Research Article

Rolling bearing fault feature selection based on standard deviation and random forest classifier using vibration signals

Moussaoui Imane(ID), Chemseddine Rahmoune(ID) and Djamel Benazzouz

Abstract

The precise identification of faults is vital for ensuring the reliability of the bearing's performance, and thus, the functionality of rotary machinery. The focus of our study is on the role that feature selection plays in improving the accuracy of predictive models used for diagnosis. The study combined the Standard Deviation (STD) parameter with the Random Forest (RF) classifier to select relevant features from vibration signals obtained from bearings operating under various conditions. We utilized three databases with different bearings' health states operating under distinct conditions. The results of the study were promising, indicating that the proposed method was not only effective but also consistent, even under time-varying conditions.

Keywords

Feature selection, standard deviation, random forest, optimization algorithm, bearing fault, diagnosis

Introduction

In recent decades, machine learning (ML) has been extensively involved in fault detection and classification problems.[1] However, the robustness of the trained model depends essentially on the quality and the quantity of the input features.[2] Therefore the optimization step becomes of tremendous potential accurately determine the perilous conditions.[3]

Feature selection (FS) arises as an essential step for the effectiveness of ML application,[4] the speed of the diagnosis process,[5] as well as for the enhancement of the predictive accuracy.[6]

Feature selection techniques fall into three main classes.[7] The first class is the filter method, which uses statistical methods to rank the features, and then removes the elements under a determined threshold.[8] This class provides a fast and efficient selection.[6] The

second class, called the wrapper class, treats the predictors as the unknown and the predictors' performance as the objective function,[8] the problem is reduced to the search algorithm.[9] Many subsets are randomly selected and then evaluated by a classifier, and the one with the maximum accuracy is picked.[8] The wrapper class is better than the filter class in terms of performance and accuracy. However, for exhaustive searching algorithms, it becomes computationally expensive.[8,10] The third type

Solid Mechanics and Systems Laboratory (LMSS), University M'hamed Bougara, Boumerdes, Algeria

Corresponding author:
Moussaoui Imane, Solid Mechanics and Systems Laboratory (LMSS), Faculty of Technology, University M'hamed Bougara, Boumerdes, Algeria, Frantz Fanon City 35000, Boumerdes, Algeria.
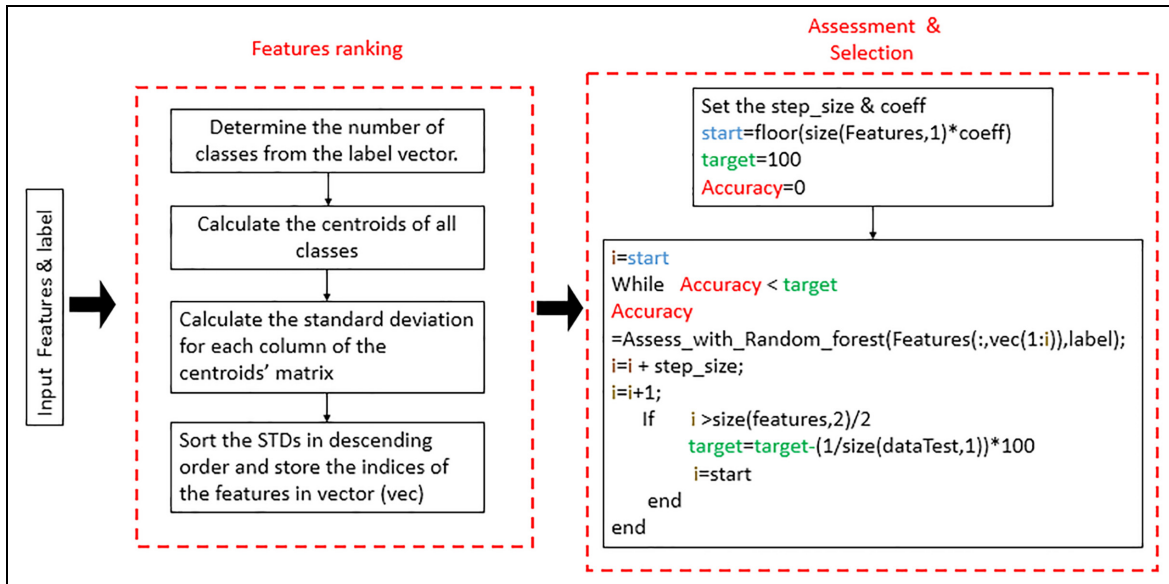Email: i.moussaoui@univ-boumerdes.dz

Figure 1. Flowchart of the proposed method.

is the hybrid or embedded class. It is a combination of the advantages of both filter and wrapper classes.[11]

Many feature selection methods are applied to the bearing fault diagnosis; provided good performances. In Peña et al.,[4] the analysis of variance (ANOVA) is used as a filter method to rank the features based on their relevance, then select the subset that yields the best accuracy through cluster validation assessment. This method provides a good classification, but it has some limitations that can be found in any real data. For example, it requires the number of samples from all classes to be equal,[12] which, whether by accident or necessity, is not always met. ANOVA requires a tiny variance within samples of the same class to be efficient.[12] However, in some cases, such as in Imane et al.,[13] the data used was collected under variable speed conditions, resulting in sparse samples of the same class. In addition to these constraints, ANOVA necessitates specific knowledge in order to interpret its results.

In Rajeswari et al.,[14] they used particle swarm optimization (PSO) for the feature selection. In Ma et al.,[15] ant colony optimization (ACO) performed the selection step. Both PSO and ACO added strength to the process of bearings' diagnosis by discarding the redundant features and preserving the relevant ones for the model training. However, PSO suffers from dimensionality issues and the demand for numerous evaluations to attain accurate results.[16] While the ACO suffers from local optimization problems.[17] In Imane et al.,[13] the cultural clan-based algorithm could select the relevant features efficiently within speed variability conditions and enhance classification accuracy. Yet its time

complexities can represent a limitation for large datasets.[18] All listed algorithms belong to the wrapper class and ensure high accuracy. However, the trade-off between the high performance and the slow execution is inescapable.[2]

In this paper, we propose a simple and efficient method for selecting the most relevant features to pave the way for a robust bearing diagnosis process. The idea came from the importance of the centroids to determine the classes. In our method, we aim to select the coordinates (the features) that cause the centroids spacing, and this can be verified by the standard deviation parameter, unlike ANOVA our method is based on the geometrical perspective and has no restrictions on the data in terms of quality or quantity. After ranking the coordinates of the centroids, random forest classifier (RF) selects the optimal subset that delivers the highest accuracy, to not rely on a distance-based classifier and ensures that the selected features are suitable for any classifier type.

The rest of the article is organized as follows: The second section describes our proposed method for using the SDT-RF selection method, and the third section represents the datasets used for testing as well as the results obtained. The final section serves as a general conclusion.

The proposed method

The flowchart in Figure 1 elucidates the method suggested for features selection used in the bearing diagnosis process. The following steps outline the proposed method.
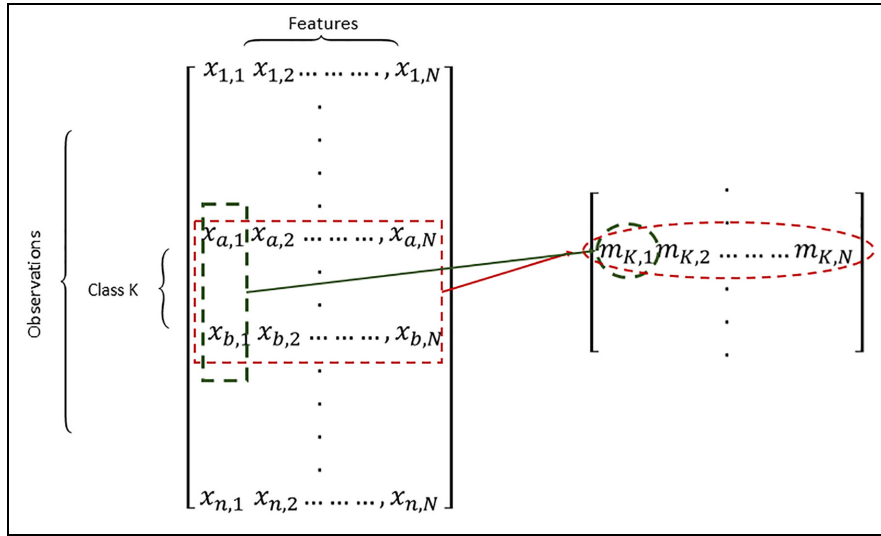
Figure 2. The calculation of the centroids' coordinates.

1. Determine the number of classes and their corresponding number of samples.
2. Calculate the centroid's coordinates for each class. C_K is the centroid of an arbitrary class K, and we calculate it as follows:

$$C_K = \frac{\sum Samples}{size(DataTest)} \quad (1)$$

Where:

$$Class_K = \begin{bmatrix} x_{1,1} & x_{1,2} & .... & x_{1,N} \\ x_{2,1} & x_{2,2} & .... & x_{2,N} \\ .... & .... & .... & .... \\ .... & .... & .... & .... \\ x_{l,1} & x_{j,2} & .... & x_{l,N} \end{bmatrix}$$

l is the size of samples in class K and N is the number of features.

we expand equation (1) into:

$$C_K = \frac{(x_{1,1}, ..., x_{1,N}) + ... + (x_{l,1}, ..., x_{j,N})}{l}$$

$$C_K = \left(\frac{x_{1,1} + ... + x_{l,1}}{l}, ..., \frac{x_{1,N} + ... + x_{l,N}}{l} \right)$$

$$C_K = \left(\frac{\sum_{j=1}^{l} x_{j,1}}{l}, ..., \frac{\sum_{j=1}^{l} x_{j,N}}{l} \right)$$

Then, the centroid's coordinates are equal to the means of the corresponding class's columns as shown in Figure 2.

3. Compute the standard deviation using equation (2) for each column of the centroids matrix.

$$Centroids = \begin{bmatrix} m_{1,1} & m_{1,2} & .... & m_{1,N} \\ m_{2,1} & m_{2,2} & .... & m_{2,N} \\ .... & .... & .... & .... \\ .... & .... & .... & .... \\ m_{p,1} & m_{p,2} & .... & m_{p,N} \end{bmatrix}$$

Where p is the number of classes,

$$STD_j = \frac{\sqrt{\sum_{i=1}^{p} (m_{i,j} - M_j)^2}}{p} \quad (2)$$

And,

$$M_j = \frac{\sum_{i=1}^{p} m_{i,j}}{p}$$

We obtain a vector S of N STD value.

$$S = \begin{bmatrix} STD_1 & STD_2 & .... & STD_N \end{bmatrix}$$

4. Sort the vector S in a descending order and save the indices of the corresponding features in vector V.
5. Execute a sequential forward selection on the indices' vector v and assess the performance of the corresponding features with the Random forest classifier. The process starts from '*start*' and stops once the accuracy reaches the *Target*.

- '*start*' is the initial index for the sequential selection. It helps to preserve time by considering the indices from 1 to *start* highly significant features.
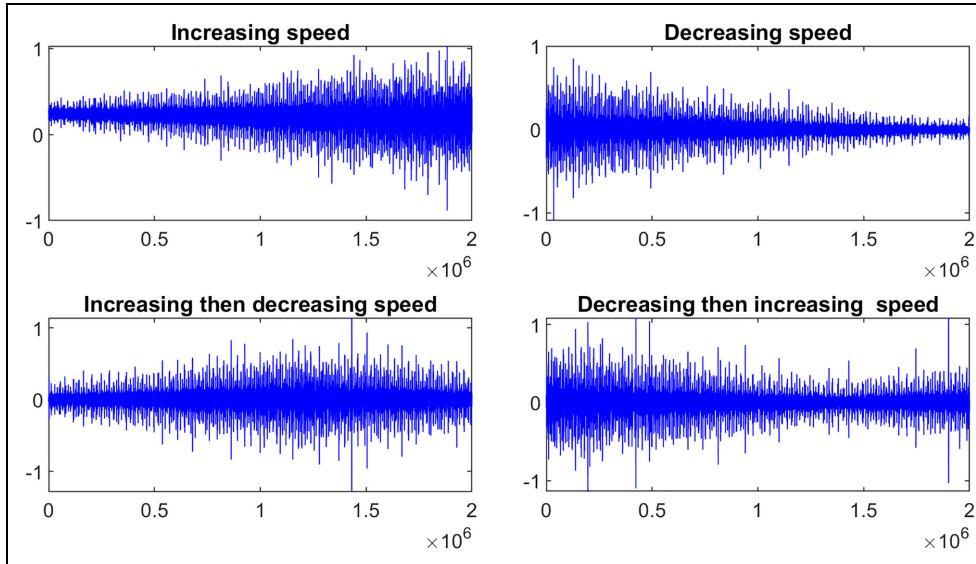
$$start = size(features) \times coef \quad (3)$$

Figure 3. Vibration signals of inner race defected bearings collected under four speed conditions.

In our application, we set the coef = 5%, assuming that the first 5% are relevant features.

- _Target_ is initially equal to 100%, it is used as a termination criterion in the selection process. If the intended accuracy is not reached with less than half of the features, the Target is adjusted using equation (4) to provide the highest possible accuracy.

$$Target = Target - \left(\frac{1}{size(DataTest)} \times 100 \right) \quad (4)$$

Experimental part

Datasets

In order to demonstrate the effectiveness of our suggested method, we conduct thorough experiments using three different databases The results of these experiments will provide valuable insights into the effectiveness of our proposed method and help us determine its potential.

Database 1. The database is called "_Bearing vibration data collected under time-varying rotational speed_," it contains three bearing health states:

- Healthy
- Inner race defect
- Outer race defect.

Operating under four rotational speed conditions to cover all possible cases of variations:

- Increasing speed
- Decreasing speed
- Increasing then decreasing speed
- Decreasing then increasing speed.

Figure 3 illustrates the data for the vibration signals that were collected while the speed varied continuously.

The bearing used is of type _ER16K_ with pitch diameter equals 38.52 mm, and nine balls with diameters equal 7.9 mm. The data is collected at a sampling rate of 200,000 Hz for 10 s for each health state under the four operating speed conditions. Three trials are repeated for each case to ensure authenticity.[19]

Database 2. _MaFaulDa_ (machinery fault data) is from a spectraQuest's machinery fault simulator (MFS) Alignment-balance-vibration (ABVT). ABVT provides vibration signals along the three axes in addition to the acoustic signal for three faulty bearings with different defective parts (outer track, rolling element, inner track).

The table below resumes the sequences for each bearing separately in two distinct positions:

- having the bearing between the rotor and the motor (underhang).
- having the rotor between the bearing and the motor (overhang).

The bearing used is of eight balls with diameters equal to 7.145 mm. Sampling rate is 50 kHz, and each sequence takes 5 s while the operating frequency ranges from 737 to 3686 rpm with steps of approximately 60 rpm. Table 1 lists all masses used for the measurements besides the number of trials for each situation.

Table 1. Characteristics of the second database 77.

State	Defect element	Masses (g)	Sequence
Normal	—	—	49
Underhang	Outer track	0	49
		6	48
		20	49
		35	42
	Rolling element	0	49
		6	49
		20	49
		35	37
	Inner track	0	50
		6	49
		20	49
		35	38
Overhang	Outer track	0	49
		6	49
		20	49
		35	41
	Rolling element	0	49
		6	49
		20	49
		35	41
	Inner track	0	49
		6	43
		20	25
		35	20

Table 2. Table of extracted features.

Feature	Equations
Mean square value	$\frac{1}{N}\sum_{i=1}^{N} x_i^2$
Minimum	$Min|x_i|$
Maximum	$Max|x_i|$
RMS	$\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$
Kurtosis	$\frac{N\sum_{i=1}^{N}(x_i-\bar{x})^4}{\left[\sum_{i=1}^{N}(x_i-\bar{x})^2\right]^2}$
Entropy	$E(S) = -\sum S_i^2 log(S_i^2)$
Standard deviation	$\sqrt{\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}}$
Mean	$\frac{\sum_{i=1}^{N} x_i}{N}$
Variance	$\frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}$
Skewness	$\frac{\sum_{i=1}^{N}(x_i-\bar{x})^3}{(n-1)\sigma^3}$
Crest factor	$\frac{Max|x(n)|}{rms}$
peak2peak	$Max(x) - Min(x)$
RSSQ	$\sqrt{\sum_{i=1}^{N} |x_i|^2}$

Database 3. The data is from the *Case Western Reserve University Bearing Data Center* website.[21] It consists of vibration signals collected from the drive end bearing of type 6205-2RS JEM SKF, deep groove ball bearing with an inside diameter of 0.9843 in and an outside diameter of 2.0472 in. The data contains four health state:

- Healthy
- Outer race defect
- Rolling element defect
- Inner race defect.

The defects are of three fault degrees 7, 14, 21 mils. The sampling rate equals 48 kHz for motor speed varies from 1797 to 1730 rpm with steps of approximately 20 rpm.

Data preprocessing. The bearing diagnosis procedure consists of three major stages: signal decomposition and feature extraction, feature selection, and classification.

we start by processing the Data. For each dataset, we followed these steps:

1. Splitting the waveform of each case into segments based on the calculated period
2. Decomposing each segment using signal decomposition technique
3. Computing the features in Table 2 for each mode of the decomposed segment
4. Repeating the steps for all segments of the signal
5. Performing the same process for all cases
6. Preserving the order and number of samples of each state.

The proposed STD-RF selection method is evaluated for its validity through a series of tests. These tests are conducted by using different signal decomposition techniques and classification methods. Moreover, the proposed STD-RF selection method is also compared to five strong optimization algorithms.

Results and discussions

On three datasets of rolling bearings collected under different conditions, we apply three signal processing techniques: Empirical Wavelet Transform (EWT), Empirical Mode Decomposition (EMD), and Maximal Overlap Discrete Wavelet Packet Transform (MODWPT). For

Table 3. Depicting the performance of the STD-RF algorithm with EWT.

TF	Dataset1		Dataset2		Dataset3	
	130		520		130	
	Selected features	Execution time (s)	Selected features	Execution time (s)	Selected features	Execution time (s)
1	14	10.20	49	13.44	17	17.71
2	23	42.19	50	12.57	17	12.77
3	14	10.28	49	16.02	17	24.26
4	14	9.95	49	10.70	20	22.49
5	16	17.53	50	13.86	21	26.40
6	14	7.01	49	15.29	20	34.09
7	16	17.53	49	11.30	17	12.83
8	16	12.61	49	12.10	20	29.96
9	19	24.65	49	11.11	23	42.90
10	20	27.06	49	15.80	22	27.39
mean	17	17.90	49	13.21	19	25.08
STD	3.09	—	0.42	—	2.27	—
% of selected features	13.07	—	9.42	—	14.61	—

Table 4. Table depicting the performance of the STD-RF algorithm with EMD.

TF	Dataset1		Dataset2		Dataset3	
	208		832		208	
	Selected features	Execution time (s)	Selected features	Execution time (s)	Selected features	Execution time (s)
1	7	16.40	132	301.20	13	22.43
2	7	31.82	132	311.70	17	50.80
3	11	25.79	130	283.97	13	19.71
4	10	25.89	129	294.46	14	47.77
5	6	13.93	132	303.46	13	15.84
6	14	46.12	130	281.63	13	18.66
7	12	34.74	132	291.80	13	16.62
8	18	62.31	132	348.93	13	14.74
9	7	15.46	132	338.05	15	32.55
10	12	31.31	135	330.55	13	33.77
mean	11	27.78	132	308.57	14	27.28
STD	3.8	—	1.64	—	1.33	—
% of selected features	5.28	—	15.86	—	6.73	—

the resulting signals, we compute the features listed in Table 2.

Then, we apply the STD-RF selection method to the obtained feature set. We consider the execution time, the number of features opted for, and the obtained accuracy.

We decompose the signal into 10 Amplitude Modulation-Frequency Modulation (AM-FM) modes for the EWT technique.

The number of intrinsic modes functions (IMF) for the EMD technique varies between 12 and 16 for the three databases. We choose 16 as the maximum value to adjust the features matrix without losing any information.

For the three datasets, the MODWPT technique extracts 16 terminal nodes.

Tables 3 to 5 present the features selected by the STD-RF method for three databases processed by EWT, EMD and MODWPT, respectively.

The tables contain the results of 10 simulations using the STD-RF selection method and the execution time for each case. As we can see, our proposed method could reduce the sets of features to less than 15% using the EWT, less than 16% using EMD and less than 10% while using MODWPT and hence help to boost the diagnosis process speed.

From Tables 3 to 5, we observe that the STD-RF's results remain in the same scope despite the signal decomposition technique tool involved in the data processing. Also, the number of selected features for the 10 simulations affirms the stability of our method in both

Table 5. Table depicting the performance of the STD-RF algorithm with MODWPT.

TF	Dataset1		Dataset2		Dataset3	
	208		832		208	
	Selected features	Execution time (s)	Selected features	Execution time (s)	Selected features	Execution time (s)
1	21	14.92	83	125.02	20	41.85
2	21	18.03	84	67.85	20	13.10
3	20	8.46	83	49.53	20	10.82
4	20	8.99	83	45.63	20	11.27
5	20	8.25	83	48.69	20	9.30
6	20	7.29	83	50.43	20	12.42
7	20	9.99	83	52.07	21	14.37
8	20	8.07	83	54.84	20	11.31
9	20	8.91	84	64.93	20	27.78
10	20	10.10	83	55.30	21	18.64
mean	20	10.30	83	61.42	20	17.08
STD	0.42	—	0.42	—	0.42	—
% of selected features	9.61	—	9.97	—	9.61	—

quantity and quality terms because of the features' ordering at the beginning of the process.

Selection techniques comparison. We put our method in a comparison with five robust optimization algorithms in bearing diagnosis field. Our method was compared to squirrel search algorithm,[21] gray wolf optimization algorithm,[22] binary coded differential evolution (BDE), Grasshopper optimization algorithm (GOA),[23] and simulated annealing (SA).[24]

Table 6 demonstrates that the *STD-RF* selection method exhibits superior performance compared to the other algorithms with respect to both accuracy and the number of selected features. Additionally, the table reveals that for the same dataset, if n simulations yield the same number of selected features, this implies that the n sets are identical, as the vector of indices **v** is consistently ordered irrespective of the initial arrangement of the data. This independence of the output from the initial data's position enhances the system's stability, unlike algorithms where the search procedure is initiated randomly and is influenced by the order of the features, leading to varying feature sets in different simulations.

Figure 4 provides a clear representation of the power of our proposed method in feature selection using the first dataset processed by the EWT technique. It reduces number of parameters involved in the classification process to just 12% without affecting the classification's accuracy.

The accuracies listed in Table 6 were assessed using the RF classifier,we have tested our proposed method using the holdout cross validation and we repeated it 10 times as an explicit 10-fold cross validation to detect any hidden variance between the 10-folds, and this

because the *k*-fold cross validation provides the average of the *k* simulations without giving an idea about the stability of the system. We spilt the data randomly into 80:20 to have larger amount of data for testing, and we repeated the process for 10 times then we calculated both the average and the STD.

Figures 5 to 7 illustrate clearly the strength of our proposed method in reducing the size of the features set compared to the total features(TF) and the outputs of strong optimization algorithms as the squirrel, gray wolf, BDE, and others, without affecting the accuracy of classification as seen in Table 6.

The accuracy of fault diagnosis can be notably enhanced by utilizing feature ranking.[25] The Figure 8 represents a histogram, which illustrates the selected features in the three datasets processed by EWT. These features are arranged in a particular order that corresponds to their importance, which is determined based on their standard deviation (STD). The histogram provides a visual depiction of the distribution of the selected features and their relative significance.

Classifiers. To determine the effectiveness of our feature selection method, we perform a thorough evaluation by testing its output with five well-established classifiers. These classifiers include K-Nearest Neighbors, Random Forest, Least-Squares Support Vector Machines, Decision Tree, and Extra-Trees. This evaluation is crucial in verifying the accuracy of the selected features and ensuring that they are capable of providing reliable results when used in the diagnosis of bearings.

Table 7 summarizes the evaluation results of the selected features by the STD-RF method from the three databases processed with the Empirical Wavelet Transform and decomposed into 10 modes.

Table 6. Number of features selected and accuracies obtained by different optimization algorithms.

Simulation	Dataset 1						Dataset 2						Dataset 3					
	Squirrel	Wolf	BDE	GOA	SA	STD-RF	Squirrel	Wolf	BDE	GOA	SA	STD-RF	Squirrel	Wolf	BDE	GOA	SA	STD-RF
1	65	50	124	55	51	16	251	112	398	65	61	49	64	38	110	54	48	17
2	57	56	120	59	53	17	247	125	442	58	57	49	69	45	94	60	48	17
3	54	50	107	58	51	15	236	148	423	61	60	50	77	50	104	52	53	17
4	71	35	122	68	48	20	231	143	415	62	60	49	65	37	75	62	52	20
5	69	42	123	55	57	14	244	108	224	69	67	50	64	44	97	58	54	21
6	59	46	117	59	59	16	244	113	441	65	63	49	66	32	72	52	53	20
7	66	32	72	52	53	20	240	131	469	67	71	49	63	45	113	56	54	17
8	70	44	122	59	60	15	251	182	437	69	67	49	59	58	98	58	51	20
9	68	54	97	62	58	16	210	192	458	60	64	49	70	38	97	56	56	23
10	73	40	126	64	53	21	229	169	248	68	63	49	60	36	95	54	56	22
Mean	65	46	117	60	54	17	238	142	395	64	63	49	66	42	96	56	53	19
Accuracy (%)	98.61	100	100	99.06	99.02	100	99.46	99.78	99.78	99.71	99.67	99.89	99.46	99.78	99.78	99.71	99.67	99.89

Table 7. Classification results of different classifiers using the selected features from the three databases.

| Simulation	Dataset 1					Dataset 2					Dataset 3				
	RF	KNN	LSSVM	ET	DT	RF	KNN	LSSVM	ET	DT	RF	KNN	LSSVM	ET	DT
1	100%	98.95%	100%	99.30%	97.54%	100%	99.20%	99.73%	100%	98.40%	99.64%	99.28%	99.64%	100%	95.35%
2	100%	98.61%	99.65%	100%	98.26%	100%	98.93%	99.73%	100%	98.93%	100%	98.57%	99.25%	100%	95%
3	100%	99.30%	99.30%	100%	96.52%	100%	98.66%	100%	100%	98.66%	100%	98.21%	99.28%	100%	95.71%
4	100%	99.65%	100%	99.30%	98.26%	100%	99.20%	100%	100%	99.46%	100%	98.92%	100%	100%	96.07%
5	99.65%	99.61%	100%	99.65%	97.28%	99.73%	99.20%	100%	100%	98.93%	99.64%	97.85%	99.28%	99.64%	96.07%
6	100%	98.95%	99.65%	100%	99.30%	100%	99.46%	100%	100%	99.73%	100%	98.92%	99.28%	100%	95.35%
7	100%	99.65%	98.95%	99.65%	98.95%	99.73%	99.20%	100%	100%	97.60%	99.64%	98.21%	100%	100%	97.14%
8	99.65%	99.65%	100%	99.65%	99.30%	100%	99.46%	100%	100%	98.66%	99.64%	97.14%	99.64%	100%	94.64%
9	99.30%	98.61%	99.65%	98.95%	99.30%	99.73%	99.73%	99.73%	100%	97.33%	99.64%	97.85%	99.64%	99.64%	96.07%
10	100%	98.95%	100%	99.30%	97.91%	100%	99.73%	99.73%	100%	98.66%	100%	98.57%	99.64%	100%	95%
Max	100%	99.65%	100%	100%	99.30%	100%	99.73%	100%	100%	99.73%	100%	97.14%	100%	100%	97.14%
Min	99.30%	98.61%	98.95%	98.65%	96.52%	99.73%	98.66%	99.73%	100%	97.33%	99.64%	97.14%	99.28%	99.64%	94.64%
Mean	99.86%	99.19%	99.69%	99.58%	98.25%	99.91%	99.27%	99.89%	100%	98.63%	99.89%	98.35%	99.60%	99.92%	95.64%
STD	0.2447	0.4306	0.4370	0.3615	0.9702	0.1304	0.3337	0.1394	0	0.7367	0.1739	0.6338	0.3162	0.1518	0.7301
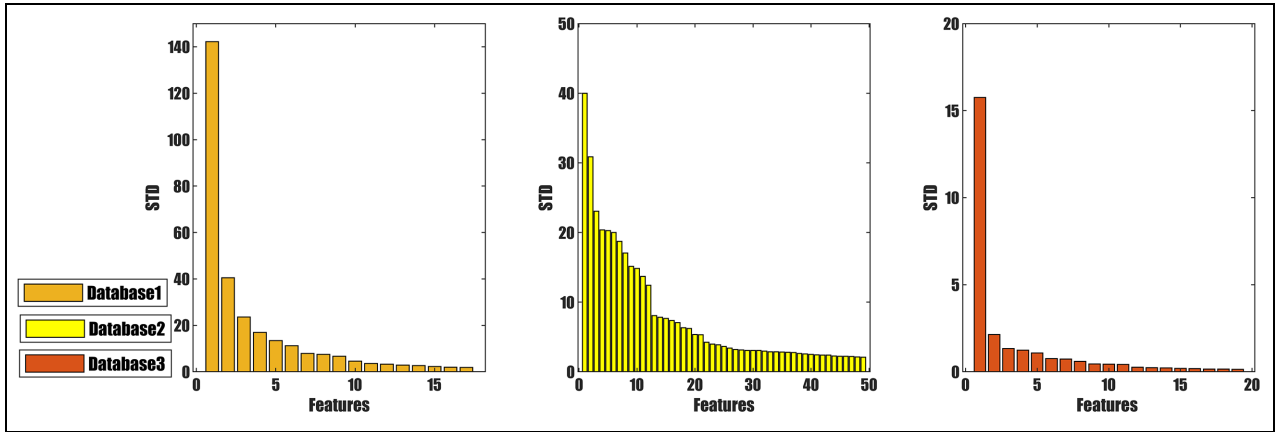
Figure 4. pie chart depicting the percentages of the selected features by different optimization algorithms.
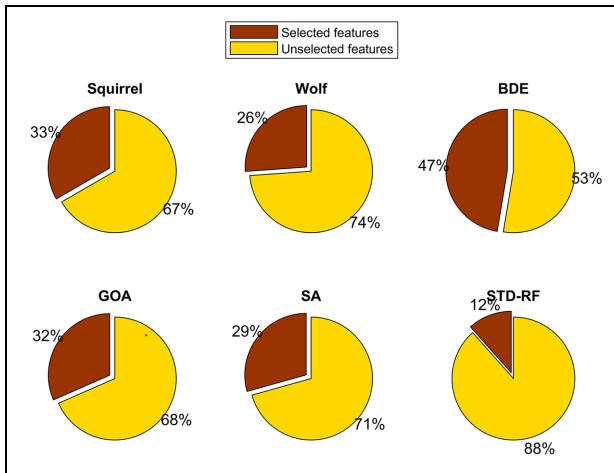

Figure 5. Comparison graph illustrating the number of selected features by different optimization methods for the first dataset.

The obtained accuracies are very promising even with a relatively weak classifier as the Decision tree,

where the mean accuracies for the three cases are 98.25%, 98.63%, and 95.64% respectively. For KNN, the accuracy approaches 100% while the rest of the classifiers could reach 100%.

Also, we notice the role of our method to retain the system's stability, where the maximum value of the STD is lower than 0.98, which is a low value and determines the robustness of the fault classification.

Conclusion

The diagnosis of bearings has gained a lot of attention due to the potential harm caused by faulty bearings. However, the accuracy of the diagnosis relies heavily on the quality of the input features used by the classifier. This is where the feature selection method comes into play. In this article, we propose an STD-RF-based feature selection method with a high ability to extract the blurred discriminative parameters for the diagnosis. We tested the selection method on three distinct databases, processed by empirical wavelet transform,
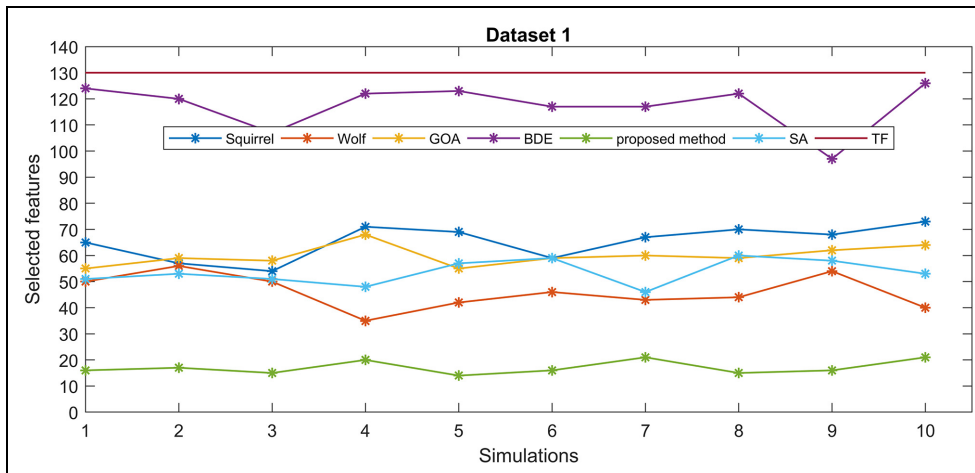

Figure 6. Comparison graph illustrating the number of selected features by different optimization methods for the second dataset.
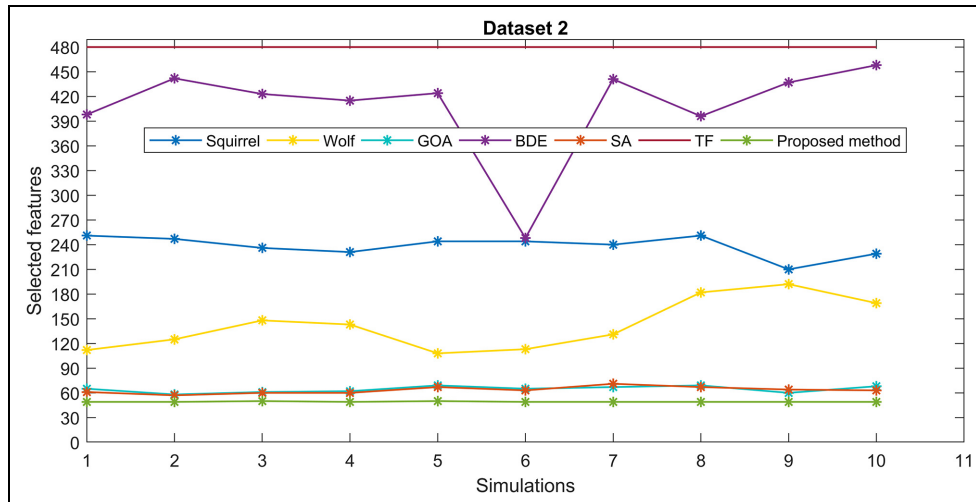
Figure 7. Comparison graph illustrating the number of selected features by different optimization methods for the third dataset.
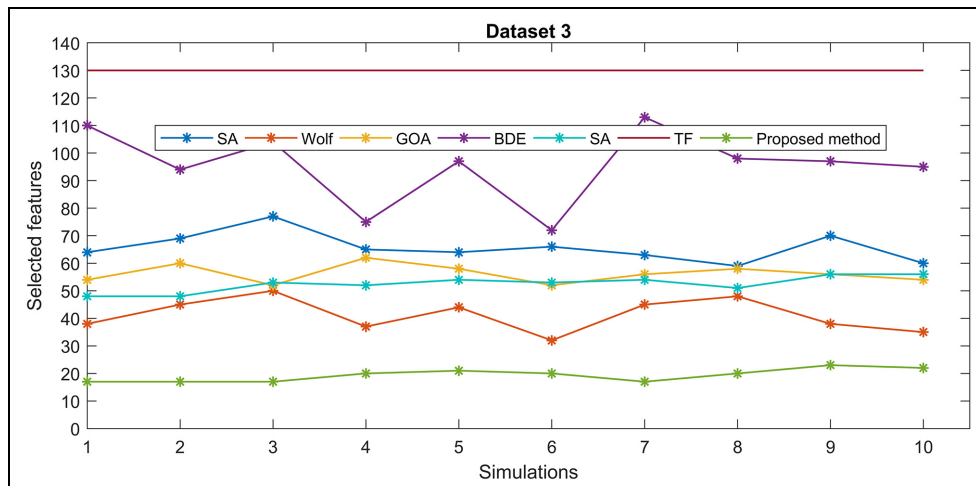

Figure 8. STD values of the selected features for the three databases processed by EWT.

empirical mode decomposition, and maximal overlap discrete wavelet packet transform. We assessed the opted set of parameters by many classifiers, such as KNN, RF, LSSVM, and others. The obtained results demonstrate the high performance of our proposed method regardless of both the signal processing technique and classifier adopted. Compared to a bunch of robust optimization techniques, the STD-RF method outperforms them in terms of accuracy, execution time, and the number of features selected. The results reveal the ability of the STD-RF-based selection method to control the time variability issue and to ensure the stability of the predictive system.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iDs

Moussaoui Imane https://orcid.org/0000-0002-1810-7469
Chemseddine Rahmoune https://orcid.org/0000-0003-0388-7455

References

1. Esakimuthu Pandarakone S, Mizuno Y and Nakamura H. A comparative study between machine learning algorithm and artificial intelligence neural network in detecting minor bearing fault of induction motors. *Energies* 2019; 12: 2105.
2. Hui KH, Ooi CS, Lim MH, et al. An improved wrapper-based feature selection method for machinery fault diagnosis. *PLoS One* 2017; 12: e0189143.

3. Bommert A, Sun X, Bischl B, et al. Benchmark for filter methods for feature selection in high-dimensional classification data. *Comput Stat Data Anal* 2020; 143: 106839.

4. Peña M, Cerrada M, Alvarez X, et al. Feature engineering based on anova, cluster validity assessment and KNN for fault diagnosis in bearings. *J Intell Fuzzy Syst* 2018; 34: 3451–3462.

5. Shunmugapriya P and Kanmani S. A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC hybrid). *Swarm Evol Comput* 2017; 36: 27–36.

6. Kumar V and Minz S. Feature selection: a literature review. *Smart Comput Rev* 2014; 4: 211–229.

7. Jović A, Brkić K and Bogunović N. A review of feature selection methods with applications. In: *2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp.1200–1205. Opatija, Croatia: IEEE.

8. Chandrashekar G and Sahin F. A survey on feature selection methods. *Comput Electr Eng* 2014; 40: 16–28.

9. El Aboudi N and Benhlima L. Review on wrapper feature selection approaches. In: *2016 international conference on engineering and MIS (ICEMIS)*, pp.1–5. Agadir, Morocco: IEEE.

10. Khalid S, Khalil T and Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. In: *2014 science and information conference*, pp.372–378. London: IEEE.

11. Veerabhadrappa M and Rangarajan L. Bi-level dimensionality reduction methods using feature selection and feature extraction. *Int J Comput Appl* 2010; 4: 33–38.

12. French A, Macedo M, Poulsen J, et al. *Multivariate analysis of variance (MANOVA)*. San Francisco State University, 2008.

13. Imane M, Rahmoune C, Zair M, et al. Bearing fault detection under time-varying speed based on empirical wavelet transform, cultural clan-based optimization algorithm, and random forest classifier. *J Vib Control* 2023; 29: 286–297.

14. Rajeswari C, Sathiyabhama B, Devendiran S, et al. Bearing fault diagnosis using wavelet packet transform, hybrid PSO and support vector machine. *Procedia Eng* 2014; 97: 1772–1783.

15. Ma M, Sun C and Chen X. Discriminative deep belief networks with ant colony optimization for health status assessment of machine. *IEEE Trans Instrum Meas* 2017; 66: 3115–3125.

16. Saini S, Bt Awang Rambli DR, Zakaria MNB, et al. A review on particle swarm optimization algorithm and its variants to human motion tracking. *Math Probl Eng* 2014; 2014: 1–16.

17. AL-Behadili HNK, Ku-Mahamud KR and Sagban R. Hybrid ant colony optimization and genetic algorithm for rule induction. *J Comput Sci* 2020; 16: 1019–1028.

18. Oloruntoba O, Cosma G and Liotta A. Clan-based cultural algorithm for feature selection. In: *2019 international conference on data mining workshops (ICDMW)*, pp.465–472. Beijing, China: IEEE.

19. Huang H and Baddour N. Bearing vibration data collected under time-varying rotational speed conditions. *Data Brief* 2018; 21: 1745–1749.

20. Case Western Reserve University bearing data center website 2021. http://csegroups.case.edu/bearingdatacenter/home;https://engineering.case.edu/bearingdatacenter

21. Jain M, Singh V and Rani A. A novel nature-inspired algorithm for optimization: Squirrel search algorithm. *Swarm Evol Comput* 2019; 44: 148–175.

22. Mirjalili S, Mirjalili SM and Lewis A. Grey wolf optimizer. *Adv Eng Softw* 2014; 69: 46–61.

23. Meraihi Y, Gabis AB, Mirjalili S, et al. Grasshopper optimization algorithm: theory, variants, and applications. *IEEE Access* 2021; 9: 50001–50024.

24. Dowsland KA and Thompson J. Simulated annealing. In: Rozenberg G, Bäck T and Kok JN (eds.) *Handbook of natural computing*. Berlin: Springer, 2012, pp. 1623–1655.

25. Vakharia V, Gupta VK and Kankar PK. A comparison of feature ranking techniques for fault diagnosis of ball bearing. *Soft Comput* 2016; 20: 1601–1619.

Appendix

List of acronyms

ML	machine learning
FS	feature selection
STD	standard deviation
RF	random forest
ANOVA	analysis of variance
PSO	particle swarm optimization
ACO	ant colony optimization
ABVT	alignment-balance-vibration
RPM	revolutions per minute
IMF	intrinsic mode function
EWT	empirical wavelet transform
EMD	empirical mode decomposition
MODWPT	maximal overlap discrete wavelet packet transform
TF	total features
KNN	K-nearest neighbors
LSSVM	least-squares support vector machines
RSSQ	root sum of squares
BDE	binary differential evolution
SA	simulated annealing
GOA	grasshopper optimization algorithm