

## Article

# Early Detection and Diagnosis of Wind Turbine Abnormal Conditions Using an Interpretable Supervised Variational Autoencoder Model

Adaiton Oliveira-Filho <sup>1,\*</sup> , Ryad Zemouri <sup>2</sup> , Philippe Cambron <sup>3</sup> and Antoine Tahan <sup>1</sup> 

<sup>1</sup> Department of Mechanical Engineering, Ecole de Technologie Supérieure, 1100 Rue Notre Dame O, Montreal, QC H3C 1K3, Canada; antoine.tahan@etsmtl.ca

<sup>2</sup> Institut de Recherche D'Hydro-Québec, 1800 Bd Lionel-Boulet, Varennes, QC J3X 1S1, Canada; zemouri.mohamedryad@hydroquebec.com

<sup>3</sup> Power Factors, 7005 Boulevard Taschereau, Brossard, QC J4Z 1A7, Canada

\* Correspondence: adaitonfilho@icloud.com

**Abstract:** The operation and maintenance of wind turbines benefit from reliable information on the wind turbine condition. Data-driven models use data from the supervisory data acquisition system. In particular, great performance is reported for artificial intelligence models. However, the lack of interpretability limits their effective industrial implementation. The present work introduces a new condition-monitoring approach for wind turbines featuring a built-in visualization tool that confers interpretability upon the model outcomes. The proposed approach is based on a supervised implementation of the variational autoencoder model, which allows the projection of the wind turbine system onto a low-dimensional representation space. Three outcomes follow from such representation: a health indicator for the early detection of abnormal conditions, a classifier providing the diagnosis status, and a visualization tool depicting the wind turbine condition as a trajectory in a 2D plot. The approach is implemented with a vast database. Two case studies demonstrate the potential of the proposed approach. The proposed health indicator detects the main bearing overtemperature 11 days before the control system alarm, one week earlier than a competing approach. Study cases illustrate that the built-in visualization tool enhances the interpretability and trust in the model outcomes, thus supporting wind turbine operation and maintenance.

**Keywords:** wind turbine; condition monitoring; variational autoencoder; SCADA data; early detection; diagnosis; model interpretability



**Citation:** Oliveira-Filho, A.; Zemouri, R.; Cambron, P.; Tahan, A. Early Detection and Diagnosis of Wind Turbine Abnormal Conditions Using an Interpretable Supervised Variational Autoencoder Model. *Energies* **2023**, *16*, 4544. <https://doi.org/10.3390/en16124544>

Academic Editors: Wei Teng and Dikang Peng

Received: 6 May 2023

Revised: 31 May 2023

Accepted: 2 June 2023

Published: 6 June 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid and assured growth of the global wind power capacity results from efforts to decarbonize energy production. This trend is supported by and depends on a competitive Levelized Cost of Energy (LCOE) for a Wind Turbine (WT). Notably, Operation and Maintenance (O&M) expenditures correspond to a significant share of the WT LCOE, with estimations ranging from 20% to 30% [1,2].

O&M in-situ interventions include inspections, preventive maintenance, and curative maintenance. Currently, scheduling of the WT O&M interventions relies on data-driven analysis to a limited extent. Typically, Performance Monitoring (PM) uses data from the WT Supervisory Control and Data Acquisition (SCADA) system to detect overall underperformance, while Condition Monitoring (CM) approaches aim at detecting and diagnosing abnormal conditions in specific critical components based on the SCADA data [3,4].

The WT CM literature pays particular attention to Artificial Intelligence (AI) models. This interest is due to the availability of ever-increasing databases from operating wind farms, not to mention the proven performance of AI-based models. Among these, the Variational Autoencoder (VAE) model (see Kingma and Welling [5]) stands out given its ability

to analyze systems characterized by a large number of features, including features with a noisy and stochastic nature, as is the case of the measures from the WT system [6–8]. Nevertheless, the effective implementation of AI-based models in WT CM requires scientific and technological gaps to be addressed, notably the lack of interpretability. Visualization tools are among the main approaches to overcoming the black-box nature of AI models [9,10]. The VAE low-dimensional latent space is used to develop visualization tools allowing one to interpret the model outcomes. Proven results are reported in applications emerging from diverse domains of study [8,11–14]. To the best of our knowledge, the use of the VAE latent space as a visualization tool is not exploited in any publication on WT CM.

The present work aims to exploit the VAE model in the definition of a CM approach for WT with a built-in visualization tool for enhanced interpretability. Precisely, the proposed approach is based on a supervised implementation of the VAE. The dimension reduction capability of the VAE model allows for the definition of a unified approach for (1) the early detection of abnormal conditions, (2) the diagnosis of the abnormal conditions, and (3) the definition of a visualization tool in which the evolution of the wind turbine condition is represented in a 2D plot.

### 1.1. Related Works

The present work introduces a supervised implementation of the VAE model that leads to the detection and diagnosis of abnormal conditions, as well as a visualization tool. Previous works are briefly reviewed below.

#### 1.1.1. Detection of Wind Turbine Abnormal Conditions Using the VAE Model

The literature on VAE-based WT detection includes two kinds of analysis. The models can focus on sub-systems and critical components or rather consider the overall WT condition. Among the former, Zhao et al. [15] use the Autoencoder (AE) model to detect abnormal conditions and anticipate failure in the gearbox, the main shaft bearing, and the generator. Wang et al. [16] use the AE model to perform CM of the WT breakage system. A wavelet-enhanced AE model is proposed by Yuan et al. [17] to detect blade icing. Hemmer et al. [18] define a VAE-based Health Indicator (HI) for the WT main bearing condition. The authors exploit, in particular, the VAE latent space, the HI being a function of the latent variables.

CM approaches analyzing the overall WT condition are often called system-wide CM. Jiang et al. [19] use a denoising AE to deal with the non-linear and stochastic WT measures. Wu et al. [20] exploit the denoising AE similarly. In both cases, the denoising AE allows the capture of different information from the features. Meanwhile, Renström et al. [21] use the VAE model in an overall WT CM approach based uniquely on SCADA data. The authors analyze the influence of the model architecture and Gaussian denoising. In particular, they conclude that the range of choices for the VAE architecture is broad, with multiple dimensions of the VAE latent space leading to similar performance. Moreover, their work reports that no noise lead to the best performance when using the SCADA 10-min measures [21].

#### 1.1.2. Diagnosis of Wind Turbine Abnormal Conditions Using the VAE Model

Diagnosis aims at characterizing the detected abnormal behaviors in terms of severity [14], category [22], location, and root causes [23]. We refer to Stetco et al. [24] for a comprehensive review of diagnosis approaches for WT.

The VAE model successfully diagnoses defaults in rolling bearings in [14]. The unsupervised training of the VAE allows for diagnosis due to the database characteristics and specific modeling choices. First, the database emerges from a controlled laboratory setting. Second, the available data include vibratory measures, which have proven accuracy in detecting abnormal conditions in rotating components [25,26]. Finally, the authors select the key features using feature engineering specific to the conditions of interest.

Roelofs et al. [23] investigated the use of the component-wise VAE reconstruction error to diagnose the WT condition. However, the classical unsupervised VAE seems insufficient

to diagnose wind turbines when only SCADA data are available [23]. The intricate nature of the features limited the success of their proposition, with a weak causal relation between some components of the reconstruction error and the abnormal behavior under analysis.

Previous works achieve good performance in diagnosis by combining the VAE with other Neural Network (NN) models in the framework of generative adversarial networks [27,28]. Liu et al. [28] introduced a diagnosis approach for the categorization of abnormal conditions in multiple WT sub-systems. The authors proposed the sparse dictionary learning-based adversarial variational auto-encoders (AVAE\_SDL) model and used only measures from the SCADA system. Scores from the reconstruction errors specific to each SCADA measure lead to the categorization of abnormalities. The AVAE\_SDL-based diagnosis correctly categorizes selected study cases, outperforming competing approaches regarding the frequency of false alarms. More recently, Zhang et al. [27] proposed a diagnosis approach for WT bearings using vibration signals. The authors reported good accuracy for multiple case studies with typical bearing vibration data and data with added noise. The conditional variational generative adversarial network (CVAE-GAN) model allowed for addressing two major problems in other vibration-based approaches, namely the imbalance of databases and the high frequency of false alarms.

### 1.1.3. VAE Model as a Visualization Tool

The VAE latent space is exploited as a visualization tool in diverse domains, including audio and speech processing [11], manufacturing [12], X-ray diffraction [13], bearing diagnosis [14], and hydrogenerator monitoring [8]. The VAE-based visualization tools are reported to outperform competing approaches such as Principal Component Analysis and t-distributed Stochastic Neighbor Embedding Projection [8,12]. Zemouri et al. [8] use the VAE to define a 2D visualization and classification tool for partial discharge, which is a consequential symptom of degradation in hydrogenerators. Proteau et al. [12] use the VAE model to detect early changes in the state of machining processes, with the corresponding latent space serving as a visualization tool. The authors recently proposed a prognosis approach from the VAE latent space [29]. Cheng and Chen [14] propose a diagnosis approach for ball bearings based on a VAE model. They show, in particular, that selecting the most sensitive features for a specific degraded condition can enhance cluster disentanglement in the latent space.

### 1.2. Main Contributions

To our knowledge, no previous works have used the VAE model to simultaneously address the WT CM and render a visualization tool for enhanced interpretability. The present work fills this gap. The three main contributions of the present paper are summarized below:

- This work introduces a new supervised implementation of the VAE model, hereafter referred to as the VAE embedded with Classification (VAEC) model, for detecting and diagnosing the WT operating conditions. The VAEC allows the representation of the WT condition in a low-dimensional and physically representative space.
- The VAEC latent space as a visualization tool for the WT condition is introduced and illustrated with two case studies. The proposed visualization tool derives directly from the VAEC model when the latent space dimension is set as two or three. The resulting 2D or 3D visualization is interpretable and is expected to enhance trust and confidence among O&M practitioners.
- A HI is introduced based on the VAEC encoding of SCADA datasets into the VAEC latent space. The proposed HI uses the Mahalanobis distance to measure how far data points are from the healthy cluster. The Exponentially Weighted Moving Average (EWMA) control chart is then used to detect trends in the daily average of the Mahalanobis distance [30]. Tests with real data show that the proposed HI allows the detection of abnormal conditions earlier than competing approaches.

### 1.3. Paper Organization

This paper is organized as follows. Section 2 describes the SCADA database from an operating wind farm. Section 3 reviews the VAE model. Section 4 introduces the VAEC model. It is then used in an approach comprising both CM and a visualization tool. Section 5 presents the methodology. Section 6 shows encouraging results from the implementation of the proposed approach in case studies. Finally, Section 7 concludes the work.

## 2. SCADA Database

This work uses data from a North American wind farm comprising over a hundred WTs. The database covers two years and four months of operation, a period that includes occurrences of multiple abnormal behaviors, some of which characterize the degradation or failure of specific critical components. The machines are horizontal and upwind and have three-blade rotors of nearly 90 m diameter and a hub height of 80 m. Each WT has a rated power of 1.85 MW and a rated wind speed of  $V_r = 13$  m/s. The WTs are pitch-controlled with cut-in wind speed  $V_{in} = 3.5$  m/s and cut-off wind speed  $V_{out} = 20$  m/s.

Temperature measures are commonly used to detect abnormal conditions in WT components when no vibratory measure is available Beretta et al. [31], which is the case for the WTs analyzed in the present study. Five interest conditions are summarized in Table 1. “HY” is the reference representing healthy operation. Four of the selected conditions consist of overtemperature at critical components. Ice accretion on blades can happen in WTs operating under icing conditions, eventually resulting in non-winterized WTs [32], as is the case for the units under analysis.

**Table 1.** WT conditions of interest to demonstrate the proposed CM approach.

ID	Description
HY	Healthy condition
BEA	Main bearing overtemperature
GBX	Gearbox oil overtemperature
GEN	Generator stator winding overtemperature
ICE	Ice accretion on blades
SWT	Gearbox oil overtemp. from thermal switch

The database made available for the present research comprises 35 SCADA measures for each WT. The SCADA system comprises multiple sensors, including geometrical, kinematic, thermal, and electrical measures. The SCADA measures are continuously acquired at frequencies that depend on each sensor and then stored following the 10-min aggregation industrial standard, i.e., each data point is the average of measures over a period of 10 minutes [33]. This data format is suitable for performance monitoring and eases data management and processing [34].

As evidenced by Cheng and Chen [14], the appropriate selection of informative features enhances the discriminative power of autoencoding models such as the VAE. In the present study, the selection of the SCADA measures was guided by the sensitivity analysis specific to each interest condition [14] and general feature selection criteria [28,34].

Each abnormal condition has key features indicating the physical nature of the condition itself. For example, the main bearing temperature  $T_{BEA}$  (°C) is the most informative feature to analyze the BEA condition. Analogously, the gearbox oil temperature  $T_{GBX-OIL}$  (°C) is essential to analyze GBX and SWT; generator temperature  $T_{GEN}$  (°C) is essential to analyze GEN. The ICE condition can be characterized from the power curve, i.e., from the pair wind speed  $WS$  (m/s) and active power  $P$  (kW) [35].

Some of the available measures add little or no information to the characterization of the WT condition. Such a lack of informative power can be related to three reasons. First, some variables are highly correlated with already selected measures [36]. For example, the battery box temperature is measured at three positions  $T_{BAX-BOXi}$ ,  $i \in \{1, 2, 3\}$ .

The cross-correlation coefficient between each pair among these three measures is one; thus, only  $T_{BAX-BOX1}$  was retained to describe the battery box temperature. Second, the 10-min average erases any significant information from measures varying at a very high frequency, such as the electrical frequency and voltage (both neglected in the model) [37]. Thirdly, some variables are too scarce, i.e., their time series present many missing or non-numerical entries [38]. Such is the case for the nacelle yaw position angle. Finally, it is worth mentioning that no vibratory or acoustic measures were available for the wind farm under analysis.

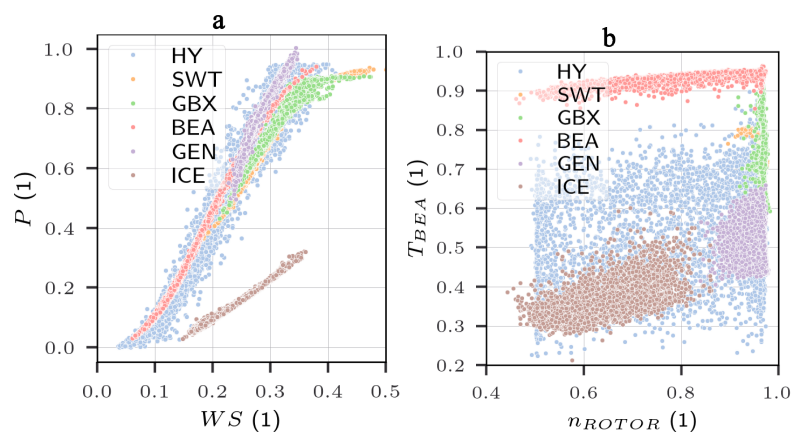
In light of the previous considerations, the present work uses the 13 key SCADA measures listed in Table 2. For each measure, the upper and lower bounds (LB and UB, respectively) were defined from the statistical analysis of the measures from all wind turbines. Such bounds have a twofold goal in pre-processing: first, filtering out physically incoherent values, i.e., values outside of the interval [LB, UB]; second, normalizing the measures into the  $[0, 1]$  interval with min-max normalization [38].

**Table 2.** SCADA measures used as features with the respective lower (LB) and upper (UB) bounds.

Measure	Symbol	Unit	LB	UB
Wind Speed	$WS$	m/s	0	31
Rotor Speed	$n_{ROTOR}$	rpm	0	18
Active Power	$P$	kW	0	2000
Ambient Temp.	$T_{AMB}$	°C	-25	45
Nacelle Temp.	$T_{NAC}$	°C	-20	70
Main Bearing Temp.	$T_{BEA}$	°C	-20	70
Gearbox Bearing Temp.	$T_{GBX-BEAR}$	°C	0	100
Gearbox Oil Temp.	$T_{GBX-OIL}$	°C	0	100
Generator Temp. Position 1	$T_{GEN1}$	°C	-10	140
Generator Temp. Position 2	$T_{GEN2}$	°C	-10	140
Generator Cooling Temp.	$T_{GEN-COOL}$	°C	-10	120
Pitch Axis Box Temp.	$T_{AX-BOX}$	°C	0	60
Battery Box Temp. Position 1	$T_{BAT-BOX1}$	°C	0	45

Besides the raw SCADA measures, the database includes SCADA log files and O&M reports. Both sources contain information on eventual underperformance root causes, degradation symptoms, and failure patterns in specific components. One can exploit such metadata to select the abnormal conditions in a given wind farm. The SCADA log files indicate exceptions on the WT operation, while the O&M reports are multi-entry forms filled out by the O&M practitioners.

Section 5.1 describes the definition and labeling of the condition-specific datasets. Figure 1 depicts the distribution of the datasets of interest in two plots.



**Figure 1.** Distribution of the datasets of interest in (a) the normalized power curve and in the (b)  $T_{BEA}$  versus  $n_{ROTOR}$  plot.

The superposition of clusters in Figure 1 illustrates the difficulties in characterizing the WT conditions based on subspaces of the high-dimensional WT physical space. Rather than analyzing such a space, the approach proposed in the present work is based on a low-dimensional representation of the WT condition.

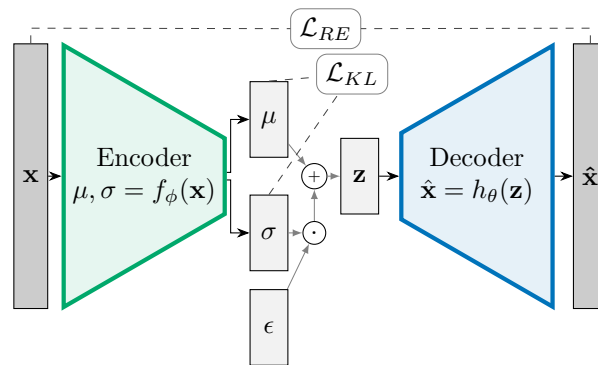
### 3. Background

A brief description of the VAE model in Section 3.1 is followed by the characterization of the VAE latent space for multiple-class databases in Section 3.2. Then, Section 3.3 reviews semi-supervised and supervised implementations of the VAE.

#### 3.1. Variational Autoencoder

Kingma and Welling [5] introduced the VAE as a generative model combining variational Bayesian and deep learning methods. We refer to [39,40] for a comprehensive presentation of the VAE. A brief description of the VAE is given below.

Let  $\mathbf{x} = [x_1, \dots, x_{n_F}]^T \in \mathbb{R}^{n_F}$  be a vector describing the physical state of an arbitrary system. The VAE is a Deep Neural Network (DNN) model that builds an approximation for  $\mathbf{x}$ , denoted  $\hat{\mathbf{x}}$ , through three transformations, namely the encoder, the reparametrization trick, and the decoder. Figure 2 illustrates the VAE architecture.



**Figure 2.** Schematic representation of the VAE model. The dimension of the input  $\mathbf{x}$  and the output  $\hat{\mathbf{x}}$  is  $n_F$  (number of features describing the system). The dimension of the latent (or code) space variable  $\mathbf{z}$  is  $n_L$ , with  $n_L < n_F$ . The loss function components  $\mathcal{L}_{RE}$  and  $\mathcal{L}_{KL}$  are highlighted.

The encoder is a DNN with parameters  $\phi$  that maps  $\mathbf{x}$  into the latent space mean  $\mu \in \mathbb{R}^{n_L}$  and standard deviation  $\sigma \in \mathbb{R}^{n_L}$ , as given by Equation (1), where  $n_L < n_F$ .

$$f_\phi : \mathbf{x} \mapsto \{\mu, \sigma\}, \mathbb{R}^{n_F} \rightarrow \mathbb{R}^{n_L} \times \mathbb{R}^{n_L} \quad (1)$$

The reparametrization trick introduces a variational Bayesian approximation to the latent space [40]. It maps the  $\mu$  and  $\sigma$  into the latent variable  $\mathbf{z} \in \mathbb{R}^{n_L}$  according to Equation (2), where  $\epsilon \sim \mathcal{N}(0, 1)$  is a  $n_L$ -dimension Gaussian vector with sample space  $\mathcal{E}$ , and  $\bullet$  is the element-wise product.

$$g : \{\mu, \sigma, \epsilon\} \mapsto \mathbf{z} = \mu + \sigma \bullet \epsilon, \mathbb{R}^{n_L} \times \mathbb{R}^{n_L} \times \mathcal{E} \rightarrow \mathbb{R}^{n_L}. \quad (2)$$

Finally, the VAE decoder is a DNN mapping the latent space variable  $\mathbf{z}$  into the VAE output  $\hat{\mathbf{x}}$ , as given by Equation (3), where  $\theta$  is the set of the decoder parameters.

$$h_\theta : \mathbf{z} \mapsto \hat{\mathbf{x}}, \mathbb{R}^{n_L} \rightarrow \mathbb{R}^{n_F}. \quad (3)$$

Training the VAE involves minimizing the loss function  $\mathcal{L}_{VAE}$  with an algorithm such as the stochastic gradient descent. The VAE loss function is given by Equation (4), where  $\mathcal{L}_{RE}$

is the reconstruction error,  $\mathcal{L}_{KL}$  is the Kullback-Liebler (KL) divergence, and the coefficient  $\beta_{kl} > 0$  is set to prevent the KL-vanishing problem [41,42].

$$\mathcal{L}_{VAE} = \mathcal{L}_{RE} + \beta_{kl} \mathcal{L}_{KL} \quad (4)$$

The reconstruction error  $\mathcal{L}_{RE}$  is given by Equation (5):

$$\mathcal{L}_{RE} = \frac{1}{n_F} \sum_{j=1}^{n_F} (x_j - \hat{x}_j)^2 \quad (5)$$

The  $\mathcal{L}_{KL}$  loss function measures the statistical distance between the latent variable distribution and the multivariate Gaussian distribution. Equation (6) gives  $\mathcal{L}_{KL}$  as a function of  $\mu$  and  $\sigma$ .

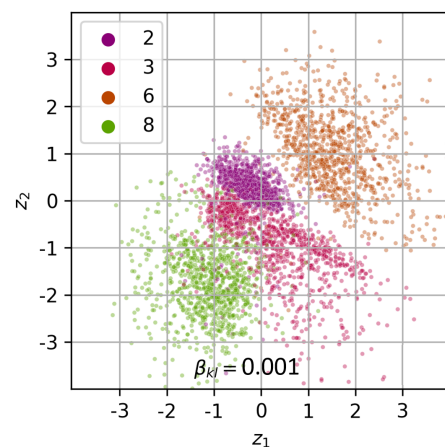
$$\mathcal{L}_{KL} = \frac{1}{2} \sum_{j=1}^{n_L} (\sigma_j^2 + \mu_j^2 - \log \sigma_j^2 - 1). \quad (6)$$

Such training forces the VAE parameters to assume values so that (i) the encoder keeps essential information from the physical space and (ii) the decoder reconstructs a good approximation for the input features from the latent space variable. Moreover, the variational approximation implies that (iii) the encoding of the training database projects points in the latent space that follow approximately a multivariate Gaussian distribution.

### 3.2. VAE Latent Space for a Multiple-Condition Database

Once the VAE is trained, the encoding of the training database projects its points into the latent space. Due to the variational approximation, datasets corresponding to a particular condition are encoded into a specific region in the latent space. The distribution of the encoded training points in the latent space depends on the characteristics of the training database and the  $\beta_{kl}$  coefficient. For instance, the encoding of a homogeneous database, i.e., whose points correspond to only one condition, results in a unique cluster of points in the latent space. A heterogeneous training database, on the contrary, projects into multiple clusters in the VAE latent space. In both cases, the set of all points together follows approximately the multivariate Gaussian distribution in the latent space. Setting a relatively large  $\beta_{kl}$  forces the latent space points to follow the Gaussian distribution.

The clusters corresponding to multiple datasets can be disjointed or entangled. Datasets sharing common patterns tend to project into superposed clusters, while very different behaviors would project into disentangled clusters. These cases are illustrated in Figure 3 for a subset of the MNIST database [43]. Notice that the clusters corresponding to the numbers 2, 3, and 8 are partially superposed.



**Figure 3.** A 2D latent space corresponding to the VAE trained on the subset {2,3,6,8} of the MNIST database with  $\beta_{kl} = 0.001$ .

Apropos, one can characterize a cluster by its distribution in the latent space. Let  $\Omega_k$  be a cluster with  $N_k$  points  $\{z_1^{\Omega_k}, \dots, z_{N_k}^{\Omega_k}\}$  in the  $n_L$ -dimension VAE latent space. The cluster's centroid  $C^{\Omega_k}$  is the average position, with coordinates  $C_\ell^{\Omega_k}, \ell \in \{1, \dots, n_L\}$ , given by Equation (7).

$$C_\ell^{\Omega_k} = \frac{1}{N_k} \sum_{i=1}^{N_k} z_{i,\ell}^{\Omega_k} \tag{7}$$

### 3.3. Semi-Supervised and Supervised Variational Autoencoder

Latent space configurations with disentangled clusters better suit the purpose of using the latent space to characterize the original system. However, the VAE-based modeling of real-world cases with heterogeneous databases often implies a latent space with entangled clusters. Multiple works address techniques to disentangle clusters in the latent space [14,44–47].

Cheng and Chen [14] use the VAE to detect and diagnose abnormal behavior in ball bearings. The authors use sensitivity analysis to guide the selection of features, which leads to the disentanglement of clusters in the latent space and to the enhanced performance of the proposed VAE-based detection approach.

Another approach to increase the disentanglement of clusters in the VAE latent space and enhance the overall VAE capabilities is to include information on the classes of subsets of the database, leading to semi-supervised and supervised implementations of the VAE. Kingma et al. [44] tackle the problem of classification when only a small share of the database is labeled. Sohn et al. [47] propose the fully supervised model Conditional VAE that allows for specific condition data generation since its decoder takes the classes as input.

More recently, Proteau et al. [12] introduced an approach based on the VAE model with two independent training steps. After training a classical VAE in the first step, an NN model is defined as the combination of VAE encoder layers (including the reparametrization trick) and a classification neural network taking the latent space as input. The final model is an ANN inheriting the encoder's architecture that allows the disentanglement of the clusters in the latent space.

## 4. Proposed Condition Monitoring Approach

The present work introduces a new supervised implementation of the VAE aiming at disentangling the multiple clusters in the latent space. This model is referred to as VAEC and is presented in Section 4.1. Ultimately, the VAEC latent space with disentangled clusters will serve the definition of both the HI and the visualization tool in Sections 4.2 and 4.3, respectively.

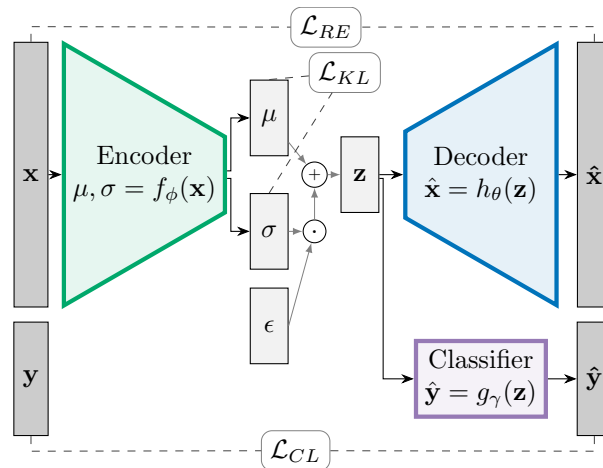
### 4.1. VAEC Model

The VAEC consists of a VAE embedded with a classification NN, as schematized in Figure 4.

In the VAEC, the classification NN takes the latent space variable  $\mathbf{z} \in \mathbb{R}^{n_L}$  as input and its output  $\hat{\mathbf{y}} \in \mathbb{R}^{n_C}$  is defined by the Softmax activation function [48]. The components  $\hat{y}_i, i \in \{1, \dots, n_C\}$ , indicate therefore the probability that the system's class is  $s_i$  among the set of classes  $\mathcal{S} = \{s_1, \dots, s_{n_C}\}$ . See Equation (8).

$$g_\gamma : \mathbf{z} \mapsto \hat{\mathbf{y}}, \mathbb{R}^{n_L} \rightarrow \left\{ \hat{\mathbf{y}} \in [0, 1]^{n_C} : \sum_{i=1}^{n_C} \hat{y}_i = 1 \right\}. \tag{8}$$





**Figure 4.** Schematic representation of the VAEC model. The input of the classifier NN is the latent space variable  $\mathbf{z}$ , and its output is the vector  $\hat{\mathbf{y}}$  with dimension  $n_C$  (number of classes). The loss function components are highlighted.

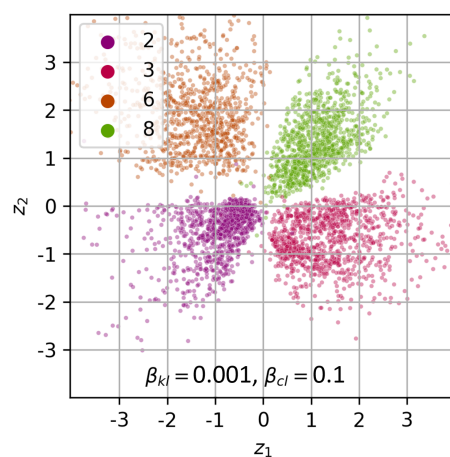
The classifier NN is identified by minimizing the classification loss function  $\mathcal{L}_{CL}$ . An usual formulation for  $\mathcal{L}_{CL}$  is the cross-entropy function given by Equation (9), where  $\mathbf{y} = \mathbf{e}_k \in \mathbb{R}^{n_C}$  results from the categorical one-hot-encoding transformation [48] corresponding to the set of labels  $\mathcal{S}$ , and  $\hat{\mathbf{y}}$  is the Softmax-shaped output of the classifier  $g_\gamma$ .

$$\mathcal{L}_{CL} = \mathbf{y} \cdot \log(\hat{\mathbf{y}}) + (\mathbf{1} - \mathbf{y}) \cdot \log(\mathbf{1} - \hat{\mathbf{y}}). \tag{9}$$

The VAEC loss function is given by Equation (10). It combines  $\mathcal{L}_{RE}$  (Equation (5)),  $\mathcal{L}_{KL}$  (Equation (6)), and  $\mathcal{L}_{CL}$  (Equation (9)). The weights  $\beta_{kl} \geq 0$  and  $\beta_{cl} \geq 0$  allow us to adjust the loss function components. Setting  $\beta_{cl} = 0$  retrieves the classical VAE.

$$\mathcal{L}_{VAEC} = \mathcal{L}_{RE} + \beta_{kl}\mathcal{L}_{KL} + \beta_{cl}\mathcal{L}_{CL} \tag{10}$$

Figure 5 depicts the VAEC latent space for the subset of the MNIST database considered previously.



**Figure 5.** The 2D latent space corresponding to the VAEC trained with the MNIST database with  $\beta_{kl} = 0.001, \beta_{cl} = 0.1$ .

It is worth noticing that the loss function coefficients  $\beta_{kl}$  and  $\beta_{cl}$  have different impacts on the latent space distribution. While the  $\beta_{kl}$  coefficient is mainly associated with the scattering of points in the latent space, the coefficient  $\beta_{cl}$  directly affects the clusters' separation. For instance, setting a relatively large  $\beta_{kl}$  value implies a latent space distribution closer to the multivariate Gaussian distribution. On the other hand, setting a large  $\beta_{cl}$  forces the

VAEC training to distinguish the different WT conditions. The proposed approach is particularly suitable for databases with labeled conditions, as is the case of WTs characterized by SCADA measures and whose operating conditions can be characterized.

4.2. Proposed Health Indicator, Detection, and Diagnosis

Once the VAEC is trained on the WT condition database, its encoder projects the training database into the latent space. For a VAEC identification chosen as a reference, one can characterize the distribution of the cluster  $\Omega_{HY} \subset \mathbb{R}^{n_L}$  corresponding to the healthy dataset  $HY$ . Let  $n_{HY} \gg 1$  be the number of points belonging to  $\Omega_{HY}$ . The vectors of the coordinates of all points  $\mathbf{z}_\ell^{HY} \in \mathbb{R}^{n_L}, \ell \in \{1, \dots, n_L\}$ , allow the estimation of the average position  $\mathbf{C}^{HY} \in \mathbb{R}^{n_L}$  and the covariance matrix  $S_{HY} \in \mathbb{R}^{n_L \times n_L}$ . The Mahalanobis distance between a point  $\mathbf{z} \in \mathbb{R}^{n_L}$  in the latent space and the  $\Omega_{HY}$  distribution is then given by Equation (11).

$$d_M(\mathbf{z}) = \sqrt{(\mathbf{z} - \mathbf{C}^{HY})^T S_{HY}^{-1} (\mathbf{z} - \mathbf{C}^{HY})} \tag{11}$$

In the latent space, a data point corresponding to a degraded condition is encoded to a position far from the  $\Omega_{HY}$  cluster, thus implying a larger distance  $d_M(\mathbf{z})$ . The HI  $I_{Dm}$  is defined as the average of the Mahalanobis distances during the period  $\mathcal{T}_k$ , as given by Equation (12), where  $n[\mathcal{T}_k]$  is the number of data points in  $\mathcal{T}_k$ .

$$I_{Dm}(\mathcal{T}_k) = \frac{1}{n[\mathcal{T}_k]} \sum_{t \in \mathcal{T}_k} d_M(\mathbf{z}(t)) \tag{12}$$

The EWMA control chart is used to detect trends in  $I_{Dm}$ . For the WT under analysis, data from operation in a healthy condition are used to estimate the reference statistics for  $I_{Dm}$ , particularly the average  $Z_M$  and the standard deviation  $s_M$ . The EWMA time series associated with the HI  $I_{Dm}$ ,  $EWMA_{Dm}$ , is then given by the initial condition  $EWMA_{Dm}(0) = Z_M$ , and the recursive relation in Equation (13) for  $k \in \{1, 2, \dots\}$ , where  $\lambda \in [0, 1]$  is the EWMA parameter that attributes weight to the previous observations of the variable under analysis.

$$EWMA_{Dm}(k) = \lambda I_{Dm}(k) + (1 - \lambda) EWMA_{Dm}(k - 1). \tag{13}$$

The Upper Control Limit (UCL) is given by Equation (14), where  $L$  is a parameter of the EWMA control chart.

$$UCL(i) = Z_M + L \cdot s_M \left( \frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]. \tag{14}$$

The base detection criterion is given by Equation (15).

$$EWMA_{Dm}(k) > UCL(k). \tag{15}$$

In the present work, each subgroup  $\mathcal{T}_k$  covers one calendar day, which corresponds to up to 144 time steps. The EWMA coefficients are set as  $\lambda = 0.1$  and  $L = 3$ . Finally, an alarm is triggered when the base detection criterion is met in three consecutive time steps.

The settings mentioned above follow practical recommendations for the EWMA control chart on Gaussian-distributed variables and analysis on multiple case studies from the WT condition database. The Mahalanobis distance and the EWMA control chart are also exploited by Renström et al. [21]. Contrary to our proposition, the previous work applied the EWMA control chart directly on the Mahalanobis distance (the corresponding distribution is approximately a  $\chi_k$ , not Gaussian) and an exceptionally small  $\lambda$  (0.007).

The VAEC classifier has a twofold purpose in the proposed approach: first, it allows adjustment of the VAEC latent space; second, it gives information on the diagnosis of the WT condition via the output  $\hat{\mathbf{y}}$ . The WT condition  $s_i \in \mathcal{S}$  is indeed given by  $\text{argmax}\{\hat{\mathbf{y}}\}$ . It is worth mentioning that the detection captures the point at which the projection of the

WT in the latent space is distanced from the healthy condition cluster  $\Omega_{HY}$ . In such a case, the projection of the WT in the latent space might move closer to any of the abnormal condition clusters under analysis. The diagnosis of the transition to a new abnormal condition follows from both the classifier output and the visualization tool.

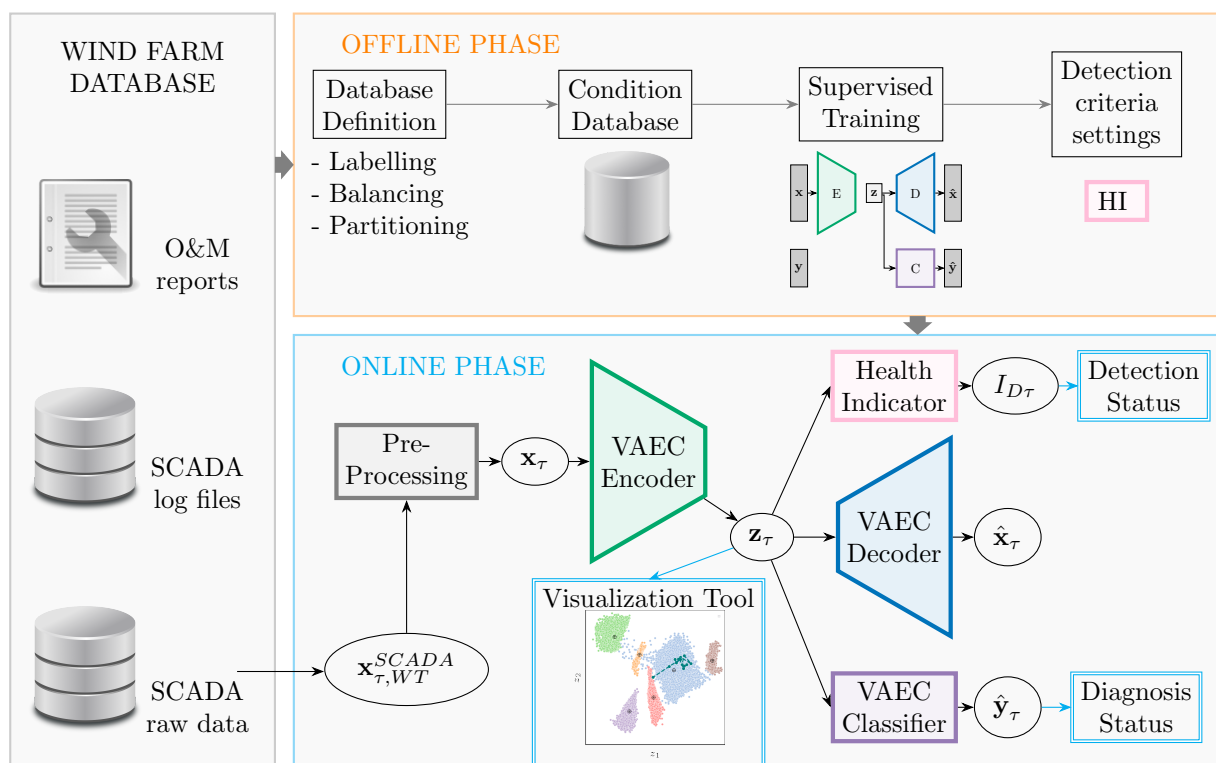
#### 4.3. Visualization Tool

Setting the VAEC latent space dimension as two or three allows us to use this low-dimensional space as a visualization tool. The proposed visualization consists of two superposed layers. The first layer consists of the clusters resulting from the VAEC encoding of the training database. It is therefore fixed once a reference VAEC is adopted. The second layer projects the successive datasets  $x_k^{SCADA}$  over the first layer. This projection is updated periodically, which allows visualization of the evolution of the WT condition in the map of conditions set by the first layer.

To improve the readability of the visualization tool for data covering larger periods, one can plot the centroids corresponding to the projection of datasets instead of individual data points. This choice can be adjusted to suit different periods of analysis and O&M requirements. A four-day time window is used in the present work. Additionally, a sequential colormap highlights the timeline in the visualization tool, with cyan dots at the beginning of the timeline and magenta dots for the latest data points. Finally, a gray dashed line connects the successive dots.

### 5. Methodology

The implementation of the proposed approach consists of two phases, namely online and offline, as schematized in Figure 6.



**Figure 6.** Flowchart identifying the phases of the proposed approach with the respective steps. The wide gray arrows symbolize steps implemented once with information from multiple machines. The black arrows indicate transformations specific to one WT and repeated periodically.

The offline phase is performed once and takes data from multiple machines of the wind farm assumed to be from the same model and to operate in similar conditions. On the other hand, the online phase is periodically fed with SCADA measures from one WT and gives an updated CM status for this specific machine. The building blocks comprising the online phase's pipeline are defined in the offline phase.

The preparation of the multiple-condition labeled database comprises filtering and normalizing the SCADA data, labeling, balancing, and partitioning. These steps are discussed in Sections 5.1 and 5.2. Section 5.3 describes the architecture, hyperparameters, and training of the VAEC.

### 5.1. Database Pre-Processing

The pre-processing of the database follows the industry practices for WT SCADA data, which includes filtering and normalization.

Measures from the SCADA system eventually present unlikely values (e.g., beyond physical limits) that can bias and distort the CM approach. Filters are therefore used to remove any physically incoherent values. Band-pass filters specific to each measure are set with the lower and upper limits from Table 2 [49]. For the database under analysis, data outside the limits in Table 2 correspond to less than 1% of all data points.

Moreover, the CM analysis considers only data corresponding to energy production. A second filtering step is therefore included to select data points verifying  $P(kW) > 0$  and  $n_{ROTOR}(rpm) > 0$ .

After the two filtering steps, each measure is normalized to the  $[0, 1]$  interval for use in the NN models. The min-max normalization technique [50] uses the limits from Table 2.

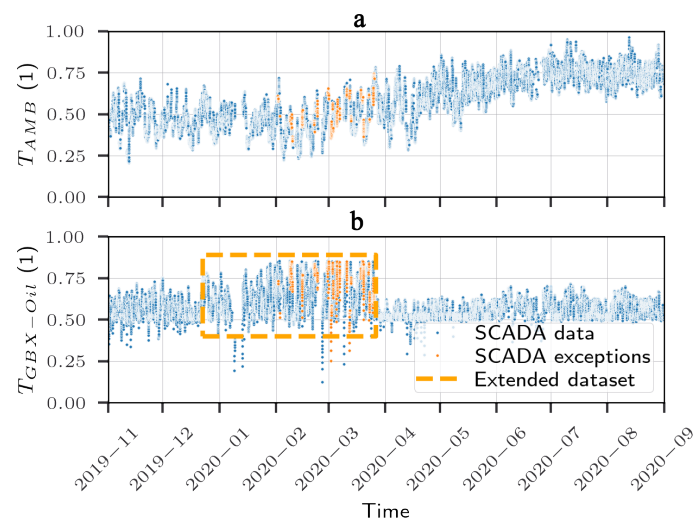
### 5.2. Data Labeling, Balancing, and Partitioning

The task of dataset labeling uses three sources of information: SCADA raw measures, SCADA log files, and O&M reports. Combining these data and metadata allows for the definition of the datasets corresponding to the conditions listed in Table 1. To detect a changing operating condition early on, the data points selected to represent a given condition might exceed the set of data points with a reported SCADA exception. For a given condition, the three steps to define the dataset are the following:

- (i) Select a subset of the WT affected by the condition of interest. This step uses the SCADA log files and the O&M reports.
- (ii) Enumerate the degradation cases from the subset of WT defined in (i). The outcome of this step is a list of cases, each identified by the WT identifier, the starting instant, and the ending instant.
- (iii) Gather the data corresponding to the cases listed in (ii). The resulting dataset comprises data points from multiple wind turbines and is supposed to represent the overall condition of interest.

Figure 7 illustrates step (ii) of the definition of the GBX dataset. The gearbox oil temperature  $T_{GBX-OIL}$  is plotted within an interval comprising the "Gearbox oil overtemperature" SCADA alarm. The threshold for abnormal behavior is manually set considering multiple samples for the same condition, which leads to the definition of the extended dataset highlighted in Figure 7.

The datasets resulting from step (iii) have different numbers of points. Therefore, a data augmentation technique based on the VAE model is used to balance the datasets [51,52]. Precisely, each homogeneous condition dataset is used to train a VAE model. Its latent space is the support to randomly generate points using the Gaussian distribution. Then, the VAE decoding of the latent space points generates the augmented dataset [52]. The target number of points for each dataset is  $n = 10,000$ .



**Figure 7.** Illustration of the definition of the dataset “Gearbox oil overtemperature” considering time series for two normalized measures: (a) ambient temperature  $T_{AMB}(1)$  and (b) gearbox oil temperature  $T_{GBX-Oil}(1)$ . Time is indicated in the format YYYY-MM.

The final database is a collection of six datasets with  $n = 10,000$  data points each. For the sake of training and evaluation, this database is partitioned into three databases: training (with 50% of all data points), validation (20%), and testing (30%).

### 5.3. VAEC Model: Architecture and Training

The proposed approach was implemented with Python (ver. 3.10). In particular, the implementation of the NN models used TensorFlow [53] and the Keras API. The VAEC architecture and training hyperparameters are described below. These settings were chosen following sensitivity analysis with the training and testing databases.

- Dimensions: input space  $n_F = 13$ ; latent space  $n_L = 2$ ; classification output  $n_C = 6$ .
- Architecture:
  - Encoder: the encoder is a DNN with three hidden layers. The number of nodes per layer is, successively, 13 (input layer), 32, 16, 8, and 2 (output layer). The encoder input layer is set with the *ReLU* activation function, while *tanh* is used in the hidden layers. Moreover, a 10% dropout layer enters after the 32-node layer to prevent overfitting.
  - Decoder: the decoder is symmetric to the encoder (nodes per layer: 2, 8, 16, 32, and 13). The decoder output layer is set with the linear activation function. All the other decoder layers are set with the *tanh* activation function.
  - Classification NN: the input of the classification NN is the latent space with dimension 2. The successive hidden layers have a decreasing number of nodes: 128, 64, 32, and 16. The *tanh* activation function is used in the input and the hidden layers. Finally, the classification output is a six-node layer using the *Softmax* activation function.
- Training: the Adam algorithm [54] is used with learning rate 0.0001, clipvalue 0.3, number of epochs 1024, and batch size 128.

## 6. Results and Discussion

The reference VAEC model is presented in Section 6.1. Two case studies in Section 6.2 demonstrate the proposed approach.

6.1. Reference VAEC Model

Training the VAE with the wind turbine condition database described in Section 2 results in latent spaces with entangled clusters. This is illustrated in Figure 8 for  $\beta_{kl} = 1$  and  $\beta_{kl} = 0.05$ .

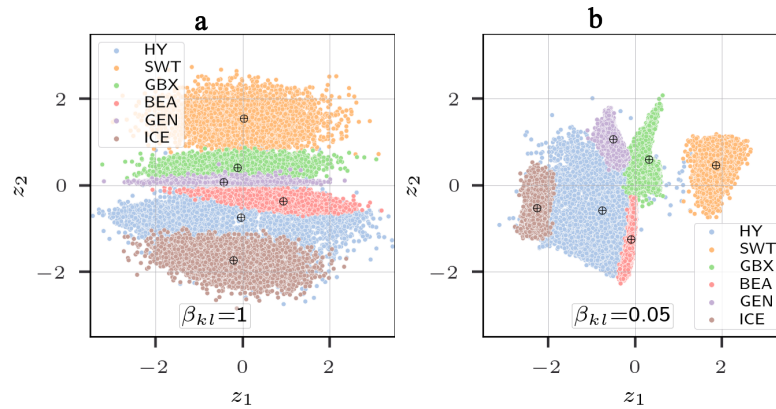


Figure 8. VAE latent space distribution for the labeled wind turbine condition database from Table 1. (a)  $\beta_{kl} = 1$  (b)  $\beta_{kl} = 0.05$ .

Adjusting  $\beta_{kl}$  to decrease the weight of the KL loss disentangles the clusters to some extent, but not sufficiently for the purposes of defining the health indicator and an intuitive visualization tool. The VAEC was designed to disentangle the clusters corresponding to the labeled database of wind turbine conditions.

Due to the randomness of the VAEC training and the reparametrization trick, multiple training instances have different outcomes. In particular, the encoding of the training database has different projections into the latent space. Comparative analysis with multiple loss function coefficients led to the choice of the hyperparameters  $\beta_{kl} = 0.05$  and  $\beta_{cl} = 10$ , which gives a classification confusion matrix with accuracy higher than 98% for multiple training instances, and a latent space with overall disentangled clusters.

Figure 9 depicts the VAEC-encoding projection and confusion matrix for one specific training of the VAEC model with the retained loss function coefficients— $\beta_{kl} = 0.05$  and  $\beta_{cl} = 10$ . The corresponding VAEC model (architecture and identified parameters) is adopted as the reference for the online CM approach.

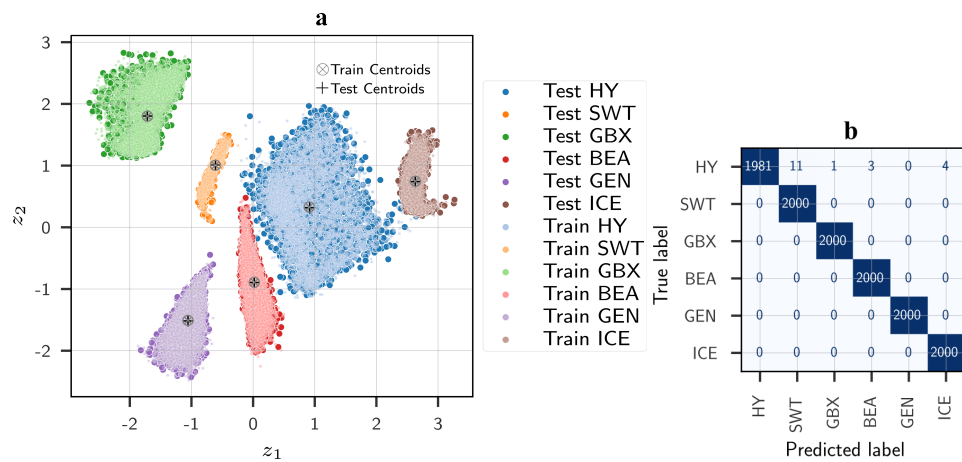


Figure 9. VAEC latent space distribution for the labeled wind turbine condition database from Table 1 with  $\beta_{kl} = 0.05$  and  $\beta_{cl} = 10$ . (a) Encoding of train and test databases. (b) Classifier confusion matrix for the test database.

Notice that the VAEC encoding of the test datasets occupies virtually the same regions as the corresponding training datasets in Figure 9a. The encoding of the training database defines the visualization tool's static layer. The VAEC classifier has accuracy close to 100% for the training instance above, as depicted in Figure 9b. Such high performance can be related to the quality of the semi-manual labeling of the datasets. It also supports the choice of the latent variable  $z$  as the input for the classifier in the VAEC model.

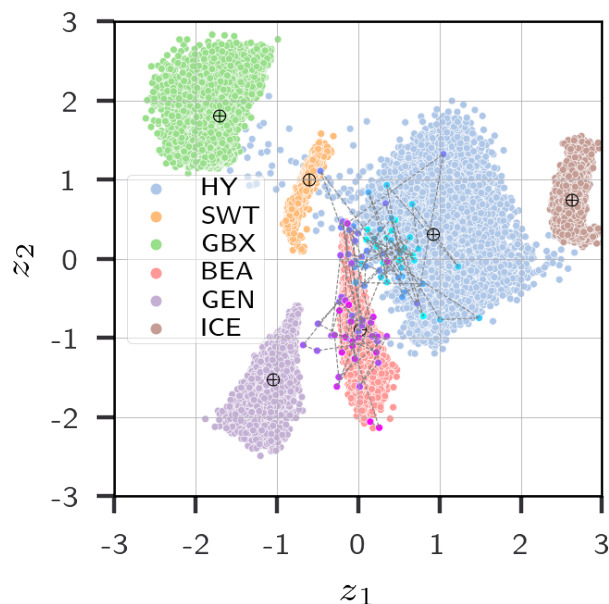
## 6.2. Case Studies

Two case studies demonstrate the proposed approach: Section 6.2.2 analyzes the case of a main bearing degradation until failure, and Section 6.2.1 presents the impact of a severe cold wave on the WT operation.

### 6.2.1. Case Study I: Main Bearing Degradation

Case Study I refers to the degradation of the main bearing of the machine  $WT_A$  within the period [1 May 2020, 1 August 2020]. This case is particularly suitable to demonstrate the proposed approach given that it starts at the healthy condition and then degrades progressively until its failure and shutdown. The machine  $WT_A$  was kept shut down for more than a month twice within 2020: first, after the shut down at  $t_{ShutDown1} = 1$  March 2020; then, months later, after the shut down at  $t_{ShutDown2} = 10$  August 2020. The available data suggest that a provisory maintenance intervention occurred in March, and a second one was performed in August after the main bearing failure. Such a chronology of O&M interventions is related to the weather conditions at the wind farm location. Indeed, historical data show that the windiest months were April, May, and June. August, on the contrary, had the lowest average wind speed.

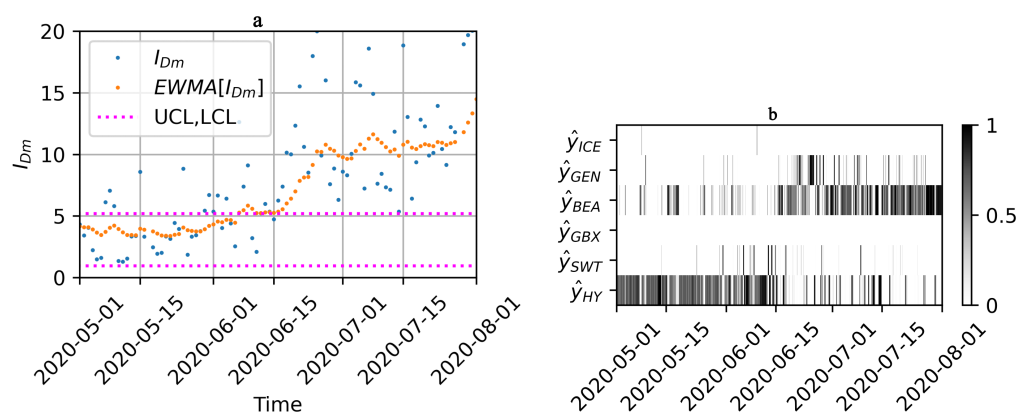
The reference VAEC model allows the projection of the data points from Case Study I, resulting in the visualization tool depicted in Figure 10.



**Figure 10.** Visualization tool for Case Study II:  $WT_A$ , [1 May 2020, 1 August 2020]. Cyan-colored markers correspond to the beginning of the timeline, and magenta-colored markers toward the end of it.

In practice, the visualization tool might be updated periodically, with the inclusion of new points in the trajectory representing the wind turbine condition. Figure 10 can be interpreted as a snapshot of the visualization tool at  $t = 1$  August 2020. It shows that the WT was initially in the healthy cluster (as the cyan markers superpose the HY cluster) and then clearly shifted to the main bearing overtemperature (as the magenta markers superpose the BEA cluster).

The detection follows from the HI  $I_{Dm}$  and the EWMA control chart, as introduced in Section 4.2. Figure 11a depicts  $I_{Dm}$  and  $EWMA[I_{Dm}]$ . These time series lead to an alarm at  $t_D = 15$  June 2020. This date is 11 days before the SCADA alarm at  $t_S = 26$  June 2020 and one week before the detection by the regression model residue-based approach proposed in Cambron et al. [55]. The latter method performs the main bearing degradation exclusively and leads to detection at  $t_{D02} = 21$  June 2020. The same period [1 October 2019, 15 December 2019] is used as the reference for the healthy condition in both estimations.



**Figure 11.** Case Study II: (a) detection, (b) diagnosis. Time is indicated in the format YYYY-MM-DD.

The diagnosis is provided by the classifier output  $\hat{y}$  depicted in Figure 11b. Each vertical line corresponds to a 10-min time step. HY is the most probable condition at the beginning of the period, with  $\hat{y}_{HY} = 1$  most of the time. After 15 June 2020, however, the most probable condition becomes the main bearing overtemperature (BEA). There is some dispersion on  $\hat{y}$ , the classifier eventually indicating other overheating conditions such as GEN and STW. The authors judge that such dispersion partially reflects the evolution of the WT's condition over time. Ultimately, one might define the diagnosis status by setting adequate post-processing on  $\hat{y}$ . The definition of the diagnosis status is left for future work.

The diagnosis in the proposed approach is based on the classifier DNN that constitutes the VAEC model. The confusion matrix from Figure 9b evidences that the VAEC properly categorizes among selected WT conditions, therefore outperforming the diagnosis based only on the unsupervised VAE model [23]. The VAEC model is simpler than both AVAE\_SDL [28] and CVAE-GAN [27]. A strict comparison of the performance of the three models is a difficult task since they use different types of inputs. For instance, the diagnosis from Zhang et al. [27] relies on vibration data unavailable in the database under analysis.

### 6.2.2. Case Study II: Impact of a Cold Wave

This case study analyzes the operation of the machine  $WT_B$  under icing conditions observed during the 2021 cold wave in North America [56]. The ambient temperature measured by the  $WT_B$ 's SCADA system reached values as low as  $T_{AMB} - 16.5$  °C in February 2021. The blades of the WT under analysis are not equipped with anti-icing systems. In case of exceptional icing conditions, careful condition monitoring is required to detect and diagnose ice accretion on blades, which should trigger protocols for curtailment or even shutdown [57].

Figure 12 depicts the visualization tool for Case Study II, i.e., data points from the machine  $WT_B$  covering the period of interest [1 December 2020, 4 March 2021]. Such a projection uses the reference VAEC model defined in Section 6.1.



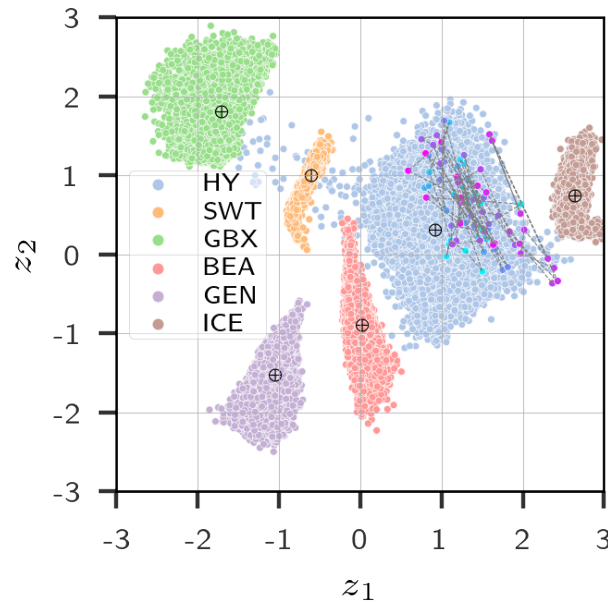


Figure 12. Visualization tool for Case Study I:  $WT_B$ , period [1 December 2020, 4 March 2021].

Figure 12 reveals a shift in the WT condition toward the ICE cluster. It is worth noticing that the trajectory in the latent space remains mostly inside the HY cluster, indicating that the ice accretion observed during the period under analysis was milder than the cases in the ICE training dataset. The power curve confirms this since the Case Study II data points remain closer to the HY cluster than the ICE cluster. More severe ice accretion is expected to shift the projection into the ICE cluster.

The proposed CM approach captures the  $WT_B$  changing condition under the 2021 cold wave. Indeed, as depicted in Figure 13a, the period of the cold wave corresponds to an increased value for  $I_{Dm}$ , with an alarm triggered at  $t_D = 15$  February 2021. The EWMA control chart used the period [5 December 2020, 5 January 2021] as a reference.

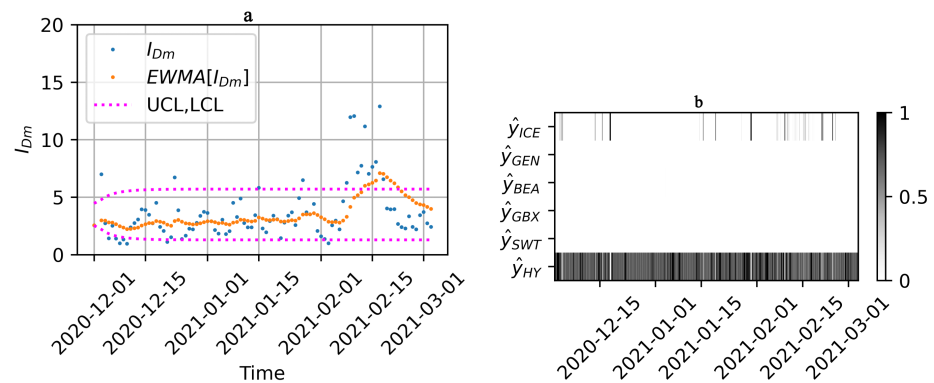


Figure 13. Case Study I: (a) detection and (b) diagnosis. Time is indicated in the format YYYY-MM-DD.

Figure 13b depicts the classifier output  $\hat{y}$ . Here,  $\hat{y}_{HY} = 1$  most of the time, except for a few time steps when  $\hat{y}_{ICE} = 1$ . The comparison with competing approaches is left for future work.

From the O&M standpoint, the actions to take in the scenario of ice accretion on blades are limited. In extreme cases, curtailment or shutdown might be necessary to prevent overloading of the rotor blades. After such an event, damage detection specific to the wind turbine blades would be recommended to assess the health condition [58,59].

## 7. Summary and Conclusions

The present work introduced a new CM approach allowing for the early detection and diagnosis of abnormal conditions in WTs. The proposed approach includes a visualization tool for enhanced interpretability.

A supervised implementation of the VAE model was introduced to project the multiple-condition SCADA system into a 2D disentangled representation space. This 2D representation is a built-in visualization tool for the AI model. Furthermore, the low-dimensional representation is representative of the dynamics of the physical measures, and it led to the definition of the proposed HI, the respective alarm criteria, and the classifier giving diagnosis information.

It was shown that each condition in the physical space was VAEC-encoded into a specific region in the latent space. In such a space, the evolution of the WT condition is expressed as a trajectory in a 2D space. The proposed  $I_{Dm}$  uses the Mahalanobis distance to measure the distance between data points and the healthy cluster distribution in the latent space. The EWMA control chart allowed for the detection of changing trends in the HI time series.

Two case studies pertinently demonstrated the potential of the proposed approach in detecting, diagnosing, and visually representing abnormal conditions in WTs. The proposed alarm criteria led to satisfactory anticipation of the SCADA alarms. In particular,  $I_{Dm}$  triggered an alarm for main bearing overtemperature 11 days earlier than the best-performing competing method. The two case studies illustrated the pertinence of the visualization tool. In practice, the visualization tool would be updated periodically in the online CM, providing a visual representation of any trends in the WT condition. Such a tool enhances the interpretability of the outcomes, therefore easing its use by O&M practitioners.

It is worth mentioning that the supervised training of the VAEC implies the need for labeled datasets. Such a task is time-consuming and costly. Fortunately, once the offline phase is completed for a subset of WTs, the online CM is inexpensive, and its parameters can be adjusted according to the in situ experience. Further research might investigate alternatives to the VAEC model, eventually a semi-supervised VAE or methods to disentangle the (unsupervised) VAE latent space. The proposed visualization tool is an advantage compared to black-box models and establishes new examples for the literature on interpretable AI. Nevertheless, metrics to evaluate the interpretability remain to be established. A future investigation should evaluate the acceptability of the proposed visualization tool among O&M practitioners. Finally, the industrial implementation of the proposed approach would require evaluating the proposed approach within an extensive selection of case studies and WT models.

**Author Contributions:** Conceptualization, A.O.-F., R.Z., P.C., and A.T.; methodology, A.O.-F., R.Z., and A.T.; software, A.O.-F., R.Z., and P.C.; validation, A.O.-F., R.Z., P.C., and A.T.; formal analysis, A.O.-F. and R.Z.; investigation, A.O.-F.; resources, P.C. and A.T.; data curation, A.O.-F. and P.C.; writing—original draft preparation, A.O.-F.; writing—review and editing, A.O.-F., R.Z., P.C., and A.T.; visualization, A.O.-F.; supervision, R.Z., P.C., and A.T.; project administration, P.C. and A.T.; funding acquisition, P.C. and A.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by MITACS through the grants MITACS Globalink Graduate (grant number IT24180-FR62006) and MITACS Acceleration (grant number IT22958-FR60047).

**Data Availability Statement:** Restrictions apply to the availability of these data. Data were obtained from Power Factors.

**Acknowledgments:** The authors acknowledge Power Factors for providing the database used in the present study.

**Conflicts of Interest:** The authors declare no competing financial interests or personal relationships that could influence the present work.

## Abbreviations

The following abbreviations are used in this manuscript:

AE	Autoencoder
AI	Artificial Intelligence
CM	Condition Monitoring
DNN	Deep Neural Network
EWMA	Exponentially Weighted Moving Average
HI	Health Indicator
KL	Kullback-Liebler
LCOE	Levelized Cost of Energy
NN	Neural Network
O&M	Operation and Maintenance
SCADA	Supervisory Control and Data Acquisition
VAE	Variational AE
VAEC	VAE Embedded with Classification
WT	Wind Turbine

## References

1. Dao, C.; Kazemtabrizi, B.; Crabtree, C. Wind turbine reliability data review and impacts on levelised cost of energy. *Wind Energy* **2019**, *22*, 1848–1871. [[CrossRef](#)]
2. Costa, Á.M.; Orosa, J.A.; Vergara, D.; Fernández-Arias, P. New tendencies in wind energy operation and maintenance. *Appl. Sci.* **2021**, *11*, 1386. [[CrossRef](#)]
3. Nicod, J.M.; Chebel-Morello, B.; Varnier, C. *From Prognostics and Health Systems Management to Predictive Maintenance 2: Knowledge, Reliability and Decision*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
4. Tautz-Weinert, J.; Watson, S.J. Using SCADA data for wind turbine condition monitoring—A review. *IET Renew. Power Gener.* **2017**, *11*, 382–394. [[CrossRef](#)]
5. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2014**, arXiv:1312.6114.
6. Helbing, G.; Ritter, M. Deep Learning for fault detection in wind turbines. *Renew. Sustain. Energy Rev.* **2018**, *98*, 189–198. [[CrossRef](#)]
7. Badihi, H.; Zhang, Y.; Jiang, B.; Pillay, P.; Rakheja, S. A Comprehensive Review on Signal-Based and Model-Based Condition Monitoring of Wind Turbines: Fault Diagnosis and Lifetime Prognosis. *Proc. IEEE* **2022**, *110*, 754–806. [[CrossRef](#)]
8. Zemouri, R.; Levesque, M.; Amyot, N.; Hudon, C.; Kokoko, O.; Tahan, S.A. Deep convolutional variational autoencoder as a 2D-visualization tool for partial discharge source classification in hydrogenerators. *IEEE Access* **2019**, *8*, 5438–5454. [[CrossRef](#)]
9. Gilpin, L.H.; Bau, D.; Yuan, B.Z.; Bajwa, A.; Specter, M.; Kagal, L. Explaining explanations: An overview of interpretability of machine learning. In Proceedings of the IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 1–3 October 2018; pp. 80–89.
10. Lipton, Z.C. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue* **2018**, *16*, 31–57. [[CrossRef](#)]
11. Tits, N.; Wang, F.; Haddad, K.E.; Pagel, V.; Dutoit, T. Visualization and interpretation of latent spaces for controlling expressive speech synthesis through audio analysis. *arXiv* **2019**, arXiv:1903.11570
12. Proteau, A.; Zemouri, R.; Tahan, A.; Thomas, M. Dimension reduction and 2D-visualization for early change of state detection in a machining process with a variational autoencoder approach. *Int. J. Adv. Manuf. Technol.* **2020**, *111*, 3597–3611. [[CrossRef](#)]
13. Banko, L.; Maffettone, P.M.; Naujoks, D.; Olds, D.; Ludwig, A. Deep learning for visualization and novelty detection in large X-ray diffraction datasets. *Npj Comput. Mater.* **2021**, *7*, 1–6. [[CrossRef](#)]
14. Cheng, R.C.; Chen, K.S. Ball bearing multiple failure diagnosis using feature-selected autoencoder model. *Int. J. Adv. Manuf. Technol.* **2022**, *120*, 4803–4819. [[CrossRef](#)]
15. Zhao, H.; Liu, H.; Hu, W.; Yan, X. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renew. Energy* **2018**, *127*, 825–834. [[CrossRef](#)]
16. Wang, L.; Zhang, Z.; Xu, J.; Liu, R. Wind Turbine Blade Breakage Monitoring with Deep Autoencoders. *IEEE Trans. Smart Grid* **2018**, *9*, 2824–2833. [[CrossRef](#)]
17. Yuan, B.; Yuan, B.; Wang, C.; Luo, C.; Luo, C.; Jiang, F.; Jiang, F.; Long, M.; Yu, P.S.; Liu, Y. WaveletAE: A Wavelet-enhanced Autoencoder for Wind Turbine Blade Icing Detection. *arXiv* **2019**, arXiv:1902.05625.
18. Hemmer, M.; Klausen, A.; Van Khang, H.; Robbersmyr, K.G.; Waag, T.I. Health indicator for low-speed axial bearings using variational autoencoders. *IEEE Access* **2020**, *8*, 35842–35852. [[CrossRef](#)]
19. Jiang, G.; Xie, P.; He, H.; Yan, J. Wind Turbine Fault Detection Using a Denoising Autoencoder With Temporal Information. *IEEE-Asme Trans. Mechatron.* **2017**, *23*, 89–100. [[CrossRef](#)]
20. Wu, X.; Jiang, G.; Wang, X.; Xie, P.; Li, X.; Li, X. A Multi-Level-Denoising Autoencoder Approach for Wind Turbine Fault Detection. *IEEE Access* **2019**, *7*, 59376–59387. [[CrossRef](#)]

21. Renström, N.; Bangalore, P.; Highcock, E. System-wide anomaly detection in wind turbines using deep autoencoders. *Renew. Energy* **2020**, *157*, 647–659. [[CrossRef](#)]
22. Lei, J.; Liu, C.; Jiang, D. Fault diagnosis of wind turbine based on Long Short-term memory networks. *Renew. Energy* **2019**, *133*, 422–432. [[CrossRef](#)]
23. Roelofs, C.M.; Lutz, M.A.; Faulstich, S.; Vogt, S. Autoencoder-based anomaly root cause analysis for wind turbines. *Energy AI* **2021**, *4*, 100065. [[CrossRef](#)]
24. Stetco, A.; Dinmohammadi, F.; Zhao, X.; Robu, V.; Flynn, D.; Barnes, M.; Keane, J.; Nenadic, G. Machine learning methods for wind turbine condition monitoring: A review. *Renew. Energy* **2019**, *133*, 620–635. [[CrossRef](#)]
25. Peeters, C.; Guillaume, P.; Helsen, J. Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy. *Renew. Energy* **2018**, *116*, 74–87. [[CrossRef](#)]
26. Barszcz, T. *Vibration-Based Condition Monitoring of Wind Turbines*; Springer: Berlin/Heidelberg, Germany, 2019.
27. Zhang, L.; Zhang, H.; Cai, G. The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–12. [[CrossRef](#)]
28. Liu, X.; Teng, W.; Wu, S.; Wu, X.; Liu, Y.; Ma, Z. Sparse dictionary learning based adversarial variational auto-encoders for fault identification of wind turbines. *Measurement* **2021**, *183*, 109810. [[CrossRef](#)]
29. Proteau, A.; Zemouri, R.; Tahan, A.; Thomas, M.; Bounouara, W.; Agnard, S. CNC machining quality prediction using variational autoencoder: A novel industrial 2 TB dataset. In Proceedings of the Prognostics and Health Management Conference, London, UK, 27–29 May 2022.
30. Roberts, S. Control Chart Tests Based on Geometric Moving Averages. *Technometrics* **1959**, *1*, 239–250. [[CrossRef](#)]
31. Beretta, M.; Julian, A.; Sepulveda, J.; Cusidó, J.; Porro, O. An ensemble learning solution for predictive maintenance of wind turbines main bearing. *Sensors* **2021**, *21*, 1512. [[CrossRef](#)]
32. Hochart, C.; Fortin, G.; Perron, J.; Ilinca, A. Wind turbine performance under icing conditions. *Wind Energy Int. J. Prog. Appl. Wind Power Convers. Technol.* **2008**, *11*, 319–333. [[CrossRef](#)]
33. *Standard IEC 61400; Wind Energy Generation Systems—Part 12-1: Power Performance Measurement of Electricity Producing Wind Turbines*. International Electrotechnical Commission: Geneva, Switzerland, 2022.
34. Pandit, R.; Astolfi, D.; Hong, J.; Infield, D.; Santos, M. SCADA data for wind turbine data-driven condition/performance monitoring: A review on state-of-art, challenges and future trends. *Wind Eng.* **2022**, *1*, 20. [[CrossRef](#)]
35. Guo, P.; Infield, D. Wind turbine blade icing detection with multi-model collaborative monitoring method. *Renew. Energy* **2021**, *179*, 1098–1105. [[CrossRef](#)]
36. Zeng, H.; Dai, J.; Zuo, C.; Chen, H.; Li, M.; Zhang, F. Correlation Investigation of Wind Turbine Multiple Operating Parameters Based on SCADA Data. *Energies* **2022**, *15*, 5280. [[CrossRef](#)]
37. Beretta, M.; Pelka, K.; Cusidó, J.; Lichtenstein, T. Quantification of the Information Loss Resulting from Temporal Aggregation of Wind Turbine Operating Data. *Appl. Sci.* **2021**, *11*, 8065. [[CrossRef](#)]
38. Marti-Puig, P.; Blanco-M, A.; Cárdenas, J.J.; Cusidó, J.; Solé-Casals, J. Effects of the pre-processing algorithms in fault diagnosis of wind turbines. *Environ. Model. Softw.* **2018**, *110*, 119–128. [[CrossRef](#)]
39. Kingma, D.P.; Welling, M. An introduction to variational autoencoders. *Found. Trends<sup>®</sup> Mach. Learn.* **2019**, *12*, 307–392. [[CrossRef](#)]
40. Doersch, C. Tutorial on variational autoencoders. *arXiv* **2016**, arXiv:1606.05908.
41. Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S.; Lerchner, A. Beta-vae: Learning basic visual concepts with a constrained variational framework. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
42. Zemouri, R.; Lévesque, M.; Boucher, É.; Kirouac, M.; Lafleur, F.; Bernier, S.; Merkhof, A. Recent Research and Applications in Variational Autoencoders for Industrial Prognosis and Health Management: A Survey. In Proceedings of the Prognostics and Health Management Conference (PHM-2022 London), London, UK, 27–29 May 2022; pp. 193–203.
43. Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Process. Mag.* **2012**, *29*, 141–142. [[CrossRef](#)]
44. Kingma, D.P.; Mohamed, S.; Jimenez Rezende, D.; Welling, M. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems 27 (NIPS 2014)*; NeurIPS: New Orleans, LA, USA, 2014; Volume 27.
45. Mathieu, E.; Rainforth, T.; Siddharth, N.; Teh, Y.W. Disentangling disentanglement in variational autoencoders. In Proceedings of the International Conference on Machine Learning. PMLR, Long Beach, CA, USA, 5–9 June 2019; pp. 4402–4412.
46. Ezukwoke, K.; Hoayek, A.; Batton-Hubert, M.; Boucher, X. GCVAE: Generalized-Controllable Variational AutoEncoder. *arXiv* **2022**, arXiv:2206.04225.
47. Sohn, K.; Lee, H.; Yan, X. Learning structured output representation using deep conditional generative models. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*; NeurIPS: New Orleans, LA, USA, 2015; Volume 28.
48. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Newton, MA, USA, 2022.
49. Oliveira-Filho, A.M.d.; Cambron, P.; Tahan, A. Condition Monitoring of Wind Turbine Main Bearing Using SCADA Data and Informed by the Principle of Energy Conservation. In Proceedings of the 2022 Prognostics and Health Management Conference (PHM-2022 London), London, UK, 27–29 May 2022; pp. 276–282.
50. Patro, S.; Sahu, K.K. Normalization: A preprocessing stage. *arXiv* **2015**, arXiv:1503.06462.

51. Tanner, M.A.; Wong, W.H. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* **1987**, *82*, 528–540. [[CrossRef](#)]
52. Chadebec, C.; Thibeau-Sutre, E.; Burgos, N.; Allasonnière, S. Data augmentation in high dimensional low sample size setting using a geometry-based variational autoencoder. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 2879–2896. [[CrossRef](#)]
53. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: A system for Large-Scale machine learning. In Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), Savannah, GA, USA, 2–4 November 2016; pp. 265–283.
54. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
55. Cambron, P.; Tahan, A.; Masson, C.; Pelletier, F. Bearing temperature monitoring of a wind turbine using physics-based model. *J. Qual. Maint. Eng.* **2017**, *23*, 479–488. [[CrossRef](#)]
56. Doss-Gollin, J.; Farnham, D.J.; Lall, U.; Modi, V. How unprecedented was the February 2021 Texas cold snap? *Environ. Res. Lett.* **2021**, *16*, 064056. [[CrossRef](#)]
57. Veers, P.; Kroposki, B.; Novacheck, J.; Gevorgian, V.; Laird, D.; Zhang, Y.; Corbus, D.; Baggu, M.; Palmintier, B.; Dhulipala, S. *Examination of the Extreme Cold Weather Event Affecting the Power System in Texas—February 2021*; Technical Report; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2021.
58. Márquez, F.P.G.; Chacón, A.M.P. A review of non-destructive testing on wind turbines blades. *Renew. Energy* **2020**, *161*, 998–1010. [[CrossRef](#)]
59. Kaewniam, P.; Cao, M.; Alkayem, N.F.; Li, D.; Manoach, E. Recent advances in damage detection of wind turbine blades: A state-of-the-art review. *Renew. Sustain. Energy Rev.* **2022**, *167*, 112723. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.