Journal of Intelligent Manufacturing manuscript No.
(will be inserted by the editor)

Hybrid data augmentation method for combined failure recognition in rotating machines

Dionísio H. C. S. S. Martins* · Amaro A. de Lima · Milena F. Pinto · Douglas de O. Hemerly · Thiago de M. Prego · Fabrício L. e Silva · Luís Tarrataca · Ulisses A. Monteiro · Ricardo H. R. Gutiérrez · Diego B. Haddad

Received: date / Accepted: date

Dionísio H. C. S. S. Martins
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: dionisiohmartins@gmail.com
ORCID: 0000-0001-7136-1119

Amaro A. de Lima
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: amaro.lima@cefet-rj.br
ORCID: 0000-0001-5397-6531

Milena F. Pinto
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: milena.pinto@cefet-rj.br
ORCID: 0000-0001-6916-700X

Douglas de O. Hemerly
International Business Machines Corporation, Rio de Janeiro, Brazil.
email: doh4rj@hotmail.com
ORCID: 0000-0003-2243-3184

Thiago de M. Prego
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: thiago.prego@cefet-rj.br
ORCID: 0000-0003-1404-4349

Fabrício L. e Silva
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: fabricio.silva@cefet-rj.br
ORCID: 0000-0002-8220-8344

Luís Tarrataca
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: luis.tarrataca@cefet-rj.br
ORCID: 0000-0001-9359-5143

Ulisses A. Monteiro
Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.
email: ulisses@oceanica.ufrj.br
ORCID: 0000-0001-8530-5775

Ricardo H. R. Gutiérrez

Abstract Rotating machines are frequently subject to a wide range of rough conditions, resulting in mechanical failures and performance degradation. Thus, it is important to apply proper failure detection and recognition techniques, such as machine learning algorithms, to prevent these issues early. In industrial environments, little data exists regarding failure conditions, which hinders the training stage of the classification algorithms responsible for classifying the failures. Therefore, this work proposes a hybrid method of data augmentation to increase the number of minority class instances in order to improve classifier performance. The approach combines the synthetic minority over-sampling and the additive white Gaussian noise techniques to create a set of artificial signals. The results show that the proposal is able to achieve better results than applying those techniques separately and also when using an undersampling strategy. For comparison purposes, four machine learning classification methods were analyzed alongside our data augmentation proposal, namely, support vector machines, K-nearest neighbors, random forest and stacked sparse autoencoder. The proposed hybrid data augmentation method associated with stacked sparse autoencoder outperformed the other models obtaining an accuracy of 100% and a processing time of 0.13 s.

Keywords Data augmentation · Combined failures recognition · Imbalance · Misalignment · Rotating Machines · Predictive maintenance.

1 Introduction

Rotating machines are widely employed in modern industry. However, they are frequently subject to a wide range of conditions, such as frequent load changes and high speeds (Qian et al., 2019) that result in performance degradation and mechanical failures (Li et al., 2020b). Consequently, a key industry issue is to provide system effectiveness and reliability through accurate fault diagnosis (Yu et al., 2019). These allow for unexpected failures and unscheduled downtime to be minimized, saving unnecessary extra costs.

Applications involving fast and intelligent fault diagnosis methods are of significant interest, as can be seen in the works (Li et al., 2019; Wang et al., 2020; Martins et al., 2021). A variety of sensors have also been employed to measure dynamic responses (Goyal et al., 2019). A possible non-invasive solution to effectively measure the different levels of degradation is through vibration signal estimation. Note that failure recognition and detection from mechanical vibration analysis enables proper maintenance measures at early

Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.
email: rhramirez@oceanica.ufrj.br
ORCID: 0000-0003-4768-3243

Diego B. Haddad
Federal Center for Technological Education of Rio de Janeiro, Rio de Janeiro, Brazil.
email: diego.haddad@cefet-rj.br
ORCID: 0000-0002-7634-5481

stages (Glowacz, 2018). The most frequent failures that affect the useful life of rotating machines are imbalance and misalignment (Bai et al., 2019; Guan et al., 2017).

Misalignment is usually due to improper installation, thermal variation, asymmetric loads, amongst others (Hujare and Karnik, 2018). These result in increased loads on bearings and couplings, the parts connected to the shaft. Misalignment usually worsens with continuous operation and requires periodical monitoring in order to be corrected (Verma et al., 2014). One possible strategy for determining misalignment is to employ vibration spectrum analysis. This is a reliable method that also enables the identification of imbalance faults. Various methodologies have been applied in the literature addressing this issue, such as (Klausen et al., 2018; Djagarov et al., 2019). For instance, (Yamamoto et al., 2016) proposed using an intelligent algorithm embedded in a Field Programmable Gate Array (FPGA) to correct imbalance faults. The work (Djagarov et al., 2019) designed a Supervisory Control and Data Acquisition (SCADA) system for monitoring electric motor failures in ships.

Other references, such as (William and Hoffman, 2011; Yu, 2019), successfully applied signal processing methods to fault detection. Recently, many authors, such as (Srinivas et al., 2019; Dekhane et al., 2020), have addressed the problem of measuring, identifying, and quantifying combined faults in rotating machines. Machine learning and statistical techniques also exist for tackling these issues, namely (Yang et al., 2019; Zhang et al., 2020a).

A review on data-driven fault severity assessment in rolling bearings was presented in (Cerrada, 2018). The work mentions a series of techniques that can be employed to assess the state of an electric engine based on digital signal processing and intelligent algorithms, namely: artificial neural networks, support vector machines, clustering, Markov models, fuzzy logic, linear discriminant analysis, Gaussian mixture models and probabilistic based approaches. One possible way for developing prognostic systems is to consider the remaining useful life of an asset, which can be estimated by fault classification techniques (Si et al., 2011). Fault classification can be divided into model-based approaches (Srinivas et al., 2019; Wang and Jiang, 2018) and data-driven methods (Dekhane et al., 2020; Li et al., 2017), the latter being the focus of this paper. Typically, statistical data-driven approaches for fault classification apply stages such as *(i)* data acquisition; *(ii)* feature extraction; *(iii)* fault identification; and *(iv)* fault severity estimation (Martins et al., 2019). Machine learning techniques are susceptible to suffer from overfitting issues. This is especially true in the case of learning from rare events (Oh and Jeong, 2020). Data augmentation schemes can be employed to reduce this issue (Li et al., 2020c).

In (Jin et al., 2021) a technique is presented based on deep learning to identify vibration signals composed of simple and combined failures related to bearing faults. The dataset used in this paper is composed of eight classes composed of three simple failures, four combined failures and one class corresponding to normal operating conditions. The algorithm employs active learn-

ing in order to overcome a lack of labeled instances. The article also proposes an automatic way of extracting features to reduce the intervention of a specialist in the initial choice of the feature set. The authors also apply a feature selection technique to choose the most relevant ones and thus reduce the number of input signals in the classifier. The algorithm achieved 100% accuracy, outperforming convolutional neural networks and long short-term memory algorithms.

In (Xiao et al., 2021) a system was designed based on deep learning using a denoising autoencoder to solve the problem of noisy domain shift in failure identification. This work made use of two datasets consisting of acoustic signals, one referring to gear faults and the other to motor faults. Noisy data was generated through additive white Gaussian noise (AWGN) and binary masking. Classification-wise, the proposed algorithm performed well even in the face of contaminated signals with high noise levels. The training time of the proposed algorithm was also lower when compared to other deep learning algorithms.

In (Shao et al., 2017) the authors propose an Auxiliary Classifier Generative Adversarial Network (ACGAN) to create new and realistic synthetic observations directly from sensor data. The method is applied for fault detection and classification in rotating machines. The authors made use of a rotor kit with one accelerometer for data gathering. Six conditions were simulated: normal, stator winding defect, imbalanced rotor, bearing defect, broken bar, and bowed rotor. The minor class had 100 samples, while the rest of the classes had 200 instances. Different training data settings using real data and generated data were used to produce 12 different scenarios. The baseline scenario employed 200 samples of real data alongside zero instances of generated data and achieved an accuracy of 99.80%. When 200 samples were used from real data in conjunction with 200 generated samples the system produced 99.93% accuracy. Classification accuracy reached 100% when 200 real samples were used alongside 600 generated ones.

In the work of (Rashid and Louis, 2019), AWGN was used to augment the positioning and movement data which were collected from GPS and gyroscope devices. The sensors were installed in heavy-duty vehicles to evaluate the optimal usage of civil construction equipment through deep learning methods. The goal of the authors was to reduce costs in civil constructions. The idea of creating a new dataset using data augmentation techniques can also be found in (Rochac et al., 2019). The authors applied AWGN to develop several new training data from an original limited set consisting of infrared camera images and further train different deep learning models. The authors gave special attention to the signal-to-noise ratio (SNR), experimenting with ten different SNR values to demonstrate the respective influence on accuracy. These results were then compared to those obtained using Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). In the latter, the authors performed experiments using SMOTE to enlarge the minority class after having undersampled the majority classes in order to analyze performance in

the ROC space. The experiments were performed with three different classifier algorithms.

In (Arslan et al., 2019), a dataset of humidity, temperature, light intensity, and air quality was preprocessed through AWGN and SMOTE data augmentation techniques and further used to train a classifier algorithm. The results suggested a better accuracy when using SMOTE than AWGN for this configuration. The work (Fernández et al., 2018) presents a literature review and approaches some of the relevant aspects of the SMOTE technique. In (Wang, 2008) the authors successfully increased classification accuracy by combining SMOTE and a Biased-SVM when applied to four other imbalanced datasets available at the UC Irvine (UCI) machine learning repository. The results suggested that classifier sensitivity to minority classes was improved by the SMOTE algorithm. It is also possible to create variations of the SMOTE technique as proposed by (Li et al., 2011). Instead of selecting the K-nearest neighbors (K-NN), the authors selected three real random samples to create a triangle. The triangle is then filled with a defined quantity of lines, and each of these lines will finally contain a defined synthetic amount of data points. This process was entitled Random-SMOTE, whose objective was to pursue a more uniform distribution of synthetic items throughout the minority class space. In (Ali et al., 2019) the influence on the model accuracy was analyzed after SMOTE was applied to enlarge the minority class of a vibration dataset. The results were comparable to the previous works, which used AWGN as the augmentation approach. The authors used a multilayer perceptron (MLP) to classify the rotating machine faults.

Variational Autoencoder is an additional data augmentation technique based on deep learning. The method allows for the reconstruction of the created examples in the data space. However, the approach is known for producing distorted reconstruction when the signal is noisy (Burks et al., 2019). The method is also difficult to train due to the required hyperparameter tuning process and the high execution computational cost (Asadi et al., 2009; Shorten and Khoshgoftaar, 2019), which requires the use of clusters and/or GPUs.

Generative artificial neural network is another deep learning method that has been used for data augmentation in several areas. However, the use of the technique has some limitations, namely: *(i)* it requires a large amount of original data to carry out training (Yu et al., 2021), which is not always possible, as is the case of this research; *(ii)* it is subject to instability and non-convergence of the algorithm in cases where the generator produces large outputs; and *(iii)* it generates examples that are not consistent with the physical nature of the real data (Shorten and Khoshgoftaar, 2019; Mikołajczyk and Grochowski, 2018).

As mentioned in the previous paragraphs, the performance of deep learning techniques is susceptible to suffer from a lack of training examples in what concerns failure conditions. Therefore, it is pertinent to propose a data augmentation method for those classes whose instances are lacking, which is: *(i)* stable when using parameter adjustment methodology; and *(ii)* does not re-

quire high-performance computational resources. The main contributions of this paper are summarized below:

1. Most of the research focusing on fault diagnosis in rotating machines only considers the identification of single faults. However, in this work, the objective is to identify and differentiate single failures from combined failures. These are situations that can occur in industrial environments. Furthermore, this task is more complex than the identification of isolated faults.
2. Compute the influence in classifier performance of preprocessing approaches such as features normalization, undersampling, and data augmentation using white noise and SMOTE.
3. Develop a novel hybrid data augmentation method using SMOTE and AWGN to increase the number of minority classes instances with the objective of improving classifier performance.

This paper is structured as follows. Section 2 presents a description of the proposed methodology, detailing the dataset as well as the feature extraction process. A theoretical foundation regarding the main concepts treated in this research is briefly explained in Section 3. Section 4 describes the effectiveness of the proposed method. The concluding remarks are reported in Section 5.

2 Case Study

Industrial rotating machines are usually involved in production processes. Production stoppage might cause significant financial losses and even damage the equipment. This makes it unfeasible to cause failures in these apparatuses for study purposes. An adequate study of the problems affecting this type of machines requires a large dataset covering different types and severities of breakdowns. Creating such a dataset can be very time consuming and even impossible for the most critical operating conditions.

In this sense, two approaches can be taken, namely: *(i)* place the rotating machine on a test bench for the purpose of inserting faults and recording the corresponding vibration signatures; and *(ii)* employ bench simulators of rotating machines. The former is impractical given the potentially high cost of the machine and elevated execution time associated with preparing and assembling the failures. As a consequence, laboratory tests are more expensive. The second approach enables the insertion of failures in a more convenient way, which results in time and execution savings (Villa et al., 2012).

As a result, the experimental bench Alignment Balance Vibration Trainer (ABVT) was employed in this study to produce simple and combined faults. This experimental bench is composed of a 0.25 hp DC motor, two rolling bearings, a thin shaft, a sliding surface, a rigid coupling, and an inertia disc positioned in the center hug configuration (between the rolling bearings), as shown in Figure 1. The simulation bench was used in an environment with a controlled temperature in the range of 22 °C to 27 °C. Before starting to record and monitor the signals, the engine was in operation for 10 minutes to ensure

that it was properly prepared. Signals that presented vibration values outside the expected range were discarded and replaced with a new recording. The module used to record the vibration and tachometer signals was the signal acquisition module (NI 9234), manufactured by National Instruments. This module converts the analog signals from the sensors into digital voltage or current signals. The main features of the module sensor are 24-bit resolution, a maximum sampling frequency of 51.2 kHz, 102 dB dynamic range, anti-aliasing filter, operating temperature range of [-40, 70] °C, and signal conditioning for piezoelectric sensors. The Labview™ software was used to implement the interface between the acquisition module and the computer. This interface enables viewing the signals of each channel during the acquisition step to avoid recording errors.

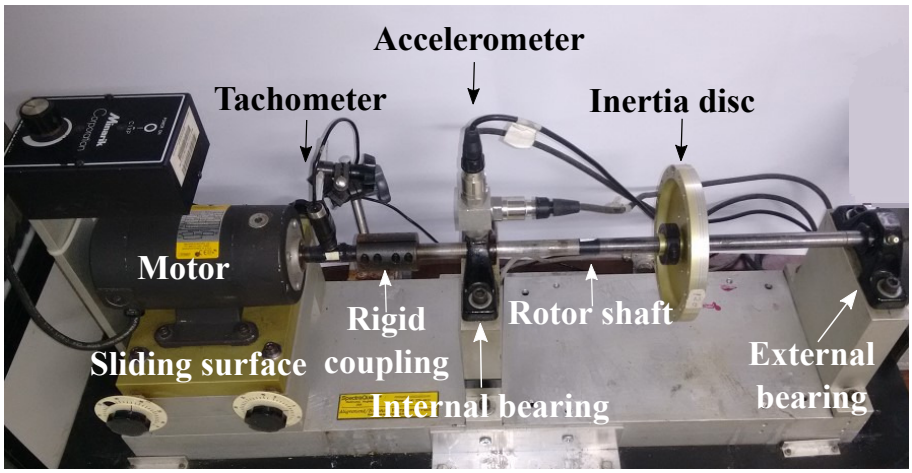

Fig. 1 ABVT Experimental bench.

The scenarios studied in this research are: *(i)* normal behavior; *(ii)* imbalanced rotor; *(iii)* imbalance rotor with added horizontal misalignment; and *(iv)* imbalance rotor with added vertical misalignment. Imbalance is provoked in the ABVT by fixing screws on the inertia disc. Vertical misalignment is produced by adding metal plates at the base of the DC motor. Horizontal misalignment is inserted by shifting the base of the motor and measuring rotational speed using a digital tachometer, as shown in Figure 2.

The vibration signals were acquired and stored. Because the acceleration signals are quite noisy, which can negatively affect the fault diagnosis stage, they were filtered by a bandpass Hamming window whose cutoff frequencies are 10 Hz and 1000 Hz. Subsequently, the discriminative characteristics of the signals were extracted as a means of reducing the amount of input information to be presented to the classifiers. The last step was to compare the classification performance behavior of four algorithms. This allowed us to better grasp of

(a) Insertion of imbalance.

(b) Insertion of horizontal misalignment.

(c) Insertion of vertical misalignment

Fig. 2 Faults insertion.

the effectiveness of the proposed hybrid data augmentation method against AWGN and SMOTE.

2.1 Dataset

Table 1 presents the details of the dataset produced, which consists of 238 signals. These were recorded by changing the motor rotational speed using 2 Hz steps in the range $f \in [16, 60]$ Hz. The maximum frequency employed is due to the operating limits of the simulation bench. The imbalance values listed in Table 1 indicate the masses (in grams (g)) that were placed on the inertia disc. The horizontal and vertical misalignments measures are in millimeters (mm) and correspond to the movement of the motor base when compared to

its initial position. The 'Label' column indicates the class that describes each scenario.

The vibration signals were measured at the internal bearing, which is closer to the DC motor. Digital data was acquired at the sampling frequency of 50 kHz for 3 seconds. Three uniaxial piezoelectric accelerometers, manufactured by IMI Sensors, were employed to obtain vibration signals in perpendicular directions: axial, horizontal, and vertical. The main characteristics of this sensor are: sensitivity (100 mV/g (20%)); frequency range ([0.27, 1000] Hz); and acceleration measurement range ([-50, 50] g, in this case g is approximately 9.8 m/s^2). In order to measure the rotational speed of the shaft motor, the tachometer MT-190 was used, which is produced by Monarch Instrument.

Table 1 Dataset description.

Scenarios	**Signals Details**		
	Quantity	**Speed (Hz)**	**Label**
Normal	115	[16, 60]	No
Imbalance (6 g)	23	[16, 60]	I
Imbalance (20 g)	18	[16, 50]	I
Imbalance (6 g) + hor. misalignment (1 mm)	23	[16, 60]	IHM
Imbalance (20 g) + hor. misalignment (1 mm)	18	[16, 50]	IHM
Imbalance (6 g) + ver. misalignment (1.27 mm)	23	[16, 60]	IVM
Imbalance (20 g) + ver. misalignment (1.27 mm)	18	[16, 50]	IVM

2.2 Feature extraction

One of the main preprocessing steps in fault diagnosis is feature extraction of the vibration signals (Razavi-Far et al., 2017; Xu et al., 2019). The fault signature can be understood as a set of symptoms associated with a defect, and these are directly related to certain features from the vibration signals (Cerrada, 2018). Feature extraction also reduces the amount of information to be used as input to the classifier. For contextualization, in this research, if this preprocessing step were not to be used, the classifier would receive 150,000 samples referring to each of the sensors used. This would unnecessarily increase the computational cost of the classification task and impair its accuracy due to the excess of information (Bramer, 2007). In this work, features are used in time and frequency domains (Pandya et al., 2013; Dhamande and Chaudhari, 2018), as shown in Table 2 where:

- $x(n)$ is the time domain vibration signal;
- N is the length of the time domain vibration signal;

- $\mathbb{E}$ denotes the expected value operator;
- $p(z_n)$ corresponds to the probability of $x(n)$ being equal to the possible values of sequence z_n;
- $s(k)$ is the vibration signal spectrum obtained by the application of Fast Fourier transform (FFT) in $x(n)$;
- K is the number of samples of $s(k)$;
- $p(z_k)$ corresponds to the probability of $x(k)$ being equal to the possible values of sequence z_k;
- R_f is the rotational speed frequency obtained by the FFT of the tachometer;
- $A_m(R_f)$ denotes the maximum value of $s(k)$ at the R_f of the rotating machine;
- N/A stands for not applicable;

with the exception of the R_f indicator, which represents only a single feature, each one of the remaining indicators in Table 2 is calculated for the axial, horizontal, and vertical directions. This results in 48 time-domain and 60 frequency-domain features, thus producing a feature vector with 109 elements.

2.3 Features normalization

In statistical studies, normalization is used to standardize data and to optimize data processing (Suarez-Alvarez et al., 2012). In machine learning, normalization plays a significant role when attributes can hinder data processing (*e.g.,* redundant or extreme values). Normalization is a way to standardize and minimize problems that originate from such dispersions or redundancies. The process allows for (Walpole and Myers, 2012): *(i)* effective data processing; and *(ii)* ignoring inconsistent data. Normalization can improve the performance of classifiers such as SVM, K-NN, and RF (Canbaz and Polat, 2019; Sikder et al., 2019).

Preliminary simulations in the dataset employed in this work show that Minimum-Maximum (min-max) normalization performs better than Z normalization. Thus, in the simulations, only the min-max normalization was applied. This technique, respectively presented in Equation (1), normalizes the values through their minimum and maximum values, separating them at fixed intervals to provide more effective processing (Polat, 2020).

Table 2 Extracted features at time and frequency domains, with $\alpha_K \triangleq \frac{\sqrt{K(K-1))}}{K-2}$.

Feature	Time	Frequency
Maximum	$T_1 = \max(x(n))$	$F_1 = \max(s(k))$
Minimum	$T_2 = \min(x(n))$	$F_2 = \min(s(k))$
Mean value	$T_3 = \sum_{n=1}^{N} \frac{x(n)}{N}$	$F_3 = \sum_{k=1}^{k} \frac{s(k)}{K}$
Variance	$T_4 = \sum_{n=1}^{N} \frac{(x(n)-T_3)^2}{N-1}$	$F_4 = \sum_{k=1}^{K} \frac{(s(k)-F_3)^2}{K-1}$
Standard deviation	$T_5 = \sqrt{T_4}$	$F_5 = \sqrt{F_4}$
Root mean square	$T_6 = \sqrt{\sum_{n=1}^{N} \frac{(x(n))^2}{N}}$	$F_6 = \sqrt{\sum_{k=1}^{K} \frac{(s(k))^2}{K}}$
Kurtuosis	$T_7 = \frac{\mathbb{E}[x(n)-T_3]}{(T_5)^4}$	$F_7 = \frac{\mathbb{E}[s(k)-F_3]}{(F_5)^4}$
Unbiased skewness	$T_8 = \frac{\sqrt{N(N-1))}}{N-2} \frac{\mathbb{E}\left[(x(n)-T_3)^3\right]}{\mathbb{E}\left[(x(n)-T_3)^2\right]^{3/2}}$	$F_8 = \alpha_K \frac{\mathbb{E}\left[(s(k)-F_3)^3\right]}{\mathbb{E}\left[(s(k)-F_3)^2\right]^{3/2}}$
Signal energy	$T_9 = \sum_{n=1}^{N} \frac{\lvert x(n)\rvert^2}{N}$	$F_9 = \sum_{k=1}^{K} \frac{\lvert s(k)\rvert^2}{K}$
Amplitude range	$T_{10} = T_1 - T_2$	$F_{10} = F_1 - F_2$
Entropy	$T_{11} = -\sum_{n=1}^{N} p(z_n) \log_2 p(z_n)$	$F_{11} = -\sum_{k=1}^{K} p(z_k) \log_2 p(z_k)$
Crest factor	$T_{12} = \frac{\max\lvert x(n)\rvert}{T_6}$	$F_{12} = \frac{\max\lvert s(k)\rvert}{F_6}$
Shape factor	$T_{13} = \frac{T_6}{\sum_{n=1}^{N} \frac{\lvert x(n)\rvert}{N}}$	$F_{13} = \frac{F_6}{\sum_{k=1}^{K} \frac{\lvert s(k)\rvert}{K}}$
Impulse factor	$T_{14} = \frac{T_1}{\sum_{n=1}^{N} \frac{\lvert x(n)\rvert}{N}}$	$F_{14} = \frac{F_1}{\sum_{k=1}^{K} \frac{\lvert s(k)\rvert}{K}}$
Margin factor	$T_{15} = \frac{\max\lvert x(n)\rvert}{\left(\sum_{n=1}^{N} \frac{\sqrt{\lvert x(n)\rvert}}{N}\right)^2}$	$F_{15} = \frac{\max\lvert s(k)\rvert}{\left(\sum_{k=1}^{K} \frac{\sqrt{\lvert s(k)\rvert}}{K}\right)^2}$
Defect factor	$T_{16} = T_1 - T_6$	$F_{16} = F_1 - F_6$
Rotational frequency	N/A	$F_{17} = R_f$
Amplitude at R_f	N/A	$F_{18} = A_m(R_f)$
Amplitude at $2R_f$	N/A	$F_{19} = A_m(2R_f)$
Amplitude at $3R_f$	N/A	$F_{20} = A_m(3R_f)$
Amplitude at $4R_f$	N/A	$F_{21} = A_m(4R_f)$

$$\mathbf{fe}_{\text{norm}} = \frac{\mathbf{fe} - \min(\mathbf{fe})}{\max(\mathbf{fe}) - \min(\mathbf{fe})}, \tag{1}$$

where $\mathbf{fe}$ is the original feature vector, $\min(\mathbf{fe})$ is the lowest value of vector $\mathbf{fe}$, $\max(\mathbf{fe})$ is the highest value of $\mathbf{fe}$ and $\mathbf{fe}_{\text{norm}}$ is the normalized $\mathbf{fe}$ vector.

3 Theoretical foundations

This section presents the theoretical background for the development of the hybrid approach, namely: Section 3.1 presents an explanation of rotating

systems and a respective dynamic model; Section 3.2 describes the imbalance whilst Section 3.3 details the misalignment effects; Section 3.4 presents the data augmentation methodology and Section 3.5 elaborates on the classification methods employed.

3.1 Mechanical Model of rotating machines

In general, a rotor-coupling-bearing system is represented by a second-order differential equation as described by (Desouki et al., 2020):

$$\mathbf{M}\ddot{\mathbf{q}} + \mathbf{C}\dot{\mathbf{q}} + \mathbf{K}\mathbf{q} = \mathbf{f}(t), \tag{2}$$

where $\mathbf{M}$ is the mass matrix, $\mathbf{C}$ is the damping matrix, and $\mathbf{K}$ is the stiffness matrix. The vector of generalized coordinates is given by $\mathbf{q}$, with its first and second derivatives with respect to time t given by $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$, respectively. While, the external forces are represented by the vector $\mathbf{f}(t)$.

Imbalance and misalignment are the main sources of vibration in rotating machinery. The vibration caused by these phenomena may destroy critical parts of the machine, depending on its amplitude. Considering those phenomena responsible for the excitation forces perceived in the coupling of the driver and driven shafts, the vector of external forces is given by (Desouki et al., 2020):

$$\mathbf{f}(t) = \mathbf{f}_{\text{imb}}(t) + \mathbf{f}_{\text{mis}}(t), \tag{3}$$

where $\mathbf{f}_{\text{imb}}(t)$ is the component due to imbalance and $\mathbf{f}_{\text{mis}}(t)$ is the component caused by parallel or angular misalignment, or even a composition of them, and t is time (Wang and Jiang, 2018; Xu and Marangoni, 1994; Wang and Gong, 2019).

3.2 Imbalance in rotating machines

According to (Desouki et al., 2020), imbalance occurs when the center of mass of a rotating assembly does not coincide with the center of rotation. The ISO 21940-1:2016 defines imbalance as a resulting condition of force transmission or vibration movement through the bearings as a result of the action of centrifugal forces (ISO, 2016). The issue is usually attributed to deformations, asymmetries, imperfections in the raw material, and assembly errors caused by an eccentric concentrated mass. The imbalance force is described by:

$$\mathbf{f}_{\text{imb}}(t) = mr\omega^2, \tag{4}$$

where m is the unbalancing mass, r is the distance from the mass center of gravity to the rotation axis, and ω is the angular velocity. Imbalance in rotating machines can be identified by applying signal processing techniques. This fault presents amplitude in the fundamental frequency of the rotational speed,

which is much higher than the amplitudes of other harmonics in the radial direction. This issue provokes high vibration amplitudes, which causes stresses in structural supports and can eventually lead to their complete failure (Bloch and Geitner, 2005).

3.3 Misalignment in rotating machines

The alignment condition on rotating machines is given by the relative position of the connected shafts. If their centerlines are coincident, forming a straight line, the rotating machine is considered aligned. Otherwise, there is misalignment, which is usually classified as parallel or offset misalignment, angular misalignment, or more commonly, a combination of both (Hujare and Karnik, 2018). The misalignment produces forces and moments, inducing radial and axial vibrations in the system, which can be represented by:

$$\mathbf{f}_{\text{mis}}(t) = \mathbf{K}_{\text{c}}\boldsymbol{\Delta}\mathbf{e}, \tag{5}$$

where $\mathbf{K}_{\text{c}}$ is the couplings stiffness matrix and $\boldsymbol{\Delta}\mathbf{e}$ is the vector of misalignments, composed by parallel and angular displacement (Wang and Jiang, 2018; Wang and Gong, 2019). It should be said that the study of rotor misalignment has been limited to a qualitative understanding of the phenomenon. This has been mostly based on experiments with scarcely successful attempts to develop an effective mathematical model that allows for a quantitative evaluation of this defect (Desouki et al., 2020; Sinha et al., 2004; Lal and Tiwari, 2018).

3.4 Data augmentation

A common issue that occurs while working with supervised data is trying to learn from imbalanced data. This usually happens due to the underrepresentation of a set of classes, *i.e.* when an uneven number of instances are used to train the machine learning algorithm (Fernández et al., 2018). These are called minority classes. This situation leads to biased models where model accuracy decreases as the imbalance ratio increases. In real-world conditions, it is to be expected to have more instances representing normal conditions than those deemed to be abnormal or defective (Chawla et al., 2002). Learning from imbalanced data has thus become an integral part of machine learning techniques (Fernández et al., 2018).

In (Fernández et al., 2018) resampling methods were presented covering undersampling and oversampling. The undersampling techniques refer to the random elimination of samples from the majority classes to make them smaller and size comparable to the smallest ones. However, this approach leads to some problems since: *(i)* important instances may be discarded, resulting in a lack of data affecting class characterization; *(ii)* higher imbalance ratio, the number of samples that will be discarded, which may reduce the ability for generalization; and *(iii)* the reduction of the training set provokes a variance increase of

the classifier (Chawla et al., 2002; Dal Pozzolo et al., 2015). In contrast, the oversampling method relies on increasing the instances of minority classes in order to make them comparable in size to the largest ones. The candidate samples are replicated based on some weight criteria.

More elaborate techniques are commonly referred to as data augmentation techniques (Fernández et al., 2018; Chawla et al., 2002), and these will be the focus of the following sections. Namely: Section 3.4.1 presents the AWGN method; Section 3.4.2 describes the SMOTE approach; the details for the hybrid data augmentation method proposed in this work can be found in Section 3.4.3.

3.4.1 Additive white gaussian noise technique

AWGN can be used in the data augmentation process, which is applied to the data space instead of the feature space, as opposed to SMOTE (Fernández et al., 2018). Figure 3 represents the AWGN method where a zero-mean Gaussian noise is added to the input vibration signal (McClaning and Vito, 2000; de Lima et al., 2013) to create a new vibration signal. The Signal-to-Noise ratio (SNR), respectively presented in Equation (6), reflects the relation between the input signal average power (P_{signal}) and the average noise power P_{noise} in dB.

$$\text{SNR}_{\text{dB}} = 10 \log \left(\frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \tag{6}$$

Due to the random character of the added noise (Diniz et al., 2010), the original input signal can be transformed as many times as needed to make the resulting polluted signal comparable in size to those of the larger classes. This can be performed by adding random noise to each new copy of the vibration signal. In this research, we employed $\text{SNR}_{\text{dB}} = 15$ dB to create the noisy signal versions.

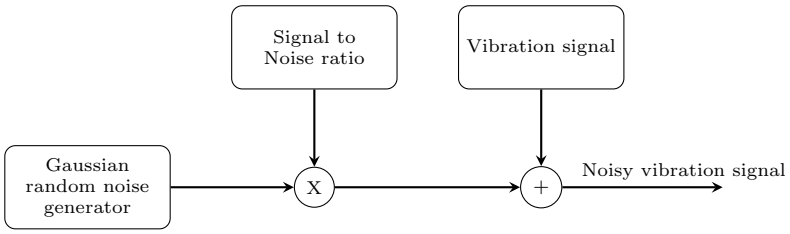

Fig. 3 AWGN signal addition scheme.

3.4.2 Synthetic minority oversampling technique

SMOTE was initially proposed in (Chawla et al., 2002) as an option to increase the proportion of minority classes in datasets. Its approach consists

in creating fictitious or synthetic observations in between two real observations. As commented in (Fernández et al., 2018; Chawla et al., 2002) this is a process applied to the feature space instead of the data space as occurs when using other oversampling methods.

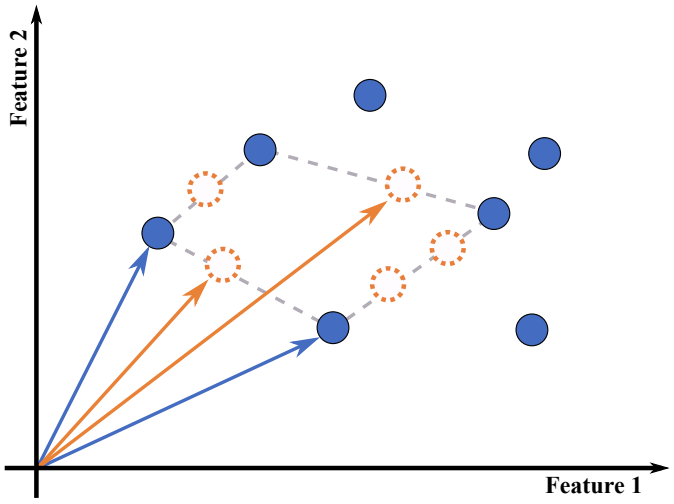

Fig. 4 Minority class feature space is represented in a simplified two dimensions scheme. The blue circles correspond to the real observation, the orange circles are the synthetics observations, the blue arrows are the real features vector and the blue arrows are the synthetics features vector.

Figure 4 presents a two-dimensional representation of the creation of the synthetic observations and the respective feature vectors. This technique can be applied to a multidimensional feature space. It is possible to create as many synthetic points as needed to make the minority class dataset size comparable or equal to the larger ones. A synthetic observation might be created between: *i)* two real observations; *ii)* a real observation and a synthetic one; and *iii)* between two previously created synthetic observations.

According to (Chawla et al., 2002), a synthetic observation can be constructed as follows: a given real feature vector, sample_i is randomly taken from the minority class dataset. In addition, one of the K nearest neighbors of a sample is randomly chosen. Subsequently, the difference, D_i between each respective feature of both vectors is calculated, and the new synthetic vector is created by summing each of feature c of the randomly chosen i samples to its corresponding $Di.G$, where G is a factor randomly chosen in the interval $[0 < G < 1]$ for each different feature c. This results in the construction of a synthetic vector between a sample and a neighbor. The aforementioned process is precisely detailed in Algorithm 1.

Algorithm 1: SMOTE algorithm

% Function syntax: $[Od] = smote(Id, S, K)$
% Id: minority class array.
% S: number of synthetic observations to add to Id.
% K: quantity of nearest neighbors
% Od: augmented class array.

$L \leftarrow$ Quantity of observations or lines of Id
$C \leftarrow$ Quantity of features or columns of Id
$Od \leftarrow Id$
for $s_c \leftarrow 1$ **to** S **do**
  $i \leftarrow$ Randomly chosen observation of Od
  $n \leftarrow$ Randomly chosen value of the K nearest neighbors
  indexes vector, using the standard K-NN algorithm, given i.
  % The code block below adds a new line to Od containing one new synthetic observation.
  $L \leftarrow L + s_c$
  for $c_c \leftarrow 1$ **to** C **do**
    $Di \leftarrow (Od(n, c_c) - Od(i, c_c))$
    $G \leftarrow$ Randomly chosen value in the interval $[0 < G < 1]$
    $Od(L, c_c) \leftarrow (Od(s_c, c_c) + (D_i.G))$
  end for
end for

3.4.3 Proposed hybrid data augmentation method

The use of SMOTE and AWGN techniques in an isolated manner to create additional instances of the minority classes are able to increase classifier performance. However, these methods can also increase overfitting (Zur et al., 2004; Santos et al., 2018), which is not desirable. Furthermore, SMOTE also has the potential to disseminate noisy information when new instances are created in unwanted positions (Cheng et al., 2019).

In order to increase the number of vibration signals and avoid overfitting, we propose a hybrid method combining SMOTE and AWGN. The purpose of applying this method is to create a set of artificial signals that have higher randomness than when applying techniques in an isolated manner. These would translate into a more robust and generalist classification model, thus decreasing the bias when compared with any one of the two data augmentation techniques employed. Two versions can be devised for the hybrid method. Namely, a first version (`version 1`) can be developed consisting in expanding only the number of instances of the minority classes without making changes to the majority classes. This procedure is illustrated in Figure 5, where:

– M_a represents the number of majority class instances;

- M_{i_1} represents the number of minority class instances obtained by feature extraction without using data augmentation techniques;
- M_{i_2} represents the number of minority class instances obtained from applying SMOTE;
- M_{i_3} represents the number of minority class instances obtained from applying AWGN.

In the first approach, M_a, M_{i_1}, M_{i_2} and M_{i_3} contain, respectively, $115, 41, 37$ and 37 instances each. Also, note that the quantity of Ma instances is equal to the sum of Mi_1, Mi_2 and Mi_3.

The second version of the method (`version 2`), presented in Figure 6, increases the number of minority class instances by x units using the AWGN technique and also modifies x signals of the majority class by adding Gaussian white noise. This way, the insertion of white noise does not become a discriminating feature between minority and majority classes. Figure 6 presents the overall details of the second approach where:

- M_{a_1} represents the number of majority class instances;
- M_{a_2} represents the number of instances modified by AWGN;
- M_{i_1} represents the number of minority class instances without using data augmentation techniques;
- M_{i_2} represents the number of minority class instances resulting from applying SMOTE;
- M_{i_3} represents the number of minority class instances resulting from applying AWGN.

In the second approach, M_{a_1}, M_{a_2}, M_{i_1}, M_{i_2} and M_{i_3} contain, respectively, $78, 37, 41, 37, 37$ instances each. Also, the sum of M_{a_1} and M_{a_2} is equal to the sum of M_{i_1}, M_{i_2} and M_{i_3}.

3.5 Classification methods

This paper compares four machine learning classification methods, namely Support Vector Machines (SVM), K-Nearest Neighbors (K-NN), Random Forest (RF) and Stacked Sparse Autoencoder (SSAE). These are, respectively, briefly described in Section 3.5.1, Section 3.5.2, Section 3.5.3 and Section 3.5.4.

3.5.1 Support Vector Machines

Support Vector Machines (SVM) is a machine learning method with a set of linear indicator functions that divides the feature space into two regions (Vapnik, 2013; Ziani et al., 2017). The method maps the original data

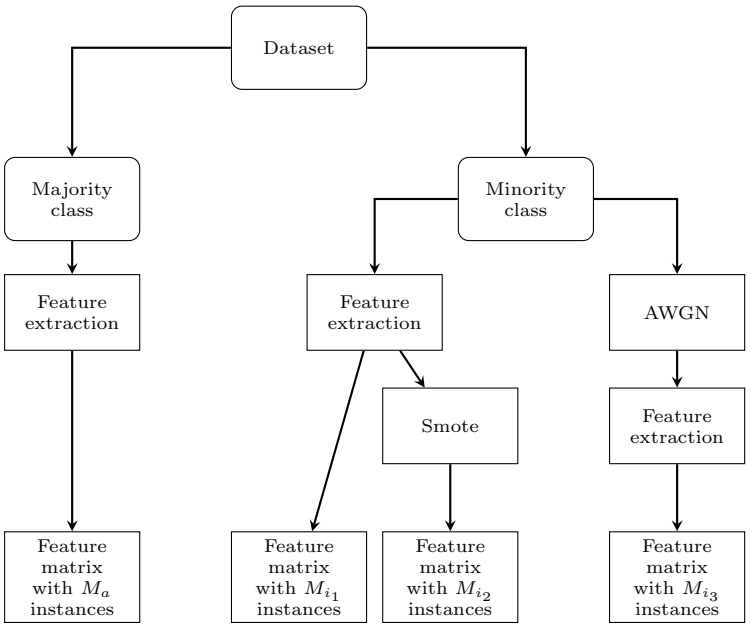

Fig. 5 Proposed hybrid method version 1.

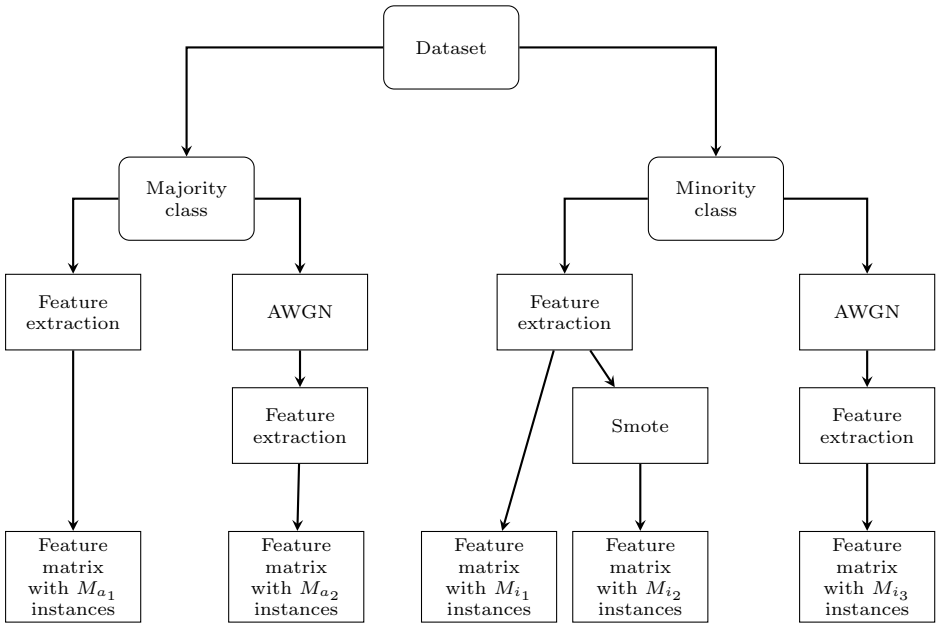

Fig. 6 Proposed hybrid method version 2.

in higher dimensional feature space (compared to the original one) using the training dataset. A hyperplane with a better discriminatory capacity is then constructed. This capacity depends on the kernel function employed, with the most common ones being the sigmoid, the radial basis, and the linear functions (Choubin et al., 2019). Usually, the radial basis function kernel tends to match the performance of the linear one (Chang et al., 2010). However, in the exploratory experiments performed in this work, the linear kernel delivered the best results. As a result, it was the one chosen for the rest of the evaluations. The linear kernel SVM also exhibits good results in the works presented in (Elangovan et al., 2011; Ruiz-Gonzalez et al., 2014).

3.5.2 K-Nearest Neighbors

K-Nearest Neighbors (K-NN) is one of the most used non-parametric methods (Yoon and Friel, 2013). This is essentially due to its simplicity of implementation. It is used to classify and cluster the nearest data vectors, with proximity being measured by some defined metric, the most common of which is the euclidean distance (also used in this work). K-NN is designed with the concept of the classification being decided by determining the majority class amongst its K closest neighbors (Xing and Bei, 2020).

3.5.3 Random Forest

Random Forest (RF) is a method of ensemble learning inspired by decision tree learning (Breiman, 2001). The method combines different decision tree predictors (with each one being statistically independent of the remaining ones) and outputs the most common predicted class. The method uses a variety of binary-ruled decisions to indicate a split in each tree (Görgens et al., 2015). Feature bagging is performed for each tree, where a random subset of the features is selected in the learning process. RF is ranked as one of the best classification methods (Fernández-Delgado et al., 2014), and its popularity growth is associated with the automation and simplicity of the algorithmic training procedure. As a result, system developers with little experience in machine learning can build classification systems with good discriminatory capacity (Fletcher and Reddy, 2016).

3.5.4 Stacked Sparse Autoencoder

AE is a deep learning algorithm consisting of neural networks whose objective is to encode and reconstruct, with the smallest possible error, the input itself in the output. It consists of two parts: an encoder and a decoder. The encoder is responsible for compressing the original data space into a new representation space, called latent space. The function of the decoder is to reconstruct the input data from the data representation in the latent space (Shao et al., 2017). The training step of the AE is unsupervised because the data labels are not provided (Li et al., 2020a). The AE can be used in several manners,

namely: *(i)* to perform feature reduction; *(ii)* to denoise data; *(iii)* to perform data augmentation; *(iv)* or classify data, as is the case in this paper (Fu et al., 2019).

A Stacked AE is a complex structure composed of a series of concatenated layers. The output of each layer is connected as an input to the next layer. In this structure, each layer is trained as an AE with the objective of reducing the error. After all layers are trained, a fine-tuning step is performed. For the classification step, the decoder layer is removed and a softmax layer is added. Due to a large number of neurons in the hidden layers, the sparse constraint is used to capture high-level representations of the data, thus its name, Stacked Sparse Autoencoder (SSAE) (Aouedi et al., 2020).

4 Results and Discussion

The main goal of this work is to identify the four classes described in Table 1, namely, No (Normal), I (Imbalance), IHM (Imbalance + Horizontal Misalignment) and IVM (Imbalance + Vertical Misalignment). In this section, the results of applying four types of classifiers are compared: SVM, K-NN, RF and SSAE in 14 different cases, which are described in Table 3.

Table 3: Cases description.

Case ID	**Description**
C_1	Employs original dataset.
C_2	Under samples majority class (No), using only 41 of the 115 available signals, so that this class has the same number of signals as the minority classes (I, IHM and IVM).
C_3	Over samples minority classes (I, IHM and IVM) using SMOTE, causing these classes to have 115 instances, which is the same number of majority class examples (No).
C_4	Over samples minority classes (I, IHM and IVM) using AWGN with SNR = 15 dB, causing these classes to have 115 instances, which is the same number of examples of the majority class (No).
C_5	Over samples minority classes (I, IHM and IVM) using AWGN with SNR = 15 dB, causing these classes to have 115 instances, which is the same number of examples of the majority class (No). In addition, white noise is added to *all* majority class examples, in order to reduce the risk of noise addition becoming a discriminatory feature. Amount of No instances is not changed.

Continues on next page

Table 3 – continued from previous page

Case ID	Description
C_6	Over samples minority classes (I, IHM and IVM) using version 1 of the hybrid method, causing these classes to have 115 instances, which is the same number of examples of the majority class (No).
C_7	Over samples minority classes (I, IHM and IVM) using version 2 of the hybrid method, causing these classes to have 115 instances, which is the same number of examples of the majority class (No). In addition, white noise is added to randomly selected majority class examples, in order to reduce the risk of noise addition becoming a discriminatory feature. Amount of No instances is not changed.
C_8	Application of features normalization plus C_1 procedures.
C_9	Application of features normalization plus C_2 procedures.
C_{10}	Application of features normalization plus C_3 procedures.
C_{11}	Application of features normalization plus C_4 procedures.
C_{12}	Application of features normalization plus C_5 procedures.
C_{13}	Application of features normalization plus C_6 procedures.
C_{14}	Application of features normalization plus C_7 procedures.

As the dataset used in this research has low cardinality, it is not recommended to use the holdout technique, which separates the data into training and test sets. In these circumstances, classifier training can result in overfitting issues, causing bias in the result (Aggarwal et al., 2018). As a result, we opted to instead apply 5-fold cross-validation, which is a stochastic partition method for training and test data. This results in a more robust and accurate prediction model (Dinov, 2018). The procedure iteratively goes through every possible training and test set combination evaluating the respective performance. This procedure is illustrated in Figure 7.

The classifiers have adjustable parameters whose selection was oriented by maximizing the highest average of intraclass relative hits. This is calculated through the sum of the correct answers of the main diagonal of the confusion matrix divided by the number of classes. The following sentences describe how the hyperparameter tuning for each classifier was performed. SVM training was performed by testing different values of the regularization term $C \in \{2^{-5}, 2^{-3}, 2^{-1}, ..., 2^{13}, 2^{15}\}$ using the linear kernel function. The training of the RF consisted of tuning the number of trees. During the training stage, the number of trees was varied from 1 to 50. The division rule used to form the nodes of the trees of RF was the Gini diversity criterion. The minimum number of observations per leaf used by the classifier was 1. In what concerns

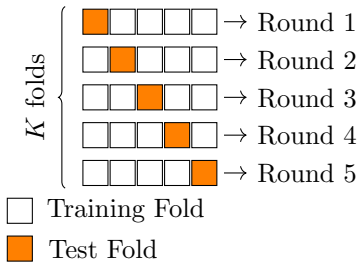

Fig. 7 K-fold representation.

the K-NN classifier, the number of neighbors was varied from 1 to 100 using the Euclidean distance to select the best value of K. Based on (Zhang et al., 2020b), the following hyperparameters were used to train SSAE with softmax classification: *(i)* three hidden layers consisting of, respectively, 100, 50, and 20 neurons; *(ii)* weight decay coefficient equal to 0.0001; *(iii)* sparsity penalty coefficient of 0.001; and *(iv)* sparsity factor set to 0.2. Seven metrics were used to measure the performance of the classifiers: classification time for one example (T), precision (P), recall (R), specificity (S), F1 Score (F1), accuracy (A), and standard deviation (SD) (Rehman et al., 2020; Kankar et al., 2011).

The following sections are organized as follows: Section 4.1 presents the results for the SVM classifier; Section 4.2 details the performance of the K-NN method; Section 4.3 describes the data obtained for the RF algorithm; and Section 4.4 lists the results for the SSAE approach.

4.1 SVM Results

Table 4 presents the SVM results for the dataset without using normalization. The data shows that using the undersampling technique (C_2) worsens SVM performance when compared against the baseline C_1. The SMOTE data-augmentation (C_3) technique causes a decrease in accuracy when compared to that of C_1. However, the other metrics evaluated are improved. The application of AWGN in C_4 and in all classes (C_5) improves precision, recall, specificity, and F1-score when compared to C_1. On the other hand, the application of these techniques worsens processing time, accuracy, and standard deviation. The application of the proposed hybrid method, version 1 (C_6) and version 2 (C_7), improves SVM performance in all the evaluated items except the processing time compared when compared against C_1.

Table 5 presents the results concerning feature normalization and showcases a significant improvement in SVM performance when compared with the described results in Table 4. Namely, all data augmentation techniques applied improved classifier performance when compared with the baseline results of C_8. Amongst the results presented, the best performing one is C_{14} which refers to the application of the second version of the hybrid method.

Table 4 SVM applied to the dataset *without* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_1	0.25	90.57	90.56	96.92	91.11	93.50	1.98
C_2	0.26	86.25	85.48	94.98	85.87	85.48	3.99
C_3	**0.23**	91.75	91.75	97.28	91.79	91.82	1.81
C_4	0.31	91.53	91.53	97.20	91.54	91.55	2.96
C_5	0.34	91.71	91.71	97.27	91.73	91.75	3.13
C_6	0.30	93.64	93.64	97.89	93.65	93.67	**1.70**
C_7	0.29	**93.76**	**93.76**	**97.92**	**93.75**	**93.74**	**1.70**

Table 5 SVM applied to the dataset *with* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_8	0.35	96.46	95.25	99.00	95.85	96.73	5.83
C_9	0.31	96.05	96.01	98.68	96.03	95.55	5.83
C_{10}	0.31	99.77	99.77	99.92	99.77	99.77	0.51
C_{11}	0.31	99.79	99.79	99.93	99.79	99.79	0.47
C_{12}	0.32	99.79	99.79	99.93	99.79	99.79	0.47
C_{13}	0.30	99.20	99.18	99.73	99.19	99.18	1.41
C_{14}	**0.25**	**99.84**	**99.84**	**99.95**	**99.84**	**99.84**	**0.36**

4.2 K-NN Results

Table 6 reports the K-NN classifier results without using feature normalization. The application of undersampling (C_2) improves K-NN performance in what concerns precision, recall, F1-score, and standard deviation. On the other hand, accuracy and specificity results are reduced when compared to the baseline (C_1). In addition, the application of oversampling techniques (C_3, C_4, C_5, C_6, C_7) caused an improvement when compared to: *(i)* the baseline results (C_1); and *(ii)* the undersampling approach (C_2). The techniques that exhibited the best results made use of AWGN (C_4 and C_5).

Table 6 K-NN applied to the dataset *without* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_1	**0.02**	57.23	57.17	86.26	57.20	67.88	9.53
C_2	0.03	62.03	61.56	83.47	61.80	61.56	5.02
C_3	0.03	89.59	89.35	96.22	89.47	89.35	2.29
C_4	0.03	**95.91**	**95.71**	**98.56**	**95.81**	**95.71**	2.59
C_5	**0.02**	95.47	95.28	98.40	95.37	95.28	**1.14**
C_6	**0.02**	88.08	87.94	95.68	88.01	87.94	3.47
C_7	0.03	91.47	91.39	96.96	91.43	91.39	2.69

Table 7 reports K-NN results for normalized features. As can be verified, the application of feature normalization improved the performance in all evaluated cases when compared to the results without normalization shown in

Table 6. The application of undersampling (C_2) improved precision, recall, F1 score and standard deviation when compared against C_1.

The application of data augmentation techniques ($C_{10}, C_{11}, C_{12}, C_{13}, C_{14}$) improved K-NN performance. The technique which presented the best result was the second version of the hybrid method (C_{14}), which resulted in an improvement of 20.92% in precision, 21.19% in recall, 5.35% in specificity, 15.26% in accuracy, 21.06% in F1-score and a reduction of 6.21% in standard deviation without requiring an increase in processing time against the baseline (C_8).

Table 7 K-NN applied to the dataset *with* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_8	0.04	78.56	78.27	94.47	78.41	84.20	7.41
C_9	**0.03**	81.64	81.69	93.38	81.67	81.69	7.19
C_{10}	0.04	98.26	98.25	99.41	98.26	98.25	**0.62**
C_{11}	0.04	99.01	99.00	99.67	99.01	99.00	1.00
C_{12}	0.04	98.23	98.19	99.39	98.21	98.19	1.70
C_{13}	0.05	98.29	98.25	99.41	98.27	98.25	2.04
C_{14}	0.04	**99.48**	**99.46**	**99.82**	**99.47**	**99.46**	1.20

4.3 RF Results

Table 8 presents RF results without feature normalization. The use of undersampling (C_2) increases performance when compared against C_1. Application of oversampling causes an improvement in performance (C_3, C_4, C_5, C_6, C_7). The best performance was derived from AWGN application in all classes (C_5) and the second version of the hybrid proposal (C_7). The latter achieved the best results, producing an improvement of 7.71% in precision, 11.46% in recall, 2.81% in specificity, 9.63% in F1-score, 8.27% in accuracy, 2.93% reduction in standard deviation, and a processing time of 0.07 s when compared against C_1.

Table 8 RF applied to the dataset *without* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_1	0.72	91.61	87.86	96.96	89.69	91.05	3.90
C_2	0.75	96.04	96.04	98.66	96.04	96.04	4.61
C_3	0.79	97.74	97.71	99.23	97.72	97.71	1.94
C_4	0.71	99.08	99.06	99.69	99.07	99.06	1.00
C_5	0.60	99.21	99.19	99.73	99.20	99.19	1.38
C_6	**0.46**	99.12	99.12	99.71	99.12	99.12	1.03
C_7	0.65	**99.32**	**99.32**	**99.77**	**99.32**	**99.32**	**0.97**

Table 9 shows RF results using feature normalization. The data demonstrate an improvement in RF performance for C_8 and C_{10} when compared

with, respectively, C_1 and C_3 of Table 8. However, RF performance for C_9, C_{11}, C_{12}, C_{13}, C_{14} was reduced when compared against, respectively, C_2, C_4, C_5, C_6, C_7 of Table 8. The results of Table 9 also show that the application of data augmentation techniques ($C_{10}, C_{11}, C_{12}, C_{13}, C_{14}$) improved RF performance when compared to C_8. The most effective techniques were: *(i)* SMOTE (C_{10}); and *(ii)* AWGN applied to all classes (C_{12}) using normalized features.

Table 9 RF applied to the dataset *with* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_8	0.47	96.30	95.01	98.95	95.65	96.56	3.57
C_9	**0.35**	94.00	93.84	97.90	93.94	93.84	4.48
C_{10}	0.63	99.06	99.05	99.68	99.05	99.05	1.02
C_{11}	0.43	98.57	98.55	99.51	98.56	98.55	2.13
C_{12}	0.43	**99.12**	**99.10**	**99.70**	**99.11**	**99.10**	**0.92**
C_{13}	0.60	96.68	96.64	98.86	96.65	96.64	2.12
C_{14}	0.39	98.49	98.49	99.49	98.49	98.49	1.86

4.4 SSAE Results

Table 10 reports SSAE results without feature normalization. The use of undersampling (C_2) reduced the specificity, F1-score, and accuracy when compared against C_1. Application of oversampling (C_3, C_4, C_5, C_6, C_7) caused an improvement in precision, recall, F1-score, and standard deviation. The best performance was derived from the AWGN application in minority classes (C_4).

Table 10 SSAE applied to the dataset *without* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_1	0.45	40.56	44.63	**80.83**	44.59	**59.09**	7.38
C_2	0.47	40.73	45.47	74.35	42.97	45.47	6.79
C_3	0.41	50.71	49.54	76.29	50.12	49.54	3.98
C_4	0.37	**53.68**	**54.07**	78.58	**53.87**	54.07	5.54
C_5	0.46	52.34	53.45	78.31	52.89	53.45	1.87
C_6	**0.33**	49.43	49.17	74.97	49.30	49.17	5.72
C_7	0.55	50.47	51.58	77.12	51.02	51.57	**0.54**

Table 11 presents the results concerning feature normalization and showcases a significant improvement in SSAE performance when compared with the described results in Table 10. The use of undersampling (C_2) reduced the performance when compared against C_1. The application of data augmentation techniques ($C_{10}, C_{11}, C_{12}, C_{13}, C_{14}$) improved SSAE performance. The technique which presented the best result was the second version of the hybrid method (C_{14}), which resulted in an improvement of 4.42% in precision,

5.12% in recall, 1.07% in specificity, 4.77% in accuracy, 3.53% in F1-score, a reduction of 3.10% in standard deviation and reduced the processing time in 0.45 seconds against the baseline (C_8).

Table 11 SSAE applied to the dataset *with* features normalization.

Cases	T (s)	P (%)	R (%)	S (%)	F1 (%)	A (%)	SD (%)
C_8	0.58	95.58	94.88	98.93	95.23	96.47	3.10
C_9	0.35	90.28	90.06	96.48	90.17	90.06	7.70
C_{10}	0.63	99.06	99.05	99.68	99.05	99.05	1.02
C_{11}	0.45	99.84	99.84	99.95	99.84	99.84	0.36
C_{12}	0.44	99.69	99.68	99.89	99.68	99.68	0.43
C_{13}	0.28	99.69	99.68	99.89	99.68	99.68	0.43
C_{14}	**0.13**	**100**	**100**	**100**	**100**	**100**	**0**

4.5 Discussion

The results also demonstrate that feature normalization is a relevant step for the K-NN, SVM, and SSAE methods, as these methods are sensitive to different feature scales. Avoiding the characteristics that have low values when compared to other ones has little influence on the decision of these classifiers. The application of the AWGN and SMOTE techniques improves the results of the four classifiers analyzed when compared to the baseline results. This is due to the small number of examples of faulty classes available in the original data sets, which hinders the individual training stages. The scarcity of machine failure signals is a frequent occurrence in real industrial environments, making the case for data augmentation approaches.

By analyzing the results it is possible to conclude that the SVM classifier achieved the best behavior when using the original dataset for both the normalized and non-normalized approaches (C_1 and C_8). Application of undersampling increased the performance of *(i)* K-NN when applied to normalized features; and *(ii)* RF when non-normalized features were used. RF performance through normalization only improved when the original dataset was employed (C_8) and when using SMOTE (C_{10}). The K-NN technique was able to deliver the fastest classification times. Overall:

- SVM exhibited the best results when using the second version of the hybrid method applied to the normalized features;
- K-NN exhibited the best results when using the second version of the hybrid method applied to normalized features;
- RF exhibited the best results when using the second version of the hybrid method applied to non-normalized features;

– SSAE exhibited the best results when using the second version of the hybrid method applied to the normalized features.

Version 2 of the proposed hybrid method is the data augmentation technique that resulted in the best performance, surpassing the application of the AWGN and SMOTE techniques individually. This shows the effectiveness of the approach when identifying combined failures in rotating machines. The hybrid proposal was able to produce new data examples with greater randomness than when using only AWGN or SMOTE. Consequently, the models generated from the hybrid approach are more generalist, resulting in an improvement in classifier performance. Figure 8 presents a radar plot comparing the performance of these classifiers.

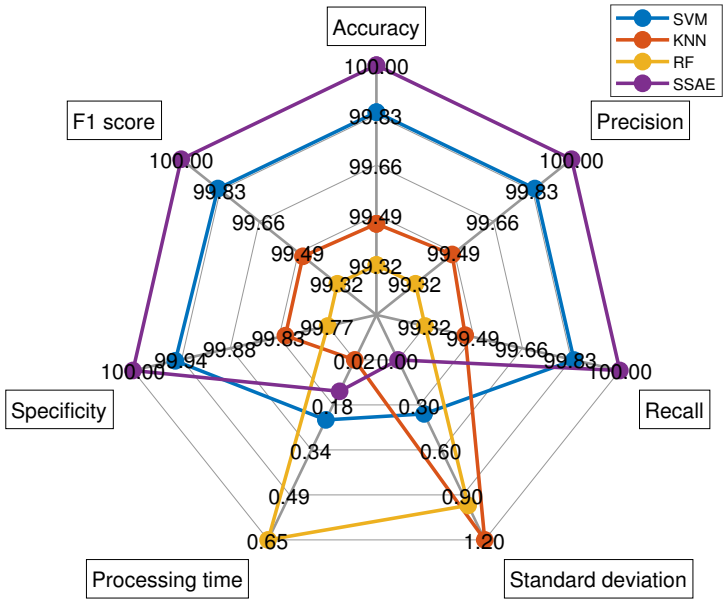

Fig. 8 Radar plot of the best classifiers: SVM using hybrid method version 2 applied to normalized features, K-NN using hybrid method version 2 applied to normalized features and; RF using hybrid method version 2 applied to non-normalized features; SSAE using hybrid method version 2 applied to normalized features.

Overall, the SSAE classifier stood out, outperforming the other ones, except for classification time where K-NN performed better. As a result, in the context of this research, the SSAE with feature normalization alongside the second version of the hybrid data augmentation proposal exhibits the best performance. In addition, it is also important to emphasize that the less time a classifier takes to identify a test example, the less complex the generated classifier model will be (Qin et al., 2021). Classification time can be a determining

factor for online fault diagnosis when deploying a classifier in an industrial setting. The data obtained show that the K-NN algorithm is recommended due to its processing speed and for exhibiting good performance among the four classifiers examined.

5 Conclusions

In this paper, a hybrid data augmentation method based on AWGN and SMOTE techniques were proposed to diagnose combined faults in rotating machines, which is a more complex task than identifying isolated failures. In industrial rotating machines, little data is available regarding faults when compared to normal operation, which leads to an imbalanced dataset. Consequently, it is necessary to use data augmentation techniques to increase the number of minority classes examples to improve classifier performance.

To validate the generalization and effectiveness of the proposed method, a comparison with 4 classifiers was performed considering 14 different cases. Each one of these tested a specific configuration such as using the original dataset, undersampling the majority class, applying feature normalization, utilizing AWGN, employing SMOTE and our hybrid proposal. The results obtained show that the latter surpassed the other approaches used in this paper. This resulted in more generalist classifier models, which improved their performance.

The best result was achieved by combining the hybrid data augmentation with the SSAE algorithm using normalized features. This method was able to achieve a processing time of 0.13 seconds whilst attaining 100% of accuracy. However, if the classifier is to be deployed in industrial applications where execution time is crucial then the K-NN classifier is a good option due to its compromise of high processing speed (0.04 seconds) and elevated accuracy (99.46%). Overall, the proposed hybrid data augmentation method is effective in improving classifier performance.

For future work, it is our intention to: *(i)* add the classes of horizontal and vertical misalignment separately; and *(ii)* the combined failure of horizontal misalignment associated with vertical misalignment. The addition of these classes will require a reevaluation of classifier performance. We also intend to use techniques such as genetic algorithms and minimum-redundancy maximum-relevancy to select the best features in order to perform dimensionality reduction. This procedure has the potential to improve classifier performance and avoid overfitting.

Acknowledgment

This work was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) Finance Code 001 and by

Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ).

Competing interests:

The authors declare that they have no conflict of interest.

Data availability

The dataset generated in this paper is available from the corresponding author on reasonable request.

References

Aggarwal CC, et al. (2018) Neural networks and deep learning. Springer

Ali MA, Bingamil AA, Jarndal A, Alsyouf I (2019) The influence of handling imbalance classes on the classification of mechanical faults using neural networks. In: 2019 8th International Conference on Modeling Simulation and Applied Optimization (ICMSAO), Manama, Bahrain, pp 1–5, DOI https://doi.org/10.1109/ICMSAO.2019.8880437

Aouedi O, Piamrat K, Bagadthey D (2020) A semi-supervised stacked autoencoder approach for network traffic classification. In: 2020 IEEE 28th International Conference on Network Protocols (ICNP), Madrid, Spain, pp 1–6, DOI 10.1109/ICNP49622.2020.9259390

Arslan M, Guzel M, Demirci M, Ozdemir S (2019) SMOTE and gaussian noise based sensor data augmentation. In: 2019 4th International Conference on Computer Science and Engineering (UBMK), Samsun, Turkey, pp 1–5, DOI https://doi.org/10.1109/UBMK.2019.8907003

Asadi R, Mustapha N, Sulaiman N, Shiri N (2009) New supervised multi layer feed forward neural network model to accelerate classification with high accuracy. European Journal of Scientific Research 33(1):163–178

Bai C, Ganeriwala SS, Sawalhi N (2019) A rational basis for determining vibration signature of shaft/coupling misalignment in rotating machinery. In: Rotating Machinery, Vibro-Acoustics & Laser Vibrometry, Volume 7, Florida, USA, pp 207–217, DOI https://doi.org/10.1007/978-3-319-74693-7_20

Bloch HP, Geitner FK (2005) Machinery component maintenance and repair. Elsevier

Bramer M (2007) Principles of data mining, vol 180. Springer

Breiman L (2001) Random forests. Machine learning 45(1):5–32, DOI https://doi.org/10.1023/A:1010933404324

Burks R, Islam KA, Lu Y, Li J (2019) Data augmentation with generative models for improved malware detection: A comparative study. In: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication

Conference (UEMCON), New York, USA, pp 0660–0665, DOI https://doi.org/10.1109/UEMCON47517.2019.8993085

Canbaz H, Polat K (2019) Fault detection of cnc machines from vibration signals using machine learning methods. In: The International Conference on Artificial Intelligence and Applied Mathematics in Engineering, Antalya, Turkey, pp 365–374, DOI https://doi.org/10.1007/978-3-030-36178-5_27

Cerrada Mea (2018) A review on data-driven fault severity assessment in rolling bearings. Mechanical Systems and Signal Processing 99(1), DOI https://doi.org/10.1016/j.ymssp.2017.06.012

Chang YW, Hsieh CJ, Chang KW, Ringgaard M, Lin CJ (2010) Training and testing low-degree polynomial data mappings via linear SVM. Journal of Machine Learning Research 11(4):1471–1490

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP (2002) SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research 16(1):321–357, DOI https://doi.org/10.1613/jair.953

Cheng K, Zhang C, Yu H, Yang X, Zou H, Gao S (2019) Grouped SMOTE with noise filtering mechanism for classifying imbalanced data. IEEE Access 7(1):170668–170681, DOI https://doi.org/10.1109/ACCESS.2019.2955086

Choubin B, Moradi E, Golshan M, Adamowski J, Sajedi-Hosseini F, Mosavi A (2019) An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Science of The Total Environment 651(1):2087 – 2096, DOI https://doi.org/10.1016/j.scitotenv.2018.10.064

Dal Pozzolo A, Caelen O, Bontempi G (2015) When is undersampling effective in unbalanced classification tasks? In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Porto, Portugal, pp 200–215, DOI https://doi.org/10.1007/978-3-319-23528-8_13

Dekhane A, Djellal A, Boutebbakh F, Lakel R (2020) Cooling fan combined fault vibration analysis using convolutional neural network classifier. In: Proceedings of the 3rd International Conference on Networking, Information Systems & Security, New York, USA, pp 1–6, DOI https://doi.org/10.1145/3386723.3387898

Desouki M, Sassi S, Renno J, Gowid SA (2020) Dynamic response of a rotating assembly under the coupled effects of misalignment and imbalance. Shock and Vibration 2020(1):1070–9622, DOI https://doi.org/10.1155/2020/8819676

Dhamande LS, Chaudhari MB (2018) Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. Measurement 125(1):63–77, DOI https://doi.org/10.1016/j.measurement.2018.04.059

Diniz PS, Da Silva EA, Netto SL (2010) Digital signal processing: system analysis and design. Cambridge University Press

Dinov ID (2018) Data science and predictive analytics: Biomedical and health applications using R. Springer

Djagarov N, Grozdev Z, Enchev G, Djagarov J (2019) Ship's induction motors fault diagnosis. In: 2019 16th Conference on Electrical Machines, Drives and

Power Systems (ELMA), Varna, Bulgaria, pp 1–4, DOI https://doi.org/10.1109/ELMA.2019.8771525

Elangovan M, Sugumaran V, Ramachandran K, Ravikumar S (2011) Effect of SVM kernel functions on classification of vibration signals of a single point cutting tool. Expert Systems with Applications 38(12):15202–15207, DOI https://doi.org/10.1016/j.eswa.2011.05.081

Fernández A, Garcia S, Herrera F, Chawla NV (2018) SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. Journal of artificial intelligence research 61(1):863–905, DOI https://doi.org/10.1613/jair.1.11192

Fernández-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we need hundreds of classifiers to solve real world classification problems? The journal of machine learning research 15(1):3133–3181

Fletcher RS, Reddy KN (2016) Random forest and leaf multispectral reflectance data to differentiate three soybean varieties from two pigweeds. Computers and Electronics in Agriculture 128(1):199 – 206, DOI https://doi.org/10.1016/j.compag.2016.09.004

Fu X, Wei Y, Xu F, Wang T, Lu Y, Li J, Huang JZ (2019) Semi-supervised aspect-level sentiment classification model based on variational autoencoder. Knowledge-Based Systems 171(1):81–92, DOI https://doi.org/10.1016/j.knosys.2019.02.008

Glowacz A (2018) Acoustic based fault diagnosis of three-phase induction motor. Applied Acoustics 137(1):82–89, DOI https://doi.org/10.1016/j.apacoust.2018.03.010

Goyal D, Pabla B, Dhami S, et al. (2019) Non-contact sensor placement strategy for condition monitoring of rotating machine-elements. Engineering Science and Technology, an International Journal 22(2):489–501, DOI https://doi.org/10.1016/j.jestch.2018.12.006

Guan Z, Chen P, Zhang X, Zhou X, Li K (2017) Vibration analysis of shaft misalignment and diagnosis method of structure faults for rotating machinery. International Journal of Performability Engineering 13(4):337–347, DOI https://doi.org/10.23940/ijpe.17.04.p1.337347

Görgens EB, Montaghi A, Rodriguez LCE (2015) A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. Computers and Electronics in Agriculture 116(1):221 – 227, DOI https://doi.org/10.1016/j.compag.2015.07.004

Hujare DP, Karnik MG (2018) Vibration responses of parallel misalignment in al shaft rotor bearing system with rigid coupling. Materials Today: Proceedings 5(11):23863–23871, DOI https://doi.org/10.1016/j.matpr.2018.10.178

ISO (2016) Mechanical vibration — rotor balancing— part 11: Procedures and tolerances for rotors with rigid behaviour. ISO 21940-11

Jin Y, Qin C, Huang Y, Liu C (2021) Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network. Measurement 173(1):108500, DOI https://doi.org/10.1016/j.measurement.2020.108500

Kankar PK, Sharma SC, Harsha SP (2011) Fault diagnosis of ball bearings using machine learning methods. Expert Systems with applications 38(3):1876–1886, DOI https://doi.org/10.1016/j.eswa.2010.07.119

Klausen A, Van Khang H, Robbersmyr KG (2018) Novel threshold calculations for remaining useful lifetime estimation of rolling element bearings. In: 2018 XIII International Conference on Electrical Machines (ICEM), Alexandroupoli, Greece, pp 1912–1918, DOI https://doi.org/10.1109/ICELMACH.2018.8507056

Lal M, Tiwari R (2018) Experimental identification of shaft misalignment in a turbo-generator system. Sādhanā 43(5):80, DOI https://doi.org/10.1007/s12046-018-0859-1

Li H, Li M, Li C, Li F, Meng G (2017) Multi-faults decoupling on turboexpander using differential-based ensemble empirical mode decomposition. Mechanical Systems and Signal Processing 93(1):267–280, DOI https://doi.org/10.1016/j.ymssp.2017.02.015

Li J, Li H, Yu JL (2011) Application of random-SMOTE on imbalanced data mining. In: 2011 Fourth International Conference on Business Intelligence and Financial Engineering, Wuhan, China, pp 130–133, DOI https://doi.org/10.1109/BIFE.2011.25

Li J, Li X, He D, Qu Y (2020a) Unsupervised rotating machinery fault diagnosis method based on integrated sae–dbn and a binary processor. Journal of Intelligent Manufacturing 95(8):1–18, DOI https://doi.org/10.1007/s10845-020-01543-8

Li P, Hu W, Hu R, Chen Z (2020b) Imbalance fault detection based on the integrated analysis strategy for variable-speed wind turbines. International Journal of Electrical Power & Energy Systems 116(1):105570, DOI https://doi.org/10.1016/j.ijepes.2019.105570

Li X, Yang X, Yang Y, Bennett I, Mba D (2019) A novel diagnostic and prognostic framework for incipient fault detection and remaining service life prediction with application to industrial rotating machines. Applied Soft Computing 82(1):105564, DOI https://doi.org/10.1016/j.asoc.2019.105564

Li X, Zhang W, Ding Q, Sun JQ (2020c) Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation. Journal of Intelligent Manufacturing 31(2):433–452, DOI https://doi.org/10.1007/s10845-018-1456-1

de Lima AA, Prego TdM, Netto SL, da Silva EA, Gutierrez RH, Monteiro UA, Troyman AC, Silveira FJdC, Vaz L (2013) On fault classification in rotating machines using fourier domain features and neural networks. In: 2013 IEEE 4th Latin American Symposium on Circuits and Systems (LASCAS), Cusco, Peru, pp 1–4, DOI https://doi.org/10.1109/LASCAS.2013.6518984

Martins DHC, Viana DP, de Lima AA, Pinto MF, Tarrataca L, e Silva FL, Gutiérrez RHR, de Moura Prego T, Monteiro UABV, Haddad DB, et al. (2021) Diagnostic and severity analysis of combined failures composed by imbalance and misalignment in rotating machines. The International Journal of Advanced Manufacturing Technology 114(9):1–16, DOI https://doi.org/10.1007/s00170-021-06873-2

Martins DHCdSS, Hemerly DO, Marins M, Lima AA, Silva FL, Prego TdM, Ribeiro FML, Netto SL, da Silva EAB (2019) Application of machine learning to evaluate unbalance severity in rotating machines. In: Proceedings of the 10th International Conference on Rotor Dynamics – IFToMM, Rio de Janeiro, Brazil, pp 144–160, DOI https://doi.org/10.1007/978-3-319-99268-6_11

McClaning K, Vito T (2000) Radio receiver design. Noble Publishing

Mikołajczyk A, Grochowski M (2018) Data augmentation for improving deep learning in image classification problem. In: 2018 international interdisciplinary PhD workshop (IIPhDW), Swinoujscie, Poland, pp 117–122, DOI https://doi.org/10.1109/IIPHDW.2018.8388338

Oh JW, Jeong J (2020) Data augmentation for bearing fault detection with a light weight CNN. Procedia Computer Science 175(1):72–79, DOI https://doi.org/10.1016/j.procs.2020.07.013

Pandya D, Upadhyay S, Harsha SP (2013) Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN. Expert Systems with Applications 40(10):4137–4145, DOI https://doi.org/10.1016/j.eswa.2013.01.033

Polat K (2020) The fault diagnosis based on deep long short-term memory model from the vibration signals in the computer numerical control machines. Journal of the Institute of Electronics and Computer 2(1):72–92, DOI https://doi.org/10.33969/JIEC.2020.21006.

Qian W, Li S, Jiang X (2019) Deep transfer network for rotating machine fault analysis. Pattern Recognition 96(1):106993, DOI https://doi.org/10.1016/j.patcog.2019.106993

Qin C, Jin Y, Tao J, Xiao D, Yu H, Liu C, Shi G, Lei J, Liu C (2021) DTCNNMI: A deep twin convolutional neural networks with multi-domain inputs for strongly noisy diesel engine misfire detection. Measurement 180(1):109548, DOI https://doi.org/10.1016/j.measurement.2021.109548

Rashid KM, Louis J (2019) Times-series data augmentation and deep learning for construction equipment activity recognition. Advanced Engineering Informatics 42(1):100944, DOI https://doi.org/10.1016/j.aei.2019.100944

Razavi-Far R, Farajzadeh-Zanjani M, Saif M (2017) An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors. IEEE Transactions on Industrial Informatics 13(6):2758–2769, DOI https://doi.org/10.1109/TII.2017.2755064

Rehman A, Naz S, Razzak MI, Akram F, Imran M (2020) A deep learning-based framework for automatic brain tumors classification using transfer learning. Circuits, Systems, and Signal Processing 39(2):757–775, DOI https://doi.org/10.1007/s00034-019-01246-3

Rochac JFR, Liang L, Zhang N, Oladunni T (2019) A gaussian data augmentation technique on highly dimensional, limited labeled data for multi-class classification using deep learning. In: 2019 Tenth International Conference on Intelligent Control and Information Processing (ICICIP), Marrakesh, Morocco, pp 145–151, DOI https://doi.org/10.1109/ICICIP47338.2019.9012197

Ruiz-Gonzalez R, Gomez-Gil J, Gomez-Gil FJ, Martínez-Martínez V (2014) An SVM-based classifier for estimating the state of various rotating components in agro-industrial machinery with a vibration signal acquired from a single point on the machine chassis. Sensors 14(11):20713–20735, DOI https://doi.org/10.3390/s141120713

Santos MS, Soares JP, Abreu PH, Araujo H, Santos J (2018) Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]. IEEE Computational intelligence magazine 13(4):59–76, DOI https://doi.org/10.1109/MCI.2018.2866730

Shao H, Jiang H, Zhao H, Wang F (2017) A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. Mechanical Systems and Signal Processing 95(1):187–204, DOI https://doi.org/10.1016/j.ymssp.2017.03.034

Shorten C, Khoshgoftaar TM (2019) A survey on image data augmentation for deep learning. Journal of Big Data 6(1):1–48, DOI https://doi.org/10.1186/s40537-019-0197-0

Si XS, Wang W, Hu CH, Zhou DH (2011) Remaining useful life estimation – a review on the statistical data driven approaches. European Journal of Operational Research 213(1):1–14, DOI https://doi.org/10.1016/j.ejor.2010.11.018

Sikder N, Bhakta K, Al Nahid A, Islam MM (2019) Fault diagnosis of motor bearing using ensemble learning algorithm with FFT-based preprocessing. In: 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), Dhaka, Bangladesh, pp 564–569, DOI https://doi.org/10.1109/ICREST.2019.8644089

Sinha JK, Lees A, Friswell M (2004) Estimating unbalance and misalignment of a flexible rotating machine from a single run-down. Journal of Sound and Vibration 272(3):967–989, DOI https://doi.org/10.1016/j.jsv.2003.03.006

Srinivas RS, Tiwari R, Kannababu C (2019) Model based analysis and identification of multiple fault parameters in coupled rotor systems with offset discs in the presence of angular misalignment and integrated with an active magnetic bearing. Journal of Sound and Vibration 450(1):109–140, DOI https://doi.org/10.1016/j.jsv.2019.03.007

Suarez-Alvarez MM, Pham DT, Prostov MY, Prostov YI (2012) Statistical approach to normalization of feature vectors and clustering of mixed datasets. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 468(2145):2630–2651, DOI https://doi.org/10.1098/rspa.2011.0704

Vapnik V (2013) The nature of statistical learning theory. Springer science & business media

Verma AK, Sarangi S, Kolekar M (2014) Experimental investigation of misalignment effects on rotor shaft vibration and on stator current signature. Journal of Failure Analysis and Prevention 14(2):125–138, DOI https://doi.org/10.1007/s11668-014-9785-7

Villa LF, Reñones A, Perán JR, de Miguel LJ (2012) Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load. Mechanical Systems and Signal Processing

29(1):436–446, DOI https://doi.org/10.1016/j.ymssp.2011.12.013
Walpole RE, Myers RH (2012) Probability & statistics for engineers & scientists. Pearson Education Limited
Wang H, Gong J (2019) Dynamic analysis of coupling misalignment and unbalance coupled faults. Journal of Low Frequency Noise, Vibration and Active Control 38(2):363–376, DOI https://doi.org/10.1177/1461348418821582
Wang HY (2008) Combination approach of SMOTE and biased-SVM for imbalanced datasets. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, pp 228–231, DOI https://doi.org/10.1109/IJCNN.2008.4633794
Wang J, Du G, Zhu Z, Shen C, He Q (2020) Fault diagnosis of rotating machines based on the EMD manifold. Mechanical Systems and Signal Processing 135(1):106443, DOI https://doi.org/10.1016/j.ymssp.2019.106443
Wang N, Jiang D (2018) Vibration response characteristics of a dual-rotor with unbalance-misalignment coupling faults: Theoretical analysis and experimental study. Mechanism and Machine Theory 125(1):207–219, DOI https://doi.org/10.1016/j.mechmachtheory.2018.03.009
William PE, Hoffman MW (2011) Identification of bearing faults using time domain zero-crossings. Mechanical systems and signal processing 25(8):3078–3088, DOI https://doi.org/10.1016/j.ymssp.2011.06.001
Xiao D, Qin C, Yu H, Huang Y, Liu C, Zhang J (2021) Unsupervised machine fault diagnosis for noisy domain adaptation using marginal denoising autoencoder based on acoustic signals. Measurement 176(1):109186, DOI https://doi.org/10.1016/j.measurement.2021.109186
Xing W, Bei Y (2020) Medical health big data classification based on KNN classification algorithm. IEEE Access 8(1):28808–28819, DOI https://doi.org/10.1109/ACCESS.2019.2955754
Xu M, Marangoni R (1994) Vibration analysis of a motor-flexible coupling-rotor system subject to misalignment and unbalance, part i: Theoretical model and analysis. Journal of Sound and Vibration 176(5):663 – 679, DOI https://doi.org/10.1006/jsvi.1994.1405
Xu Q, Lu S, Jia W, Jiang C (2019) Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning. Journal of Intelligent Manufacturing (6):1–15, DOI https://doi.org/10.1007/s10845-019-01522-8
Yamamoto GK, da Costa C, da Silva Sousa JS (2016) A smart experimental setup for vibration measurement and imbalance fault detection in rotating machinery. Case Studies in Mechanical Systems and Signal Processing 4(1):8–18, DOI https://doi.org/10.1016/j.csmssp.2016.07.001
Yang B, Lei Y, Jia F, Xing S (2019) An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. Mechanical Systems and Signal Processing 122(1):692–706, DOI https://doi.org/10.1016/j.ymssp.2018.12.051
Yoon J, Friel N (2013) Efficient estimation of the number of neighbours in probabilistic K-Nearest Neighbour classification. ArXiv abs/1305.1002

Yu G (2019) A concentrated time–frequency analysis tool for bearing fault diagnosis. IEEE Transactions on Instrumentation and Measurement 69(2):371–381, DOI https://doi.org/10.1109/TIM.2019.2901514

Yu K, Lin TR, Ma H, Li H, Zeng J (2019) A combined polynomial chirplet transform and synchroextracting technique for analyzing nonstationary signals of rotating machinery. IEEE Transactions on Instrumentation and Measurement 69(4):1505–1518, DOI https://doi.org/10.1109/TIM.2019.2913058

Yu K, Lin TR, Ma H, Li X, Li X (2021) A multi-stage semi-supervised learning approach for intelligent fault diagnosis of rolling bearing using data augmentation and metric learning. Mechanical Systems and Signal Processing 146(1):107043, DOI https://doi.org/10.1016/j.ymssp.2020.107043

Zhang S, Zhang S, Wang B, Habetler TG (2020a) Deep learning algorithms for bearing fault diagnostics—a comprehensive review. IEEE Access 8(1):29857–29881, DOI https://doi.org/10.1109/ACCESS.2020.2972859

Zhang Y, Li X, Gao L, Chen W, Li P (2020b) Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method. Measurement 151(1):107232, DOI https://doi.org/10.1016/j.measurement.2019.107232

Ziani R, Felkaoui A, Zegadi R (2017) Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized fisher's criterion. Journal of Intelligent Manufacturing 28(2):405–417, DOI https://doi.org/10.1007/s10845-014-0987-3

Zur R, Jiang Y, Metz C (2004) Comparison of two methods of adding jitter to artificial neural network training. International Congress Series 1268(1):886–889, DOI https://doi.org/10.1016/j.ics.2004.03.238