Hybrid CNN-Integrated LSTM for Fault Detection and Diagnosis of Wind Turbines

Bisrat Gebreamlak

Master's Thesis

UNIVERSITY OF EASTERN FINLAND

School of Computing

Computer Science

May 2024

UNIVERSITY OF EASTERN FINLAND, Faculty of Science and Forestry

Joensuu School of Computing

Computer Science

Student, Bisrat Gebrealmlak: Hybrid CNN-Integrated LSTM for Fault Detection and Diagnosis of Wind Turbines

Supervisor of the master's thesis: Professor Xiao-Zhi Gao

May 2024

Abstract: The study has been based on wind turbulence fault detection and diagnosis with the help of CNN integrated and LSTM models based on machine learning approaches. There has been observed to be an option of several types and variations of explanations that do not create or display any form of turbines as the initial expenses of installation of detectors and data collection have been regarded to be good for data collection and analysis. The combination of CNN and LSTM has been noted to have exceptional performance and a reduction in the possibilities of errors, crucial for this subject area. The separation of information among all kinds of groups or distances has been designed to be known to be as well as designed to be flexible for an integrated clustering process. The SCADA system data has been utilized and based on it all forms of critical analysis have been done to the study. Based on the findings, it can be noted that the proposed fault detection and diagnosis approach based on the LSTM-CNN model has highlighted effectiveness in terms of accuracy, precision, and F1-score. Thus, integrated supervised and unsupervised learning models that improve decision-making accuracy and generate the optimal predictive model to understand the effective diagnosis of faults in wind turbines can be designed.

Keywords

Machine learning, Diagnosis, Defects, Wind turbines, Wind energy, Deep learning, Time-series analysis, Forecasts

Acknowledgments

I would like to thank my supervisor, Prof. Xiao-Zhi Gao, for his patient supervision, invaluable help, and continuous encouragement during my studies at UEF (University of Eastern Finland) (University of Eastern Finland). Prof. Gao's remarkable knowledge and academic expertise have been instrumental to my progress, and I feel exceptionally fortunate to have had such a dedicated and responsive mentor. His willingness to provide prompt feedback and genuine concern for my academic development has been invaluable, making my research journey both productive and enjoyable.

A special note of appreciation goes to my parents. Their teachings, support, and encouragement have been fundamental to me. They have instilled in me the values of perseverance, hard work, and integrity, and their belief in my abilities has been truly uplifting.

My heartfelt thanks go to my family and friends for their constant assistance. They have been my guiding light, providing the emotional strength and encouragement needed to overcome challenges. Their constant presence and faith in my potential have given me the confidence to pursue my goals and succeed. The joy and understanding they have shared with me have been essential in maintaining a balanced and positive outlook throughout my academic journey.

List of Figures

List of Abbreviations

Abbreviations	Full Form
CNN	Convolutional neural network
LSTM	Long Short-Term Memory
AoA	Angle of Attack
EU	European Union

Table of Contents

1. Introduction

Mucchielli et al. (2021) stated that the EU Renewable Energies Directive, Ireland must have source 40% of its power from renewable sources by 2020 to be able to meet the legally mandated aim of generating 16% of its yearly demand for energy from resources. Ireland boasts one of Europe's greatest wind resources in addition to a developed wind energy sector. Because of that, it is anticipated that solar power can achieve a huge portion of this goal. Compared to other machines that spin, wind turbine parts could encounter significant levels of strain during their service life because of the very unpredictable pressures they experience from diverse and turbulent wind conditions. As a result, up to 30% of the cost of producing wind power is spent on upkeep and operation. In the wind turbine strategy, where wind turbines frequently must run independently within remote locations, it could prove impossible to conduct regular visual examinations, making the capacity to remotely observe element quality more essential compared to other industries. A wind turbine yearly maintenance budget may be heavily depleted by corrective repair required for unexpected breakdowns, which can be exceedingly costly. It may be avoided with scheduled preventive maintenance, which entails periodic checks and repairs. The lifespan of the parts could not have reached their maximum potential at the time they were replaced or repaired, and the price of more regular downtime for inspection could be incredibly significant. Nevertheless, this remains the case in some needless expenses.

1.1 Background

Baboli et al. (2020) stated that the Condition-based maintenance, potentially CBM, was a technique in which a successful servicing selection is taken, when necessary, through continuously monitoring the state of the machines to identify potential or emerging problems. When compared to planned maintenance for wind turbines, such a method can reduce service expenditures by up to 20–25%. Prognostic analysis, which estimates an element's remaining usable life (RUL), is another capability of CBM. This may make it possible to arrange servicing duties even more precisely. Wind turbine had the condition monitoring systems (CMSs) mostly comprised of vibration detectors attached to turbine sub-assemblies for localized monitoring. Visual strain measurements and oil particles counts are occasionally used in conjunction with vibration sensors. Such data is routed to a centralized information processing system, where exclusive software is used for analysis. An alert is then generated if an early defect was found. Nevertheless, considering such apparent advantages, the wind energy sector has not fully adopted CBM and predictive technology.

There are several explanations for each turbine that could be spent on the initial expenses of installing detectors in addition to data collecting and analysis. Commercial wind turbine CMSs have not performed as well as anticipated because of the inherent dangers and errors with CM approaches, even though CMSs were extensively effective for various applications. It has resulted in false alarms that could be expensive because they need human examinations and delay. Even more concerning is the fact that in certain instances they have not performed well enough to identify early flaws. Poor upkeep might result in catastrophic breakdown of linked systems and equipment. Although the goal of wind turbine CMSs is to install more sensors as a way provide comprehensive forecasting about turbine component parts, the turbine currently has multiple detectors connected to the Supervisory Control and Data Acquisition (SCADA) system. A major attempt was made recently to use CM methods on wind turbines through the analysis of information gathered by SCADA systems. A broad range of evaluations, including active and reactive electricity, A device current and voltages, anemometer-measured speed of the wind, the engine shaft velocity, power source, gear system, and the roof temperatures, are included in SCADA data, which is usually collected at intervals of ten minutes to minimize the information transmitted bandwidth and storage. It is feasible to determine whether a defect has emerged or whether the generator could soon enter a period of low efficiency simply running statistical examinations on several patterns throughout that information. All the above can be accomplished without incurring additional expenses from equipping the engine using additional detectors offshore (Kang et al. 2020).

1.1.1 Wind Turbine Failure Modes

Yu et al. (2018) studied thoroughly in the investigation of the frequency of various methods of failure overall windmill elements and individual parts including its effects of interruptions, Figure 1.1 presents the findings of a Failure Mode Effects Analysis (FMEA) over windmills. It is evident that the electrical supply had the most impact on the complete breakdown rate. It translates to a contribution to total downtime on the examined turbines of less than 40%. This information is derived from research conducted by the EU FP7 Wind project, which was carried out by a group of academics and technological specialists, including wind business participants.

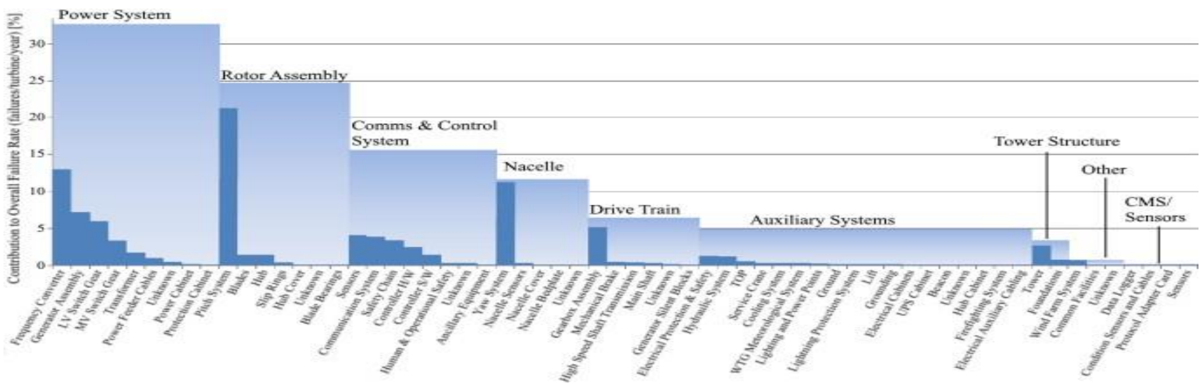

Figure 1.1: Failure Mode Effects Analysis (FMEA) in wind turbine (Source: Yu et al. 2018)

1.1.2 Power Curve

Aziz (2020) determined that the power-wind velocity connection with a particular turbine of the energy output of the turbine is displayed on this graph, sometimes referred to as an efficiency curve, as an indicator of hub height and wind speed. A key indicator of wind turbine functioning is the output waveform of the windmill illustrated in Figure 1.2 (a). The potential trajectories from different windmill designs vary depending on the circumstances of operation that they were created intended, usually within a particular band of wind velocity. Three key elements on this graph could be connected to a turbine's performance at varying wind speeds. The minimal acceptable wind velocity when a turbine starts producing electricity was called the cutting speed. The velocity where the maximum nominal generator output is achieved is known as the recommended speed. The highest speed at which the turbine can generate electricity is known as the cut-out velocity. Engineering and safety restrictions limited that; however certain turbine types enable a restricted power output above it by cleverly adjusting the blade pitch angle. The vendor will often provide the power curve for a certain generator type according to an operational parameter that was assured.

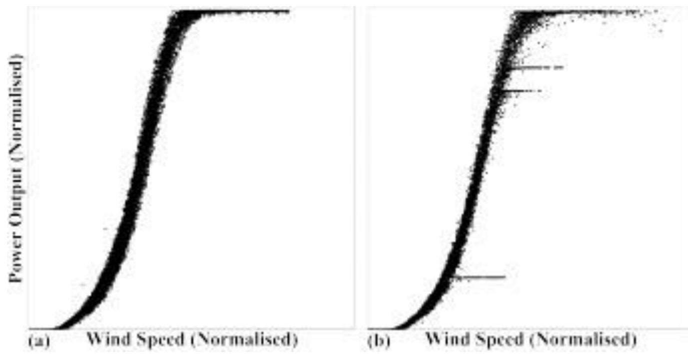

Figure 1.2: Power Curve of Wind Turbine (Source: Aziz, 2020)

A vital way of determining when a windmill is operating appropriately was to compare the power it generates at a certain velocity of the wind to the power curve that was provided. Nonetheless, that power curve is normally developed by the supplier in accordance with industry standards, such as IEC 61400-12-1. Additionally, they are created under typical working conditions using a particular approach that cannot be replicated at a windmill that is already in operation. Since turbines are typically installed on locations with different topographies and wind patterns, any divergence from the manufacturer's power curve may not be a sign of a malfunctioning windmill but rather could be the result of multiple external variables. A novel power curve may be built and utilized as a visual aid for tracking performance in coming years by utilizing data collected during a specific turbine's free of errors optimum operation. Since the landscape and wind direction at a particular location will mostly stay the same over time, any variations in the energy curve's distinctive shape could be attributed to modifications in the windmill itself, which an expert can clearly identify as the root cause of a particular incipient fault referred in the Fig. 2 (b). A specialist could clarify the distinctive form of this power curve by stating that improper regulator parameters resulted in reduced power output (Moreno et al. 2020).

1.1.3 Review of SCADA-based CM systems

Aziz et al. (2021) stated that many efforts were made in previous years to apply analytical and neural network techniques to computerize the diagnosis process. Several methods are available to estimate the power curve in typical operating scenarios. After that, an aggregate remainder can be generated periodically by comparing this to the data obtained online. When the residual rises above a particular level, there may be a turbine issue. They employed that the copulate statistical for modeling the strength curve during typical circumstances and show that a divergence beyond that might potentially be used in the probable future to indicate a Figure 2 shows the power curves for the windmill under both (a) ideal operating conditions and (b)malfunctioning controller values. Kernel approaches are used to simulate the power curve; variations from this correlate to times when the turbine performs poorly because of ice, controller problems, power de-rating, and noise reduction mode operating. But the approach does not distinguish between these errors. They carried out a comparable investigation by modeling the power curve with Gaussian Processes methods. They were successful in demonstrating a decline of efficiency that started shortly before the turbine's primary gear failed. Once more, though, this approach lacked diagnostic skills. A brand-new method for power curve modeling was created. It exists to find the average power production across multiple wind speed "containers." Later, an interim power curve is built by extending an interpolation between those two points. The next step is to move this curve up and down by different degrees to obtain optimum boundaries. After filtering out any areas that fall separate from specified ranges, the procedure is repeated whenever a framework that adequately represents actual functioning is discovered. The next step was to create smart alert limitations to identify unusual behavior in the future. This approach did not identify a specific problem; rather, it indicated improper operation. Using metrics for performance other than the strength curve is an extension of the previously mentioned techniques. Additional metrics including turbine heat, rotor rotation rate, transmission oil temperatures, and generator containing temperatures were added in the CNN training process.

Zhang and Lang (2020) illustrated how their framework effectively deteriorates prior to failure. Different research used winds that became against power production, rotor velocity, and blades pitch angle to assess performance. This provided a useful performance statistic for the turbine, but the research's purview did not include defect diagnostics. They have successfully shown fault categorization and restricted failure forecasting utilizing a larger range of SCADA information. For evaluation of its efficacy on defect diagnosis along with prediction, a variety of models, including neural networks, enhancing trees, support vector machines, as well as traditional classifications regression trees, was constructed. The literature makes it abundantly evident that with only SCADA data, it is possible to identify certain indications of a major component's full breakdown decades before advanced. However, forecasting of over thirty minutes ahead of time remains extremely difficult on fewer significant but frequently occurring failures that also lead to reduced turbine efficiency, including energy eating, blades pitch, or deflector defects. The aim in our study was to expand the prediction capacity for these kinds of errors. In the most effective attempt to achieve that, which was previously discovered that the boosted tree approach outperformed other techniques, such as Support Vector Machines (SVMs), however the researchers failed to elaborate on the particulars of the models that were employed. SVMs are a popular and effective method for resolving classification issues, unfortunately. The SVM's fundamental tenet is that, using designated training information, a choice boundary that separates opposed categories was drawn. To avoid any excessive fitting, certain numbers of scores could be incorrectly categorized. They are particularly appropriate for this situation, in which there is an extensive along with dynamic interaction among a great deal of variables (such as the numerous values gathered by a SCADA system). They have shown excellent results when utilized for defect detection and tracking of conditions across various sectors.

Lei et al. (2019) explored that technological advancement develops, business machinery becomes more intelligent and integrated, that improves the composition layout of industrial equipment. Simultaneously, this modification also significantly escalates the challenge of fault detection, rendering the disassembling fault diagnostic approach impossible to implement in the present scenario. There are now two different approaches to defect diagnosis: the first uses techniques like mechanisms examination, signaling evaluation, and consequently; the second uses machine learning and additional techniques. There are several drawbacks to the former technique when it comes to fault identification because it relies heavily on past information. For the latter, deep learning and other techniques serve primarily to diagnose faults. The above diagnostic technique requires no prior exposure at all, or at least minimal previous understanding. Time-frequency analysis and process analysis are the primary methods used in the first diagnostic approach to examine faults. It used cepstrum technology to obtain bearing defect characteristics, and they were successful in producing a great anti-noise performance.

Xiang et al. (2021) used practical mode decomposition and ambient spectrum analysis to present a novel defect diagnostic technique for bearings that rotate. This technique performed atmospheric spectrum assessment on the components following the decomposition of the experimental mode. They suggested an automated aperture spectrum analysis-based defect diagnostic approach. The second approach to defect diagnostics mostly uses machine learning techniques like SVM and BP artificial intelligence. The fault diagnosis technique integrating DS evidence concepts and general set theory had produced outstanding results in the research of fault detection of specialized equipment. A hybrid framework based on multi-envelope instructional optimization that combines variation mode decomposition and support vector machines was presented for the investigation on containing fault diagnosis, and it could deliver satisfactory results in bearing fault classification. A fault diagnosis model based on period minimal squares support vector machines proposed in response to the inadequate analyzing caused by traditional deep learning computations when analyzing a great deal of bearing data. This model efficiently alleviated the challenge by combining the least squares support vector machine method with the HHL technique in quantum computing. Algorithms for machine learning are also effective in different domains.

1.1.4 Combining of CNN and LSTM

Wang and Yin, (2024) combined a hybrid fault diagnosis model integrating decision frameworks and a fast CNN was suggested and used to investigate transformer in the fault diagnosis. The technique's generalization is shown through experiments. In fault identification, artificial intelligence is making significant strides and reduces the need for previous research partially when used for problem detection. However, when dealing with complex algorithms or substantial quantities of data, the conventional deep learning approach comes with major drawbacks. During 2006, the idea of deep learning emerged. The advancement of information technology resulted in a massive rise in computing resources and the practical use of deep learning in several fields, therefore enhancing people's lives. Deep belief networks (DBN), generative adversarial networks (GAN), recurrent neural networks (RNN), Convolutional neural networks (CNN) constitute typical network topologies in defect diagnostics.

1.3 Problem Statement

The current architecture of fault monitoring conditions and fault detection in wind turbines and other energy systems has undergone certain challenges and difficulties concerning the old existing methodologies. The existing methods typically relied more on crucial features and expert knowledge. Furthermore, these techniques consistently failed to identify the complex long-term relationships that are present in time-series data (Lei et al. 2019). The effectiveness of current methods in defect detection is more limited because they are unable to completely simulate the complex behavior of time-domain signals and sometimes need manual feature engineering. An advanced and novel method that can effectively capture long-term interdependence and autonomously identify classifiable features from multivariate time-series data needs to be developed to minimize these challenges (Yin et al. 2020). Experiments on limited datasets further demonstrate the resilience of the framework, indicating its potential influence on improving the accuracy and efficiency of problem diagnostics in energy systems.

1.4 Research Objectives

The formulated research objectives of the study are as follows:

- To study the various fault detection and related diagnosis of wind turbine using machine learning approach
- To examine the fault detection methods of wind turbine using integrated machine learning approach
- To evaluate the fault diagnosis methods of wind turbine using novel effective machine learning classifier.

1.5 Research Questions

The framed research questions based on the research objectives are as follows:

1. What are the various fault detection and related diagnosis of wind turbine using machine learning approach?

2. How to examine the fault detection methods of wind turbine using integrated machine learning approach?

3. How to evaluate the fault diagnosis methods of wind turbines using novel effective machine learning classifier?

1.6 Organization of the Thesis

The thesis has been divided into four key sections, namely the introduction, literature review, methodology, results and discussion and conclusion. The reason for doing the same lies in the ground that each chapter has a specific significance, and this can be helpful in achieving a more concrete and needed conclusion.

2. Literature Review

Elizondo et al. (2019) considering that wind energy is one of the cleanest energy sources that is growing quickly and provides encouragement for the growth of renewable energy globally, wind power has attracted a lot of interest in the last 10 years. China has monetized the use of green power, increased the employment of renewable energy, and shown its resolve to reach the emission peak and carbon neutrality goals of 2030 and 2060, respectively, to accomplish the targets for emission maximal and carbon neutrality. The country is the catalyst behind the growth of alternative energy globally and places a high value on new energy, particularly wind power generation. As to the data provided by the Global Wind Energy Commission (GWEC), the nation's newly added capacity attained 65.1 GW in 2019. Large-scale wind turbine production and deployment has created enormous prospects for the market economy's growth, but it has also mentioned significant issues pertaining to energy security, cost-effectiveness, and dependability. On the one hand, wind turbines (WTs) frequently fail due to their distant locations, prolonged operation in challenging circumstances, and their capacity to withstand erratic variations on temperatures, speed of the wind, wave motion, and workload.

Zhong et al. (2018) illustrated that the power grid has the largest percentage of WT element defect percentage, preceding a control network and detector. However, the necessity of problem identification is further highlighted by the substantial expense of operation and maintenance (OM) associated with wind turbines. Defect identification and prompt WT maintenance can minimize significant financial losses. Given the justifications, WT defect diagnostic and notification ought to be carried out. The process is advised to use the artificial neural network-based fault diagnostic approach to identify the operational circumstances of the wind turbines (WTs) to limit downtime, lower OM costs, and increase the turbine lifespan. Several worldwide including national professionals as well as scientists released certain effective faulty detection techniques across multiple elements following the development of the fault diagnosis technological age. To carry out real-time defect conclusions, analytically system-based WT failure detection techniques must evaluate and model the system. Such errors are frequently directly connected to WT model settings shown in figure 2.1.

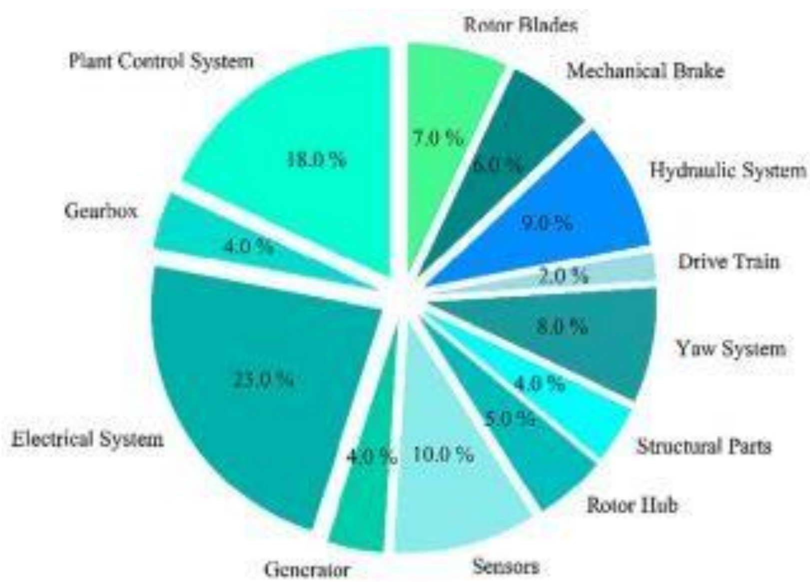

Figure 2.1: Fault Rate of Wind Turbine Component (Source: Zhong et al. 2018)

Cho et al. (2018) distinguished that modeling is used to improve faulty identification precision in accordance with greater knowledge of WT's problem diagnostic procedure. However, the technique of using analytical model-based WT fault diagnosis techniques possesses trouble guaranteeing the preciseness of responsibility evaluation because it models the inner component of the WTs over estimated states and live approximate using structure remainders. Thus, the above procedure is not enough to ensure durability, and unavoidable mistakes and unidentified interfering factors happen. Specialist expertise within wind power-related domains is necessary for expertise-based WT defect diagnostic techniques (Yang et al. 2016). The degree of knowledge and skills of WT fault diagnostic specialists, who possess the capacity for self-learning and acknowledgment, determines how accurate the fault identification findings are. When WT is operating, knowledge-based techniques for diagnosing WT faults are unable to learn new things through the actual instances that are identified in a result, could occurred inadequate accurate diagnosis (Liu et al. 2017).

Data-driven WT fault diagnostic techniques leverage information extraction technologies to uncover essential information that can be used to define the problem and healthy phases of the equipment while requiring previous encounters. This allows for real-time troubleshooting. Both substantial delayed information and updated information are present in the WT supervisory control and data acquisition (SCADA) system. To collect precise fault characteristics and achieve real-time WT fault detection, information mining must be used and analyzed. ML, multidimensional statistical analysis, signal evaluation, and information merging techniques are examples of data-driven WT problem diagnostic techniques (Yin et al. 2014).

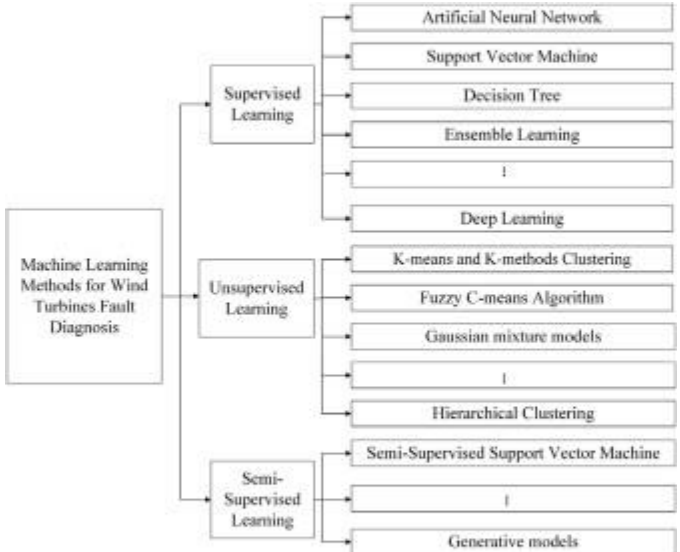

Figure 2.2: ML Method for Wind Turbine Fault Diagnosis (Source: de Azevedo et al. 2016)

The ML-based WT fault diagnostic techniques are categorized under three main categories like supervised, unsupervised, and semi-supervised learning approaches. A thorough assessment of research on the ML-based fault identification technique on machine learning remains lacking, despite the publication in certain studies on WT fault diagnostics and tracking of conditions. Consequently, the current work offers a comprehensive and relevant state-of-the-art evaluation of recent research on machine learning methods and approaches that have been used to WT defect diagnostics. This study analyzes the advantages and disadvantages of current approaches, highlights the difficulties, and makes recommendations for future research directions in the area. It also describes the study methodologies employed in WT fault evaluation as seen in Figure 2.2 (de Azevedo et al. 2016).

2.1 Fault Diagnosis of Wind Turbine

Habibi et al. (2019) stated that the number of nations used to carry out studies regarding WT technology, and the US and Europeans are making improvements towards problem detection and forecasting. For instance, large renewable energy production businesses utilize Siemens' SCADA system extensively. China's wind power sector started later than that of Europe and the United States, yet over the past few decades, study of WT finding faults has accelerated. The techniques for diagnosing WT faults were thoroughly researched because the introduction and growth of ML and AI in the past few years shown in tfigure2.3 depicts the WT structure (Dao et al. 2018).

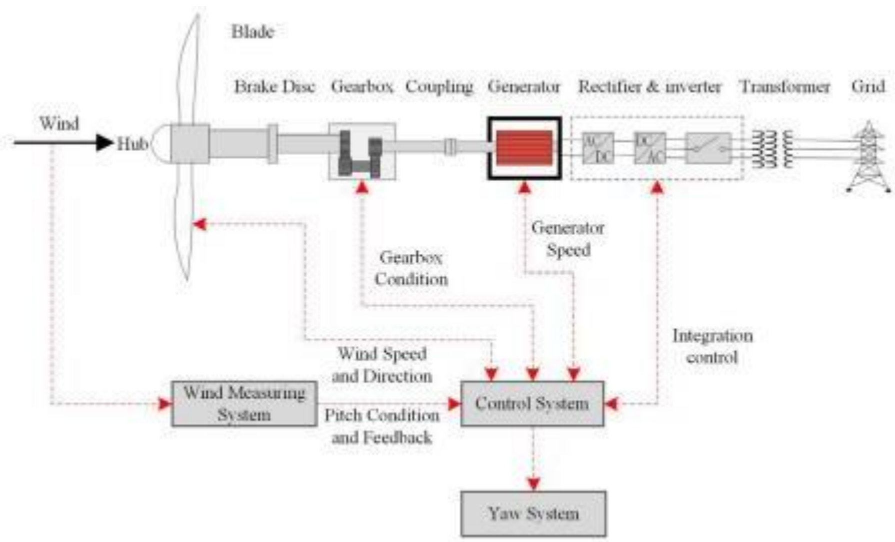

Figure 2.3: Structure of Wind Turbine (Source: Habibi et al. 2019)

Lin et al. (2016) mentioned that the wind wheel, gearbox, generator, converter, horizontal and pitch systems, hydrodynamic and combined control mechanisms, and other systems are the key parts of WT. The wind wheel is essential to WT's converting energy, and WT's effectiveness and security are affected by operating reliability. The failure rate of the wind wheel and other elements rises with increased WT running duration, which has a negative impact on WT's operational efficiency. The conclusion skills are significantly difficult because, on an irregular condition, the incidence components of the WT breakdown at the power source would extend throughout the range proportionate to the movement.

2.2 Fault detection methods of wind turbine using integrated machine learning approach.

The outcomes of space representation impacted by dictionary learning have some limitations and were addressed by Yang et al. (2017). To overcome this, an innovative approach to diagnose faults was proposed by this study. This approach relies on sparse representation and shift-invariant dictionary learning (SIDL) for effectively identifying and generating wind turbines. The outcomes extracted the impulse signal that are directly proportional to the actual signal and obtained the sparse coefficients in accordance with SIDL (Yang et al. 2017).

An important unit in the conversion of WT energy is the converter. This converter allows the output current of WT to traverse to the grid. This output current is having constant amplitude and frequency while traversing. In most cases, high-pressure and elevated temperature working conditions impact the converters due to their nature of low stability. Hence, operating the converter for a prolonged period leads to any irreversible faults in the WT systems. A study conducted by Toubakh and Sayed-Mouchaweh (2016) examined the faults in the converter due to any changes in parameter. The authors suggested an WT converter approach to diagnose faults derived from hybrid dynamic classifier. This approach helped to observe the converters' standard performance in the discrete form in the case of non-functioning parameter. The changes in parameter can be utilized to diagnose faults at the initial stage of the WT converter (Toubakh and Sayed-Mouchaweh, 2016). Another WT converter-based approach for diagnosing faults was proposed by Liang et al. (2020). This approach processed the output voltages' total empirical mode decomposition and obtained a set of implicit mode functions. Along with this, it measured the standard entropy based on the statistical characteristic of implicit model function. Moreover, this approach was performed to diagnose the faults in the fan system where the diagnostic attributes are described using the extracted information. The result of this approach exhibited a greater accuracy rate for diagnosis of 99.57 per cent (Liang et al. 2020).

Another critical element of WT energy is the yaw system. The engine room of wind turbine spins around the centerline of tower with the help of this yaw system. This system maintains the vertical movement of winds and the scanning surface of the wind wheels. Due to the fluctuations in loading functions and in severe operating circumstances, malfunctions in yaw systems arise. This will also impact the WTs' efficiency in generating power. A fault diagnosis approach for the WT yaw system was proposed by Pei et al. (2018) for the qualitative evaluation of the zero-setting errors in the systems. This approach has analyzed the power features under distinct angles in the yaw system to identify the zero point shifting errors. Triggering these zero points shifting errors leads to prompt identification of errors that can enhance the efficiency of WT. However, this can occur only if the pre-determined threshold values are lesser than the measurement error values in the yaw angles (Pei et al. 2018). Another signal analysis-based WT yaw system fault diagnosis approach was proposed by Ouanas et al. (2018) concerning the faults in the yaw system. The yaw drive provides the inverter signal. By filtering this signal, the excessive unnecessary information can be eliminated using the empirical mode decomposition and discrete wavelet transform techniques. The efficiency of the approach has been verified by detecting the faults in the Hilbert transform (Ouanas et al. 2018).

A WT device for controlling the speed and helpful in adjusting any changes in power through modifying the angle of attack (AoA) in the WT blades is called "pitch control system." The complex internal pitch system structure and the external wind occurrences may lead to the collapse of units, damage to blades, and irregular output power. The rate of failures in this system can be high. To overcome this, most of the scholars have suggested approaches to diagnose pitch control system faults. One such study was conducted by Habibi et al. (2017) that used a non-linear model to propose an WT pitch system approach for diagnosing faults. They designed the optimal desired state to address the issue of increasing extraction of energy. For checking the practical aspects of this proposed approach, the researchers performed various experiments. Another study on this topic was performed by Lan et al. (2018). This study used the pitch system's function for indicating faults and the changing phase-by-phase sliding state observer. The WTs' pitch fault can be identified by effectively dealing with the non-linear distribution functions of faults (Lan et al. 2018).

Another crucial element of WT is the hydraulic system. It plays a significant function in the WTs' transmission chain braking, pitch control, and yaw systems and works in any circumstances like high-altitude, open-air, and all-weather situations. However, maintaining this system has become quite challenging since it may lead to spool jamming and oil leakage. A Petri net model-based WT hydraulic system approach for detecting any faults was proposed by Yang et al. (2011). Initially, a model was established using the Petri net theory for every distinct WT hydraulic pitch system's operating state. Then, it leads to build a fault Petri net model. Afterward, the Petri net was calculated along with the fault qualitative analysis results in obtaining the reliability index for the system. This was an effortless process and has an extensive application potential and can be used to the WT hydraulic system to diagnose faults (Yang et al. 2011).

2.2.1 Machine Learning Methods for Wind Turbine Fault Diagnosis

The term "Machine learning (ML)" describes the process by which a computer, without the assistance of any humans, trains an inductive model from a small quantity of data and then utilizes that model to make future choices (Stetco et al. 2019; Clifton et al. 2013). To diagnose the wind turbine faults, ML applications can be used, as argued by Leahy et al. (2016). This consists of objective functions, models, as well as inputs and outputs. The sample data for WT can be given like x = $\{x_1, x_2, x_3, \ldots x_n\}$ (where 'x' denotes the data set having 'n' number of data samples) and 'y' denotes the fault category. The output training sample for this:

$$(\{x_i, y_i\}_1^M \in \{x, y\})$$

The above equation can be used for training the model and obtaining the approximate value 'f(x)' for the actual value 'y.' Also, the total number of training samples is represented by 'M,' the loss function is represented by 'L,' and the interconnection between 'x' and 'y' can be represented by 'y^*.'

$$\begin{aligned} y^* &= argmin_f\left(E_{xy}L\left(y, f\left(x\right)\right)\right) \\ &= argmin_f\left(E_x\left(E_y\left(L\left(y, f\left(x\right)\right)\right)|x\right)\right) \qquad (1) \end{aligned}$$

Calculating the mean value of loss of training set can help identify 'empirical risk.' ML methods need to reduce this empirical risk. Log loss, absolute value, square, and 0-1 functions are some of the loss functions often utilized. ML methods, however, have an issue of overfitting. Hence, reducing structural and empirical risks has become essential. The complexity of the models can be measured using the regular term J(f). Ridge and Lasso regression are commonly used regular terms. The below equation is the final optimized objective function.

$$Obj = min\frac{1}{M}\sum_{i=1}^{M} L\left(y, f\left(x\right)\right) + \lambda J\left(f\right) \qquad (2)$$

According to Lei et al. (2020), ML methods can be categorized into three sub-categories: supervised, unsupervised, and semi-supervised methods. This study also discussed all the categories of ML methods in the context of wind turbine fault diagnosis.

2.3 Supervised Learning Methods for Wind Turbine Fault Diagnosis

According to Schwenker and Trentin (2014) and Zhou (2018), supervised ML method involves modifying classifier parameters from the known class samples for achieving the expected outcomes. The goal of supervised ML method is to provide the computer common mapping principles of input to output by providing it with sample inputs and the necessary outputs.

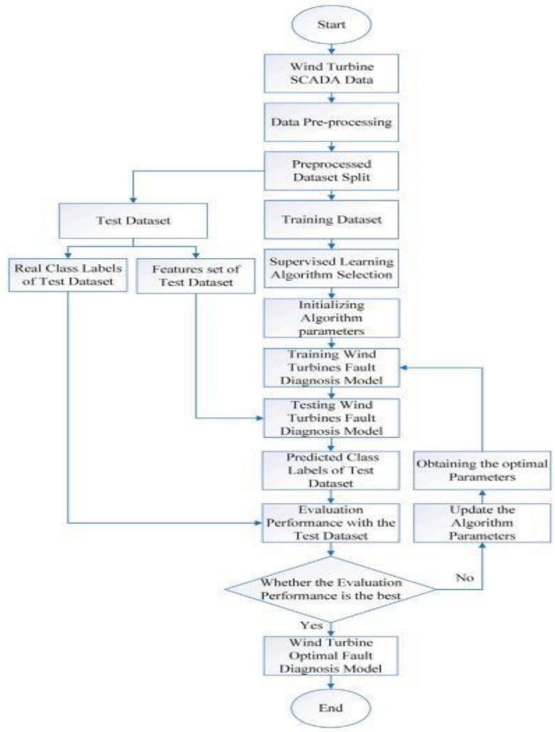

Figure 2.4: Flowchart of WT Fault Diagnosis based on Supervised Learning Method (Source: Jiménez et al. 2019)

This type of ML methods can be extensively applied in diagnosing the faults in the WT. According to Jiménez et al. (2019), supervised ML methods offer unique algorithms based on the type of issues (Figure 2.4). Firstly, the research object taken was fault diagnosis of WT; then, data was obtained from supervisory control and data acquisition (SCADA) systems; afterwards, it divides the data sets for training, evaluation, and test sets; pre-processing of datasets was then performed; and finally, the data was normalized after processing of missing values. Secondly, the dataset was trained by choosing the right ML modeling algorithm. Thirdly, the quality of the model was then evaluated using the test set. Lastly, continuous optimization of WT fault diagnosis leads to produce a precise classification of faults (Jiménez et al. 2019).

2.3.1 Artificial Neural Network

According to Agatonovic-Kustrin and Beresford (2000) and Xi et al. (2020), the most employed supervised ML method is the Artificial Neural Network (ANN). This network contains multiple numbers of neurons and can be classified into three layers: input, output, and hidden layers. As noted by Samanta and Al-Balushi (2003) and Saravanan and Ramachandran (2010), the ANN methods have been extensively used in fault diagnosis. A faulty device can be identified by establishing the relationship mapping between categories and characteristics of faults by learning fault samples which are already known. The ANN-based fault diagnosis WT model involving three layers are shown in figure 2.5. From the figure 7, x_1, x_2, … x_n denotes the WTs' input characteristics, the total WT fault types is denoted by 'm' and the total input characteristics sample is denoted by 'n'.

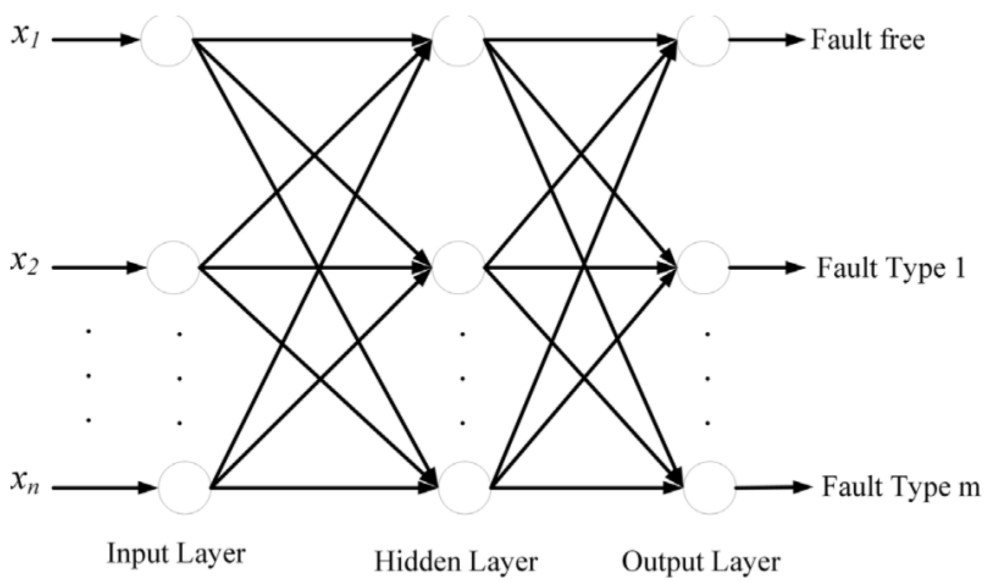

Figure 2.5: ANN-based WT Fault Diagnosis Model (Source: Tang et al. 2021)

Radial basis function (RBF), self-organizing map (SOM) and adaptive resonance theory (ART) networks are the commonly used supervised ANN methods (Tang et al. 2021).

An approach for diagnosing faults in the main bearing of WT based on ANN models was proposed by Zhang and Wang (2014). This study used the data obtained from SCADA systems. The faults in the main bearing of WT have been identified as quick as possible by calculating the variations in the actual and theoretical parameter values (Zhang and Wang, 2014).

A hybrid approach of RBF and ART networks was proposed in another study conducted by Bielecki et al. (2014). This approach produced high efficiency in real-time by reducing the detection time of faults in the WT, monitoring the WTs' current working conditions through Internet, and finding any faults as quick as possible (Bielecki et al. 2014). Nevertheless, ANN algorithms are not able to diagnose faults accurately since not all faulty information can be collected by the wind farm's real engineering. Hence, the SOM model was proposed by Zhao et al. (2015) to diagnose the WT faults. Here, the network is trained with the WTs' data samples in normal state. Based on the output neuron's position in the output layers of ANN, identification of any malfunctions in the WT has become possible. Consequently, the fault diagnosis of WT produces efficient outcomes including strong robustness through this proposed SOM model (Zhao et al. 2015). Nonetheless this SOM model offers strong robustness and high accuracy rate, it needs more time to train the model and various modeling parameters. Although the Wind power industry was established in China only in the later part of 1980s, numerous studies have been conducted under the area of fault diagnosis of WT. On the one hand, there have been insufficient data samples with regards to fault diagnosis in the WT. On the other hand, the fault diagnosis classification accuracy has been severely impacted by the WT data samples' completeness and accuracy. This has been seen as the current key problem that restricts the growth of ANN models in the study area of fault diagnosis of WT (Tang et al. 2021).

2.3.2 Support Vector Machine

Vapnik (1999) established the kernel-based ML technique known as support vector machine (SVM) and can be mostly applied in classification and regression analysis. The primary goal is to increase the margin between two hyperplanes that divide two datasets in a multidimensional space. To construct a decision hyperplane that balances model complexity and empirical risk, reduction of structural risks is ensured by formulating SVM. SVM has been mostly applied in non-linear issues where the isolation margin between positive and negative samples are increased and a classification hyperplane is constructed as a decision plane (Deka, 2014).

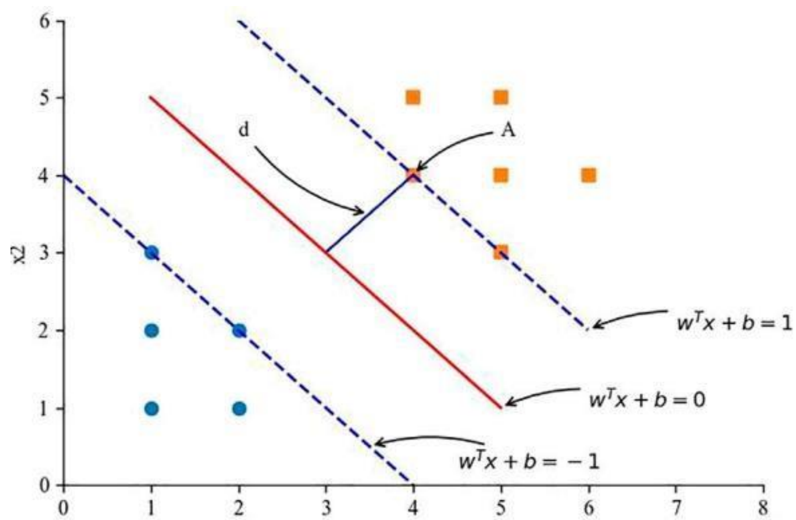

Figure 2.6: Support Vector Machine (Source: Tang et al. 2021)

In figure 2.6, ‘b’ is the constant, and ‘W’ is the normal vector that represents hyperplane using the following formulas: $w^Tx + b = 1$; $w^Tx + b = 0$; $w^Tx + b = -1$.

To separate the data points, a hyperplane needs to be found for SVM algorithms where the minimal geometric distance is the maximum point. A convex quadratic issue with a universal optimal solution can be considered as having a solution process like the SVM. Hence, SVM can be extensively utilized in the realm of diagnosing faults. Raw data is mapped to a high-dimensional space using the inner product kernel function in SVM to create linear data. Nevertheless, it has been quite challenging to model big data in WT. The precision rate of fault diagnosis can also be impacted by choosing the kernel parameters. Additionally, it is challenging to ensure that multi-type fault diagnosis of WT (Tang et al. 2021).

A manifold learning and Shannon wavelet SVM-based fault diagnosis approach for WT was proposed by Tang et al. (2014). A feature set with high dimensions was created to extract the mixed-domain features. This created feature set was then compressed using manifold learning into an eigenvector with low dimensions. Shannon wavelet SVM uses this eigenvector as input for diagnosing the faults in the WT gearbox (Tang et al. 2014). Another study used multi-kernel SVM, and a random subspace identification-based WT fault diagnosis approach was proposed for optimizing the individual kernel parameter by Zhao et al. (2018). This proposed approach outperformed other conventional SVM in diagnosing the WT faults. In another study, multi-SVM algorithms were utilized for diagnosing the renewable energy power grid faults that exhibited a greater precision rate (Liu et al. 2020).

2.3.3 Decision Tree

According to Safavian and Landgrebe (1991), a decision tree (DT) has been made up of several judgment nodes and the tree structure creates a model for regression or classification. A simple DT-based fault diagnosis WT model is shown in figure 2.7.

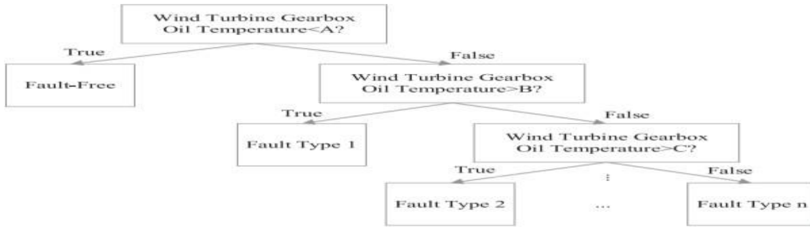

Figure 2.7: WT Fault Diagnosis Method based on Decision Tree (Source: Tang et al. 2021)

A DT algorithm-based fault diagnosis approach for DT connected with grids was proposed by Benkercha and Moulahoum (2018). This approach yielded a greater accuracy rate and prediction performance. Another DT-based approach to diagnosing faults in WT was proposed by Abdallah et al. (2018). This study used trained and integrated DT classifiers, and constant and high rate of data samples collected from WT. While DT algorithms have been easier to adopt compared to other ML algorithms, it is less effective at handling missing values. More number of fault-free type samples and a smaller number of fault-type samples can be seen in the DT-based fault diagnosis approach of WT. When data handling that has varying sizes in samples across diverse areas, DT is likely to select features with extra numerical values for information gain. This leads to overfitting and limited usages in diagnosing the faults in WT (Tang et al. 2021).

2.3.4 Ensemble Learning

According to Polikar (2012) and Liu et al. (2019), as ensemble members, adjusting and training several base learners to become strong learners is the fundamental principle of ensemble learning. The mean performance of resultant learners is high when compared with other ensemble members. Subsequently, loss function has been optimized for establishing a model and increase the fault classification performance. Boosting and bagging ensemble methods are commonly used ensemble learning methods (Breiman, 1996).

The bagging ensemble method, also known as bootstrap aggregating, has been mostly used for statistical and regression classification. In this algorithm, the samples are returned to obtain a new set of data and for each dataset, the base learners are efficiently trained or combined. Overfitting issues have been prevented and variance has decreased in this algorithm. Random forest (RF) is an example algorithm of the bagging algorithm (Breiman, 1996). An RF-based WT gearbox fault diagnosis approach was proposed by Cabrera et al. (2015). Firstly, the decomposition of the wavelet packet extracted the vibration signal's conditional variables to serve as classification problems' input features. Afterward, the optimal mother wave set is then identified by approximating the parameter space to choose the most effective feature using the RF classifiers' internal ranking. Finally, WT gearbox faults are diagnosed using the RF algorithm (Cabrera et al. 2015). Another study was conducted to improve the accuracy of diagnosing faults for the WT gearbox. This study used a deep RF fusion (DRFF) approach (Li et al. 2016). The wavelet packet transforms' parameter values are characterized using two deep Boltzmann machines. Moreover, the RF algorithm is used to combine the integrated DRFF model and the two deep Boltzmann machines' output. The outcomes of this study demonstrated that DRFF algorithm outperforms traditional RF in the fault diagnosis for WT gearboxes (Li et al. 2016).

Boosting, on the other hand, modifies the algorithm by assigning a high priority to the faulty classification, leading to considerable enhancements in classification performance (Freund and Schapire, 1996). Overfitting issues in this boosting algorithm can be prevented by focusing on the reduction of bias. LightGBM and XGBoost are some examples of boosting algorithms. An XGBoost- and RF-based approach for the effective diagnosis of faults was proposed by Zhang et al. (2018). Using XGBoost, the ensemble classifiers are trained for every individual fault by utilizing the features that rank highest after RF algorithm ranks them based on significance. This proposed approach prevents the issue of overfitting. The validity of this approach was verified through experiments. Adaptive LightGBM-based approach for improving the accuracy rate in diagnosing faults for the WT gearbox was suggested by Tang et al. (2020). High coefficient information is used to analyze the correlation of the sample data of WT for accomplishing the selection of features in the detection of faults. Concurrently, the WT gearbox fault diagnosis approach employed Bayesian optimization and LightGBM methods. The results of this study indicated that the proposed approach exhibits a low missing detection rate and false alarm rate (Tang et al. 2020).

Ensemble learning has been extensively applied in many areas for its greater ability to diagnose faults as promptly as possible. Nevertheless, certain ensemble learning methods present over-fitting issues, convergence speed is slow, and in some cases, weak learners need to be dependent on each other. Hence, some factors must be considered while employing ensemble learning algorithms, such as weights, number of base learners and iterations (Tang et al. 2021).

2.4 Unsupervised Learning Methods for Wind Turbine Fault Diagnosis

Unsupervised training was the process by which a computer learns unprocessed data to uncover the underlying structure, clarify the salient characteristics, and categorize the data clustering is a representative approach. Several commonly used unsupervised learning methods are derived from the clustering technique. Unsupervised teaching educated the data of unlabeled specimens instead of analyzing class-labeled examples like supervised learning about the approaches. They reduced the interclass deviation from the participant collection organized based on sample resemblance which establishes the model. Unsupervised learning techniques have become popular throughout troubleshooting, data extraction, computer graphics, and various other fields. They possessed the ability of classifying and predicting data from trials by uncovering concealed patterns and connections within the information collection. Unsupervised learning usually used a variety of techniques, including the Gaussian combination model, fuzzy C-means (FCM), hierarchical clustering technique, K-means algorithm, and others (Zhang and Zhou 2018a).

2.4.1 K-Means

Wang et al. (2016) studied that the K-means clustering is a directly unsupervised learning technique that divided n data points into K clusters, about every result belonging to the cluster that has the closest mean. With the goal recognize the K beginning locations from the information as the grouping center, and they first initialize the centers of clusters. Next, then compute the amount of time between all points and the exact center and allocate it to the closest cluster. Third, to reduce the internal total, they compute the points of the cluster. Lastly, until the focal points of every cluster remain unchanged, the distribution and updating procedures are repeated. K-means clustering is finished if each point is assigned to an identical group that previously. For instance, the K-means method was applied to the WT data collection to try to arrange the information. The five various kinds of outcomes from clustering are displayed in Figure 2.9.

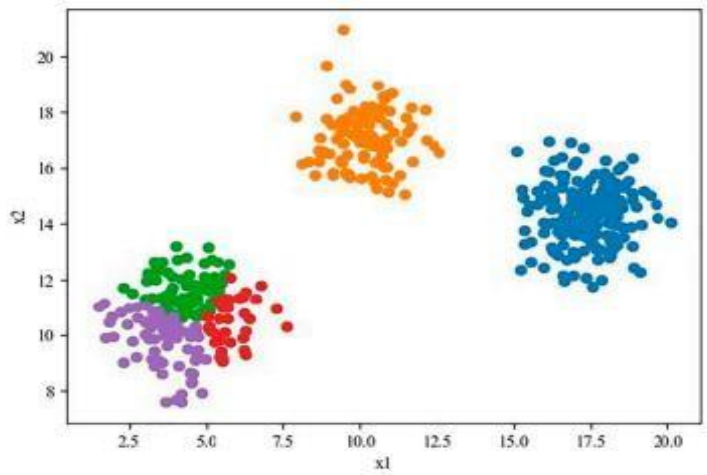

Figure 2.9: K-Means Algorithm (Source: Wang et al. 2016)

Yesilbudak (2016) introduced a K-means clustering technique for rolled bearings defect diagnostics, which overcomes K-means' sensitivity to the choice of starting cluster centers. Characteristics acquired using simulation waveforms are used to pick the beginning locations. Investigations on three diverse types of bearings for identifying flaws demonstrate the effectiveness of this technology in classifying defects of unsupervised learning procedure developed using kernel technology. It can distinguish between various non-linear processing options, efficiently identify errors, and lower the false alarm rate.

2.4.2 Fuzzy C-Means

Hu et al. (2019) defined that the information using the FCM method could represent a part of more than one cluster. The fundamental principle was to minimize the connection between different clusters and enhance the similarities across items assigned to an identical cluster. They presented a fault diagnostic approach based on global local mean deconstruction and FCM clustering in the WT transmissions failure identification. This approach provides simplistic execution as well as producing excellent diagnostic findings. The identified sample was grouped using the FCM aggregation, and the experimental population was determined and categorized. The WT fault diagnostic techniques need monitoring as well as instruction based on knowing fault specimens from the past, but gathering known fault specimens is costly and takes time. The approach based on kernel FCM (KFCM) clustering used for the defect identification of the WT gearbox, considering the incomplete features of identified specimens in WT. The specimens containing identified patterns are categorized using the KFCM grouping technique, and the categorization concentrate with each recorded defect is found. To determine if the newly collected data specimens are part of the known defects, similarity characteristics are also computed. This technique can precisely and successfully diagnose both recognized and undiscovered WT problems.

2.4.3 Hierarchical Clustering

Liu and Ge, (2018) learned that unsupervised training, hierarchy clustering is a statistical technique that creates a model by organizing groupings into a hierarchical structure. A "tree diagram," or a tree's framework with branches and leaves, should be used to illustrate the sequential clustering approach. When it comes to clustering types of trees, the base nodes of the group are at the highest point of the tree, while the initial informational points of various groups are at the lowest level. As seen in Figure 2.10, the hierarchy clustering approach consists of two processes: divisive top-down clustering, which starts at the base and repeatedly separates the groups, and agglomerative hierarchy groupings which starts at the branches and gradually integrates clusters.

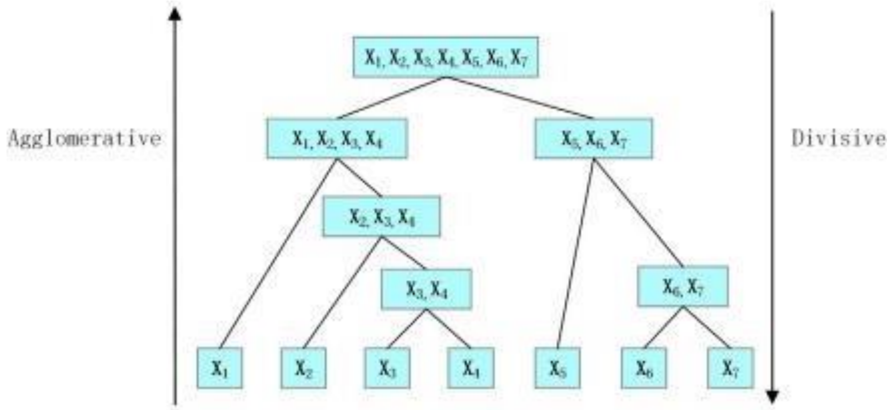

Figure 2.10: Hierarchy Clustering Structure (Source: Liu and Ge, 2018)

The separation among the information locations for multiple groups can be determined using the distances defined by Euclid using the hierarchy of clustering approach. To detect various transmission illnesses, we presented an error evaluation technique that utilized flexible multiple habitats structural adjustments and enhanced organizational entropy. They also used the hierarchy collection technique to lessen sound in the problem characteristics obtained from the signal. They introduced an ordered cluster selection-based weighted random woodland approach for identifying faults in intricate manufacturing processes. Continuous fault recognition could be made simpler by using the pyramidal clustering approach for online picking models (Lin et al. 2020).

2.4.4 Gaussian Mixture Model

Zhang and Dimitrov (2023) stated that the Gaussian Mixture Model (GMM) is derived from a combination of limited amount of likelihood models with unidentified variables and assumes that every data point followed a Gaussian pattern. With the goal to execute the allocation of data, GMM could be defined simply as a method of matching an ordered set of many distributional dispersion models. To identify WT gearbox breakdowns and determine the adverse log-likelihood of a gearbox's bearings vibrating output section that reflects a properly functioning gearbox. They presented a Gaussian-based hybrid model and non-stationary windmill gearing noise signals should be handled using this technique. In the multi-parameter medical evaluation system that takes dynamically operating circumstances into account, considering the extremely complicated and unpredictable operating characteristics of WT.

The operating state of WT gets assessed after the pick of characteristic values and GMM-based multi-regime simulation, which can detect WT problems. They presented a fault diagnostic technique for WT that utilized a GMM randomized variable framework in reaction to recurrent WT failures. Using an GMM-based randomized variable structure, the movement impulses of WT that vary periodically depending on surroundings and operational circumstances are retrieved, and the predictive parameters are established. The most problem levels and kinds were accurately identified by the approach offering notable performance gains. The enormous number of WT information samples could be handled by GMM with effectiveness, and its sluggish resolution and high computation cost are drawbacks (Avendaio-Valencia and Fassois 2017).

2.5 Semi-supervised Learning Methods for Wind Turbine Fault Diagnosis

Zhang and Zhou, (2018b) defined the two types of learning that semi-supervised training was a method of instruction that find the certain commonalities between unlabeled and labeled specimens with the goal to assist identify the features of the predictive model and transfer declared from marked to raw data. By mixing labeled and unidentified specimens throughout the information set selection procedure could improve the precision of training. The semi-supervised Support Vector Machine (SVM), generative model-based, disagreement-based, and graph-based approaches are the four primary concepts for semi-supervised training. They worked on co-training and laid the foundation for disagreement-based semi-supervised instruction that is mostly employed in human-computer interaction and has less impact by the non-convexity of the lost function and the volume of information.

2.5.1 S3VM

Liu et al. (2020b) presented that the SVM was created for partially supervised learning or S3VM. Marking samples without labeling to optimize the gap after the division of the hyper-plane is the main concept of S3VM. A Transductive Support Vector Machine (TSVM) is the most often utilized S3VM. The strategy's main concept may be defined by five phases using data labeled, an SVM classifier trained in the initial stage. In the second stage, unlabeled data classification outcomes are projected with Support Vector Machines (SVM) using the preexisting labeled sample and unidentified instance to retrain SVM. In the third phase, it seeks to identify the opposing labeling in the projected raw data which would be inaccurate for the marked samples to switch the label.

Mao et al. (2020) repeated the subsequent phases in the final phase unless the best S3VM classification is found. The unlabeled samples must be tagged, and the classification outcomes predicted utilizing the S3VM algorithm in the fifth phase. The discipline of WT defect diagnostics makes a great deal of the S3VM methodologies. Roller bearings defect diagnostic technique based on S3VM technique uses a small number of labeled data to create a representation with an excellent classification performance. They employed semi-supervised construction, presented a digital approach to facilitate initial identification of flaws of bushings with a lower incidence of false alarms and better discriminatory capacity towards beginning fault characteristics.

2.5.2 Generative Models

Li et al. (2018) explained the method of tracking were capable of the desired outcomes for problem identification and mitigation with this approach. They developed a semi-parametric technique for asynchronous motor defect diagnostic and identification that employs the hidden Markov model to accomplish automated identifications. They followed the training phase of each probabilistic (non-parametric stage) hidden Markov model, which forms the basis of the hidden Markov model classifiers (parameter stage). In the two matrix sets are computed using statistical inference to address the efficiency issue in the fault classification procedure. They anticipated premature transmission component mistakes, introduced a fault detection technique based on a hidden semi-Markov modeling and a multivariate Bayesian control strategy (Xin et al. 2018).

3. Methodology

This chapter was focused on providing an actual implementation method of a recommended technique to implement suitable machine learning approaches in predicting the faults in wind turbines. It provides an implementation strategy with detailed steps that cover a wide range of subjects like study design, data collection, feature selection, machine learning methods, model building, and evaluation metrics. This chapter provides an implementation of the chosen machine learning model in addition to highlighting practical issues and challenges that come up in the execution phase to get reliable and accurate predictions in fault types.

3.1 Research Design

To determine the association between multiple fault modes and the prevalence of fault diagnosis, retrospective methodology was used in this study to investigate the trends and associated fault components with fault detection. However, it used the already available dataset from the SCADA system, including Scada_data, Fault_data, and Status_data. It was a perfect fit to adopt the retrospective method for the study goals of predicting the faults in wind turbines (Wen et al. 2022). With the examination of large data, long-term trends can be explored based on large data for analysis. Moreover, this research design would be beneficial in terms of money- and time-efficiency since it does away with the requirement for lengthy follow-up periods included in prospective studies.

In this study, the available prior data may be effectively useful in creating precise models for the prognostication of fault detection based on the integrated retrospective research approach. It could be possible to evaluate different fault modes that improve the understanding of the diagnosis of faults and provide the information for fault diagnosis measures. The retrospective design methodology was aligned with the research objectives that allowed the data collection and analysis of large datasets with the potential detection of faults in wind turbine systems.

3.2 Data Collection

The method incorporated in predicting faults for wind turbines would be impacted by how data collection was performed. This section provided a detailed explanation of the data sources used in the research and the criteria of sample size. The SCADA system data was used in this study due to the comprehension, availability, and pertinence of large data sets, including fault components, fault modes, and their status information (Letzgus, 2015). Moreover, the sample size of the dataset should be large enough to provide the necessary statistical power for accurate findings. The presumptions on the information variability, impact of size, and confidence levels were required to be considered in evaluating the sample size. The large sample size of data should ensure that the statistical analysis would provide important results based on sample estimates and power evaluation.

3.3 Feature Selection

It is crucial to implement the prediction of fault detection and diagnosis models for wind turbine systems. The features should be chosen to find out the strongest features that connect the prediction of fault modes. Initially, we sorted out which attributes were most significant for this study based on earlier studies and the data that was already available. Accordingly, the chosen attributes were considered in the dataset. To reduce the extraneous attributes, Principal Component Analysis (PCA) was used as a popular method for feature extraction. The approaches like wrappers, filters, and embedded programmers have used methods. Before feature selection, preprocessing of data was performed to ensure the dataset’s integrity and consistency. Moreover, the model can predict the most valuable metrics for prediction of fault detection and diagnosis through the careful selection of key features and filtering. This study used the PCA method as a feature selection to determine the significant traits (Wu et al. 2022). However, preprocessing involved feature extraction or selection, and cleaning of the data which served as resources for model building and evaluation further (Figure 3.1). Accordingly, it is essential to choose the most helpful features and ensure the accuracy of the used data in the optimization method for enhancing the prediction ability and the understanding of fault detection and diagnosis.

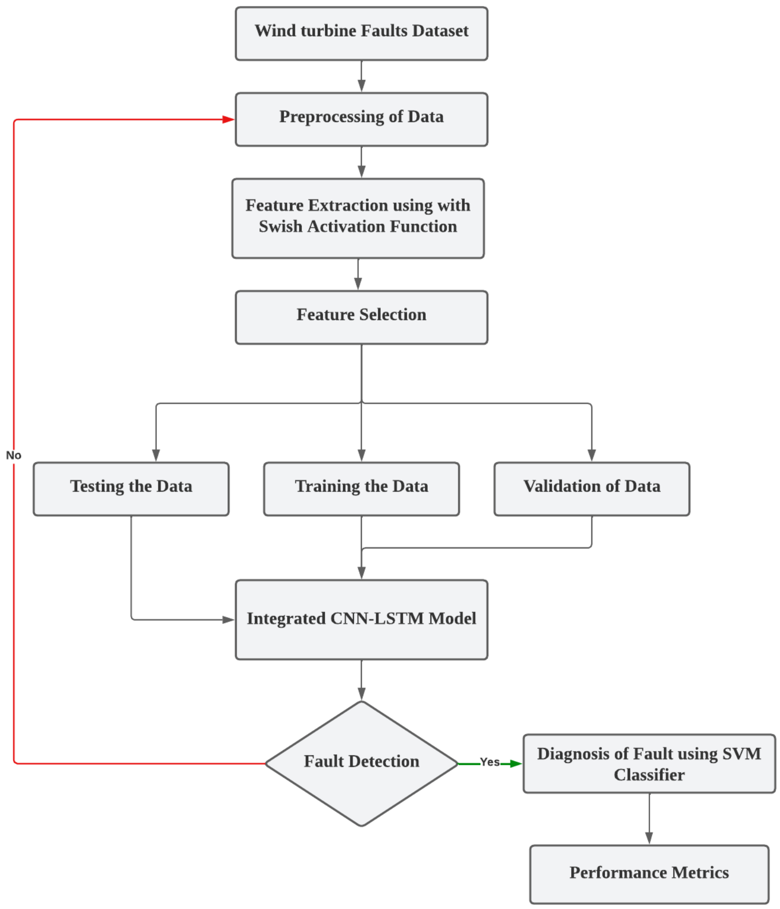

Figure 3.1: Proposed Methodology (Source: Author)

3.4 Machine Learning Approaches

With the careful selection and implementation of efficient machine learning approaches, we could design a prediction model for fault detection and diagnosis in wind turbine systems. This section discussed detailed information about employed machine learning algorithms in the study in addition to the rationale and how they achieved the proposed objectives. Several types of supervised and unsupervised learning algorithms have been used in energy systems that provide unique advantages and drawbacks based on multiple attributes and research questions. The Long Short-Term Memory (LSTM) is a type of unsupervised learning that was often used or widely used in energy systems to address faults and diagnosis methods (Appiah et al. 2019). It was highly beneficial to examine the attributes of fault detection that predict the probability and development of wind turbine systems. Moreover, convolutional neural networks (CNN) also demonstrated efficacy in performing the classification tasks that involve both nonlinear and linear cases under a supervised learning approach (Rahman et al. 2021). These models could be useful in managing large-scale datasets successfully by providing accurate performances. Therefore, the LSTM and CNN models were chosen in this study due to their benefits in handling linear and non-linear data in the wind turbine energy systems.

3.5 Deep Learning

LeCun et al. (2015) initially proposed the concept of "deep learning (DL)". DL is the subset of ML algorithms, and the structure follows a multi-level deep network that uses a specific training algorithm with its own data samples. For the identification of distributed data feature representation, high-level abstract representation is formed by collectively combining the low-level feature representation (LeCun et al. 2015). Many fields broadly employ these DL algorithms such as fault diagnosis, data mining, image processing, etc. (Helbing and Ritter, 2018; Goodfellow et al. 2016; Schmidhuber, 2015). Convolutional neural networks (CNN), deep auto-encoder networks (DAE), and deep belief networks (DBN) are some of the common DL algorithms (Jiang et al. 2018).

An innovative fault diagnosis approach for WT gearbox was proposed by Qin et al. (2018). This approach employed improved logical Sigmoid unit and DBN algorithm. This combined approach increased the diagnosis accuracy rate by extracting input signals' impulse components. Soft-thresholding, kurtosis index and optimized Morlet wavelet transform were used in this approach.

This proposed approach outperformed the conventional Sigmoid approach (Qin et al. 2018). A DAE-based DL approach for WT fault and anomaly diagnosis was proposed by Zhao et al. (2018). In this, SCADA system was used to collect data and a deep automatic encoder network model was formed by Boltzmann machine. The residual of this formed model helps in prompt diagnosis of WT faults by identifying its physical location. Due to the WT's changing operational states and significant interference of noise, the precision rate of WT fault diagnosis has been reduced (Zhao et al. 2018). To overcome such issues, a concurrent CNN-based WT fault diagnosis approach was proposed by Chang et al. (2020). No prior information is needed for the raw WT data; instead, they follow adaptive learning. This facilitates greater ability for generalization and precision rates. The fault diagnosis of WT offers improved precision rate since various branches' convolutional layers choose kernels with identical levels and at varied sizes (Chang et al. 2020). Despite its greater accuracy of fault diagnosis and learning ability, DL algorithms need a lot of processing power and data, which is expensive and requires a lot of hardware. Hence, addressing these drawbacks has been the primary concern, as argued by Tang et al. (2021).

3.5.1 Convolution neural network

As noted by LeCun et al. (1989), the convolutional neural network (CNN) applications have been widely used since its establishment in 1989. It is a type of DL algorithm. The faults can be classified, and data features are extracted efficiently in this CNN algorithm. A hybrid approach based on CNN-MLP applications for diagnosing the faults in the rolling bearing was proposed by Sinitsin et al. (2022). This approach used combined input data for performing the rolling bearing fault diagnosis. Another CNN-based feature learning approach to monitor states in the rotating instrument issues was proposed by Janssens et al. (2016). This proposed approach outperformed the traditional random forest classifiers and feature selection algorithms. According to Guo et al. (2023), dedicated equipment has been extensively applied in many kinds of vehicles and is considered essential to control the regular functioning of special vehicles. Conventional ML algorithms have been mostly used by former researchers since the acquisition of data has been difficult for dedicated equipment. Hence, a simulation platform was used to obtain the fault data and diagnose it using the DL algorithm in the study conducted by Guo et al. (2023). CNN-LSTM-based fault diagnosis approach was proposed by the researchers. The spatial-temporal data characteristics were retrieved by integrating conventional CNN and modern LSTM networks. Moreover, the ability to extract key features was improved by using the CBAM. Data from the hardware-inloop dedicated equipment simulation application was used to train and verify the model. The classification techniques of fault and the proposed approaches' parameters were verified, and it has been established that this proposed approach possesses effective classification of fault diagnosis of dedicated equipment. Moreover, this approach outperforms the other different CNN models (Guo et al. 2023).

CNN is made up of three layers: activation, pooling, and convolution layers. Of these, convolution layer acts as a core layer of the CNN. Inside the convolution layer, there exists a convolution operation. The process of extracting feature information from the data has been made possible through this convolution operation that acts as the inner product of discrete filter matrix and data parts. Different data features can be extracted using different convolution kernels and impacted by the kernel size. As a result, both are essential to guaranteeing the normal operation of the convolution layer (Guo et al. 2023). The below formula is used to measure convolution operation: γ.

$$x_j^{\gamma} = f\left(\sum_{i=0}^{C-1} x_i^{\gamma-1} * k_{ij}^{\gamma} + b_j^{\gamma}\right)$$

where, the neural network layer count is represented by 'γ', the bias matrix is represented by b, the respective convolution kernel's weight matrix is represented by k, number of input kernels data is represented by 'C', the data element of the γ-1 layer is x, the jth eigenmatrix of the current layer is represented by x, and the activation function is f(x).

Downsampling function can be implemented by pooling layer by modifying the input data filter and selecting the mean or highest sliding window data value as the pooling unit's output. The decision of whether to transmit signals to following neurons and what can be included in it has been determined by the activation function. The most common activation functions include ReLU, Tanh, and Sigmoid functions (Guo et al. 2023). γj

Sigmoid

$$f(x) = \frac{1}{1+\mathrm{e}^{-x}}$$

Tanh

$$f(x) = \frac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x}+\mathrm{e}^{-x}} = \frac{2}{1+\mathrm{e}^{-2x}} - 1$$

ReLU

$$f(x) = \begin{cases} \max(0,x) & ,x \geq 0 \\ 0 & ,x < 0 \end{cases}$$

3.5.2 Long-Short Term Memory

One type of RNN algorithm that has been useful to process long-time data series is called Long-Short Term Memory (LSTM). Cell state and gate mechanism are added in the LSTM algorithm to efficiently handle the information for a long-term and outperforms the other traditional RNN algorithms (Guo et al. 2023). The LSTM architecture is shown in figure 2.8.

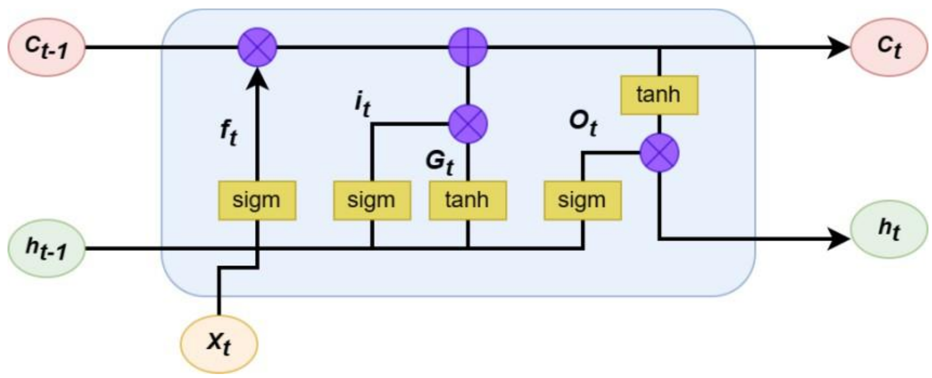

Figure 2.8: LSTM Architecture (Source : Guo et al. 2023)

LSTM stores more essential information for a longer period by adding cell state 'C_t.' A change in input can adjust this information in a dynamic manner. The gate mechanism executed the operation to process information. Three gates are present as shown in figure 10, including output gate, input gate, and the memory gate from right to left.

The below formula can be used to calculate the forgetting door 'f_t':

$$f_t = \text{sigmoid}\ (w_f * [h_{t-1}, x_t] + b_f)$$

where, the current input is represented by 'x_t,' the previous unit output is represented by 'h_{t-1}', bias vector is represented by 'b_f', and the weight matrix is represented by 'w_f'.

The following formulas are used in the first stage:

$$i_t = \text{sigmoid}\ (w_i * [h_{t-1}, x_t] + b_i)$$

$$G_t = \tanh\ (w_g * [h_{t-1}, x_t] + b_g)$$

where, the bias vector is represented by 'b_i,' and 'b_s' and the weight matrix are represented by 'w_g,' and 'w_i.' Choosing which information is to retain has been the aim of this first stage.

Updating the cell state is the purpose of the second stage. The following formula can be used for this:

$$C_t = C_{t-1} * f_t + i_t * G_t$$

The following two formulas can be used for the output gate:

$$O_t = \text{sigmoid} (w_0 * [h_{t-1}, x_t] + b_0)$$

$$h_t = O * \tanh (C_t)$$

where the bias vector is represented by ‘b_0’, and the weight matrix is represented by ‘w_0’.

It was argued that other than the research in fault diagnosis, LSTM offers ideal solution for dependency issue between unique words in NLP (Natural Language Processing) due to its nature of effective processing of information sequence for a long time (Guo et al. 2023).

3.6 Model Development

It is crucial to develop an accurate model for fault detection and diagnosis of wind turbines. For that, the Scada and Fault datasets were combined since they have suitable attributes for our study like fault components and fault modes. Then, the dataset was categorized into training, validation, and test sets. Based on the training set, the characteristics of input data were given and the corresponding target variables that indicate the existence or absence of faults. The model's hyperparameters were adjusted in the validation process and the results were evaluated throughout learning.

Based on optimization methods like random gradient descent or linear gradient descent, the difference between the two models was adjusted frequently using the weights of a model. Moreover, accurate predictions can be provided by integrating two models CNN and LSTM

(Figure 3.1). The efficacy of a model was evaluated at distinct phases based on different metrics like accuracy, prediction, and sensitivity (Duchesne et al. 2020). These measures were useful to find out whether the model is accurate enough in the classification of wind turbine systems with and without faults. The chosen models were trained, and their effectiveness was evaluated based on metrics like accuracy, precision, recall, and F1-score. This stage was used to establish the foundation for a thorough examination and understanding of the predictive power of proposed machine learning models in fault detection and diagnosis of wind turbine systems.

4. Results

As discussed above, three models were chosen to provide the best possible machine learning classifier for evaluating the fault diagnosis of wind turbines. However, deep learning and supervised machine learning models like CNN, LSTM, and integrated LSTM-CNN for fault detection methods were used in this study. They included AF (Air-cooling Fault), EF (Excitation Fault), FF (Feeding Fault), GF (Generator Heating Fault), MF (Mains-Failure Fault), and NF (No Fault). In this chapter, data findings and results were analyzed using Python based on precision, recall, and f1-scores to determine the novel effective machine learning approach of integrated LSTM-CNN to predict the fault detection of wind turbines. In addition, model development, selection of different machine learning algorithms, and statistical analysis were performed using evaluation metrics to build the most resilient models in the detection of fault diagnosis methods of wind turbines.

4.1 Model Building

When choosing specific best classifiers, we should be concerned about their distinctiveness and evaluate the performance of individual models. An integrated CNN-based LSTM deep learning model has been used as a classifier to predict fault detection and achieve the diagnosis of wind turbines. Three different datasets, including scada_data, fault_data, and status_data were used to evaluate the measures for fault detection accuracy based on the deep learning approach of the CNN-LSTM model. These datasets included information on more than 60 wind turbine components recorded by SCADA, fault types of wind turbines, and status of wind turbines with descriptions, respectively. However, six different faults were identified such as AF (Air-cooling

Fault), EF (Excitation Fault), FF (Feeding Fault), GF (Generator Heating Fault), MF (Mains-Failure Fault), and NF (No Fault) in wind turbines. The chosen machine learning approach was implemented to test the accuracy, precision, and recall scores for these six fault types and evaluate the diagnosis methods of faults in wind turbines.

4.2 Time-Series Analysis

In the analysis of time series, the given data plot configures the size of the figure with 10*4 and the y-axis was set to the limit between 0 and 4. The data status, spanning from 2014 January to 2016 January, exhibits a trend of consistent values ranging from -0.0 to 0.4 (Figure 4.1). In the period from 2014 April and 2015 April, the data of Scada displayed a range of narrower that fluctuated between -3.0 to 3.0 which suggests various contexts of operations compared to the data status. The data of Scada notably showed a shift of significance around 2015 January, which indicates the change in the potential dynamics of the system. The fault data concurrently cover 2014 June to 2015 January, demonstrating the constrained range of -1.0 to 1., with a correlation of possibility between the occurrences of the fault and the anomalies observed in the Scada data. This analysis of interaction underscores the nuanced need to understand the interplay of the status, Scada, and fault data to unravel the patterns and their causal potential relationships. More investigation into the time intervals and their operations in corresponding conditions enhances the warranted reliability of the system and performance.

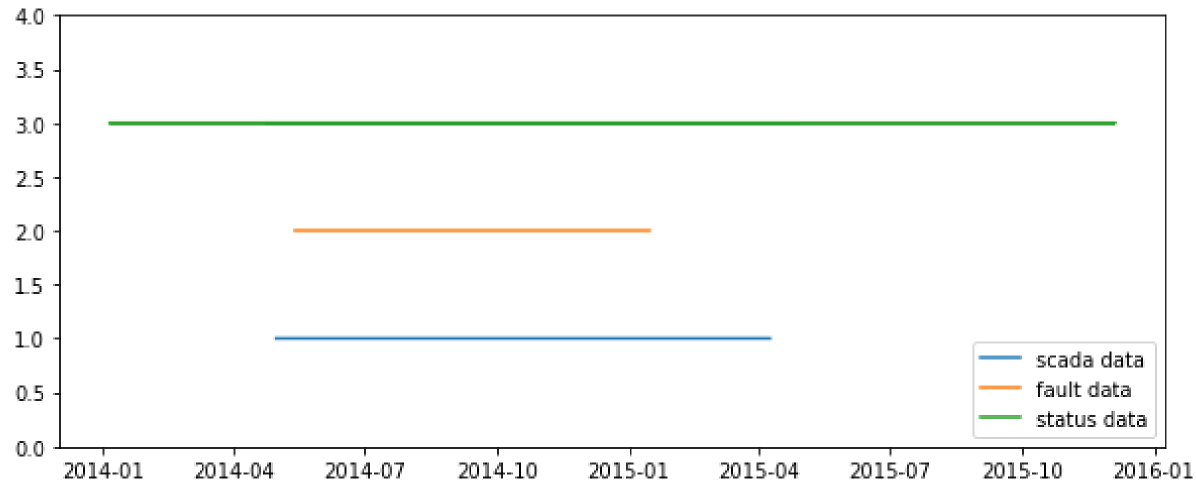

Figure 4.1: Time-Series Analysis of Scada, Fault, and Status Data

The time series of all three datasets were compared to determine the fluctuations (Figure 4.1) from Jan 2014 to Jan 2016. It depicted the constant time series data for Scada, fault, and status data sets. Therefore, these results indicated that there were no fluctuations or changes involved in information about wind turbine components, fault types of wind turbines, and the status of wind turbines over a period from Jan 2014 to Jan 2016.

4.2.1 Maximum Power

Scada_data was evaluated for the time series of WEC (Wave Energy Converter): maximum power (Figure 4.2) over a period between May 2014 to Apr 2015. In contrast to the results mentioned in Figure 4.1, the time-series analysis for WEC: max. power showed that the peak value in March 2015 while the lowest values of max. power were observed in May 2014 and April 2015.

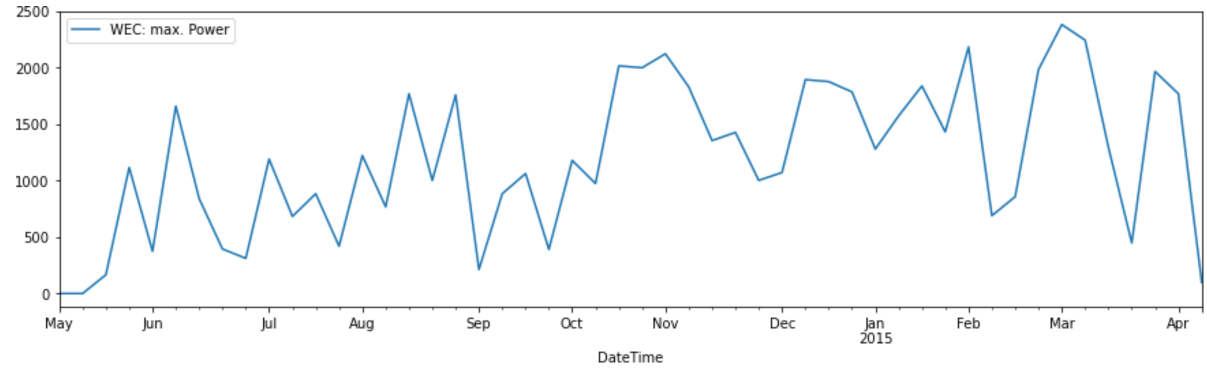

Figure 4.2: Time-Series Analysis of WEC: Max Power from Scada Data

4.2.2 Production

Scada_data was evaluated for the time series of WEC: production (kWh) (Figure 4.3) over a period between May 2014 to Apr 2015. Contrary to the results mentioned in Figure 4.1, the time-series analysis for WEC: production showed an increasing trend and resulted in the peak value in April 2015 while the lowest production was observed in May 2014.

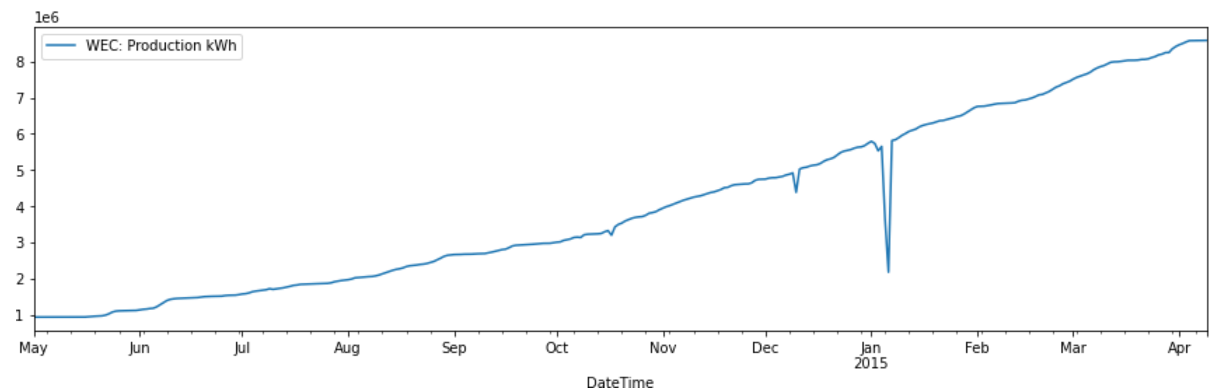

Figure 4.2: Time-Series Analysis of WEC: Production from Scada

To determine the fault predictions for different fault components, we combined the datasets of Scada and Fault and proceeded to the exploratory data analysis.

4.3 Exploratory Data Analysis

The combined datasets of Scada and Fault were analysed to evaluate the average predictions of faults for different fault components. However, the fault types were AF, EF, FF, GF, MF, and NF against fault components, including rotor temperature 2, stator temperature 1 & 2, Nacelle ambient temperature, Nacelle ambient temperature 1 & 2, Nacelle cabinet temperature, main carrier temperature, rectifier cabinet temperature, fan inverter cabinet temperature, yaw inverter cabinet temperature, ambient temperature, tower temperature, control cabinet temperature, transformer temperature, RTU: ava. Setpoint 1, inverter averages, inverter std. dev, group, and frozen.

Table 4.1: Fault Average Predictions for Different Fault Components

Fault Components	**Faults**					
	AF	**EF**	**FF**	**GF**	**MF**	**NF**
Rotor temp. 2	55.7	99.9	65.9	34.9	52.5	52.6
Stator temp. 1	68.9	101.6	70.4	42.8	66.4	60.7
Stator temp. 2	68.3	100.5	69.7	42.5	65.9	60.2
Nacelle ambient temp. 1	15.6	14.5	11.0	12.1	14.8	12.5
Nacelle ambient temp. 2	15.6	14.3	10.9	11.9	14.7	12.4
Nacelle temperature.	19.3	17.4	14.1	14.1	17.9	16.3
Nacelle cabinet temperature.	22.9	20.7	17.8	19.4	21.0	19.9
Main carrier temperature.	20.5	19.6	15.9	13.3	19.1	16.6
Rectifier cabinet temperature.	32.6	30.4	26.7	30.8	30.0	30.4
Yaw inverter cabinet temperature.	27.7	27.3	22.6	21.1	25.9	24.3
Fan inverter cabinet temperature.	32.0	29.9	26.3	24.9	30.1	28.8
Ambient temperature.	16.8	14.9	11.4	12.6	16.8	13.4
Tower temperature.	28.1	28.9	21.6	16.3	27.2	23.1
Control cabinet temperature.	36.5	38.8	31.6	24.9	35.2	31.8

Transformer temperature.	46.9	64.1	51.1	30.2	45.3	43.9
RTU: ava. Setpoint 1	3042.2	3049.4	3047.5	2973.4	2940.2	2988.2
Inverter averages	31.8	34.2	30.1	23.0	30.3	27.8
Inverter std dev	1.4	1.6	1.6	1.1	1.5	1.9
group	150.0	251.0	297.0	285.5	95.7	1591.3
Frozen	0.1	0.0	0.0	0.0	0.2	0.1

The highest mean value was observed for Excitation Fault (EF) with 99.9 in the fault component of rotor temperature 2 while it was lower generator heating fault (GF) with 34.9 (Table 4.1). Both stator temperature 1 and stator temperature 2 fault components were evident that they recorded the highest mean score for fault type of excitation fault (EF) with 101.6 and 100.5, respectively. Nacelle ambient temperature, Nacelle ambient temperature 1 & 2, and Nacelle cabinet temperature were expressed higher for air cooling fault type (AF) with mean scores of 19.3, 15.6, and 22.9, respectively. They were lower in the case of feeding fault (FF) type. Moreover, main carrier temperature, rectifier cabinet temperature, yaw inverter cabinet temperature, and fan inverter cabinet temperatures also showed the highest mean scores with the air-cooling fault (AF), i.e., 20.5, 32.6, 27.7, and 32, respectively. In addition, these fault components resulted in the lowest mean scores for generator heating fault (GF) except for rectifier cabinet temperature which expressed at lower conditions in case of feeding fault (FF). The ambient temperature was higher on average for AF (air cooling fault) and MF (mains failure fault) with a mean value of 16.8 and lowest for FF. Tower temperature, control cabinet temperature, transformer temperature, RTU: ava. Setpoint 1, and inverter averages were higher for excitation fault with 28.9, 38.8, 64.1, 3049.4, and 34.2, respectively (Table 4.1). They resulted in the lowest values except for RTU, such as 16.3, 24.9, 30.2, and 23, respectively for GF. RTU had to have the lowest mean of 2940.2 for MF. The inverter std. dev and group were highest under no-fault type, i.e., 1.9 and 1591.3, respectively. In the case of frozen, the mean score was reported as 0.2 which was highest for MF while it was lower for EF, FF, and GF with a 0-mean score.

4.4 Analysis of LSTM and CNN-based Machine Learning Models

Initially, an unsupervised learning algorithm of LSTM (Long Short-term Memory) and a supervised learning model of CNN (Convolutional Neural Network) were employed for the training dataset to predict the performance of fault detection. Accordingly, these models were used to evaluate the accuracy of fault detection individually and compare their accuracies with the proposed model of integrated LSTM and CNN-based algorithm with a five-fold cross-validation method for diagnosis of faults in wind turbine machines.

Table 4.2: Accuracies of LSTM, CNN, and Integrated CNN-LSTM Models for Different Faults

		Accuracy	Precision	Recall	F1-score
Long-Short Term Memory (LSTM)	AF	0.70	0.62	0.65	0.63
	EF		0.49	0.49	0.49
	FF		0.64	0.61	0.63
	GF		1.00	1.00	1.00
	MF		0.29	0.29	0.29
	NF		0.95	0.90	0.92
Convolutional Neural Networks (CNN)	AF	0.69	0.47	0.40	0.43
	EF		0.58	0.84	0.69
	FF		0.73	0.58	0.65
	GF		1.00	1.00	1.00
	MF		0.12	0.14	0.13
	NF		0.83	0.76	0.79
	AF	0.75	1.00	0.25	0.40

Integrated LSTM-CNN-based Algorithm	EF		0.60	1.00	0.75
	FF		1.00	0.51	0.68
	GF		1.00	1.00	1.00
	MF		0.00	0.00	0.00
	NF		0.75	0.93	0.83

When compared to the results of LSTM, CNN, and integrated LSTM-CNN models for predicting the fault detection in different fault types like AF, EF, FF, GF, MF, and NF, an improved accuracy of 75% was observed for the integrated LSTM-CNN model than the individual models like LSTM and CNN (Table 4.2). Although the accuracy was lower for CNN (69%) in comparison with LSTM (70%), the precision and F1-score values were improved with CNN for excitation fault (EF) and feed fault (FF) conditions. However, precision was increased from 49% to 58%, and F1-score from 49% to 69% in the case of excitation fault while they were improved outcomes from 64% to 73% and 63% to 65% in feed fault type, respectively. With the integrated LSTM-CNN model, the precision and F1-scores were further improved for EF (60% and 75%, respectively) and FF (100%, and 68%, respectively) (Table 4.2). Overall, the novel algorithm of the integrated LSTM-CNN classifier showed improved performance in the detection of faults in wind turbines.

In comparison between CNN-LSTM it is quite evident that the need to completely change and it is completely evident that the training time is highly proportional to all aspects. Based on the training times, the CNN-LSTM can be noted to be a much faster approach. According to Aksan et al. (2023), weather parameters among the datasets can be helpful in detecting and finding out any form of variant which can increase the training time. Thus, the input size of a data has been associated with being more can affect the training time of the powerflow grids. On the other hand, in contrast, it can be determined that the abilities and the ways things can get affected can be studied using the grid cluster which when trained can be useful for power dataflow. Experimental results from numerous studies have furthermore proven that the The Convolutional-Long Short-Term Memory-Deep Neural Network (CLDNN) model has been associated with being a convenient and accurate model as it has the reduced error rate of around 4-6% of LSTM which can better improve the robustness among larger datasets and affect different environmental conditions. This CNN-LSTM model can be predicted with the usage of factors such as house energy consumption, achievement of prediction performance from various previous aspects and predict areas such as power consumption which can improve robustness.

4.5 Summary of Data Findings

The chosen three different ML algorithms, such as LSTM, CNN, and combined LSTM-CNN models were evaluated against accuracy, precision, recall, and F1-score using Python. The results depicted that the proposed novel algorithm of LSTM-CNN based model was more accurate with 75% than the other algorithms like LSTM with 70% and CNN with 69% accuracy in predicting the fault detection under AF, FF, EF, GF, MF, and NF. In addition, the proposed model showed enhanced performance of precision and F1-score with 60% and 100%; and 75% and 68%, respectively under Excitation Fault and Feed Fault modes, respectively. Therefore, a novel algorithm of integrated LSTM-CNN could be able to predict fault detection and diagnose the faults efficiently in wind turbines.

5. Discussion and Conclusions

The integration of Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks identifies the fault wind turbine and diagnoses the approach of the advantages that combine both designs. For this reason, to identify these problems it is essential at an early stage to decrease the cost of maintenance and downtime, hence the complex systems of wind turbines move with various parts. CNNs are specifically suited for structural health to assess the wind turbine components for their ability to extract information from multidimensional data. The uses of CNN can learn and organize the patterns with spatial arrangements associated with different ideal conditions of operation by the sensor data that receives the variations in temperature and patterns of vibration. This primary feature of extraction improves the accuracy of the overall diagnostic of the system by focusing on the relevant things.

5.1 Discussion of Results

The analysis of exploratory data reveals a set of comprehensive variables related to the system of wind turbines, that range from factors of the environment to the temperatures of components and different metrics of performance. The fault data, which includes AF, EF, FF, GF, MF, and NF, serves as an indicator of critical system reliability and health. Analyzing these occurrences of fault temperatures alongside different variables like rotor temperatures, nacelle ambient temperature, and stator temperature provides insight into correlations between potential operational conditions and fault events. The various temperature measurements across the various components such as carrier rectifier, yaw inverter cabinet, and control cabinet suggest a need for an understanding of the thermal dynamics systems. The numerical values of presence associated implies each variable for the aspect of quantitative to the EDA, allow an analysis of further statistics and modeling. The environmental factors including ambient temperature and tower temperature underscore the external conditions impacting turbine performance. The metric Group with values like 1591.343 and 149.952, further warrants clarification for the dataset significance. Overall, this provides EDA with a deeper foundation of investigations into the relationships between fault occurrences, temperature variations, and other parameters operation, and finally the contribution of more approaches of informed to the system of wind turbine maintenance and optimization.

In this thesis, a data-driven approach was chosen for fault detection and diagnosis based on combined SCADA and fault data. The report of classification in the data of fault provides an evaluation of the comprehensive performance of the model across five distinct fault modes: generator heating fault (gf), mains failure fault (mf), feeding fault (ff), air cooling fault (af), and excitation fault (ef). The precision, recall, F1 score, and support metrics reported that insights of valuable in the ability of models to classify the correct instances of each fault type. The proposed integrated LSTM-CNN model was better than the other competitive models like CNN and LSTM algorithms in fault detection based on SCADA and fault data systems. This is due to the combined advantages of supervised and unsupervised algorithms that enhance the overall performance and accuracy of the model. The LSTM and CNN models were provided the lowest accuracies with 70% and 69%, respectively compared to the novel model (combined LSTM-CNN algorithm), which resulted in an accuracy of 75% (Table 4.2) under different fault modes like AF, EF, FF, GF, MF, and NF. These improved results were aligned with the study of Gong et al. (2019), in which the proposed novel CNN model showed better performance in the detection of faults. Moreover, the precision and F1-scores were also greatly increased for the Excitation Fault (EF) mode with 60% and 75%, respectively with integrated LSTM-CNN model than CNN (58% and 69%) and LSTM (49% and 49%) models. On the contrary, the study of Jiang et al. (2018) evaluated that the CNN-based model showed improved results of F1-scores in the detection of faults and diagnosis in wind turbines compared to other ML algorithms like SVM, LR, KNN, and DT. Similarly, the improved precision and F1-scores were observed under the Feed Fault (FF) model with 100% and 68%, respectively than LSTM (64% and 63%) and CNN (73% and 65%) algorithms. Based on these comparative advantages, the integrated LSTM-CNN model can be recommended for application in wind turbines to detect and diagnose faults for failure prediction.

5.2 Conclusion and Future Research

This thesis presented an improved LSTM-based CNN model for fault detection and diagnosis in wind turbines. The proposed approach of fault detection model based on machine learning was addressed in a way that the extracted and selected features were introduced based on principal component analysis (PCA) as an input for the LSTM-based CNN for classification purposes. The proposed novel model's effectiveness was validated in terms of accuracy, precision, and F1-score with two different machine learning approaches LSTM and CNN. We used the combined SCADA and fault data sets that included the fault components with several fault modes, including AF (air cooling fault), GF (generator heating fault), MF (Mains Failure Fault), FF (Feeding Fault), EF (Excitation Fault), and NF (No Fault) to evaluate the performance of the novel approach, i.e. integrated LSTM-based CNN. The data findings were evident that the proposed fault detection and diagnosis approach based on the LSTM-CNN model showed effectiveness in terms of accuracy, precision, and F1-score.

When using the proposed model, the fault diagnosis accuracy showed reduced outcomes specifically in the case of MF (Mains Failure Fault) and AF (Aircooling Fault) modes. Thus, the future research direction is to improve the supervised learning-based approaches to update the model and improve the classification results. Another future research direction is to improve the machine learning approaches dealing with uncertainties in wind turbines based on time-series analysis. In addition, integrated supervised and unsupervised learning models can be developed to improve decision-making accuracy and generate the optimal predictive model to provide the effective diagnosis of faults in wind turbines.

References

1. Abdallah, I., Ntertimanis, V., Mylonas, C., Tatsis, K., Chatzi, E., Dervilis, N., Keith, W. and Eoghan, M., (2018). Fault diagnosis of wind turbine structures using decision tree learning algorithms with big data. *Safety and Reliability–Safe Societies in a Changing World*, pp.3053-3061.

2. Agatonovic-Kustrin, S. and Beresford, R., (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, *22*(5), pp.717-727.

3. Aksan, F., Li, Y., Suresh, V. and Janík, P. (2023). CNN-LSTM vs. LSTM-CNN to Predict Power Flow Direction: A Case Study of the High-Voltage Subnet of Northeast Germany. *Sensors*, 23(2), pp.901–901. doi: https://doi.org/10.3390/s23020901.

4. Appiah, A.Y., Zhang, X., Ayawli, B.B.K. and Kyeremeh, F., 2019. Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis. *IEEE Access*, *7*, pp.30089-30101.

5. Avendaño-Valencia, L.D. and Fassois, S.D., (2017). Damage/fault diagnosis in an operating wind turbine under uncertainty via a vibration response Gaussian mixture random coefficient model-based framework. Mechanical Systems and Signal Processing, 91, pp.326-353.

6. Aziz, U., (2020). Power curve-based wind turbine fault detection: A critical performance comparison and proposition of a multi-turbine approach (Doctoral dissertation, Université Grenoble Alpes [2020]).

7. Aziz, U., Charbonnier, S., Bérenguer, C., Lebranchu, A. and Prevost, F., (2021). Critical comparison of power-based wind turbine fault-detection methods using a realistic framework for SCADA data simulation. Renewable and Sustainable Energy Reviews, 144, p.110961.

8. Baboli, P.T., Raeiszadeh, A., Babazadeh, D. and Meiners, J., (2020), September. Two-stage condition-based maintenance model of wind turbine: From diagnosis to prognosis. In *2020 IEEE International Smart Cities Conference (ISC2)* (pp. 1-6). IEEE.

9. Benkercha, R. and Moulahoum, S., (2018). Fault detection and diagnosis based on C4. 5 decision tree algorithms for grid connected PV system. *Solar Energy*, *173*, pp.610-634.

10. Bielecki, A., Barszcz, T., Wójcik, M. and Bielecka, M., (2014). Hybrid system of art and rbf neural networks for classification of vibration signals and operational states of wind turbines. In *Artificial Intelligence and Soft Computing: 13th International Conference*, *ICAISC 2014, Zakopane, Poland, June 1-5, 2014, Proceedings*, *Part I 13* (pp. 3-11). Springer International Publishing.

11. Breiman, L., (1996). Bagging predictors. *Machine learning*, *24*, pp.123-140.

12. Cabrera, D., Sancho, F., Sánchez, R.V., Zurita, G., Cerrada, M., Li, C. and Vásquez, R.E., (2015). Fault diagnosis of spur gearbox based on random forest and wavelet packet decomposition. *Frontiers of Mechanical Engineering*, *10*, pp.277-286.

13. Chang, Y., Chen, J., Qu, C. and Pan, T., (2020). Intelligent fault diagnosis of wind turbines via a deep learning network using parallel convolution layers with multi-scale kernels. *Renewable Energy*, *153*, pp.205-213.

14. Cho, S., Gao, Z., and Moan, T. (2018). Model-based Fault Detection, Fault Isolation and Fault-Tolerant Control of a Blade Pitch System in Floating Wind Turbines. Renew. Energ. 120, 306–321. doi: 10.1016/j.renene.2017.12.102

15. Clifton, A., Kilcher, L., Lundquist, J.K. and Fleming, P., (2013). Using machine learning to predict wind turbine power output. *Environmental research letters*, *8*(2), p.024009.

16. Dao, P. B., Staszewski, W. J., Barszcz, T., and Uhl, T. (2018). Condition Monitoring and Fault Detection in Wind Turbines Based on Cointegration Analysis of SCADA Data. Renew. Energ. 116, 107–122. doi: 10.1016/j.renene.2017.06.089

17. De Azevedo, H. D. M., Araújo, A. M., and Bouchonneau, N. (2016). A Review of Wind Turbine Bearing Condition Monitoring: State of the Art and Challenges. Renew. Sustainable Energ. Rev. 56, 368–379. doi: 10.1016/j.rser.2015.11.032

18. Deka, P.C., (2014). Support vector machine applications in hydrology: a review. *Applied soft computing*, *19*, pp.372-386.

19. Duchesne, L., Karangelos, E. and Wehenkel, L., 2020. Recent developments in machine learning for energy systems reliability management. *Proceedings of the IEEE*, *108*(9), pp.1656-1676.

20. Elizondo, J. L., Rivera, M., and Wheeler, P. (2019). "Wind Energy Development and Technology in theWorld: A Brief Overview,"in 2019 IEEE CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON) (IEEE), 1–5. doi:10.1109/CHILECON47746.2019.8987714

21. Freund, Y. and Schapire, R.E., (1996), July. Experiments with a new boosting algorithm. In *icml* (Vol. 96, pp. 148-156).

22. Gao, Z. and Liu, X., (2021). An overview on fault diagnosis, prognosis, and resilient control for wind turbine systems. *Processes*, *9*(2), p.300.

23. Gong, W., Chen, H., Zhang, Z., Zhang, M., Wang, R., Guan, C. and Wang, Q., 2019. A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion. *Sensors*, *19*(7), p.1693.

24. Goodfellow, I., Bengio, Y. and Courville, A., (2016). *Deep learning*. MIT press.

25. Guo, Z., Hao, Y., Shi, H., Wu, Z., Wu, Y. and Sun, X., (2023). A fault diagnosis algorithm for the dedicated equipment based on the CNN-LSTM mechanism. *Energies*, *16*(13), p.5230.

26. Habibi, H., Rahimi Nohooji, H. and Howard, I., (2017). Power maximization of variable-speed variable-pitch wind turbines using passive adaptive neural fault tolerant control. *Frontiers of Mechanical Engineering*, *12*, pp.377-388.

27. Helbing, G. and Ritter, M., (2018). Deep Learning for fault detection in wind turbines. *Renewable and Sustainable Energy Reviews*, *98*, pp.189-198.

28. Hu, Y., Zhang, S., Jiang, A., Zhang, L., Jiang, W. and Li, J., (2019). A new method of wind turbine bearing fault diagnosis based on multi-masking empirical mode decomposition and fuzzy C-means clustering. Chinese Journal of Mechanical Engineering, 32(1), pp.1-12.

29. Janssens, O., Slavkovikj, V., Vervisch, B., Stockman, K., Loccufier, M., Verstockt, S., Van de Walle, R. and Van Hoecke, S., (2016). Convolutional neural network-based fault detection for rotating machinery. *Journal of Sound and Vibration*, *377*, pp.331-345.

30. Jiang, G., He, H., Yan, J. and Xie, P., (2018). Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox. *IEEE Transactions on Industrial Electronics*, *66*(4), pp.3196-3207.

31. Jiménez, A.A., Muñoz, C.Q.G. and Márquez, F.P.G., (2019). Dirt and mud detection and diagnosis on a wind turbine blade employing guided waves and supervised learning classifiers. *Reliability Engineering & System Safety*, *184*, pp.2-12.

32. Kang, J., Wang, Z. and Guedes Soares, C., (2020). Condition-based maintenance for offshore wind turbines based on support vector machine. *Energies*, *13*(14), p.3518.

33. Lan, J., Patton, R.J. and Zhu, X., (2018). Fault-tolerant wind turbine pitch control using adaptive sliding mode estimation. *Renewable Energy, 116,* pp.219-231.

34. Leahy, K., Hu, R.L., Konstantakopoulos, I.C., Spanos, C.J. and Agogino, A.M., (2016). Diagnosing wind turbine faults using machine learning techniques applied to operational data. In *2016 ieee international conference on prognostics and health management (icphm)* (pp. 1-8). IEEE.

35. LeCun, Y., Bengio, Y. and Hinton, G., (2015). Deep learning. *nature*, *521*(7553), pp.436-444.

36. LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D., (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, *1*(4), pp.541-551.

37. Lei, J., Liu, C. and Jiang, D., (2019). Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renewable energy*, *133*, pp.422-432.

38. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A.K., (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, *138*, p.106587.

39. Letzgus, S., 2015. SCADA-data analysis for condition monitoring of wind turbines.

40. Li, C., Sanchez, R.V., Zurita, G., Cerrada, M., Cabrera, D. and Vásquez, R.E., (2016). Gearbox fault diagnosis is based on deep random forest fusion of acoustic and vibratory signals. *Mechanical Systems and Signal Processing, 76,* pp.283-293.

41. Liang, J., Zhang, K., Al-Durra, A. and Zhou, D., (2020). A novel fault diagnostic method in power converters for wind power generation system. *Applied energy*, *266*, p.114851.

42. Lin, Y., Tu, L., Liu, H., and Li, W. (2016). Fault Analysis of Wind Turbines in China. Renew. Sustainable Energ. Rev. 55, 482–490. doi: 10.1016/j.rser.2015.10.149

43. Lin, Z., Chen, Z., Qu, C., Guo, Y., Liu, J. and Wu, Q., (2020). A hierarchical clustering-based optimization strategy for active power dispatch of large-scale wind farm. International Journal of Electrical Power & Energy Systems, 121, p.106155.

44. Liu, H., Chen, C., Lv, X., Wu, X. and Liu, M., (2019). Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods. *Energy Conversion and Management*, *195*, pp.328-345.

45. Liu, K., Dong, W., Dong, H., Wei, J. and Xiao, S., (2020). A Complex Fault Diagnostic Approach of Active Distribution Network Based on SBS-SFS Optimized Multi-SVM. *Mathematical Problems in Engineering*, *2020*.

46. Liu, K., Dong, W., Dong, H., Wei, J., and Xiao, S. (2020b). A Complex Fault Diagnostic Approach of Active Distribution Network Based on SBS-SFS Optimized Multi-SVM. Math. Probl. Eng. 2020, 1–12. doi:10.1155/2020/ 8423571.

47. Liu, X., Gao, Z., and Chen, M. Z. Q. (2017). Takagi-Sugeno Fuzzy Model Based Fault Estimation and Signal Compensation with Application to Wind Turbines. IEEE Trans. Ind. Electron. 64 (7), 5678–5689. doi:10.1109/TIE.2017.2677327.

48. Liu, Y., and Ge, Z. (2018). Weighted Random Forests for Fault Classification in Industrial Processes with Hierarchical Clustering Model Selection. J. Process Control. 64, 62–70. doi: 10.1016/j.jprocont.2018.02.005

49. Mao, W., Tian, S., Fan, J., Liang, X., and Safian, A. (2020). Online Detection of Bearing Incipient Fault with Semi-supervised Architecture and Deep Feature Representation. J. Manufacturing Syst. 55, 179–198. doi:10.1016/ j. jmsy.2020.03.005

50. Moreno, S.R., Coelho, L.D.S., Ayala, H.V. and Mariani, V.C., (2020). Wind turbines anomaly detection based on power curves and ensemble learning. IET Renewable Power Generation, 14(19), pp.4086-4093.

51. Mucchielli, P., Bhowmik, B., Ghosh, B. and Pakrashi, V., (2021). Real-time accurate detection of wind turbine downtime-an Irish perspective. *Renewable Energy*, *179*, pp.1969-1989.

52. Ouanas, A., Medoued, A., Mordjaoui, M., Lebaroud, A. and Sayad, D., (2018). Fault diagnosis in yaw drive induction motor for wind turbine. *Wind Engineering*, *42*(6), pp.576-595.

53. Pei, Y., Qian, Z., Jing, B., Kang, D. and Zhang, L., (2018). Data-driven method for wind turbine yaw angle sensor zero-point shifting fault detection. *Energies*, *11*(3), p.553.

54. Polikar, R., (2012). Ensemble learning. *Ensemble machine learning: Methods and applications*, pp.1-34.

55. Qin, Y., Wang, X. and Zou, J., (2018). The optimized deep belief networks with improved logistic sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines. *IEEE Transactions on Industrial Electronics*, *66*(5), pp.3814-3824.

56. Rahman, M.M., Shakeri, M., Tiong, S.K., Khatun, F., Amin, N., Pasupuleti, J., and Hasan, M.K., 2021. Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability*, *13*(4), p.2393.

57. Safavian, S.R. and Landgrebe, D., (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), pp.660-674.

58. Samanta, B. and Al-Balushi, K.R., (2003). Artificial neural network-based fault diagnostics of rolling element bearings using time-domain features. *Mechanical systems and signal processing*, *17*(2), pp.317-328.

59. Saravanan, N., and Ramachandran, K.I., (2010). Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN). *Expert systems with applications*, *37*(6), pp.4168-4181.

60. Schmidhuber, J., (2015). Deep learning in neural networks: An overview. *Neural networks*, *61*, pp.85-117.

61. Schwenker, F. and Trentin, E., (2014). Pattern classification and clustering: A review of partially supervised learning approaches. *Pattern Recognition Letters*, *37*, pp.4-14.

62. Shi, H., Wei, A., Xu, X., Zhu, Y., Hu, H. and Tang, S. (2024). A CNN-LSTM based deep learning model with high accuracy and robustness for carbon price forecasting: A case of Shenzhen's carbon market in China. *Journal of Environmental Management*, [online] 352, p.120131. doi: https://doi.org/10.1016/j.jenvman.2024.120131.

63. Sinitsin, V., Ibryaeva, O., Sakovskaya, V. and Eremeeva, V., (2022). Intelligent bearing fault diagnosis method combining mixed input and hybrid CNN-MLP model. *Mechanical Systems and Signal Processing*, *180*, p.109454.

64. Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J. and Nenadic, G., (2019). Machine learning methods for wind turbine condition monitoring: A review. *Renewable energy*, *133*, pp.620-635.

65. Tang, B., Song, T., Li, F. and Deng, L., (2014). Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. *Renewable Energy*, *62*, pp.1-9.

66. Tang, M., Zhao, Q., Ding, S.X., Wu, H., Li, L., Long, W. and Huang, B., (2020). An improved lightGBM algorithm for online fault detection of wind turbine gearboxes. *Energies*, *13*(4), p.807.

67. Tang, M., Zhao, Q., Wu, H., Wang, Z., Meng, C. and Wang, Y., (2021). Review and perspectives of machine learning methods for wind turbine fault diagnosis. *Frontiers in Energy Research*, *9*, p.751066.

68. Toubakh, H. and Sayed-Mouchaweh, M., (2016). Hybrid dynamic classifier for drift-like fault diagnosis in a class of hybrid dynamic systems: Application to wind turbine converters. *Neurocomputing*, *171*, pp.1496-1516.

69. Vapnik, V., (1999). *The nature of statistical learning theory*. Springer science & business media.

70. Wang, K., Qi, X., Liu, H. and Song, J., (2018). Deep belief network-based k-means cluster approach for short-term wind power forecasting. Energy, 165, pp.840-852.

71. Wang, T. and Yin, L., (2024). A hybrid 3DSE-CNN-2DLSTM model for compound fault detection of wind turbines. Expert Systems with Applications, 242, p.122776.

72. Wen, X., Jaxa-Rozen, M. and Trutnevyte, E., 2022. Accuracy indicators for evaluating retrospective performance of energy system models. *Applied Energy*, *325*, p.119906.

73. Wu, R.M., Zhang, Z., Yan, W., Fan, J., Gou, J., Liu, B., Gide, E., Soar, J., Shen, B., Fazal-e-Hasan, S. and Liu, Z., 2022. A comparative analysis of the principal component analysis and entropy weight methods to establish the indexing measurement. *PloS one*, *17*(1), p.e0262261.

74. Xi, L., Wu, J., Xu, Y. and Sun, H., (2020). Automatic generation control based on multiple neural networks with an actor-critic strategy. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(6), pp.2483-2493.

75. Xiang, L., Wang, P., Yang, X., Hu, A. and Su, H., (2021). Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism. *Measurement*, *175*, p.109094.

76. Xin, G., Hamzaoui, N., and Antoni, J. (2018). Semi-automated Diagnosis of Bearing Faults Based on a Hidden Markov Model of the Vibration Signals. Measurement 127, 141–166. doi: 10.1016/j.measurement.2018.05.040

77. Yang, B., Liu, R. and Chen, X., (2017). Fault diagnosis for a wind turbine generator bearing via sparse representation and shift-invariant K-SVD. *IEEE Transactions on Industrial Informatics*, *13*(3), pp.1321-1331.

78. Yang, T., Pen, H., Wang, Z., and Chang, C. S. (2016). Feature Knowledge Based Fault Detection of Induction Motors through the Analysis of Stator Current Data. IEEE Trans. Instrum. Meas. 65 (3), 549–558. doi:10.1109/ TIM.2015.2498978

79. Yang, X., Li, J., Liu, W. and Guo, P., (2011). Petri net model and reliability evaluation for wind turbine hydraulic variable pitch systems. *Energies*, *4*(6), pp.978-997.

80. Yesilbudak, M., (2016), November. Clustering analysis of multidimensional wind speed data using k-means approach. In 2016 IEEE international conference on renewable energy research and applications (ICRERA) (pp. 961-965). IEEE.

81. Yin, S., Wang, G., and Karimi, H. R. (2014). Data-driven Design of Robust Fault Detection System for Wind Turbines. Mechatronics 24 (4), 298–306. doi: 10.1016/j.mechatronics.2013.11.009

82. Yu, D., Chen, Z.M., Xiahou, K.S., Li, M.S., Ji, T.Y. and Wu, Q.H., (2018). A radically data-driven method for fault detection and diagnosis in wind turbines. *International Journal of Electrical Power & Energy Systems*, *99*, pp.577-584.

83. Zhang, D., Qian, L., Mao, B., Huang, C., Huang, B. and Si, Y., (2018). A data-driven design for fault detection of wind turbines using random forests and XGboost. *Ieee Access*, *6*, pp.21020-21031.

84. Zhang, S., and Lang, Z.Q., (2020). SCADA-data-based wind turbine fault detection: A dynamic model sensor method. Control Engineering Practice, 102, p.104546.

85. Zhang, T., and Zhou, Z. H. (2018a). “Optimal Margin Distribution Clustering,” in Proceedings of the AAAI Conference on Artificial Intelligence.

86. Zhang, T., and Zhou, Z.-H. (2018b). “Semi-Supervised Optimal Margin Distribution Machines,” in Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence: IJCAI 2018, 3104–3110. doi:10.24963/ijcai.2018/431

87. Zhang, X. and Dimitrov, N., (2023). Extreme wind turbine response extrapolation with Gaussian mixture model. Wind Energy Science Discussions, 2023, pp.1-16.

88. Zhang, Z.Y. and Wang, K.S., (2014). Wind turbine fault detection based on SCADA data analysis using ANN. *Advances in Manufacturing*, *2*, pp.70-78.

89. Zhao, H., Liu, H., Hu, W. and Yan, X., (2018). Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renewable energy*, *127*, pp.825-834.

90. Zhao, L., Pan, Z., Shao, C. and Yang, Q., (2015). Application of SOM neural network in fault diagnosis of wind turbine. In *International Conference on Renewable Power Generation (RPG 2015)* (pp. 1-4). IET.

91. Zhong, M., Xue, T., and Ding, S. X. (2018). A Survey on Model-Based Fault Diagnosis for Linear Discrete Time-Varying Systems. Neurocomputing 306, 51–60. doi: 10.1016/j.neucom.2018.04.037

92. Zhou, Z.H., (2018). A brief introduction to weakly supervised learning. *National science review*, *5*(1), pp.44-53.