Anomaly Detection and Diagnosis of Wind Turbines Using Deep Learning Techniques: Aeolian Wind Speed Case Study

Abstract. In recent years, wind turbine condition monitoring based on Supervisory Control and Data Acquisition (SCADA) systems has attracted considerable scientific research interest. Frequently reported challenges include the fact that most wind turbine SCADA parameters are highly dependent on the operating conditions, such as wind speed, wind direction, and LV Active Power, along with the control actions imposed on the wind turbine. Thus, strict and effective data quality control of the SCADA data is crucial. Besides that, intelligent anomaly detection for wind turbines using artificial intelligence techniques has been extensively researched and yielded significant results. Likewise, the usage of machine/deep learning techniques is widely spread and has been implemented in the wind industry in the last few years. The development of sophisticated deep learning now allows improvements in anomaly detection from historical data. In this paper, we present a wind speed anomaly detection approach using LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), and GRU (Gated Recurrent Unit) to detect the minimum and maximum values of wind speed. The approach applies the information in supervisory control and data acquisition systems of Aeolian wind speed. This comprehensive approach offers a promising avenue for the precise anomaly detection of Aeolian wind speed, providing practicians with a reliable tool for accurate diagnosis, critical for timely intervention. Real cases from a wind farm have confirmed the feasibility and advancement of the proposed Deep Learning models, while also discussing the effects of various applied parameters.

Keywords: Wind Energy, Wind Turbine, Anomaly Detection, SCADA Dataset, Deep Learning, CNN, LSTM, GRU.

1. Introduction

In recent years, investments in green energy have yielded large growth in the energy sector both in terms of economy and research opportunities. With its renewable and clean characteristics, wind energy has demonstrated its competitiveness among all kinds of energies [1]. Besides that, wind energy is one of the most important renewable energy sources and has gained much attention due to the recent energy crisis. Moreover, many countries use wind turbines to produce electricity, considered a clean energy technology, and a friendly alternative to fossil fuels to minimize carbon footprints. Wind energy is possibly one of the game-changers in future decarbonization scenarios in Algeria. According to [2], renewable energy sources namely Wind Energy (WE) are projected to grow substantially in the coming decades and play major roles in achieving energy sustainability. Wind energy, or wind power, is created using a wind turbine, a device that channels the power of the wind to generate electricity. These contemporary marvels come in various sizes and forms, and each is made to collect the wind's kinetic energy effectively. Wind turbines have been widely deployed to convert wind energy into electricity. As a result, numerous cutting-edge data-driven technologies have begun to find applications in the wind farm life cycle, with the operation and maintenance cycle receiving special attention due to its high cost and complexity. According to [3], the condition monitoring of wind turbines is of increasing importance as the size and remote locations of wind turbines used nowadays make the technical availability of the turbine very crucial. For [4; 5], various types of condition monitoring sensors are installed in different wind turbine (WT) components, and their multi-dimensional state parameters, such as wind speed are recorded and saved by the WT supervisory control and data acquisition (SCADA) system. Moreover, the SCADA (Supervisory Control and Data Acquisition) system accumulates a large amount of data that contains the health conditions of the wind turbines. Many methods have been proposed to optimize wind farm power output control and predictive maintenance. In wind turbines, anomaly detection can be considered a good classification problem where the task is to label the incoming data as either healthy or unhealthy (due to anomalies). Anomaly detection is used in several areas, for example, in security areas to detect malicious behavior such as intrusion or fraud, and in industrial areas to detect anomalies occurring in manufacturing processes. Anomaly detection typically involves dataset classification by using Artificial intelligence models. Moreover, the world is experiencing considerable technological advances in the renewable energy sector thanks to Artificial Intelligence (AI) [6]. Deep learning (DL), considered the primary means to achieve AI, is to provides modelling rules to a computer system to gain information from data without explicit human programming [7]. Deep Learning is a machine learning paradigm based on deep neural networks that has shown great success in various applications over recent years. Deep learning techniques have demonstrated themselves as a prominent field of study within a data-driven framework over the last decade by addressing numerous challenging problems in healthcare [24], sentiment analysis [5], opinion mining, malware detection [10], and other real-world applications. Moreover, DL has been increasingly used for data analyses and for gaining additional knowledge from data (e.g., Prediction and Detection of Anomalies) [8]. Various deep learning methods have been developed in the literature to predict wind turbines in recent decades. The deep learning method has been used in the field of renewable energy since it provides a feasible method for not only linear correlations but also nonlinear dynamic prediction and correlation processes. For example, an intelligent anomaly detection method based on deep learning networks has been receiving increasing attention. In other words, various AI solutions have been proposed to predict and detect turbine faults, assisting in diagnosis and allowing technicians to determine when to perform preventive maintenance. According to [9], large wind turbines have a higher failure rate compared to thermal and hydroelectric turbines due to the challenging external environment and complex operating conditions. According to [10], Deep ML solutions have become popular in different fields, and their application in wind turbines has obtained promising results.

Combined with the characteristics of wind turbines' SCADA data, this paper proposes deep learning techniques (CNN, LSTM, and GRU) for anomaly detection of wind turbines (Aeolian Wind Speed). For this study, previously unlabeled SCADA data from wind turbines was used to pre-train these deep learning models to extract implied features. As per our objective, we present CNN, LSTM, and GRU models for the prediction of minimum and maximum wind speed values, as well as the prediction of power output. In this way, we generate the best anomaly detection of wind speed by comparing the results of three models, namely LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), and GRU (Gated Recurrent Unit). Specifically, we set the following scientific goals:

- Definition of minimum and maximum thresholds for anomaly detection (min_threshold, max_threshold);
- Added an "Anomaly" column to indicate anomalies based on thresholds;
- Displaying wind speed and generated power values for detected anomalies and non-anomalous events;
- Calculation of detected anomalies and non-anomalies number;
- To design and develop deep learning based anomaly detection models for wind speed and generated power values utilizing past datasets;
- To evaluate and examine the prediction capabilities of the proposed deep learning models (CNN, LSTM, and GRU), performance assessment metrics such as Mean-Square Error (MSE), Root Mean-Square Error (RMSE), Mean Absolute Error (MAE), Pearson Correlation (PC), and Coefficient of determination (R^2) are used to compare the performance of the deep learning models.;
- This study compares the proposed model (CNN, LSTM, and GRU) with state-of-the-art to show how our models are more efficient than existing studies.

The remainder of the paper is organized as follows: Related Work section 2 presents the related works. The architecture section describes the architecture of the proposed approach. Case Study section describes the case study and results of the proposed approach. The conclusion section concludes our paper.

2. Related Work

The industry 4.0 has created a paradigm shift in how industrial equipment could be monitored and diagnosed with the help of emerging technologies such as artificial intelligence (AI) techniques. AI-driven troubleshooting tools play an important role in high-efficacy diagnosis and monitoring processes, especially for systems consisting of several components including wind turbines (WTs). A lot of researchers have conducted studies, new frameworks, and designs regarding anomaly detection of aeolian wind speed using various algorithms, methodologies, techniques, and procedures, which will be considered as part of the theoretical framework of this research, enabling afterward the construction of the conceptual framework of the study. Cui et al. [11] presented an anomaly detection approach using machine learning to achieve condition monitoring for wind turbines. The approach applies the information in supervisory control and data acquisition systems. The proposed approach has been tested with the data experience of a 2MW wind turbine in Sweden. The result demonstrates that the approach can detect possible anomalies before the failure occurrence. In the study done by [12], proposes a novel Deep Small-World Neural Network (DSWNN) based on unsupervised learning to detect the early failures of wind turbines. The DSWNN model is a combination of a deep auto-encoder network and a small-world neural network, which are more accurate in simulating the dynamic behavior of wind turbines by working on a closer level of mimicking the working process of a natural brain. In the study done by [13], an anomaly detection method for gearbox oil temperature using SCADA data is proposed based on Sparse Bayesian Learning (SBL) and hypothesis testing (HT). Then, the anomaly can be detected by observing whether the actual temperature value falls into the estimated interval at an enough high possibility, which can be checked by using HT. Besides that, Nie et al. [14] an auto-encoder-based solution named denoising stacked feature enhanced auto-encoder with dynamic feature enhanced factor for fault diagnosis of wind turbines. In their approach, feature enhancement relies on a competition and enhancement policy that prioritizes neurons with higher activation values, suppressing the neurons with lower ones. This approach aims to increase the neurons' specialization, leading the DAE to extract discriminative features. Roelofs et al. [15] introduced a novel method: ARCANA. The authors used ARCANA to identify the possible root causes of anomalies detected by an autoencoder. It describes the reconstruction process as an optimization problem that aims to remove anomalous properties from an anomaly considerably. The proposed method is applied to an open data set of wind turbine sensor data, where an artificial error was added to the wind speed sensor measurements to acquire a controlled test environment. Hoffmann et al. [16] presented a Semi-Supervised Deep Learning approach for anomaly detection of Wind Turbine generators based on vibration signals. The proposed solution is integrated into an IoT Platform as a real-time Workflow. The Workflow is responsible for the whole detection process when a new sample is inserted in the IoT Platform, performing data gathering, preprocessing, feature extraction, and classification. In another study by [17], a new system named LSTM-based VAE-WGAN was established to address the challenge of small and noisy wind turbine datasets. Moreover, the proposed LSTM-VAE-WGAN system with the two-stage adversarial semi-supervised training approach achieved the best performance with the earliest alarm point and highest $F1$-score. According to the authors, the similarity between the model-fit distribution and true distribution was quantified using the Wasserstein distance, enabling complex high-dimensional data distributions to be learned. Additionally, an adaptive identification method of abnormal data (AIMAD) in the wind and solar power stations is proposed by [18], including the bidirectional one-sided quartile and double DBSCAN method to deal with unevenly distributed abnormal data. The operation data of 30 wind farms and 8 solar plants in China are taken as examples to verify the effectiveness and superiority of the proposed method. Du et al. [19] proposed a denoising autoencoder (DAE) based anomaly detector and performed anomaly root cause analysis using sparse estimation. For anomaly detection, a deep denoising autoencoder is learned with normal history data, with enhanced robustness compared to the conventional autoencoder. In another study by [20], is to explore the preprocessing of normal data sets using the WPT HPF-PCA method and the detection of outliers using the LSTM-AE model. The key idea was to transform the data using WPT and HPF and perform dimensionality reduction using PCA to extract the characteristics of the data. In the experiments, the authors trained and evaluated the model using different normal datasets with sequential step preprocessing, and evaluated the model's performance by comparing the reconstruction loss with the actual outliers.

In the artificial intelligence studies, Ding et al. [21] proposed a remote real-time monitoring anomaly recognition method for power system equipment based on artificial intelligence technology. This article used four wind turbines as experimental objects and collected data on their parameters such as speed, temperature, current, and voltage. After data collection, preprocessing and feature extraction were performed on the data, and a model based on CNN was selected for anomaly detection. Amini et al. [22] evaluated, compared, and contrasted eight different artificial neural networks (ANNs) models for the diagnosis of WTs that predict the machinery's system failure based on internal components' sensor signals and generation temperature. The authors developed a system that predicted the output of the WT's generator temperature with 2 months in advance measurement prediction. The following Table. 1 gives a summary of all literature reviews.

Table 1. Literature review of SCADA data of wind turbines by various authors.

Authors and Paper	Model	Dataset
Cui et al. [11]	Machine Learning	SCADA Data
[12]	Novel Deep Small-World Neural Network (DSWNN)	SCADA Data of Wind Turbines
[13]	LSTM-SDAE and XGBoost	Historical SCADA Data
Nie et al. [14]	Novel Autoencoder with Dynamic Feature	The bearing vibration signals of Case Western Reserve University (CWRU)
Roelofs et al. [15]	Novel Method ARCANA	SCADA data of wind farms for 2016 and 2017

Hoffmann et al. [16]	Deep Autoencoder (DAE)	The monitoring systems collected data according to the instrumentation of the wind turbines.
[17]	LSTM-based VAE-WGAN	Small and Noisy Wind Turbine Datasets
[18]	DBSCAN method	Wind and Solar Power Stations Dataset
Du et al. [19]	Denoising AutoEncoder (DAE)	Normal History Dataset
[20]	LSTMAutoencoder (LSTM-AE)	Vibration Data of Wind Turbines
Ding et al. [21]	Artificial Intelligence Technology (CNN)	Operational data of four wind turbine equipment
Amini et al. [22]	Artificial Intelligence Neural Network	SCADA data recording of nine WTs

In Table 1, it is evident that researchers have conducted experiments utilizing a diverse assortment of pre-trained deep learning models across multiple SCADA datasets of Wind Turbines, with a particular emphasis on SCADA data, small and noisy wind Turbine data, and vibration data. These investigations have yielded a spectrum of accuracy scores. Consequently, our research focuses on the SCADA Data of Wind Turbines, aiming to enhance its accuracy to a level that meets acceptable standards.

3. Approach Proposed

As per our objective and motivations, this study is associated with some background ideas and research efforts as shown in Figure. 1. Briefly, especially using deep learning models for anomaly detection and supporting it with Aeolian wind speed processing has been remarkable ideas to follow. In general, widely followed automatic detection approaches performed with Deep Learning models such as LSTM (Long Short-Term Memory), CNN (Convolutional Neural Network), and GRU (Gated Recurrent Unit) have been directed to the detection of the minimum and maximum values of wind speed as well as predict the power produced by wind turbines (see Fig. 1).

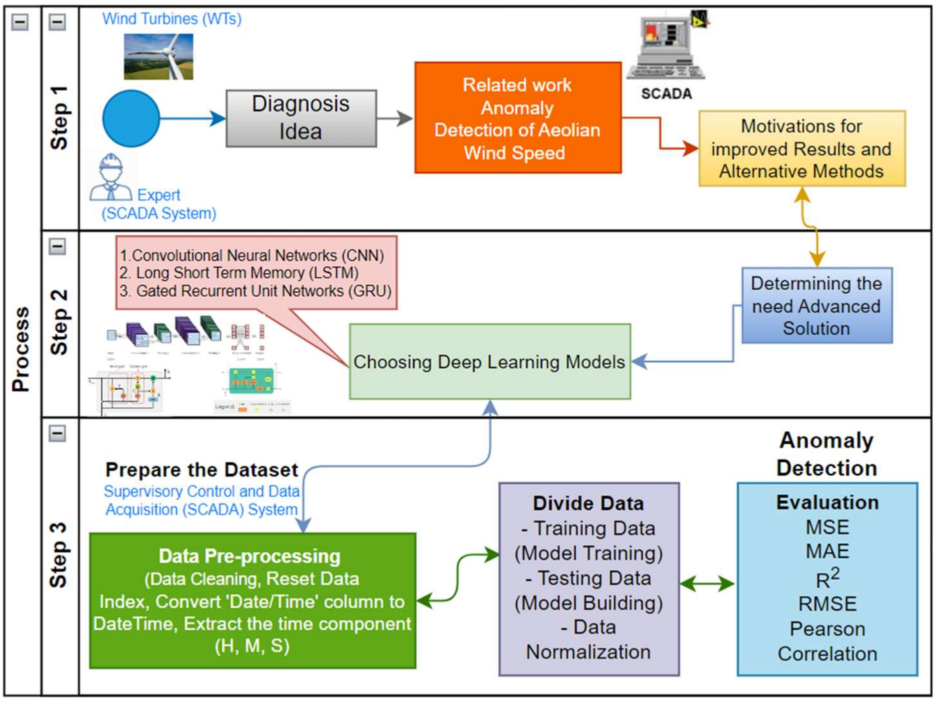

Fig. 1. Ideas and research efforts on the background of this study

In the context mentioned above, this study followed an easy-to-design data pre-processing and deep learning approach for the anomaly detection of Aeolian wind speed by considering wind turbines' SCADA data, Wind Power Data, and Smart sensors as input data. Initially, previously unlabeled SCADA data from wind parc and smart sensors are collected, and further, the collected data are pre-processed using a feature engineering technique. Our proposed approach is then used to train and classify the SCADA dataset. Finally, some evaluation criteria are used to gauge performance (see Fig. 2).

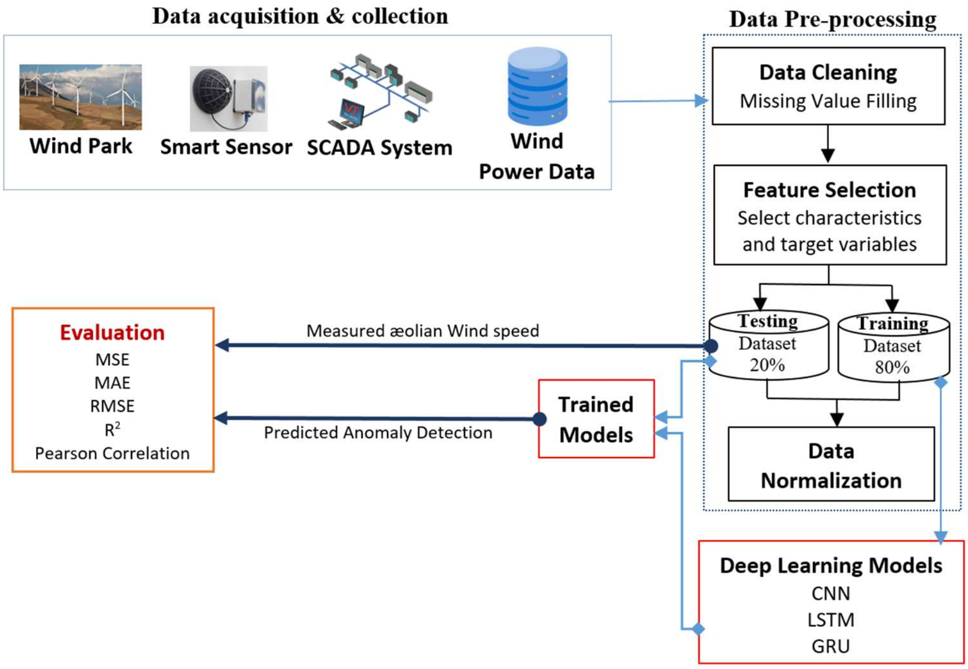

Fig. 2. Main Steps of the proposed approach

3.1. SCADA Data Collection and Dataset Analysis

In the literature, the wind turbines are connected to the SCADA system, which collects data about the wind turbines continuously. The SCADA system supervises the operating status of the wind turbines and protects them from extreme loads. In Wind Turbines, SCADA Systems measure and save metrological datasets for 10-minute intervals [23]. According to [24], the SCADA system typically monitors hundreds of variables with a low sampling frequency, ranging from every few seconds to minutes. The logged data includes details such as the direction of the wind and the speed of the wind. This dataset has been used for analyzing the performance of algorithms (Machine/Deep Learning) used for anomaly detection of aeolian wind speed. The Wind Turbine SCADA, wind power, and smart sensor datasets used in this paper to test the accuracy of the proposed algorithms are taken from the Kaggle website[1]. The current dataset is organized in per-year comma-separated values (.csv) files from January 1, 2018, to December 31, 2018. The wind farm is located at Yalova in the north-western region of Turkey [25]. These wind turbines are equipped with a SCADA unit, and five attributes are reported at 10-minute intervals (See Table 2):

Table 2. The Wind Turbine SCADA Dataset Features

Feature	Description
Date/Time	10-minute intervals (timestamp of the observation).
LV Active Power (kw)	The power generated by the turbine for that moment.
Wind Speed (m/s)	The wind speed at the hub height of the turbine (the wind speed that the turbine uses for electricity generation).
Theoretical_Power_Curve (KWh)	The theoretical power values that the turbine generates with that wind speed which is given by the turbine manufacturer.
Wind Direction (°)	The wind direction at the hub height of the turbine (wind turbines turn to this direction automatically).

Moreover, Table 3 presents the raw dataset collected by the SCADA system. The data set shown in Table 3 holds a total of 50,530, 10 min of measurements of Wind Speed, Active Power, Theoretical Power, and Wind Direction. For the resulting dataset, Wind Speed, Wind Direction, and Active Power were extracted and utilized to develop the anomaly detection models.

Table 3. Raw dataset stored by the Wind Turbine SCADA systems.

[1]The Wind Turbine SCADA Dataset [Online]. Available: https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset/data

Date	Time	Active Power (kw)	Wind Speed (m/s)	Theoretical_Power _Curve (KWh)	Wind Direction (°)
January 01, 2018	00:00	380.047	5.311	416.328	259.994
January 01, 2018	00:10	453.769	5.672	519.917	268.641
January 01, 2018	00:20	306.376	5.216	390.900	272.564
January 01, 2018	00:30	419.645	5.659	516.127	271.258
January 01, 2018	00:40	380.650	5.577	491.702	265.674
January 01, 2018	00:50	402.391	5.604	499.436	265.674
…..	…..	…..	…..	…..	…..
December 31, 2018	23:50	2820.466	9.979	2779.18	82.274

In terms of data analysis, the researchers considered the wind speed as a temporal data series to predict the intensity of the wind in the coming days. The time-series data is transformed into a visual representation, where all its characteristics and distinguishable elements are noticeable. As well, the researchers can see regions where the wind blows the most, and model corroboration in terms of actual energy generation, anomalies, and wind behavior in terms of speed and direction. Furthermore, the dataset has been divided into two parts: Training and Testing. The training data consists of 80% of the dataset. The checking and testing data, on the other hand, consist of 20% of the dataset.

3.2. Data Preprocessing

Data pre-processing is crucial in Deep Learning (DL) because "Better data beats fancier algorithms". Data preprocessing is an important aspect of data preparation that prioritizes data quality analysis. It is the basis for the accuracy and reliability of the prediction model [6; 10]. It should be noted that data preprocessing is an essential step to eliminate invalid data before undertaking the modeling of SCADA data. These data will reduce model accuracy and should be removed before model training. The SCADA data used for this study has an extensive range of wind turbine parameters. Additionally, wind turbine SCADA data is adopted, and several parameters are selected based on physics knowledge and correlation coefficient analysis for normal behavior modeling. The dataset contains different outliers, for example, negative power values, abnormal wind speed, and out-of-range values. Data processing is a technique mostly used in DL to convert the data into a desired and meaningful. There are several useless or missing data in the collected sequences because of the acquisition failure and unscheduled downtime. So the raw data will be preprocessed first before training the model to assure accuracy. Due to the failure during the transmission and storage, some missing values and blanks break the integrality of the raw data. The features with less contribution are removed and the model accuracy is calculated again. These steps are repeated until we get the best model. This will continue until the maximum number of folds is reached.

In this study, we create a data frame to associate feature names with their importance. In this regard, we see that Wind turbine systems can't generate any power if the wind speed is less than 4 m/s. Then, when the wind speed is larger than 4 m/s to 11 m/s, the relation between them is linear meaning that increasing the wind speed allows turbines to generate more power. Finally, after the wind speed passes 11 m/s, the power generated is saturated at 3600 KWh (see Fig. 3).

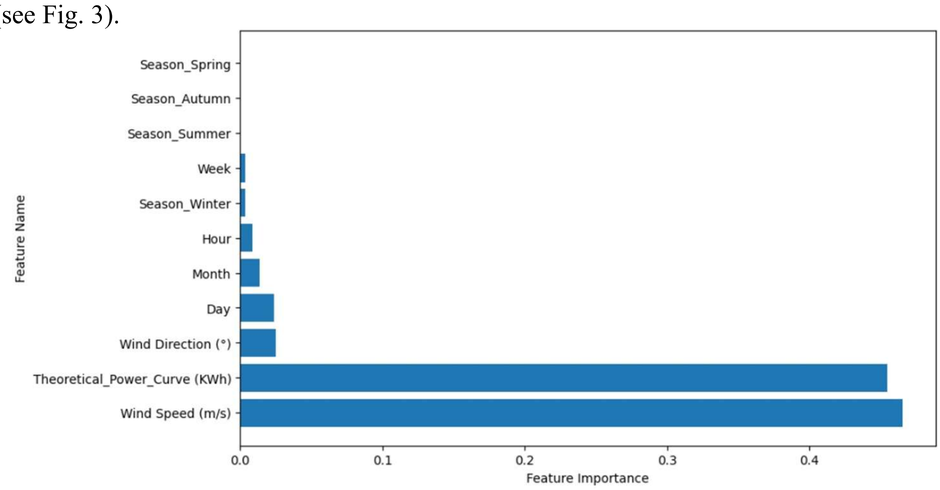

Fig. 3. Sort features by importance

In this Python data visualization, we worked with the Pandas *scatter_matrix* method to visualize the relationship between multiple variables in a dataset at once, namely to explore trends between the two variables (*Wind Speed* *(m/s) and LV ActivePower (kW)*). For data normalization, this means the training data will be used to estimate the minimum and maximum. The *Scikit-learn* scaler is a fundamental tool that helps standardize numerical data within a specific range, making it suitable for deep learning algorithms that are sensitive to feature scaling. The *MinMaxScaler()* function scales each feature individually so that the values have a given minimum and maximum value, with a default of 0 and 1.

4. Deep Learning Architecture Overview

Deep Learning (DL) is a part of an artificial neural network technique and a subclass of machine learning [26; 27]. Moreover, DL is part of a broader family of machine learning methods based on learning data representations [26; 28]. Deep learning is an advanced sub-field of machine learning, which advanced Machine Learning closer to Artificial Intelligence. According to [29], multiple layers in deep learning algorithms are used for a higher feature level in the input dataset. Furthermore, Deep Conventional Neural Network (DCNN) is a class of deep, feed-forward artificial neural networks that has successfully been applied to analyzing visual imagery.

4.1. Convolutional Neural Networks (CNNs)

In most literature, a neural network comprises diverse layers associated with each other, working on the structure and function of the human brain. It learns from large volumes of information and uses complex calculations to prepare a neural network. In recent years, the Convolutional Neural Network (CNN) has been a type of artificial neural network mainly used in data processing with grid-like topologies, such as image recognition and classification [30]. Likewise, CNN is designed based on a convolutional layer, which is considered the core building block of a CNN. It presents multiple parameters that include a group of learnable kernel filters. Each filter was convolved across the width and height of the input volume [27]. CNNs use convolution in at least one of their layers, instead of a general matrix multiplication, as do the feed-forward deep neural networks studied in the previous chapters. Compared to other classification algorithms, the preprocessing required in a CNN is considerably lower [29]. CNN explicitly assumes the input is an image and reacts it into its architecture. CNN usually contains a Convolutional layer, a Pooling layer, and a fully connected layer. Convolutional layers and Pooling layers are stacked on each other; fully connected layers at the top of the network output the class probabilities (see Fig. 4). The advantages of CNNs include feature selection, weight sharing, and pooling mechanisms. For these reasons, the network may recognize increasingly complex elements, including objects, faces, and other fields of data science, including big data analysis [31].

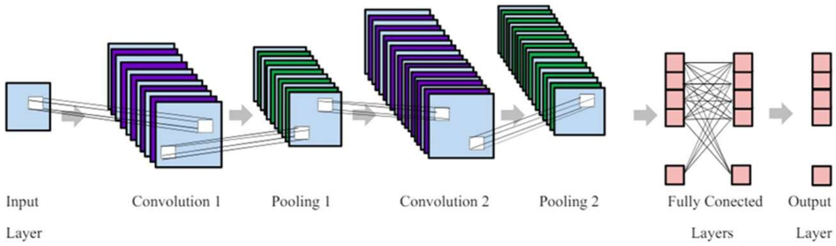

Fig. 4. Convolutional Neural Networks (CNNs) Architecture [31]

4.2. Long Short-Term Memory (LSTM)

The long short-term memory (LSTM) method was proposed by [32] to deal with the gradient vanishing problem in RNN. It comprises recurrent-network units that keep track of long and short-term values, store data in memory cells, and are more adept at recognizing and using long-range context [24; 31]. Moreover, the LSTM method can effectively extract features from non-linear time series data and has been widely used in big data analysis, image processing, and speech recognition. According to [34], LSTM can be used as a complex nonlinear unit to construct a larger deep neural network, which can reflect the effect of long-term memory and has the ability of deep learning. LSTM network consists of an input layer, an output layer, and a plurality of hidden layers, which is composed of memory cells. The LSTM network architecture consists of three parts, as shown in the figure below, and each part performs an individual function (see Fig. 5).

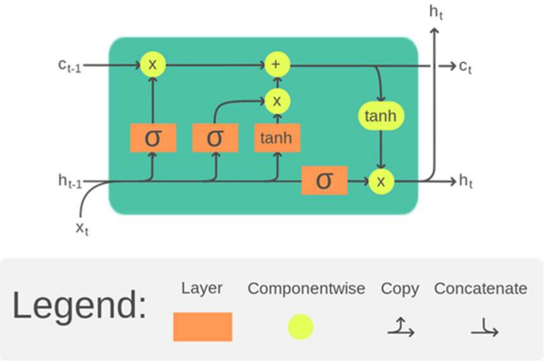

Fig. 5. The Long Short-Term Memory (LSTM) Architecture [34]

In this respect, an LSTM has a hidden state where H_{t-1} represents the hidden state of the previous timestamp and H_t is the hidden state of the current timestamp. In addition to that, LSTM also has a cell state represented by C_{t-1} and $C_{(t)}$ for the previous and current timestamps, respectively.

4.3. Gated Recurrent Unit (GRU)

Gated recurrent unit (GRU) is similar to LSTM but has fewer gates and is a variant of the RNN architecture where gates are used to control the flow of information between neurons [35]. For [36], the Gated recurrent unit (GRU) network is regarded as an updated version of LSTM with a simple structure including a memory cell and gate units. According to [37], the GRU neural network is a circular network structure that determines the current output information through the input information at the current moment and the output information at the previous moment. Formally, Fig. 6 shows that the reset gate is used to control the degree of ignoring the information of the previous moment, and the update gates control whether the status of GRU is updated and how many of the gating units are updated. Therefore, the output information at each moment in the GRU neural network depends on past information. Therefore, its chain attribute is closely related to the sequential labeling problem and is applied to the word segmentation task. The activation gate h_t of the GRU at time t is a linear interpolation between the previous activation h_{t-1} and the next activation $h_t\%$. For this, the equation of GRU can be described as:

$$r_t = \sigma(W_r x_t + U_r h_{t-1}) \quad (1)$$

$$z_t = \sigma(W_z x_t + U_r h_{t-1}) \quad (2)$$

$$h_t\% = tanh(W_h x_t + U_h r_t . h_{t-1}) \quad (3)$$

$$h_t = (1 - z_t)h_{t-1} + z_t h_t\% \quad (4)$$

Where r_t is the reset gate determining the number of ignored prior information. Where σ is the activation function sigmoid, which ranges from 0 to 1. x_t represents the input of the memory unit, z_t is the update gate which determines the number of information input to the next state cell. W_t, W_h *and* W_z represent weight vectors corresponding to the gates in the memory unit, respectively.

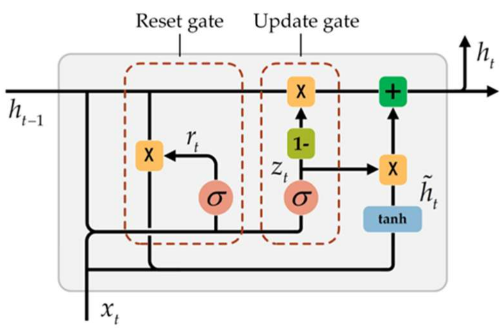

Fig. 6. The Gated Recurrent Unit (GRU) neural network unit structure [37]

5. Experimental and Results

In the literature, most of the research works that apply deep learning for anomaly detection of aeolian wind speed use SCADA datasets to train the model, meaning a huge burden for the experts to label the SCADA data accordingly. In this section, we present the performance results of the wind turbine SCADA dataset and other datasets. Lastly, we compared the performance of our model with previous works.

1.5. Hardware Specifications

The experimental environment of this paper is Windows 10 system, Python 3.6.2, Tensorflow 1.11.0, and Jupyter. In the hardware device section, the CPU is an Intel I9-11900k CPU and an NVIDIA 3070 with 12 GB of VRAM. The primary software configuration included Python compiler, Spyder 4.0.1 editor, deep learning framework PyTorch, and uses the neural network library Keras 2.2.4, Numpy 1.22.3, Pandas 1.4.2, SciPy 1.8.0, Scikit learn 1.0.2, and Matplotlib 3.1.3.

2.5. Performance Measures

A range of statistical techniques was employed to assess the deep learning-based architecture's prediction. To this end, the three standard metrics, the mean absolute error (MAE), root mean square error (RMSE), mean squared error (MSE), and coefficient of determination (R^2), are used to measure the model prediction performance [25; 38; 39]. These measures are defined as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\widehat{x_i} - x_i| \qquad (5)$$

$$RMAE = \sqrt{\sum_{i=1}^{N}(\widehat{x_i} - x_i)^2/N} \qquad (6)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \widehat{x_i})^2 \qquad (7)$$

$$R^2 = 1 - \frac{\sum_{i=0}^{N}(x_i - \widehat{x_i})^2}{\sum_{i=0}^{N}(x_i - \overline{x_i})} \qquad (8)$$

Where x_i is the actual value of wind power, $\widehat{x_i}$ is the predicted value of wind power, and R^2 value ranges between 0 and 1, and higher value corresponds to high performance.
The MAE measures the average magnitude of the errors by considering the absolute value, which presents the accuracy of the prediction. The RMSE measures the forecasting error by differencing the prediction and the actual value, which is squared, average, and then followed by a square root. In the case of MAE and MSE the lower value of prediction corresponds to high accuracy.

3.5. Results and Discussions

In this section, we apply the proposed scheme to detect Wind Turbine (WT) anomalies using the above SCADA data. To illustrate the superiority of deep learning models, we also consider several popular models and compare their performance in exploring the possible nonlinear mechanism. This section presents the results of our deep learning models using tables and graphs containing the research questions and the categories associated with each question. The feasibility of a learning model is limited by assessing the performance of prediction models using multiple assessment measures. This study evaluated the model performance according to the indicators described in Section 2.5. First, we present a deep learning model for the prediction of minimum and maximum wind speed values, as well as the prediction of power output. In this scenario, Table 3 presents the proposed algorithm results, the evaluated thresholds, and the respective metrics. The proposed DL models have achieved high performance, with an R^2 value of 82.37% on the training dataset. In addition, MSE, MAE, and RMSE values are 0.0242, 0.1180, and 0.1555, respectively, and a good prediction result is obtained with Pearson Correlation metrics (see Table 4). This shows that the model is not overfitting and can capture the connection between the data in the newly acquired test dataset well.

Table 4. The evaluation metrics values of the achieved results using the DL models

Model	Evaluation of Performance Using Training Data					Evaluation of performance using Testing Data					Data Train Time (sec.)
	MSE	MAE	R^2	RMSE ($\mu g/m^3$)	PC (r)	MSE	MAE	R^2	RMSE ($\mu g/m^3$)	PC (r)	
LSTM	0.036	0.3211	0.736	0.1904	0.8842	0.037	0.325	0.734	0.199	0.886	5.02
CNN	0.147	0.3211	0.736	0.3715	0.0766	0.175	0.323	0.727	0.376	0.079	5.01
GRU	0.024	0.1181	**0.823**	0.1556	0.9105	0.024	0.190	0.824	0.158	0.912	**4.02**

When the results are examined, the model is able to analyze the data well and shows a successful prediction capability. This is due to the ability of the GRU-based deep learning method to capture long-term dependencies. Compared to classical deep learning-based methods, GRU can also be considered a variation on the LSTM because both are designed similarly and, in some cases, produce equally excellent results. This allows the model to make predictions based on previous data. The MSE, MAE, R^2, RMSE, and Pearson Correlation reported were 0.024, 0.12, 82%, 0.16, and 0.91, respectively. The performance measures shown in Table 3 display that the GRU model has a higher R^2 value than the LSTM and CNN models. Furthermore, the accuracy measures obtained indicate that the GRU model outperforms the other machine learning models tested, with an accuracy of 96.9%, defining the prediction capabilities of the model.

For the identification of normal behavior, it is well-known that the power curve is typically employed to measure the performance of a WT. As seen in the graph, the red points depict the relationship between the wind, where the x-axis represents the wind speed in m/s, and the y-axis represents the active power in kwh (Scatter Diagram: wind speed vs LV Active Power) (see Fig. 7). An increment in power is observed with the increase in wind speed. However, there are some outliers where the power is zero, even with a high wind speed.

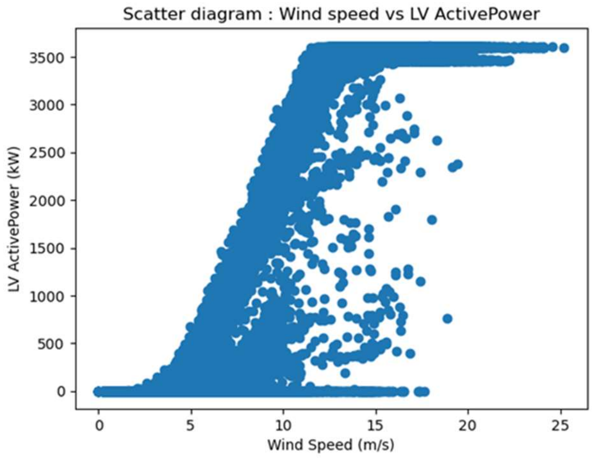

Fig. 7. The scatter plot of LV Active Power vs Wind Speed and the identified normal behavior interval

On the other hand, the baseline models included in this experiment are the deep neural network (CNN, LSTM, and GRU) models. The parameters of these models are presented in the Table. 5.

Table 5. Parameters setting of the deep neural network (CNN, LSTM, and GRU)

Batch size	Learning rate	Epochs	Optimizer	Activation function used in the output	Activation function used in the hidden
64	0.0001	50	Adam	Linear	ReLU

In this context, the results show that our deep learning models provide good performance for detecting wind speed anomalies and predicting power output for training and validation phases (see Fig. 8).

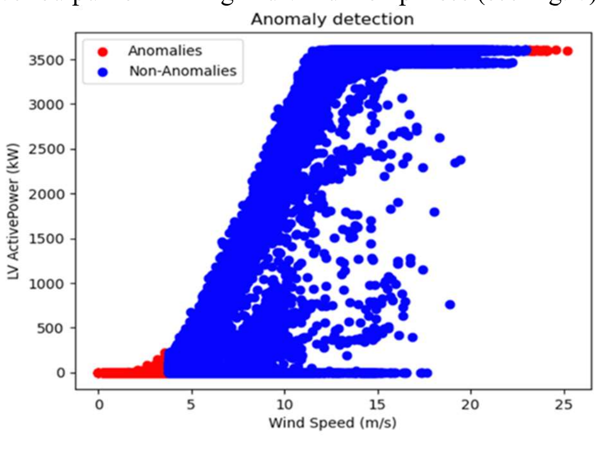

Fig. 8. The anomaly detection of LV active power vs wind speed

For efficiency assessment in this study, samples from running conditions, which are used for validation and test phases, are illustrated in Fig.9, which represents different prediction models. Figure 8 shows the actual active power and the predicted power values by the proposed architecture for the date range 01.12.2018-01.12.2019 on the time axis graph. The actual data and the predicted data are given in the same figure (see Fig. 9). In this context, the results show that the GRU and LSTM models perform better than the CNN model in predicting the active power. Moreover, the LSTM model exhibits strong overall performance with favorable performance metrics, while the GRU model stands out with its high MSE and Pearson correlation. The proposed model performed well by overlapping with the actual value. In summary, anomaly detection of Aeolian wind speed includes the process of continuously observing if certain indicators deviate from normal behaviors, which can detect anomalies of Aeolian wind speed in the early stage. If the indicators consistently surpass the threshold for an extended period of time, the WT will be regarded as abnormal and undergo a shutdown to check the components and resolve the malfunction.

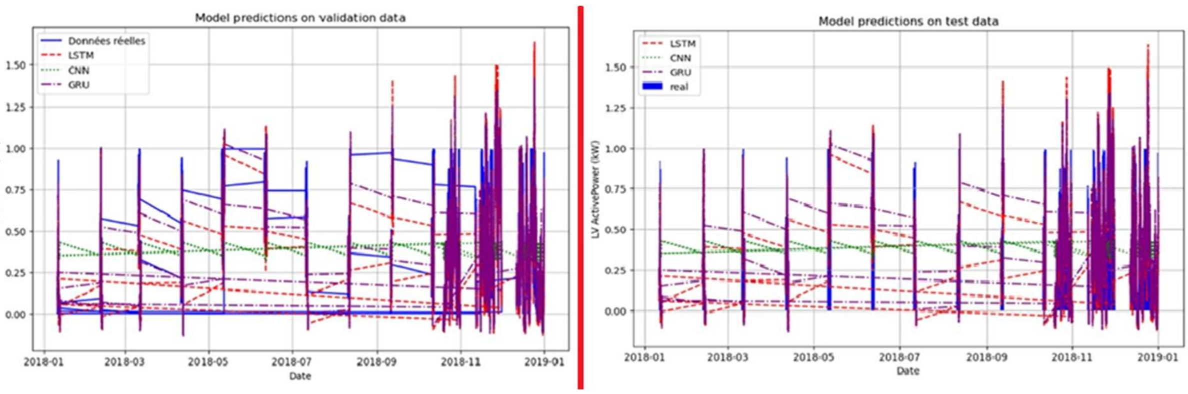

Fig. 9. The prediction model on the validation and test data

As occurs in many applications, the amount of data available for training and testing is limited. Unfortunately, it is necessary, in order to build good models, to use as much available data as possible. Additionally, small validation data sets will give noisy estimates of predictive performance. It is expected that deviations from a normal state can be seen in the 10-minute average data of the sensors and the operational data, and this is indicated as an anomaly by the trained GRU. The result demonstrates that the approach can detect possible anomalies before the failure occurrence. In addition, since this research was conducted on a limited dataset, we plan to demonstrate its validity on various datasets and real-world wind farm environments. Continued technical and research efforts to address these issues are expected to open up new possibilities and contribute to the broader field of artificial intelligence, namely, deterministic artificial intelligence. In addition, anomaly detection needs to be performed using datasets containing different types of anomaly data in order to classify and predict anomaly types.

4. 5. Comparison against State-of-the-Art Methods

Interestingly, these techniques provide insights into the model's decision-making processes, contributing to a more interpretable and transparent understanding of the detections and predictions. In conclusion, we conducted a comparative study of our proposed approach with other existing anomaly detection of Aeolian wind speed models based on the approach used in experimentation, performance metrics used for evaluation, and percentage accuracy achieved (see Table 6).

Table 6. Comparison of the Aeolian wind speed anomaly detection proposed model with existing anomaly detection and classification methodologies.

Techniques applied	Dataset used	Performance Metrics
Machine Learning	SCADA Data	244 alarms from 35 different types are recorded in the SCADA alarm logs
Novel Deep Small-World Neural Network (DSWNN)	SCADA Data of Wind Turbines	Accuracy of DSWNN 97,36%
LSTM-SDAE and XGBoost	Historical SCADA Data	MSE 0.0143 MAE 0.0132 R2 96 %

		RMSE 0.01778
Novel Autoencoder with Dynamic Feature	The bearing vibration signals of Case Western Reserve University (CWRU)	Accuracy 97,14%
Novel Method ARCANA	SCADA data of wind farms for 2016 and 2017	MAE 0.097
Deep Autoencoder (DAE)	Data collected by the monitoring systems according to the wind turbine instrumentation.	MAE 0.1735 Accuracy 99,70% Precision 97,80% Recall 100%
LSTM-based VAE-WGAN	Small and Noisy Wind Turbine Datasets	MAE 0.008 RMSE 0.002 Reconstruction Error 99.7% $F1$ score 0.8381
DBSCAN method	Wind and Solar Power Stations Dataset	MAE 0,2214 MSE 0,012 RMSE 0,1622
Denoising AutoEncoder (DAE)	Normal History Dataset	MAPE 9.69 MSE 5.29 MAE 8.45 R2 15.65
LSTMAutoencoder (LSTM-AE)	Vibration Data of Wind Turbines	Accuracy of 97.44%
Artificial Intelligence Technology (CNN)	Operational data of four wind turbine equipment	Accuracy of CNN 95,36%
Artificial Intelligence Neural Network	SCADA data recording of nine WTs	Accuracy of CNN 99,8%
Proposed Approach (Deep Learning)	**Wind Turbine SCADA Dataset**	**MSE 0.024** **MAE 0.12** **R2 82 %** **RMSE 0.16**

4. Conclusion

On the basis of the scientific literature reviewed by the authors, we can conclude that the most frequently found AI techniques for aeolian wind speed anomaly detection are deep learning algorithms, appearing in the total articles, respectively. As explained in the introduction section, different deep learning architectures have been successfully used in detecting the anomaly of aeolian wind speed. In addition, MSE, MAE, R^2, and RMSE are the most commonly used anomaly detection model evaluation metrics. Considering the complex, changeable working environment of wind turbines, an anomaly detection deep learning models is proposed in this paper, which is employed for anomaly detection of aeolian wind speed. The study focused on anomaly detection within a single wind farm, using different amounts of tuning data. We used data from the supervisory control and data acquisition system to improve wind turbine fault detection accuracy. These architectures are capable of successfully capturing complex relationships between data. However, increasing the number of layers may increase the computational cost.

To further analyze the anomaly of aeolian wind speed based on SCADA data, owners and operators of WTs must provide detailed information and high-frequency data. Ideally, this information should be made accessible to researchers, even if under confidentiality agreements. However, Deep Learning is computationally expensive and may require high computational resources while working with large datasets, and this is critical from the perspective of renewable energy applications. Despite the above relevant findings, further research is needed to improve the model. Moreover, more complex and complete cases are needed to verify the wide applicability of the developed models. In the future, optimization algorithms can be used to enhance accuracy.

References

1. Sahu, B.K. (2017). Wind Energy Developments and Policies in China: A Short Review. *Renewable and Sustainable Energy Reviews*, *81*(1), 1393–1405. https://doi.org/10.1016/j.rser.2017.05.183
2. Porté-Agel, F., Bastankhah, M., Shamsoddin, S. (2019). Wind-turbine and Wind-farm Flows: A Review. *Boundary-Layer Meteorology*, *174* (1), 1–59. https://doi.org/10.1007/s10546-019-00473-0
3. Meik, S., Ilmar, F. S. (2014). Wind Turbine Condition Monitoring based on SCADA Data using Normal Behavior Models. Part 2: Application examples. *Applied Soft Computing*, *14*(C), 447-460. https://doi.org/10.1016/j.asoc.2013.09.016
4. Yang, W., Court, R., Jiang, J. (2013). Wind Turbine Condition Monitoring by the Approach of SCADA Data Analysis. *Renewable Energy*, 53, 365–376. https://doi.org/10.1016/j.renene.2012.11.030
5. Brahami, M., Abdeldjouad, F. Z., Sabri, M. (2022). Multi-Class Sentiment Classification for Healthcare Tweets Using Supervised Learning Techniques. *International Journal of Service Science, Management, Engineering, and Technology*, *13*(1), 1-23. http://doi.org/10.4018/IJSSMET.298669
6. Brahami, M., Sabri, M., Kedidir, C., Dehbi, F. (2022). Sentiment Analysis for Product Recommendation on Amazon's Mobile Phone Reviews using Machine Learning Techniques. *International Aegean Conference on Innovation Technologies and Engineering*. September 02-04, 2022, Ege University, Izmir, Türkiye, Proceedings Book by Iksad Global Publishing House, Edited by Bahri Başaran & Aysun Baltaci, ISBN : 978-625-8213-60-7
7. Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., Nenadic, G. (2019). Machine Learning Methods for Wind Turbine Condition Monitoring: A review. *Renewable Energy* 133:620–35. https://doi.org/10.1016/j.renene.2018.10.047
8. Clottey, R. N., Yaokumah, W., Appati, J. K. (2021). Modelling and Evaluation of Network Intrusion Detection Systems Using Machine Learning Techniques. *International Journal of Intelligent Information Technologies (IJIIT)*, *17*(4), 1–19. http://doi.org/10.4018/IJIIT.289971
9. Zhang, C., Yang, T. (2023). Anomaly Detection for Wind Turbines Using Long Short-Term Memory-Based Variational Autoencoder Wasserstein Generation Adversarial Network under Semi-Supervised Training. *Energies*, *16*(19):7008. https://doi.org/10.3390/en16197008
10. Brahami, M., Abdallah El Hadj, M. I., Matta, N. (2023). Towards an Approach of Android Malware Detection using Stacked AutoEncoder and Convolutional Neural Networks. *International Journal of Intelligent Information Technologies (IJIIT)*, *19*(1), 1-22. http://doi.org/10.4018/IJIIT.329956
11. Cui, Y., Bangalore, P., & Tjernberg, L. B. (2018). An Anomaly Detection Approach Based on Machine Learning and SCADA Data for Condition Monitoring of Wind Turbines. *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 24-28 June, Boise, ID, USA, 1-6. doi:10.1109/pmaps.2018.8440525
12. Li, M., Wang, S., Fang, S., Zhao, J. (2020). Anomaly Detection of Wind Turbines Based on Deep Small-World Neural Network. *Applied Sciences*. *10*(4):1243. 1-15. https://doi.org/10.3390/app10041243
13. Zeng, X. J., Yang, M., & Bo, Y. F. (2020). Gearbox oil Temperature Anomaly Detection for Wind Turbine based on Sparse Bayesian Probability Estimation. *International Journal of Electrical Power & Energy Systems*, 123, 106233. https://doi.org/10.1016/j.ijepes.2020.106233
14. Nie, X., Liu, S., Xie, G. (2020). A Novel Autoencoder with Dynamic Feature Enhanced Factor for Fault Diagnosis of Wind Turbine. *Electronics*, *9*(4):600. https://doi.org/10.3390/electronics9040600
15. Roelofs, C.M.A., Lutz, M.-A., Faulstich, S., Vogt, S. (2021). Autoencoder-based Anomaly Root Cause Analysis for Wind Turbines, *Energy and AI*, *4* (2021) 100065, 1-9. https://doi.org/10.1016/j.egyai.2021.100065
16. Hoffmann, J. L. C., Horstmann, L. P., Lucena, M. M., Medeiros de Araujo, G., Fröhlich, A. A., and Nishioka, M. H. N. (2021). Anomaly Detection on Wind Turbines Based on a Deep Learning Analysis of Vibration Signals. Applied Artificial Intelligence, 35(12), 893-913. https://doi-org.sndl1.arn.dz/10.1080/08839514.2021.1966879
17. Zhang, C.; Yang, T. (2023). Anomaly Detection for Wind Turbines Using Long Short-Term Memory-Based Variational Autoencoder Wasserstein Generation Adversarial Network under Semi-Supervised Training. *Energies*, 16, 7008, 1-18. https://doi.org/10.3390/en16197008
18. Wang, H., Zhang, N., Du, E., Yan, J., Han, S., Li, N., Li, H., Liu, Y. (2023). An Adaptive Identification Method of Abnormal Data in Wind and Solar Power Stations. *Renewable Energy*, *208*, 76-93. https://doi.org/10.1016/j.renene.2023.03.081
19. Du, S., Wan, Y., Zhang, C., Zhang, S. (2023). Anomaly Root Cause Analysis for Wind Turbines Based on Denoising Autoencoder and Sparse Estimation. Proceedings of the *IEEE 12th Data-Driven Control and Learning Systems Conference (DDCLS'2023)*, 12-14 May, Xiangtan, China. https://doi.org/10.1109/DDCLS58216.2023.10166052

20. Lee, Y., Park, C., Kim, N., Ahn, J., and Jeong, J. (2024). LSTM-Autoencoder Based Anomaly Detection Using Vibration Data of Wind Turbines. Sensors, 24(9):2833, 1-24. https://doi.org/10.3390/s24092833
21. Ding, X., Gong, Y., Wang, C., and Zheng, Z. (2024). Artificial intelligence-based abnormal detection system and method for wind power equipment. International Journal of Thermofluids, 21 (100569), 1-9. https://doi.org/10.1016/j.ijft.2024.100569
22. Amini, A., Kanfoud, J., and Gan, T-H. (2022). An Artificial Intelligence Neural Network Predictive Model for Anomaly Detection and Monitoring of Wind Turbines Using SCADA Data. Applied Artificial Intelligence, 36(1), 880-894. https://doi.org/10.1080/08839514.2022.2034718
23. Delgado, I., Fahim, M. (2020). Wind Turbine Data Analysis and LSTM-Based Prediction in SCADA System. *Energies*, *14*(1), 1-22. https://doi.org/10.3390/en14010125
24. Singh, U., Rizwan, M. (2022). SCADA System Dataset Exploration and Machine Learning-based Forecast for Wind Turbines. *Results in Engineering*, *16* (100640), 1-9. https://doi.org/10.1016/j.rineng.2022.100640.
25. Çelebi, S. B., Karaman, Ö. A. (2023). Multilayer LSTM Model for Wind Power Estimation in the SCADA System. *European Journal of Technic*, *13*(2), 116-122. https://doi.org/10.36222/ejt.1382837
26. Qiu, J., Zhang, J., Luo, W., Pan, L., Nepal, S., Xiang, Y. (2021). A Survey of Android Malware Detection with Deep Neural Models. *ACM Computing Surveys*, *53*(6), 1–36. https://doi.org/10.1145/3417978
27. Brahami, M., Dermane, Z., Kebir, N. E-H., & Matta, N. (2022). Diabetic Retinopathy Classification Using Hybrid Deep Learning Approach. *SN Computer Science*, *3*(5), 1-15. https://doi.org/10.1007/s42979-022-01240-8
28. Ykhlef, A., Labri, N.S. Brahami, M. (2024). Supervised learning techniques for blood product prediction in patients with hematologic diseases: a multi-center study in Western Algeria. *International Journal of Information Technology*, *16*(5), 1-26. https://doi.org/10.1007/s41870-024-01928-5
29. Brahami, M., Abdeldjouad, F. Z., Sabri, M., Semaoune, K., Matta, N. (2022). Forecasting Supply Chain Demand Approach Using Knowledge Management Processes and Supervised Learning Techniques. *International Journal of Information Systems and Supply Chain Management*, *15*(1), 1–21. https://doi.org/10.4018/IJISSCM.2022010103
30. Montesinos López, O. A., Montesinos López, A., Crossa, J. (2022). *Convolutional Neural Networks*. Chapter 13. In the book: Multivariate Statistical Machine Learning Methods for Genomic Prediction. Springer, Cham, 533–577. ISBN : 978-3-030-89010-0. https://doi.org/10.1007/978-3-030-89010-0_13
31. Ma, W., Qiu, Z., Song, J., Li, J., Cheng, Q., Zhai, J. et al. (2018). A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta* 248:1307–1318. https://doi.org/10.1007/s00425-018-2976-9
32. Hochreiter, S., Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, *9*(8), 1735-80. http://doi.org/10.1162/neco.1997.9.8.1735
33. Veena, H.N., Patil, K.K., Vanajakshi, P., Ambore, A., & Gowda, N. C. (2024). An Enhanced RNN-LSTM Model for Fundus Image Classification to Diagnose Glaucoma. *SN Computer Science*, *5*(514), 1-12. https://doi.org/10.1007/s42979-024-02867-5
34. Xiaoyun, Q., Xiaoning, K., Chao, Z., Shuai, J., Xiuda, M. (2016). Short-term prediction of wind power based on deep Long Short-Term Memory. Proceedings of the *IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC'2016)*, 25-28 October, Xi'an, China. 10.1109/appeec.2016.7779672
35. Hu, C., Wu, Q., Li, H., Jian, S., Li, N., Lou, Z. (2018). Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. *Water*, *10*(11), 1-19. https://doi.org/10.3390/w10111543
36. Batur Dinler, Ö., Aydin, N. (2020). An Optimal Feature Parameter Set Based on Gated Recurrent Unit Recurrent Neural Networks for Speech Segment Detection. *Applied Sciences*, *10*(4), 1-23. https://doi.org/10.3390/app10041273
37. Wang, Y., Han, J., Zhang, T., Qing, D. (2021). Speech Enhancement from Fused Features based on Deep Neural Network and Gated Recurrent Unit Network. *EURASIP Journal on Advances in Signal Processing*, *104* (2021), 1-19. https://doi.org/10.1186/s13634-021-00813-8
38. Li, P., Luo, A., Liu, J., Wang, Y., Zhu, J., Deng, Y., Zhang, J. (2020). Bidirectional Gated Recurrent Unit Neural Network for Chinese Address Element Segmentation. *ISPRS International Journal of Geo-Information*, *9*(11):635, 1-19. https://doi.org/10.3390/ijgi9110635
39. Zhao, H., Liu, H., Hu, W., Yan, X. (2018). Anomaly Detection and Fault Analysis of Wind Turbine Components Based on Deep Learning Network. *Renewable Energy*, *127*, 825–834. https://doi.org/10.1016/j.renene.2018.05.024