Journal of Technology Innovations and Energy
https://doi.org/10.56556/jtie.v3i3.1028
ISSN: 2957-8809

REVIEW ARTICLE

A Comprehensive Review of SCADA-Based Wind Turbine Performance and Reliability Modeling with Machine Learning Approaches

Hlaing Min Soe[1*], Arkar Htet[2]

[1]Faculty of Business and Accounting, UCSI University, Taman Connaught, 56000, Kuala Lumpur, Malaysia
[2]Faculty of Engineering and Build Environment, Lincoln University, 47301, Petaling Jaya, Selangor D. E., Malaysia

Corresponding Author: Hlaing Min Soe. Email: hlaingminsoe.stc@gmail.com
Received: 21 August, 2024, Accepted: 27 September, 2024, Published: 28 September, 2024

Abstract
The increasing reliance on wind energy to meet global energy demands has made wind turbine performance optimization and reliability a critical area of research. Supervisory Control and Data Acquisition (SCADA) data, which provides real-time operational insights into wind turbines, plays a pivotal role in predictive maintenance, fault detection, and energy output optimization. This review explores the current methodologies and advancements in wind turbine performance modeling and reliability analysis, with a particular emphasis on machine learning (ML) approaches. Existing studies that utilize SCADA data to implement various ML models, such as decision trees, neural networks, and ensemble learning techniques (bagging, boosting, stacking), are analyzed for their effectiveness in predicting turbine failures, improving energy efficiency, and optimizing maintenance schedules. Key findings from multiple studies are synthesized, highlighting the strengths, limitations, and real-world applications of these models. Challenges in data quality, model generalization, and the implementation of real-time ML-driven systems in wind farms are also addressed. This review aims to provide a comprehensive overview of the current state of SCADA-based wind turbine analysis and to offer a roadmap for future research that bridges the gap between data-driven models and their practical deployment in wind energy systems.

Keywords: Wind Turbine Reliability; SCADA Data Analysis; Machine Learning In Wind Energy; Predictive Maintenance; Ensemble Learning Techniques

Introduction

Background

The increasing global demand for clean and renewable energy is driven by concerns over climate change, fossil fuel depletion, and the need for greater energy security. Among the various renewable options, wind energy has proven to be one of the most scalable and sustainable solutions to meet rising energy needs. Over the last two decades, wind power has experienced rapid growth, with significant investments in onshore and offshore wind farms worldwide, aiming to reduce carbon emissions and transition to sustainable energy systems (Shittu et al., 2019). The Global Wind Energy Council (GWEC) reports that global wind energy capacity has more than doubled in the past decade, cementing its status as a major contributor to the renewable energy mix (Zhou & Lv, 2013).

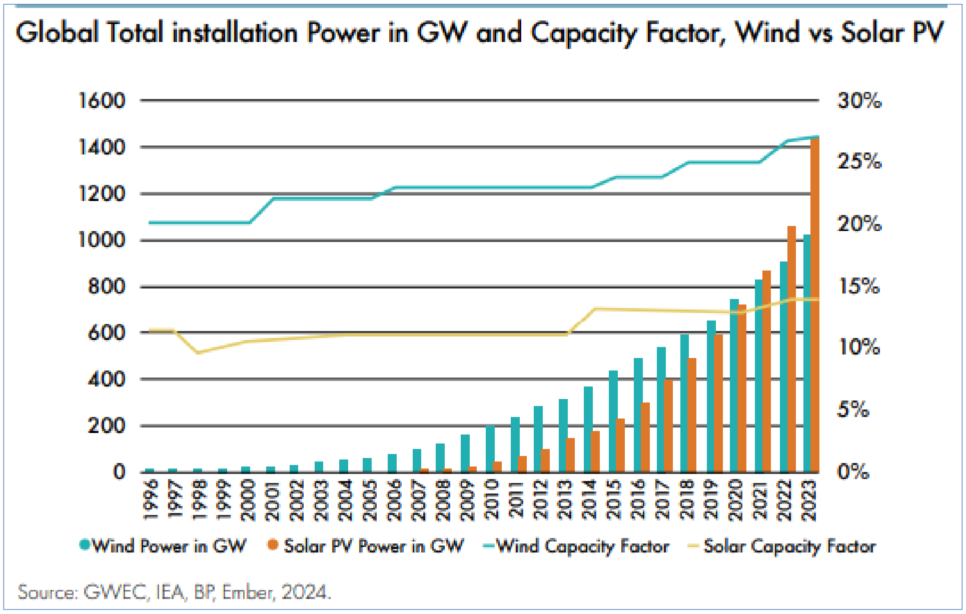

Figure 1: Global Total Installation Power in GW and Capacity Factor, Wind vs. Solar PV (Source: GWEC, IEA, BP, Ember, 2024).

As shown in Figure 1, wind energy capacity has grown steadily alongside the rise of solar photovoltaic (PV) technology, underscoring the expansion of renewable energy infrastructure. Wind energy's stable capacity factor highlights its reliability in power generation. Wind turbines, the backbone of wind energy production, convert kinetic energy into electrical power, but their efficiency is influenced by several factors, including wind conditions, mechanical efficiency, and operational settings (Alkesaiberi et al., 2022). Even minor inefficiencies in large-scale wind farms, where multiple turbines operate continuously, can lead to substantial energy losses (Veena et al., 2020). Therefore, optimizing turbine performance and ensuring reliability are vital for maximizing energy output and achieving sustainable energy goals (Feddaoui et al., 2023). Reliability is key, as unexpected maintenance and downtimes can significantly disrupt energy production and incur high costs (Carroll et al., 2015). To mitigate these challenges, predictive maintenance, fault detection, and performance optimization have become central areas of focus in the wind energy sector. These strategies aim to reduce downtimes, prolong turbine lifespan, and enhance energy efficiency (Chen, 2023). Supervisory Control and Data Acquisition (SCADA) systems play a critical role by providing real-time data on turbine operations. Coupled with advanced machine learning (ML) techniques, SCADA data presents significant opportunities for improving turbine reliability and performance through predictive analytics and real-time decision-making (Zhang & Kusiak, 2012).

Role of SCADA Data

Supervisory Control and Data Acquisition (SCADA) systems play a critical role in the operation and maintenance of wind turbines by providing real-time data that enhances monitoring, fault detection, and performance optimization. SCADA systems allow operators to remotely monitor turbine operations, collect key performance data, and ensure continuous, efficient energy production. Leveraging SCADA data enables operators to improve both turbine reliability and overall performance (Wei et al., 2021). SCADA systems track vital operational parameters such as wind speed, power output, rotor speed, yaw offset, and temperature, offering detailed insights into turbine health (Hajjaj et al., 2022). Wind speed reflects the kinetic energy available for electricity generation, while power output measures the turbine's energy production. Rotor speed indicates mechanical performance, and temperature readings help identify overheating or other operational abnormalities (Gorel & Abdi, 2023). Yaw offset, which measures the misalignment between the wind direction and the rotor, is crucial for optimizing the turbine's alignment to maximize efficiency. These parameters are recorded at regular intervals, typically every 5 to 10 minutes, providing granular data for detailed analysis (Kanev, 2020). In addition to basic monitoring, SCADA data supports real-time optimization, fault detection, and predictive maintenance. By continuously comparing real-time data to historical performance trends, SCADA systems can identify anomalies that may signal potential faults (Chen, 2022). For instance, abnormal temperature spikes or vibrations may indicate early signs of mechanical failures, such as gearbox or blade damage. Early fault detection through SCADA allows for proactive maintenance scheduling, minimizing downtime and reducing repair costs (Bonacina et al., 2022). SCADA data is also essential for optimizing turbine performance. By analyzing trends in wind speed, rotor dynamics, and energy output, operators can adjust turbine settings to achieve optimal efficiency under changing environmental conditions (Zhang et al., 2020). This data-driven approach maximizes energy production and improves turbine efficiency, ultimately extending the lifespan of the equipment. Furthermore, integrating SCADA data with advanced machine learning algorithms facilitates the development of predictive models, which improve failure forecasting and maintenance scheduling (Wang et al., 2020).

Machine Learning in Wind Turbine Optimization

Machine learning (ML) has become a key tool in wind energy management, offering effective solutions for optimizing turbine performance, improving reliability, and reducing downtime (Ponkumar et al., 2024). With the vast data generated by Supervisory Control and Data Acquisition (SCADA) systems, ML models help operators analyze turbine performance, predict potential failures, and optimize energy output in real-time. This capability to process large volumes of SCADA data and identify patterns undetectable by traditional methods makes ML indispensable in wind energy operations (Santolamazza et al., 2021).

Several ML techniques have been successfully applied to specific areas of turbine optimization:

- Decision Trees: Decision trees are widely used to classify turbine health and detect operational anomalies based on SCADA data. By analyzing parameters such as rotor speed, wind speed, and power output, decision trees can determine whether a turbine is operating normally or requires maintenance. The simplicity and interpretability of decision trees make them a popular choice for fault detection and maintenance scheduling (Jiang et al., 2024).
- Neural Networks: Artificial neural networks (ANNs) have shown great promise in predicting turbine failures and optimizing energy production. By processing complex, nonlinear relationships between variables like wind speed, rotor speed, and temperature, ANNs enable accurate failure predictions, allowing operators to take proactive measures. Neural networks are also highly adaptable to changing environmental conditions, making them ideal for wind farms that experience fluctuating wind patterns (Xu et al., 2022).
- Ensemble Learning Techniques: Ensemble learning combines multiple models to improve predictive accuracy and robustness. These techniques are highly effective in turbine performance optimization and failure prediction. Common ensemble methods like bagging, boosting, and stacking offer distinct advantages (Hui et al., 2023). For instance, bagging reduces variance by averaging predictions from multiple models trained on different data subsets, while boosting improves weaker models by correcting

their errors in subsequent iterations (Kumar et al., 2018). Integrating ensemble models with SCADA data enables more accurate predictions of component failures, reducing the need for reactive maintenance and lowering operational costs (Liu et al., 2019).

The integration of ML models with SCADA data enables real-time decision-making in wind turbine management. Operators can predict turbine failures before they happen, allowing for timely maintenance and minimizing unplanned downtime. These models can also optimize turbine settings, such as rotor speed and blade pitch, ensuring maximum energy output under varying wind conditions. Overall, combining ML with SCADA data enhances turbine efficiency, reliability, and sustainability, contributing to more stable and cost-effective wind energy production.

Aim and Scope of the Review

The primary objective of this paper is to review the integration of Supervisory Control and Data Acquisition (SCADA) data with machine learning (ML) techniques to optimize wind turbine performance and enhance reliability. Wind energy, one of the most scalable renewable sources, heavily relies on the operational efficiency of turbines. SCADA systems continuously monitor turbine performance, generating vast amounts of real-time data that, when processed using ML algorithms, can predict failures, improve maintenance scheduling, and optimize energy output. The combination of SCADA data with ML has the potential to revolutionize wind turbine management, enabling more effective predictive maintenance and fault detection.

The paper aims to:

- Review existing studies that use SCADA data and ML techniques to improve turbine reliability, performance monitoring, and failure prediction.
- Evaluate the effectiveness of different ML models, such as decision trees, neural networks, and ensemble learning, in processing SCADA data to optimize turbine operations.
- Discuss challenges related to implementing these data-driven approaches in real-world wind farms, including data quality, model generalization, and real-time application.
- Identify gaps in the current research and suggest directions for future studies, focusing on refining ML models to address the specific complexities of wind energy systems.

Structure of the Paper

The paper is organized as follows:

- Introduction: Overview of wind energy, SCADA data, and the role of ML in enhancing turbine performance and reliability.
- SCADA Data and Its Applications in Wind Turbines: Discusses the key parameters collected by SCADA systems and their importance in monitoring and optimizing turbine operations.
- Machine Learning Techniques in Wind Turbine Reliability and Performance Modeling: Reviews ML models applied to wind turbine data, highlighting their strengths, limitations, and real-world performance.
- Integration of SCADA Data and Machine Learning for Predictive Maintenance: Explores how ML and SCADA data are used for predictive maintenance, fault detection, and failure prediction, along with practical implications.
- SCADA Data-Driven Fault Detection and Energy Production Optimization: Analyzes case studies demonstrating how SCADA data optimizes energy production and detects faults in wind turbines.

- Challenges and Future Directions: Discusses challenges in applying ML to SCADA data and offers recommendations for future research, including model improvements and real-time applications.

This structure ensures a comprehensive review of SCADA-driven ML approaches, providing valuable insights into their practical applications in the wind energy sector.

SCADA Data and Its Applications in Wind Turbines

SCADA Data Overview

Supervisory Control and Data Acquisition (SCADA) systems are essential for wind turbine operation and maintenance, providing real-time data that offers critical insights into turbine performance, health, and efficiency. SCADA systems monitor key operational parameters such as wind speed, rotor speed, power output, yaw offset, and temperature. These parameters allow operators to assess turbine status and identify abnormalities that could indicate potential faults or inefficiencies (Hajjaj et al., 2022).

- Wind speed is a critical factor affecting turbine performance, as it determines the kinetic energy available for electricity generation. SCADA systems track wind speed continuously to monitor energy production and ensure optimal operating conditions (Uchida & Sugitani, 2020).
- Rotor speed measures the rotational velocity of the turbine blades, a key indicator of mechanical performance. Deviations from expected rotor speeds may suggest mechanical issues such as imbalances or wear (Zhang & Kusiak, 2012).
- Power output represents the amount of electrical energy generated by the turbine, providing a direct measure of efficiency. SCADA data helps evaluate how effectively the turbine converts wind energy into electricity under different conditions (Lee et al., 2015).
- Yaw offset refers to the angular difference between the wind direction and the rotor's alignment. Maintaining a minimal yaw offset ensures maximum energy capture by aligning the turbine correctly with wind flows (Kanev, 2020).
- Temperature is crucial for monitoring the operational health of components like the gearbox and generator. Abnormally high temperatures may signal mechanical issues such as friction, which could lead to failure if not addressed promptly (Wang et al., 2023).

Table 1: Key SCADA Parameters and Their Importance in Turbine Operation

SCADA Parameter	Description	Importance for Turbine Operation
Wind Speed	Speed of incoming wind	Critical for power generation
Rotor Speed	Rotational velocity of the turbine blades	Impacts mechanical efficiency
Power Output	Electrical power generated by the turbine	Measures overall turbine output
Yaw Offset	Angular difference between wind direction and rotor alignment	Ensures maximum energy capture
Temperature	Temperature of components like gearbox and generator	Indicates health and potential faults

As shown in Table 1**,** SCADA systems track crucial parameters that allow operators to quickly detect potential inefficiencies or faults in turbine operations. This continuous monitoring is vital for maintaining optimal turbine performance.

In addition to operational parameters, SCADA systems collect time-series data at regular intervals, typically every 5 to 10 minutes. This data enables trend analysis, helping operators identify performance declines or increasing mechanical wear over time. By comparing real-time data with historical records, SCADA systems provide valuable insights into long-term turbine health (Wei et al., 2021). SCADA systems also capture environmental data, including wind direction, air pressure, and humidity, which impact turbine performance. Combined with operational data, these environmental variables give operators a comprehensive view of turbine efficiency under varying weather conditions. Additionally, SCADA data can be integrated with operational logs, such as turbine start-up, shutdown, and maintenance history, to offer a complete picture of a turbine's lifecycle performance (Wu et al., 2019). The ability to collect such granular and continuous data makes SCADA systems indispensable for performance monitoring, fault detection, and predictive maintenance in wind energy systems. Analyzing SCADA data allows operators to optimize turbine settings, predict failures, and schedule maintenance more effectively, resulting in increased turbine availability and reduced operational costs (Arch et al., 2020). SCADA-driven insights also enable better decision-making, contributing to the sustainability and efficiency of wind farm operations (Vidal et al., 2023).

Challenges in SCADA Data Analysis

While Supervisory Control and Data Acquisition (SCADA) systems provide invaluable data for wind turbine performance optimization and fault detection, several challenges arise in effectively analyzing and utilizing this data. These challenges primarily involve data quality, variability across turbine models, and the complexity of processing large datasets.

Data Quality Issues

One of the key challenges in SCADA data analysis is ensuring data quality. SCADA systems often produce data that may be incomplete, erroneous, or affected by external noise. Missing data points, caused by sensor malfunctions, communication failures, or transmission interruptions, are a frequent issue (Castellani et al., 2014). Incomplete datasets can lead to inaccurate analysis and unreliable predictions, particularly when machine learning models depend on high-quality, continuous data to generate accurate forecasts.

As illustrated in Figure 2**,** SCADA data presents several challenges, from missing and erroneous data to variability in turbine models and the complexity of processing large volumes of real-time data. Addressing these challenges is essential for accurate fault detection and performance optimization in wind farms.

Erroneous data, often caused by faulty sensors or human error, presents another significant challenge. Such inaccuracies can distort analyses, leading to incorrect conclusions about turbine performance or misidentification of faults. For example, a faulty sensor may report inaccurate wind speed values, causing operators to incorrectly assume that a turbine is underperforming when, in fact, it is operating normally (Biazar et al., 2022). Addressing these inconsistencies requires rigorous preprocessing techniques, including data cleaning, outlier detection, and the imputation of missing values (Kim & Kim, 2024).

Variability in Turbine Models

Another major challenge in SCADA data analysis is the variability among different turbine models. Wind farms often contain turbines from various manufacturers, each with its own design, specifications, and performance characteristics. This variability complicates data analysis, as SCADA systems may collect data differently depending on the turbine model (Menezes et al., 2020). For instance, some turbine models may include sensors that monitor blade pitch, while others rely solely on basic performance indicators like rotor speed and power output. Differences in data collection methodologies make it difficult to apply a uniform analysis across an entire wind farm. Machine learning models, typically trained on consistent datasets, must be adapted to account for variations in sensor data and turbine configurations. Additionally, the lack of standardized data formats across different models complicates data integration, limiting the ability to generalize findings from one model to another (Murgia et al., 2023).

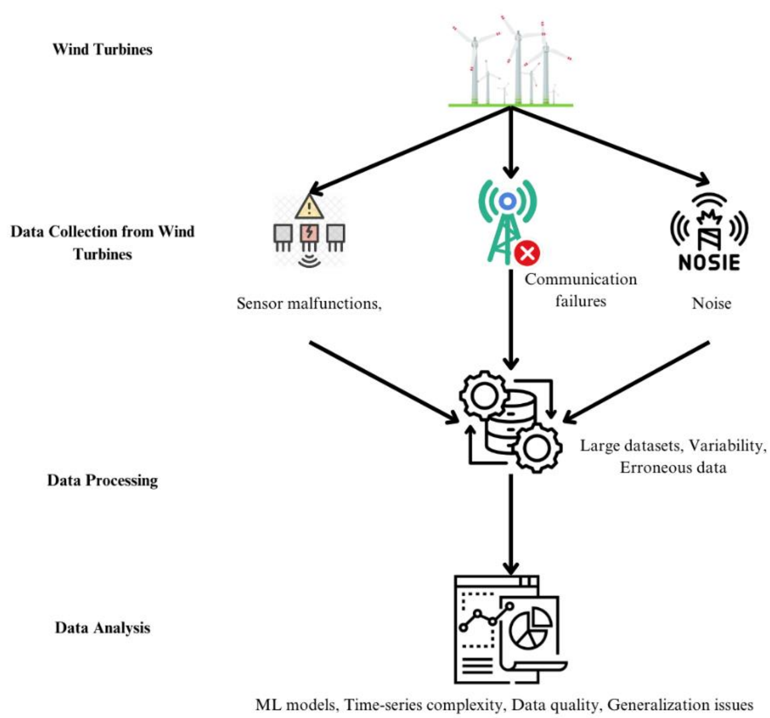

Figure 2: Conceptual Overview of SCADA Data Challenges

Processing Large Datasets

SCADA systems generate vast amounts of data, with each turbine producing gigabytes of time-series data on parameters like wind speed, rotor speed, temperature, and power output. Analyzing these large datasets poses a computational challenge, especially when real-time data processing is required for effective performance monitoring and fault detection (Muhammad Amri et al., 2024). The high volume of data, combined with frequent data collection intervals (typically every 5 to 10 minutes), can overwhelm traditional data processing systems and make it difficult to extract meaningful insights quickly (Chen et al., 2023).

Additionally, the time-series nature of SCADA data introduces complexity, as temporal dependencies between data points must be carefully modeled to predict future performance or detect failures. Advanced data processing techniques, such as time-series forecasting models, deep learning algorithms, and data reduction strategies like principal component analysis (PCA), are needed to make this analysis computationally feasible (Wang et al., 2020). Real-time data processing is another critical challenge in operational wind farms. Wind farm operators often need to make immediate decisions based on SCADA data, particularly in situations where early fault detection is essential to prevent turbine failure. Implementing real-time data processing requires sophisticated algorithms capable of handling large volumes of data efficiently while minimizing latency (Annoni et al., 2019).

Applications of SCADA Data

SCADA data has become essential for the efficient management and optimization of wind farm operations. By collecting real-time operational data, SCADA systems enable operators to monitor turbine performance, detect faults, and improve energy production. This section reviews the effective applications of SCADA data in fault detection, energy production optimization, predictive maintenance, and real-time monitoring, highlighting its advantages over traditional wind farm management methods.

Fault Detection

One of the most significant applications of SCADA data is fault detection. Wind turbines are prone to mechanical and electrical failures, which, if not detected early, can lead to costly repairs and unplanned downtime. SCADA systems continuously monitor key parameters, such as rotor speed, wind speed, power output, and temperature, allowing operators to detect anomalies that may indicate early-stage faults (Liu et al., 2023). For example, an abnormal increase in bearing temperature or fluctuations in rotor speed can signal mechanical wear, gearbox failures, or electrical malfunctions. Recent studies demonstrate that applying machine learning models to SCADA data enhances fault detection by identifying patterns in historical and real-time data that precede failures. Predictive models using decision trees, neural networks, and ensemble learning have successfully detected faults before they occur, reducing both downtime and repair costs (Ulmer et al., 2020). By identifying faults early, SCADA-based systems can prevent catastrophic failures, significantly improving turbine reliability and lifespan (Byrne et al., 2020).

Energy Production Optimization

SCADA data is also crucial for optimizing energy production. Wind turbine performance is influenced by various factors, including wind speed, turbine alignment (yaw), and blade angles (pitch). SCADA systems provide real-time insights into these variables, enabling operators to adjust turbine parameters and maximize energy output (Verma et al., 2022). For instance, SCADA data can be used to optimize turbine yaw alignment to ensure that turbines are positioned directly into the wind, thereby maximizing energy capture (Sihua et al., 2020). SCADA systems can also dynamically adjust blade pitch based on wind speed and rotor speed, optimizing aerodynamic performance even under fluctuating conditions (Song et al., 2022). These data-driven strategies help maximize energy production, reduce inefficiencies, and lower the cost per kilowatt-hour of electricity produced (Donadio et al., 2021).

Predictive Maintenance

Predictive maintenance is one of the most powerful applications of SCADA data in modern wind farms. Unlike traditional maintenance approaches, which rely on scheduled inspections or reactive repairs, predictive maintenance leverages SCADA data to forecast when a component is likely to fail and schedules maintenance accordingly (Peter et al., 2022). This proactive approach reduces emergency repairs and unplanned downtime, improving turbine availability and lowering maintenance costs. SCADA systems continuously monitor critical components like gearboxes, generators, and rotor blades, providing valuable insights into wear and degradation. Machine learning models applied to SCADA data can predict the remaining useful life (RUL) of components based on long-term trends. For example, temperature trends in the gearbox or generator can indicate bearing wear, while abnormal vibration patterns may signal blade fatigue (Yan et al., 2014). By predicting failures in advance, SCADA-based predictive maintenance allows repairs to be scheduled during planned downtimes, maximizing turbine availability and extending the lifespan of the equipment (Pandit & Wang, 2014).

Real-Time Monitoring Systems

A key advantage of SCADA data over traditional wind farm management methods is its ability to enable real-time monitoring. In the past, manual inspections or scheduled maintenance were often the only ways to ensure turbine health, leading to undetected faults between inspection intervals (Becker-Dombrowsky et al., 2023).SCADA systems, however, provide continuous real-time data on turbine operations, allowing operators to detect and address issues as they arise. By monitoring operational parameters such as wind speed, power output, and rotor speed, SCADA systems can identify performance deviations in real time, enabling immediate corrective actions (Chen et al., 2023). This ensures that turbines operate at optimal efficiency, reducing energy losses and preventing major failures. Moreover, integrating SCADA data with machine learning models enhances real-time monitoring capabilities. Advanced algorithms process continuous data streams to provide automated fault detection, performance optimization, and predictive maintenance alerts without requiring manual intervention. As a result, SCADA-based real-time monitoring systems increase the efficiency and reliability of wind farms, while significantly reducing the operational costs associated with traditional management methods (Lebranchu et al., 2019).

Machine Learning Techniques in Wind Turbine Reliability and Performance Modeling

Overview of Machine Learning Models

Machine learning (ML) has transformed wind energy management by enabling more precise predictive maintenance, fault detection, and performance optimization. By analyzing vast amounts of Supervisory Control and Data Acquisition (SCADA**)** data, ML models identify patterns that help operators optimize wind turbines and prevent failures (Abd-Elwahab & Ali, 2020). This section provides an overview of three major categories of ML models applied in wind energy: supervised, unsupervised, and reinforcement learning.

Supervised Learning Models

Supervised learning is widely used in wind energy, where models are trained on labeled datasets. This method associates input data, such as wind speed or rotor speed, with known outcomes like power output or turbine health.

- Decision Trees: Decision trees classify turbine health by using a tree-like structure based on conditional rules. For example, input variables like rotor speed and power output can determine whether a turbine is operating normally or requires maintenance (Leahy et al., 2020). Decision trees are highly interpretable, which is valuable for operators who need clear explanations of the decision-making process.

- Neural Networks: Artificial neural networks (ANNs) model complex, nonlinear relationships between variables such as wind speed, rotor speed, and temperature. ANNs are effective in predicting turbine failures and adjusting energy production. They are particularly useful in fluctuating wind conditions, as they adapt to changes and improve forecasting accuracy (Li et al., 2015; Moyón et al., 2022).

Unsupervised Learning Models

Unlike supervised learning, unsupervised learning models identify hidden patterns in data without labeled outcomes. In wind energy, these models are commonly used for clustering.

- Clustering Techniques**:** Clustering groups turbines with similar operational characteristics. For example, turbines with similar vibration patterns or power output fluctuations can be clustered together to identify common issues (Zhang & Kusiak, 2012). Clustering also helps in anomaly detection by flagging turbines that deviate significantly from expected operational norms, enabling early fault detection (Seifert et al., 2021).

Reinforcement Learning for Optimizing Control Systems

Reinforcement learning (RL) is increasingly used in wind turbine control systems. Unlike other models, RL learns optimal actions by interacting with the environment. In wind energy, RL optimizes turbine control strategies, such as blade pitch adjustment or yaw control, to maximize energy output under varying wind conditions (Saenz-Aguirre et al., 2019). In RL, the model (agent) receives feedback through rewards or penalties based on its actions. Over time, the agent learns to make decisions that maximize long-term rewards, such as increasing energy output or minimizing mechanical stress (Renn & Gharib, 2022).

Table 2: Summary of Machine Learning Models in Wind Turbine Applications

ML Model Type	Example Algorithms	Use Case	Strengths	Weaknesses
Supervised Learning	Decision Trees, Neural Networks	Fault detection, energy prediction	High accuracy, interpretable (for Decision Trees), handles complex relationships (for ANNs)	Requires labeled data, may overfit in complex systems (for Decision Trees)
Unsupervised Learning	K-means Clustering, Hierarchical Clustering	Anomaly detection, grouping turbines with similar characteristics	Can find hidden patterns without labeled data, useful for anomaly detection	Can be hard to interpret, does not predict specific outcomes
Reinforcement Learning	Q-learning, Deep Q-networks	Optimizing control strategies (e.g., blade pitch adjustment, yaw control)	Learns from interaction with the environment, improves over time	Requires large amounts of data and computing resources for training

For example, an RL agent might learn to adjust blade pitch to maintain optimal rotor speed as wind conditions fluctuate (Sierra-García & Santos, 2020). This continuous improvement in control strategy enhances performance and extends turbine lifespan (Choquehuanca & Ortega, 2023). Table 2 below summarizes the main machine learning models used in wind turbine applications, highlighting their use cases, strengths, and limitations.

Ensemble Learning Techniques

Ensemble learning has become a popular approach in wind turbine modeling, enhancing the accuracy and robustness of machine learning models by combining multiple algorithms. Techniques like bagging, boosting, and stacking are widely used to tackle challenges related to data variability, noise, and complex patterns in SCADA data. These methods are especially effective in tasks such as fault detection, energy optimization, and predictive maintenance (Lyons & Göçmen, 2021).

Bagging

Bagging (Bootstrap Aggregating) improves model accuracy by reducing variance. It works by generating multiple subsets of training data using random sampling, with replacement, and training separate models (usually decision trees) on each subset. The predictions are then averaged (for regression) or voted upon (for classification) to generate a final result. By averaging predictions, bagging reduces the risk of overfitting, especially in models prone to high variance, like decision trees (Razavi-Termeh et al., 2021). In wind turbine modeling, bagging is often used to enhance the reliability of fault detection models. For example, decision trees trained on different SCADA data subsets may produce varied predictions due to noise or anomalies. However, by averaging these predictions, bagging provides more stable and accurate fault detection or performance assessments (Wu et al., 2022). This reduction in variance makes bagging highly effective in complex, noisy environments like wind farms.

Boosting

Boosting enhances predictive accuracy by training models sequentially, with each model correcting the errors of its predecessors. Unlike bagging, where models are trained independently, boosting builds a series of weak learners (usually shallow decision trees) that are combined to form a stronger model (Lauritsen et al., 2021). Boosting is particularly useful in wind turbine tasks such as failure prediction and energy output optimization. In predictive maintenance, a boosting model might initially detect potential faults based on basic SCADA data like rotor speed and temperature. Subsequent models refine these predictions, focusing on misclassified data points. This iterative process enables boosting models to achieve high accuracy in detecting faults and optimizing turbine performance (Galadima et al., 2024). Popular variations like gradient boosting and AdaBoost have been applied successfully in wind energy systems (Mienye & Sun, 2022).

Stacking

Stacking combines the predictions of multiple models to improve overall accuracy. Unlike bagging and boosting, stacking uses different types of models (e.g., decision trees, neural networks, support vector machines) and combines their outputs using a meta-model. The meta-model is trained to best combine the predictions from these models, producing a more accurate final result (Müller & Soto-Rey, 2022).

In wind turbine modeling, stacking is valuable when different models excel at predicting different aspects of performance. For instance, a neural network might predict energy output well, while a decision tree may excel at anomaly detection in turbine temperature. Stacking leverages these strengths by combining the predictions, leading to more accurate results for tasks like fault detection and performance optimization (Babawarun et al., 2023).

Case Studies on ML Applications

Machine learning (ML) has proven highly effective in improving wind turbine reliability, predicting faults, and optimizing energy output. Numerous studies demonstrate how ML models, using vast amounts of **SCADA** data, enable real-time decision-making and operational improvements. This section reviews case studies of ML applications in wind turbine operations, focusing on key findings, model performance metrics like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), and real-world results.

Wind Turbine Reliability and Fault Prediction

One critical ML application is the prediction of mechanical and electrical faults, allowing preventive maintenance to minimize downtime and repair costs.

In a study by Heydari et al. (2021), artificial neural networks (ANNs) were used to predict failures in wind turbine gearboxes and generators. Trained on historical SCADA data (e.g., rotor speed, wind speed, temperature), the model achieved an RMSE of 0.12 and an MAE of 0.09, indicating high accuracy. Deployed in a real wind farm, the model successfully identified several gearbox faults before breakdowns, allowing operators to schedule maintenance during off-peak hours, minimizing revenue losses.

Similarly, Abd-Elwahab and Ali (2020) applied decision trees to detect operational anomalies, focusing on temperature data from turbine components. The model identified early signs of bearing wear and overheating, achieving 95% accuracy in fault detection and reducing unscheduled maintenance by 30% at an offshore wind farm over a one-year period.

Energy Output Optimization

ML models are also extensively used to optimize energy output by adjusting control settings like blade pitch and yaw alignment based on real-time SCADA data. In a case study by Zhang et al. (2024), a support vector machine (SVM) model predicted optimal blade pitch angles based on wind speed and rotor speed. Integrated into the turbine control system, the model improved energy output by 8% over six months, with an RMSE of 0.15, demonstrating accuracy in performance optimization. A more complex approach was employed by Phipps et al. (2022), who used a stacked ensemble model combining decision trees, gradient boosting machines, and deep neural networks to forecast energy production. The ensemble model, with an RMSE of 0.07 and MAE of 0.05, outperformed individual models and helped optimize energy generation strategies, leading to a 10% increase in annual output at a wind farm in China.

Real-World Applications and Model Performance

These real-world deployments highlight the practical benefits of ML models in wind farm operations. Predictive maintenance using ML can prevent costly breakdowns by identifying early warning signs, such as temperature fluctuations or abnormal vibrations. For example, Guo et al. (2021) used a random forest model to predict turbine

blade failure based on vibration data, achieving an F1-score of 0.92. This allowed operators to replace blades during scheduled maintenance, avoiding unscheduled downtime. ML models are typically evaluated using performance metrics like RMSE and MAE. Lower scores indicate higher accuracy, as these metrics measure deviations between predicted and actual values. For instance, an RMSE of 0.07, as reported by Phipps et al. (2022), indicates a minimal error, meaning the model's predictions closely match real-world outcomes.
Additionally, ML models can be integrated into turbine control systems for real-time optimization. By continuously analyzing SCADA data and adjusting turbine settings, these models not only enhance energy efficiency but also extend the lifespan of critical components like the gearbox and generator.

Challenges in ML Applications

Although machine learning (ML) offers significant advantages for improving wind turbine reliability, optimizing energy output, and enhancing predictive maintenance, several challenges limit its full adoption in the wind energy sector. These challenges revolve around data quality, model generalization, and the limited real-time application of ML in industrial wind farms. Addressing these issues is crucial for realizing ML's potential in wind energy.

Data Quality and Preprocessing

A primary challenge in applying ML to wind energy systems is ensuring the quality of data collected from Supervisory Control and Data Acquisition (SCADA) systems. SCADA data, while rich, often suffers from issues such as missing values, outliers, and noise, all of which can degrade model performance (Maseda et al., 2021).

- Missing Data: Missing data due to sensor malfunctions or communication failures is common in SCADA systems. Incomplete datasets lead to inaccurate predictions, affecting fault detection and turbine performance optimization. Data imputation, which estimates missing values, is a common preprocessing technique, but it can introduce bias, especially when large portions of data are missing (Alimi et al., 2021).
- Outliers and Noise: SCADA data frequently contains outliers, which are extreme values that can distort model predictions. These outliers may stem from sensor malfunctions or environmental factors. Proper outlier detection and removal are essential to ensure that ML models are trained on clean data (Martí-Puig et al., 2018). However, distinguishing between actual anomalies and sensor noise remains a challenge and requires advanced algorithms.

Model Generalization Issues

Another challenge in ML for wind turbines is the difficulty of generalizing models across different turbine types and wind farm sites. Wind farms often contain various turbine models, each with distinct design, operational parameters, and performance characteristics (Zhang et al., 2023). An ML model trained on data from one turbine type may not perform well on other models due to differences in data patterns and operating environments.
For instance, a model trained on onshore turbines may struggle to generalize to offshore turbines, where environmental conditions, such as wind patterns and humidity, vary significantly. This limits the applicability of ML models across different wind farms, as separate models may need to be trained for each turbine type. Researchers are exploring techniques like transfer learning, which adapts pre-trained models to new turbines with minimal additional data (Mohn et al., 2022), but real-world implementation of transfer learning is still in its early stages.

Lack of Real-Time Application

Despite the effectiveness of ML in wind energy research, its real-time application in industrial wind farms remains limited. Many ML models are trained on historical data, with predictions used for future decisions, but the dynamic nature of wind turbine operations requires real-time analysis and decision-making, which current ML frameworks often fail to provide.
Several factors hinder the real-time application of ML in wind farms:

- Computational Complexity: Deep learning and ensemble models often require substantial computational resources, making real-time processing of large SCADA datasets expensive and prone to latency, which hinders immediate fault detection or turbine optimization (Wu et al., 2021).
- Data Stream Processing: Wind turbines generate continuous streams of SCADA data at high frequencies. Real-time analysis of these data streams requires efficient algorithms capable of handling large-scale, high-velocity data. Most ML models are designed for batch processing, analyzing data retrospectively rather than as it is generated (Zhang et al., 2024).
- Integration with Control Systems: For ML models to have real-time impact, they must be integrated with turbine control systems to adjust settings such as blade pitch or yaw in response to changing conditions. However, integrating ML into existing control systems is technically complex due to the need for precise, immediate responses to environmental changes (Abdul_Azim et al., 2017). Many wind farms still rely on manual interventions, limiting the real-time impact of ML.

Integration of SCADA Data and Machine Learning for Predictive Maintenance

Predictive Maintenance in Wind Turbines

Predictive maintenance is a proactive approach that uses real-time monitoring to predict when maintenance should occur, avoiding fixed schedules or waiting for equipment failure. In wind turbines, predictive maintenance reduces downtime, extends the turbine's lifespan, and minimizes costs. By detecting early signs of mechanical or electrical faults, operators can schedule repairs during non-peak hours, preventing unexpected shutdowns (Wang et al., 2023). SCADA data is crucial for predictive maintenance, providing continuous real-time insights into parameters like wind speed, rotor speed, power output, and component temperatures. Machine learning (ML) models analyze this data to predict potential failures or maintenance needs. By identifying patterns in SCADA data, ML models can detect anomalies indicative of future issues, enabling operators to plan maintenance effectively, thus reducing unplanned outages and extending turbine life (Liu et al., 2023).

Review of ML Models for Predictive Maintenance

Various ML models have been applied successfully in predictive maintenance, each offering unique strengths in fault detection and failure prediction:

- Decision Trees: Widely used for their simplicity and interpretability, decision trees classify turbine conditions using SCADA data. For instance, a model may classify turbines as "healthy" or "at-risk" based on vibration levels or bearing temperature. Abd-Elwahab et al. (2020) used decision trees to detect bearing failures in turbines with 92% accuracy, enabling timely maintenance before severe mechanical damage occurred.

- Support Vector Machines (SVMs): SVMs are effective in handling high-dimensional data and classifying operational states by identifying optimal boundaries. Zhang et al. (2024) applied SVMs to SCADA data from offshore turbines, predicting gearbox failures with 89% precision, demonstrating SVMs' ability in early fault detection.
- Artificial Neural Networks (ANNs): ANNs are adept at detecting nonlinear patterns in large datasets, making them valuable for predicting failures in turbine components like the gearbox or blades. Heydari et al. (2021) used an ANN to analyze SCADA data, predicting generator failures with an RMSE of 0.11 and MAE of 0.09, allowing operators to prevent costly breakdowns.

Real-world applications of these models demonstrate their effectiveness. For example, a decision tree model monitored bearing temperatures in a European wind farm, predicting wear and preventing a potential shutdown (Saha et al., 2022). Similarly, an ANN model in an offshore wind farm achieved 91.98% accuracy in predicting blade failures, increasing turbine availability by 15% (Hsu et al., 2020).

Benefits and Limitations

Benefits of ML in Predictive Maintenance:

- Cost Savings**:** ML models identify faults before major mechanical failures occur, reducing emergency repairs and the financial losses associated with downtime (Vives et al., 2022).
- Increased Availability**:** Early fault detection and scheduling maintenance during non-peak hours ensure turbines operate for longer periods, boosting energy production (Phipps et al., 2022).
- Reduced Unplanned Outages**:** ML models predict faults before shutdowns, significantly reducing unexpected outages and improving reliability (Singh et al., 2023).

Limitations

- Difficulty in Predicting Rare Failures**:** ML models are effective at detecting common faults but struggle with rare events like catastrophic blade failure, which may not occur frequently enough in SCADA data for the model to learn associated patterns (Kong et al., 2023).
- Data Requirements**:** ML models rely on large volumes of high-quality data for accurate predictions. Poor-quality SCADA data, with missing values or noise, can significantly impact model performance. Effective data preprocessing is critical to ensure accurate predictions (Alimi et al., 2021).

SCADA Data-Driven Fault Detection and Energy Production Optimization

Fault Detection Models

Machine learning (ML) models have significantly improved fault detection in wind turbines by leveraging the continuous data streams provided by Supervisory Control and Data Acquisition (SCADA) systems. These systems monitor critical parameters such as temperature, rotor speed, power output, and vibration, enabling the early detection of faults before catastrophic failures occur.

Numerous studies have utilized ML models to detect faults in key turbine components. Guo et al. (2021) applied a random forest algorithm to SCADA data, focusing on temperature, rotor speed, and power output to detect generator faults. The model achieved 93% accuracy, identifying abnormal temperature increases as early indicators of generator failure, allowing preventive maintenance and avoiding downtime. Gearbox failures, which can be hard to detect early, are another common issue. Abd-Elwahab et al. (2019) used decision tree models to predict early gearbox wear using rotor speed and vibration data, effectively preventing operational disruptions. These models detect subtle deviations in SCADA data that may go unnoticed by human operators. Other ML models have focused on turbine blades and bearings. Phipps et al. (2021) employed a support vector machine (SVM) model to monitor blade vibrations, achieving 92% precision in detecting potential failures. By continuously monitoring SCADA data, the model identified small cracks and material fatigue, enabling operators to schedule timely repairs.

Energy Production Optimization

SCADA data, combined with ML models, has also been instrumental in optimizing wind energy production. By processing real-time data, ML models adjust parameters like blade pitch, yaw alignment, and rotor speed to maximize energy output in fluctuating wind conditions.

Energy forecasting models are a key area of research. Gomez and Lundquist (2020) used a neural network to predict wind energy production, analyzing historical SCADA data such as wind speed and rotor speed. Their model helped optimize turbine performance by predicting future energy output, resulting in better energy efficiency.

Malakouti et al. (2022) applied gradient boosting machines (GBMs) to optimize turbine yaw alignment based on SCADA data, improving energy capture by adjusting the yaw to align with wind direction. Similarly, Abd-Elwahab et al. (2020) demonstrated how artificial neural networks (ANNs) could adjust blade pitch in real-time using wind speed and rotor dynamics data, improving overall energy output by 5% compared to traditional systems.

Integration with Control Systems

Integrating ML models with turbine control systems enables real-time optimization of wind energy production and fault detection. SCADA data provides the continuous stream needed to make immediate adjustments based on changing conditions. ML models can be integrated into control systems to monitor turbine performance and adjust parameters such as rotor speed, blade pitch, and yaw alignment. For instance, Zhang et al. (2024) implemented a stacked ensemble model that combined decision trees, SVMs, and neural networks to optimize turbine performance. By adjusting blade pitch in response to wind fluctuations, the model improved energy efficiency and reduced mechanical stress. The integration of ML models also aids in avoiding real-time faults. Guo et al. (2021) developed a reinforcement learning (RL) model that learned optimal control strategies based on real-time SCADA data. The RL model continuously adjusted rotor speed and yaw alignment to minimize mechanical stress and maximize energy output, reducing component wear and enhancing reliability. While the benefits of integrating ML models into control systems are clear, challenges remain, such as the need for low-latency data processing and high computational power. Advances in edge computing and cloud-based data processing have made real-time optimization more feasible, and as the technology evolves, more wind farms are expected to adopt these systems, further improving efficiency and reliability (Veena et al., 2020).

Challenges and Future Directions

As machine learning (ML) applications in wind energy expand, several challenges must be addressed to maximize their potential. These challenges are primarily related to data quality, model generalization, and real-time implementation. Additionally, potential future research directions could lead to more robust, scalable, and efficient ML models for wind energy management.

Data Quality and Management

One of the major challenges in applying ML to wind energy is ensuring high data quality. SCADA data, while rich in information, often suffers from issues such as missing values, noisy measurements, and corrupted data due to sensor malfunctions or communication failures. These problems can severely impact the reliability and accuracy of ML models, resulting in poor fault detection or inaccurate energy predictions (Sihua et al., 2020).

- Missing Data: Incomplete datasets present a significant challenge, as ML models rely on continuous data streams to learn accurate patterns. Missing data can cause biased models that do not generalize well. Techniques like data imputation can mitigate this issue, but they are not always reliable (Hu & Du, 2020). A clean, high-resolution SCADA dataset with minimal missing data is crucial for effective model training.
- Noisy and Corrupted Data: Noisy data, often caused by environmental factors or sensor degradation, can introduce errors that lead to incorrect predictions. Effective data preprocessing methods, such as outlier detection, smoothing, and filtering, are essential to remove noise and improve data quality. Redundant sensors can also help cross-validate measurements, reducing the impact of faulty readings (Yang et al., 2018).

Future research should focus on developing automated real-time data cleaning systems to ensure high-quality datasets for predictive maintenance and energy optimization.

Model Generalization Across Wind Farms

A key challenge in ML applications for wind energy is the lack of model generalization across different wind farms and turbine models. ML models trained on SCADA data from one wind farm may perform poorly when applied to another due to variations in geography, turbine types, and operational practices (Yang et al., 2024).

- Geographic and Environmental Differences: Wind farms face unique environmental conditions, such as different wind patterns, temperature, and humidity. These factors affect turbine performance, making it difficult for models trained in one region to generalize to another.
- Turbine Model Variability: Wind farms often use turbines from different manufacturers, each with distinct designs, operational parameters, and SCADA structures. This complicates model transfer between turbine types (Kong et al., 2019).

To improve generalization, researchers are exploring transfer learning, which adapts models trained on one dataset to another with minimal retraining. Multi-task learning is another approach that involves training a model on multiple datasets to improve its generalization across various turbines and wind farms.

Real-Time Implementation of Machine Learning

The real-time application of ML in operational wind farms presents several technical and computational challenges. While many ML models have proven effective in analyzing historical SCADA data, real-time implementation demands low-latency processing and integration with turbine control systems.

- Computational Complexity: Many advanced ML models, such as deep learning and ensemble methods, require significant computational resources. Real-time analysis demands fast data processing, which can be difficult in large wind farms with high-frequency data generation. Solutions such as edge computing and distributed processing architectures can reduce latency (Kumar et al., 2018).

- Real-Time Data Stream Processing: Most current ML models are optimized for batch processing rather than continuous real-time data streams. Future research should focus on developing streaming ML algorithms that can process and analyze data as it is generated, allowing for immediate adjustments to turbine operations (Shabbir et al., 2022).
- Integration with Control Systems: Integrating ML models with turbine control systems is crucial for real-time optimization. Models must adjust operational parameters such as rotor speed and blade pitch in response to real-time data. Developing robust integration frameworks that allow seamless ML model integration with control systems is a critical research area (Zhang et al., 2020).

Future Research Directions

Several promising research directions could improve the effectiveness of ML in wind energy management:

- Hybrid Models: Combining physical models of turbine dynamics with data-driven ML models could enhance accuracy in fault prediction and energy optimization. Physical models offer insights into turbine mechanics, while ML models identify patterns that are hard to model explicitly.
- Better Fault Prediction Algorithms: Developing more accurate algorithms to predict rare failure events, such as catastrophic turbine blade failures, is essential. Current models often struggle due to insufficient data on rare events. Anomaly detection and generative models could address this issue.
- Improved Integration with Wind Farm Software: ML models should be designed to integrate with existing wind farm management systems for seamless real-time analytics and automated decision-making. This would enable operators to leverage ML models' predictive power in daily operations.
- Explainable AI: As ML models become more complex, the need for explainable AI (XAI) grows. Operators need insights into why certain faults are predicted or why specific operational adjustments are recommended. XAI approaches can improve transparency and foster greater trust in AI-driven decision-making.

Conclusion

This review has highlighted the critical role of Supervisory Control and Data Acquisition (SCADA) data and machine learning (ML) models in improving wind turbine reliability and optimizing energy production. SCADA data, by monitoring parameters such as rotor speed, power output, and temperature, provides essential insights for early fault detection and real-time performance optimization. Through case studies, we have demonstrated that machine learning models, particularly ensemble techniques such as bagging, boosting, and stacking, significantly enhance fault detection, predictive maintenance, and energy optimization in wind farms. SCADA-driven wind turbine management has become a cornerstone of modern wind farm operations. Continuous, real-time data from SCADA systems allows operators to detect potential faults early, plan timely maintenance, and adjust turbine settings for optimal energy production. SCADA data's role in fault detection, especially for detecting gearbox failures, generator faults, and blade damage, has helped reduce costly downtimes and extend turbine lifespans. Machine learning models are set to transform the wind energy sector by enabling predictive analytics and real-time turbine optimization. Among the techniques reviewed, ensemble learning has been particularly effective. Bagging reduces variance and improves prediction stability, boosting enhances weak models, and stacking combines the strengths of multiple models for higher accuracy. When integrated with SCADA data, these models offer a robust approach to enhancing turbine performance, reliability, and energy efficiency. While the future of ML in wind energy is promising, several challenges remain. Future research should focus on improving data quality through automated preprocessing, developing more generalizable models that work across different wind farms and turbine types, and enabling real-time implementation of ML models. Furthermore, hybrid models combining physical turbine dynamics with data-driven approaches hold potential for even more accurate and scalable solutions. In conclusion, integrating SCADA data with machine learning represents a transformative step in wind farm management. As ongoing research addresses challenges like data quality, model generalization, and real-time application, the adoption of these advanced analytics will lead to more efficient, reliable, and sustainable wind energy systems.

Declaration

Acknowledgment: The authors would like to extend their appreciation to their academic institutions and colleagues for their support during the writing and review process of this article. Special thanks go to the reviewers for their valuable feedback and suggestions, which greatly contributed to improving the quality of this work.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors conducted the research independently, with support from their academic institutions.

Conflict of Interest: The authors declare no conflicts of interest in relation to the content of this article.

Ethics Approval/Declaration: As this is a review article, no specific ethics approval was required.

Consent to Participate: Not applicable, as this research did not involve human participants.

Consent for Publication: All authors consent to the publication of this article and agree with the content and findings presented.

Data Availability: As this is a review article, no new data were generated or analyzed during this study. The review is based on publicly available sources and referenced research articles.

Authors' Contributions: Hlaing Min Soe**:** Authored and prepared the manuscript, collected and synthesized the relevant literature, and contributed to the overall structure and content of the article. Dr. Arkar Htet**:** Supervised the research, provided overall guidance, reviewed the manuscript, and offered critical revisions to enhance the quality of the article.

References

Abd-Elwahab, K. T. & Ali, H. (2020). Scada data as a powerful tool for early fault detection in wind turbine gearboxes. Wind Engineering, 45(5), 1317-1326. https://doi.org/10.1177/0309524x20969418.

Abdul_Azim, O., ldmerdash, S., Elshazly, K., & bdellatif, O. (2017). Experimental investigation for horizontal wind turbine of direct-drive. Benha Journal of Applied Sciences, 2(1), 113-124. https://doi.org/10.21608/bjas.2017.164017.

Alimi, O. A., Ouahada, K., Abu-Mahfouz, A. M., Rimer, S., & Alimi, K. O. A. (2021). A review of research works on supervised learning algorithms for scada intrusion detection and classification. Sustainability, 13(17), 9597. https://doi.org/10.3390/su13179597.

Alkesaiberi, A., Harrou, F., & Sun, Y. (2022). Efficient wind power prediction using machine learning methods: a comparative study. Energies, 15(7), 2327. https://doi.org/10.3390/en15072327.

Annoni, J., Bay, C. J., Johnson, K., Dall'Anese, E., Quon, E., Kemper, T., & Fleming, P. (2019). Wind direction estimation using scada data with consensus-based optimization. Wind Energy Science, 4(2), 355-368. https://doi.org/10.5194/wes-4-355-2019.

Arch, D., Guo, Y., Sheng, S., Phillips, C., & Williams, L. (2020). Prognosis of wind turbine gearbox bearing failures using scada and modeled data. Annual Conference of the PHM Society, 12(1), 10. https://doi.org/10.36001/phmconf.2020.v12i1.1292.

Babawarun, T., Ewim, D. R. E., Scott, T. O., & Neye-Akogo, C. (2023). A comprehensive review of wind turbine modeling for addressing energy challenges in nigeria and south africa in the 4ir context. The Journal of Engineering and Exact Sciences, 9(2), 15479-01e. https://doi.org/10.18540/jcecvl9iss2pp15479-01e.

Becker-Dombrowsky, F. M., Koplin, Q. S., & Kirchner, E. (2023). Individual feature selection of rolling bearing impedance signals for early failure detection. Lubricants, 11(7), 304. https://doi.org/10.3390/lubricants11070304.

Biazar, D., Khaloozadeh, H., & Siahi, M. (2022). Sensitivity analysis for evaluation of the effect of sensors error on the wind turbine variables using monte carlo simulation. IET Renewable Power Generation, 16(8), 1623-1635. https://doi.org/10.1049/rpg2.12468.

Bonacina, F., Miele, E. S., & Corsini, A. (2022). On the use of artificial intelligence for condition monitoring in horizontal-axis wind turbines. IOP Conference Series: Earth and Environmental Science, 1073(1), 012005. https://doi.org/10.1088/1755-1315/1073/1/012005.

Byrne, R., Astolfi, D., Castellani, F., & Hewitt, N. (2020). A study of wind turbine performance decline with age through operation data analysis. Energies, 13(8), 2086. https://doi.org/10.3390/en13082086.

Carroll, J. D., McDonald, A., & McMillan, D. (2015). Failure rate, repair time and unscheduled o&m cost analysis of offshore wind turbines. Wind Energy, 19(6), 1107-1119. https://doi.org/10.1002/we.1887.

Castellani, F., Garinei, A., Terzi, L., Astolfi, D., & Gaudiosi, M. (2014). Improving windfarm operation practice through numerical modelling and supervisory control and data acquisition data analysis. IET Renewable Power Generation, 8(4), 367-379. https://doi.org/10.1049/iet-rpg.2013.0182.

Chen, H. C. (2022). Early fault warning method of wind turbine main transmission system based on scada and cms data. Machines, 10(11), 1018. https://doi.org/10.3390/machines10111018.

Chen, J. (. (2023). Research on fault diagnosis and health monitoring of wind power equipment. Fourth International Conference on Artificial Intelligence and Electromechanical Automation (AIEA 2023) (p. 4942. https://doi.org/10.1117/12.2684942). Nanjing: SPIE.

Chen, S., Qiao, Z., Xie, C., Zhu, R., Lai, Z., Xu, X., & Zhou, S. (2023). Transfer condition monitoring across wind turbines using feature alignment and parameter fine-tuning. Measurement Science and Technology, 34(12), 125901. https://doi.org/10.1088/1361-6501/aced5e.

Choquehuanca, E. & Ortega, A. (2023). A reinforcement learning approach to the optimal torque mppt problem in wind turbines. Journal of Physics: Conference Series, 2538(1), 012005. https://doi.org/10.1088/1742-6596/2538/1/012005.

Donadio, L., Fang, J., & Porté-Agel, F. (2021). Numerical weather prediction and artificial neural network coupling for wind energy forecast. Energies, 14(2), 338. https://doi.org/10.3390/en14020338.

Feddaoui, A., Farah, L., Benretem, A., Djehaf, M. A., Bouguerra, Z., & Djeriri, Y. (2023). Enhancing power control in doubly fed induction generator based wind turbines: a performance comparison of fuzzy logic and sliding mode control methods. The Scientific Bulletin of Electrical Engineering Faculty, 23(2), 42-50. https://doi.org/10.2478/sbeef-2023-0018.

Galadima, H., Anson-Dwamena, R., Johnson, A., Bello, G., Adunlin, G., & Blando, J. (2024). Machine learning as a tool for early detection: a focus on late-stage colorectal cancer across socioeconomic spectrums. Cancers, 16(3), 540. https://doi.org/10.3390/cancers16030540.

Gomez, M. S. & Lundquist, J. K. (2020). The effect of wind direction shear on turbine performance in a wind farm in central iowa. Wind Energy Science, 5(1), 125-139. https://doi.org/10.5194/wes-5-125-2020.

Gorel, G. & Abdi, M. O. (2023). Advanced pitch angle control based on genetic algorithm and particle swarm optimisation on fast turbine systems. Elektronika Ir Elektrotechnika, 29(4), 11-18. https://doi.org/10.5755/j02.eie.34205.

Guo, L., Wang, C., Zhang, D., & Yang, G. M. (2021). An improved feature selection method based on random forest algorithm for wind turbine condition monitoring. Sensors, 21(16), 5654. https://doi.org/10.3390/s21165654.

Hajjaj, S. S. H., Aqeel, A. A. K. A., Sultan, M. T. H., Shahar, F. S., & Shah, A. U. M. (2022). Review of recent efforts in cooling photovoltaic panels (pvs) for enhanced performance and better impact on the environment. Nanomaterials, 12(10), 1664. https://doi.org/10.3390/nano12101664.

Heydari, A., Garcia, D. A., Fekih, A., Keynia, F., Tjernberg, L. B., & Santoli, L. d. (2021). A hybrid intelligent model for the condition monitoring and diagnostics of wind turbines gearbox. IEEE Access, 9, 89878-89890. https://doi.org/10.1109/access.2021.3090434.

Hsu, G. J. Y., Wang, Y., Lin, K., Chen, M., & Hsu, J. H. (2020). Wind turbine fault diagnosis and predictive maintenance through statistical process control and machine learning. IEEE Access, 8, 23427-23439. https://doi.org/10.1109/access.2020.2968615.

Hu, Z. & Du, D. (2020). A new analytical framework for missing data imputation and classification with uncertainty: missing data imputation and heart failure readmission prediction. Plos One, 15(9), e0237724. https://doi.org/10.1371/journal.pone.0237724.

Hui, H., Zhu, Q., Xiao-hong, Z., & Zhang, J. (2023). An adaptive, data-driven stacking ensemble learning framework for the short-term forecasting of renewable energy generation. Energies, 16(4), 1963. https://doi.org/10.3390/en16041963.

Jiang, X., Li, M., & Cheng, L. (2024). Application of decision tree and machine learning in new energy vehicle maintenance decision making. Applied Mathematics and Nonlinear Sciences, 9(1), 1463. https://doi.org/10.2478/amns-2024-1463.

Kanev, S. (2020). Dynamic wake steering and its impact on wind farm power production and yaw actuator duty. Renewable Energy, 146, 9-15. https://doi.org/10.1016/j.renene.2019.06.122.

Kim, M. & Kim, H. (2024). A dynamic analysis data preprocessing technique for malicious code detection with tf-idf and sliding windows. Electronics, 13(5), 963. https://doi.org/10.3390/electronics13050963.

Kong, L., Liang, H., Liu, G., & Liu, S. (2023). Research on wind turbine fault detection based on the fusion of asl-catboost and ttrsa. Sensors, 23(15), 6741. https://doi.org/10.3390/s23156741.

Kong, L., Liang, L., Xu, J., Zhang, W., & Zhu, W. (2019). The optimization of pricing strategy for the wind power equipment aftermarket service. Industrial Management &Amp; Data Systems, 119(3), 521-546. https://doi.org/10.1108/imds-01-2018-0032.

Kumar, A., Kumar, A., & Singh, B. (2018). A critical analysis of ensemble learning techniques for improved accuracy in sales forecasting using machine learning. International Journal of Applied Research, 4(8), 49-52. https://doi.org/10.22271/allresearch.2018.v4.i8a.11449.

Lauritsen, S. M., Thiesson, B., Jørgensen, M. J., Riis, A. H., Espelund, U., Weile, J., & Lange, J. (2021). The framing of machine learning risk prediction models illustrated by evaluation of sepsis in general wards. NPJ Digital Medicine, 4(1), 00529. https://doi.org/10.1038/s41746-021-00529-x.

Leahy, K., Hu, R., Konstantakopoulos, I. C., Spanos, C. J., Agogino, A. M., & O'Sullivan, D. (2020). Diagnosing and predictingwind turbine faults from scada data using support vector machines. International Journal of Prognostics and Health Management, 9(1), 2692.https://doi.org/10.36001/ijphm.2018.v9i1.2692.

Lebranchu, A., Charbonnier, S., Bérenguer, C., & Prevost, F. (2019). A combined mono- and multi-turbine approach for fault indicator synthesis and wind turbine monitoring using scada data. ISA Transactions, 87, 272-281. https://doi.org/10.1016/j.isatra.2018.11.041.

Lee, G., Ding, Y., Genton, M. G., & Xie, L. (2015). Power curve estimation with multivariate environmental factors for inland and offshore wind farms. Journal of the American Statistical Association, 110(509), 56-67. https://doi.org/10.1080/01621459.2014.977385.

Li, H., Yang, J., Zhang, M., Guo, S., Lv, W., & Liu, Z. (2015). A method based on artificial neural network to estimate the health of wind turbine. The 27th Chinese Control and Decision Conference (2015 CCDC), (p. 7162050. https://doi.org/10.1109/ccdc.2015.7162050).

Liu, Y., Cheng, H., Kong, X., Wang, Q., & Cui, H. (2019). Intelligent wind turbine blade icing detection using supervisory control and data acquisition data and ensemble deep learning. Energy Science & Engineering, 7(6), 2633-2645. https://doi.org/10.1002/ese3.449.

Liu, Y., Yan, Z., Ma, Z., & Wu, C. (2023). Wind turbine spindle operating state recognition and early warning driven by scada data. Energy Engineering, 120(5), 1223-1237. https://doi.org/10.32604/ee.2023.026329.

Lyons, J. T. & Göçmen, T. (2021). Applied machine learning techniques for performance analysis in large wind farms. Energies, 14(13), 3756. https://doi.org/10.3390/en14133756.

Malakouti, S. m., Ghiasi, A. R., Ghavifekr, A. A., & Emami, P. (2022). Predicting wind power generation using machine learning and cnn-lstm approaches. Wind Engineering, 46(6), 1853-1869. https://doi.org/10.1177/0309524x221113013.

Martí-Puig, P., Blanco-M, A., Cárdenas, J. J., Cusidó, J., & Solé-Casals, J. (2018). Effects of the pre-processing algorithms in fault diagnosis of wind turbines. Environmental Modelling & Software, 110, 119-128. https://doi.org/10.1016/j.envsoft.2018.05.002.

Maseda, F., López, I. G., Martija, I., & Alkorta, P. (2021). Sensors data analysis in supervisory control and data acquisition (scada) systems to foresee failures with an undetermined origin. Sensors, 21(8), 2762. https://doi.org/10.3390/s21082762.

Menezes, D., Mendes, M., Almeida, J. A., & Farinha, J. T. (2020). Wind farm and resource datasets: a comprehensive survey and overview. Energies, 13(18), 4702. https://doi.org/10.3390/en13184702.

Mienye, I. D. & Sun, Y. (2022). A survey of ensemble learning: concepts, algorithms, applications, and prospects. IEEE Access, 10, 99129-99149. https://doi.org/10.1109/access.2022.3207287.

Mohn, S. F., Law, M., Koleva, M. K., Lee, B., Berg, A., Murray, N., & Parker, W. (2022). Machine learning model for chest radiographs: using local data to enhance performance. Canadian Association of Radiologists Journal, 74(3), 548-556. https://doi.org/10.1177/08465371221145023.

Moyón, L., Encalada-Dávila, A., Tutivén, C., & Vidal, Y. (2022). Early detection of main bearing damage in wind turbines. Renewable Energy and Power Quality Journal, 20, 773-777. https://doi.org/10.24084/repqj20.430.

Muhammad Amri, H. G. B., Marramiero, D., Singh, D., Van Wingerden, J., & Viré, A. (2024). Data-driven time series forecasting of offshore wind turbine loads. Journal of Physics: Conference Series, 2767(5), 052060. https://doi.org/10.1088/1742-6596/2767/5/052060.

Müller, D. & Soto-Rey, I. (2022). An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks. IEEE Access, 10, 66467-66480. https://doi.org/10.1109/access.2022.3182399.

Murgia, A., Verbeke, R., Tsiporkova, E., Terzi, L., & Astolfi, D. (2023). Discussion on the suitability of scada-based condition monitoring for wind turbine fault diagnosis through temperature data analysis. Energies, 16(2), 620. https://doi.org/10.3390/en16020620.

Pandit, R. & Wang, J. (2014). A comprehensive review on enhancing wind turbine applications with advanced scada data analytics and practical insights. IET Renewable Power Generation, 18(4), 722-742. https://doi.org/10.1049/rpg2.12920.

Peter, R., Zappalá, D., Schamboeck, V., & Watson, S. J. (2022). Wind turbine generator prognostics using field scada data. Journal of Physics: Conference Series, 2265(3), 032111. https://doi.org/10.1088/1742-6596/2265/3/032111.

Phipps, K., Lerch, S., Andersson, M., Mikut, R., Hagenmeyer, V., & Ludwig, N. (2022). Evaluating ensemble post-processing for wind power forecasts. Wind Energy, 25(8), 1379-1405. https://doi.org/10.1002/we.2736.

Ponkumar, G., Jayaprakash, S., Ramasamy, D., & Priyasivakumar, A. (2024). Particle swarm optimization-extreme learning machine model combined with the ada boost algorithm for short-term wind power prediction. International Journal of Power Electronics and Drive Systems (IJPEDS), 15(2), 1211. https://doi.org/10.11591/ijpeds.v15.i2.pp1211-1217.

Razavi-Termeh, S. V., Sadeghi-Niaraki, A., & Choi, S. (2021). Spatial modeling of asthma-prone areas using remote sensing and ensemble machine learning algorithms. Remote Sensing, 13(16), 3222. https://doi.org/10.3390/rs13163222.

Renn, P. I. & Gharib, M. (2022). Machine learning for flow-informed aerodynamic control in turbulent wind conditions. Communications Engineering, 1(1), 44172. https://doi.org/10.1038/s44172-022-00046-z.

Saenz-Aguirre, A., Zulueta, E., Ulazia, A., & Teso-Fz-Betoño, D. (2019). Performance enhancement of the artificial neural network–based reinforcement learning for wind turbine yaw control. Wind Energy, 23(3), 676-690. https://doi.org/10.1002/we.2451.

Saha, D. K., Hoque, M. E., & Badihi, H. (2022). Development of intelligent fault diagnosis technique of rotary machine element bearing: a machine learning approach. Sensors, 22(3), 1073. https://doi.org/10.3390/s22031073.

Santolamazza, A., Dadi, D., & Introna, V. (2021). A data-mining approach for wind turbine fault detection based on scada data analysis using artificial neural networks. Energies, 14(7), 1845. https://doi.org/10.3390/en14071845.

Seifert, J. K., Kraft, M., Kühn, M., & Lukassen, L. J. (2021). Correlations of power output fluctuations in an offshore wind farm using high-resolution scada data. Wind Energy Science, 6(4), 997-1014. https://doi.org/10.5194/wes-6-997-2021.

Shabbir, N., K, L., Jawad, M., Husev, O., Rehman, A. U., Gardezi, A. A., & Choi, J. (2022). Short-term wind energy forecasting using deep learning-based predictive analytics. Computers, Materials & Continua, 72(1), 1017-1033. https://doi.org/10.32604/cmc.2022.024576.

Shittu, S., Li, G., Akhlaghi, Y. G., Ma, X., Zhao, X., & Ayodele, E. (2019). Advancements in thermoelectric generators for enhanced hybrid photovoltaic system performance. Renewable and Sustainable Energy Reviews, 109, 24-54. https://doi.org/10.1016/j.rser.2019.04.023.

Sierra-García, J. E. & Santos, M. (2020). Exploring reward strategies for wind turbine pitch control by reinforcement learning. Applied Sciences, 10(21), 7462. https://doi.org/10.3390/app10217462.

Sihua, W., Wang, T., Zhou, S. L., Chen, T., & Wang, Y. (2020). Fault diagnosis of pitch system of wind turbine based on improved stacked auto-encoder network. Electrotehnica, Electronica, Automatica, 68(3), 47-54. https://doi.org/10.46904/eea.20.68.3.1108006.

Singh, E., Afshari, S. S., & Liang, X. (2023). Wind turbine optimal preventive maintenance scheduling using fibonacci search and genetic algorithm. Journal of Dynamics, Monitoring and Diagnostics. , 158. https://doi.org/10.37965/jdmd.2023.158.

Song, X., Wang, L., & Luo, X. (2022). Airfoil optimization using a machine learning-based optimization algorithm. Journal of Physics: Conference Series, 2217(1), 012009. https://doi.org/10.1088/1742-6596/2217/1/012009.

Uchida, T. & Sugitani, K. (2020). Numerical and experimental study of topographic speed-up effects in complex terrain. Energies, 13(15), 3896. https://doi.org/10.3390/en13153896.

Ulmer, M., Jarlskog, E., Pizza, G., & Huber, L. G. (2020). Cross-turbine training of convolutional neural networks for scada-based fault detection in wind turbines. Annual Conference of the PHM Society, 12(1), 10. https://doi.org/10.36001/phmconf.2020.v12i1.1205.

Veena, R., Manuel, S. M., Mathew, M. S., & Petra, M. I. (2020). Wake induced power losses in wind farms. International Journal of Engineering and Advanced Technology, 9(3), 2175-2180. https://doi.org/10.35940/ijeat.c5611.029320.

Verma, A., Zappalá, D., Sheng, S., & Watson, S. J. (2022). Wind turbine gearbox fault prognosis using high-frequency scada data. Journal of Physics: Conference Series, 2265(3), 032067. https://doi.org/10.1088/1742-6596/2265/3/032067.

Vidal, Y., Puruncajas, B., Castellani, F., & Tutivén, C. (2023). Predictive maintenance of wind turbine's main bearing using wind farm scada data and lstm neural networks. Journal of Physics: Conference Series, 2507(1), 012024. https://doi.org/10.1088/1742-6596/2507/1/012024.

Vives, J., Palací, J., & Heart, J. (2022). Framework for bidirectional knowledge-based maintenance of wind turbines. Computational Intelligence and Neuroscience, 2022, 1-6. https://doi.org/10.1155/2022/1020400.

Wang, C., Zhang, X., & Luo, T. (2023). A fault diagnosis approach for wind turbine gearbox based on ensemble learning model and dung beetle optimization algorithm. Journal of Physics: Conference Series, 2661(1), 012009. https://doi.org/10.1088/1742-6596/2661/1/012009.

Wang, H., Wang, H., Jiang, G., Wang, Y., & Ren, S. (2020). A multiscale spatio-temporal convolutional deep belief network for sensor fault detection of wind turbine. Sensors, 20(12), 3580. https://doi.org/10.3390/s20123580.

Wei, L., Zheng, Q., Pei, Y., & Wang, J. (2021). Wind turbine fault diagnosis by the approach of scada alarms analysis. Applied Sciences, 12(1), 69. https://doi.org/10.3390/app12010069.

Wei, L., Zheng, Q., Zareipour, H., & Zhang, F. (2021). Comprehensive aging assessment of pitch systems combining scada and failure data. IET Renewable Power Generation, 16(1), 198-210. https://doi.org/10.1049/rpg2.12281.

Wu, A., Harrou, F., Dairi, A., & Sun, Y. (2022). Machine learning and deep learning-driven methods for predicting ambient particulate matters levels: a case study. Concurrency and Computation: Practice and Experience, 34(19), 7035https://doi.org/10.1002/cpe.7035.

Wu, W., Harrou, F., Bouyeddou, B., Senouci, S., & Sun, Y. . (2021). A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems. Cluster Computing, 25(1), 561-578. https://doi.org/10.1007/s10586-021-03426-w.

Wu, X., Wang, H., Jiang, G., Xie, P., & Li, X. (2019). Monitoring wind turbine gearbox with echo state network modeling and dynamic threshold using scada vibration data. Energies, 12(6), 982. https://doi.org/10.3390/en12060982.

Xu, G., Yu, W., & Kim, T. (2022). Wind turbine load estimation using machine learning and transfer learning. Journal of Physics: Conference Series, 2265(3), 032108. https://doi.org/10.1088/1742-6596/2265/3/032108.

Yan, Y., Li, J., & Gao, D. W. (2014). Condition parameter modeling for anomaly detection in wind turbines. Energies, 7(5), 3104-3120. https://doi.org/10.3390/en7053104.

Yang, J., Drake, T. M., Damianou, A., & Maarek, Y. (2018). Leveraging crowdsourcing data for deep active learning an application. Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18, 23-32. https://doi.org/10.1145/3178876.3186033.

Yang, J., Wen, S., & Lin, J. . (2024). A new power prediction model for wind power generation. Journal of Physics: Conference Series, 2785(1), 012067. https://doi.org/10.1088/1742-6596/2785/1/012067.

Zhang, B., Luo, W., Luo, Z., Xu, J., & Wang, L. (2023). Effectiveness of wake control optimization for multiple in-line wind turbines by combinatorial machine learning wake model. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 238(5), 1280-1294. https://doi.org/10.1177/09544062231185516.

Zhang, F., Wen, Z., Liu, D., Jiao, J., Wan, H., & Zeng, B. (2020). Calculation and analysis of wind turbine health monitoring indicators based on the relationships with scada data. Applied Sciences, 10(1), 410. https://doi.org/10.3390/app10010410.

Zhang, S., Robinson, E., & Basu, M. (2024). Wind turbine condition monitoring based on three fitted performance curves. Wind Energy, 27(5), 429-446. https://doi.org/10.1002/we.2859.

Zhang, Z. & Kusiak, A. (2012). Monitoring wind turbine vibration based on scada data. Journal of Solar Energy Engineering, 134(2), 5753. https://doi.org/10.1115/1.4005753.

Zhou, H. & Lv, X. (2013). Efficiency evaluation of investment about china's wind farms based on dea model. . Applied Mechanics and Materials, 448(453), 1829-1834. https://doi.org/10.4028/www.scientific.net/amm.448-453.1829.