IEEE Open Journal of
Industry Applications

Received 18 December 2023; revised 27 February 2024; accepted 13 March 2024. Date of publication 26 March 2024;
date of current version 4 April 2024. The review of this article was arranged by Associate Editor M. Di Nardo.

Digital Object Identifier 10.1109/OJIA.2024.3380249

Fault Detection Based on Vibration Measurements and Variational Autoencoder-Desirability Function

RONY IBRAHIM [1], RYAD ZEMOURI [2], ANTOINE TAHAN [1], BACHIR KEDJAR [1] (Member, IEEE), AREZKI MERKHOUF [2] (Senior Member, IEEE), AND KAMAL AL-HADDAD [1] (Life Fellow, IEEE)

[1]École de Technologie Supérieure, Montréal, QC H3C 1K3, Canada
[2]Centre de Recherche d'Hydro-Québec (CRHQ), Varennes, QC J3X 1S1, Canada

CORRESPONDING AUTHOR: RONY IBRAHIM (e-mail: rony.ibrahim.1@ens.etsmtl.ca)

This work was supported in part by the CRSNG under Grant RDCPJ 535809-18, in part by the InnovÉÉ under Grant R13-1805, in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), in part by the Innovation en énergie électrique (InnovÉÉ), Hydro-Québec, in part by the Opal-RT, the industrial partners of the project, and in part by the École de Technologie supérieure (ETS).

ABSTRACT In the field of electrical machines maintenance, accurate and timely diagnosis plays a crucial role in ensuring reliability and efficiency. Variational autoencoder (VAE) techniques have emerged as a promising tool for fault classification due to their robustness in handling complex data. However, the inherent nondeterministic aspect of the VAE creates a significant challenge as it leads to varying cluster locations for identical health states across different machines. This variability complicates the creation of a standardized applicable diagnostic tool and challenges for the implementation of effective real-time health monitoring and prognostics. Addressing this issue, a novel approach is proposed wherein a desirability function-based term is integrated into the cost function of the VAE. The enhancement achieved by this approach arises from the standardization of classification, guaranteeing that analogous faults are assigned to identical geolocations within a 2-D user-friendly space. This method's efficacy is validated through two separate case studies: one analyzing vibration data from two diverse designs of large existing hydrogenerators, and the other utilizing vibration data sourced from an open-access dataset focused on bearing fault. The findings of both studies show that the model can cluster 97% of similar faults into preset zones, compared with 40% when the desirability term is excluded.

INDEX TERMS Desirability function, diagnosis, fault detection, large hydrogenerators, variational autoencoder (VAE), vibration.

I. INTRODUCTION

Electrical machines (EM), including induction machines (IM), synchronous generators (SG), and salient pole synchronous generators (SPSG), represent intricate systems that involve multiple physical phenomena. It is essential to implement early-stage fault detection and diagnosis techniques to maintain a stable and dependable energy supply while minimizing economic losses. Different types of faults can manifest in EM, notably affecting the rotor and stator windings, bearings, or the stator core [1]. Electrical faults encompass rotor interturn short circuit (RITSC), for example, while mechanical issues include eccentricities (static, dynamic, or hybrid), bearing faults (inner race fault, outer race fault, etc.), and broken damper bar (BDB). These degradation phenomena, whether occurring individually or in combination, have been examined in prior studies, such as in [2] and [3]. In recent years, various physical entities, such as the air gap flux [4], line current [5], torque [6], vibration [7], acoustics [8], stray flux [9], and more, have found applications in fault diagnosis. This trend has been notably highlighted in [1] and [10], especially with the advent of artificial intelligence (AI) and deep learning (DL) techniques that are summarized in [11] and [12]. Through the integration of AI with these physical entities, the approach leverages AI-based methods to achieve

more precise and dependable fault diagnosis, with the added advantage of adaptability to new data.

The existing literature underscores the significance of vibration signals for fault diagnosis, primarily due to their noninvasive nature [2], [3]. These signals have been employed for detecting RITSC in small SPSG. This has been achieved through a combination of experimental measurements and numerical modeling, as demonstrated in studies such as [13] and [14]. In addition, vibration signals have proven effective in identifying eccentricities as in [15].

Recent research in the domain of large hydrogenerator diagnosis has centered on the utilization of DL techniques, specifically neural networks (NN), with a particular emphasis on the variational autoencoder (VAE), a derivative of the conventional autoencoder (AE) introduced by Kingma [16]. The power of VAE to classify partial discharge sources and patterns in large hydrogenerators has been proved in [17] and [18]. Moreover, the VAE has demonstrated its ability to detect faults at an earlier stage in comparison with statistical monitoring techniques, particularly in the context of vibration analysis, as highlighted in [19]. In [20], a novel approach was adopted, combining VAE with sparse dictionary learning, to achieve early anomaly detection based on the reconstruction error of a model trained on real vibratory signals from large hydrogenerators. Furthermore, this VAE-based approach has been applied to assess various levels of RITSC severity within large hydrogenerators, providing a user-friendly 3-D visualization, as described in [7]. In addition, this model has been used for the detection of diverse rotor faults, each exhibiting varying degrees of severity, within a user-friendly 2-D space, as explained in [7]. Notably, a novel monitoring metric based on the Euclidean distance within the latent space was introduced, proving its effectiveness in detecting faults at earlier stages compared with alternative methods based on the root mean square (rms) of the signal and the mean squared error (MSE) of the model. It is also worth mentioning that it has also demonstrated its ability to discern faulty data in scenarios involving gearboxes, cutting tool machines, and bearings, as exemplified in [21], [22], [23] and [24], respectively.

The above references emphasize the importance of vibration signals in health monitoring and fault diagnosis. Despite ongoing attempts, models developed, notably for hydrogenerators, can only be applied to one machine design at a time. When these models are applied to diverse machine designs, inconsistencies in the locations of identical fault clusters are detected due to the nondeterministic aspect of the latent space. This difficulty not only hinders the development of a universal diagnostic tool, but also impedes the implementation of real-time health monitoring and future prognosis. As a result, the creation of a standardized and relevant model is necessary.

Therefore, the focus of this research is building upon the findings presented in [7] so as to refine the current diagnostic approach reliant on vibration data and the VAE model for fault detection and health monitoring.

To address these issues, the key contribution of the study is the integration of the desirability function into the model's cost function, as described in [25]. This function uses desirability scores calculated from the polar coordinates of all places in the obtained reduced space to identify similar fault clusters across several machines in a predetermined area. As a result, a consistent model adaptable to many machine types is used, which improves faults' detection accuracy and, ultimately, overall system dependability and performance.

As a result, two case studies are considered to emphasize the importance of incorporating this term.

1) In an initial case study, two large hydrogenerators at Hydro-Quebec, referred to as Machines 1 and 2 with varying capacities, are under analysis. The vibration dataset used is a combination of real vibration data gathered onsite and faulty ones generated by the numerical model of the machine.
2) Another case study revolves about vibration data from an open-access bearings dataset that contains healthy and faulty signals generated though experimentation.

The rest of this article is organized as follows. The methodology applied in this work is presented in Section II. Section III details the two case studies analyzed mentioned above. The results are explored and discussed in Section IV, highlighting findings with and without the incorporation of the desirability term for both scenarios. Finally, Section V concludes this article with a summary and highlights the significant contributions of this study.

II. METHODOLOGY

A. VAE ARCHITECTURE AND TRAINING METHODOLOGY

A VAE model is used in a manner akin to the approach outlined in [22]. The training and validating the model is based on two distinct datasets, denoted as Set #1 and Set #2, followed by an evaluation of its clustering performance using Set #3. The model is built with Python 3.8.8 and visual studio code 1.70.2, a user-friendly code editor with complete Python support. The code executes the model using the Keras library, linked with TensorFlow. The computational tasks are carried out using a PC with a 3.6 GHz processor, 16 physical cores, 128 GB of RAM, and a 500 GB SSD hard drive.

The architecture of the VAE model is presented in Fig. 1. Fig. 2 provides an overview of the overall architecture and a detailed structure of the model.

1) The encoder is a deep neural network (DNN), with parameters ϕ, that maps the input data x ($x \in \mathbb{R}^n$) into the latent space of standard deviation σ and mean μ. The encoder network consists of nine convolutional layers, each with 128, 64, and 32 nodes, respectively. Every hidden layer uses an activation function, which is chosen according to the data set and the range of variance in the input data (*tanh*, *sigmoid*, or *ReLU*). In addition, a 10% dropout layer is added after the 32-node layer to prevent over fitting.
2) The latent space is a dimensionally reduced space ($\mathbb{R}^p$ where p is the latent space dimension, $p = 2$ or 3), consisting of z_i components, where ϵ is a variable following a Gaussian distribution $\epsilon \sim N(0, 1)$: a clustering can be observed, as it can be seen in Fig. 1.

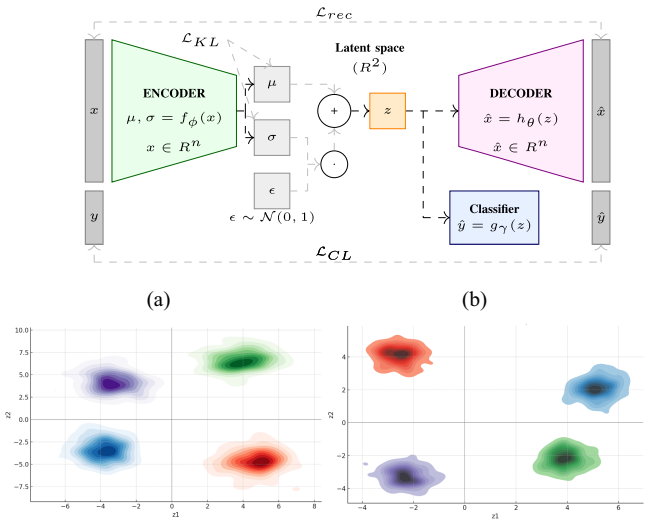

FIGURE 1. Training and validation process. (a) Latent Space-Machine 1. (b) Latent Space-Machine 2.

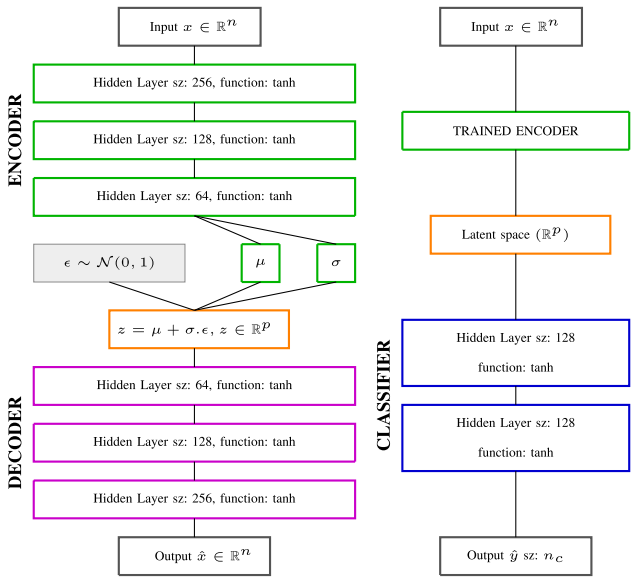

FIGURE 2. Flowchart algorithm (inspired by [22]).

The two obtained latent spaces, in Fig. 1(a) and (b), correspond to the application of the model on two different Machines 1 and 2, respectively.

3) The decoder is a NN that decodes the latent representation to reconstruct the original input $\hat{x}$, where θ is the set of the decoder parameters. The decoder's architecture is constructed by mirroring the convolutional layers of the encoder network.
4) A NN softmax classifier has been included to enhance the distribution of the latent space. Its input is the latent space variable z_i (with dimension 2). The successive hidden two conventional layers have 128 nodes each, and the *tanh* activation function is used in the input and the hidden layers. The output is the vector $\hat{y}$ with dimension n_c, where n_c represents the number of classes.

Algorithm 1: RMS of the Latent Space Variables.

Data:
$z_j^{(i)}$ is the j^{th} element of i^{th} dimension of the latent variables $\mathbf{z} \in \mathbb{R}^p$
$W_j^{(i)} = \{z_j^{(i)}\}_j^{j+N}$ is a moving window containing N elements $z_j^{(i)}$
M is the length of $z_j^{(i)}$
for $i \in \{1, ..., p\}$ **do**
  for $j \in \{1, ..., M - N\}$ **do**
    $RMS_j^{(i)} = \sqrt{\frac{1}{N}\sum_j^{j+N}(z_j^{(i)})^2}$

The training methodology uses time-variation signals and an Adam optimizer algorithm [26], with an appropriate learning rate, number of epochs, clip-value, and batch size, depending on the study case.

The cost function L_{VAE} is given by (1a) as a linear combination of different terms, where:

1) The first term, as defined in (1b), represents the reconstruction error (L_{rec}) and aims to maximize the maximum likelihood estimates (MLE), as in [16].
2) The second term, as expressed in (1c), is the Kullback–Leibler loss term (L_{KL}) and serves to assess the similarity between the encoded latent space and a Gaussian distribution, where n_s denotes the total number of samples, as introduced in [17].
3) The third term is the binary crossentropy cost function, defined in (1d), and its objective is minimization. In this context, the variable y is a result of the categorical one-hot-encoding transformation, whereas $\hat{y}$ corresponds to the output of the classifier.

Typically, $\beta_{\text{KL}} = 10^{-3}$ and $\beta_{\text{CL}} = 10^{-1}$ are the usual values used to obtain balanced components [27].

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{rec}} + \beta_{\text{KL}}\mathcal{L}_{\text{KL}} + \beta_{\text{CL}}\mathcal{L}_{\text{CL}} \tag{1a}$$

$$\mathcal{L}_{\text{rec}} = MSE(x, \hat{x}) = \frac{1}{n_s}(\|x - \hat{x}\|_2)^2 \tag{1b}$$

$$\mathcal{L}_{\text{KL}} = \sum_{i=1}^{n_s} \sigma^2{}_i + \mu^2{}_i - \log(\sigma^2{}_i) - 1 \tag{1c}$$

$$\mathcal{L}_{\text{CL}} = \sum_{i=1}^{n_s} y \log(\hat{y}) + (1 - y)\log(1 - \hat{y}). \tag{1d}$$

For the latent space, a more comprehensive analysis of the latent space is conducted, when injected the signals of Set #3, involving the computation of the rms for each latent variable z_i using Algorithm 1, where p represents the dimension of the latent space ($p = 2$ in this study).

B. DESIRABILITY FUNCTION TERM INJECTION

As depicted in the latent spaces obtained in Fig. 1 of two different machines, the clusters representing the same fault in

the two distinct machines are situated in distinct areas within both latent spaces. This separation is the result of the inherent nondeterministic nature of the latent space. Consequently, it becomes imperative to devise a standardized method that facilitates the merging of these clusters into a shared region to facilitate the diagnostic and prognostic processes. Consequently, the concept of desirability, a quantitative measure frequently employed in probability and statistics, is used [25]. A desirability function is a mathematical tool utilized to quantify the desirability or quality of a particular outcome or solution in the context of an optimization problem [28], [29], [30]. It assigns a numerical score to a solution, with higher values signifying more desirable outcomes. To account for this, a fourth term has been introduced to the cost function, as defined in (2a).

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{rec}} + \beta_{\text{KL}}\mathcal{L}_{\text{KL}} + \beta_{\text{CL}}\mathcal{L}_{\text{CL}} + \beta_{\mathcal{D}}\mathcal{L}_{\mathcal{D}} \tag{2a}$$

$$\mathcal{L}_{\mathcal{D}} = \frac{1}{\mathcal{D} + \epsilon} \tag{2b}$$

$$\mathcal{D} = \frac{1}{n}\sum_{i=1}^{n} \mathcal{D}_P \tag{2c}$$

$$\mathcal{D}_P = \sqrt{{\mathcal{D}_{R_P}}^2 + {\mathcal{D}_{\theta_P}}^2}. \tag{2d}$$

Every point, denoted as $P(z_1, z_2)$, within the latent space shown in Fig. 1(a), can be expressed in polar coordinates using (3). For each combination of R and θ, there are two corresponding desirability values, $\mathcal{D}_{R_P}$ for the radius R and $\mathcal{D}_{R_\theta}$ for the angle θ. Separate graphs for $\mathcal{D}_R$ and $\mathcal{D}_\theta$ are generated for each health state, with their attributes determined by predefined limits outlined in Fig. 3. These limits are defined according to the specific requirements set forth by experts involved in the analysis. Consequently, the desirability scores $\mathcal{D}_{R_P}$ and $\mathcal{D}_{R_\theta}$ are computed based on this framework.

$$R = \sqrt{z_1^2 + z_2^2} \tag{3a}$$

$$\theta = \arctan\left(\frac{z_2}{z_1}\right) \tag{3b}$$

Following (2d), the desirability of each point, denoted as $\mathcal{D}_P$, is calculated. This process is iteratively applied to all points within the latent space, and the overall desirability score, denoted as $\mathcal{D}$, is determined using (2c), where n represents the total number of points. Consequently, the desirability term, denoted as $\mathcal{L}_{\mathcal{D}}$, is computed as specified in (2b). This specific formulation, denoted as Derringer desirability function [31], is chosen among numerous alternatives, as detailed in [25]. In particular, the nominal-the-best formulation is chosen, in which the predicted response value is expected to approach a predetermined target value. The value of $\beta_{\mathcal{D}}$ should be set as high as possible to maximize the impact of the desirability score obtained, aiming for it to be of a similar magnitude as the other terms in the cost function. In addition, the term ϵ is used to ensure method convergence, particularly when the desirability value is close to zero. In this context, ϵ is set to 2.24×10^{-16}, in accordance with the IEEE754 standard.

The models are currently trained using the formulation provided in (2a), and the resulting latent representations of both machines are displayed in Fig. 4, with the same fault for both machines occupying a predefined region.

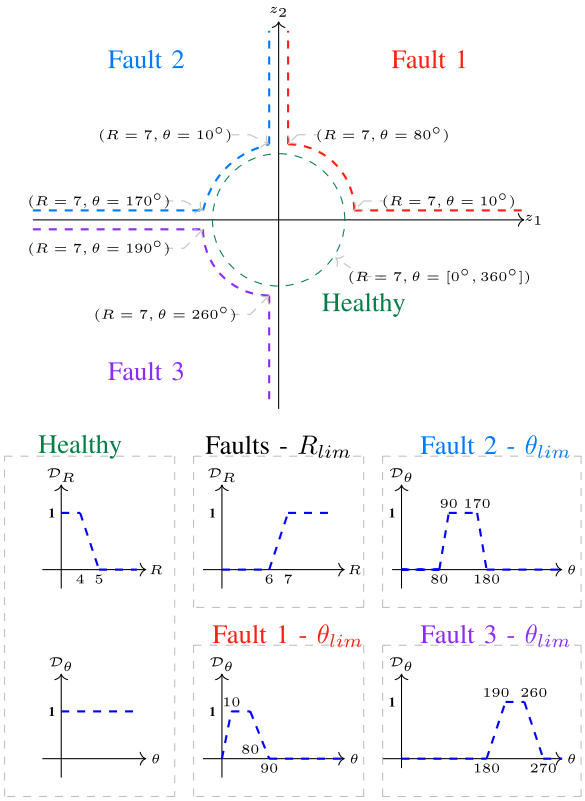

FIGURE 3. Predefined limits based on polar coordinates, radius R and angle θ.

TABLE 1 Physical Characteristics of Machines 1 and 2

	Machine 1	Machine 2
Number of poles N_p	76	56
Speed [RPM]	94.7	128.5
Power [MVA]	80	30.5
Number of slots	396	384
Power Factor	0.85	0.85

III. STUDY CASES

A. FIRST STUDY CASE–LARGE HYDROGENERATOR

1) DATASET

In this study, the vibration signals of two large hydrogenerators, installed in the parks of Hydro-Québec (QC, Canada), are taken into consideration. The characteristics of these two machines are summarized in Table 1.

Vibration measurements are conducted using an accelerometer in a campaign organized by the utility company Hydro-Quebec. These measurements have been preprocessed to improve the data analysis. Initially, they are verified and

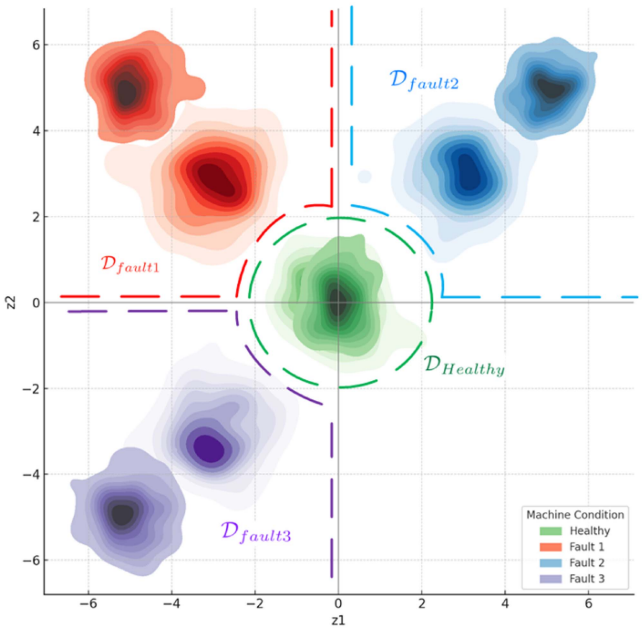

FIGURE 4. Latent space clustering upon the injection of the desirability term.

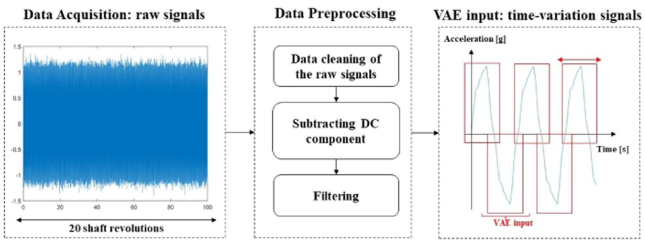

FIGURE 5. Preprocessing methodology.

cleaned. As a result, saturated and interrupted measurements are eliminated, knowing that the correctness of the outputs is dependent on the integrity of the input data. The first second of the measured signal is then eliminated to avoid data overlap with previous observations stored in the buffer. A 2 kHz low-pass filter is then used to filter out high-frequency interference. The preprocessing methodology is illustrated in Fig. 5, and it is composed of three steps: data acquisition, data processing, and VAE feeding input signals.

The primary objective of this study is to collect a substantial amount of data under various faulty conditions, which will be used to train the AI model being developed. Obtaining actual labeled faulty data was unfeasible due to the inability to shut down a functioning hydrogenerator. As a solution, a numerical model of an ideal hydrogenerator is created, based on the specifications found in real plans and documentation [7]. This model is implemented using ANSYS WORKBENCH software on a computer equipped with an Intel(R) Core(TM) $i9-12900KF$, 3.2 GHz computer with 128 GB RAM.

One of the main factors contributing to vibrations in EM is distortions in the magnetic field. The objective is to derive the electromagnetic forces and moments acting on the stator teeth from the electromagnetic model. These results serve as input data for the mechanical model. Consequently, a comprehensive model that integrates electromagnetism with vibration dynamics has been designed and fine-tuned using actual vibration measurements of the two machines.

To create an exhaustive database, various types of rotor faults, including RITSC, BDB, and static eccentricity (SE) are simulated to generate the fault pattern in the case of each fault. In the scope of this study, the following faults are taken into consideration, with only one degree of severity of each.

1) RITSC: The severity of the fault is represented by a decrease in the number of turns in one pole of the numerical model. In this study, 24% of the total number of turns per pole were examined, which equated to four short-circuited turns.
2) BDB: The material of the relevant bar is replaced with one that has nearly zero conductivity. This study takes into consideration 1BDB, which corresponds to one BDB in the same pole of the numerical model.
3) SE: Depending on the severity level, the numerical model shifts the rotor components by a specific percentage within the airgap. This study considers a 10% SE.

These patterns are injected into the actual healthy vibration signals collected in situ from both machines: The database of each machine now comprises data from both healthy and faulty signals. More detailed description of the model can be found in [7].

2) MODEL APPLICABILITY

Two distinct models are developed for each machine based on datasets containing both healthy and faulty time-variation signals. The input for each model is based on the length of a single pole of the signal, with all signals being decomposed accordingly. Hence, the database of each machine comprises $N_H = 72$ signals (including both healthy and faulty ones), each containing $n_s = 1140$ samples, and the sampling rate is set at 5 kHz. The database is divided into three sets: Sets 1, 2, and 3, which are designated for training (Set #1), validation (Set #2), and testing (Set #3), respectively. The training dataset is used to train the VAE and fine-tune its hyperparameter values. The validation dataset verifies the model's capacity to rebuild unknown data, which is measured by a low reconstruction error. Furthermore, the test dataset is used to map unknown inputs in the latent space and investigate data clustering, hence, improving fault monitoring.

The parameters of each model are individually adapted to suit their specific requirements, such as the input size and the models' hyperparameters: These variables are selected following a sensitivity analysis of the training and validation databases: the Adam algorithm is used with a learning rate of 0.0001, a number of epochs of 1024, and a batch size of 128.

Then, the same desirability function is incorporated into the cost function of both models. The outcomes of this integration are then executed and analyzed, as detailed in Section IV. To further validate the proposed method, a crossvalidation

TABLE 2 Predefined Limits for the Radius R and the Angle θ for Study Case 2

	Healthy	Inner race defect	Outer race defect
Radius R	< 1.5	> 2	> 2
Angle θ	$[0°, 360°]$	$[10°, 80°]$	$[90°, 180°]$

performance test is conducted. During this test, all predefined limits are interchanged between the models, and the results are collected for a comprehensive comparison.

B. SECOND STUDY CASE - NASA BEARING DATASET

1) DATASET

To enhance the validation of the suggested methodology, the publicly available NASA Bearing Dataset [32] is utilized. This setup involved mounting four bearings on a shaft, which was driven at a consistent rotational speed of 2000 RPM by an AC motor connected with rubber belts. A radial force of 6000 pounds was exerted on both the shaft and bearings through a spring system. The bearings were subjected to forced lubrication. Consistent with the practices in [32], accelerometers were installed on the housing of the bearings.

This study focuses on Bearing 2's dataset. The dataset comprises $N_H = 30$ signals, each containing $n_s = 980$ samples, and the sampling rate is set at 20 kHz. Before treating the signals, they undergo verification and cleaning, ensuring the removal of any saturation aberrations. The same preprocessing steps, presented in Fig. 5, have been applied. The dataset is divided into three subsets: Set A, Set B, and Set C. Each subset includes a mix of healthy and faulty signals, with particular attention given to inner and outer race defects.

2) MODEL APPLICABILITY

Set A is utilized for training the model, Set B for validation, and Set C serves to evaluate the model's clustering performance. The model's parameters have been optimized for this particular study: These variables are selected following a sensitivity analysis of the training and validation databases: the Adam algorithm is employed with a learning rate of 0.0001, a number of epochs of 256, and a batch size of 32.

The predefined limits adopted for this study case are presented in Table 2. Subsequently, the desirability function is computed and integrated into the cost function. The results of this integration are then evaluated against those obtained without the integration of the desirability function, as elaborated in Section IV.

IV. RESULTS

A. STUDY CASE - LARGE HYDROGENERATORS

1) TRAINING COST FUNCTION—DESIRABILITY FUNCTION INJECTION

To address the issue where clusters representing the same fault do not occupy the same region in both latent spaces, a desirability term has been introduced into the cost function, as outlined in (2). The computation methodology, as described in Section II-B, yields a desirability score of $\mathcal{D} = 0.689$. Upon its incorporation into (2b), this score will decrease. Given that the objective is to maximize the $\mathcal{D}$ score, the parameter $\beta_{\mathcal{D}}$ has been set to 1.25. Consequently, both machine models have now been trained using (2a).

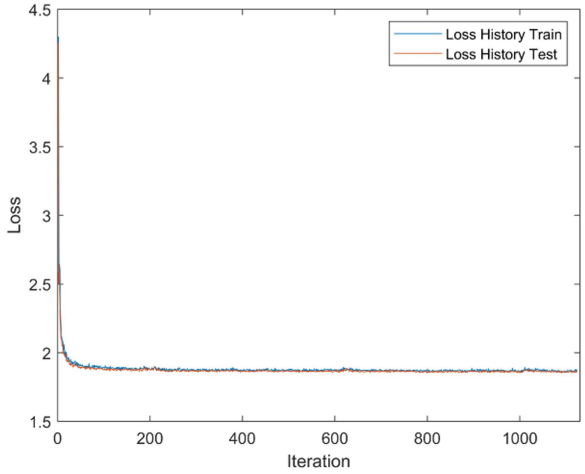

FIGURE 6. Training's cost function with desirability term included.

The cost function obtained after training the model, which includes the desirability term, is presented in Fig. 6. Notably, the curve demonstrates a clear trend of convergence, where the loss steadily decreases over the training iterations. This convergence is a vital indicator of the model's ability to fine-tune its parameters and align its predictions with the desired outcomes. It signifies that the model has reached a stable state, indicating that further training may not yield significant improvements, and the model is well-fitted to the data.

2) LATENT SPACE CLUSTERING

Following the methodology outlined in Section II-A, the latent visualization of both machines in a 2-D space, without incorporating the desirability term, is presented in Fig. 7(b). Distinct clusters are visible, each distinguished by a unique color representing various cases. The healthy cluster is depicted in shades of green, whereas the RITSC is indicated by shades of red, SE by shades of blue, and BDB by shades of magenta.

This visual representation of the clusters underscores the technique's effectiveness in capturing and segregating different cases, even when faults share the same frequency pattern, such as SE and RITSC, which are located in the same direction in the latent space. This observation aligns with findings from the literature, confirming that these two faults have the same frequency pattern in the vibration signal, as documented in [13] and [33]. Moreover, despite sharing the same pattern, these faults exhibit varying amplitudes, which the VAE successfully clusters, as depicted in Fig. 7(b), highlighting the VAE's potential for fault clustering.

This study's findings are consistent with previous research and show a user-friendly diagnostic capability, distinguishing

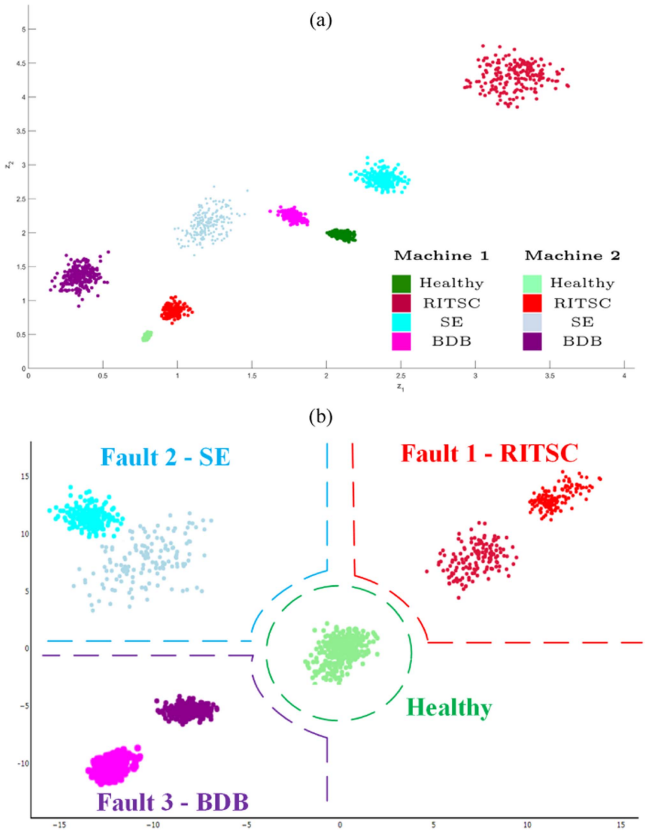

FIGURE 7. Latent space visualization ($p = 2$). (a) Without desirability term. (b) With desirability term.

it from established methods such as the Kernel classifier, support vector machine (SVM), and NN employed in [34], [35], and [36], respectively. These approaches have been applied for IM or SG, or in the context of SPSG, as examined in [37], where several classifiers including K-nearest neighbors, radial basis function SVM, linear SVM, and multilayer perceptron were evaluated.

However, while the latent space displays distinct clusters, a comparison reveals that these clusters do not occupy the same positions or regions. This poses a challenge for robust real-time health monitoring of the machines and prognostics. It indicates that the applied methodology may not be generalized and may not be suitable for all types of large hydrogenerator designs.

Hence, by considering the desirability term and following the steps outlined in Algorithm 1, the resulting latent spaces for both machines are depicted in Fig. 7(b). The following conclusions can be drawn.

1) The healthy clusters of both machines (shades of green) are centered at the origin.
2) The RITSC clusters of both machines (shades of red) are situated in the zone between 0° and 80° at a distance greater than 7, i.e., 10.
3) The SE clusters of both machines (shades of blue) are located in the zone between 90° and 170° at a distance greater than 7, i.e., 8.

TABLE 3 Cross-Validation Test Configurations for R_{lim}

	Configuration 1	Configuration 2	Configuration 3
Healthy	< 7	< 5	< 5
RITSC	> 10	> 7	> 7
SE	> 10	> 7	> 7
BDB	> 10	> 7	> 7

TABLE 4 Cross-Validation Test Configurations for θ_{lim}

	Configuration 1	Configuration 2	Configuration 3
Healthy	$[0°, 360°]$	$[0°, 360°]$	$[0°, 360°]$
RITSC	$[0°, 80°]$	$[180°, 270°]$	$[90°, 180°]$
SE	$[90°, 170°]$	$[0°, 80°]$	$[180°, 270°]$
BDB	$[180°, 270°]$	$[90°, 170°]$	$[0°, 80°]$

4) The BDB clusters of both machines (shades of magenta) are positioned in the zone between 180° and 270° at a distance greater than 7, i.e., 8.5.

These conclusions align with the predetermined constraints outlined in Fig. 3, demonstrating the efficacy of the suggested approach and its capability to cluster faults of the same type in two distinct designs of large hydrogenerators within the same region. Hence, the proposed method holds significant importance in improving the maintenance and operational efficiency of these machines.

3) PERFORMANCE STUDY

To assess the efficacy of the proposed model, a crossvalidation test is carried out. For each scenario, the pre-set R_{limits} and θ_{limits} are interchanged according to the three configurations outlined in Tables 3 and 4, respectively.

The results are presented in Fig. 8(a), (b), and (c), representing the three configurations. Each configuration consistently showcases the model's ability to classify faults in the areas outlined in Tables 3 and 4. This uniform performance across different configurations suggests that the model is dependable and likely to yield satisfactory outcomes.

To further evaluate the model's performance, the model employed on Machine 1 is tested on Machine 2, both with and without the desirability term injected. The two confusion matrices obtained, shown in Fig. 9(a) and (b), is used to evaluate and validate the model's classification performance.

The detection performance is evaluated by computing the F_1 and Acc scores using (4a) and (4b), respectively. TP, TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives, respectively. TP represents the number of defective signals correctly identified as anomalies; TN represents the number of healthy signals misidentified as healthy data; FP represents the number of healthy signals incorrectly classified as anomalies; and FN represents the number of faulty signals misidentified as healthy data. These scores demonstrate the model's ability to

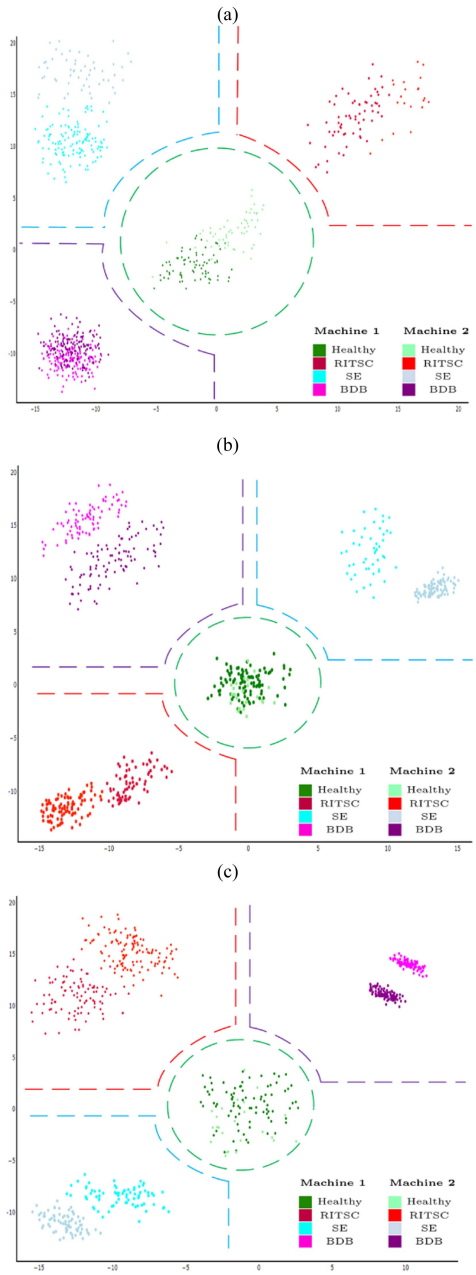

FIGURE 8. Latent space visualization for the different configurations (with $p = 2$). (a) Configuration 1. (b) Configuration 2. (c) Configuration 3.

correctly identify unusual data.

$$F_1 = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})} \tag{4a}$$

$$\text{Acc} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}. \tag{4b}$$

The F_1 and Acc coefficients indicate a result of 24.8% and 43.7%, respectively, for the model that does not use the desirability, whereas a result of 98% and 95% when the desirability is included. This means that the model that includes the desirability term is able to accurately identify a large proportion of

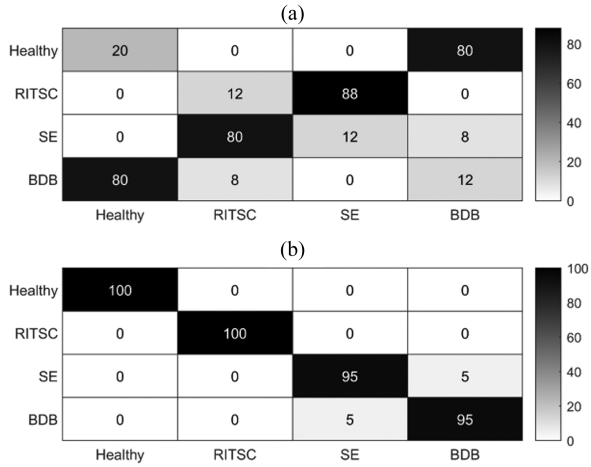

FIGURE 9. Confusion matrix obtained. (a) No desirability included. (b) Desirability included.

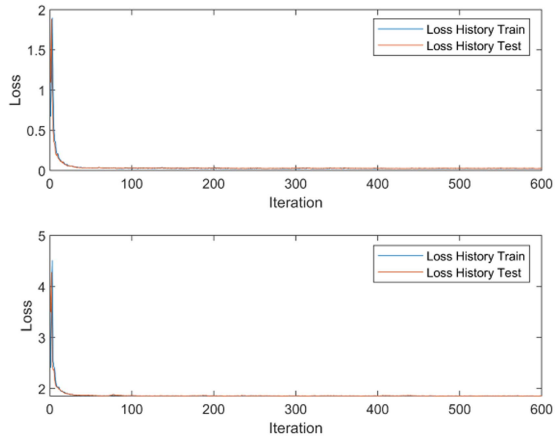

FIGURE 10. Training's cost function. (a) Without desirability term. (b) With desirability term.

TP and has a low rate of FP and FN, which proves the model's capacity to perform the classification.

B. STUDY CASE—NASA BEARING DATASET

1) TRAINING COST FUNCTION

The training and validation cost functions, both with and without the inclusion of the desirability term, are shown in Fig. 10(a) and (b), respectively. The curve distinctly exhibits a converging pattern, where the loss consistently diminishes through successive training iterations. This trend underscores the model's proficiency in refining its parameters and calibrating its predictions to match the targeted results.

2) LATENT SPACE CLUSTERING

Fig. 11(a) and (b) displays the latent space representations of Set C's signals, both with and without the desirability term. These figures feature clearly defined clusters, each distinguished by unique color codes representing different cases. The healthy state is shown in green, whereas the inner and outer race defects conditions are indicated by red and blue,

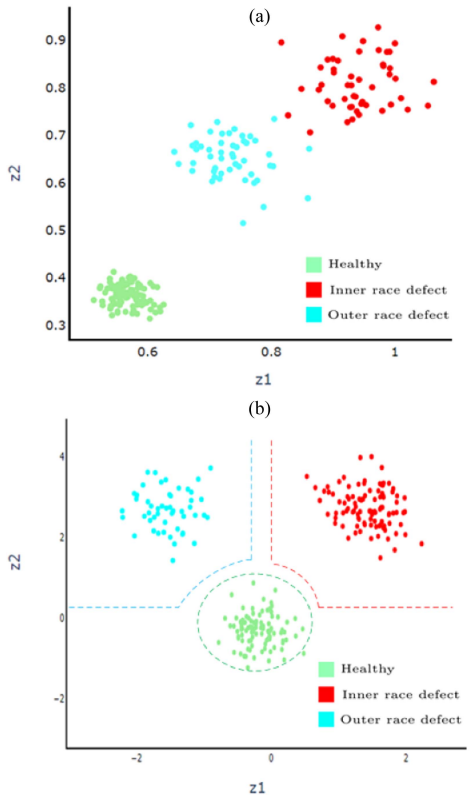

FIGURE 11. Latent space visualization (with $p = 2$). (a) Without desirability term. (b) With desirability term.

respectively. Despite the evident clustering in the latent space, as shown in Fig. 11(a), a comparison with the predefined regions in Fig. 3 reveals a disparity in the positions and areas occupied by these clusters. However, the introduction of the desirability term in Fig. 11(b) demonstrates the effectiveness of this approach, confirming its ability to cluster faults within predetermined regions.

The calculation of the F_1 score and Acc coefficient disclose a 95% value for both, validating, therefore, the efficacy of the proposed approach.

V. CONCLUSION

This study introduces a method for facilitating the detection of faults in EM using the VAE technique and vibration measurements. The challenge with VAE lies in its nondeterministic nature, which results in different clustering of the same health states in various machines. This variability challenges the development of a universally effective diagnostic tool for real-time health monitoring and prognostics.

To overcome this issue, the study suggests adding a Derringer desirability function-based component to the VAE's cost function. This modification markedly enhances the tool's capability to uniformly classify identical fault types in a specific area of the 2-D space, thereby improving usability and standardization. The improved diagnostic tool's effectiveness is demonstrated through two case studies. The first investigates the vibration data from two different types of large hydrogenerators, whereas the second assesses bearing fault data from a publicly accessible dataset. This method demonstrates significant potential in streamlining and unifying fault diagnosis in these scenarios.

REFERENCES

[1] M. Mostafaei and J. Faiz, "An overview of various faults detection methods in synchronous generators," *IET Electric Power Appl.*, vol. 15, no. 4, pp. 391–404, 2021. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/elp2.12031

[2] S. Nandi, H. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors–A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Dec. 2005.

[3] I. Sadeghi, H. Ehya, J. Faiz, and H. Ostovar, "Online fault diagnosis of large electrical machines using vibration signal – A review," in *Proc. Int. Conf. Optim. Elect. Electron. Equip. Int. Aegean Conf. Elect. Machines Power Electron.*, 2017, pp. 470–475.

[4] H. C. Dirani, A. Merkhouf, A.-M. Giroux, B. Kedjar, and K. A.-Haddad, "Impact of real air-gap nonuniformity on the electromagnetic forces of a large hydro-generator," *IEEE Trans. Ind. Electron.*, vol. 65, no. 11, pp. 8464–8475, Nov. 2018.

[5] Y. Merizalde, L. H.-Callejo, O. D.-Pérez, and V. A.-Gómez, "Diagnosis of wind turbine faults using generator current signature analysis: A review," *J. Qual. Maintenance Eng.*, vol. 26, no. 3, pp. 431–458, Oct. 2019. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/JQME-02-2019-0020/full/html

[6] A. Gao, Z. Feng, and L. Ming, "Permanent magnet synchronous generator stator current am-fm model and joint signature analysis for planetary gearbox fault diagnosis," *Mech. Syst. Signal Process.*, vol. 149, 2021, Art. no. 107331. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327020307172

[7] R. Ibrahim et al., "Non-invasive detection of rotor inter-turn short circuit of a hydrogenerator using AI-based variational autoencoder," *IEEE Trans. Ind. Appl.*, vol. 60, no. 1, pp. 28–37, Jan./Feb. 2024.

[8] F. Dao, Y. Zeng, Y. Zou, X. Li, and J. Qian, "Acoustic vibration approach for detecting faults in hydroelectric units: A review," *Energies*, vol. 14, no. 23, 2021, Art. no. 7840. [Online]. Available: https://www.mdpi.com/1996-1073/14/23/7840

[9] H. Bechara, R. Zemouri, B. Kedjar, A. Merkhouf, K. A.-Haddad, and A. Tahan, "Non-invasive detection of rotor inter-turn short circuit in large hydrogenerators by using stray flux measurement combined with convolutional variational autoencoder analysis (CVAE)," *IEEE Trans. Ind. Appl.*, vol. 60, no. 1, pp. 196–205, Jan./Feb. 2024.

[10] T. Paul, "Review of condition monitoring of rotating electrical machines," *IET Electr. Power Appl.*, vol. 2, no. 4, pp. 215–247, 2008.

[11] B. A. Tama, M. Vania, S. Lee, and S. Lim, "Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals," *Artif. Intell. Rev.*, vol. 56, pp. 4667–4709, 2023. [Online]. Available: https://link.springer.com/10.1007/s10462-022-10293-3

[12] B. Rezaeianjouybari and Y. Shang, "Deep learning for prognostics and health management: State of the art, challenges, and opportunities," *Measurement*, vol. 163, 2020, Art. no. 107929. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S026322412030467X

[13] M. Cuevas, R. Romary, J.-P. Lecointe, and T. Jacq, "Non-invasive detection of rotor short-circuit fault in synchronous machines by analysis of stray magnetic field and frame vibrations," *IEEE Trans. Magn.*, vol. 52, no. 7, Jul. 2016, Art. no. 8105304.

[14] G. L. Rødal, "Online condition monitoring of synchronous generators using vibration signal," *Master's thesis*, Norwegian Univ. Sci. Technol., Trondheim, Norway, 2020. [Online]. Available: https://hdl.handle.net/11250/2778212

[15] Y.-L. He et al., "Analysis of stator vibration characteristics in synchronous generators considering inclined static air gap eccentricity," *IEEE Access*, vol. 11, pp. 7794–7807, 2022.

[16] K. Diederik and M. Welling, "Auto-encoding variational bayes," Machine Learning Group, Universiteit van Amsterdam, Amsterdam, The Netherlands, Tech. Rep., 2014.

[17] R. Zemouri, M. Lévesque, N. Amyot, C. Hudon, O. Kokoko, and S. A. Tahan, "Deep convolutional variational autoencoder as a 2D-visualization tool for partial discharge source classification in hydrogenerators," *IEEE Access*, vol. 8, pp. 5438–5454, 2020.
[18] O. Kokoko, C. Hudon, M. Lévesque, N. Amyot, and R. Zemouri, "Comparison of an automatic classification of partial dischage patterns for large hydrogenerator," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2021, pp. 1–6.
[19] R. Ibrahim et al., "Anomaly detection for large hydrogenerators using the variational autoencoder based on vibration signals," in *Proc. Int. Conf. Elect. Machines*, 2022, pp. 1609–1615.
[20] R. Zemouri, R. Ibrahim, and A. Tahan, "Hydrogenerator early fault detection: Sparse dictionary learning jointly with the variational autoencoder," *Eng. Appl. Artif. Intell.*, vol. 120, 2023, Art. no. 105859. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095219762300043X
[21] G. S. Martin, E. L. Droguett, V. Meruane, and M. d. C. Moura, "Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis," *Struct. Health Monit.*, vol. 18, no. 4, pp. 1092–1128, 2019, doi: 10.1177/1475921718788299.
[22] A. Proteau, R. Zemouri, A. Tahan, and M. Thomas, "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.*, vol. 111, no. 11, pp. 3597–3611, Dec. 2020, doi: 10.1007/s00170-020-06338-y.
[23] M. Hemmer, A. Klausen, H. V. Khang, K. G. Robbersmyr, and T. I. Waag, "Health indicator for low-speed axial bearings using variational autoencoders," *IEEE Access*, vol. 8, pp. 35 842–35 852, 2020.
[24] Y. Xiao, H. Shao, J. Wang, S. Yan, and B. Liu, "Bayesian variational transformer: A generalizable model for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 207, 2024, Art. no. 110936. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327023008440
[25] N. R. Costa, J. Lourenço, and Z. L. Pereira, "Desirability function approach: A review and performance evaluation in adverse conditions," *Chemometrics Intell. Lab. Syst.*, vol. 107, no. 2, pp. 234–244, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169743911000797
[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.
[27] A. Proteau, A. Tahan, R. Zemouri, and M. Thomas, "Predicting the quality of a machined workpiece with a variational autoencoder approach," *J. Intell. Manuf.*, vol. 34, pp. 719–737, Aug. 2023, doi: 10.1007/s10845-021-01822-y.
[28] Y. Mao, Y. Gao, Y. Gao, W. Cheng, and Y. Wang, "Simplified desirability level metrics for estimation performance evaluation," in *Proc. 21st Int. Conf. Inf. Fusion*, 2018, pp. 877–882.
[29] K. Sonam, S. Sutavani, S. R. Wagh, and N. M. Singh, "Optimal control of probabilistic boolean networks: An information-theoretic approach," *IEEE Access*, vol. 9, pp. 157 068–157 082, 2021.
[30] M. Zheng, J. Yu, H. Teng, Y. Cui, and Y. Wang, *Extension of Probability-Based Multi-Objective Optimization in Condition of the Utility With Desirable Value*. Berlin, Germany: Springer, Aug. 2023, pp. 61–69.
[31] G. Derringer and R. Suich, "Simultaneous optimization of several response variables," *J. Qual. Technol.*, vol. 12, pp. 214–219, 1980.
[32] NASA bearing dataset, 2006. [Online]. Available: www.kaggle.com/datasets/vinayak123tyagi/bearing-dataset
[33] R. Ibrahim, R. Zemouri, B. Kedjar, A. Tahan, A. Merkhouf, and K. A.-Haddad, "Rotor fault diagnosis in a hydrogenerator based on the stator vibration and the variational autoencoder," in *Proc. IEEE Int. Electric Machines Drives Conf.*, 2023, pp. 1–7.
[34] L. Wang, Y. Li, and J. Li, "Diagnosis of inter-turn short circuit of synchronous generator rotor winding based on volterra kernel identification," *Energies*, vol. 11, no. 10, Sep. 2018, Art. no. 2524. [Online]. Available: http://www.mdpi.com/1996-1073/11/10/2524
[35] M. E.-Saadawi and A. Hatata, "A novel protection scheme for synchronous generator stator windings based on SVM," *Protection Control Modern Power Syst.*, vol. 2, no. 3, pp. 1–12, Jul. 2017.
[36] M. Doseděl, L. Kopečný, M. Kozovský, J. Hnidka, and Z. Havránek, "Detection of the interturn shorts of a three-phase motor using artificial intelligence processing vibration data," in *Proc. 20th Int. Conf. Mechatronics - Mechatronika*, 2022, pp. 1–5.
[37] H. Ehya, T. N. Skreien, and A. Nysveen, "Intelligent data-driven diagnosis of incipient interturn short circuit fault in field winding of salient pole synchronous generators," *IEEE Trans. Ind. Inform.*, vol. 18, no. 5, pp. 3286–3294, May 2022.

RONY IBRAHIM was born in Lebanon, in 1996. He received the B.Sc. degree in electrical engineering and the M.Sc.A. degree in renewable energies from Saint-Joseph University in Beirut, Beirut, Lebanon, in 2019 and 2020, respectively. He is currently working toward the Ph.D. degree with École de Technologie Supérieure (ÉTS), Montreal, QC, Canada.

RYAD ZEMOURI was born in Algiers, Algeria, in 1974. He received the Engineering degree in electronics and control science from the University Mouloud Mammeri of Tizi-Ouzou, Tizi Ouzou, Algeria, in 1998, and the postgraduate Diploma, and the Ph.D. degree in the same domain from the Franche-Comté University of Besançon, Besançon, France, in 2000 and 2003, respectively.

In 2003, he was an Associate Professor with the Conservatoire National des Arts et Métiers, Paris, France. In 2022, he joined the Research Center of Hydro Quebec (CRHQ) where he is currently a Researcher. His research interests include machine learning for safety-critical prognosis, diagnosis, and health management.

ANTOINE TAHAN was born in Syria, in 1966. He received the B.Sc.A., M.Sc.A. degrees in mechanical engineering and Ph.D. degree in electrical engineering from the University of Laval, Quebec City, QC, Canada, in 1991, 1992, and 1998, respectively.

Since 2004, he has been a Full Professor with the Department of Mechanical Engineering, École de Technologie Supérieure (ÉTS), Montreal, QC, Canada, and an active member of the engineering laboratory for products, processes, and systems (LIPPS). He is specialized in dimensional metrology and geometrical characterization, tolerance management, and propagation of uncertainties. He is working on the propagation of uncertainties in engineering models, diagnostic and prognostic tools for complex electromechanical systems. He contributes to the development of probabilistic approaches to detect the state of degradation and damage to improve reliability and design. His research interests include the aerospace and energy industries.

BACHIR KEDJAR (Member, IEEE) received the B.Sc. degree in electrotechnical from the Ecole Nationale Polytechnique, Algiers, Algeria, in 1984, the M.Sc. degree in electrical engineering from the Universite Pierre et Marie Curie, Paris, France, in 1988, and the Ph.D. degree in power electronics from the Universite des Sciences et Techniques de Lille Flandres-Artois, Lille, France, in 1988.

He is currently with the Electrical Engineering Department, Ecole de Technologie Supérieure, Montréal, QC, Canada, as a Research Assistant and a Lecturer. His research interests include power electronics converters, power quality, and electric machines including their modeling, simulation, and control

AREZKI MERKHOUF (Senior Member IEEE) received the B.Sc. Eng. degree from the University of Algeria, Algiers, Algeria, in 1990, the M.Sc. degree from the University of Pierre and Marie Curie, Paris, France, in 1992, and the Ph.D. degree from the University of Sherbrooke, Sherbrooke, QC, Canada, in 1999, all in electrical engineering.

In 1999, he joined GE Energy Power Systems, where he was a member with the R&D team until 2008. In 2008, he joined the Research Center of Hydro Quebec (CRHQ), where he is now a Senior Researcher and a Project Leader on the diagnosis of large electrical machines. He has authored more than 65 scientific papers. His research interests include generator and motor design and optimization, generator acoustics and vibrations, analytical and numerical simulations of large rotating electrical machines, and numerical and analytical computation of electromagnetic fields.

KAMAL AL-HADDAD (Life Fellow, IEEE) received the B.Sc.A. and M.Sc.A. degrees in electrical engineering from the University of Québec à Trois-Rivières, Trois-Rivières, QC, Canada, in 1982 and 1984, respectively, and the Ph.D. degree in electrical engineering from the Institute National Polytechnique, Toulouse, France, in 1988.

Since 1990, he has been a Professor with the Electrical Engineering Department, École de Technologie Supérieure (ÉTS), Montreal, QC, Canada, where he has been the holder of the senior Canada Research Chair in Electric Energy Conversion and Power Electronics since 2002. He is a Consultant and has established very solid link with many Canadian and international industries working in the field of power electronics, electric transportation, aeronautics, and telecommunications. He successfully transferred and implemented dozens of technologies to Canadian and international companies. His research interests include highly efficient static power converters, harmonics and reactive power control using hybrid filters, voltage level multiplier, resonant and multilevel converters, including the modeling, control, and development of prototypes for various industrial applications in electric traction, renewable energy, power supplies for drives, telecommunication, etc.

Dr. Al-Haddad was the recipient of the 2014 IEEE IES Dr.-Ing. Eugene Mittelmann Achievement Award and the IEEE Medal in Power Engineering. He is a Member of the Academy of Sciences and Fellow of the Royal Society of Canada and a Fellow Member of the Canadian Academy of Engineering. He is IEEE IES President 2016 and 2017, an Associate editor of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, and IES Distinguished Lecturer. He has been elected the IEEE 2022 division VI Director.