing their advantages and contributions to this field.

METHOD

PRISMA Methodology

This systematic literature review follows the PRISMA methodology. PRISMA is a widely recognized approach to improving systematic reviews and meta-analyses' transparency, completeness, and reliability. The method provides a structured framework for conducting and reporting literature reviews.

Search Strategy

This study employs a systematic approach to conduct a comprehensive literature review on fault detection in wind turbines, specifically concept drift algorithms. The literature search is based on three primary keywords: "wind turbine," "fault detection," and "concept drift." The search process is structured in three phases:

Initial Phase: Extensive search is conducted using the Web of Science database, provides broad, multidisciplinary coverage of scientific literature.

Secondary Phase: The search is expanded to include IEEE Xplore, a database specializing in information technology and engineering literature.

Final Phase: The literature search is investigated using ScienceDirect, which offers access to a wide range of scientific and technical publications to ensure comprehensive coverage.

Data Extraction and Analysis Plan

It is planned to extract the data and analysis methods of each literature, including the following aspects: research questions, research objectives, research methods, results, data sources, relevance to wind turbine fault detection, relevance to concept drift, and use of distance metrics. A keyword pairwise query method is employed to ensure a comprehensive literature analysis. This approach facilitates the examination of intersections between fault detection, concept drift, and distance metric applications in the context of wind turbine research.

Literature Screening

The literature search process involves a comprehensive keyword search across multiple databases, as illustrated in Figure 1. The search yielded the following results: 140 relevant publications were retrieved from the Web of Science, while IEEE Xplore and ScienceDirect yielded 1,383 and 2,404 relevant publications, respectively. After the initial search, a two-step refinement process was applied, identifying and removing 556 duplicate entries across the databases. Additionally, 1,894 publications published before 2018 were excluded to focus on recent research.

Upon further scrutiny of the 1,477 publications, it was revealed that 835 publications primarily focused on vibration-related topics and their various aspects, including vibration analysis, data, experiments, datasets, monitoring, signals, signal analysis, noise, and spectrum and frequency band identification. While these topics are generally relevant to wind turbine research, they do not align directly with the specific focus of this study. Consequently, these 835 publications were excluded from further analysis. After this screening process, 642 publications remained for potential inclusion in the review.

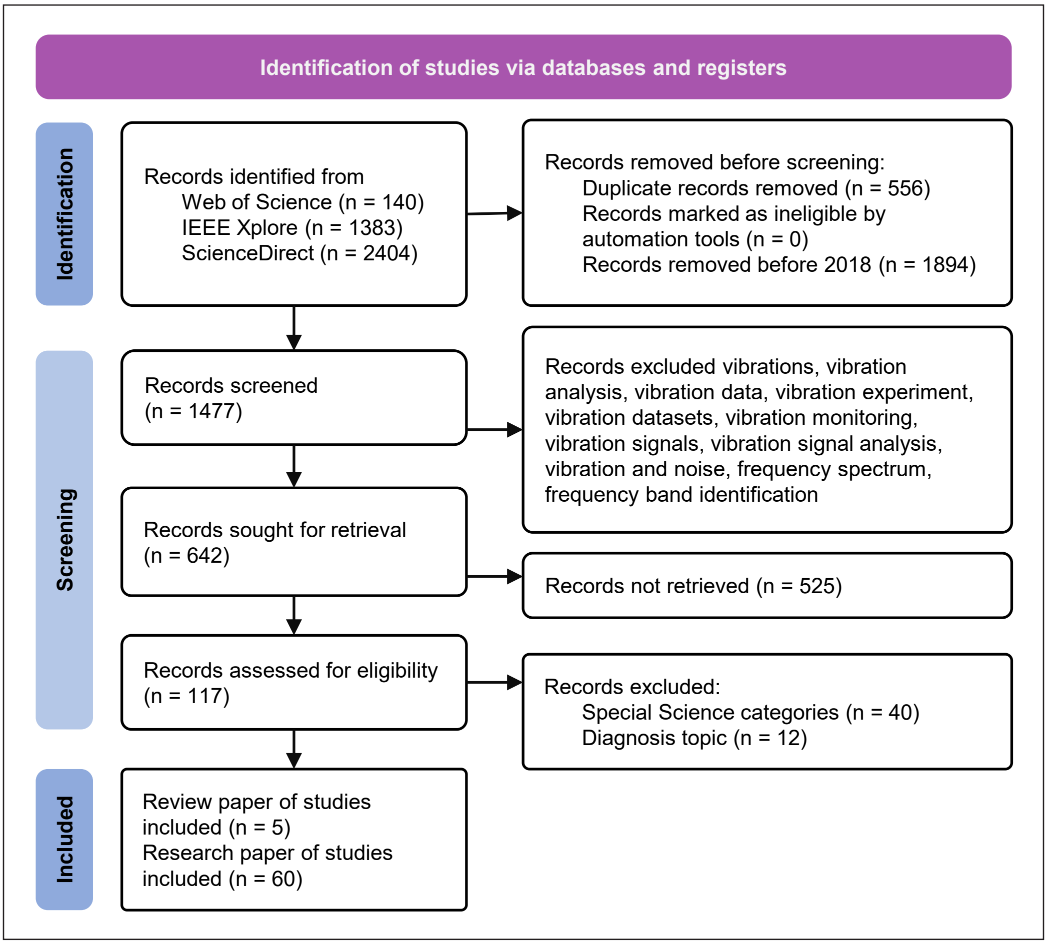

Figure 1. Identification of studies via databases and register

The final screening process of the 642 potentially relevant publications involves several steps. First, 525 publications were deemed inaccessible. After their removal, 117 publications remained. Subsequently, 40 publications categorized as specialized science were found irrelevant to this review, along with 12 publications focused on diagnostic topics. These 52 publications were excluded. Ultimately, 65 publications were deemed relevant for this review, comprising five review papers and 60 research papers.

Analysis of Basic Literature Information

After the screening process, a comprehensive analysis of the 65 selected publications related to wind turbine fault detection will be conducted. This analysis will focus on literature type, publication year, and the first author's country, providing a clear overview of the research status in the field.

Analysis of Literature Types

A total of 65 literature items were selected and divided into conference papers and journal papers. According to the statistics, there are 11 conference papers and 54 journal papers.

Analysis of Publication Dates

Figure 2 illustrates the annual distribution of conference and journal papers from 2018 to the present. This visualization reveals several important trends and insights. The red bar chart representing conference papers exhibits an irregular pattern across the years, with no discernible trend. This irregularity can be attributed to conference papers, which are more sensitive to short-term factors or specific events in the field. The right blue bars are the statistical results of journal papers. The statistical number of journal papers shows an increasing trend, indicating that the scientific research resources invested in this field are increasing yearly.

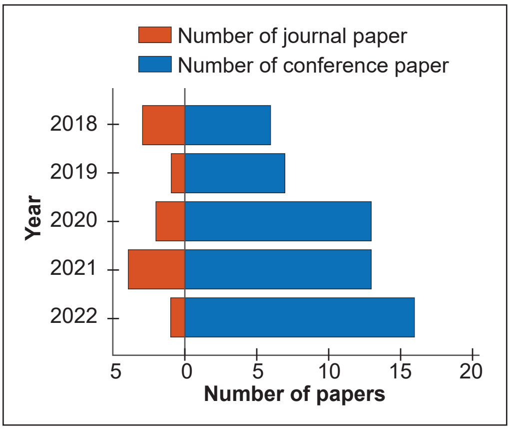

Figure 2. Statistics of different literature types

Analysis of the First Author Country

Figure 3 illustrates first-author countries' distribution for conference and journal papers. This visualization reveals several important trends and insights. China (CN) leads in both journal and conference papers. This aligns with China's recent aggressive push for new

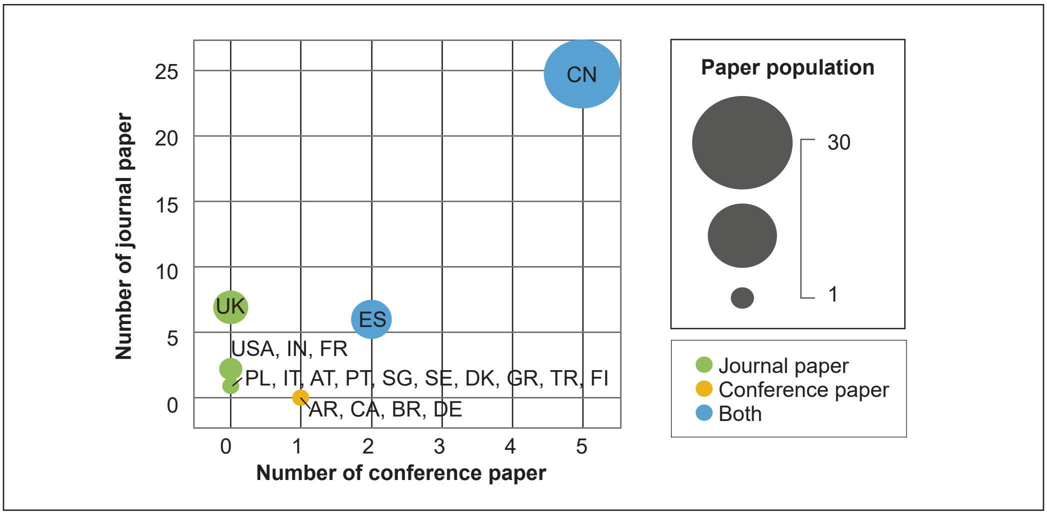

Figure 3. Statistical number for country analysis of the first author

energy policies and investments in renewable energy research. It also reflects China's growing wind energy sector and the need for advanced fault detection technologies. The United Kingdom (UK) and Spain (ES) show significant research activities. It can be attributed to their geographical conditions favorable for wind energy, long-standing traditions in wind power utilization and strong governmental support for renewable energy research. The United States (USA), India (IN), and France (FR) each contribute two journal articles. It indicates a broader global interest in wind turbine fault detection, albeit at a lower intensity compared to the leading countries.

RESULTS

Fault Analysis of Wind Turbine Components

Based on the analysis of the selected publications, it is evident that in post-2014, virtually all newly installed large wind turbines globally have adopted variable pitch, variable speed and constant frequency technology. These modern wind turbines can be broadly categorized into doubly fed asynchronous wind turbines and direct drive permanent magnet wind turbines. The primary distinction between these types is the presence or absence of a substantial transmission device—the gearbox. The research on wind turbine component fault detection primarily focuses on several key components: generators, transmission chains (including gearboxes and spindles), yaw systems, variable pitch systems, blades, and electronic components, by systematically classifying and screening the collected literature, resulting in the comprehensive overview presented in Table 2.

An analysis of Table 2 reveals significant research on faults related to generators, gearboxes, and blades within wind turbine systems. Further examination of these three critical components indicates a prevalent focus on utilizing SCADA attributes to detect and analyze abnormal temperature conditions in generators and gearboxes. This trend is evidenced by numerous studies (Jia et al., 2021; Liu et al., 2020; Qu et al., 2021; Velandia-Cardenas et al., 2021; Wang et al., 2022; Wang, Zhao et al., 2021; Xu et al., 2019). The following conclusions can be drawn from the literature in Table 2.

(a) Faults in wind turbine components such as generators and gearboxes have significant impacts, incurring high costs in terms of time, spare parts, and repairs for replacement and maintenance. Focusing on abnormal temperature faults in gearboxes and generators can effectively contribute to reducing operational expenses. This approach enables early detection of potential issues, allowing for timely interventions and preventive maintenance, thereby minimizing the risk of catastrophic faults and optimizing the overall maintenance strategy.

(b) Gearboxes and generators, as typical rotating equipment in wind turbines, predominantly exhibit gradual faults. The characteristic of these faults is a slow temperature drift in the rotating components, eventually leading to malfunctions.

Table 2
Components fault of wind turbine

Components	Fault type	Reference
Generator	Fault of bearing inner raceway, outer raceway, and rolling element for generator bearing	Tang et al., 2022; (Yang, Liu et al., 2022
	Generator bearing temperature Generator front bearing temperature overrun fault Generator rear bearing temperature overrun fault Generator damages the front and rear bearings et al.	Wang et al., 2022 Jia et al., 2021 Wang, Zhao et al., 2021
	Rotor winding aging	Zhang & Lang, 2020
Gearbox	Gearbox pump damaged Gearbox bearings damaged Gearbox noise	Latiffianti et al., 2022
	Abnormal gearbox temperature rise	Liu et al., 2020; Velandia-Cardenas et al., 2021; Qu et al., 2021; Xu et al., 2019
	Gearbox pitting, broken tooth	Du et al., 2022
	Gearbox oil pressure difference anomaly	Bo et al., 2019
	Gearbox lubricant pressure anomaly	Wang et al., 2017; Yang & Zhang, 2021b
	The low temperature of the gearbox oil Gearbox frequency converter no feedback Gearbox oil flow no feedback Gearbox bearing 1 PT100 error	Chacon et al., 2020
	Gearbox high-speed stage bearing fault.	McKinnon et al., 2020
Main bearing	Bearing over temperature warning	Herp et al., 2020; Wang, Zhao et al., 2021; Xiao et al., 2022
Pitch system	Encoder Failure Slip Ring Failure Electric Motor Failure	Wei et al., 2020
	Hydraulic hoses and oil replacement Hub oil leakage Block replacement at blade B Block leakage in blade B Replacement of blade valve Nitrogen accumulator (No 4) replacement of Blade Blade tracking error during stop/operation of Blade Replacement of hyd. Cylinder	Korkos et al., 2022
	High air content in oil Pump wear Hydraulic leakage	Tutiven et al., 2018
	The pitch gear fault of blade 1	Tao et al., 2019
Blade	Blade icing fault	Tong et al., 2022; Yi et al., 2021; Velasquez et al., 2021; Aziz et al., 2022; Aziz et al., 2021

Table 2 *(continue)*

Components	Fault type	Reference
	Blade breakages or blade rupture	Zhao et al., 2021; Yang & Zhang, 2021a Wang, Zhang et al., 2018
	Blade contamination	Velasquez et al., 2021
Yaw system	Yaw misalignment	Pandit & Infield, 2018; Aziz et al., 2021)
Electronic component	Short-circuits of generator	Sousa et al., 2018 Wang, Ma et al., 2018
Sensor	Sensor faults	Kavaz & Barutcu, 2018

This gradual temperature change presents potential features that can be modeled to predict faults in advance (Wang et al., 2017; Yang & Zhang, 2021b). Consequently, early fault detection studies focusing on abnormal temperature attributes are common in this field.

(c) Ice faults are the most extensively studied issues affecting wind turbine blades (Aziz et al., 2021, 2022; Tong et al., 2022; Velasquez et al., 2021; Yi et al., 2021). The primary reason for this focus is that ice accumulation directly reduces the active power output of wind turbines under identical wind speed conditions. This phenomenon significantly affects the power generation capacity of wind farms, resulting in decreased overall efficiency and productivity.

(d) The primary focus is on gradual deterioration in the literature concerning the main bearing faults. Researchers typically utilize the main bearing temperature signal as the main data source for analysis to assess the potential presence of faults. This approach enables the detection of subtle changes in bearing performance over time, facilitating the early identification of developing issues (Herp et al., 2020; Wang, Zhao et al., 2021; Xiao et al., 2022).

(e) Pitch system faults are relatively complex, encompassing multiple components and signal attributes. These faults are generally associated with hydraulic devices and motors within the pitch control mechanism.

(f) Yaw system fault studies primarily focus on yaw misalignment issues. The main objectives of these investigations are to optimize wind energy capture and reduce the stress impact on wind turbines.

Analysis of Modeling Types for Fault Detection

A comprehensive review of the literature on early fault detection in wind turbines has been conducted, utilizing the methodologies employed in research as a basis for classification. The literature search focused on four primary categories: machine learning models, deep learning models, statistical probability models, and other models. Table 3 presents a systematic compilation of the fault detection modeling methods utilized in the reviewed literature, offering a structured overview of the current state of research in this field.

The following conclusions can be drawn based on the analysis presented in Table 3.

(a) From the perspective of literature statistics, machine learning approaches are more prevalent than other types. The predominantly applied models include Support

Table 3
Analysis of modelling types for fault detection

Modeling type	Algorithm	Reference
Machine learning	SVM	Velandia-Cardenas et al., 2021; Qu et al., 2021; Mammadov et al., 2021
	Support vector regression	Díaz et al., 2020; Tao et al., 2019
	One-Class Support Vector Machine (OCSVM), Isolation Forest (IF), Elliptical Envelope (EE).	McKinnon et al., 2020
	XGboost, AdaBoost	Liu et al., 2020; Mammadov et al., 2021; Velandia-Cardenas et al., 2021; Zhang et al., 2018; Trizoglou et al., 2021
	Random forest regressive	Turnbull et al., 2022; Zenisek et al., 2019; Zhang et al., 2018
	Sparse isolation encoding forest	Du et al., 2022
	Sparse Bayesian Learning (SBL) algorithm	Bo et al., 2019
	Optimized relevance vector machine (RVM) regression	Wei et al., 2020
	Adaptive neuro-fuzzy inference system (ANFIS) technique	Korkos et al., 2022
	Gaussian Process (GP) models	Pandit & Infield, 2018; Pandit & Infield, 2019
	Semisupervised extreme learning machine (SS-ELM) algorithm	Tong et al., 2022
	Minority clustering Synthetic minority oversampling technique	Yi et al., 2021
	Higher Order Statistics-Bayes classifiers	Sousa et al., 2018
	Quantile regression neural networks	Xu et al., 2019
	Improved principal component analysis	Zhang et al., 2021; Wang, Ma et al., 2018; Pozo et al., 2018
Deep learning	Dual-stage attention-based recurrent neural network	Yang, Liu et al., 2022
	Cascade SAE & LightGBM	Wang et al., 2022
	Secondary decomposition, reinforcement learning and SRU network	Liu et al., 2021

Table 3 *(continue)*

Modeling type	Algorithm	Reference
	Stacked long-short-term memory with multi-layer perceptron (SLSTM-MLP)	Xiao et al., 2022
	RUL Recurrent Neural Network	Herp et al., 2020
	Conditional convolutional autoencoder	Yang & Zhang, 2021a
	LSTM-SAE, CNN-SAE	Fotiadou et al., 2020
	DAE, CNN, residual attention module (RAM)	Jia et al., 2021
	Deep Neural Networks	Wang et al., 2017
	Deep Autoencoder	Wang, Zhang et al., 2018
	Joint variational autoencoder (JVAE)	Yang & Zhang, 2021b
	Multi-Channel CNN	Mohammadi et al., 2020
Probabilistic statistical model	Optimized relevance vector machine	Wei et al., 2020
	Sparse heteroscedastic Gaussian Process regression	Rogers et al., 2020
	Discrete digital model	Tang et al., 2022
	Base pattern Probability Mass Function (PMF)	Peña et al., 2021
Other methods	Failure Modes, Effects, and Criticality Analysis	Catelani et al., 2020
	Ensemble Fuzzy Classifier	Pratama et al., 2018
	Symbolic Regression (SR)	Zenisek et al., 2019

Vector Machines (SVM) (Mammadov et al., 2021; Qu et al., 2021; Velandia-Cardenas et al., 2021), Boost algorithm (Liu et al., 2020; Mammadov et al., 2021; Trizoglou et al., 2021; Velandia-Cardenas et al., 2021; Zhang et al., 2018) and decision tree (Turnbull et al., 2022; Zenisek et al., 2019; Zhang et al., 2018). These studies investigate early fault detection using machine learning from various application aspects. However, there has been a notable decrease in publications utilizing these methods over the past two years.

(b) Research on wind turbine fault detection utilizing deep learning algorithms has gained momentum, which aligns with the broader surge in deep learning research. The proportion of literature focusing on early fault detection in wind turbines based on deep learning approaches is steadily increasing. Various deep learning architectures have been employed, including Deep Neural Networks (DNN), Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), Recurrent Neural Networks (RNN), Stacked Autoencoders (SAE), Convolutional Neural Networks (CNN), and Variational Autoencoders (VAE), among others (Yang, Liu et al., 2022; Wang et al., 2022; Xiao et al., 2022; Liu et al., 2021; Yang & Zhang, 2021b; Wang et al., 2017; Yang & Zhang, 2021b). Their multilayered and complex designs characterize the structure and mechanism of deep learning algorithms. Furthermore, there is a growing demand for computing resources due to increasing computational requirements.

(c) The method based on statistical probability continues to play a significant role in early fault detection in wind turbines, primarily due to the stochastic nature of wind energy resources (Peña et al., 2021; Rogers et al., 2020; Tang et al., 2022; Wei et al., 2020). These approaches leverage statistical tools such as confidence intervals (Wei et al., 2020) and probability density functions (Peña et al., 2021) to assess the operational status of wind turbines and determine whether they are in an abnormal state.

Analysis of Data Source

Through analysis of selected literature, wind turbine fault data sources can be categorized into four distinct types: open datasets, real SCADA data, simulation data, and experimental platform data. Open datasets are primarily accessible and downloadable from online sources. Real SCADA data consists of actual operational data obtained through project collaborations or direct involvement in wind turbine maintenance activities. Simulation data is generated using software-based simulations on hardware platforms, allowing for the introduction of manual fault simulations and the collection of diverse fault data. Experimental platform data is derived from custom-built hardware simulation platforms, where researchers simulate wind turbine faults and collect fault signals using tailored data acquisition systems to create fault data samples. Table 4 presents a comprehensive analysis of these data source categories.

Through the screening and classification of data sources in Table 4, the following conclusions can be drawn:

Table 4
Data source information

Dataset type	Numbers of unit	Description	Location	Reference
Open datasets	5 WTs	EDP open data, 2-year time span, 10-min time resolution,	--	Latiffianti et al., 2022
	5 WTs	EDP open data, 2-year time span, 10-min time resolution	--	Jia et al., 2021
	2 WTs	From the China Industrial Big Data Competition	Beijing, China	Tong et al., 2022
	3 WTs	European project OPTIMUS, 40 variables every 10 min with 101,752 samples	--	Chacon et al., 2020
Real SCADA data	30 WTs, 22 healthy WTs	A sampling interval of 5min	mid-eastern China.	Yang, Liu et al., 2022
	4 WTs	10-min granularity and 101 features	Prince Edward Island (PEI), Canada	Mammadov et al., 2021

Table 4 *(continue)*

Dataset type	Numbers of unit	Description	Location	Reference
	2 WTs	The 30s per data instance	Inner Mongolia, China	Liu et al., 2020
	3 WTs	Case 1: data per 1s Case 2: data per 1s Case 3: a section of data every 10min	--	Wang et al., 2022
	3 WTs	Every 10 min, 40 attributes	Spain	Zhang & Lang, 2020
	3 WTs	2 MW WTs	--	Liu et al., 2021
	--	Sampling interval 1 min	--	Bo et al., 2019
	21 WTs	10-min averaged SCADA data, 2 separate months of data	Europe	McKinnon et al., 2020
	132 WTs	Sampled each 10 min as averages of the past 10 min interval	--	Herp et al., 2020
	24 WTs	10-min interval-sampling SCADA data	Southern China	Wei et al., 2020
	5 WTs	2.3MW WTs, 10 years data, 10-min sampling interval	North-western Finland	Korkos et al., 2022
	1 WT	2.3 MW Siemens turbine, SCADA data with 10-min sampling interval	--	Pandit & Infield, 2018; Pandit & Infield, 2019
	117 WTs	The Ningxia dataset was sampled in 30-s intervals, and the dataset of Shandong and Anhui was sampled in 10 min intervals	Ningxia; Shandong; Anhui, China	Yang & Zhang, 2021a
	13 WTs	10 min sampling intervals, 2019 – 2020	North of Perú	Velasquez et al., 2021
	1 WTs	Onshore 2MW	Aragon, Spain	Catelani et al., 2020
	5 WTs	2.3 MW Enercon E-70 WTs, 10 min sampling intervals	El Hierro-Canary Islands- Spain	Díaz et al., 2020
	--	3 MW direct-drive turbine, 10 min intervals	South coast of Ireland	Fotiadou et al., 2020
	1 WT	a 2 MW direct-driven WT with cut-in, rated, and cut-out wind speeds of 3, 11, and 25 m/s, SCADA data sampling interval 10min	Lu Hejin wind farm in Chenzhou, southern China	Xiao et al., 2022
	--	10-min interval SCADA data, 4 wind farms	Hebei; Liaoning; Shanxi, China	Yang & Zhang, 2021b
	--	10-min interval, 3 SCADA datasets	Mainland China	Yang, Wang et al., 2022
	92 WTs	10-min interval SCADA data, 6 wind farms	Hebei; Liaoning; Shanxi; Shanxi; Shandong, China	Wang et al., 2017

Table 4 *(continue)*

Dataset type	Numbers of unit	Description	Location	Reference
	60 WTs	4wind farms, with 10-min SCADA data.	Shandong, Anhui, Ningxia, Tianjin, China	Wang, Zhang et al., 2018
	3 WTs	1.5MW, sampling rate of 1Hz	East China	Quanlin et al., 2020
Simulation data	--	1 kHz current signal, sampling time 5.5 s	--	Tang et al., 2022
	1 WT	FAST models, a barge-offshore version	NREL	Tutiven et al., 2018; Pozo et al., 2018
	--	Damaged gear, Cracked gear	MATLAB platform	Agasthian et al., 2019
Experiment platform data	wind turbine platform	Sampling frequency rate of 100 kHz and a sampling duration of 20 s.	--	Du et al., 2022
	--	Data was sampled at 5 kHz, with a 14-bit resolution	--	Sousa et al., 2018

(a) The analysis of data sources reveals that faults in critical wind turbine components, particularly generators and gearboxes, substantially impact operational efficiency and costs. When replacement or maintenance is required, these components are associated with high time, spare parts, and repair costs.

(b) The analysis of data sources reveals a relatively small number of studies based on open datasets (Chacon et al., 2020; Jia et al., 2021; Latiffianti et al., 2022; Tong et al., 2022), with the primary open databases including EDP (Jia et al., 2021; Latiffianti et al., 2022) and OPTIMUS1 (Chacon et al., 2020). Notably, there is a significant lack of comparative analysis between different research methods across these studies, which limits the ability to assess the relative effectiveness of various fault detection approaches when applied to common datasets.

(c) The literature review reveals that studies based on real SCADA data constitute the largest proportion of research, with contributions from diverse geographical regions, including Asia (Yang, Liu et al., 2022), Europe (McKinnon et al., 2020), South America (Velasquez et al., 2021), and North America (Mammadov et al., 2021). However, the significant data variability between studies and the inherent privacy constraints of SCADA datasets limit the generalizability of research findings, confining their applicability primarily to case-specific contexts. Notably, within the geographical distribution of SCADA data sources, China emerges as the predominant contributor, as evidenced by multiple studies (Liu et al., 2020; Quanlin et al., 2020; Wei et al., 2020; Xiao et al., 2022; Yang, Liu et al., 2022; Yang & Zhang, 2021a).

(d) Simulation data predominantly relies on NREL's FAST software (Pozo et al., 2018; Tutiven et al., 2018), with various fault types generated through parameter adjustments in the model. MATLAB/Simulink software is also employed for simulation purposes (Agasthian et al., 2019). However, the effectiveness of this method is directly related to the accuracy of the model and the similarity of wind farms.

(e) Experimental platforms utilizing self-designed research setups have collected fault data with remarkably high sampling frequencies, typically exceeding 1000 Hz (Du et al., 2022; Sousa et al., 2018). This high-frequency data acquisition necessitates subsequent sampling or preprocessing of the original data.

Research Question Analysis

This review has significantly enhanced the understanding of early fault detection in wind turbines through a comprehensive analysis of the literature from the perspectives of fault types, modeling methods, and data sources. However, a more focused examination of the literature is necessary to directly address the three research questions posed in this paper. This discussion aims to provide explicit answers to these questions through a targeted analysis of the relevant studies, synthesizing the insights gained from the diverse approaches and methodologies employed in wind turbine fault detection research.

RQ1: What is the Current Trend in Advanced Fault Detection Methods for Wind Turbines?

With the prosperous development of new energy power generation worldwide, scholars researching early fault detection in wind turbines have rapidly increased. Researchers have explored diverse approaches, including strategy design optimization and enhanced sensor deployment, to pre-emptively identify potential faults. The scope of wind turbine early fault detection research is remarkably broad, encompassing areas such as wind speed-active power curve fitting, temperature signal modeling of rotating components, and the development and threshold optimization of early fault detection indicators. While each of these research domains merits a dedicated review, this paper focuses specifically on wind turbine fault research directions. The subsequent analysis synthesizes the research trends in early fault detection of wind turbines, concentrating on modeling algorithms and fault typologies.

Modeling Algorithm. Based on the analysis of previous modeling methods, it is evident that deep learning and machine learning remain the predominant research approaches for early fault detection in wind turbines. Specifically, the rapid advancements in deep learning have enabled the accurate characterization of nonlinear data, proving highly effective in early fault detection. However, the poor explainability of deep learning inevitably poses challenges in the field of early fault detection for wind turbines. Machine learning approaches maintain a predominant position in this domain, primarily leveraging SVM and related algorithms, regression-based methods, and Boost-related techniques. These approaches offer clear principles and straightforward implementations, contributing to their widespread adoption in wind turbine fault detection research.

Fault type. Wind turbine fault type studies can be broadly categorized into two primary domains: temperature-related and electrical-related faults, which dominate the landscape of wind turbine fault detection research. More attributes are included in fault detection research with the development of SCADA systems. Notably, temperature-related faults typically manifest as gradual processes rather than sudden occurrences, aligning their study with the research paradigm of concept drift algorithms.

Furthermore, analysis of data sources for wind turbine fault detection reveals a trend towards increasing sampling frequencies in SCADA systems, potentially enhancing fault detection accuracy. The literature demonstrates a wide range of sampling frequencies, spanning from 10 minutes to 1 second, indicating that the operation and maintenance of current wind power are developing rapidly, and there is an urgent need for a high-precision early detection model.

RQ2: What is the Current Research Status of Concept Drift Methods in Wind Turbine Fault Detection?

The concept drift methods have laid a solid foundation for early fault detection in wind turbines, although the approaches to research vary significantly among studies. The literature on concept drift has been reviewed, and the findings are summarized in Table 5.

The following conclusions can be drawn from the aforementioned literature review.

(a) Several studies characterize wind turbines' normal operating state and employ algorithms to identify "new" data or "novel" types. These approaches detect states that differ from the normal operating conditions in wind turbines but do not explicitly address the issue of concept drift (Bilendo et al., 2021; Du et al., 2022).

(b) Some detection methods integrate machine learning techniques with concept drift algorithms to create comprehensive strategies for identifying potential faults. For instance, some researchers utilize KNN and SVM in their research approach (Hu et al., 2021). Others employ Dynamic AdaBoost as their modeling method (Lin et al., 2019).

(c) The concept drift method has not yet become a primary research focus in early fault detection for wind turbines. Moreover, the screened literature does not prominently feature typical concept drift algorithms such as EDDM (Early Drift Detection Method) and ADWIN (ADaptive WINdowing).

Table 5
Literature related to concept drift

Problem	Aim/Objective	Main Research methods	Result	References
Detection of the data shift with unlabeled data	Fault diagnosis in an evolving environment	K-nearest neighbor (KNN) classifier SVM classifier	Classification accuracy, parsimony and easiness	Hu et al., 2021
Monitoring ball-bearings without previous ball-bearing RTF data	Online fault detection and prognosis	Hidden Markov models polynomial regression model	Online predictive health assessment	Puerto-Santana et al., 2022
Gearbox fault diagnosis	Fault Severity Diagnosis	Distance metric and the concept detection	KL divergence has been the appropriate metric	Peña et al., 2021
Modeling deteriorated because of the non-stationarities of industrial data	Labeling the potential concept to detect a fault	Statistical detectors and window-based approaches	Different types of drifts – sudden, gradual, recurrent – can be classified	Martinez et al., 2018
Hard fault detection with Big Imbalance Industrial IoT Data	An Ensemble Learning Method of Offline Classifiers	Dynamic AdaBoost	High accuracy rate, over 94%	Lin et al., 2019
Traditional incremental learning model high update frequency	Controlling the incremental update by detecting the concept drift	Shared nearest neighbors (AILSNN)	Higher accuracy than 1DCNN with less training time	Wang, Sun et al., 2021
Wind turbine degradation evaluation	Using information granules to indicate the health state	Concepts extraction using fuzzy c-means clustering	Deterioration was most visible for higher wind speeds	Jastrzebska et al., 2022
Anomaly detection and novel fault discrimination for WTs	Unsupervised method from anomaly detection to novel fault discrimination.	Sparse isolation encoding forest	High diagnostic accuracy	Du et al., 2022
Fault detection by the "drive-train" signal	Effectively detect faults without any prior knowledge	K-means + LDA + ANN	High accuracy by processing the fault-candidates concept	Bilendo et al., 2021
Industrial radial fans predictive maintenance	Identify wear and tear	Linear Regression (LR), Random Forest Regression (RF), Symbolic Regression (SR)	Sufficiently large forecast horizon, predict the drift early	Zenisek et al., 2019
Concept classification for data streams under complex environments	Built upon an evolving classifier	pENsemble	High accuracy and complexity	Pratama et al., 2018

RQ3: What are the Advantages of Distance Metrics for Wind Turbine Fault Detection and Concept Drift?

Distance metrics play a significant role in both early fault detection and concept drift research. In studies of early fault detection based on concept drift algorithms, various types of data, including values, datasets, and indicators, are frequently compared and analyzed. The choice of distance metric is fundamental to the effectiveness of this research. An analysis of literature related to distance metrics has been conducted, and the results are presented in Table 6.

Different distance metric methods exhibit distinct advantages in wind turbine fault detection. KL divergence performs well in comparing data distributions, and it is particularly suitable for diagnosing the severity of faults. Hellinger distance plays a crucial role in concept drift detection. Mahalanobis distance shows advantages in anomaly detection in high-dimensional spaces. The diversity of these distance measurement methods allows researchers to select the most suitable method based on specific problems, thereby enhancing the accuracy and efficiency of fault detection.

After summarizing the literature from Table 6, the following conclusions can be drawn:

(a) Distance metrics are employed in diverse ways across the reviewed literature. For example, some researchers measure different data distributions through KL divergence (Peña et al., 2021). Other researchers obtain better accuracy by measuring the distance between different patterns (Hu et al., 2021).

(b) The role of distance metrics in early fault detection based on concept drift strategies varies. For instance, Mahalanobis distance can be applied to the original data in the preprocessing stage (Renstrom et al., 2020). Additionally, distance metrics can serve as performance indicators in the residual processing stage after modeling (Jastrzebska et al., 2022).

Distance metric plays a crucial role in wind turbine fault detection. They provide the foundation for data analysis and directly influence the accuracy and efficiency of fault detection. By choosing appropriate distance metric methods, researchers can identify changes in data patterns more accurately, thereby detecting potential faults promptly.

State-of-the-Art Performance in Wind Turbine Fault Detection

Reviewing and summarizing the current state-of-the-art research performance is essential before exploring emerging trends and future directions in wind turbine fault detection. This approach provides readers with an overview of the latest achievements in the field and establishes a foundation for subsequent discussions. This discussion will be on comparative analysis of recent, highly representative studies encompassing various fault detection methods and their effectiveness across different fault scenarios.

Table 6
Metric distance analysis

Problem	Aim/Objective	Main Research Methods	Distance Metric Application	Result	References
Detection of the data shift with unlabeled data	Fault diagnosis in an evolving environment	K-nearest neighbor (KNN) classifier SVM classifier	Euclidean distance for pattern similarity measurement	Classification accuracy, parsimony and easiness	Hu et al., 2021
Monitoring ball-bearings without previous ball-bearing RTF data	Online fault detection and prognosis	Hidden Markov models polynomial regression model	Page test and the Chernoff bounds to detect the new data pattern	Online predictive health assessment	Puerto-Santana et al., 2022
Gearbox fault diagnosis	Fault severity diagnosis	Distance metric and the concept detection	KL divergence for comparing data distributions	KL divergence has been the appropriate metric	Peña et al., 2021
Modeling deteriorated because of the non-stationarities of industrial data	Labeling the potential concept to detect a fault	Statistical detectors and window-based approaches	Hellinger distance for concept drift detection	Different types of drifts, sudden, gradual, and recurrent, can be classified	Martinez et al., 2018
Early warning for gearbox failure	Anticipate failure events with a good lead time	Combination of the use of LoMST and a CUSUM approach	Improved CUSUM method is applied to early warning	Average warning lead time is 55 days	Latiffianti et al., 2022
Wind turbine degradation evaluation	Using information granules to indicate the health state	Concepts extraction using fuzzy c-means clustering	Euclidean distance in fuzzy clustering for health state evaluation	Deterioration was most visible for higher wind speeds	Jastrzebska et al., 2022
System-wide anomaly detection for WTs	Monitor the whole component with only one model	Deep autoencoders	Mahalanobis distance processed by EWMA	Code sizes 18 and 24 were the most capable with the detection ability	Renstrom et al., 2020
Fault detection for WTs	Wind-power curve-related NBM	Boosted stacked regressor	Performance metric, MAE is applied	5 types of faults are detected	Bilendo et al., 2022

Table 7 summarizes the key performance indicators of the representative studies.

Through performance analysis of these state-of-the-art methods, it can draw the following observations:

(a) Various advanced modeling techniques show excellent performance across different fault types. For instance, the dual-stage attention-based recurrent neural network (Yang, Liu et al., 2022) achieves the highest accuracy for generator bearing faults, while the Joint Variational Autoencoder (JVAE) (Yang & Zhang, 2021b) demonstrates high F1-Score and low false positive rate for gearbox lubricant pressure anomalies.

(b) Different performance criteria are used across studies, making direct comparisons challenging. Common metrics include accuracy, F1-score, detection rate, and specific measures such as ICS and CRR for certain fault types.

(c) Some methods show promising results for specific fault types. For example, the LoMST and CUSUM approach (Latiffianti et al., 2022) achieves a 100% detection rate for gearbox bearing damage, while the SS-ELM algorithm (Tong et al., 2022) performs well on imbalanced datasets for blade icing fault detection.

(d) Despite these advancements, challenges remain in practical applications. These may include the need for large amounts labeled data and real-time processing capabilities.

Table 7

State-of-the-art performance in wind turbine fault detection methods

Fault type	Modeling type	Performance criteria	reference
Generator bearing Fault	Discrete digital model	Best ICS, CRR achieved	Tang et al., 2022
	Dual-stage attention-based recurrent neural network	Highest accuracy (Acc), recall (Rec), and F_{β}-score among the compared algorithms	Yang, Liu et al., 2022
Gearbox lubricant pressure anomaly	Joint variational autoencoder (JVAE)	F1-Score: 0.914 PR > 97%, FPR < 1%	Yang & Zhang, 2021b
Gearbox bearings damaged	Combination of the use of LoMST and a CUSUM approach	100% detection rate	Latiffianti et al., 2022
Pitch system fault	Adaptive neuro-fuzzy inference system (ANFIS) technique	F1-score 86% F1-score of Pitch faults detection task	Korkos et al., 2022
Blade icing fault	Semi-supervised extreme learning machine (SS-ELM) algorithm	MCC, G-mean, F1_score. Bset imbalance dataset test performance	Tong et al., 2022
Multitask fault detection	Continual Learning, digital twin	F1_score, RMSE Best continual task performance	Yang, Wang et al., 2022

Emerging Trends and Future Directions

As wind turbine technology continues to advance and data science rapidly evolves, the field of wind turbine fault detection is undergoing a significant transformation. In recent years, several emerging technologies and methods have shown great potential to further improve the accuracy and efficiency of fault detection.

Firstly, the application of digital twin(Feng, Ji, Zhang, et al., 2023) technology in wind turbine fault detection is on the rise. As demonstrated by the continuous learning framework proposed by Wang (Yang, Wang et al., 2022), digital twins can provide a unified platform for multiple modeling tasks, including gearbox fault detection, blade fracture detection, and wind power prediction. This approach enhances model generality and improves its adaptability under various operating conditions.

Secondly, the integration of physics-informed and data-driven methods is becoming an important trend (Feng, Ji, Zhang, et al., 2023). While currently primarily applied to vibration data analysis, the concept of this approach can be extended to SCADA data analysis. For instance, incorporating knowledge from physical models into deep learning networks could potentially improve model interpretability and generalization capabilities.

Furthermore, multi-scale feature fusion and novel deep-learning neural network structures (such as gated recurrent units) have shown promising results in bearing health management (Mohammadi et al., 2020; Ni et al., 2024; Xiao et al., 2022). These techniques have the potential to be applied to wind turbine SCADA data analysis to enhance fault detection accuracy and predictive capabilities.

Lastly, with the development of 5G technology and edge computing, the prospects for real-time big data processing and analysis in wind turbine fault detection are broad. It could lead to faster and more precise fault detection systems, reducing maintenance costs and improving wind turbine reliability.

These emerging trends indicate future directions in wind turbine fault detection. Researchers should closely monitor these areas to drive further advancements in the field.

CONCLUSION

This paper provides a comprehensive overview of the current state of wind turbine fault detection research. Fault detection is an important field of study in the wind turbine industry, as it has the potential to reduce operational and maintenance costs.

This paper uses the PRISMA method to select 65 relevant wind turbine fault detection literature. The paper presents essential information by analyzing the literature, including literature type, publication year, and author nationality. Furthermore, by thoroughly analyzing and synthesizing these 65 literature items, the study investigated fault detection components, modeling methods, and data sources. From this comprehensive review, the following conclusions are drawn:

(a) In the component analysis of wind turbine early fault detection, it can be concluded that most research focuses on the abnormal temperature attributes of generators and gearboxes. Research on blade faults is also common, while studies on other component faults are less prevalent.
(b) In current wind turbine fault detection research, modeling methods primarily rely on machine learning and deep learning techniques. Most studies focus on combining algorithms and parameter optimization. However, research on improving the algorithms themselves is relatively scarce.
(c) Most of the research literature relies on real SCADA data for analysis, with a significant proportion originating from China. In contrast, studies utilizing open datasets and other sources are comparatively scarce.

After thoroughly reviewing and organizing the literature, this paper addresses research questions regarding research trends, concept drift, and distance metrics in wind turbine fault detection.

(a) Wind turbine fault detection primarily relies on machine learning and deep learning techniques, with temperature and electrical faults being the predominant focus. As SCADA systems have evolved, a broader range of attributes has been incorporated into early fault detection research.
(b) Concept drift algorithms represent a minor portion of wind turbine early fault detection research. They typically achieve fault detection objectives through integration with other methods. These algorithms show considerable potential and suitability for studying early fault detection in wind turbines.
(c) The distance metric plays a fundamental role in modeling and is one of the key elements for early fault detection. Improving the distance metric can enhance the effectiveness of early fault detection.

Research Gaps and Future Work

As the wind turbine fault detection field rapidly evolves, it is crucial to identify existing research gaps and outline future directions. This discussion synthesizes the key challenges that emerged from comprehensive literature review and proposes potential avenues for future research.

This study employs the PRISMA method to systematically review existing literature, revealing several key research gaps and future challenges in wind turbine fault detection.

(a) Data Quality and Availability

The paucity of openly accessible, high-quality, large-scale SCADA datasets significantly impedes cross-method comparisons.

Current research predominantly utilizes data from specific wind farms, resulting in limited generalizability across diverse geographical regions and turbine types.

(b) Algorithm Adaptability and Robustness
Existing methodologies demonstrate insufficient adaptability to diverse operational conditions and environmental fluctuations.
The current state-of-the-art exhibits limited efficacy in detecting and identifying rare or novel fault types.
(c) Multi-source Data Fusion
Comprehensive approaches are lacking for effectively integrating SCADA data, vibration data, meteorological data, and other multi-source information.
Inadequate research on how to collaboratively analyze data with different sampling frequencies and characteristics.
(d) Real-time Capability and Computational Efficiency
Numerous advanced deep learning algorithms exhibit substantial computational complexity, thereby challenging their feasibility and efficacy for real-time monitoring applications in wind turbine fault detection systems.
Lack of research on lightweight fault detection algorithms suitable for edge computing environments.
(e) Interpretability
The inherent opacity of deep learning methodologies, often referred to as the "black box" phenomenon, significantly impedes their credibility and interpretability in practical fault detection applications.
There is a notable deficiency in research on effectively synthesizing domain knowledge from physical models with data-driven approaches.

Further analysis shows that it cannot cover the current status of wind turbines' overall fault detection research. The aspects that need to be improved are as follows.

(a) There are few horizontal comparisons of relevant literature on temperature parameter modeling, and relevant horizontal comparison research plans to carry out. The research focuses on data preprocessing methods, selecting suitable SCADA attributes for different faults, and threshold determination methods.
(b) Early fault detection for wind turbines based on SCADA data often focuses on temperature and electrical signals, while pressure and torque signals have received less study and analysis.
(c) The literature on wind turbine early fault detection based on non-parametric methods needs to be further studied.
(d) Current literature on the fusion of SCADA data and other data for fault detection should be studied
(e) Further literature review and summarization of its developmental characteristics are necessary to incrementally update the model for early fault detection in wind turbines.
(f) The classification method and modeling method need to be further improved.

ACKNOWLEDGMENTS

The authors sincerely thank Universiti Teknologi MARA (UiTM) and the College of Computing, Informatics and Mathematics for providing the resources to carry out this research. Sincere thanks also go to our team members, Dr. Zainura Idrus and Dr. Raseeda Hamzah, for their valuable input throughout the research implementation process. Their technical assistance is greatly appreciated. Finally, the authors deeply appreciate the helpful comments from the editor and the anonymous reviewers. Their constructive feedback has significantly improved the quality of this paper. The authors would also like to thank the Journal Support Fund under the Institute of Postgraduate Studies, Universiti Teknologi MARA.

REFERENCES

Agasthian, A., Pamula, R., & Kumaraswamidhas, L. A. (2019). Fault classification and detection in wind turbine using cuckoo-optimized support vector machine. *Neural Computing & Applications*, *31*(5), 1503–1511. https://doi.org/10.1007/s00521-018-3690-z

Aziz, U., Charbonnier, S., Berenguer, C., Lebranchu, A., & Prevost, F. (2021). Critical comparison of power-based wind turbine fault-detection methods using a realistic framework for SCADA data simulation. *Renewable & Sustainable Energy Reviews*, *144*, Article 110961. https://doi.org/10.1016/j.rser.2021.110961

Aziz, U., Charbonnier, S., Berenguer, C., Lebranchu, A., & Prevost, F. (2022). A multi-turbine approach for improving performance of wind turbine power-based fault detection methods. *Energies*, *15*(8), Article 2806. https://doi.org/10.3390/en15082806

Badihi, H., Zhang, Y., Jiang, B., Pillay, P., & Rakheja, S. (2022). A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis. *Proceedings of the IEEE*, *110*(6), 754–806. https://doi.org/10.1109/JPROC.2022.3171691

Bilendo, F., Badihi, H., Lu, N., Cambron, P., & Jiang, B. (2021, September 17-20). *An intelligent data-driven machine learning approach for fault detection of wind turbines*. [Paper presentation]. 6th International Conference on Power and Renewable Energy (ICPRE), Shanghai, China. https://doi.org/10.1109/ICPRE52634.2021.9635340

Bilendo, F., Badihi, H., Lu, N., Cambron, P., & Jiang, B. (2022). Power curve-based fault detection method for wind turbines. *IFAC-PapersOnLine*, *55*(6), 408–413. https://doi.org/10.1016/j.ifacol.2022.07.163

Bo, Y. F., Zeng, X. J., Yang, M., & Zhu, Y. (2019). Anomaly detection for wind turbine gearbox oil pressure difference based on SCADA data. *IOP Conference Series: Earth and Environmental Science 354,* Article 012115. https://doi.org/10.1088/1755-1315/354/1/012115

Catelani, M., Ciani, L., Galar, D., & Patrizi, G. (2020). Risk assessment of a wind turbine: A new FMECA-based tool with RPN threshold estimation. *IEEE Access*, *8*, 20181–20190. https://doi.org/10.1109/ACCESS.2020.2968812

Chacon, A. M. P., Ramirez, I. S., & Marquez, F. P. G. (2020). False alarms analysis of wind turbine bearing system. *Sustainability*, *12*(19), Article 7867. https://doi.org/10.3390/su12197867

Dhanola, A., & Garg, H. C. (2020). Tribological challenges and advancements in wind turbine bearings: A review. *Engineering Failure Analysis*, *118*, Article 104885. https://doi.org/10.1016/j.engfailanal.2020.104885

Díaz, S., Carta, J. A., & Castañeda, A. (2020). Influence of the variation of meteorological and operational parameters on estimation of the power output of a wind farm with active power control. *Renewable Energy*, *159*, 812–826. https://doi.org/10.1016/j.renene.2020.05.187

Du, W., Guo, Z., Li, C., Gong, X., & Pu, Z. (2022). From anomaly detection to novel fault discrimination for wind turbine gearboxes with a sparse isolation encoding forest. *IEEE Transactions on Instrumentation and Measurement*, *71*, Article 2512710. https://doi.org/10.1109/TIM.2022.3187737

Feng, K., Ji, J. C., Ni, Q., & Beer, M. (2023). A review of vibration-based gear wear monitoring and prediction techniques. *Mechanical Systems and Signal Processing*, *182*, Article 109605. https://doi.org/10.1016/j.ymssp.2022.109605

Feng, K., Ji, J. C., Zhang, Y., Ni, Q., Liu, Z., & Beer, M. (2023). Digital twin-driven intelligent assessment of gear surface degradation. *Mechanical Systems and Signal Processing*, *186*, Article 109896. https://doi.org/10.1016/j.ymssp.2022.109896

Fernandes, M., Corchado, J. M., & Marreiros, G. (2022). Machine learning techniques applied to mechanical fault diagnosis and fault prognosis in the context of real industrial manufacturing use-cases: A systematic literature review. *Applied Intelligence*, *52*(12), 14246–14280. https://doi.org/10.1007/s10489-022-03344-3

Fotiadou, K., Velivassaki, T. H., Voulkidis, A., Skias, D., De Santis, C., & Zahariadis, T. (2020). Proactive critical energy infrastructure protection via deep feature learning. *Energies*, *13*(10), Article 2622. https://doi.org/10.3390/en13102622

Herp, J., Pedersen, N. L., & Nadimi, E. S. (2020). Assessment of early stopping through statistical health prognostic models for empirical RUL estimation in wind turbine main bearing failure monitoring. *Energies*, *13*(1), Article 83. https://doi.org/10.3390/en13010083

Hu, Y., Baraldi, P., Di Maio, F., Liu, J., & Zio, E. (2021). A method for fault diagnosis in evolving environment using unlabeled data. *Proceedings of the Institution of Mechanical Engineers Part O-Journal of Risk and Reliability*, *235*(1), 33–49. https://doi.org/10.1177/1748006X20946529

Jastrzebska, A., Hernández, A. M., Nápoles, G., Salgueiro, Y., & Vanhoof, K. (2022). Measuring wind turbine health using fuzzy-concept-based drifting models. *Renewable Energy*, *190*, 730–740. https://doi.org/10.1016/j.renene.2022.03.116

Jia, X., Han, Y., Li, Y., Sang, Y., & Zhang, G. (2021). Condition monitoring and performance forecasting of wind turbines based on denoising autoencoder and novel convolutional neural networks. *Energy Reports*, *7*, 6354–6365. https://doi.org/10.1016/j.egyr.2021.09.080

Kavaz, A. G., & Barutcu, B. (2018). Fault detection of wind turbine sensors using artificial neural networks. *Journal of Sensors*, *2018*(1), Article 5628429. https://doi.org/10.1155/2018/5628429

Korkos, P., Linjama, M., Kleemola, J., & Lehtovaara, A. (2022). Data annotation and feature extraction in fault detection in a wind turbine hydraulic pitch system. *Renewable Energy*, *185*, 692–703. https://doi.org/10.1016/j.renene.2021.12.047

Latiffianti, E., Sheng, S., & Ding, Y. (2022). Wind turbine gearbox failure detection through cumulative sum of multivariate time series data. *Frontiers in Energy Research*, *10*, Article 904622. https://doi.org/10.3389/fenrg.2022.904622

Lin, C. C., Deng, D. J., Kuo, C. H., & Chen, L. (2019). Concept drift detection and adaption in big imbalance industrial IoT data using an ensemble learning method of offline classifiers. *IEEE Access*, *7*, 56198–56207. https://doi.org/10.1109/ACCESS.2019.2912631

Liu, H., Yu, C., & Yu, C. (2021). A new hybrid model based on secondary decomposition, reinforcement learning and SRU network for wind turbine gearbox oil temperature forecasting. *Measurement*, *178*, Article 109347. https://doi.org/10.1016/j.measurement.2021.109347

Liu, Y., Wu, Z., & Wang, X. (2020). Research on fault diagnosis of wind turbine based on SCADA data. *IEEE Access*, *8*, 185557–185569. https://doi.org/10.1109/ACCESS.2020.3029435

Mammadov, E., Farrokhabadi, M., & Cañizares, C. A. (2021, October 18-21). *AI-enabled predictive maintenance of wind generators.* [Paper presentation]. IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe), Espoo, Finland. https://doi.org/10.1109/ISGTEurope52324.2021.9640162

Martinez, I., Viles, E., & Cabrejas, I. (2018). Labelling drifts in a fault detection system for wind turbine maintenance. In J. DelSer, E. Osaba, M. N. Bilbao, J. J. SanchezMedina, M. Vecchio, & X. S. Yang (Eds.), *Intelligent Distributed Computing XII* (pp. 145–156). Springer. https://doi.org/10.1007/978-3-319-99626-4_13

Márquez, F. P. G., & Chacón, A. M. P. (2020). A review of non-destructive testing on wind turbines blades. *Renewable Energy*, *161*, 998–1010. https://doi.org/10.1016/j.renene.2020.07.145

McKinnon, C., Carroll, J., McDonald, A., Koukoura, S., Infield, D., & Soraghan, C. (2020). Comparison of new anomaly detection technique for wind turbine condition monitoring using gearbox SCADA data. *Energies*, *13*(19), Article 5152. https://doi.org/10.3390/en13195152

Mohammadi, H. G., Arshad, R., Rautmare, S., Manjunatha, S., Kuschel, M., Jentzsch, F. P., Platzner, M., Boschmann, A., & Schollbach, D. (2020, September 8-11). *DeepWind: An accurate wind turbine condition monitoring framework via deep learning on embedded platforms*. 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), Vienna, Austria's. https://doi.org/10.1109/ETFA46521.2020.9211880

Ni, Q., Ji, J. C., Feng, K., Zhang, Y., Lin, D., & Zheng, J. (2024). Data-driven bearing health management using a novel multi-scale fused feature and gated recurrent unit. *Reliability Engineering & System Safety*, *242*, Article 109753. https://doi.org/10.1016/j.ress.2023.109753

Ni, Q., Ji, J. C., Halkon, B., Feng, K., & Nandi, A. K. (2023). Physics-informed residual network (PIResNet) for rolling element bearing fault diagnostics. *Mechanical Systems and Signal Processing*, *200*, Article 110544. https://doi.org/10.1016/j.ymssp.2023.110544

Pandit, R. K., & Infield, D. (2018). SCADA-based wind turbine anomaly detection using gaussian process models for wind turbine condition monitoring purposes. *IET Renewable Power Generation*, *12*(11), 1249–1255. https://doi.org/10.1049/iet-rpg.2018.0156

Pandit, R. K., & Infield, D. (2019). Comparative analysis of gaussian process power curve models based on different stationary covariance functions for the purpose of improving model accuracy. *Renewable Energy*, *140*, 190–202. https://doi.org/10.1016/j.renene.2019.03.047

Peña, M., Lanzarini, L., Cerrada, M., Cabrera, D., & Sánchez, R. V. (2021, October 12-15). *Data-driven gearbox fault severity diagnosis based on concept drift*. IEEE Fifth Ecuador Technical Chapters Meeting (ETCM), Cuenca, Ecuador. https://doi.org/10.1109/ETCM53643.2021.9590689

Pozo, F., Vidal, Y., & Salgado, O. (2018). Wind turbine condition monitoring strategy through multiway PCA and multivariate inference. *Energies*, *11*(4), Article 749. https://doi.org/10.3390/en11040749

Pratama, M., Pedrycz, W., & Lughofer, E. (2018). Evolving ensemble fuzzy classifier. *IEEE Transactions on Fuzzy Systems*, *26*(5), 2552–2567. https://doi.org/10.1109/TFUZZ.2018.2796099

Puerto-Santana, C., Bielza, C., Diaz-Rozo, J., Ramirez-Gargallo, G., Mantovani, F., Virumbrales, G., Labarta, J., & Larranaga, P. (2022). Asymmetric HMMs for online ball-bearing health assessments. *IEEE Internet of Things Journal*, *9*(20), 20160–20177. https://doi.org/10.1109/JIOT.2022.3173064

Qu, F., Liu, J., Liu, X., & Jiang, L. (2021). A multi-fault detection method with improved triplet loss based on hard sample mining. *IEEE Transactions on Sustainable Energy*, *12*(1), 127–137. https://doi.org/10.1109/TSTE.2020.2985217

Quanlin, Z., Xiaoxiao, Z., & Chenggang, H. (2020, December 18-20). *An automatic data cleaning and operating conditions classification method for wind turbines scada system*. 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China. https://doi.org/10.1109/ICCWAMTIP51612.2020.9317436

Renstrom, N., Bangalore, P., & Highcock, E. (2020). System-wide anomaly detection in wind turbines using deep autoencoders. *Renewable Energy*, *157*, 647–659. https://doi.org/10.1016/j.renene.2020.04.148

Rogers, T. J., Gardner, P., Dervilis, N., Worden, K., Maguire, A. E., Papatheou, E., & Cross, E. J. (2020). Probabilistic modelling of wind turbine power curves with application of heteroscedastic gaussian process regression. *Renewable Energy*, *148*, 1124–1136. https://doi.org/10.1016/j.renene.2019.09.145

Sousa, P. H. F. D., Nascimento, N. M. M., Filho, P. P. R., & Medeiros, C. M. S. D. (2018, July 8-13). *Detection and classification of faults in induction generator applied into wind turbines through a machine learning approach*. [Paper presentation]. International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil. https://doi.org/10.1109/IJCNN.2018.8489521

Tang, H., Dai, H. L., & Du, Y. (2022). Bearing fault detection for doubly fed induction generator based on stator current. *IEEE Transactions on Industrial Electronics*, *69*(5), 5267–5276. https://doi.org/10.1109/TIE.2021.3080216

Tao, L., Siqi, Q., Zhang, Y., & Shi, H. (2019). Abnormal detection of wind turbine based on SCADA data mining. *Mathematical Problems in Engineering*, *2019*(1), Article 5976843. https://doi.org/10.1155/2019/5976843

Tong, R., Li, P., Gao, L., Lang, X., Miao, A., & Shen, X. (2022). A novel ellipsoidal semisupervised extreme learning machine algorithm and its application in wind turbine blade icing fault detection. *IEEE Transactions on Instrumentation and Measurement*, *71*, Article 3523116. https://doi.org/10.1109/TIM.2022.3205920

Trizoglou, P., Liu, X., & Lin, Z. (2021). Fault detection by an ensemble framework of extreme gradient boosting (XGBoost) in the operation of offshore wind turbines. *Renewable Energy*, *179*, 945–962. https://doi.org/10.1016/j.renene.2021.07.085

Turnbull, A., Carroll, J., & McDonald, A. (2022). A comparative analysis on the variability of temperature thresholds through time for wind turbine generators using normal behaviour modelling. *Energies*, *15*(14), Article 5298. https://doi.org/10.3390/en15145298

Tutiven, C., Vidal, Y., Acho, L., & Rodellar, J. (2018). Fault detection and isolation of pitch actuator faults in a floating wind turbine. *IFAC-PapersOnLine*, *51*(24), 480–487. https://doi.org/10.1016/j.ifacol.2018.09.620

Velandia-Cardenas, C., Vidal, Y., & Pozo, F. (2021). Wind turbine fault detection using highly imbalanced real SCADA data. *Energies*, *14*(6), Article 1728. https://doi.org/10.3390/en14061728

Velasquez, R. M. A., Tataje, F. A. O., & Ancaya-Martinez, M. D. C. E. (2021). Early detection of faults and stall effects associated to wind farms. *Sustainable Energy Technologies and Assessments*, *47*, Article 101441. https://doi.org/10.1016/j.seta.2021.101441

Wang, B., Sun, N., Wang, Z., & Han, G. (2021, November 26-28). *An adaptive incremental learning algorithm based on shared nearest neighbors in fault detection*. [Paper presentation] Computing, Communications and IoT Applications (ComComAp), Shenzhen, China. https://doi.org/10.1109/ComComAp53641.2021.9652989

Wang, L., Jia, S., Yan, X., Ma, L., & Fang, J. (2022). A SCADA-data-driven condition monitoring method of wind turbine generators. *IEEE Access*, *10*, 67532–67540. https://doi.org/10.1109/ACCESS.2022.3185259

Wang, L., Zhang, Z., Long, H., Xu, J., & Liu, R. (2017). Wind turbine gearbox failure identification with deep neural networks. *IEEE Transactions on Industrial Informatics*, *13*(3), 1360–1368. https://doi.org/10.1109/TII.2016.2607179

Wang, L., Zhang, Z., Xu, J., & Liu, R. (2018). Wind turbine blade breakage monitoring with deep autoencoders. *IEEE Transactions on Smart Grid*, *9*(4), 2824–2833. https://doi.org/10.1109/TSG.2016.2621135

Wang, X., Zhao, Q., Yang, X., & Zeng, B. (2021). Analysis of long-term temperature monitoring of multiple wind turbines. *Measurement & Control*, *54*(5–6), 627–640. https://doi.org/10.1177/00202940211013061

Wang, Y., Ma, X., & Qian, P. (2018). Wind turbine fault detection and identification through PCA-based optimal variable selection. *IEEE Transactions on Sustainable Energy*, *9*(4), 1627–1635. https://doi.org/10.1109/TSTE.2018.2801625

Wei, L., Qian, Z., & Zareipour, H. (2020). Wind turbine pitch system condition monitoring and fault detection based on optimized relevance vector machine regression. *IEEE Transactions on Sustainable Energy*, *11*(4), 2326–2336. https://doi.org/10.1109/TSTE.2019.2954834

Xiao, X., Liu, J., Liu, D., Tang, Y., & Zhang, F. (2022). Condition monitoring of wind turbine main bearing based on multivariate time series forecasting. *Energies*, *15*(5), Article 1951. https://doi.org/10.3390/en15051951

Xu, Q., Fan, Z., Jia, W., & Jiang, C. (2019). Quantile regression neural network-based fault detection scheme for wind turbines with application to monitoring a bearing. *Wind Energy*, *22*(10), 1390–1401. https://doi.org/10.1002/we.2375

Yang, L., Wang, L., Zheng, Z., & Zhang, Z. (2022). A continual learning-based framework for developing a single wind turbine cybertwin adaptively serving multiple modeling tasks. *IEEE Transactions on Industrial Informatics*, *18*(7), 4912–4921. https://doi.org/10.1109/TII.2021.3130721

Yang, L., & Zhang, Z. (2021a). A conditional convolutional autoencoder-based method for monitoring wind turbine blade breakages. *IEEE Transactions on Industrial Informatics*, *17*(9), 6390–6398. https://doi.org/10.1109/TII.2020.3011441

Yang, L., & Zhang, Z. (2021b). Wind turbine gearbox failure detection based on SCADA data: A deep learning-based approach. *IEEE Transactions on Instrumentation and Measurement*, *70*, Article 3507911. https://doi.org/10.1109/TIM.2020.3045800

Yang, Q., Liu, G., Bao, Y., & Chen, Q. (2022). Fault detection of wind turbine generator bearing using attention-based neural networks and voting-based strategy. *IEEE/ASME Transactions on Mechatronics*, *27*(5), 3008–3018. https://doi.org/10.1109/TMECH.2021.3127213

Yi, H., Jiang, Q., Yan, X., & Wang, B. (2021). Imbalanced classification based on minority clustering synthetic minority oversampling technique with wind turbine fault detection application. *IEEE Transactions on Industrial Informatics*, *17*(9), 5867–5875. https://doi.org/10.1109/TII.2020.3046566

Zenisek, J., Holzinger, F., & Affenzeller, M. (2019). Machine learning based concept drift detection for predictive maintenance. *Computers & Industrial Engineering*, *137*, Article 106031. https://doi.org/10.1016/j.cie.2019.106031

Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B., & Si, Y. (2018). A data-driven design for fault detection of wind turbines using random forests and XGboost. *IEEE Access*, *6*, 21020–21031. https://doi.org/10.1109/ACCESS.2018.2818678

Zhang, K., Tang, B., Deng, L., & Yu, X. (2021). Fault detection of wind turbines by subspace reconstruction-based robust kernel principal component analysis. *IEEE Transactions on Instrumentation and Measurement*, *70*, Article 3515711. https://doi.org/10.1109/TIM.2021.3075742

Zhang, S., & Lang, Z. Q. (2020). SCADA-data-based wind turbine fault detection: A dynamic model sensor method. *Control Engineering Practice*, *102*, Article 104546. https://doi.org/10.1016/j.conengprac.2020.104546

Zhao, H., Chen, G., Hong, H., & Zhu, X. (2021). Remote structural health monitoring for industrial wind turbines using short-range doppler radar. *IEEE Transactions on Instrumentation and Measurement*, *70*, Article 8002609. https://doi.org/10.1109/TIM.2021.3053959