Wind Turbine Hybrid Physics-based Deep Learning Model for a Health Monitoring Approach Considering Provision of Ancillary Services

Nezmin Kayedpour (iD), Jixiang Qing (iD), Jolan Wauters (iD),
Jeroen D. M. De Kooning (iD), Ivo Couckuyt (iD), and Guillaume Crevecoeur (iD)

Abstract—Assessing the overall condition of wind turbines in operation is challenging due to their intricate nature. This becomes even more complicated when wind turbines provide ancillary services and respond to grid requirements under curtailment modes. Multiple models are required to effectively evaluate the wind turbines' healthy condition, which can be unmanageable and impractical, particularly for large-scale wind farms. This article proposes a novel hybrid physics-based deep learning framework to accurately approximate the time-varying correlation between control sequences and system response, reflecting the aerodynamic nonlinearity of the 5-megawatt offshore wind turbine model, designed and tested by the National Renewable Energy Laboratory (NREL). Another layer of this study's novelty relies on proposing a computationally efficient weakly supervised method that uses the hybrid structure to detect degradations and anomalies considering curtailment operation. Then, a self-learning classification approach is employed to iteratively update the best-tuned classifier, dynamically learning unforeseen abnormalities from brand-new anomalies during active operations. The proposed anomaly detection strategy deals with system uncertainties, such as wind stochasticity, power curve variations, and different sparsity levels in the datasets. The results of the proposed approach show promise in improving health monitoring performance, leading to a more efficient and accurate assessment of the overall condition of wind turbines.

Index Terms—WT, frequency containment reserve (FCR), anomaly detection, hybrid models, adaptive neuro-fuzzy inference system (ANFIS), long short-term memory (LSTM), support vector machine (SVM), temporal convolutional network (TCN).

I. INTRODUCTION

WITH the growing number of wind energy conversion systems, ensuring reliable operation that considers grid balancing provision through ancillary services while lowering maintenance costs and reducing downtime is necessary. However, wind turbines (WTs) are complicated systems containing multiple subsystems. Anomalies and faults can occur due to various factors, leading to failure. Therefore, it is crucial to have a condition monitoring system (CMS) that can detect these issues early on to improve the system's reliability and to lower the Levelized Cost Of Energy (LCOE) [1].

The fast development of big data techniques plays an evolutionary role in WT Health monitoring and predictive maintenance strategies. In recent years, there has been a significant amount of research interest in monitoring the conditions of WTs using hybrid modeling, measured data, and deep learning methods to carry out a wide range of tasks, from detecting WT anomalies [1], [2], gearbox fault diagnosis [3]–[7], blade icing detection [8], [9], and predicting failures [10] to estimating components remaining useful life [11]–[14].

Among the latest studies, some efforts aim to extract high-level features from data, obviating the need for specialized domain knowledge. In [15], a deep learning classifier for WT gearboxes employs a stacked auto-encoder. Similarly, [16] uses an adaptive algorithm to capture evolving sensor weights for WT health assessment. In [17], an interactive spatiotemporal model extracts features, while [18] introduces an adaptive gated attention mechanism for fault feature extraction. Nevertheless, emphasis is placed on achieving accurate modeling, given the intricate and aerodynamic nature of WT data connections to unveil hidden nonlinear patterns, ensuring an overall recognition of rapidly altering operating conditions, i.e., maximum power point tracking (MPPT) mode, power regulation, and the transitions zone, and improving the overall effectiveness of WT condition monitoring, fault diagnosis, and lifetime prognosis [19].

Data-driven methods and sensor data analytics are powerful tools to mirror and predict the system performance [20]. A machine learning-based surrogate structure as a virtual model and a proxy to the actual high-fidelity model of an existing system can emulate the behavior of a physical system depending on the entity's design and operation. [21] proposes using a surrogate model to calculate extreme WT tower loads using various signals and a suitable simulation tool. Also, surrogate models based on polynomial chaos expansion (PCE) and Kriging are suggested in [22] to approximate WT fatigue loads. Long short-term memory (LSTM) is a recurrent neural network (RNN) that is often used in WT surrogate modeling due to its ability to capture temporal dependencies and sequential patterns in time series data, which is common in WT operational data [23], [24]. Another robust and effective method to estimate the nonlinear behavior of dynamical operations is the adaptive neuro-fuzzy inference system (ANFIS), which enables expressing uncertain

N. Kayedpour, J. Wauters, J. D. M. De Kooning, and G. Crevecoeur are with the Department of Electromechanical, Systems & Metal Engineering, Ghent University, Tech Lane Ghent Science Park - Campus Ardoyen, Technologiepark-Zwijnaarde 131, B-9052 Ghent, Belgium. N. Kayedpour, J. Wauters, J. D. M. De Kooning, and G. Crevecoeur are also with FlandersMake@UGent - Corelab MIRO. J. Qing and I. Couckuyt are with the Department of Information Technology (INTEC), IMEC, IDLab, Ghent University, Tech Lane Ghent Science Park – Zwijnaarde 126, B-9052 Ghent, Belgium (email: nezmin.kayedpour@ugent.be; jixiang.qing@ugent.be; jolan.wauters@ugent.be; jeroen.dekooning@ugent.be; ivo.couckuyt@ugent.be and guillaume.crevecoeur@ugent.be).

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT

2

circumstances in the form of rules by using the (if-then) decision-making mechanism [25]. ANFIS and fuzzy logic approaches can show superior performance, short execution time, and accuracy, especially in WT applications that possess stochastic aerodynamic characteristics [26]–[29]. Introducing a convolutional neural network (CNN) has broadened the possibilities for time series prediction methods and made them no longer limited to RNNs. CNNs offer the advantage of parallel processing and the ability to expand their receptive fields. This capability allows CNNs to access a more comprehensive historical context, potentially mitigating issues associated with long-term dependencies. The temporal convolutional network (TCN) represents an advanced refinement of the CNN architecture. It does so by employing dilated causal convolutions to uncover important historical information [30]. Interestingly, enlarging the receptive field using dilated causal convolutions results in only a minor increase in the network's layer count and parameter complexity. As a result, TCN excels in the domain of time series prediction, surpassing the predictive capabilities of standard CNNs [30]. Promising outcomes of TCN in predictive tasks, such as predicting the RUL of rotating machinery and forecasting wind speed intervals for WTs, are discussed in [14], [31], [32].

The Transformer model is another deep learning architecture, which recently has become very popular due to its effectiveness in capturing long-range dependencies in sequences [33], [34]. The attention-based transformer model addresses some of the limitations of recurrent and convolutional neural networks in handling sequential data and is recently developed for remaining useful life (RUL) estimation approaches [35]. However, despite the mentioned benefits of Transformers, their adoption for real-time inference is hindered by demanding computational requirements [36]. This complexity in deployment poses a notable challenge for practical applications where data might be limited.

Nevertheless, applying pure data-driven techniques to WT modeling faces significant limitations, emerging from insufficient data quality and quantity of vital parameters like wind and rotor speed, power, and control performance metrics [37]. By juxtaposing power predictions with real-time data and combining physics-based and data-driven methods, a comprehensive and accurate representation of the turbine's behavior can result in successful flag-up deviations, particularly in scenarios involving stochastic processes or varying grid conditions. However, the current hybrid methods often focus on limited parameters and elements, solely on key components such as the gearbox, the generator, blade bearings, or power quality [38]. Furthermore, power prediction within WTs plays a pivotal role in wind energy management and power forecasting. When combined with data analysis and monitoring methods, this predictive capacity not only aids energy generation management and grid integration but also facilitates the detection and diagnosis of significant powertrain degradation and failures [38]. On the other hand, accurate modeling and a standalone condition monitoring system (CMS) for the entire WT can be costly, requiring extra investments [1]. While significant progress has been made through these techniques, a comprehensive evaluation of the entire WT is often challeng-

ing to attain [1], [39]. The issue becomes even more complex when the selected monitoring indicator fails to detect the fault signatures, as these could be masked by the condition parameters of another component or different operating policies, such as power degradation related to the Frequency Containment Reserve (FCR) provision [40], which has not been taken into consideration in the previous studies. Multiple models have been produced due to various studies attempting to tackle the issue by combining various methods that increase the system's complexities. However, they are found to be less efficient and more costly to manage, particularly in large floating offshore wind farms [41]. This underscores the need for an integrated approach that comprehensively addresses the intricacies and challenges of WT health assessment and condition monitoring.

This study addresses three key challenges that hinder a comprehensive evaluation of WT performance. Firstly, existing methods focusing on single turbine components fail to provide an overall performance assessment, disregarding complex interdependencies within the system [38]. Secondly, managing multiple models for different aspects of performance is often inefficient and computationally expensive [1]. Thirdly, the impact of WT curtailment operations on system health for grid balancing purposes is often overlooked [42]. To bridge these gaps, we propose a hybrid framework that predicts the WT overall performance by coupling physical equations representing the estimation of controlled parameters responding to various operating conditions and a deep learning surrogate model predicting WT aerodynamic behavior, allowing a more accurate reflection of healthy behavior. We explored the predictive capabilities of three different deep learning approaches commonly discussed in literature, i.e., ANFIS, LSTM, and TCN, ensuring the surrogate model's accuracy and computational efficiency. Throughout our investigation, we aimed to highlight the distinctive characteristics of each method, shedding light on their performance differences. This study particularly addresses the third gap by explicitly considering curtailment-related degradation that can be falsely detected as performance degradation or any other anomalies. Additionally, we employed a weakly supervised approach using limited labeled faulty data to address the challenges posed by supervised and unsupervised methods and their requirements of extensive labeled data, which are not always practically available [37], [43].

The key contributions of this article can be outlined as follows:

- A hybrid framework is proposed that accurately predicts the WT's overall healthy performance by approximating the electrical power and rotational speed, not only considering the stochastic nature of wind speed but also complex correlations between control sequences of pitch-generator torque and system response in turbulent wind. These parameters are physically described and integrated into a deep learning surrogate model to effectively capture the system's nonlinearities across various stochastic conditions and operational modes. Employing the proposed hybrid structure facilitates anomaly detection by distinguishing between normal and abnormal states, indicating deviations (residuals).

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 3

- A self-learned classification approach with an iterative framework is investigated that improves the classifier's performance by dynamically updating newly labeled anomalies from former successful classifications. Support Vector Machines (SVMs) are employed for classifying the faultiness and degradation, incorporating coherent features that can be extracted from the plant's main observables, i.e., electrical power, and rotational speed, in time and frequency domains for seamless integration into the system's condition indicators.

- This study also considers a wide range of anomalies and degradation scenarios, including degradation due to a curtailment operation providing FCR, blade pitch control failure, yaw misalignment, and Permanent Magnet Synchronous Generator (PMSG) abnormalities. The suggested approach addresses the intricacies and inter-dependencies present in WT performance under varying grid requirements and operating conditions. The methodology's validation was performed on a realistic 5MW offshore floating WT using the NREL FAST software. This simulation integrates detailed models of the WT's nonlinear aerodynamics, providing a realistic environment for comprehensive assessment.

This article is organized as follows: Section II discusses the hybrid physics-based deep learning structure. Section III describes the proposed condition monitoring approach. The performance assessment, data, simulations, and results are presented in Section IV. Finally, in Section V, the findings are discussed, and conclusions are drawn.

II. HYBRID PHYSICS-BASED DEEP LEARNING MODEL

A. Underlying WT physical system

This study employs a 5MW offshore WT with variable blade-pitch-to-feather configuration and an operational control approach based on power-production regulation using pitch and torque control systems. To accurately model the behavior of the WT, each subsystem is described and modeled separately. This involves developing a detailed model of the aerodynamics, control system, and electrical characteristics of the Permanent Magnet Synchronous Generator (PMSG). These models are then integrated into a closed-loop system to study WT's dynamic behavior in various operating conditions.

1) Dynamic model: To investigate the dynamic behavior of WTs, the study employs TurbSim to generate time series data for the three Cartesian wind components within a dimensional grid. This data is generated based on statistical models, effectively simulating the full-field wind speed distribution. Subsequently, the generated wind data undergoes analysis to assess its spectral and coherence properties in the frequency domain. By applying an inverse Fourier transform, the data is transformed into wind speed time series, a crucial preparatory step for its integration into the time domain-oriented FAST simulation tool. This comprehensive approach ensures the faithful representation of wind conditions, ultimately facilitating a robust exploration of WT dynamics and performance [44], [45]. The WT captures a total amount of
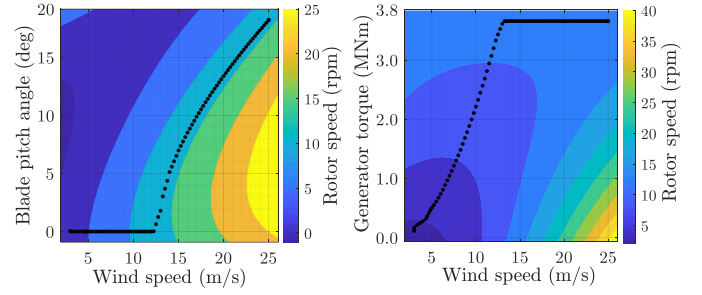

Figure 1: Nonlinear mapping between wind speed, WT generator torque, blade pitch angle, and rotational speed.

mechanical power P_m, and the mechanical torque T_m of the WT can be described using the following relationships:

$$P_m = Av^3C_p(\lambda, \theta), A = \frac{1}{2}\rho\pi R^2, \tag{1}$$

$$T_m = Av^3C_p(\lambda, \theta) \cdot \frac{1}{\omega_r}, \tag{2}$$

where C_p represents the power coefficient, ρ is the air density, R denotes the blade length and θ is the pitch angle. The tip-speed ratio $\lambda = \omega_r R / v$ is a function of wind v and rotational speed ω_r. FAST implements the Blade Element Momentum (BEM) theory and simulates the nonlinear equations of motion. It also determines the WT's aerodynamic and structural response to wind-inflow conditions in time, which is advantageous for developing control designs and analysis [46].

2) Generator: The dynamic equivalent model of the PMSG can be formulated in the q,d rotating reference frame:

$$V_d = R_sI_d + L_d\frac{dI_d}{dt} - N_p\omega_r L_qI_q, \tag{3}$$

$$V_q = R_sI_q + L_q\frac{dI_q}{dt} + N_p\omega_r(L_dI_d + \Phi_m), \tag{4}$$

where R_s is the stator-winding resistance, L_d and L_q are the d-axis and q-axis stator-inductances and Φ_m is the flux linkage. V_d and I_d are the d-axis stator voltage and current. V_q and I_q are the q-axis stator voltage and current. N_p is the number of pole pairs. The generator torque and electrical power can be formulated as follows:

$$T_g = \frac{3}{2}N_P\left(\Phi_mI_q + (L_d - L_q)I_dI_q\right), \tag{5}$$

$$P_e = \frac{3}{2}[V_dI_d + V_{sq}I_{sq}]. \tag{6}$$

The WT system's equation of mechanical motion is formulated as follows:

$$T_m - T_g = J \cdot \frac{d\omega_r}{dt} + B_f \cdot \omega_r, \tag{7}$$

where, J represents the overall moment of inertia, while B_f denotes the coefficient associated with viscous friction. The machine's realistic dynamics and losses, including machine inductances, the armature reaction effect, stator winding copper losses, and iron core losses, are considered and included in the efficiency curve as proposed in [47].
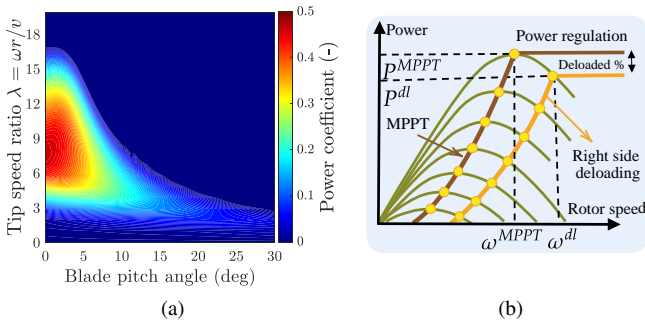
THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT
4

Figure 2: a) WT optimal power coefficient for normal operation. b) WT power modification for curtailed operation.

3) Control system and FCR provision: The WT control design includes two main controllers: a generator-torque controller and a full-span rotor-collective blade-pitch controller. These controllers operate across all operational regions. In wind speeds below the rated level, the pitch angle is set at zero degrees, and the torque controller optimally maximizes wind power extraction by keeping λ and consequently C_p at the optimal level, which is $C_p^{opt} = 0.482$ and $\lambda_{opt} = 7.55$ respectively. Conversely, for wind speeds above the rated level, the pitch controller maintains rotational speed using a gain-scheduled PI controller. In this operating region, there is no rotor acceleration by applying a rated generator torque T_{g-equ} that cancels the aerodynamic torques at equilibrium. Figure 1 illustrates the nonlinear relationships among wind speed, generator torque, blade pitch angle, and rotational speed. The pitch controller adapts to the nonlinear aerodynamic characteristics at different operating points, as determined through linearization analysis in FAST. The control system includes a transition zone between partial and full load to ensure a smooth transition between maximum power point tracking and power regulation. The pitch angle θ can be approximated considering the pitch control proportional K_p and integral K_i, to keep the rotational speed at the rated value ω_{ref}:

$$\theta \approx \theta_{\text{ref}} = K_p \delta\omega_r + K_i \int_0^t \delta\omega_r dt, \qquad (8)$$

$$\delta\omega_r = \omega_{\text{ref}} - \omega_r, \qquad (9)$$

where, ω_{ref} is the rated rotational speed of 12.1 rpm for 5MW offshore WT. The proportional K_p and integral K_i scheduled gains that are calculated by multiplying the gain correction factor $GK(\beta) = \frac{1}{1+\beta/\beta_K}$ to constant values, considering the aerodynamic pitch sensitivity of WT $\delta P/\delta\beta$ [46]. Moreover, the control system also governs the yaw angle during normal operation, ensuring the nacelle remains aligned with the wind direction. This dynamic adjustment minimizes mechanical stress, enhances energy output, and mitigates potential damage from extreme wind conditions. In curtailment mode, WTs are intentionally operated at less than their maximum power output capacity. In curtailment mode, a power reserve is needed to let the WT respond to grid frequency variations with primary control architecture[48]. This means that the turbine's blades and generator torque are adjusted to capture less energy from the wind than they would at their optimal operating

conditions. The power coefficient in curtailment mode C_p^{cur} would generally be lower than the optimal power coefficient C_p^{opt} in full-healthy operating mode by applying a deloaded factor β (in percentage) that determines WT contribution in FCR market, which can be formulated as follows:

$$C_p^{\text{cur}} = \beta \cdot C_P^{\text{opt}}. \qquad (10)$$

Fig.2a shows WT optimal power coefficient in relationship with tip speed ratio and blade pitch angle. To let the WT respond to grid frequency variations, a supplementary FCR control loop is used, and the rotor speed-power lookup table is modified by shifting the WT's operating point to the right side of the Maximum Power Point Tracking (MPPT) curve [49]. The curtailed operation lets the WT to regulate the deloaded power P^{dl}, considering the adjustment of rated rotor speed in deloading mode ($\delta\omega_r^{\text{dl}} = \alpha\omega_{\text{ref}} - \omega_r$), applying the deloaded generator torque T_{g-equ}^{dl}. Figure 2b shows WT power modification for curtailed operation as discussed in [48]. The WT overspeeding factor α ensures sufficient reserve at sub-optimal performance while enabling the WT to respond to grid frequency variations Δf proportionally, even when wind speed is at the below-rated value.

B. Hybrid framework

Recently, a robust baseline approach has been suggested in literature that leverages a predictive model and estimates a healthy power curve, achieving a normal behavior model (NBM) based on measured wind speeds for assessing the overall health condition of the entire WT [1]. However, several abnormalities can induce similar degradation impacts on the power curve while impacting rotational speed differently. For instance, increasing and decreasing rotor speed above or below rated wind speed has the same electrical power degradation effect. Therefore, yaw misalignment in below-rated wind conditions can be misinterpreted as curtailment degradation if we only consider the power cure prediction. However, observing the rotational speed more accurately indicates the related deviation. This is mainly because in yaw misalignment the rotor speed drops below the MPPT curve, while it will be moved to above the MPPT curve for curtailment reasons to avoid losing kinetic energy that can be used for supporting inertial response [48].

On the other hand, PMSG abnormalities are more evident by observing the electrical power, and they have less impact on the WT's rotational speed. Therefore, in the proposed hybrid framework, we suggest closely monitoring both the WT power and rotational speed to evaluate the overall performance of the WT in different operating conditions. Additionally, the proposed architecture aims to approximate the time-varying correlation between control inputs and system response while ensuring the process dynamics are enforced within the network. The baseline approach only considers wind speed as the main independent input. However, the mentioned controlled parameters can directly impact the aerodynamic nonlinearity of the system. They can be estimated by governing physical equations that provide the pitch and generator torque signals based on the central independent input, wind speed, and rotational speed as a feedback signal. The mathematical equations

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

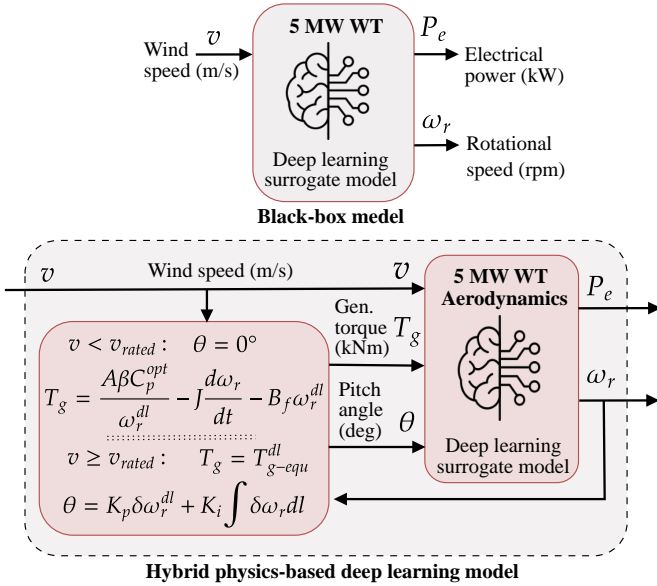IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 5

Figure 3: Proposed hybrid framework.

are adjusted to respond to wind speed variations, and the active power, which should be decided according to the power grid fluctuations at the supplementary FCR control loop. Then, a nonlinear mapping of multiple inputs, i.e., wind speed, blade pitch angle, and generator torque, into multiple outputs, i.e., rotor speed and mechanical power, is carried out by training a deep learning surrogate model, using a data-set gathered offline in various possible operating conditions.

Figure 3 presents the suggested hybrid framework alongside a black-box data-driven model. The baseline balck-box model predominantly relies on wind speed as its main input and is commonly studied in the literature as a benchmark for predicting WT power curve [50]. The hybrid model, on the other hand, goes a step further by estimating pitch and generator torque both below and above the rated wind speed, utilizing wind and rotor speed estimations. Subsequently, it predicts the aerodynamic behavior of the WT by establishing a relationship between healthy control parameters and wind speed.

In the following subsection, we go into the methodological aspects of the applied deep learning surrogate model and the structure design of ANFIS, LSTM, and TCN as the three potential networks with strong abilities to learn nonlinear behavior with stochastic and fast dynamic characteristics. This exploration will provide a deeper understanding of their mechanism and their respective applications in the discussed hybrid model. Furthermore, we will conduct a comprehensive assessment to evaluate and compare the effectiveness and practicality of these deep-learning approaches in accurately capturing the intricate nonlinearities inherent in the WT.

C. Structure of the surrogate models

1) Adaptive neuro-fuzzy inference system (ANFIS): A hybrid learning algorithm of both the least-squares method and backpropagation learning is used to train the network and optimize the parameters of a fuzzy model capable of handling both quantitative and qualitative criteria. The nonlinear mapping is carried out using the Takagi–Sugeno inference

model employing fuzzy if-then rules [26], [29], [51]. In the fuzzification layer of the ANFIS structure, the Gaussian Membership Functions (MF) of the crisp inputs are created by:

$$\mu_{A_i}(T_g), \ \mu_{B_j}(v), \ \mu_{C_k}(\theta), \ i, j, k = 1, ..., n, \quad (11)$$

$$\mu_x = e^{-\left(x - \frac{a_i}{b_i}\right)^2}, \quad (12)$$

where μ_{A_i}, μ_{B_j} and μ_{C_k} are the MFs of fuzzy sets, which have a Gaussian form characterized by the variance a_i and center b_i of the MF. In the rule layer, each node output is denoted by the fuzzy inference system representing the firing strength of a rule W_p, which is calculated by the multiplication of incoming signals (12). The purpose of the normalization layer is to normalize the weight function using (13).

$$W_p \begin{cases} \mu_{A_i}(T_g) \cdot \mu_{B_j}(v), \\ \mu_{B_j}(v) \cdot \mu_{C_k}(\theta), \qquad i, j, k = 1, ..., n, \quad (13) \\ \mu_{A_i}(T_g) \cdot \mu_{C_k}(\theta), \end{cases}$$

$$\bar{W}_p = \frac{W_p}{\Sigma W_p}, \quad p = 1, ..., m. \quad (14)$$

In the defuzzification layer, the output of nodes will be defined as the product of first order polynomials f_p and normalized firing strength $\bar{W}_p$, where f_p represents the fuzzy If–then rules:

$R1$: If $T_g = A_n$ and $v = B_n$, Then $f_n = \alpha_n T_g + \beta_n v + r_n$,

$R2$: If $v = B_n$ and $\theta = C_n$, Then $f_m = \beta_n v + \gamma_n \theta + r_n$,

$R3$: If $T_g = A_n$ and $\theta = C_n$, Then $f_m = \alpha_n v + \gamma_n \theta + r_n$,

where $\{\alpha, \beta, \gamma, r\}$ is the resultant attribute set, which belongs to each node. Finally, one node represents the sum layer, which calculates the total output summation of all arriving signals by:

$$Y = \sum_{p=1}^{m} \bar{W}_p f_p. \quad (15)$$

In this study, the Fuzzy C-means (FCM) method is used to compute the membership degrees, minimizing the following objective function:

$$\mathcal{J}_q = \sum_{l=1}^{N} \sum_{s=1}^{G} u_{ls}^q \|x_l - G_s\|^2, \quad 1 < q < \infty, \quad (16)$$

where N is the size of the data set, q is a weighted index, u_{ls} is the degree of membership of x_l in the cluster q, which is lth of d-dimensional measured data. G_s is the d-dimension center of the cluster. A Genetic Algorithm (GA) is used to find the optimal weighting exponent q value for the FCM algorithm. The q value determines the fuzziness degree in the clustering process and affects the performance. The goal is to partition the dataset into the desired number of classes and to calculate the cluster centers and membership degrees that assign data points to clusters.

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT

6

2) Long short-term memory (LSTM) network: The LSTM network is a Recurrent Neural Network (RNN) based architecture that has been commonly used for sequence regression-related application problems, e.g., [52], [53]. The main advantage of LSTM over standard RNN is that it is able to capture not only the short-term temporal relations in sequence but also the long-term relationship. This has been achieved by its specifically designed network structure, as illustrated in Fig.4. For a specific LSTM unit, if we denote the input at time step t as $\boldsymbol{X}_t := [v_t, T_{g_t}, \theta_t]$, the flow within the LSTM unit $\text{LSTM} : \boldsymbol{X}_t \to \text{LSTM}(\boldsymbol{X}_t)$ can be represented as:

$$\begin{aligned}
\boldsymbol{f}_t &= \sigma(\boldsymbol{W}_f[\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_f), \\
\boldsymbol{i}_t &= \sigma(\boldsymbol{W}_i[\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_i), \\
\hat{\boldsymbol{C}}_t &= \tanh(\boldsymbol{W}_{\hat{C}}[\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_{\hat{C}}), \\
\boldsymbol{o}_t &= \sigma(\boldsymbol{W}_o[\boldsymbol{h}_{t-1}, \boldsymbol{X}_t] + \boldsymbol{b}_o), \\
\boldsymbol{C}_t &= \boldsymbol{f}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \odot \hat{\boldsymbol{C}}_t, \\
\boldsymbol{h}_t &= \boldsymbol{o}_t \odot \tanh(\boldsymbol{C}_t),
\end{aligned} \tag{17}$$

where σ, $\tanh$ represents the sigmoid and tanh activation functions respectively; $\boldsymbol{W}_f, \boldsymbol{W}_o, \boldsymbol{W}_{\hat{C}}$ represent the weights and $\boldsymbol{b}_f, \boldsymbol{b}_i, \boldsymbol{b}_{\hat{C}}, \boldsymbol{b}_o$ denote the biases. $[\cdot]$ represents the concatenation operation and $\odot$ is the Hadamard product. $\boldsymbol{o}_t$ is the output gate's activation vector, $\boldsymbol{C}_t$ is the output cell state and $\boldsymbol{h}_t$ is the hidden state that will be provided for $t + 1$ steps recurrently. Importantly, we remark that $\boldsymbol{C}_t$ is the key factor that enables capturing long-term patterns. With the elaborated LSTM unit, in order to increase the flexibility of the network to capture the potential complex temporal correlations between the input and output of the surrogate model, we construct an LSTM-based network architecture illustrated in Fig.4: the constructed architecture consists of 3 LSTM layers, with two dense layers afterward. We note that this network architecture is empirically determined as leading to a satisfying performance in reality. For input $\boldsymbol{X}_t$ any time steps t, this architecture has the following output:

$$\boldsymbol{h}_t = \text{LSTM}_3(\text{LSTM}_2(\text{LSTM}_1(\boldsymbol{X}_t))), \tag{18}$$

$$\boldsymbol{y}_t = \text{Dense}_2(\text{Dense}_1(\boldsymbol{h}_t)), \tag{19}$$

where LSTM gives one LSTM layer, Dense $:= I(\boldsymbol{W}\boldsymbol{x} + \boldsymbol{b})$ represents one fully connected layer, where the linear activation function is used. One hundred twenty-eight hidden units within each LSTM layer are used for the other hyperparameter of the network. The hidden unit for the first dense layer is 128, and 3 for the second layer.

To optimize the weights of the proposed LSTM network architectures, we utilize the mean squared error as the objective function commonly used in a regression problem. The Adam optimizer [54] is utilized as a stochastic gradient-based optimizer. The model is implemented utilizing Keras under Tensorflow 2, and the parameters are trained with 100 epochs with a mini-batch size of 250.
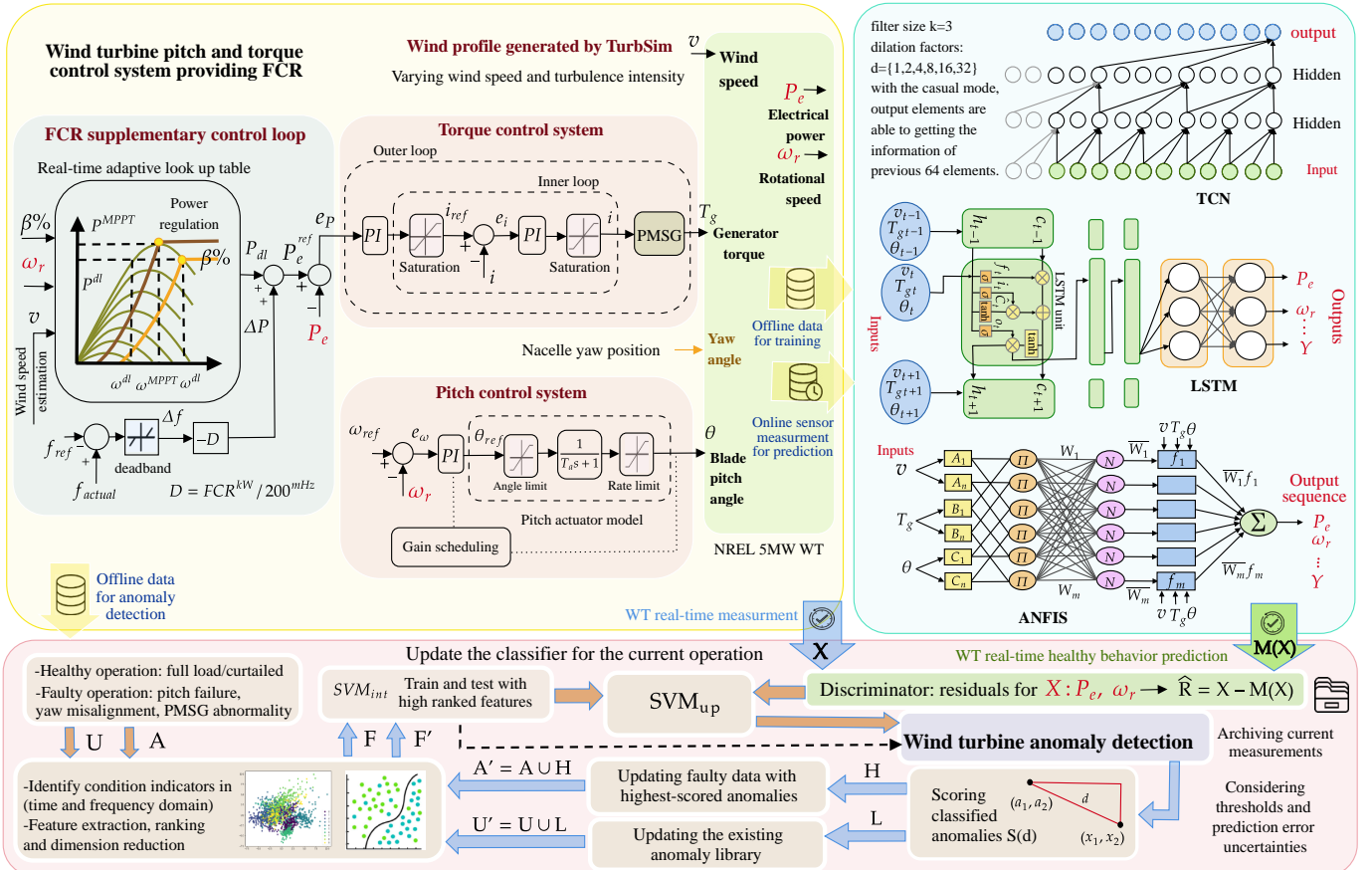

Figure 4: The proposed health monitoring framework using the hybrid model.

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 7

3) Temporal convolutional neural network (TCN): TCN is a convolutional type neural network. It works similarly to a standard convolutional neural network while the convolution operates on the time series. The convolutional kernel can either be causal (as depicted in Fig. 4), preventing any information leakage from the future time step, or use these parts of information if in a feasible scenario. With stacked dilated convolutions, the model is able to effectively expand the receptive field of the convolutional network, hence getting the information from a very long context. More formally, for a 1-dimensional time series input, and filters defined as $f \in \{0, \ldots, k-1\}$, the dilated convolution on an arbitrary element s of the sequence is defined as:

$$F(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{(s-d \cdot i)} \qquad (20)$$

where the element s will incorporate the information of the past up to $k(d-1)$ elements. For a more in-depth description of the neural network, we refer to the paper [30]. The TCN network structure to compare in this research consists of 3 stacked TCN layers. The filter size is set to 32, we use a kernel k with size 3, and a dilation of $d = \{1, 2, 4, 8, 16, 32, 64\}$ is utilized. For simplicity, not all hidden layers are shown in the TCN Structure in Fig. 4. A residual block is utilized (as shown in Fig. 1.b of [30]). After passing the 3 stacked TCN layers, a dense layer is set to extract the final output time series. We follow the same training routine as training LSTM and ANFIS models.

III. HEALTH MONITORING

This section introduces a health monitoring approach that utilizes the surrogate models proposed in Section II-C to assess the current condition of a WT and detect and diagnose anomalies. Figure 4 outlines the workflow of this approach, which commences with a data-gathering step that reflects the WT's operation under both healthy and faulty conditions. This step provides the model with measured data for anomaly detection and performance assessment. Once the surrogate model is created, the proposed hybrid structure can mimic the healthy behavior of the system and discern normal and abnormal behavior from the calculated deviations, i.e., the residuals. The next step involves using the extracted features that can be identified and incorporated into the system's condition indicators. A classifier is then trained using a small set of labeled anomaly data. In an iterative process, the highest-scored anomalies detected will be used to update the classifier by introducing more faulty sets to the initial dataset through an automatic or manual labeling method. Additionally, unknown anomalies (lowest or zero-scored data points) will be added to the existing library and updated by repeating the feature extraction and dimension reduction step. Finally, the classification model is updated using the brand-new archived dataset from the current operation, considering thresholds and the uncertainties of prediction errors. This updated model enables more accurate detection and diagnosis of anomalies in the system.
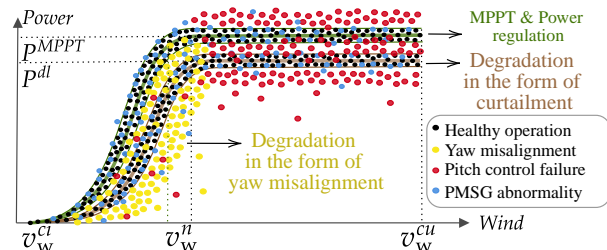

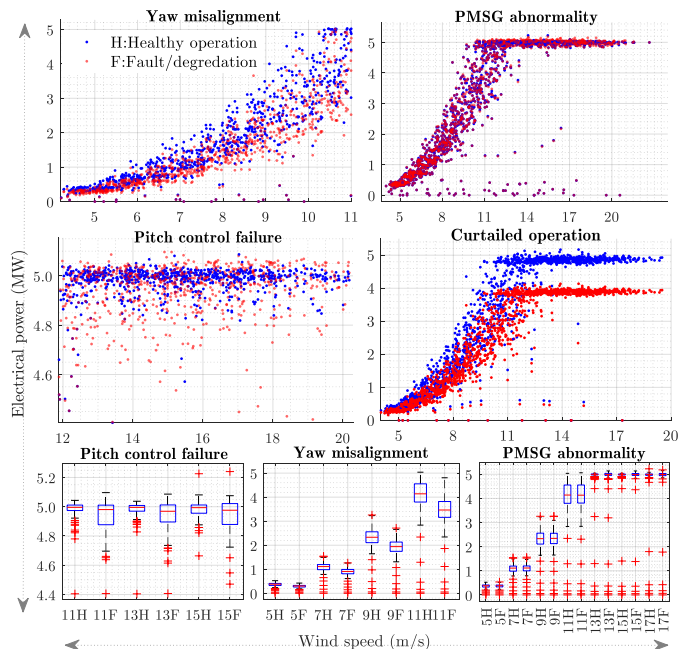Figure 5: Degradation in the form of anomalies/curtailment.

Figure 6: WT operation in abnormal conditions.

A. Anomaly and degradation scenarios

The suggested health monitoring approach is developed and assessed in various working conditions with different sources of faultiness, which have a high chance of occurrence and can be falsely interpreted as normal degradation in deloading operations with curtailment. As shown in Fig.5, two control failures, i.e., blade pitch angle and nacelle yaw position error, are considered, affecting the rotational speed and causing electrical power degradation in full and partial load regions, respectively. In this study, the blade pitch failure mode occurs when one or two blade pitch motor mechanisms fail to respond to the control signals, lock the blade at a certain position, and stop creating pitch-angle demands. Yaw misalignment is also implemented for yaw position errors from 5 to 20° when the WT is not fully facing the wind. This occurs when the wind direction changes and the yaw control system fails to orient the WT rotor towards the wind direction properly.

Moreover, PMSG abnormality is considered to assess the performance of the proposed condition monitoring approach in all operating regions. The presence of an electrical disturbance in the PMSG may occur at any operational condition and may interrupt or degrade electrical power. Abnormal behavior in the PMSG is attained by adding nontracked order and random
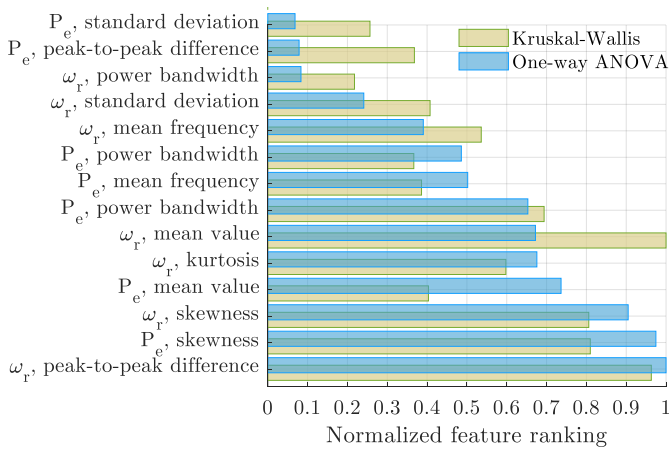
Figure 7: Features sorted by their rank.

noise to the back electromotive force (EMF) with a noise power of 1% to 5% of the EMF voltage and 10ms sample time. The data in healthy and faulty states are obtained by running numerous simulations in the incoming flow field with all operating ranges of wind speed and turbulence intensity levels of 5% to 15%. Figure 6 reveals the scatter plot of the WT healthy and 20% curtailed operation ($\beta = 80\%$) as well as the mentioned anomalies with boxplots demonstrating the anomaly locality, distribution, and skewness. The electrical power degradation is evident in pitch failure and yaw misalignment in full load and partial regions. The PMSG abnormality subtly impacts the electrical power in all operating regions compared to the two other anomalies. However, its impact hardly appears in the rotational speed time-series signal.

B. Feature extraction and dimension reduction

When calculating condition indicators as summary statistics, it is crucial to consider the system features that differentiate normal operations from abnormal behaviors, including degradation in the form of curtailment. A good understanding of the system is necessary to select appropriate condition indicators in two or multiple dimensions, and some experimentation may be required. In this study, we analyze several features in both the time and frequency domains for the electrical power and rotational speed signals, which can be combined to create condition indicators that capture the overall "unusualness"

of the data. The effectiveness of each feature in differentiating normality and abnormality is estimated and ranked using one-way ANalysis Of VAriance (ANOVA) and Kruskal-Wallis [55]. Figure 7 shows all the features used in this study, which are ranked by their importance. This work employs Principal Component Analysis (PCA) for efficiently reducing feature dimensions to enhance computational efficiency.

C. Classification

In this study, Support Vector Machines (SVM) are used to classify the obtained feature vectors, comparing the current observable value with the corresponding healthy value provided by the surrogate models. Then faulty conditions are estimated and the occurred faultiness is identified. Different kernel functions, which map the input samples into a higher dimensional space using a nonlinear function $\phi(\cdot)$ and soft margin hyperplanes separating the data in the higher dimensional space, are considered. The SVM, in general, solves the following quadratic optimization problem:

$$\min_{W,b,\xi} \quad \frac{1}{2}\|W\|_2^2 + C\sum_{i=1}^{N}\xi_i \quad , \qquad (21)$$
$$s.t. \quad y_i(W^T\Phi(x_i) + b) \geqslant 1 - \xi_i$$

where (x_i, y_i) denotes the training set, ξ is the slack variable that allows the hard margin to be violated, W and b are N-dimensional vectors, and the offset defines the hyperplane equation. The parameter C controls the trade-off between achieving a larger margin and minimizing the number of misclassifications. Then, a kernel function $k(X_s, X_s')$ is explicitly defined to calculate the inner product in the image of the nonlinear mapping function $\xi(\cdot)$. The Gaussian, quadratic, and cubic kernel functions can be written as follows:

$$k(X_s, X_s') = \exp\left(-\frac{\|X_s - X_s'\|_2^2}{2\sigma}\right), \qquad (22)$$

$$k(X_s, X_s') = 1 - \frac{\|X_s - X_s'\|_2^2}{\|X_s - X_s'\|_2^2 + C}, \qquad (23)$$

$$k(X_s, X_s') = (X_s^T X_s' + 1)^3, \qquad (24)$$

where X_s and X_s' are two arbitrary samples and σ is the kernel width. Fig.8 illustrates the two-dimensional condition
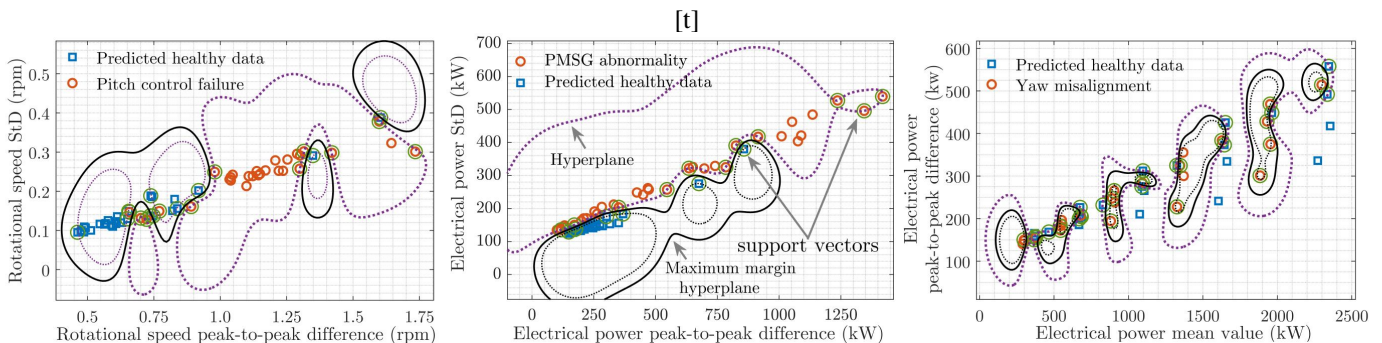
[t]

Figure 8: Employing the SVM classification on lowest-ranked features.

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 9

indicators and the Gaussian kernel SVM classification performance for the lowest-ranked features, which have complex distributions. Although the proposed classifier is robust enough and still able to give appropriate boundaries, it results in local performance. Therefore, the first ten high-ranked features are selected to achieve global performance and reduce computational complexity. The hyperparameters are tuned to find a hyperplane that separates the data perfectly into faulty and healthy classes and reduces misclassification errors.

D. Enhancing Classification using surrogate model

Algorithm 1 Anomaly detection using surrogate model

Require: Measured data X, anomaly librery U
1: Create surrogate model M
2: Discriminator: Calculate residuals $\hat{R} \leftarrow (X - M(X))$
3: Extract features $F : f_1, ..., f_n$ from residuals $\hat{R}$
4: Apply one-way ANOVA to select effective features F
5: Create condition indicators I incorporating features F'
6: Train SVM_{int} using small set of labeled anomalies A for faulty scenarios from anomaly library U
7: Applying different kernel functions
8: Optimize SVM hyperparameters for the best kernel
9: **while** not converged **do**
10:     Calculate anomaly scores S for all data points in X
11:     Sort the data points in X based on their corresponding scores in S in **descending** order: SortIndices $\leftarrow$ argsort(S)
12:     Select the r highest-scored anomalies: $H \leftarrow \{X[\text{SortIndices}[i]]\}_{i=1}^{r}$
13:     Augment labeled dataset A with H to obtain A' : $A' = A \cup H$
14:     Update classifier SVM_{up} using A'
15: **end while**
16: Identify unknown anomalies: Sort the data points in X based on their corresponding scores in S in **ascending** order: SortIndices $\leftarrow$ argsort(S)
17: Select the r' lowes-scored anomalies: $L \leftarrow \{X[\text{SortIndices}[i]]\}_{i=1}^{r'}$
18: Add L to the anomaly library $U' : U \cup L$
19: Update feature space: perform feature extraction and dimension reduction on U to obtain reduced feature set F'
20: Update SVM_{int} hyperparameters
21: Repeat steps 8 to 15
22: **return** SVM_{up}

The anomaly detection given in algorithm 1 aims to detect anomalies in measured data by utilizing the hybrid model that mimics the system behavior. After creating and training the surrogate model using the offline data, the residuals between the measured data and the hybrid model's predictions will be calculated. Then, features are then extracted from the residuals, and practical features are selected using one-way ANOVA. Condition indicators incorporating these features in the time and frequency domain are created, and an initial classifier, SVM_{int}, is trained using a small set of labeled anomalies under different operating conditions. The algorithm

iteratively updates the classifier by selecting and adding the highest-scored anomalies to the labeled dataset, followed by retraining the classifier. The anomaly scoring relies on a distance-based scoring method in which the abnormalities are characterized by being significantly distant from the majority of the data points, considering the Euclidean distance metric. Then, unknown anomalies, identified as the lowest-scored data points, are added to the anomaly library and undergo feature extraction and dimension reduction. The SVM_{int} hyperparameters are updated based on the reduced feature set. The iteration continues until convergence is achieved. The final output is an updated self-learned classifier, SVM_{up}, capable of accurately detecting anomalies in the system. In general, the algorithm aims to improve the strength and effectiveness of the classifier over time by experiencing more abnormalities. The iterative nature of the algorithm allows the classifier to learn from new anomalies selected and added to the labeled dataset in each iteration. By continuously updating the classifier with increasing irregularities, it becomes more robust and adaptive to different types of abnormalities present in the system. This iterative learning process helps enhance the classifier's ability to detect anomalies and improve its overall performance over time.

IV. SIMULATION RESULTS

A. Prediction accuracy

In order to train the deep learning models, the time-consuming, computationally expensive simulations are carried out offline to generate a training dataset for numerous ranges of mean wind speed and turbulence intensities. The WT behavior is monitored for 600s in each simulation with a sampling rate of 100s. The steady-state operation of the WT is used for the training data. As a result, 894 training datasets are obtained, each having a length of 55000 samples. In order to provide a more efficient dataset for model training, each data sequence is truncated to new sequences, each with a length of 1000 samples, resulting in an expanded training data set with 49170 sequences. Finally, the ANFIS, LSTM and TCN models are trained using 100 epochs. In order to evaluate the performance of the models, we use another 100 original-length data as test data to measure the performance of the predictive model. The prediction results for the baseline black box model vs. the proposed hybrid physics-based deep learning model are given in Table I, comparing the employed deep learning methods, i.e., ANFIS Grid Partitioning (GP), Subtractive Clustering (SC), Fuzzy C-Means (FCM), optimized FCM using Genetic Algorithm (GA), LSTM and TCN. The Root Mean Square Error (RMSE) of both observable predictions, i.e., rotational speed and electrical power, are quantified as follows:

$$\text{RMSE} = \frac{1}{55000} \sum_{t=1}^{55000} \left(\sqrt{\sum_{i=1}^{100} \left(y_{\text{predicted}} - y_{\text{target}}\right)^2} \right). \tag{25}$$

The prediction accuracy of the proposed hybrid framework is validated while separately incorporates three distinct architectures, i.e., ANFIS, LSTM, and TCN. While, the ANFIS model excels in handling complex, nonlinear relationships

Table I: Predicting accuracy of the baseline black box Model vs. proposed hybrid physics-based deep learning model.

Model	Prediction Error of ω_r (rpm)		Prediction Error of P_e (kW)		Maximum Inference Time
	Baseline	Hybrid	Baseline	Hybrid	
ANFIS-GP [29], [51]	0.40	0.1196	139.11	41.7667	0.53 sec
ANFIS-SC [29], [51]	0.38	0.1263	137.47	42.6894	0.14 sec
ANFIS-FCM [29], [51]	0.38	0.1321	134.11	45.3367	0.14 sec
ANFIS-FCM/GA [29], [51]	0.31	0.0891	97.15	26.1521	0.41 sec
LSTM [23], [52], [53]	0.21	0.0704	74.82	16.8951	0.78 sec
TCN [14], [32]	0.60	0.1528	176.21	103.63	0.25 sec

and adapting to dynamic data patterns, LSTM addresses the vanishing gradient problem and effectively captures temporal dependencies of the main observables' time series. TCN, designed for sequence modeling, efficiently processes the WT sequential data, focusing on capturing long-range dependencies between estimated wind speed, pitch, and torque control sequences. Although these architectures differ in handling data characteristics, parameter connections, and computational requirements, the hybrid approach leverages the strengths of each model to enhance predictive accuracy for anomaly detection across various conditions and operational modes.

The studied ANFIS-based, LSTM, and TCN architectures demonstrate significantly reduced RMSE values for rotor speed and electrical power prediction within the hybrid framework, which integrates data-driven and physics-based information. This signifies a substantial enhancement in the proposed prediction strategy compared to the baseline approach. However, it is worth mentioning that the LSTM outperforms ANFIS and TCN in terms of accuracy despite a longer inference time. On the other hand, introducing genetic algorithms (GA) to optimize FCM parameters significantly improves the accuracy and sensitivity of the ANFIS-GP/SC model to the stochasticity of wind speed while reducing execution time. In the case of TCN, its performance excels when trained for extended epochs, and it can rival the LSTM when the number of epochs is increased to around 300. However, the number of training epochs has been maintained at 100 for all models to mitigate over-fitting risks and ensure a fair comparison.

The improved accuracy across all deep learning approaches within the hybrid framework underscores the practicality and versatility of the proposed method. Nevertheless, selecting among these deep learning surrogate models should be influenced by the size of the training dataset and a trade-off between the required inference time and prediction accuracy. These findings highlight the superiority of the hybrid physics-based modeling approach, which adeptly captures the intricate dynamics of WT systems, resulting in more precise predictions of rotor speed and power output compared to the baseline black box model, with a substantial 67.97% reduction in average RMSE.

The following subsection discusses the application of the proposed hybrid model in the anomaly detection framework for health monitoring purposes.

B. Anomaly detection performance

In this section, the healthy prediction of the ANFIS model is fed into the weakly supervised health monitoring method for

Binary Decision Tree

True class \ Predicted	HO	PA	PF	YM
HO	307 >99%	0	1 <1%	0
PA	0	112 73%	14 9%	28 18%
PF	1 1%	13 7%	161 91%	1 1%
YM	1 1%	16 10%	1 1%	136 88%

SVM (Gaussian kernel)

True class \ Predicted	HO	PA	PF	YM
HO	304 >99%	2 1%	2 1%	0
PA	0	145 94%	3 2%	6 4%
PF	0	3 2%	172 98%	1 1%
YM	0	2 1%	0	152 99%

HO: Healthy Operation   PF: Pitch Failure
PA: PMSG Abnormality   YM: Yaw Misalignment

Figure 9: Performance of the classifiers for the first type data: the baseline approach with the BDT and the proposed approach with the self-learned SVM.

two types of operating conditions. The first type represents the healthy and faulty operation of the WT in partial and full-load regions without considering the transition zone. The healthy predicted data needs to be differentiated from all the single anomalies or the combination of PMSG abnormalities with either pitch failure or yaw misalignment. The confusion matrices are shown in Fig.9, using the health codes, i.e., Healthy Operation including the curtailment (HO), PMSG Abnormality (PA), Pitch Failure (PF), and Yaw Misalignment (YM). These indicate the best performance of the updated SVM created by the proposed self-learning classification strategy, compared with the conventional classification approach that suggests a Binary Decision Tree (BDT) to classify anomalies without considering the updated learning approach. Even though the BDT presents rather satisfactory accuracy among all applied classifiers (BDT and SVMs with different kernel functions discussed in TableII), the results show employing the proposed approach gives an updated SVM with Gaussian kernel function that significantly improves anomaly detection performance. The second type of data used for evaluating the proposed algorithm includes WT operation in the transition zone, where all the anomalies are likely to occur while the control system performance degrades due to the frequent switching between pitch and torque control mechanisms, supporting the FCR provision of the deloaded WT. The performance of the SVM with the best kernel functions, including execution time, minimum prediction speed, total misclassification cost, and minimum accuracy, are presented in Table II. Moreover, the Receiver Operating Characteristic (ROC) curves and the area under the

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

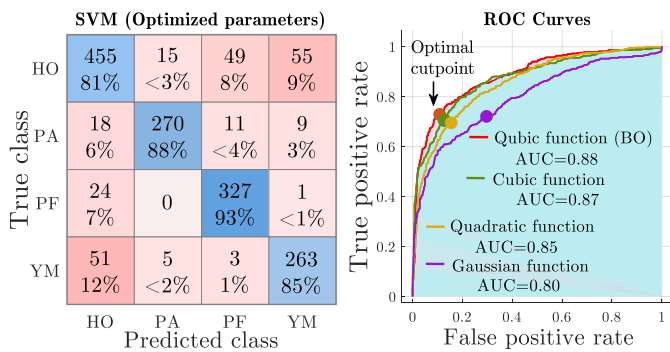IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 11

Figure 10: Anomaly detection for the second type data: the proposed approach with the self-learned SVM and optimized parameters.
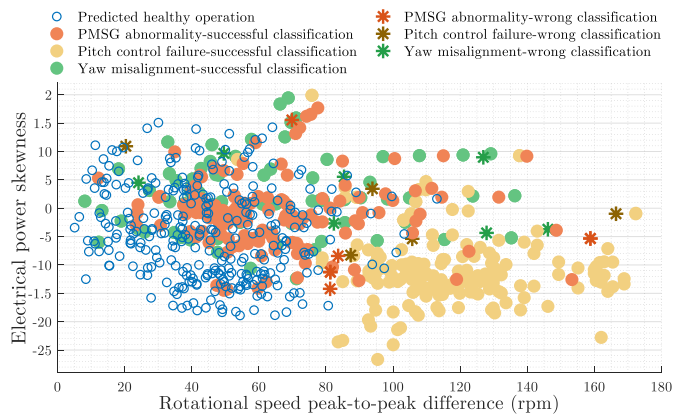

Figure 11: Anomaly detection for the first data type applying the proposed approach.

ROC curve (AUC) for different searched kernel functions are illustrated in Fig.10, indicating that the cubic kernel function has the best performance. Applying the Bayesian Optimization (BO) algorithm, with a wide range of searches between 0.001 and 1000 for kernel scale and box constraint, can improve the search efficiency and increase the execution time, which may be less practical from a computational point of view. Fig.11 demonstrates successful classifications and highlights incorrect classifications within the scatter plot of the two-dimensional updated feature space. This showcases the robust performance of the proposed anomaly detection method, even when faced with challenges such as sparsity and lack of linearity in the data points.

The presence of successful classifications in the scatter plot validates the effectiveness of the proposed approach in accurately identifying healthy behaviors. The algorithm identifies and correctly classifies instances that exhibit patterns and characteristics indicative of normal behavior. This demonstrates the ability of the method to capture and understand complex relationships within the data despite the inherent challenges posed by sparsity and nonlinearity.

Additionally, incorrect classifications in the scatter plot highlight the method's ability to detect anomalies that deviate from the expected patterns. These incorrect classifications represent instances where the algorithm identifies data points

as anomalous, even though they may appear similar to healthy data points that represent normal degradations due to TI and power curtailment. This demonstrates the algorithm's sensitivity to subtle variations and its capability to identify anomalies that might not be apparent through conventional methods. As

Table II: The SVM kernel tricks for the second type data

Kernel function	Execution time (sec)	Min prediction speed (obs/sec)	Total cost of misclassification	Minimum accuracy
Linear	6.061	16000	542	65.2%
Gaussian	3.279	11000	501	67.8%
Quadratic	19.235	12000	396	74.6%
Cubic	29.96	12000	332	78.7%
Cubic-BO	33.86	12000	293	81.2%

the confusion matrix illustrates in Fig.10, a global optimization result is achieved by involving the kernel tricks for the updated SVM. The pitch failure detection shows outperformance, while the yaw misalignment detection has the lowest accuracy. The signal shapes of electrical power and rotational speed in yaw misalignment, unlike the PMSG abnormality and the pitch failure, do not deviate significantly. In the yaw misalignment scenario, the anomaly appears in the form of degradation and lower electrical power efficiency. Also, this kind of abnormality may occur because of aerodynamic degradation due to a high level of turbulence intensity or frequent transients from torque to pitch control action. Also, it can be challenging for the algorithm to distinguish the curtailment and the aerodynamic degradation from yaw misalignment in the transition zone. Nevertheless, by comparing the results shown in Fig.9 and 10, the proposed approach gives a better result for both data types. Although the second dataset type appears to be more challenging due to the inclusion of data from the transition zone, the proposed self-learning classification performance is still satisfactory compared to the baseline approach applied to the first dataset type without considering operations in the transition zone. This observation indicates the overall improvement of the proposed anomaly detection in the presence of different sparsity levels in the dataset and different degradation scenarios.

C. Computational efficiency trade-offs

By decreasing the time length of the sliding window over the observable signals and recording at a low sampling frequency, the detailed transient behavior of the system can be captured and predicted by the deep learning models and, therefore, automatically translated into features for a more realistic and improved classification. To comprehensively analyze the temporal dynamics of the system's behavior, it is essential to employ multiple time windows spanning from seconds to minutes. Shorter time windows prove more effective for identifying rapid fluctuations or anomalies, such as those associated with PMSG abnormalities or degradations due to turbulent wind conditions, which can occur within a short time frame. In contrast, longer time windows offer a broader perspective, facilitating the detection of gradual performance decline or persistent issues like pitch or control failures.

Nevertheless, reducing the window's duration comes with a trade-off. It intensifies the computational workload and may hinder the optimization process when searching for the best SVM kernel hyperparameters. Thus, when applying the proposed approach, it is crucial to carefully consider the choice of the time window, taking into account the specific types of anomalies and a trade-off between accuracy and computational feasibility.

This study investigated various time lengths of sliding windows for the proposed anomaly detection method, and the most favorable results emerged within the 10-50-second window size. The first data type exhibited the best performance with a 50-second time window. Remarkably, even when we reduced the time window to just 10 seconds for the second data type, we observed an improvement in classification performance. Therefore, since the proposed approach compromises fast and efficient computation, the ANFIS-FCM and SVM-Gaussian kernel function with the lowest inference and execution time can be considered an initial setting for real-time health monitoring approaches. On the other hand, the LSTM or TCN and SVM-BO are better alternatives for a reasonably large window of time, i.e., more than 100s. These setting options can provide a fully automated fault correction algorithm with sufficient time to decide different logic and arrangements depending on various operational conditions, consequently leading to a more dedicated health monitoring system.

D. Limitations and prospects

In future research, noteworthy challenges deserve closer attention. One significant challenge is data availability and quality for predicting healthy behavior, which needs a more thorough examination. This includes finding solutions for dealing with limited data, creating effective methods for marking unusual events during the initial stages of model training, and managing the computational demands of complex models, especially when working with larger wind farms. Moreover, the aging effects, another form of degradation, should be considered in predicting healthy operation. This means the hybrid model should be adjusted, knowing the aging factors, or updated using the most recent datasets. Addressing these challenges is fundamental for making progress in WT health monitoring.

Additionally, applying the proposed methodology to different types of WTs and various environmental conditions can benefit transfer learning techniques, allowing the model to adapt to different situations. By understanding how to transfer and adjust knowledge across different types of turbines and environmental settings, it is possible to fully realize the potential of the proposed approach and make it useful in a broader range of real-world applications.

V. CONCLUSION

In conclusion, this study introduces a hybrid physics-based deep learning modeling approach that advances the field of WT health monitoring and anomaly detection, particularly in the context of providing Frequency Containment Reserve (FCR). The contributions of this study are three-fold: First,

a hybrid framework is presented that accurately predicts WT health by capturing the intricate interplay between stochastic wind speed fluctuations and complex correlations between control sequences (pitch and generator torque) and system responses. The proposed hybrid structure's practicality in predicting two main observables, i.e., WT electrical power and rotational speed, shows improvements compared to the baseline black box approach. This modeling approach enhances anomaly detection by effectively distinguishing normal and abnormal states. Second, this research introduces a self-learning approach with an iterative framework, demonstrating notable improvements in classifier performance. For employing Support Vector Machines (SVM) classification, coherent features are extracted from crucial observables in both time and frequency domains, enhancing the accuracy of condition indicators. Third, a comprehensive range of anomaly and degradation scenarios are considered, including those resulting from curtailment operations for FCR provision, blade pitch control failures, yaw misalignment, and Permanent Magnet Synchronous Generator (PMSG) abnormalities. The results demonstrate that the proposed health monitoring approach has improved performance and can detect anomalies that may be falsely classified as healthy but still possess some level of degradation due to turbulent intensities or deloading operations for FCR provision.

This work generally contributes to advancing wind energy system monitoring and predictive maintenance strategies by comprehensively evaluating WT health and performance. It considers intricate operational conditions and the interdependencies among control sequences, enhancing the interpretability of anomaly detection and management of wind energy conversion systems. However, challenges such as data constraints, labeling anomalies for initial training, and model complexity that can be computationally intensive for larger wind farms should be further investigated in future studies.

REFERENCES

[1] F. Bilendo, H. Badihi, N. Lu, P. Cambron, and B. Jiang, "A normal behavior model based on power curve and stacked regressions for condition monitoring of wind turbines," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–13, 2022.

[2] M. Wang, C. Wang, A. Hnydiuk-Stefan, S. Feng, I. Atilla, and Z. Li, "Recent progress on reliability analysis of offshore wind turbine support structures considering digital twin solutions," *Ocean Engineering*, vol. 232, p. 109168, 2021.

[3] Z. Wang, X. Jin, and Z. Xu, "An adaptive condition monitoring method of wind turbines based on multivariate state estimation technique and continual learning," *IEEE Transactions on Instrumentation and Measurement*, 2023.

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT

13

[4] X. Yu, B. Tang, and K. Zhang, "Fault diagnosis of wind turbine gearbox using a novel method of fast deep graph convolutional networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[5] L. Yang and Z. Zhang, "Wind turbine gearbox failure detection based on scada data: A deep learning-based approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2020.

[6] G. Jiang, H. He, P. Xie, and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017.

[7] L. Lu, Y. He, Y. Ruan, and W. Yuan, "Wind turbine planetary gearbox condition monitoring method based on wireless sensor and deep learning approach," *IEEE transactions on instrumentation and measurement*, vol. 70, pp. 1–16, 2020.

[8] L. Wang, Y. He, K. Shao, Z. Xing, and Y. Zhou, "An unsupervised approach to wind turbine blade icing detection based on beta variational graph attention autoencoder," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[9] D. Zhang, W. Tian, X. Cheng, F. Shi, H. Qiu, X. Liu, and S. Chen, "Fedbip: A federated learning based model for wind turbine blade icing prediction," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[10] Z.-X. Hu, Y. Wang, M.-F. Ge, and J. Liu, "Data-driven fault diagnosis method based on compressed sensing and improved multiscale network," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 4, pp. 3216–3225, 2019.

[11] T. Xiahou, Z. Zeng, and Y. Liu, "Remaining Useful Life Prediction by Fusing Expert Knowledge and Condition Monitoring Information," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2653–2663, 2020.

[12] H. Zhao, H. Liu, Y. Jin, X. Dang, and W. Deng, "Feature extraction for data-driven remaining useful life prediction of rolling bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–10, 2021.

[13] Z. Kong, X. Jin, Z. Xu, and B. Zhang, "Spatio-temporal fusion attention: a novel approach for remaining useful life prediction based on graph neural network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[14] K. He, Z. Su, X. Tian, H. Yu, and M. Luo, "Rul prediction of wind turbine gearbox bearings based on self-calibration temporal convolutional network," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[15] F. Cheng, J. Wang, L. Qu, and W. Qiao, "Rotor-current-based fault diagnosis for dfig wind turbine drivetrain gearboxes using frequency analysis and a deep classifier," *IEEE transactions on industry applications*, vol. 54, no. 2, pp. 1062–1071, 2017.

[16] J. Peng, A. Kimmig, Z. Niu, J. Wang, X. Liu, D. Wang, and J. Ovtcharova, "Wind turbine failure prediction and health assessment based on adaptive maximum mean discrepancy," *International Journal of Electrical Power & Energy Systems*, vol. 134, p. 107391, 2022.

[17] Q. Lyu, J. Liu, Y. He, X. Wang, and S. Wu, "Condition monitoring of wind turbines with implementation of interactive spatio temporal deep learning networks," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[18] Z. Su, X. Zhang, Y. Han, S. Wang, and M. Luo, "Adaptive gated attention network with weighted metric enhancement for fault diagnosis of wind turbine gearbox," *IEEE Transactions on Instrumentation and Measurement*, 2023.

[19] H. Badihi, Y. Zhang, B. Jiang, P. Pillay, and S. Rakheja, "A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis," *Proceedings of the IEEE*, vol. 110, no. 6, pp. 754–806, 2022.

[20] S. Chakraborty, S. Adhikari, and R. Ganguli, "The role of surrogate models in the development of digital twins of dynamic systems," *Applied Mathematical Modelling*, vol. 90, pp. 662–681, 2021.

[21] M. S. Nielsen and V. Rohde, "A surrogate model for estimating extreme tower loads on wind turbines based on random forest proximities," *Journal of Applied Statistics*, vol. 49, no. 2, pp. 485–497, 2022.

[22] R. M. Slot, J. D. Sørensen, B. Sudret, L. Svenningsen, and M. L. Thøgersen, "Surrogate model uncertainty in wind turbine reliability assessment," *Renewable Energy*, vol. 151, pp. 1150–1162, 2020.

[23] P. Wang, Y. Li, and G. Zhang, "Probabilistic power curve estimation based on meteorological factors and density lstm," *Energy*, vol. 269, p. 126768, 2023.

[24] 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," *Renewable Energy*, vol. 172, pp. 829–840, 2021.

[25] D. Karaboga and E. Kaya, "Adaptive network based fuzzy inference system (ANFIS) training approaches: a comprehensive survey," *Artificial Intelligence Review*, vol. 52, no. 4, pp. 2263–2293, 2019.

[26] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, M. E. Orchard, and M. Saif, "Critical wind turbine components prognostics: A comprehensive review," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9306–9328, 2020.

[27] X. Hu, Y. Li, L. Jia, and M. Qiu, "A Novel Two-stage Unsupervised Fault Recognition Framework Combining Feature Extraction and Fuzzy Clustering for Collaborative AIoT," *IEEE Transactions on Industrial Informatics*, 2021.

[28] A. B. Asghar and X. Liu, "Adaptive neuro-fuzzy algorithm to estimate effective wind speed and optimal rotor speed for variable-speed wind turbine," *Neurocomputing*, vol. 272, pp. 495–504, 2018.

[29] B. Bilal, K. H. Adjallah, A. Sava, K. Yetilmezsoy, and E. Kıyan, "Wind power conversion system model identification using adaptive neuro-fuzzy inference systems: A case study," *Energy*, vol. 239, p. 122089, 2022.

[30] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[31] H. Sun, M. Xia, Y. Hu, S. Lu, Y. Liu, and Q. Wang, "A new sorting feature-based temporal convolutional network for remaining useful life prediction of rotating machinery," *Computers and Electrical Engineering*, vol. 95, p. 107413, 2021.

[32] Z. Gan, C. Li, J. Zhou, and G. Tang, "Temporal convolutional networks interval prediction model for wind speed forecasting," *Electric Power Systems Research*, vol. 191, p. 106865, 2021.

[33] E. G. S. Nascimento, T. A. de Melo, and D. M. Moreira, "A transformer-based deep neural network with wavelet transform for forecasting wind speed and wind energy," *Energy*, vol. 278, p. 127678, 2023.

[34] Y. Wang, H. Xu, M. Song, F. Zhang, Y. Li, S. Zhou, and L. Zhang, "A convolutional transformer-based truncated gaussian density network with data denoising for wind speed forecasting," *Applied Energy*, vol. 333, p. 120601, 2023.

[35] J. Zhang, X. Li, J. Tian, H. Luo, and S. Yin, "An integrated multi-head dual sparse self-attention network for remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 233, p. 109096, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095183202300011X

[36] L. Liu, Z. Qu, Z. Chen, F. Tu, Y. Ding, and Y. Xie, "Dynamic sparse attention for scalable transformer acceleration," *IEEE Transactions on Computers*, vol. 71, no. 12, pp. 3165–3178, 2022.

[37] X. Li and W. Zhang, "Physics-informed deep learning model in wind turbine response prediction," *Renewable Energy*, vol. 185, pp. 932–944, 2022.

[38] M. Benbouzid, T. Berghout, N. Sarma, S. Djurović, Y. Wu, and X. Ma, "Intelligent condition monitoring of wind power systems: State of the art review," *Energies*, vol. 14, no. 18, p. 5967, 2021.

[39] K. Zhang, B. Tang, L. Deng, and X. Yu, "Fault detection of wind turbines by subspace reconstruction-based robust kernel principal component analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.

[40] F. K. Moghadam, V. Chabaud, Z. Gao, and S. Chapaloglou, "Power train degradation modelling for multi-objective active power control of wind farms," *Forschung im Ingenieurwesen*, vol. 87, no. 1, pp. 13–30, 2023.

[41] A. Otter, J. Murphy, V. Pakrashi, A. Robertson, and C. Desmond, "A review of modelling techniques for floating offshore wind turbines," *Wind Energy*, vol. 25, no. 5, pp. 831–857, 2022.

[42] A. Castellani, S. Schmitt, and S. Squartini, "Real-World Anomaly Detection by Using Digital Twin Systems and Weakly Supervised Learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4733–4742, 2020.

[43] M. Zheng, J. Man, D. Wang, Y. Chen, Q. Li, and Y. Liu, "Semi-supervised multivariate time series anomaly detection for wind turbines using generator scada data," *Reliability Engineering & System Safety*, vol. 235, p. 109235, 2023.

[44] B. J. Jonkman, "TurbSim user's guide: Version 1.50," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2009.

[45] B. Jonkman and L. Kilcher, "Turbsim user's guide: version 1.06. 00," *National Renewable Energy Laboratory: Golden, CO, USA*, 2012.

[46] J. Jonkman, S. Butterfield, W. Musial, and G. Scott, "Definition of a 5-MW reference wind turbine for offshore system development," National Renewable Energy Lab (NREL), Golden, CO (United States), Tech. Rep., 2009.

THIS IS THE ACCEPTED MANUSCRIPT. The layout may differ from the published version.

IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT 14

[47] J. D. M. De Kooning, T. L. Vandoorn, J. Van de Vyver, B. Meersman, and L. Vandevelde, "Displacement of the maximum power point caused by losses in wind turbine systems," *Renewable Energy*, vol. 85, pp. 273–280, 2016.

[48] N. Kayedpour, J. D. De Kooning, A. E. Samani, L. Vandevelde, and G. Crevecoeur, "An adaptive operational strategy for enhanced provision of frequency containment reserve by wind turbines: Data-driven based power reserve adjustment," *Electric Power Systems Research*, vol. 223, p. 109564, 2023.

[49] X. Ge, X. Zhu, Y. Fu, Y. Xu, and L. Huang, "Optimization of reserve with different time scales for wind-thermal power optimal scheduling considering dynamic deloading of wind turbines," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 4, pp. 2041–2050, 2022.

[50] Y. Wang, X. Duan, R. Zou, F. Zhang, Y. Li, and Q. Hu, "A novel data-driven deep learning approach for wind turbine power curve modeling," *Energy*, vol. 270, p. 126908, 2023.

[51] B. Bilal, K. H. Adjallah, A. Sava, K. Yetilmezsoy, and M. Ouassaid, "Wind turbine output power prediction and optimization based on a novel adaptive neuro-fuzzy inference system with the moving window," *Energy*, vol. 263, p. 126159, 2023. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S0360544222030456

[52] D.-E. Choe, H.-C. Kim, and M.-H. Kim, "Sequence-based modeling of deep learning with lstm and gru networks for structural damage detection of floating offshore wind turbine blades," *Renewable Energy*, vol. 174, pp. 218–235, 2021.

[53] N. Zhang, S.-L. Shen, A. Zhou, and Y.-F. Jin, "Application of LSTM approach for modelling stress–strain behaviour of soil," *Applied Soft Computing*, vol. 100, p. 106959, 2021.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[55] T. V. Hecke, "Power study of anova versus Kruskal-Wallis test," *Journal of Statistics and Management Systems*, vol. 15, no. 2-3, pp. 241–247, 2012.

Nezmin Kayedpour received her Ph.D. in Electromechanical Engineering in 2023 at Ghent University, Ghent, Belgium. She holds a post-doctoral research position in the Electrical Energy Laboratory within the Department of Electromechanical, Systems, and Metal Engineering at Ghent University. Her research focuses on renewable energy advancements by integrating hybrid modeling, data-driven control strategies and multi-criteria decision-making processes for optimizing smart grids and flexibility provision. She is also a member of the Flanders-Make@UGent MIRO core lab.

Jixiang Qing received the M.S. degree in aircraft design engineering from Northwestern Polytechnical University, China, in 2019, and the Ph.D. degree in engineering from the IDLab, Ghent University, Ghent, Belgium, in 2023. His current research interests include Bayesian Optimization, Machine Learning, and their applications.
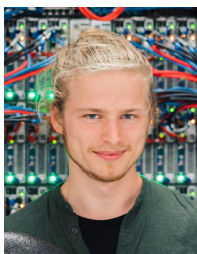

Jolan Wauters received his master's and Ph.D. degree in electromechanical engineering from Ghent University, Ghent, Belgium in 2015 and 2020, respectively. He was the recipient of a personal postdoctoral research grant from Research Foundation – Flanders (FWO) in 2022. His research focuses on advancing model-based system design by incorporating high-fidelity modelling, uncertainty propagation, and control strategies techniques in the development cycle of dynamical systems. This is realized by means of innovative surrogate-assisted multidisciplinary optimization routines. Validation of the developed methodologies is performed on mechatronic applications coupled to computational fluid mechanics.

Jeroen D. M. De Kooning (M'09, SM'20) was born in Kapellen, Belgium, in 1987. He received the Bachelor, Master and Ph.D. degrees in electromechanical engineering from Ghent University, Belgium, in 2008, 2010 and 2015 respectively. Since 2019, he is assistant professor at the Department of Electromechanical, Systems and Metal Engineering at the Faculty of Engineering and Architecture of Ghent University, Belgium. In 2022, he was a visiting professor at the Lappeenranta University of Technology, Finland. He is a member of the Dynamical Systems and Control (DySC) research group and the Flanders-Make@UGent MIRO core lab. He conducted research on current waveform shaping techniques for permanent magnet synchronous machines, and optimal control and design of renewable energy systems. His present research interests include modelling, optimization and control of mechatronic systems, drivetrains and manufacturing machines in an Industry 4.0 context, with a particular interest in digital twins.
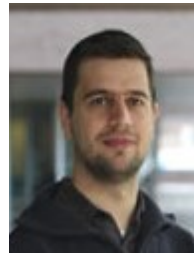
Ivo Couckuyt (Member, IEEE) received the M.S. degree in computer science from the University of Antwerp (UA), Belgium, in 2007, and the Ph.D. degree in engineering from the IDLab, Ghent University, Ghent, Belgium, in 2013. In October 2007, he was with the research group Computer Modelling and Simulation (COMS). Since 2021, he has been an Associate Professor with the IDLab Research Group, Ghent University, working on automation in machine learning, data analytics, data-efficient machine learning, and surrogate modling

Guillaume Crevecoeur Guillaume Crevecoeur (°1981) received his Master and PhD degree in Engineering Physics from Ghent University in 2004 and 2009, respectively. He received a Research Foundation Flanders postdoctoral fellowship in 2009 and was appointed Associate Professor at Ghent University in 2014. He is member of Flanders Make in which he leads the Ghent University activities on machines, intelligence, robotics and control. With his team, he conducts research on system identification and nonlinear control for mechatronics, industrial robotics and energy systems. His goal is to endow physical dynamic systems with improved functionalities and capabilities when interacting with uncertain environments, other systems and humans.