symmetry

MDPI

Article

Machine Learning Techniques for Multi-Fault Analysis and Detection on a Rotating Test Rig Using Vibration Signal

Iulian Lupea [1,*] and Mihaiela Lupea [2]

[1] Faculty of Industrial Engineering, Robotics and Production Management, Technical University of Cluj-Napoca, 400641 Cluj-Napoca, Romania
[2] Faculty of Mathematics and Computer Science, Babes-Bolyai University, 400084 Cluj-Napoca, Romania
* Correspondence: iulian.lupea@mep.utcluj.ro

Abstract: Machine health monitoring of rotating mechanical systems is an important task in manufacturing engineering. In this paper, a system for analyzing and detecting mounting defects on a rotating test rig is developed. The test rig comprises a slender shaft with a central disc, supported symmetrically by oscillating ball bearings. The shaft is driven at constant speed (with tiny variations) through a timing belt. Faults, such as the translation of the central disc along the shaft, the disc eccentricity, and defects on the motor reducer position or timing belt mounting position, are imposed. Time and frequency domain features, extracted from the vibration signal, are used as predictors in fault detection. This task is modeled as a multi-class classification problem, where the classes correspond to eight health states: one healthy and seven faulty. Data analysis, using unsupervised and supervised algorithms, provides significant insights (relevance of features, correlation between features, classification difficulties, data visualization) into the initial dataset, a balanced one. The experiments are performed using classifiers from MATLAB and six feature sets. Quadratic SVM achieves the best performance: 99.18% accuracy for the set of all 41 features extracted from *X* and *Y* accelerometer axes, and 98.93% accuracy for the subset of the 18 most relevant features.

Keywords: fault detection; vibration signal; slender shaft; machine learning; multi-class classification

Citation: Lupea, I.; Lupea, M. Machine Learning Techniques for Multi-Fault Analysis and Detection on a Rotating Test Rig Using Vibration Signal. *Symmetry* **2023**, *15*, 86. https://doi.org/10.3390/sym15010086

Academic Editors: Zhiwen Chen, Hao Luo and Chao Cheng

Received: 17 November 2022
Revised: 17 December 2022
Accepted: 24 December 2022
Published: 28 December 2022

1. Introduction

Rotating machinery constitute an important class of machines in the manufacturing and automation industries. Machine health monitoring is already an important task in modern manufacturing systems. It is vital to detect, diagnose and predict faults associated with the machine elements and to estimate the remaining time until failure. Data-driven approaches have been extensively studied for the robust fault diagnosis of rotating machines [1,2]. Extracted features from measured vibration signals give early indications of mechanical problems such as imbalances, parallel/angular misalignments, bent or cracked shafts, oil whirl/whips, looseness of parts, rubs, blade passes, faults on transmission belts, bearings and gears defects [2–4]. A broad frequency range has to be observed from very low frequencies associated with subsynchronous components of the signal for loose parts or oil whirl to high frequencies such as the structural resonances excited by a faulty bearing or the gear mesh. This bottom-up or measured data-driven solution is a new paradigm compared to traditional top-down physics-based modeling. The aim of the fault detection problem is to identify the health state of a machine from a set of possible faulty states and a healthy state. A so-called conventional data-driven machine health monitoring starts with operator feature selection/extraction then continues with model training and testing by using shallow machine learning algorithms. Instead of this, deep-learning-based machine health monitoring systems are able to automatically learn internal representations from raw vibration signals and predict targets; hence, less expert knowledge and human labor is required [5].

Time series signals are in general the raw data generated by sensors used for data acquisition. The most used feature extraction methods from stationary vibration signals are those methods that extract statistical time-domain features and frequency-domain features [2–4,6]. In typical rolling element bearings, the extraction of statistical features from time-domain data has been found to be the most used approach [7]. Frequency domain features extraction, especially by using the fast Fourier transform (FFT) is very important, often sufficient for stationary signals processing. Apart from the spectral coefficients, some current statistical features in frequency domain are mean frequency, spectral skewness, spectral kurtosis, first-to-fourth order moments of the FFT coefficients and others. The methods of analysis in the time-frequency domain are used to extract characteristics from the non-stationary signals generated by and measured on the rotating machinery. Such methods are the short-time Fourier transform (STFT), continuous wavelet transform (CWT), wavelet packet decomposition (WPD), the Wigner–Ville distribution and other methods [8–10]. In STFT, by taking successive time slices of the vibration signal and applying the FFT to each, a spectrogram is generated. A limitation of STFT is their application for signals with slow variations in time. Many articles have been published for stationary or near-stationary signals. In [11] structural faults are detected by observing the low frequency FFT spectrum associated with the first several harmonics of the rotational frequency of a rotating machine. Fluctuation frequency bands about the theoretical harmonic values are considered for detection of faults such as unbalance, looseness or misalignment under varied operating conditions. In [12] typical faults of fatigue crack and shaft misalignment under noisy environment are observed in a rotor test bench. The raw vibration signal from displacement sensor is denoised with special combined methods, then an optimized convolutional neural network for fault classification is employed. In [13] vibration signals from tri-axial MEMS accelerometers (0.5 to 1.6 kHz frequency bandwidth on X and Y axis) and Arduino Mega 2560 microcontroller under LabVIEW are used for acquisition. Statistical time domain features and an artificial neural network (ANN) for fault diagnosis with MATLAB are proposed for the health monitoring of a belt-and-pulley test rig. The proper installation, service and preventive maintenance of a large palette of transmission belts including synchronous belts is noted in [14]. Advantages over chain and gear transmissions are mentioned. Diagnosis of belt-only faults and the case of the superimposion of these faults on an unbalanced shaft are studied in [15] by using the vibration time and frequency domain features. In [10] the interest is in the detection of idler faults on belt conveyors. Signals from accelerometers are decomposed with WPD. The extracted features are associated with the energies of different frequency bands. The features are employed to train a support vector machine (SVM). In the literature research into the health monitoring and fault detection of toothed or synchronous belts with machine learning techniques is not extensive. A review of various artificial intelligence algorithms for fault diagnosis of rotating machinery with advantages, limitations and industrial applications can be found in [16]. A wide range of faults and severity levels in rotating machines is approached by using deep learning and synthetic data augmentation through virtual sensors in [17]. This is applicable for data imbalance and data scarcity in some health states where the virtual sensors artificially augment the vibration data.

In the current work seven artificial faults are introduced in a rotating machinery test rig that is designed to be reconfigurable and capable of testing various mounting and structural faults. The current configuration comprises a slender shaft with a central disc placed symmetrically with regards to the supporting oscillating bearings and a timing belt transmission from the motor reducer. The proposed multi-fault analysis and detection system applies machine learning techniques to distinguish the machinery state from eight possible health states, one healthy and seven faulty. This is a data-driven approach that learns to predict a health state based on features extracted from data (vibration signals) that were previously measured for all states under observation. The vibration signals acquired from a triaxial accelerometer are processed, then time and frequency domain parameters (features) are extracted. The multi-fault detection task is modeled as a multi-class clas-

sification problem. The data analysis stage that precedes classification aims at revealing significant insights (relevance of features, correlation between features, classification difficulties, data visualization) into the initial dataset and selection of the most informative features. The experiments are performed using a set of classifiers from MATLAB 2022 and several selected feature sets with the goal of finding the best classification model that predicts the eight health states.

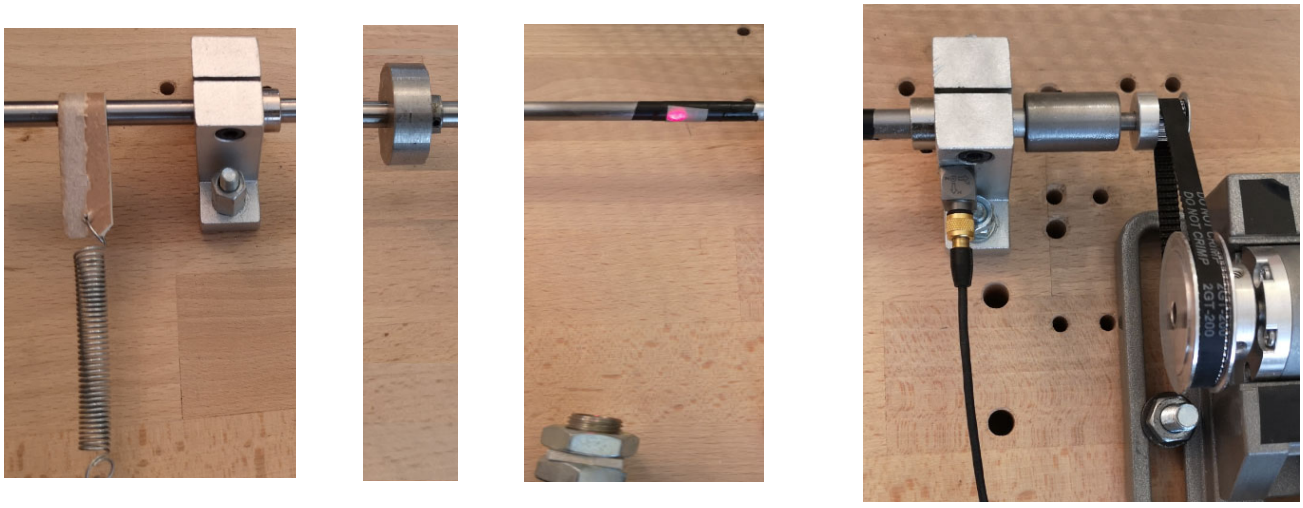
The rest of the paper is organized as follows. In Section 2 the test rig used for experiments, the data acquisition process and the faults are presented. Section 3 contains a brief description of the proposed multi-fault analysis and detection system. In Section 4 the time-domain and frequency-domain features extracted from the vibration signals are described. The subject of Section 5 is data analysis, an important step prior to classification. Section 6 is dedicated to the presentation of the classification results achieved by the classifiers applied to solve the multi-fault detection problem. Conclusions are drawn in the last section.

2. Experimental Set-Up and Defects

This section is dedicated to the presentation of the test rig used for experiments, the data acquisition process and the description of the defects.

2.1. Test Rig and Data Acquisition

For the experiments, a test rig with common machine elements such as a motor, a reducer, a timing belt with the associated pulleys and a flexible shaft is used (Figure 1d). The shaft is mounted on double raw ball oscillating bearings.



(**a**) (**b**) (**c**) (**d**)

Figure 1. Test rig: (**a**) friction load; (**b**) central disc; (**c**) tachometer; (**d**) motor reducer, timing belt, accelerometer.

The slender shaft is driven by a brushed Pololu DC motor (12 Volts) and a proper motor speed controller. The motor is coupled with a metal gearbox reducer of three stages and of 19:1 transmission ratio. The reducer's first stage is a helical gear pair for reducing noise and vibrations, followed by two stages of spur gears. From the reducer the rotation is transferred to the flexible shaft through a timing synchronous belt, type GT2-200 (2 mm tooth pitch, width of 6 mm, length of 200 mm), often used for 3D printers and robotic arms. The driver and the driven pulleys of z_1 = 60 teeth and z_2 = 18 teeth, respectively, are made of aluminum (Figure 1d). The associated rotation speeds (Hz) are n_1 and n_2 respectively. By manually translating the motor reducer (which is caught in a vice) the belt tension is adjusted. The calibrated flexible shaft of 0.008 m diameter has a length of 0.9 m between the ball bearings. The total length of the shaft is one meter. A disc is mounted at the middle of the shaft; hence, the central disc (Figure 1b) is 450 mm apart from the support with the accelerometer. The first flexural mode of vibration of the shaft with the central disc

is at the frequency f_1 of 17.56 Hz. The associated first natural circular frequency ω_1 is derived [15] according to the Ritz method [18], where the boundary conditions are pin joints at each support. Some other properties of the system are the bending rigidity of the shaft at $EI = 42.22$ Nm^2 (E is the Young's modulus of elasticity, I is the area moment of inertia), the shaft mass at $M = 0.35$ kg and $m = 0.055$ kg is the disc mass.

For the vibration data acquisition, a dynamic acquisition board NI USB-4431 type connected to a laptop is used. The board is provided with four analog-to-digital converters, each of 24 bit, and maximum sampling rate of 102.4 kS/s. The first three channels of the board are used for recording the output signal of the piezoelectric accelerometer. The B&K triaxial DeltaTron accelerometer (10 mV/g, with a frequency range of roughly 2 Hz to 7 kHz, a mass of 4 g, and a 4-mV supply current) is glued to the bearing support (X axis is horizontal, Y axis is axial, and, along the slender shaft, the Z axis is vertical). The tachometer (connected to the fourth acquisition channel) points a laser beam to the rotating shaft (Figure 1c). By using a tensioned plastic ribbon wrapped around the shaft (Figure 1a), a small friction load is imposed on the shaft.

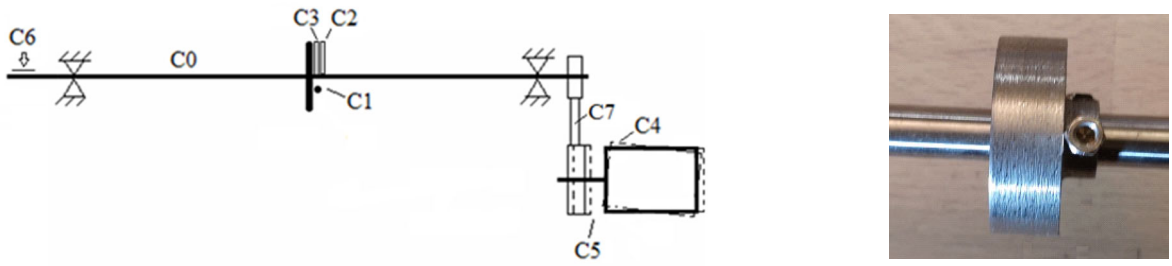
A LabVIEW application with acquisition on four analog channels, three for the accelerometer and one for the tachometer, is employed. The tachometer generates one voltage pulse per shaft revolution. The acquisition sampling rate is 5000 samples per second on each channel resulting a theoretical frequency bandwidth of up to 2.5 kHz for the vibration signals. Vibration data coming from the two radial axes, the axial axis and from the tachometer are saved in log files. The recording time is 500 s for each health state. One hundred observations per state are considered, each of 5 s duration.

2.2. Fault Conditions

In this study seven mounting defects (faults) on the test rig are under consideration. The eight health states (one healthy and seven faulty) of the system will be represented as classes: C0, C1, C2, and up to C7 (see Table 1) in the multi-class classification problem that models the multi-fault detection task. The faults are roughly drawn in Figure 2a, along a sketch of the test rig.

Table 1. List of classes—health states.

Class	Health State Description	Class	Health State Description
C0	Healthy state	C4	Motor reducer clockwise rotation
C1	Eccentric mass on the central disc	C5	Motor reducer translation
C2	Translation#1 of the central disc	C6	Increased friction load
C3	Translation#2 of the central disc	C7	Timing belt erroneous mounted



(**a**) (**b**)

Figure 2. (**a**) Test rig sketch with eight healthy states; (**b**) rotor and eccentric mass.

The healthy state, denoted by C0, corresponds to the initial state of the system, in which the central disc is at the middle of the shaft. The disc is fixed to the shaft with two M3 screws.

A small eccentric mass of one gram attached radially to the central disc (Figure 2b) corresponds to the first fault (class C1). An eccentric mass changes the radial vibration amplitude of the system and the statistical features of the signal measured at the bearing support and also the power spectrum lines which are mainly at the level of the first order $1\times$ turning speed n_2, where n_2 is the rotation speed (rps) of the shaft. The second fault (class C2) is described by the translation of the central disc of about 8 mm toward the motor reducer. The third fault (class C3) corresponds to a smaller translation of the central disc of 4 mm toward the motor reducer. The fourth system fault (class C4) is a clockwise rotation of the motor reducer resulting in a small angle between the two sheaves of the timing belt and a tiny decrease of the timing belt tension. The fifth system fault (class C5) is the motor reducer translation referring to the timing belt, more precisely the offset misalignment of sheaves (without eccentric mass on the central disc, similar to all the other faults excepting the first one). The offset is about 3.5 mm outward, obtained by the motor reducer translation from the initial (nominal) position. This fault type produces a high vibration amplitude at $1 \times f_{sv}$ (rps) of the particular sheave rotational speed (f_{sv}), predominantly in the axial direction. For comparison, the width of the timing belt is 6 mm and the width of the teeth of the sheaves is about 7.3 mm. The effect of the fault is an increased overall vibration level causing accelerated wear for the belt and the sheaves. Wear or pulley/sheave misalignment is regularly indicated by a high vibration amplitude at the timing belt frequency f_b, which is subsynchronous with respect to n_2 and is calculated using Equation (1).

$$f_b = D_1 \pi \, n_1 / L_b = D_2 \pi \, n_2 / L_b \tag{1}$$

where n_1, n_2 are the rotational speeds of the sheaves, D_1, D_2 are the sheaves pitch diameters, and L_b is the belt length. $f_b = 0.18 \, n_2$ and n_2 is about 12.5 Hz (measured by the tachometer). At least 6 revolutions of the belt should be contained in an analyzed block data. The block data lasts for 5 s. The sheave-timing belt mesh frequency (Hz) is defined in Equation (2).

$$f_{mesh} = z_2 n_2 \tag{2}$$

The sixth fault (class C6) is an increased friction load obtained by the extra stretch of the helical spring with 10 mm from 55 mm to 65 mm. A small friction load at the shaft level is maintained for all states except for the class C6 where the friction force is slightly increased. For the seventh fault (class C7) the belt is mounted vice versa with the teeth in exterior (assembling error).

The proposed faults are modifying the dynamics of the mechanical system; hence, each fault changes the vibration signal in a particular manner at a particular measuring spot; some effects are briefly mentioned in the sequel. An increased eccentric mass on the central disc will increase the first order peak of the frequency domain (C1 fault). The translation of the central disc toward a bearing support will increase the load on that support (C2 and C3 faults). A translation or rotation of the motor reducer will affect the timing belt tension and the friction belt sheaves, adding axial forces on the shaft (classes C4 and C5); this increases the importance of the axial (Y) axis of the vibration signal. The increased friction load (C6) while kipping the constant speed will increase tension on the mechanical structure and the timing belt reversal (C7) will modify the tension in the belt and the power spectrum at the mesh frequency. Analytical and numerical models of the system, including the effect of the proposed faults, could reveal important aspects of the acceleration at the measuring location by the accelerometer and could be an alternative for far less practical and less performant fault detection methods.

The shaft rotation speed (n_2) during the experiments is 750 rpm (12.5 rps), with small variations about the nominal value. Being relatively far below the first critical speed, the behavior of the system is considered to be unaffected by the first flexural mode of vibration of the shaft (even when the central mass m is slightly modified by adding about one gram at the C1 class observed in the study).

One hundred (100) observations are recorded for each of the eight health states. Each registration lasts for five seconds and gathers 25,000 samples for a sampling frequency of 5000 samples per second.

3. Multi-Fault Analysis and Detection System

The present approach proposes a system that analyzes, models and solves the fault detection task as a multi-class classification problem. The vibration signals measured for all eight health states (classes) are processed, and features (parameters) are extracted and used as further predictors in the classification process. The architecture of the proposed multi-fault analysis and detection system, depicted in Figure 3, is composed of seven stages briefly presented below and described in detail in the following sections.

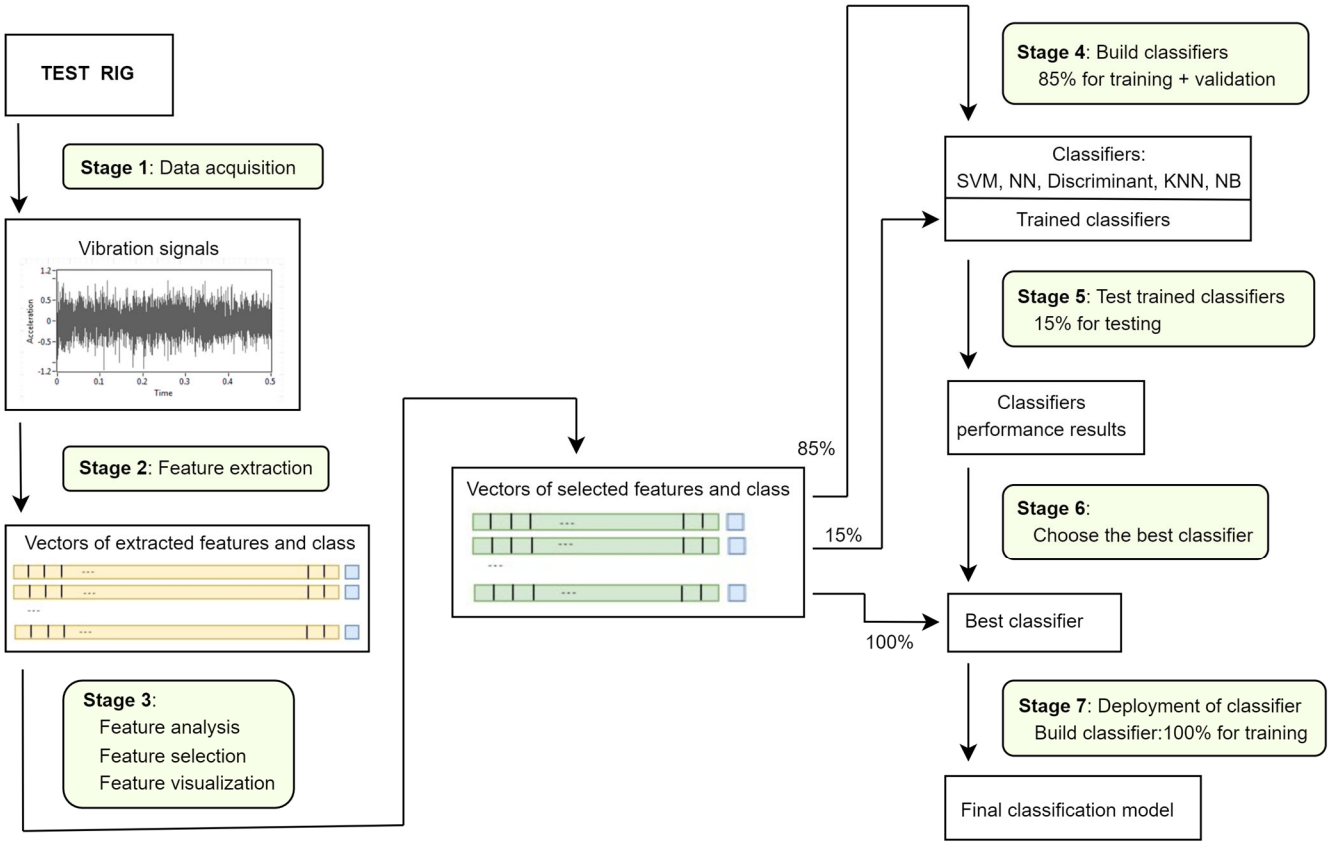


Figure 3. Architecture of the multi-fault analysis and detection system.

1. Data acquisition: a LabVIEW application is considered to measure the vibration signals (Section 2.1).
2. Feature extraction: the vibration signals are processed using a LabVIEW program and statistical time-domain and frequency-domain features are extracted. The dataset used further contains a vector of features and the corresponding health state (class) for each observation.
3. Feature analysis and selection: unsupervised and supervised techniques implemented in MATLAB are applied to explore and visualize the dataset of extracted features with the aim of identifying the most informative features that will be used as predictors in classification; several feature sets are proposed.
4. Build classifiers: a set of classifiers [19] (*support vector machine*—SVM, *neural networks*—NN, *k-Nearest Neighbor*—kNN, *discriminant-based classifiers*, *Naïve Bayes*—NB) from MATLAB are trained using 85% instances from the dataset for different feature sets.

5. Test trained classifiers: the classifiers built at the previous stage are tested using the 15% unseen instances from the dataset.
6. Choose the best classifier: the best classification model (classifier type + feature set) is selected, based on the ac-curacies of all employed classifiers and the feature sets.
7. Deployment of classifier: the best classifier is trained on the full dataset restricted to the selected set of features, the final classification model is built and then exported to be used for future predictions of faults.

4. Feature Extraction

To distinguish the eight health states, an initial set of twenty-one parameters (features) are proposed and extracted from the measured observations employing a LabVIEW application. The first subset of parameters contains time-domain parameters, and the other parameters are from the frequency domain.

The first parameter is the average of the shaft speed n_2 (rotations per second, Hz) for each time window of five seconds, the duration of a registration.

The next parameters are time-domain statistical parameters: median value, crest factor, peak-to-peak of the signal, mean value (μ), mean of absolute values, standard deviation (σ), variance (σ^2), root mean square (RMS), 3rd moment about mean, 4th moment about mean, skewness and kurtosis, all from the horizontal (radial, X axis) and axial (Y axis) accelerometer time signal. RMS includes the mean value of the signal. When the mean value of the signal is zero, RMS equals standard deviation. Crest factor is the ratio of the maximum absolute value of the signal to the RMS value. For a signal with N sample values the mth order moment is calculated using Equation (3). For m = 2, the moment about the mean (μ) equals the population variance (σ^2). Skewness (x_{SK}), defined in Equation (4), expresses the asymmetrical behavior of the vibration signal through its probability density function (PDF). For a negative value of the skewness the left side of the probability density function graph is longer than the right side, while for a positive value the right side of the probability density function graph is longer than the right side. For a symmetric signal the skewness is zero. Kurtosis (x_{KU}) is defined in Equation (5) and expresses the peak measurement (peakedness) of the distribution of the input sequence. Normal distribution has a kurtosis value of 3. A value less than 3 indicates a flatter distribution than normal, whereas a value greater than 3 indicates a sharper distribution than normal.

$$\sigma_x^m = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^m \tag{3}$$

$$x_{SK} = \sum_{i=1}^{N}(x_i - \mu)^3 / N\,\sigma_x^3 \tag{4}$$

$$x_{KU} = \sum_{i=1}^{N}(x_i - \mu)^4 / N\,\sigma_x^4 \tag{5}$$

The peak (PS1) of the power spectrum in a narrow interval of spectral lines about the shaft rotational frequency n_2 (Hz) is the first spectral parameter. PS1 is followed by three harmonics (PS2, PS3, PS4). Similarly, the harmonics are obtained by taking the peaks of the power spectrum in a narrow interval centered at two times, three times and four times of PS1.

Four spectral features [20], the spectral centroid (C), spectral spread (S), spectral skewness and spectral kurtosis, are added. C is a frequency-weighted sum normalized by the unweighted sum (Equation (6)). S is the standard deviation around the spectral centroid (Equation (7)). A_k is the kth power spectral coefficient magnitude and f_k is the associated frequency. Spectral skewness is computed from the third order moment and spectral kurtosis is calculated from the fourth order moment. The last parameter is the power spectrum peak, at the driven sheave timing belt mesh frequency.

$$C = \sum_{k=1}^{N/2} f_k A_k / \sum_{k=1}^{N/2} A_k \tag{6}$$

$$S = \sqrt{\sum_{k=1}^{N/2} (f_k - C)^2 A_k / \sum_{k=1}^{N/2} A_k} \quad (7)$$

The analysis time window is 5 s and the sampling rate (f_s) is 5 kHz. The spectrum resolution (interval between adjacent spectral lines) is $\Delta f = f_s/N = 0.2$ Hz, where N = 25,000 samples.

The features extracted from the vibration signals measured on the horizontal direction (X accelerometer axis) and the axial direction (Y accelerometer axis) are presented in Table 2.

Table 2. List of extracted features.

Feature Description	**Feature Notation**		**Feature Description**		**Feature Notation**	
shaft rotation frequency	F1					
Time-Domain Features	***X* Axis**	***Y* Axis**	**Frequency-Domain Features**		***X* Axis**	***Y* Axis**
median	F2	F22	Power Spectrum	PS1	F13	F33
peak to peak	F3	F23		PS2	F14	F34
mean of absolute values	F4	F24		PS3	F15	F35
mean value	F5	F25		PS4	F16	F36
standard deviation	F6	F26	spectral centroid		F17	F37
root mean square	F7	F27	spectral spread		F18	F38
crest factor	F8	F28	spectral skewness		F19	F39
3rd momentum	F9	F29	spectral kurtosis		F20	F40
4th momentum	F10	F30	PS at mesh frequency		F21	F41
skewness	F11	F31				
kurtosis	F12	F32				

The result of the feature extraction stage is a balanced dataset, a total of 800 vectors (instances) used in data analysis and classification. For each of the eight health states (classes), 100 vectors with 42 elements (41 features and the class) are generated. This initial dataset will be further analyzed using unsupervised and supervised techniques as a prior step to classification.

5. Data Analysis

An in-depth data analysis, with the goal of obtaining significant insights (relevance of features, correlation between features, classification difficulties, data visualization) into the initial dataset, is an important stage prior to classification modelling. The analysis stage is conducted on two subsets of the extracted features: *Sx_21* = {F1, . . . , F21} containing all 21 features on X axis and *Sxy_41* = {F1, . . . , F41}, all 41 features on X and Y axes of the accelerometer. The objective is to select the most informative features and use them further in the multi-class classification task that solves the proposed multi-fault detection problem.

5.1. Feature Analysis and Selection

A feature standardization [21] preprocessing step transforms data with high differences between their ranges into new features with a mean of zero and a standard deviation of one.

In the feature selection process the most useful, relevant features, from those extracted, are identified, with the aim of decreasing the computational complexity and improving the performance of machine learning models. Even if the dimensionality of the feature space is not so high (21 or 41 features), it is important to find a reduced subset of independent features that still accurately describes the fault detection problem. In the present approach the feature selection is based on two analysis steps: (1) scoring the features based on their

relevance as predictors in classification and (2) finding linear/monotonic dependencies between features.

Relief-based algorithms [22,23] deal with the estimation of the quality of the features used in classification tasks. They are distance-based, requiring feature standardization, and are applied in a supervised analysis processing step prior to modeling. The basic idea is to identify how well the features' values distinguish between neighboring instances in the feature space. The features' weights (scores) are in the interval [−1,1] and are calculated in an iterative process. The scores are updated based on feature value differences between an instance and the *k*-nearest neighbors from its class and also the *k*-nearest neighbors from each of the other classes. The weight of a feature decreases/*increases* when it has differences in values for neighboring instances of the same/*different* class. The features with negative weights are considered irrelevant and positive weights are assigned to informative, relevant features. The target is the ranking of the features according to their relevance to the output/predicted results.

The MATLAB function *relieff* works for multi-class problems and can deal with incomplete and noisy data. In this approach it is applied on standardized data and with the parameter $k = 10$. The provided result is the ranking of the initial features. The lists of features in the decreasing order of their relevance (importance) as predictors in classification:

- for *Sx_21*: *F14, F13, F21, F5, F4, F1, F7, F6, F18, F12, F10*, F19, F17, F20, F16, F8, F3, F15, F2, F9, F11.
- for *Sxy_41*: *F4, F7, F6, F37, F39, F38, F13, F40, F10, F18, F27, F26, F25, F30, F28, F24, F1, F17, F19, F14, F32, F5, F21, F20, F23, F12, F33, F34*, F8, F16, F41, F3, F9, F11, F2, F22, F36, F29, F31, F15, F35.
- for *Sy_21*: *F38, F1, F37, F32, F39, F28, F25, F36, F40, F23, F26, F27, F41, F24, F30*, F34, F33, F31, F22, F29, F35.

In the lists, the features with positive weights are in italics. These represent the most informative features and will therefore be good candidates as predictors in classification.

After finding the most relevant features, the next step is to restrict them to a subset of independent features. A statistic-based feature analysis using Pearson correlation coefficient aims to identify linear relationships between pairs of features. The symmetric matrix of the Pearson coefficients [24] provided by the *correcoef* function in MATLAB for *Sxy_41* set is presented in Figure 4 in the form of a heatmap. The results are the following:

- very strong positive linear dependencies between any two features from the set {F4, F6, F7, F10} with coefficient values greater than 0.97.
- *Pearson_coef*(F9, F11) = 0.99 and *Pearson_coef*(F17, F18) = 0.983, suggesting a very strong positive correlation between the features from the corresponding pairs.
- F19 and F20 are negatively correlated based on the Pearson coefficient with a value of −0.922.
- similar correlation relations are identified for the same pairs of features but extracted from *Y* axis: *Pearson_coef*(*a*,*b*) > 0.97 for any *a*,*b* ϵ {F24, F26, F27, F30}, *Pearson_coef*(F29, F31) = 0.97, *Pearson_coef*(F37, F38) = 0.98, and *Pearson_coef*(F39,F40) = −0.835.
- no other significant correlations (values in $(-1,-0.8) \cup (0.8, 1)$) were obtained.

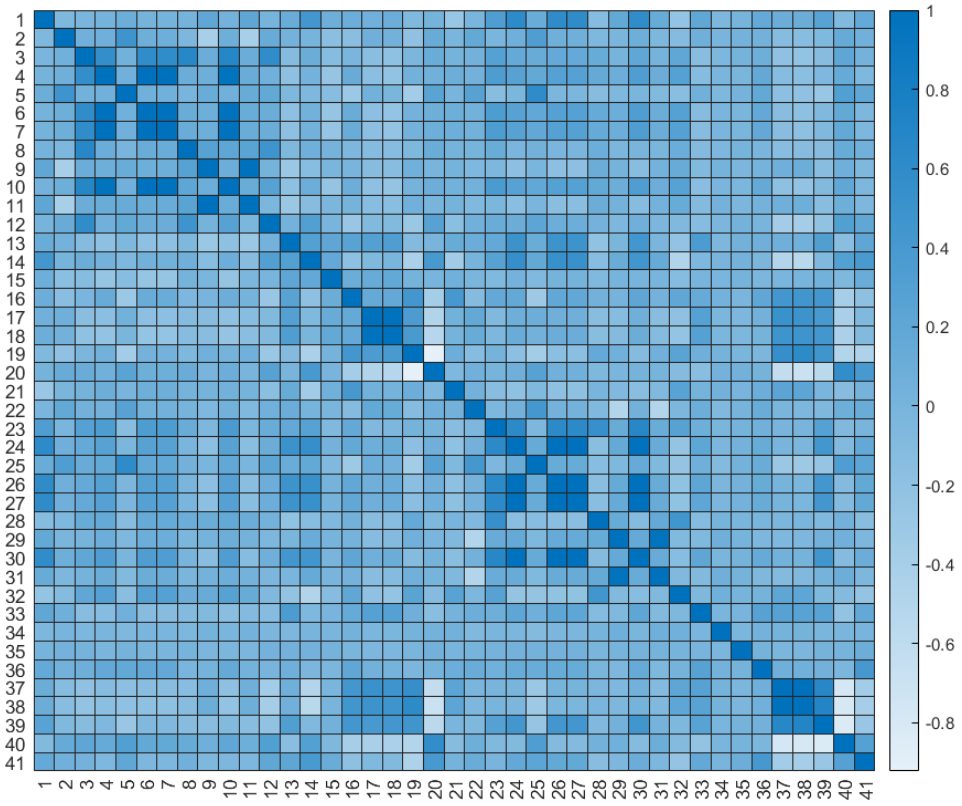


Figure 4. Pearson coefficients for *Sxy_41*.

Based on the previous analysis steps, the feature selection begins with the elimination of the negative weighted ones. For the remaining features, according to the above correlations, from each group of correlated features the most important one is kept, and the others are discarded. Thus, F18 is chosen from {F17, F18} and from the group {F4, F6, F7, F10}, F4 is selected as being the most informative feature according to the ranking provided by the *relieff* function for both, *Sx_21* and *Sxy_41*. The set {F24, F26, F27, F30} is reduced to F27, the feature F19 is chosen from {F19, F20}, F37 from {F37, F38} and F39 from {F39, F40} in *Sxy_41*. The reduced subsets of selected features containing the most relevant and independent features are as follows:

- from *Sx_21*, 8 features are selected: *Sx_sel8* = {F1,F4,F5,F12,F13,F14,F18,F21}.
- from *Sxy_41*, 18 features are selected:

$$Sxy\_sel18 = Sx\_sel8 \cup \{F19,F23,F25,F27,F28,F32,F33,F34,F37,F39\} = \{F1,F4,F5,F12,F13,F14,F18,F19,F21,F23,F25,F27,F28,F32,F33,F34,F37,F39\}$$

- from *Sxy_sel18*, four features (F12, F23, F33, F34) with very small positive weights are eliminated and 14 features are selected:

$$Sxy\_sel14 = \{F1,F4,F5,F13,F14,F18,F19,F21,F25,F27,F28,F32,F37,F39\}$$

To compare the importance of the features extracted from *X* axis versus those on *Y* axis and the frequency-domain versus the time-domain features, the weights provided by *relieff* are used. Let $W_{freq}(fs)$, $W_{time}(fs)$, $W_x(fs)$ and $W_y(fs)$ be the sums of the weights assigned to the features of the set "fs" with the properties from frequency-domain and from time-domain, extracted from the *X* axis and the *Y* axis, respectively. W(F1) is the weight of the feature F1-shaft rotation frequency. In Table 3 the values express the importance (%) of the features with the corresponding properties in the selected feature sets.

Table 3. Importance of features.

	Selected Feature Set (fs)		
	Sx_sel8	***Sxy_sel18***	***Sxy_sel14***
W(F1)	6.75%	5%	5%
W_{freq}(fs)	70.5%	54%	55%
W_{time}(fs)	22.75%	41%	40%
W_x(fs)	93.25%	48.5%	49.5%
W_y(fs)	–	46.5%	45.5%

After this supervised feature selection process, prior to classification, the following conclusions can be drawn: (1) in all three selected feature sets the importance of frequency-domain features prevail over the time-domain features; (2) the selected features from *X* axis are slightly more relevant than those selected from *Y* axis.

5.2. Data Exploration and Visualization

The initial dataset can be explored with a very intuitive and simple unsupervised method using the *classification difficulty* [25] per class. Informally this measure reveals how difficult it is to detect a class in that dataset and provides hints regarding the future classification performance. Based on the representations of the instances in an *n*-dimensional space of a subset of features, the Euclidian distance is used to calculate the neighbors of an instance. CD(fs,c) is called the classification difficulty of the class "c", in the dataset restricted to the feature set "fs" and is calculated using Equation (8):

$$CD(fs, c) = d_c^{fs}/n_c \quad (8)$$

where d_c^{fs} counts the instances of "c" with the nearest neighbor in the feature space defined by "fs" pertaining to another class and n_c is the total number of instances of class "c". The dataset contains 100 (n_c) instances for each class.

The CD(fs,c) value is a percentage representing how many instances of "c" are not in the nearest neighborhood (in the feature space "fs") of its class, out of all instances of "c", therefore suggesting that these instances might be misclassified later in the classification process. The higher the CD(fs,c), the less distinguishable is "c" in the feature space "fs". This measure helps to compare the classification difficulties of different classes in the same feature space and also the classification difficulties of a class in different feature spaces. The initial dataset is a balanced one, so an overall classification difficulty per feature set can be defined: CD(fs) = average{CD(fs,c) | c ϵ {C0, . . . ,C7}}.

Table 4 presents the classification difficulties per classes and per feature sets and the results are as follows:

- Using only the features extracted from *X* axis (*Sx_21*) the classes C0, C3, C4, C5, C7 have classification difficulties ≤ 0.03. Therefore, in this feature space at most 3% of the instances of these classes might be misclassified and 97% will be correctly classified. By applying performant classification algorithms better classification results (higher than 97%) are expected. The classes C1 and C6 have difficulties slightly higher than the other classes, so for these classes a slight drop in classification performance is expected. The least detectable class in the feature space *Sx_21* is C2.
- The classification difficulties in the feature space *Sy_21* are higher than the difficulties in the other feature spaces for all classes, except for the C7 fault. Thus, the lowest classification results are expected if only the features extracted from *Y* axis are used as predictors.
- Adding the features on *Y* axis to the features on *X* axis decreases the classification difficulties for C2 and C6 and thus improving the classification results.

- Comparing the CD(fs) for all feature sets one note that the feature selection step proved to be beneficial in decreasing the classification difficulties.

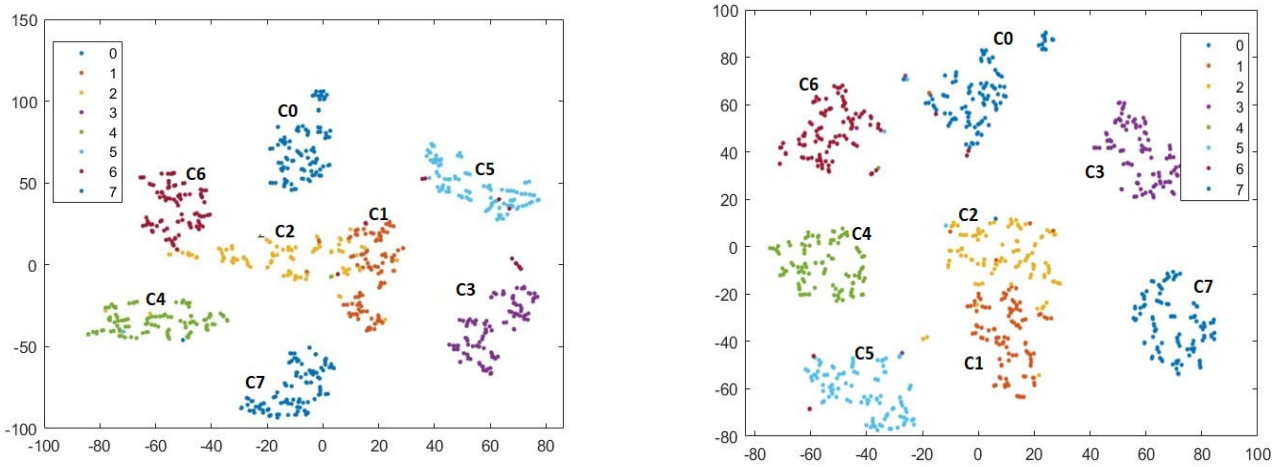
Table 4. Classification difficulties per classes and per feature set.

Feature Set (fs)	Classification Difficulties: CD (fs,c)								CD (fs)
	Classes (c)								
	C0	C1	C2	C3	C4	C5	C6	C7	
Sx_21	0.00	0.07	0.21	0.03	0.01	0.01	0.10	0.03	0.0575
Sy_21	0.35	0.19	0.25	0.12	0.18	0.40	0.21	0.03	0.2163
Sxy_41	0.00	0.01	0.12	0.02	0.02	0.06	0.05	0.01	0.0475
Sx_sel8	0.00	0.04	0.15	0.00	0.01	0.01	0.05	0.01	0.0338
Sxy_sel18	0.02	0.10	0.05	0.03	0.01	0.04	0.04	0.01	0.0375
Sxy_sel14	0.00	0.10	0.10	0.02	0.01	0.01	0.05	0.01	0.0378

These conclusions provide a naïve overview of the multi-class classification task and a graphical representation of the dataset in different feature spaces would be useful.

An algorithm for dimensionality reduction, *t-Distributed Stochastic Neighbor Embedding (t-SNE)* [26], is applied to visualize the input data. This is an unsupervised non-linear technique that embeds high-dimensional points into low dimensions. The neighborhood relation between points in the initial space is preserved in the embedded space, so natural clusters in the high-dimensional space are visualized in a low-dimensional space (2D or 3D). The graphical representations in the low-dimensional space help to check how separable (detectable) are the classes in the initial space.

The feature spaces chosen for a comparison at the graphical level are *Sx_sel8* and *Sxy_sel18*, because they have the lowest classification difficulties (Table 4). Figure 5a,b depict 2D *t-SNE* projections (Euclidian distance) of *Sx_sel8* and *Sxy_sel18* respectively. There are no notable differences in these graphical representations. In both feature spaces the best detectable classes are C0, C3, C4, C5, C7, and the classes C1 and C2 are not so well separable. In *Sxy_sel18* the class C6 is better detectable than in the *Sx_sel8* feature space. These remarks are consistent with the conclusions based on the classification difficulties and they will be compared with the multi-class classification results.



(**a**) (**b**)

Figure 5. 2D *t-SNE* projections of the dataset in the feature spaces: (**a**) *Sx_sel8*; (**b**) *Sxy_sel18*.

6. Classification Results and Discussions

Different classifiers (SVM, NN, kNN, NB, discriminant-based) from MATLAB 2022a have been employed in experiments with the aim of solving the proposed multi-fault detection problem.

The input dataset containing 800 observations is a balanced one, with 100 observations for each class (health state). A classification model is built using the *training + validation subset*, 85% of the initial dataset, 680 instances (85 from each class) and a five-fold cross-validation to prevent the overfitting. From the total of 800 initial instances, the remaining 120 instances (15% from each class) represent the *testing subset*. In the testing stage the performance of a classification model is reported using *accuracy*, which is an overall evaluation metric expressing the probability that an arbitrary observation (instance) is correctly classified.

For a comparison of the classification results the experiments were conducted on six feature sets. In the first three experiments all the features extracted separately from *X* axis (*Sx_21*), *Y* axis (*Sy_21*) and both axes (*Sxy_41*) of the accelerometer were used as predictors in the classification process. The next three experiments were performed to check whether the reduced subsets of selected features (*Sx_sel8*, *Sxy_sel18* and *Sxy_sel14*) still accurately describe the fault detection problem.

The classification results achieved by the employed classifiers are presented in Table 5 as a confidence interval ($CI = ma \pm 1.96\ \sigma/\sqrt{n}$) [27] at the 95% confidence level. Ten runs (n = 10) were performed for each classifier and for all subsets of features. In each run the initial dataset was randomly split in *training + validation* subset (85% from each class) and *testing* subset (15% from each class). A mean accuracy (*ma*), and a standard deviation (σ) for the 10 test accuracies were calculated.

Table 5. Classification results.

Classifier Type	Classifier Model	Feature Subset *CI*(%) of the Mean Accuracy over 10 Runs					
		***Sx_21* *CI* (%)**	***Sx_sel8* *CI* (%)**	***Sy_21* *CI* (%)**	***Sxy_41* *CI* (%)**	***Sxy_sel18* *CI* (%)**	***Sxy_sel14* *CI* (%)**
SVM	Quadratic	98.09 ± 0.6	98.08 ± 0.78	89.75 ± 2.01	99.18 ± 0.49	98.93 ± 0.69	98.93 ± 0.43
	Cubic	97.26 ± 0.64	97.34 ± 0.77	88.92 ± 2.47	98.93 ± 0.55	98.67 ± 0.65	97.83 ± 0.66
	Linear	97.93 ± 1.04	97.66 ± 1.06	89.84 ± 1.32	98.66 ± 0.7	98.34 ± 0.65	98.17 ± 0.73
	Medium Gaussian	97.76 ± 0.6	97.99 ± 0.79	88.35 ± 2.26	98.58 ± 0.6	98.16 ± 0.91	97.83 ± 0.78
Discriminant	Quadratic	98.17 ± 0.91	98.08 ± 0.88	92.57 ± 2.19	98.03 ± 0.69	97.82 ± 0.56	98.16 ± 0.47
	Linear	97.41 ± 0.99	96.06 ± 0.86	94.29 ± 1.17	98.5 ± 0.54	97.84 ± 0.77	97.59 ± 0.79
Ensemble	Subspace discriminant	97.13 ± 0.96	95.71 ± 1.03	91.09 ± 2.07	98.73 ± 0.56	98.09 ± 0.67	97.77 ± 0.65
NN	Wide	97.59 ± 0.56	97.00 ± 0.88	86.26 ± 3.05	98.93 ± 0.27	97.99 ± 0.74	98.24 ± 0.57
	Medium	97.49 ± 0.77	96.76 ± 1.11	87.51 ± 3.02	98.57 ± 0.55	97.84 ± 0.6	98.42 ± 0.62
kNN	Weighted	97.18 ± 0.87	96.49 ± 0.86	86.00 ± 3.18	98.15 ± 0.8	96.42 ± 0.91	96.17 ± 1.3
NB	Kernel	97.08 ± 0.69	96.56 ± 0.9	86.07 ± 4.75	98.00 ± 0.83	96.97 ± 1.57	97.38 ± 1.21

The configurations of the classifiers are: (1) SVM–kernel functions: linear/quadratic/cubic/medium Gaussian, box constraint level: 1, standardized data, multiclass method: one-vs-one; (2) Discriminant quadratic/linear–covariance structure: full; (3) Ensemble–subspace discriminant, 30 learners; (4) neural network wide and medium–first fully connected layer: size 100 and 25 respectively, iteration limit: 1000, activation function: ReLU, final fully connected layer with eight (number of classes) neurons and Softmax function for prediction, standardized data; (5) kNN weighted–number of neighbors: 10, distance metric: Euclidean, distance weight: squared inverse, standardized data; and (6) Naïve Bayes–distribution: kernel.

Based on the performance results of the classifiers the conclusions are the following:

- All employed classifiers working with all 21 features extracted from *X* axis have a mean accuracy > 97%, quadratic SVM and quadratic discriminant achieving the best accuracies > 98%.
- From the initial 21 features on *X* axis, the most relevant, eight, were selected and used for prediction and the accuracy results are very close to those obtained for 21 features; for the SVM classifiers and the quadratic discriminant the difference is less than 0.27%.
- The features from *Y* axis are less effective predictors than the similar features on *X* axis and the best results are obtained by the discriminant-based classifiers with mean accuracy between 91–94.3%.
- A small improvement of accuracy (in the interval (0.7, 1.6)) is obtained when the classifiers use all 41 features extracted from *X* and *Y* axes. All mean accuracies > 98%, quadratic SVM being the most accurate, with an accuracy of 99.18%, and then cubic SVM and wide NN with an accuracy of 98.93%.
- The 18 or 14 selected features from *Sxy_41* proved to be very good predictors in almost all classification models. SVM classifiers were the best suited for solving the fault detection task in these feature subsets, with accuracies slightly lower (<0.42%) than for 41 features.
- The very good classification accuracies are consistent with the results obtained in the analysis stage: very low classification difficulties per classes and per feature sets and a good detectability of classes in the *t-SNE* representations.

In Table 6 are depicted the test confusion matrices in one run (one random split of the dataset: 85% *training + validation*, 15% *testing*) from 10 runs performed, for two classification models based on quadratic SVM using as feature sets: (a) *Sx_sel8* and (b) *Sxy_sel18*. Excellent results were obtained in both feature spaces, accuracy of 98.3% and 99.2%, respectively. It should be noted that the eight selected features on *X* axis manage to discriminate perfectly the instances from the classes C0, C2, C3, C4, C5, C7, but C1 and C6 are confused with C2, a result that is consistent with Figure 5a where C1, C2 and C6 are not very well separable. In the feature space *Sxy_sel18* the *t-SNE* representation (Figure 5b) shows that class C6 is well separable, a result that is validated by the confusion matrix in Table 6b.

Table 6. Confusion matrices for quadratic SVM: (a) *Sx_sel8* feature set; (b) *Sxy_sel18* feature set.

120 Tested Instances	**(a)** *Sx_sel8*							
	Predicted Class							
True Class	**C0**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
C0	**15**							
C1		**14**	**1**					
C2			**15**					
C3				**15**				
C4					**15**			
C5						**15**		
C6			**1**				**14**	
C7								**15**

Accuracy
= (15 + 14 + 15 + 15 + 15 + 15 + 14 + 15)/120 = **0.983**

120 Tested Instances	**(b)** *Sxy_sel18*							
	Predicted Class							
True Class	**C0**	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
C0	**15**							
C1		**14**	**1**					
C2			**15**					
C3				**15**				
C4					**15**			
C5						**15**		
C6							**15**	
C7								**15**

Accuracy
= (15 + 14 + 15 + 15 + 15 + 15 + 15 + 15)/120 = **0.992**

In the deployment stage the classification model based on quadratic SVM was trained on the full dataset restricted to the feature set *Sxy_sel18* and a final model was built. This option was based on the idea of choosing less features as predictors and still achieved a very good accuracy.

In the proposed fault detection task, data patterns usually remain constant, meaning that the statistical properties of the defects (classes) do not change significantly over time.

In this case an offline training (based on an initial dataset) was the best solution and then the final model could be used for online predictions on the test rig. If some changes in the statistical properties of the defects occur, then the overall performance (accuracy) in prediction is affected. Periodically the consistent results of the predictions can be used to update the model, adapt it to the new data (delivered in batches) and a new model can be generated.

7. Conclusions

In this paper certain mounting defects of a rotating test rig are investigated by using a proposed data-driven multi-fault analysis and detection system. The test rig assembly comprises a slender shaft supported by oscillating ball bearings with a central disc. The shaft is driven by a DC motor reducer through a timing belt and a small variation in the shaft speed during experiment about the nominal constant value. Faults split in seven classes are imposed on the test rig, such as: disc eccentric mass, two-disc incorrect positioning along the shaft, two motor reducer positioning (rotation, translation), erroneous belt mounting and an unexpected increased load on the shaft. An accelerometer records the vibration on a bearing support, close to the shaft drive system. Twenty-one features on X axis and forty-one features on the combined X and Y axes, on time and frequency domains, were extracted from the accelerometer signal and the tachometer. Six sets of features (three full and three selected feature sets) were gathered and used in the classification process that solves the proposed multi-fault detection problem.

The contributions and conclusions of the paper are presented below.

- A data-driven multi-fault analysis and detection system using machine learning techniques with MATLAB was proposed.
- A data analysis stage was completed that aimed at obtaining important insights (relevance of features, correlation between features, classification difficulties, data visualization) into the initial dataset containing statistical features from time and frequency domains extracted from X and Y axes of the accelerometer. Using unsupervised and supervised analysis methods, from 41 extracted features three subsets of features (*Sx_sel8*, *Sxy_sel18* and *Sxy_sel14*) were selected and they proved to be good predictor candidates for the multi-class classification task. In these feature spaces the classes C0, C3, C4, C5 and C7 are very easily detectable, but C1, C2 and C6 are not very well separable.
- A comparison of the importance of the features in the three selected feature spaces provided the results: the frequency-domain features prevail over the time-domain features and the selected features from X axis are slightly more relevant than those selected from Y axis.
- In the classification process all the employed classifiers trained on the dataset restricted to the 21 features extracted from radial X axis achieved excellent results, with a mean accuracy > 97%. The performance results were slightly improved by 0.7% to 1.7% by adding all 20 features from Y axis, a total of 41 predictors. The best mean accuracy, 99.18%, was obtained by quadratic SVM for all 41 features and 98.08% for 21 features extracted from X axis.
- For the proposed set of faults, the features from the radial X axis are more relevant than those on the axial Y axis, as was assumed. The best classification results obtained using only the features on Y axis were provided by the discriminant-based classifiers with mean accuracy in the interval (91%, 94.3%).
- The three selected feature subsets (*Sx_sel8*, *Sxy_sel18* and *Sxy_sel14*), containing the relevant features proved to accurately describe the health states, the classification results are very close to those obtained for the full sets of features. At the level of classes, C0, C3, C4, C5 and C7 are always correctly classified. Prediction errors, not more than 2%, occur for C1, C2 and C6 classes in *Sx_sel8*. The features on Y axis in *Sxy_sel18* and *Sxy_sel14* manages to detect C6 but still has some problems with C1

and C2 classes which could come from the shaft speed variation during observations. These results validate the conclusions in the analysis stage.

- From the set of all employed classifiers the best performance results (accuracies) were achieved by the SVM classifiers and the neural networks for the full and selected feature sets.
- For the deployment stage the option was to choose the best classifier, quadratic SVM, and *Sxy_sel18* feature set, with a mean accuracy of 98.93%. This final classification model is best suited for solving the proposed multi-fault detection problem and will be used for future fault predictions.

The proposed system could be used to detect assembling defects at a production line test station where the component under test contains parts similar to those in the test rig. Further work should be oriented so as to develop the proposed analysis and a detection system for a real production equipment. Joint time-frequency analyses for variable speed devices containing extra gear faults are in progress.

Author Contributions: Conceptualization, I.L. and M.L.; methodology, I.L. and M.L.; validation, I.L. and M.L.; formal analysis, I.L. and M.L.; investigation, I.L. and M.L.; resources, I.L. and M.L.; data curation, I.L. and M.L.; writing—original draft preparation, I.L. and M.L.; writing—review and editing, I.L. and M.L.; visualization, I.L. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The experimental vibration data is available under request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Umbrajkaar, A.M.; Krishnamoorthy, A.; Dhumale, R.B. Vibration Analysis of Shaft Misalignment Using Machine Learning Approach under Variable Load Conditions. *Shock. Vib.* **2020**, *4*, 1–12. [CrossRef]
2. Ahmed, H.; Nandi, A.K. *Condition Monitoring with Vibration Signals—Compressive Sampling and Learning Algorithms for Rotating Machines*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2020. [CrossRef]
3. Brandt, A. *Noise and Vibration Analysis*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2011.
4. Scheffer, C.; Girdhar., P. *Practical Machinery Vibration Analysis and Predictive Maintenance*; Elsevier: Amsterdam, The Netherlands, 2004.
5. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its application to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
6. Randall, R.B. *Vibration-Based Condition Monitoring—Industrial, Aerospace and Automotive Applications*; John Wiley & Sons Ltd.: Hoboken, NJ, USA, 2011.
7. Caesarendra, W.; Tjahjowidodo, T. A Review of Feature Extraction Methods in Vibration-Based Condition Monitoring and Its Application for Degradation Trend Estimation of Low-Speed Slew Bearing. *Machines* **2017**, *5*, 21. [CrossRef]
8. Jardine, A.K.S.; Lin, D.; Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech. Syst. Signal Process.* **2006**, *20*, 1483–1510. [CrossRef]
9. Rodrigues, C.E.; Nascimento, C.L.; Rade, D.A. Machine Learning Techniques for Fault Diagnosis of Rotating Machines Using Spectrum Image of Vibration Orbits. In Proceedings of the Annals of XXII Congresso Brasileiro de Automática (CBA), virtual, São Paulo, Brazil, 23–26 December 2020; Volume 2, pp. 1–7. [CrossRef]
10. Li, W.; Wang, Z.; Zhu, Z.; Zhou, G.; Chen, G. Design of Online Monitoring and Fault Diagnosis System for Belt Conveyors Based on Wavelet Packet Decomposition and Support Vector Machine. *Adv. Mech. Eng.* **2013**, *5*, 1–10. [CrossRef]
11. Xue, H.; Wang, H.; Chen, P.; Li, K.; Song, L. Automatic diagnosis method for structural fault of rotating machinery based on distinctive frequency components and support vector machines under varied operating conditions. *Neurocomputing* **2013**, *116*, 326–335. [CrossRef]
12. Zhao, W.; Hua, C.; Dong, D.; Ouyang, H. A Novel Method for Identifying Crack and Shaft Misalignment Faults in Rotor Systems under Noisy Environments Based on CNN, 19. *Sensors* **2019**, *19*, 5158. [CrossRef] [PubMed]
13. Jaber, A.; Ali, K.M. Artificial Neural Network Based Fault Diagnosis of a Pulley-Belt Rotating System. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2019**, *9*, 544–551. [CrossRef]
14. 20087_E2_Preventive_Maintenance_Manual. Available online: https://industrialbeltdrives.com/wp-content/uploads/2017/11/20087_E2_PREVENTIVE_MAINTENANCE_MANUAL.pdf (accessed on 15 May 2022).
15. Lupea, I.; Lupea, M. Fault Detection on a Rotating Test Rig based on Vibration Analysis and Machine Learning. *Proc. Rom. Acad. Ser. A* **2022**, *23*, 151–160.

16. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
17. Khan, A.; Hwang, H.; Kim, H.S. Synthetic Data Augmentation and Deep Learning for the Fault Diagnosis of Rotating Machines. *Mathematics* **2021**, *9*, 2336. [CrossRef]
18. Buzdugan, G.; Fetcu, L.; Radeş, M. *Vibraţiile Sistemelor Mecanice*; Editura Academiei RSR: Bucureşti, Romania, 1975.
19. Kubat, M. *An Introduction to Machine Learning*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2021.
20. Mathworks. Spectral Descriptors. Available online: www.mathworks.com/help/audio/ug/spectral-descriptors.html (accessed on 16 June 2022).
21. Han, J.; Kamber, M.; Pei, J. Data Transformation and Data Discretization. In *Data Mining: Concepts and Techniques*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2011; pp. 111–118.
22. Kira, K.; Rendell, L.A. The feature selection problem: Traditional methods and a new algorithm. *AAAI Assoc. Adv. Artif. Intell.* **1992**, *2*, 129–134.
23. Urbanowics, R.J.; Meeker, M.; Cava, W.; Olson, R.S.; Moore, J.H. Relief—Based Feature Selection: Introduction and review. *J. Biomed. Inform.* **2018**, *85*, 189–203. [CrossRef] [PubMed]
24. Tuffery, S. *Data Mining and Statistics for Decision Making*; John Wiley and Sons: Hoboken, NJ, USA, 2011.
25. Boetticher, G.D. Improving the Credibility of Machine Learner Models in Software Engineering. In *Advances in Machine Learning Applications in Software Engineering*; IGI Global: Hershey, PA, USA, 2007; pp. 52–72. [CrossRef]
26. Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
27. Brown, L.; Cai, T.; DasGupta, A. Interval estimation for a binomial proportion. *Stat. Sci.* **2001**, *16*, 101–133. [CrossRef]