23rd International Conference on Renewable Energies and Power Quality (ICREPQ'25)
Tenerife (Spain), 25th to 27th June 2025
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.23, September 2025

Wind turbine anomaly detection using a support vector machine optimized with a Bayesian algorithm.

D. Coronel[1], C. Guevara [2] and M. Santos [3]

[1] Computer Science Faculty
University Complutense of Madrid, Spain
dcoron01@ucm.es

[2] Quantitative Methods Department
CUNEF University, Spain
cesar.guevara@cunef.edu

[3] Institute of Knowledge Technology
Complutense University of Madrid, Spain
msantos@ucm.es

Abstract. Wind energy has experienced significant growth in recent years; however, it still faces challenges in operation and maintenance, which impact energy efficiency and lead to high costs. This study proposes an anomaly detection model for wind turbines based on a support vector machine (SVM), optimized using Bayesian search. The model was trained using vibration data from the Gearbox Reliability Collaborative (GRC) project of the National Renewable Energy Laboratory (NREL), specifically from the High-Speed Shaft Upwind Bearing Radial and High-Speed Shaft Downwind Bearing Radial sensors, with 40,000 records evenly distributed between normal and anomalous conditions. The proposed model achieved an overall fault detection accuracy of 78.95% and 78.50% for the respective sensor data. Bayesian optimization facilitated the fine-tuning of the hyperparameters in the classification technique, enhancing the model's anomaly detection capability. Furthermore, the use of vibration data enabled the identification of critical fault patterns in turbine operation, contributing to the improvement of wind turbine efficiency and reliability.

Key words. Support Vector Machines, Anomalies, Wind Turbines, Bayesian algorithm.

1. Introduction

In recent years, the global adoption of renewable energy sources has increased significantly, driving major advancements in energy sustainability and playing a crucial role in reducing environmental impact [1]. Among these sources, wind energy has emerged as a promising alternative to fossil fuels due to its capacity to generate electricity efficiently, reliably, and with minimal environmental impact by harnessing wind power [2].

According to the Global Offshore Wind Report 2024 published by the Global Wind Energy Council (GWEC), global wind power capacity reached a record high of 1 terawatt (TW) in 2023, representing a 24% increase compared to 2022 [3].

Despite significant technological advancements, wind energy continues to face substantial challenges in its operation and maintenance (O&M) processes, leading to high maintenance costs and impacting power generation efficiency. These costs account for approximately 10% to 15% of onshore wind farm revenues and 20% to 25% for offshore turbines over their 20-year operational lifespan [4]. Key O&M challenges include failures in electrical and control systems, blades, hydraulic systems, generators, and gearboxes [5], [6].

Several studies have investigated the detection of turbine anomalies, often utilizing measured vibration signals as a key diagnostic tool [7]. For instance, in [8], a method combining Long Short-Term Memory (LSTM) neural networks and Autoencoders is proposed for anomaly detection in a 3 MW wind turbine located on the south coast of Ireland. The results of this study achieved an accuracy of 92.95%.

Similarly, in [9], a multi-channel convolutional neural network (MCNN) was employed to detect blade angle anomalies, surface damage, and fractures based on data from a triaxial transducer, which recorded vibration signals under normal conditions and three specific fault states, achieving an accuracy of 87.8%.

Furthermore, the study presented in [10] introduced an anomaly detection model based on a deep convolutional autoencoder enhanced by fault instances (Triplet-ConvDAE). This model was trained using SCADA data collected from a Mongolian wind farm, analyzing two wind turbines over the period November 2015 to January 2016. The results demonstrated classification accuracies of 87.2% and 87.0% for the two turbines.

Given the significance of anomaly detection in generators and its potential to reduce maintenance costs while improving wind turbine operation, this study proposes an approach based on support vector machines (SVM) combined with a Bayesian optimizer for hyperparameter tuning. The dataset used in this study was obtained from the Gearbox Reliability Collaborative (GRC) project [3] of the National Renewable Energy Laboratory (NREL), which includes sensors installed at various locations within the wind turbine gearbox to collect operational data. The model's performance is evaluated using widely recognized metrics, including accuracy, F1-score, recall, and the confusion matrix.

This document is structured as follows: Section 2 describes the materials used in this study. Section 3 outlines the proposed strategy for anomaly detection. Section 4 presents the obtained results and compares them with existing studies in the literature. Finally, Section 5 summarizes the conclusions drawn from this study and suggests potential directions for future research.

2. Materials

This study utilized a dataset from a wind turbine test rig, provided by the Gearbox Reliability Collaborative (GRC) project of the National Renewable Energy Laboratory (NREL) [3] The dataset was collected from accelerometers installed at various locations within the wind turbine drive system, including the ring gear and the bearings of the low-, intermediate-, and high-speed shafts, among others. Each accelerometer was connected to a SCADA-based data acquisition system, operating at a sampling rate of 40 kHz per channel.

This dataset comprises 10 subsets, each corresponding to a different wind turbine equipped with 9 sensors, recording approximately 2.4 million samples per sensor, classified as either “Healthy” or “Damaged”. For this study, 20,000 records obtained from the High-Speed Shaft Upwind Bearing Radial Sensor were processed and analyzed, with 10,000 classified as “Healthy” and the remaining 10,000 as “Damaged”. A similar procedure was followed for the data from the High-Speed Shaft Downwind Bearing Radial Sensor.

Table I provides a summary of the dataset used in this study, organized according to turbine condition, sensor type, and data format.

Table 1. High-Speed Shaft Sensor – Vibration Data

Status	Sensor Position	Data type	Example	Unit
Healthy	Upwind	Integer 2 decimals	1.65	m/s²
	Downwind		2.95	
Damaged	Upwind		15.93	
	Downwind		10.07	

Figure 1 illustrates the location of each sensor within the gearbox. The High-Speed Shaft Upwind Bearing Radial Sensor measures vibrations at the windward bearing of the high-speed shaft, while the High-Speed Shaft Downwind Bearing Radial Sensor monitors radial vibrations in the epicyclic gear train support.

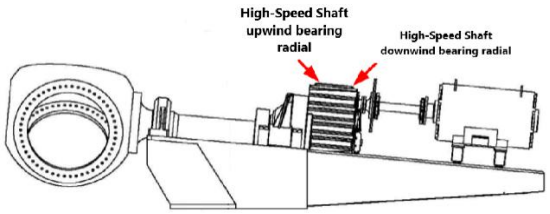

Fig.1. Location of vibration measurement sensors [3].

Figure 2 presents an example of vibratory acceleration signals recorded by the High-Speed Shaft Upwind Bearing Sensor in the wind turbine gearbox. Figure 2a depicts normal operating conditions, with acceleration values ranging approximately between ±15 m/s², while Figure 2b illustrates anomalous conditions or faults, with a reduced acceleration range of approximately ±8 m/s².

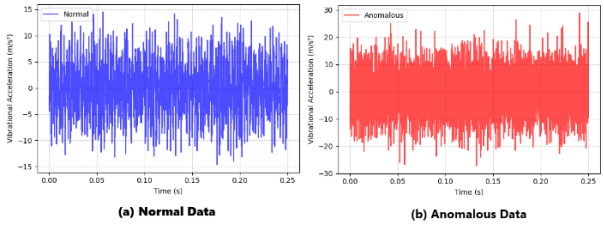

Fig.2. Sample records of the High-Speed Shaft upwind bearing sensor: a) normal and b) anomalous.

Figure 3 presents vibratory acceleration signals recorded by the High-Speed Shaft Downwind Bearing Radial Sensor in the wind turbine gearbox. Figure 3a depicts normal operating conditions, with acceleration values ranging approximately between ±7 m/s², while Figure 3b illustrates anomalous conditions or faults, with an increased acceleration range of approximately ±23 m/s².

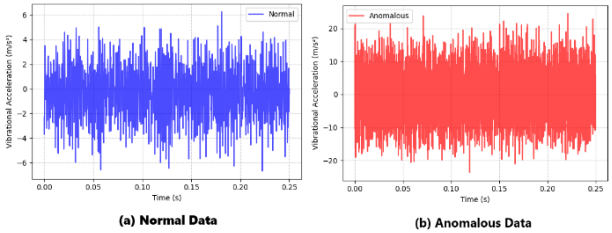

Fig.3. Sample records of the High-Speed Shaft downwind bearing sensor: a) normal and b) anomalous.

The sampling frequency was 40 kHz, where each High-Speed Shaft vibration sensor oscillated between ±7 m/s² under normal Econditions and up to ±23 m/s² under fault conditions, as shown in Figures 2 and 3. These changes between samples provide fundamental information for fault detection.

3. Anomaly detection algorithm

The proposed wind turbine anomaly detection algorithm utilizes a support vector machine (SVM) model [11], optimized using the Bayesian Optuna algorithm [12]. The algorithm is structured into several phases: data acquisition and preprocessing, hyperparameter optimization, model training, and performance evaluation.

A. Phase 1: Reading and preparation of the data

In this phase, the dataset is imported from .csv files containing sensor measurements under both normal and abnormal conditions. Subsequently, labels are assigned to each sample according to Equation (1).

$$y_i = \begin{Bmatrix} 1\ si\ X_i \in X_{normal} \\ -1\ si\ X_i \in X_{anomalous} \end{Bmatrix} \quad (1)$$

Where X_{normal} represents the normal data, corresponding to the expected behavior of the turbine, and is assigned the label $y = 1$. Conversely, $X_{anomalous}$ represents faulty conditions and is assigned the label $y = -1$.

Once the data were labeled, the X_{normal} and $X_{anomalous}$ sets were concatenated for normalization. Z-score normalization was then applied using Equation (2) to enhance the distinction between normal and abnormal data.

$$X_{Normalized} = \frac{X - \mu}{\sigma} \quad (2)$$

Where X represents the original feature value, μ is the mean of the feature, and σ is the standard deviation.
Finally, the dataset was divided into training and test sets, allocating 70% for training and 30% for testing.

B. Phase 2: Hyperparameter optimization of the SVM

In this phase, the hyperparameters of the SVM model are optimized using Optuna, a Bayesian optimization framework. The goal of this optimization is to identify the optimal combination of parameters that maximizes the model's performance metric or minimizes the associated error.

The hyperparameters to be optimized are C, which controls the trade-off between margin maximization and classification error, and γ, which determines the influence of each data point on the classification process.
The Bayesian optimization process is mathematically defined by Equation (3).

$$(C^*, \gamma^*) = \underset{C,\gamma}{\arg\max}\, f(C, \gamma) \quad (3)$$

Where C and γ are assigned following a uniform linear distribution, enabling the exploration of both small and large values with equal probability (Equation 4).

$$\begin{matrix} C \in [10^{-2}, 10^{2}] \\ \gamma \in [10^{-4},\ 10^{-1}] \end{matrix} \quad (4)$$

In addition, the function f(C, γ) measures the model performance by a Gaussian process (equation 5).

$$f(C, \gamma) \sim GP(\mu(C, \gamma), k((C, \gamma), (C', \gamma'))) \quad (5)$$

Where μ(C, γ) represents the predictive mean, and k((C, γ), (C′, γ′)) denotes the covariance (kernel), which quantifies the similarity between points in the search space.

Equation (3) iteratively evaluates multiple hyperparameter combinations within these ranges (Equation 4) until the optimal configuration is determined.

The obtained results are as follows: for the High-Speed Shaft Upwind Bearing Sensor data, C = 0.4651 and γ = 0.0483; while for the High-Speed Shaft Downwind Bearing Sensor data, C = 0.4961 and γ = 0.0483.

C. *Phase 3: Support Vector Machines (SVM) model definition and model evaluation.*

Once the optimal values of C and γ for the model have been determined, the training process is conducted using Equation (5).

$$f(\mathrm{x}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \alpha_{\mathrm{i}} y_{\mathrm{i}} \mathrm{K}(\mathrm{x_i}, \mathrm{x}) + \mathrm{b} \quad (5)$$

Where x represents the new point to be classified, xi are the support vectors, αi are the corresponding coefficients, and yi are the labels assigned to each support vector (+1 for normal and -1 for anomalous). The function K(xi, x) denotes the kernel function, while b is the bias term that adjusts the class separation. In this process, the hyperparameter C influences the values of αi , while γ affects the kernel function K(xi, x).

Therefore, if $f(x) \geq 0$, the data point is classified as normal. Conversely, if $f(x) < 0$, it is classified as an anomalous.

In this phase, the trained model is evaluated using performance metrics such as accuracy, F1-score, recall, and the confusion matrix to assess its effectiveness in distinguishing between normal and abnormal conditions.

4. Fault detection with optimized SVM

For each sensor, 70% of the total dataset 14,000 records) was used for training, while the remaining 30% (6,000 records) was assigned to the test set. This division is commonly used in artificial intelligence, as it aims to provide sufficient data for model training without compromising its generalization ability during evaluation [13].

In order to evaluate the repeatability and robustness of the proposed approach, the entire experimental process was

repeated 10 times for each sensor. Additionally, in the automatic hyperparameter optimization process (Bayesian algorithm), a limit of up to 1000 different hyperparameter combinations was set in order to find the best configuration.

Figure 4 presents the confusion matrix illustrating the performance of the algorithm in classifying normal and anomalous data.

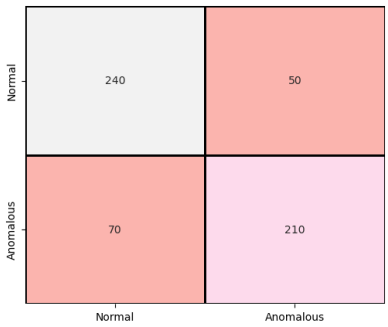

Fig.4. Confusion matrix for data error detection of the High-Speed Shaft Upwind Bearing Radial sensor.

This confusion matrix reports a total of 240 true negatives (TN) and 210 true positives (TP), indicating that the model correctly classifies a significant proportion of normal and anomalous cases. However, the presence of 50 false positives (FP) and 70 false negatives (FN) suggests that classification errors still persist.

Table 2 presents the results of the proposed algorithm using data from the High-Speed Shaft Upwind Bearing Radial Sensor, including accuracy, which represents the overall classification success rate; recall, which quantifies the proportion of correctly identified positive cases; and the F1-score, which is the harmonic mean of precision and recall.

Table 2. - Results obtained with data from the High-Speed Shaft Upwind Bearing Radial Sensor.

Class	Precision	Recall	F1-score
Normals	0.774	0.82	0.799
Anomalous	0.8076	0.75	0.777

The model achieved an overall accuracy of 78.95%, with an accuracy of 80.77% and a recall of 75.00% in anomaly detection, demonstrating a strong capability for correctly identifying anomalous events. Similarly, the normal class exhibited an accuracy of 77.42% and a recall of 82.76%, reflecting reliable performance in detecting fault-free states. Additionally, the execution time was 7.8 seconds, indicating that this approach provides a simpler and faster alternative, facilitating its application in large-scale data analysis.

Similarly, for the High-Speed Shaft Downwind Bearing Radial Sensor, 70% of the total dataset (14,000 records) was used for training the model, while the remaining 30% (6,000 records) was allocated to the test set.

Figure 5 presents the confusion matrix, illustrating the performance of the proposed algorithm in classifying normal and anomalous data using measurements from the High-Speed Shaft Downwind Bearing Radial Sensor.

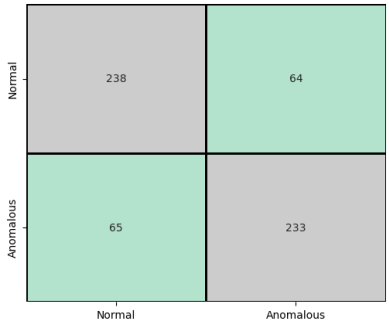

Fig.5. Confusion matrix for data error detection of the High-Speed Shaft downwind bearing radial sensor.

This confusion matrix demonstrates a balanced performance of the model in detecting anomalies, with 238 true negatives (TN) and 233 true positives (TP), indicating that the model correctly classifies a significant proportion of normal and anomalous cases. However, the presence of 64 false positives (FP) and 65 false negatives (FN) suggests that classification errors persist.

Table 3 presents the results of the algorithm using data from the High-Speed Shaft Downwind Bearing Radial Sensor.

Table 3. - Results obtained with data from the High-Speed Shaft downwind Bearing Radial sensor.

Class	Precision	Recall	F1-score
Normals	0.7855	0.7881	0.7868
Anomalous	0.7845	0.7819	0.7832

The model achieved an overall accuracy of 78.50% in classifying both categories. The similar values of precision, recall, and F1-score for both classes suggest that the dataset is of sufficient quality to train a reliable model. However, the presence of false negatives indicates challenges in correctly identifying certain anomalous instances.

Furthermore, the execution time was 8.1 seconds, demonstrating that this approach provides a computationally efficient and scalable solution, facilitating its application in large-scale data analysis.

Compared to previous studies, the proposed model demonstrates competitive performance in wind turbine anomaly detection. Hansi Chen [10] achieved an accuracy of 92.95% using LSTM and Autoencoders, Wang Meng-Hui [11] obtained 87.8% with MCNN, and Jiarui Liu [12] reported accuracies of 87.2% and 87.0% with Triplet-ConvDAE. In this study, accuracies of 78.95% and 78.50% were obtained using the High-Speed Shaft Upwind and Downwind Bearing Radial Sensors.

Although the accuracy values are lower than those reported in previous studies, this approach is distinguished by its computational efficiency and simplicity, as it operates directly with real data without requiring complex computational processes. Additionally, the processing time is short, averaging 7.94 seconds, further reinforcing its efficiency and making it a viable alternative for the analysis of large-scale datasets.

5. Conclusions and future work

This study examined the effectiveness of a support vector machine (SVM)-based model, optimized using a Bayesian search algorithm, for wind turbine anomaly detection based on vibration data from two different sensors: the High-Speed Shaft Upwind Bearing Radial Sensor and the High-Speed Shaft Downwind Bearing Radial Sensor.

The application of support vector machines (SVM) in this study demonstrated effective classification of wind turbine anomalies by identifying patterns in vibration data. The optimization of hyperparameters using a Bayesian algorithm improved the model's accuracy and enhanced the balance in fault classification, reducing the risk of overfitting. Collectively, these factors contribute to a robust and reliable approach for wind turbine fault detection.

As future research directions, several improvements are proposed, including the application of advanced feature extraction techniques for vibration signals, such as wavelet transform, to enhance the identification of patterns associated with failures. Additionally, the integration of multiple data sources will enable a more comprehensive and accurate analysis of the operational status. Another potential enhancement involves extracting information to explain the factors that influenced the anomaly classification process [14].

Acknowledgement

This work has been partially funded by the Spanish Ministry of Science and Innovation, under the MCI/AEI/FEDER project number PID2021-123543OBC21.

References

[1] T. Matsui, K. Yamamoto, and J. Ogata, "Anomaly detection for wind turbine damaged due to lightning strike," *Electric Power Systems Research*, vol. 209, Aug. 2022, doi: 10.1016/j.epsr.2022.107918.

[2] R. Pandit, D. Infield, and M. Santos, "Accounting for Environmental Conditions in Data-Driven Wind Turbine Power Models," *IEEE Trans Sustain Energy*, vol. 14, pp. 168–177, Jan. 2023, doi: 10.1109/TSTE.2022.3204453.

[3] Global Wind Energy Council, "Global Wind Report 2024." Accessed: Dec. 19, 2024. [Online]. Available: https://gwec.net/global-wind-report-2024/

[4] J. Chen, J. Li, W. Chen, Y. Wang, and T. Jiang, "Anomaly detection for wind turbines based on the reconstruction of condition parameters using stacked denoising autoencoders," *Renew Energy*, vol. 147, pp. 1469–1480, Mar. 2020, doi: 10.1016/J.RENENE.2019.09.041.

[5] E. S. Miele, F. Bonacina, and A. Corsini, "Deep anomaly detection in horizontal axis wind turbines using Graph Convolutional Autoencoders for Multivariate Time series," *Energy and AI*, vol. 8, May 2022, doi: doi.org/10.1016/j.egyai.2022.100145.

[6] R. Pandit, M. Santos, and J. E. Sierra-García, "Comparative analysis of novel data-driven techniques for remaining useful life estimation of wind turbine high-speed shaft bearings," *Energy Sci Eng*, Oct. 2024, doi: 10.1002/ese3.1911.

[7] M. Serrano-Antoñanazas, J. E. Sierra-Garcia, M. Santos, and M. Tomas-Rodríguez, "Identification of Vibration Modes in Floating Offshore Wind Turbines," *J Mar Sci Eng*, vol. 11, Oct. 2023, doi: 10.3390/jmse11101893.

[8] H. Chen, H. Liu, X. Chu, Q. Liu, and D. Xue, "Anomaly detection and critical SCADA parameters identification for wind turbines based on LSTM-AE neural network," *Renew Energy*, vol. 172, pp. 829–840, Jul. 2021, doi: 10.1016/j.renene.2021.03.078.

[9] M.-H. ; Wang *et al.*, "Fault Detection of Wind Turbine Blades Using Multi-Channel CNN," *Sustainability 2022, Vol. 14, Page 1781*, vol. 14, no. 3, p. 1781, Feb. 2022, doi: 10.3390/SU14031781.

[10] J. Liu, G. Yang, X. Li, Q. Wang, Y. He, and X. Yang, "Wind turbine anomaly detection based on SCADA: A deep autoencoder enhanced by fault instances," *ISA Trans*, vol. 139, pp. 586–605, Aug. 2023, doi: 10.1016/J.ISATRA.2023.03.045.

[11] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: 10.1016/j.neucom.2019.10.118.

[12] S. M. Lundberg, G. G. Erion, and S.-I. Lee, "An interpretable model for landslide susceptibility assessment based on Optuna hyperparameter optimization and Random Forest," *Geomatics, Natural Hazards and Risk*, vol. 15, no. 1, Apr. 2024, doi: 10.1080/19475705.2024.2347421.

[13] V. R. Joseph, "Optimal ratio for data splitting," *Stat Anal Data Min*, vol. 15, pp. 531–538, Aug. 2022, doi: 10.1002/sam.11583.

[14] S. Cuéllar, M. Santos, F. Alonso, E. Fabregas, and G. Farias, "Explainable anomaly detection in spacecraft telemetry," *Eng Appl Artif Intell*, vol. 133, Jul. 2024, doi: 10.1016/j.engappai.2024.108083.