Applied Acoustics 187 (2022) 108505

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Optimal sensor placement for identifying multi-component failures in a wind turbine gearbox using integrated condition monitoring scheme

S V V S Narayana Pichika [a,*], Ruchir Yadav [a], Sabareesh Geetha Rajasekharan [a], Hemanth Mithun Praveen [a], Vamsi Inturi [a,b]

[a] Department of Mechanical Engineering, BITS-Pilani Hyderabad Campus, India
[b] Mechanical Engineering Department, Chaitanya Bharathi Institute of Technology (A), Hyderabad, India

ARTICLE INFO

Article history:
Received 12 June 2021
Received in revised form 1 October 2021
Accepted 25 October 2021

Keywords:
Condition monitoring
Grey wolf optimizer
Multi-component defects
Optimal sensor placement
Wind turbine gearbox

ABSTRACT

Wind turbine gearbox has a high failure frequency and downtime, and therefore, several sensors are installed to perform condition monitoring to reduce the operation and maintenance costs. A gearbox can have infinite sensor nodal positions, but, in reality, the positioning of sensors is limited to a finite number of locations. Also, sensor location influences the quality of the data captured by the sensors, which is of key importance in a condition monitoring system. Hence selection of optimal sensor placement (OSP) is a challenging task which needs to be addressed. When the sensor type changes, the measurement response changes, and hence the OSP methodologies based on the measured responses may not work well. For addressing this, an optimization method based on statistical features is proposed to find the optimum sensor placement (OSP). In order to evaluate the effectiveness of the proposed method, experiments are conducted on a laboratory scale model of wind turbine gearbox considering multi-component faults and an integrated condition monitoring scheme. Variational mode decomposition and the Spearmen correlation coefficient are used to processes raw acoustic and vibration signals. Feature extraction is performed to obtain nine statistical features, and a mathematical objective function is constructed as a function of these features. Grey wolf optimizer is employed to find the optimal values of the statistical features. Fault classification is performed using Random forest (RF) and deep multi-layer perceptron (MLP) algorithms. Optimal sensor network identified by the above method reported classification accuracy of 86.88% and 88.34% for RF and MLP, respectively. As a result, the number of sensors reduced from eight to five. The proposed method can be used as an effective technique for OSP problems.

1. Background & literature survey

The current climatic conditions and change in climate policies have attracted various countries worldwide to adopt renewable technologies to fight climate change. Wind energy technology has gained much attention and investments in the renewable energy markets in the past few decades. World wind energy association (WWEA) reported a 29% global growth in wind energy installed capacity from 2015 to 2020 [1]. Wind turbines (WT) operate under changing weather conditions and are often subjected to severe operating conditions. Eventually, the failures occurring in any sub-system of the wind turbine, may lead to damage in other components and might further result in catastrophic failures. Among the different sub-systems of the wind turbine, the gearbox has the highest failure rate and downtime [2]. The WT gearbox transmits the power from low-speed rotor shaft to the high-speed shaft that drives the generator, thus enduring fluctuation loads from both the turbine and generator sides. Inaccurate installation, manufacturing error, and abrasion may cause distributed defects, whereas fatigue loads may cause localized defects such as gear tooth crack, bearing race cracks, and bearing spall [3]. Frequent breakdown of the gearbox sub-systems (either gears or bearings) is common due to the harsh environment created during operations, leading to immense operation and maintenance costs.

Condition monitoring (CM) has gained significant attention in recent years in the field of preventive maintenance in reducing operation cost, downtime of wind turbines, and maintenance costs. CM monitors the real-time machinery health condition, helps in making optimal maintenance decisions, and improves machine reliability. Defect identification and quantification in machinery components are achieved with the help of sophisticated signal processing and decision support algorithms in a CM system. Various

* Corresponding author.
E-mail address: p20200042@hyderabad.bits-pilani.ac.in (S V V S Narayana Pichika).

https://doi.org/10.1016/j.apacoust.2021.108505
0003-682X/

authors have implemented different CM techniques based on vibration, quality of lubrication oil, temperature, acoustic signals, and motor current [4–13]. Individual CM techniques showed certain limitations; for example, individual vibration analysis and wear debris based CM analysis can only diagnose about 30–40% of faults [14], shifting focus toward integrated condition monitoring (ICM). This combines different CM techniques to improve the fault diagnosis and prognosis of the systems under study. Vamsi et al. [15] proposed an ICM based on vibration, acoustic and lubricating oil analysis and observed that the ICM scheme provided better classification accuracy than individual CM techniques while diagnosing the bearing faults of a multi-stage WT gearbox. Liya et al. [16] performed ICM on a two-stage gearbox and achieved an average fault recognition rate of 97.7% while detecting the gear tooth defects.

In ICM, data from different sensors are fused and utilized for the diagnosis of defects. However, for most of the ICM schemes, data from different sensors may not be commensurate, making sensor-level data fusion highly challenging. Feature-level data fusion overcomes this limitation by fusing the features obtained by post-processing the raw signals using signal processing approaches. Various signal processing approaches have been proposed by researchers for machinery fault diagnosis that includes traditional basis expansion methods and adaptive mode decomposition methods [17]. Empirical mode decomposition (EMD), local mean decomposition (LMD), ensemble EMD (EEMD), and variable mode decomposition (VMD) are a few adaptive mode decomposition methods. Among these, VMD algorithm proposed by Dragomiretskiy et al. [18] is robust against noise and decomposes a complicated multi-component signal into constituent amplitude modulation and frequency modulation (AM-FM) components. VMD is used as a signal processing tool in the fault diagnosis of rolling elements. Kumar et al. [19] proposed a dynamic time warping algorithm (DTW) to select intrinsic mode functions (IMFs) of VMD for bearing fault diagnosis. DTW was able to select appropriate IMFs than the fault correlation factor. Li et al. [20] optimized VMD based on kurtosis and resonance frequency to extract bearing fault features. The optimized VMD was able to extract early fault signals. Li et al. [21] utilized frequency band entropy to select IMFs of VMD to diagnose bearing faults. Zhong et al. [22] proposed theoretical propositions to demonstrate the effects of impulsive noise on kurtosis and negative entropy. The health of the system was monitored by developing a weighted residual regression-based health index. The features extracted from the processed signals can be further used in either machine-learning (ML) based and deep learning (DL) based fault classification. Long et al. [23] proposed a hybrid deep learning architecture to diagnose the faults in industrial robots. Sparse auto-encoder (SAE) is integrated with the support vector machine (SVM) to build a hybrid SAE-SVM based fault diagnosis model. Among different ML and DL classification algorithms, Random forest (RF) and multi-layer perceptron (MLP) have shown good classification accuracy while classifying between the various defect severity levels of bearing race and gear tooth faults [24–29].

It is a common observation from previous research that quality of signals acquired by sensors affects the fault diagnosis accuracy and stability. Thus, the effectiveness of the CM system is dependent on the sensor distribution and signal quality. Furthermore, sensor placement affects the accuracy of the obtained information. With an increase in the number of sensors, the quality of data improves. However, this is restricted by space restrictions and high installation and operation costs. Thus, optimal sensor placement (OSP) is a significant paradigm in the CM. Pan et al. [30] proposed a model-based particle swarm optimization algorithm to identify the optimal layout of the acceleration sensor (six instead of ten) during the fault diagnosis of a gearbox. Zhao et al. [31] proposed an improved discrete shuffled frog leaping algorithm (ID-SFLA) for sensor network optimization (three instead of eighteen) in a gearbox considering bearing and gear tooth faults. Also, it was reported that the proposed ID-SFLA showed better searching ability and convergence performance than the genetic algorithm and SLFA. Vanraj et al. [32] reported a response surface methodology (RSM) based OSP for a two-stage gearbox considering gear tooth fault and identified the sensor's optimal vertical and horizontal positions. GuiLan et al. [33] conducted numerical simulation-based OSP studies on a planetary gearbox and identified the minimal sensor network that can isolate and diagnose the gear tooth defects.

However, most OSP strategies reported in the existing literature are limited to single component fault (either gear tooth or bearing race fault) and utilized only a single CM technique. Albeit, in the real world, wind turbine gearboxes are experiencing multi-component failures (gear tooth and bearing race defects together) and multiple types of defects. Besides, individual CM techniques showed limited diagnosing capabilities. In addition to that, it is challenging to obtain the labelled historical data or simulate the data through model-based investigations, which can be further be used for performing the OSP studies. This paper attempts to address these issues by proposing an OSP through Grey wolf optimizer (GWO) on an ICM scheme data for diagnosing the multi-component failures of a wind turbine gearbox. The main contribution of this work is thus an optimization methodology to find the OSP. The developed method uses statistical features in the optimization process and hence is independent of change in the type of data and sensors unlike the earlier approaches. The same is demonstrated on a WT gearbox using an ICM considering multi-component faults. The raw vibration and acoustic signals are collected, processed using the VMD approach, and various statistical features are extracted. Later, an objective function is developed using the extracted features and is optimized using GWO, and these optimized statistical features are used to identify the OSP. Finally, the fault classification is performed using RF and MLP algorithms. The presented OSP methodology is used to find the optimum number and location of sensors to perform fault diagnosis in a WT gearbox. The rest of the paper is organised as follows. Experimental setup and procedure are detailed in Section 2. Signal processing, feature extraction, and optimization methodology are set forth in Section 3. Fault classification results and relevant discussions are reported in Section 4. Finally, conclusions are drawn in Section 5.

2. Experimental setup & test procedure

Condition monitoring investigations were performed on a scaled wind turbine gearbox, essentially the same as used by Vamsi et al. [34]. The overall gear ratio of the gearbox is 48:1 and has three-stage parallel spur gear pairs, namely high-speed stage (HSS), intermediate-speed stage, and low-speed stage (LSS). Since the study is aimed at defect diagnosis rather than power generation, the generator is replaced by a 1 hp AC three-phase synchronous motor and is coupled to the HSS of the gearbox using a jaw coupler. The HSS is driven by the motor, which further drives the ISS and LSS. All the shafts are braced using radial insert ball bearings. Table 1 shows the details of the different components of the gearbox. The speed input to the motor is given through variable frequency drive (VFD). Piezoelectric accelerometers (PCB make, 1 Hz to 11 kHz frequency range with 10 mV/g) mounted on the top of the bearing housing are used to acquire vibration data. Microphones (free-field, GRAS make, 40PH, 10 Hz to 20 kHz) are mounted on the casing of the gearbox using 3D-printed mounts to acquire acoustic signals. The sensors are connected through a data logger (NI 9234) to the computer.

Table 1
Details of the gearbox components.

Parameter	LSS	ISS	HSS
Teeth on gear	60	80	80
Gear pitch diameter, mm	180	240	240
Teeth on pinion	20	20	20
Pressure angle, deg	20	20	20
Bearing pitch diameter, mm	25	20	25
Number of balls	8	8	8
Ball diameter, mm	9	8	9
Gear module, mm	3		

Fig. 1 and Fig. 2 shows the experimental test rig and schematic of the sensor location of the wind turbine gearbox, respectively. In Fig. 2, notations B, P, and G correspond to bearing, pinion, and gear, respectively. Locations of the sensors are represented by the notation 'L', and the respective numbering for all notations are as shown in Fig. 2.

Localized defects of the bearing and the pinion are the primary modes of failure [11], and hence are considered in this current study. The defects are artificially seeded, and wire cut electro-discharge machining (EDM) is used to create the faults on the bearings and the pinions. Gear tooth crack is created in axial direction using EDM and is considered as the gear fault, refer Fig. 3. Outer race fault and inner race fault are the two bearing local defects considered in this study. EDM is used to create these faults in the axial direction on the bearing outer race and inner race. Table 2 gives the dimensions of the seeded faults. Bearing faults are shown in Fig. 4.

Since the prime focus of this investigation is to identify the optimal sensor layout for the diagnosis of multi-component defects, the authors attempted with the hard-to-detect cases mentioned in Antoniadou et al. [35]. The gearbox is tested for stationary loading with a constant operating speed of 50% of the maximum motor speed, i.e., at 720 rpm. Table 3 shows the different combinations of the faults that are considered in this study. The speed of the motor is controlled using a VFD. Sensor location, bearings, and gears inside the gearbox follow the name convention as shown in Fig. 2. Vibration and acoustic data are acquired at 12 kHz sampling frequency with 16,384 as the sample length. For each fault combination, the size of the dataset is 100 × 131,072.

3. Methodology

In the present investigation, the gearbox is equipped with eight sensors (five accelerometers and three microphones) as shown in the Fig. 2 to enable ICM, and the corresponding initial sensor network diagram considered is shown in Fig. 5, where the gearbox component locations are as shown in Fig. 2. Vibration and acoustic signals are collected using the respective sensors for the fault conditions listed in the Table 3. Variational mode decomposition (VMD) is used for signal processing. Statistical feature extraction is done after signal processing to find useful features. An objective function is constructed based on the extracted statistical features, and Grey wolf optimizer (GWO) is used to find the optimized values of these statistical features. A sensor network is selected based on the optimized values of these statistical features. Classification is performed using machine learning algorithms for the sensor sets. A sensor network with the highest classification accuracy is selected as the optimal sensor network. The proposed OSP methodology is shown in Fig. 6. A general introduction to VMD, determination of VMD mode number, selection of sensitive intrinsic mode functions (IMFs) using Spearman correlation, feature extraction, and optimization methodology are discussed in the following sections.

3.1. Signal processing

As discussed in the earlier section, VMD is used for post-processing the raw vibration and acoustic signals. VMD algorithm is robust against noise and decomposes a complicated multi-component signal into constituent amplitude modulation and frequency modulation (AM-FM) components. It requires determining parameters such as mode number and penalty prior to performing VMD analysis, which is detailed in the subsequent sections. Modes are sub-signals that produce input with specific sparsity properties, and an optimal mode number needs to be estimated so as to reduce the noise in the reconstructed signal. The penalty is introduced to increase reconstruction fidelity. Acquired vibration and acoustic signals are processed using VMD after finding the above influential parameters to get a series of IMFs. IMFs are AM-FM signals that have to be adequately extracted from the signal to keep its properties and physical meaning of signal unaltered. In this investigation, the correlation between the original signal and the IMFs is analyzed using Spearman correlation coefficient and used to further screen the IMF as detailed in the following sections.

3.1.1. VMD mode number determination

The mode number of VMD needs to be determined to estimate the IMFs. The principle of the Kurtosis maximum is used to determine mode number (K). To initialize the VMD process, the penalty parameter (α) is given first, and then it is optimized with the opti-

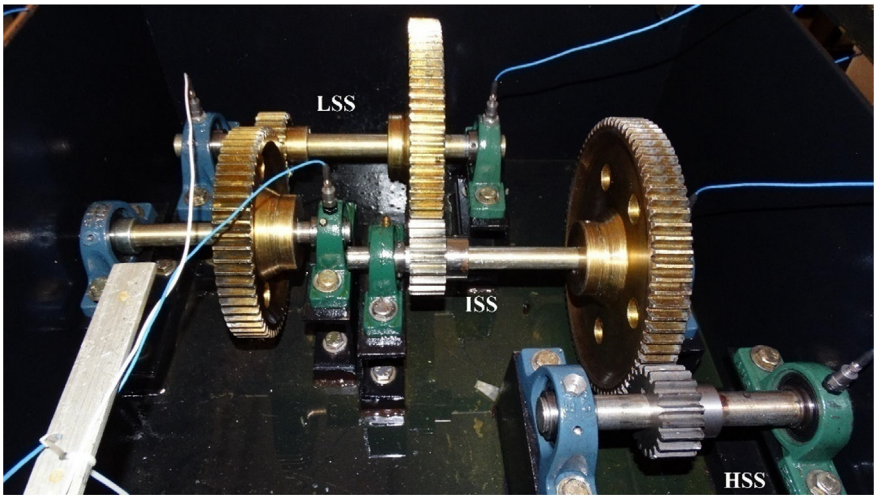


Fig. 1. Wind turbine gearbox test rig.

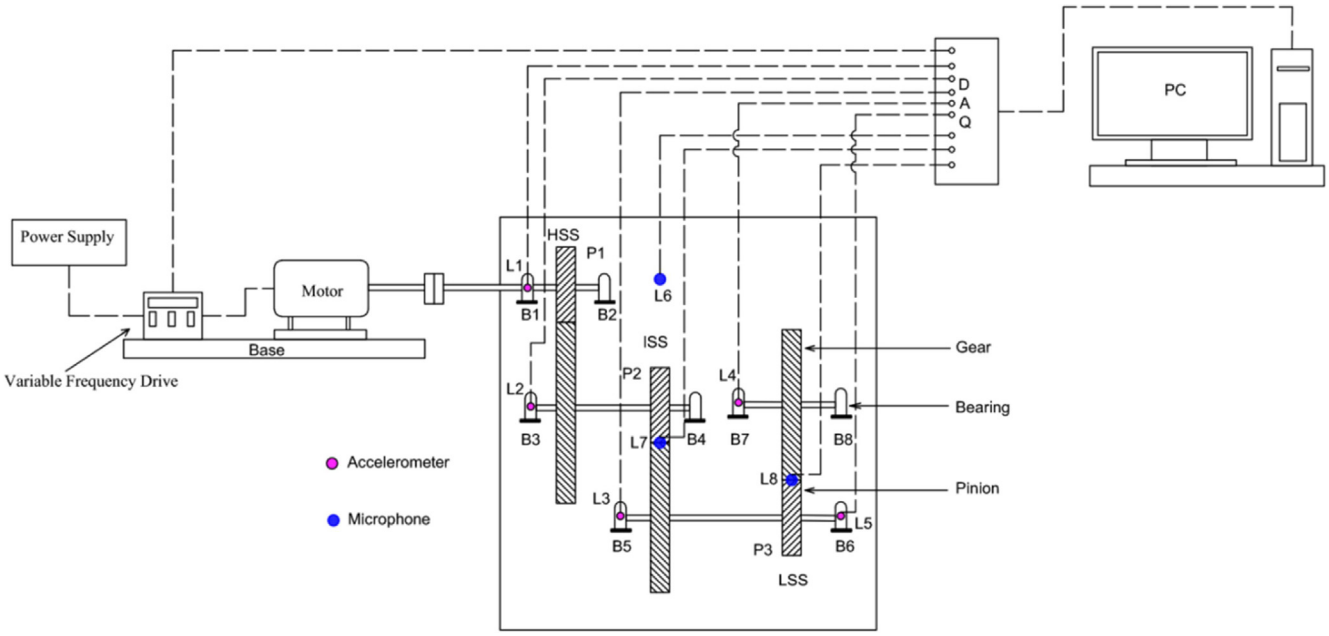


Fig. 2. Wind turbine gearbox diagram showing sensor location.

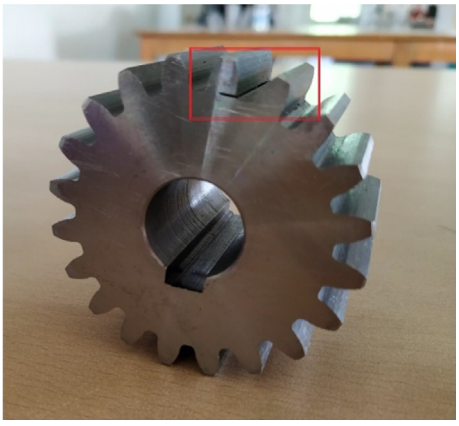


Fig. 3. Gear tooth root crack.

Table 2
Fault dimensions.

Name	Dimensions (length × depth × thickness)
Healthy	————
Cracked pinion tooth	30 mm × 3 mm × 0.25 mm
Outer race fault	14 mm × 1.4 mm × 0.25 mm
Inner race fault	30 mm × 2.2 mm × 0.25 mm

mized K^*, which corresponds to the mode number, and is discussed below [36]. For optimizing the mode number K, take the penalty factor $\alpha = 2000$ and bandwidth $\tau = 0$ as the default value, and mode number is initialized as $K = 2$, Xueli et al. [36]. The search range of K is set in the range two to fifteen with the search step equal to one, Tang et al. [37]. K VMD modes ($K \in [2, 15]$) are performed on the original signal. For each IMF, the value of Kurtosis is calculated under this mode number K and is recorded as

$$K_n = (K_1, K_2, \cdots\cdots, K_m) \tag{1}$$

where, for each IMF the Kurtosis value is indicated by K_n where the subscript 'n' denotes mode number during optimization. A local maximum K_a is obtained as

$$K_a = max(K_1, K_2, \cdots\cdots, K_m) \tag{2}$$

There are fifteen local maximum as the K is set in range two to fifteen. Therefore, we obtain

$$K_{a,n}^{max} = (K_{a,1}^{max}, K_{a,2}^{max}, \cdots\cdots.., K_{a,16}^{max}) \tag{3}$$

The global maximum K_b^{max} is then obtained as

$$K_b^{max} = \max(K_{a,1}^{max}, K_{a,2}^{max}, \cdots\cdots.., K_{a,16}^{max}) \tag{4}$$

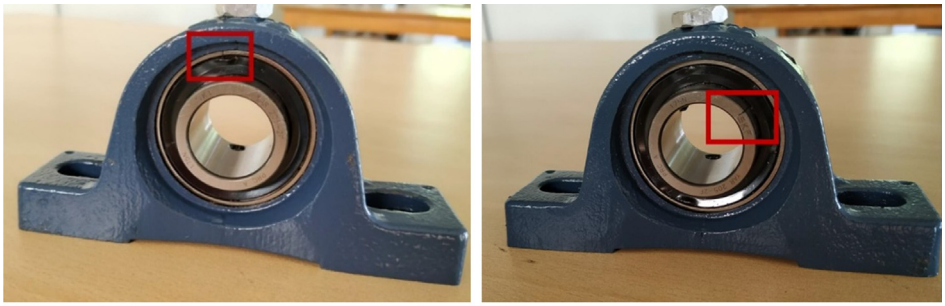
This gives the optimal mode number and is denoted by K^*. The number of IMFs generated is equal to K^*.

The procedure as mentioned above is implemented on the raw vibration and acoustic signals, which are acquired using five accelerometers and three microphones, respectively. VMD method is used to decompose the original signal from each sensor, and for each case, two to fifteen IMFs are obtained. A representative case, case C_1, is detailed in this section to explain the process. The raw sensor signals are subjected to VMD with the initial mode number $K = 2$, and the K search range is set - two to fifteen with a search step size value equal to one, as explained earlier. The maximum kurtosis values are obtained for mode numbers $K = 7$ and $K = 10$ for the sensors L_1 and L_7, respectively. So, for the sensors L_1 and L_7, the optimal mode number is taken as $K^* = 7$ and $K^* = 10$, respectively. VMD with seven and ten modes are used to analyse the original signals from the sensors L_1 and L_7 and correspondingly same number of IMFs are generated, respectively. Fig. 7 shows the time-domain waveforms of the IMFs for sensor L_1.

Further, the characteristic IMFs should be selected to reconstruct the signal which is explained in the next section.

3.1.2. Characteristic IMF selection

All the generated IMFs may not contain useful information and has to be screened. Spearman correlation coefficient, also known as Spearman's rank correlation coefficient, is used to analyse the correlation between the IMF and the original signal to screen the IMFs. Spearman correlation coefficient gives the statistical dependence between two variables in a non-parametric measurement. The coefficient uses a monotonic function to assess the quality of the relationship between two variables. If each variable is a perfect monotone function, the Spearman correlation has a value



(a) (b)

Fig. 4. Bearing defects: (a) Outer race fault, (b) Inner race fault.

Table 3
Fault combinations.

S. No.	Fault combination	Case label
1	Cracked pinion tooth (at HSS, ISS, and LSS) + Outer race fault (at LSS) + Inner race fault (at HSS)	C_1
2	Cracked pinion tooth (at HSS, ISS, and LSS) + Inner race fault (at HSS)	C_2
3	Cracked pinion tooth (at HSS, ISS, and LSS) + Outer race fault (at LSS)	C_3
4	Cracked pinion tooth (at ISS) + Inner race fault (at HSS)	C_4
5	Cracked pinion tooth (at ISS) + Outer race fault (at LSS)	C_5
6	Cracked pinion tooth (at LSS) + Outer race fault (at LSS)	C_6

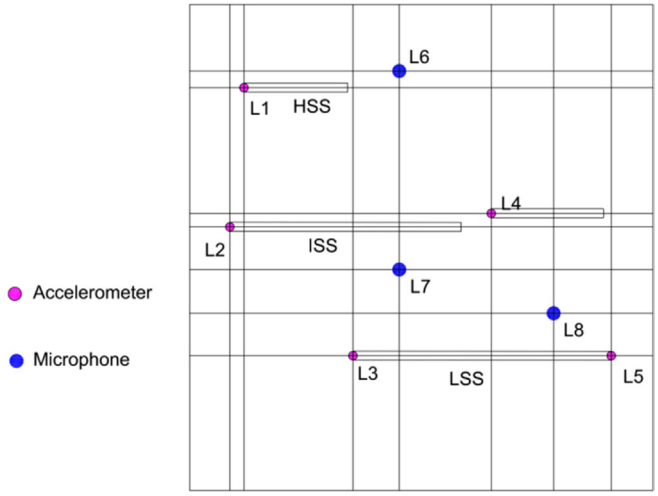


Fig. 5. Initial sensor network diagram.

+1 or −1. The Spearman coefficient is non-parametric and can be used with both discrete and continuous variables [38]. The original data X_i and Y_i from a sample of size N are converted into ranks R_{Xi} - and R_{Yi}, respectively. The Spearman correlation coefficient R_S is given as

$$R_S = \frac{cov(R_{Xi}, R_{Yi})}{\sigma_{R_{Xi}}, \sigma_{R_{XYi}}} \tag{5}$$

where covariance of the rank variables is represented by $cov(R_{Xi}, R_{Yi})$ and $\sigma_{R_{Xi}}$ and $\sigma_{R_{XYi}}$ are the standard deviations of the rank variables. IMFs having less similarity with the original signals are eliminated by a threshold Φ [39]. An IMF is selected if the R_S value is greater than Φ otherwise, it is abandoned. The threshold is defined as follows.

$$\Phi = \sqrt{\frac{\sum_{i=1}^{k} \left(R_{S,i} - \overline{R}_s\right)^2}{k}} \tag{6}$$

where R_S is the Spearman correlation coefficient, $\overline{R}_s$ is the mean value of R_S and k is the number of IMFs. The selected IMFs are further used for signal reconstruction from which the features are extracted.

Firstly, by using formula (5), the R_S value of each IMF is calculated. Fig. 8 (a) and 8 (b) gives the R_S values obtained for the IMFs of the L_1 and L_7, respectively. To select an IMF with high similarity with the original signal, R_S should be greater than Φ, given by equation (6). The Φ values are calculated using equation (6) and have values, 0.2101 and 0.1027 for the IMF components of the L_1 and L_7, respectively. The selected IMFs are combined to reconstruct the signal, and the remaining IMFs are omitted. Original signals and reconstructed signals for sensors L_1 and L_7 are shown in Fig. 9.

The process is repeated for all the cases (Table 3 and the reconstructed signals are constructed. Statistical features are extracted from the reconstructed signal which is explained in the next section.

3.2. Feature extraction

The statistical features, namely, standard deviation (std), maximum amplitude, peak2peak, clearance factor, kurtosis, skewness, shape factor, impulse factor, and crest factor, are extracted from the reconstructed signals. The final dataset consists of ten columns, out of which nine columns are features, and the *10th* column is a class label. This data is used for finding OSP and perform fault classification that is discussed in the subsequent sections. These statistical features carry information that helps in the identification of the state of the components. The accuracy of the fault identification depends on the quality of the statistical features. Therefore, to improve the quality of the information, a novel method based on statistical features is proposed and is detailed in section 3.3.

3.3. Proposed optimization methodology

A methodology based on statistical features and Grey wolf optimizer is proposed to optimize the sensor placement. Mirjalili at al. [40] designed GWO by mathematically modelling the social hierarchy and hunter behavior of Grey wolves. GWO has three main steps of hunting, searching for prey, encircling prey, and attacking prey. The detailed GWO process is given by Mirjalili at al. [40]. The objective function is formulated as given below

Min Z = $\sum_{i=1}^{9} W_i \frac{S_{i,j}}{S_{i,min}}$; j = (1, 2, 3, ... n)

Subjected to

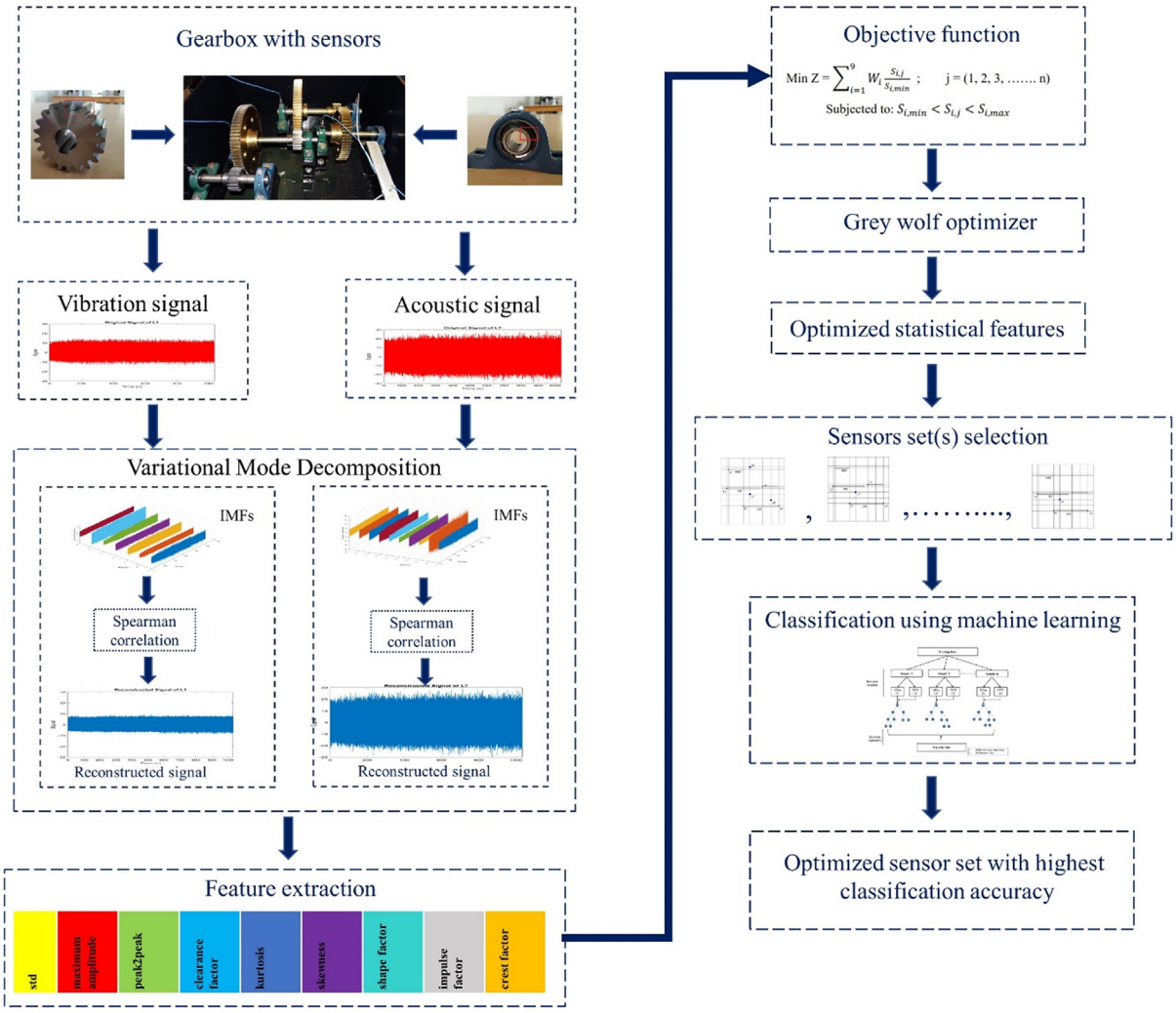


Fig. 6. Proposed OSP methodology.

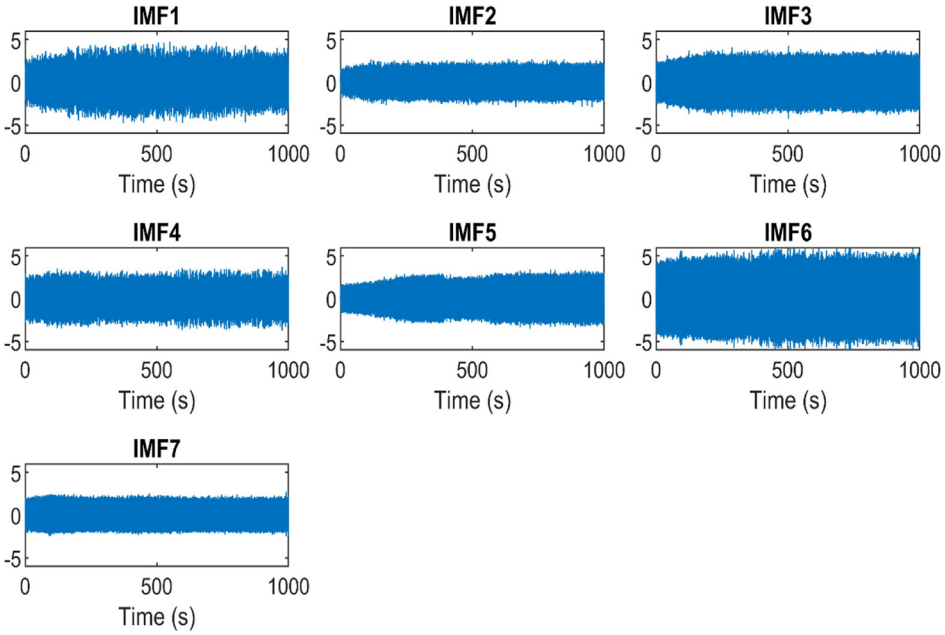


Fig. 7. VMD results for sensor L_1.

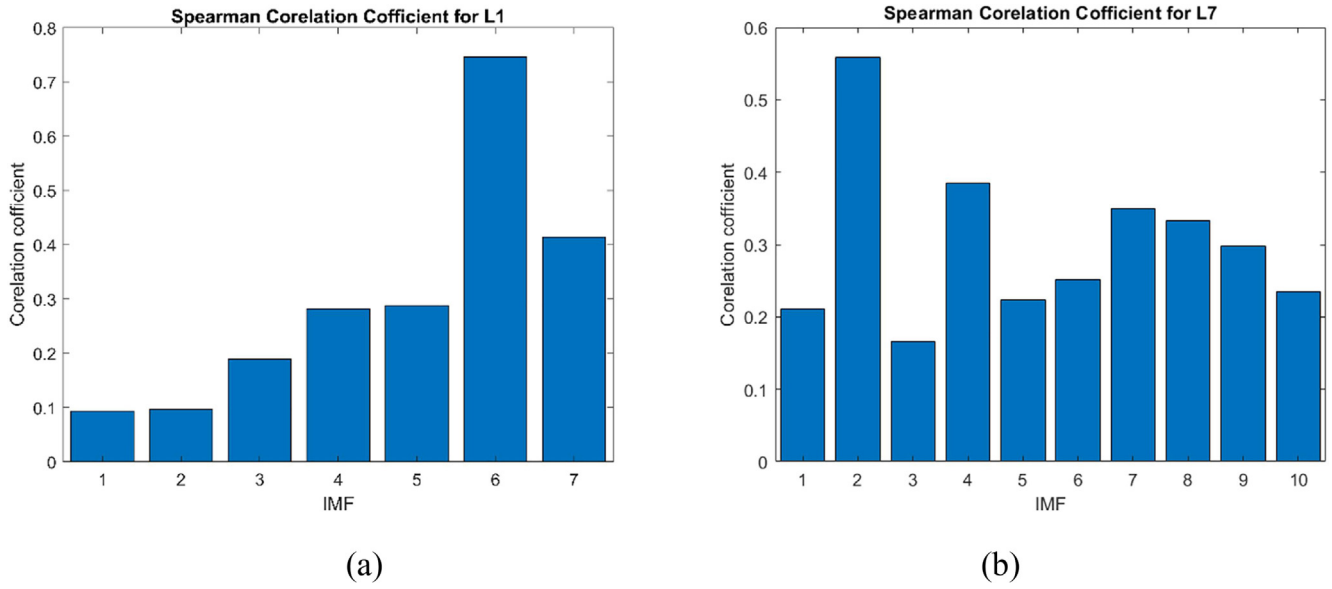


Fig. 8. Spearman correlation coefficient for: (a) sensor L_1 and (b) sensor L_7.

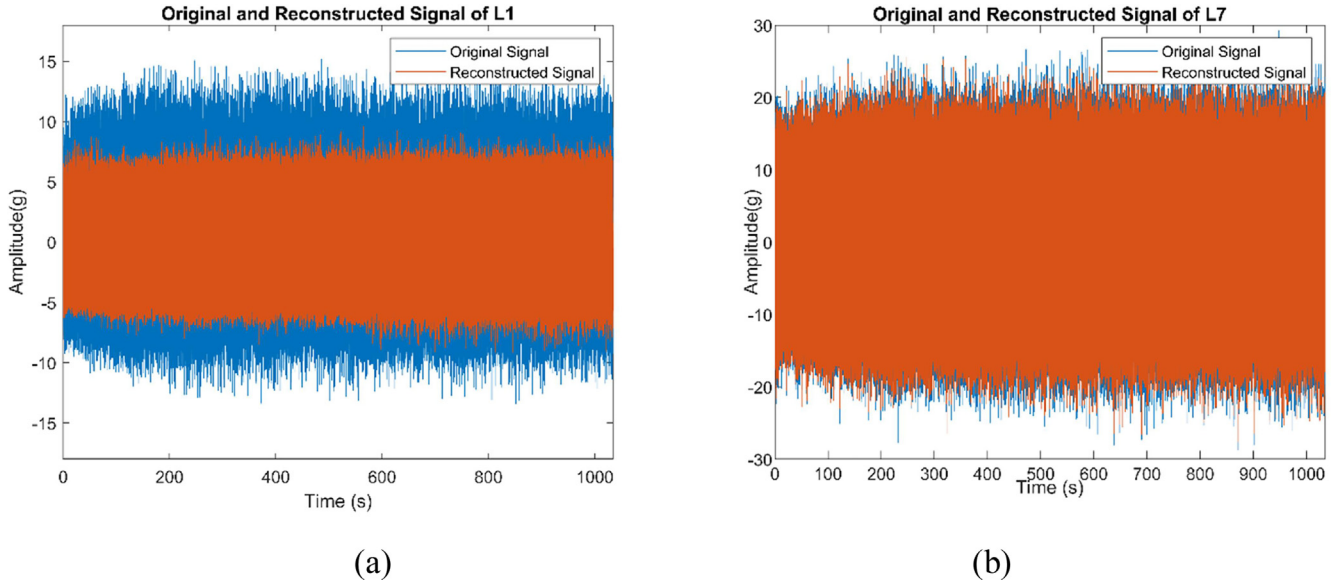


Fig. 9. Original and reconstructed signals for: (a) sensor L_1 and (b) sensor L_7.

$$S_{i,min} < S_{i,j} < S_{i,max} \tag{7}$$

where the S is the statistical feature, W_i is the weight factor given to each variable, $S_{i,j}$ is the value of each feature at index j, and $S_{i,\min}$ and $S_{i,\max}$ are the minimum and maximum values of statistical feature i for all the fault cases, respectively.

Eq. (7) is a convex function and can be verified by checking the Hessian matrix. However, the constraints involved are non-smooth/non-convex sets. A sample set is shown in Fig. 10. These problems, when solved using mathematical programming techniques, suffers local optimality or curse of dimensionality [41,42]. The dimensions of the problem may increase with iterations and diverge instead of converging. Multiple feasible regions and multiple local optimal points may exist within each region of constraints, as shown in Fig. 10. The computational time will be large in finding an optimal value among the global optimum across all feasible regions. Also, the non-linearities inside the gearbox due to varying operating conditions may not be captured by using conventional mathematical programming techniques. As such intelligent algorithms are used in the present study, which can overcome such limitations as they work with real datasets.

In this work, two machine learning models are used for selecting feature importance, namely, logistic regression and random forest decision tree. Hyperparameter tuning is the process of choosing the optimal combination of hyperparameters that reduces the loss function and improves the model accuracy. Upon hyperparameter tuning, the random forest showed high classification accuracy of 75%, whereas logistic regression showed 38% accuracy. So, a random forest decision tree algorithm is considered for finding the weights used in the objective function. The weight factors are found using random forests based on Gini impurity and are shown in Fig. 11. Gini impurity is the likelihood of incorrectly classifying the data point in a dataset and is used to decide the optimal splitting of the root node. Its value ranges between '0' to '0.5'.

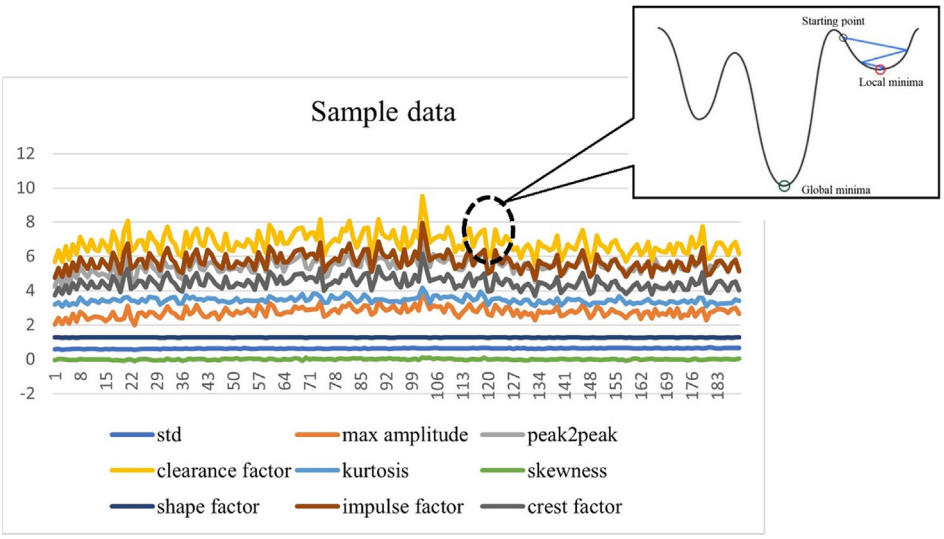


Fig. 10. Distribution of statistical features showing local optima and global optima.

Lower the Gini, better the split [43]. Further, the GWO algorithm is used to solve for optimal values of the statistical features. The optimized statistical features were used to search for a sensor network that has the optimal values in their statistical feature dataset for all the fault cases considered. Upon selecting the sensor network, fault classification is performed for each of these sensor networks. A sensor network with maximum classification accuracy is considered as the solution to the OSP problem. The optimization algorithm proposed in this study is given as follows.

Proposed optimization algorithm
Initialize Xi = [$S_{1,j}S_{2,j}S_{3,j}S_{4,j}S_{5,j}S_{6,j}S_{7,j}S_{8,j}S_{9,j}$]; (i = 1, 2, . . ., n), (j = 1, 2, . . ., m).
Apply GWO to solve Eq. (7).
Find optimal values of the statistical features.
Search for sensors sets that have optimized statistical features values.
Apply fault classification using individual sensor network from step 4.
Select a sensor network that has highest classification accuracy from step 5.

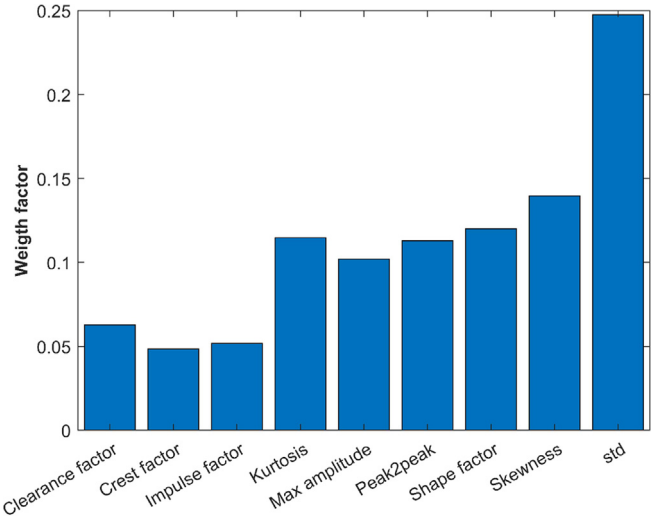


Fig. 11. Weight factor for statistical features.

These weights are given as input to equation (7), and GWO is used to find the optimal values of the statistical features. The optimal values of the statistical features are tabulated in Table 4. These optimal values are further used to select the sensors sets to find the optimal sensor network. A search is conducted in the main feature dataset to identify the sensors with the optimal feature values having 10% percentage variance. This search aims to identify the sensor with the optimized statistical values from GWO. Upon searching, the sensor network solutions are found for different fault conditions as M_1 = {L_2, L_3, L_4, L_5, L_7}, M_2 = {L_2, L_3, L_5, L_7}, M_3 = {L_2, L_3, L_4, L_5}, M_4 = {L_4, L_5, L_7, L_8}, and M_5 = {L_2, L_3, L_4, L_7}. The sensor networks M_1 and M_2 are shown in Fig. 12. To find the optimal sensor network, the classification accuracy of the sensor network is checked using random forest and deep multi-layer perceptron classification algorithms and is discussed in the next section.

4. Fault classification & result analysis

To check the robustness of the proposed optimization method, random forest (RF) with decision tree as its base learner and deep multi-layer perceptron (MLP) were used for fault classification. Random forest, a supervised machine learning algorithm, is an ensemble of random trees that are trained with the bagging method, shown in Fig. 13. The general idea of the bagging method is that random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction. Data points from each sample were chosen randomly with replacement, and those that failed to be part of the sample are out of bag points

Table 4
Optimal feature value after GWO.

Features	Optimum value
std	0.27874
Skewness	1.05136
Shape factor	2.15374
Kurtosis	6.26544
Peak2peak	3.63318
Max amplitude	0.028559
Clearance factor	1.30565
Impulse factor	5.16253
Crest factor	3.95397

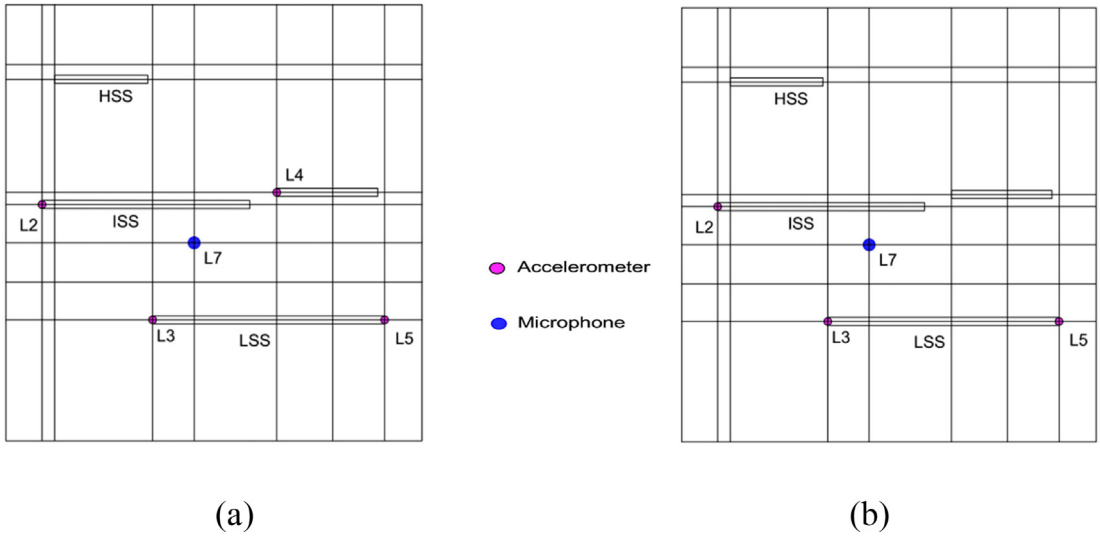


Fig. 12. Sensor network diagram for: (a) M_1 (b) M_2.

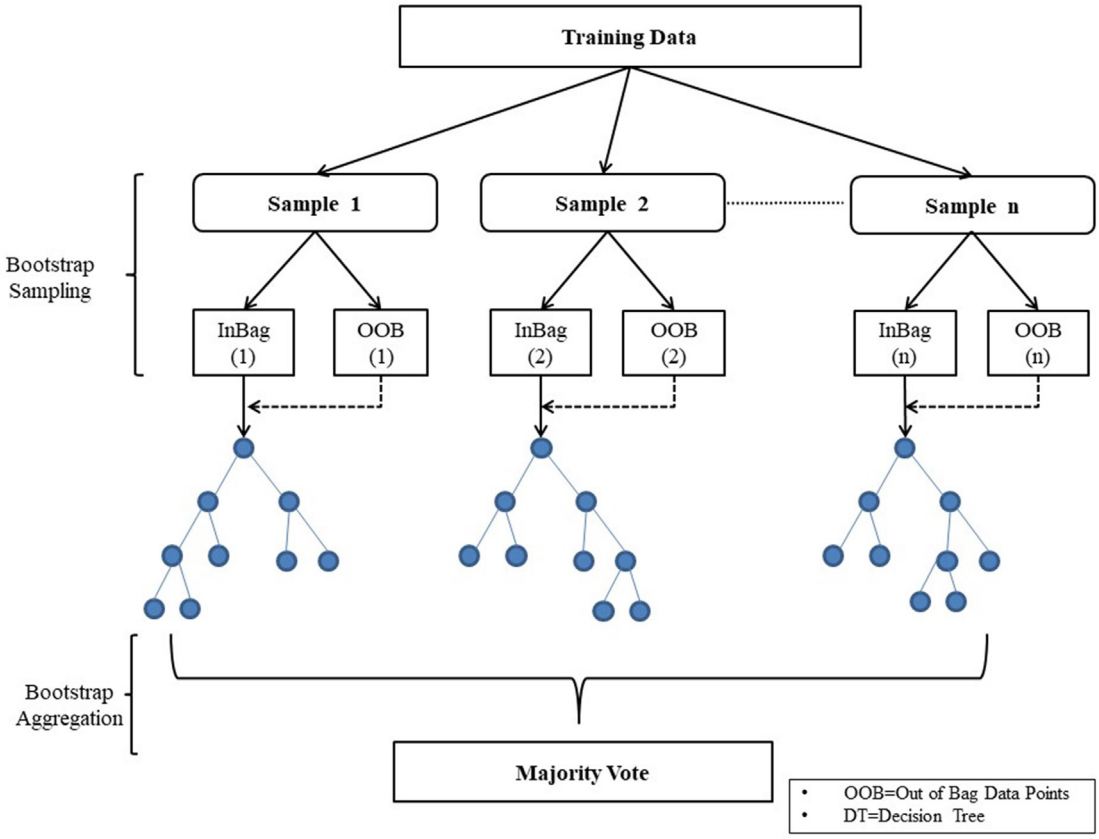


Fig. 13. Random forest algorithm flowchart.

(OOB), and it serves as the cross-validation for random forest models. OOB gives a low variance model and prevents data leakage. InBag points are chosen from each sample and used to train the decision tree. Deep multi-layer perceptron (MLP) is a class of feed-forward artificial neural networks (ANN) that consists of interconnected neurons transferring information to each other. MLP is a feed-forward neural network with one or more hidden layers. It consists of an input layer, at least one hidden layer and an output layer. Multiple layers of MLP and its non-linear activation function distinguish it from a linear perceptron and the data that is not linearly separable. The MLP architecture used in this work is shown in Fig. 14.

Hyperparameter tuning is the selection of optimal hyperparameters whose value controls the learning process. It minimizes a predefined loss function to give better results. The wrapper is used to connect the *scikit* learn [44] python library to *Keras* [45] for hyperparameters tuning, and *GridSearchCV* (10-fold cross-validation) is used to find different hyperparameters. Table 5 and Table 6 gives

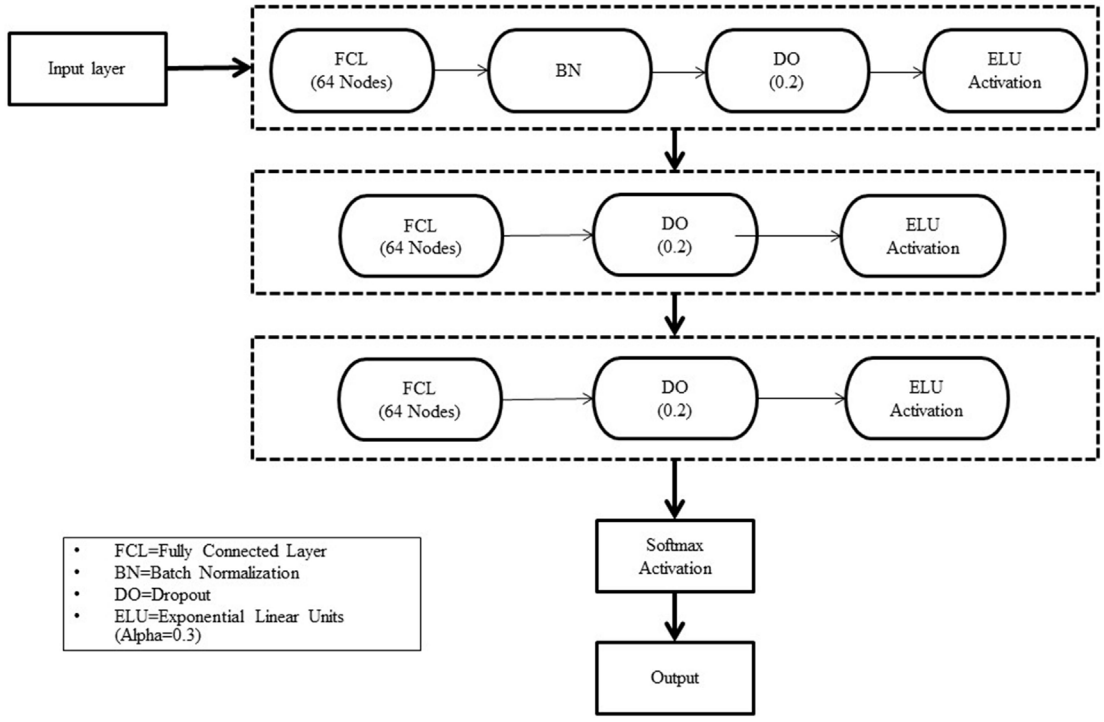


Fig. 14. Deep multi-layer perceptron representation.

Table 5
Hyperparameters for random forest algorithm.

Hyperparameter	Corresponding input
n_estimators	80
criterion	'Gini'
min_samples_split	3
max_depth	7

Table 6
Hyperparameters for deep multi-layer perceptron algorithm.

Hyperparameter	Corresponding input
No. of epoch	200
Batch size	128
Number of hidden layers	3
Dropout	0.2, 0.2, 0.2
Number of Neurons in the Hidden Layer	64, 64, 64
Neuron Activation Function	ELU
Optimization Algorithm	Adam
Learning rate	0.001

the details of hyperparameters for RF and MLP, respectively. In RF, as the depth increases, the possibility of having few points on leaf nodes increases and the interpretability of the model will decrease, thereby increasing the chances of overfitting of models due to noisy points. Dropout is a regularization technique of MLP to avoid overfitting, thus generalizing the model. Model is likely to result in better performance when dropout is used on a larger network, allowing the model to learn independently.

Fault classification is performed using the sensor network solutions obtained from the previous section using both RF and MLP algorithms. The classification algorithms are trained using 80% data, and the remaining 20% is used as test data. Train data (80%) is further divided into training and validation datasets, and the remaining 20% data is test data used to evaluate the performance of the trained model. In the training process, k-fold cross-validation is used, so that model is trained on different subsets of train data, resulting in a more generalized model further minimizing the chances of overfitting. In the present study, 10-fold cross-validation is used on the data sets. In the MLP model, three hidden layers are present in between the input and output layers. Before training the model, *GridSearchCV* is used to find the optimum hyperparameters on the training data set. During training, the model is evaluated on a validation dataset after each epoch. If the performance of the model on the validation dataset starts to degrade (e.g., loss begins to increase or accuracy begins to decrease), then the training process is terminated at that point. The corresponding classification accuracies obtained from RF and MLP algorithms are reported in Table 7. The classification accuracy of the initial sensor network, M_i, is also included in Table 7. From Table 7, the highest fault classification accuracy of 86.88% and

Table 7
Fault classification accuracy for RF and MLP algorithms.

Sensor set	Placement	Number	RF (%)	MLP (%)
M_1	L_2, L_3, L_4, L_5, L_7	5	86.88	88.34
M_2	L_2, L_3, L_5, L_7	4	82.03	83.20
M_3	L_2, L_3, L_4, L_5	4	83.47	82.65
M_4	L_4, L_5, L_7, L_8	4	85.64	86.11
M_5	L_2, L_3, L_4, L_7	4	85.06	85.23
M_i	L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8	8	75.00	78.90

88.34% is obtained for the sensor network M_1 using RF and MLP, respectively. This is due to the best match between the optimized statistical features values and the values of the statistical features for this sensor network. The least fault classification accuracy of 75% and 78.9% is observed for the initial sensor network, M_i, respectively, for RF and MLP. From step 5 of the proposed optimization algorithm discussed in section 3.3, the optimal sensor network can be identified as $M^* = \{L_2, L_3, L_4, L_5, L_7\}$ and is shown in Fig. 12 (a).

5. Conclusion

Sensor placement is of key importance in the construction and implementation of an effective condition monitoring system. In this paper, a novel sensor placement optimization (OSP) algorithm based on statistical features and a Grey wolf optimizer is presented for a wind turbine gearbox. An integrated condition monitoring scheme is performed using vibration and acoustic sensors. Experiments are conducted using multicomponent faults. Signal processing is performed using variational mode decomposition (VMD) analysis and Spearman coefficient correlation analysis followed by statistical feature extraction. GWO is used to find the optimum values of these statistical features, and these values are used to find the optimal sensor networks. The optimal sensor network and location is found based on the ability to classify faults accurately. Random forest (RF) and deep multi-layer perceptron (MLP) algorithms are used as classification algorithms. The following conclusions can be drawn:

The optimal configuration of the sensors can be obtained by optimizing the values of the statistical features. As a result, the number of sensors was reduced from eight to five.

Sensor network influences the quality of the statistical features and hence the fault classification accuracy.

Compared with the initial configuration of the sensors, the classification accuracy for the optimized sensor network increased by 15.84% and 11.96% for RF and MLP, respectively.

MLP showed higher fault classification accuracy than RF in all the cases.

The optimized sensor network showed the highest fault classification accuracy of 86.88% and 88.34% for RF and MLP, respectively.

In this current investigation, data-driven algorithms are employed for selecting the optimal sensor network through the optimal values of the statistical features. Since the goal of data-driven algorithms in this work is to find optimal sensor networks, improving and tuning the classification accuracy of the data-driven models is not considered. Further investigations are required to check consistency in results when the dataset changes. The influence of ICM schemes and extracted features on the proposed method also needs to be evaluated. The future investigations will also evaluate the external load variations (either electrical load or brake mechanism) similar to the real-time operating conditions.

CRediT authorship contribution statement

S.V.V.S. Narayana Pichika: Conceptualization, Methodology, Investigation, Writing – original draft. **Ruchir Yadav:** Software, Data curation. **Sabareesh Geetha Rajasekharan:** Supervision, Project administration, Writing – review & editing. **Hemanth Mithun Praveen:** Investigation, Methodology. **Vamsi Inturi:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Worldwide Wind Capacity Reaches 744 Gigawatts – An Unprecedented 93 Gigawatts added in 2020 - World Wind Energy Association. https://wwindea.org/worldwide-wind-capacity-reaches-744-gigawatts/ (accessed Jun. 02, 2021).

[2] Salameh JP, Cauet S, Etien E, Sakout A, Rambault L. Gearbox condition monitoring in wind turbines: a review. Mech Syst Signal Process 2018;111:251–64. https://doi.org/10.1016/j.ymssp.2018.03.052.

[3] Kankar PK, Sharma SC, Harsha SP. Fault diagnosis of ball bearings using machine learning methods. Expert Syst Appl 2011;38(3):1876–86. https://doi.org/10.1016/j.eswa.2010.07.119.

[4] Gómez MJ, Marklund P, Strombergsson D, Castejón C, García-Prada JC. Analysis of vibration signals of drivetrain failures in wind turbines for condition monitoring. Exp Tech 2021;45(1). https://doi.org/10.1007/s40799-020-00387-4.

[5] Peng Z, Kessissoglou NJ, Cox M. A study of the effect of contaminant particles in lubricants using wear debris and vibration condition monitoring techniques. Wear 2005;258(11–12):1651–62. https://doi.org/10.1016/j.wear.2004.11.020.

[6] Zeng XJ, Yang M, Bo YF. Gearbox oil temperature anomaly detection for wind turbine based on sparse Bayesian probability estimation. Int J Electr Power Energy Syst 2020;123(April). https://doi.org/10.1016/j.ijepes.2020.106233.

[7] Chen D, Lin J, Li Y. Modified complementary ensemble empirical mode decomposition and intrinsic mode functions evaluation index for high-speed train gearbox fault diagnosis. J Sound Vib 2018;424:192–207. https://doi.org/10.1016/j.jsv.2018.03.018.

[8] Yao J, Liu C, Song K, Feng C, Jiang D. Fault diagnosis of planetary gearbox based on acoustic signals. Appl Acoust 2021;181:108151. https://doi.org/10.1016/j.apacoust.2021.108151.

[9] Elasha F, Greaves M, Mba D, Fang D. A comparative study of the effectiveness of vibration and acoustic emission in diagnosing a defective bearing in a planetry gearbox. Appl Acoust 2017;115:181–95. https://doi.org/10.1016/j.apacoust.2016.07.026.

[10] Bravo-Imaz I, Davari Ardakani H, Liu Z, García-Arribas A, Arnaiz A, Lee J. Motor current signature analysis for gearbox condition monitoring under transient speeds using wavelet analysis and dual-level time synchronous averaging. Mech Syst Signal Process 2017;94:73–84. https://doi.org/10.1016/j.ymssp.2017.02.011.

[11] Vamsi I, Sabareesh GR, Penumakala PK. Comparison of condition monitoring techniques in assessing fault severity for a wind turbine gearbox under non-stationary loading. Mech Syst Signal Process 2019;124:1–20. https://doi.org/10.1016/j.ymssp.2019.01.038.

[12] Garcia Fernandez P, de-Juan A, Diez-Ibarbia A, Sanchez-Espiga J, Fernandez del Rincon A. Acoustic intensity technique applied to monitor planetary gears. Appl Acoust 2021;172. https://doi.org/10.1016/j.apacoust.2020.107621.

[13] Schmidt S, Heyns PS, Gryllias KC. A methodology using the spectral coherence and healthy historical data to perform gearbox fault diagnosis under varying operating conditions. Appl Acoust 2020;158:107038. https://doi.org/10.1016/j.apacoust.2019.107038.

[14] Peng Z, Kessissoglou N. An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis. Wear 2003;255(7–12):1221–32. https://doi.org/10.1016/S0043-1648(03)00098-X.

[15] Inturi V, Sabareesh GR, Supradeepan K, Penumakala PK. Integrated condition monitoring scheme for bearing fault diagnosis of a wind turbine gearbox. JVC/J Vib Control 2019;25(12):1852–65. https://doi.org/10.1177/1077546319841495.

[16] Yu L, Yao X, Yang J, Li C. Gear fault diagnosis through vibration and acoustic signal combination based on convolutional neural network. Information 2020;11(5). https://doi.org/10.3390/info11050266.

[17] Feng Z, Zhang D, Zuo MJ. Adaptive Mode Decomposition Methods and Their Applications in Signal Analysis for Machinery Fault Diagnosis: A Review With Examples. IEEE Access 2017;5:24301–31. https://doi.org/10.1109/ACCESS.2017.2766232.

[18] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Trans Signal Process 2014;62(3):531–44. https://doi.org/10.1109/TSP.2013.2288675.

[19] Kumar PS, Kumaraswamidhas LA, Laha SK. Selecting effective intrinsic mode functions of empirical mode decomposition and variational mode decomposition using dynamic time warping algorithm for rolling element bearing fault diagnosis. Trans Inst Meas Control Aug. 2018;41(7):1923–32. https://doi.org/10.1177/0142331218790788.

[20] Li H, Liu T, Wu X, Chen Q. Application of optimized variational mode decomposition based on kurtosis and resonance frequency in bearing fault feature extraction. Trans Inst Meas Control Sep. 2019;42(3):518–27. https://doi.org/10.1177/0142331219875348.

[21] Li H, Liu T, Wu X, Chen Q. An optimized VMD method and its applications in bearing fault diagnosis. Meas J Int Meas Confed 2020;166:108185. https://doi.org/10.1016/j.measurement.2020.108185.
[22] Zhong J, Wang D, Guo J, Cabrera D, Li C. Theoretical investigations on kurtosis and entropy and their improvements for system health monitoring. IEEE Trans Instrum Meas 2021;70:1–10. https://doi.org/10.1109/TIM.2020.3031125.
[23] Long J, Mou J, Zhang L, Zhang S, Li C. Attitude data-based deep hybrid learning architecture for intelligent fault diagnosis of multi-joint industrial robots. J Manuf Syst 2020. https://doi.org/10.1016/J.JMSY.2020.08.010.
[24] Jedliński Ł, Jonak J. Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform. Appl Soft Comput 2015;30:636–41. https://doi.org/10.1016/j.asoc.2015.02.015.
[25] Lopez-Flores DR, Duran-Gomez JL, Chacon-Murguia MI. A mechanical sensorless MPPT algorithm for a wind energy conversion system based on a modular multilayer perceptron and a processor-in-the-loop approach. Electr Power Syst Res 2020;186:106409. https://doi.org/10.1016/j.epsr.2020.106409.
[26] Tang X, Gu X, Rao L, Lu J. A single fault detection method of gearbox based on random forest hybrid classifier and improved Dempster-Shafer information fusion. Comput Electr Eng 2021;92:107101. https://doi.org/10.1016/j.compeleceng.2021.107101.
[27] Li C, Sanchez RV, Zurita G, Cerrada M, Cabrera D, Vásquez RE. Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals. Mech Syst Signal Process 2016;76–77:283–93. https://doi.org/10.1016/j.ymssp.2016.02.007.
[28] Wei Y, Yang Y, Xu M, Huang W. Intelligent fault diagnosis of planetary gearbox based on refined composite hierarchical fuzzy entropy and random forest. ISA Trans 2021;109:340–51. https://doi.org/10.1016/j.isatra.2020.10.028.
[29] Singh A, Parey A. Gearbox fault diagnosis under fluctuating load conditions with independent angular re-sampling technique, continuous wavelet transform and multilayer perceptron neural network. IET Sci Meas Technol 2017;11(2):220–5. https://doi.org/10.1049/iet-smt.2016.0291.
[30] Pan H, Wei X. Optimal placement of sensor in gearbox fault diagnosis based on VPSO, *Proc. – 2010 6th Int. Conf. Nat. Comput. ICNC 2010*, vol. 7, no. Icnc, pp. 3383–3387, 2010, doi: 10.1109/ICNC.2010.5583680.
[31] Zhao Z, Xu Q, Jia M. Sensor network optimization of gearbox based on dependence matrix and improved discrete shuffled frog leaping algorithm. Nat Comput 2016;15(4):653–64. https://doi.org/10.1007/s11047-015-9515-4.
[32] Vanraj S, Dhami S, Pabla BS. Optimization of sound sensor placement for condition monitoring of fixed-axis gearbox, *Cogent Eng*, 4 (1), 2017, 10.1080/23311916.2017.1345673.
[33] GuiLan W, HongShan Z, ShuangWei G, ZengQiang Mi. Numeric optimal sensor configuration solutions for wind turbine gearbox based on structure analysis. IET Renew Power Gener 2017;11(12):1597–602. https://doi.org/10.1049/iet-rpg.2016.0157.
[34] Inturi V, Shreyas N, Chetti K, Sabareesh GR. Comprehensive fault diagnostics of wind turbine gearbox through adaptive condition monitoring scheme. Appl Acoust 2021;174:107738. https://doi.org/10.1016/j.apacoust.2020.107738.
[35] Antoniadou I, Manson G, Staszewski WJ, Barszcz T, Worden K. A time–frequency analysis approach for condition monitoring of a wind turbine gearbox under varying load conditions. Mech Syst Signal Process Dec. 2015;64–65:188–216. https://doi.org/10.1016/J.YMSSP.2015.03.003.
[36] An X, Yang J. Denoising of hydropower unit vibration signal based on variational mode decomposition and approximate entropy. Trans Inst Meas Control 2016;38(3):282–92. https://doi.org/10.1177/0142331215592064.
[37] Tang G, Luo G, Zhang W, Yang C, Wang H. Underdetermined blind source separation with variational mode decomposition for compound roller bearing fault signals, *Sensors*, 16 (6), 2016, 10.3390/s16060897.
[38] Research Design and Statistical Analysis: Third Edition - 3rd Edition. https://www.routledge.com/Research-Design-and-Statistical-Analysis-Third-Edition/Myers-Well-Jr/p/book/9780805864311 (accessed Jun. 02, 2021).
[39] Shang H, Li Y, Xu J, Qi B, Yin J. A novel hybrid approach for partial discharge signal detection based on complete ensemble empirical mode decomposition with adaptive noise and approximate entropy. Entropy 2020;22(9). https://doi.org/10.3390/e22091039.
[40] Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. Adv Eng Softw 2014;69:46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.
[41] Yang H-T, Yang P-C, Huang C-L. Evolutionary programming based economic dispatch for units with non-smooth fuel cost functions. IEEE Trans Power Syst 1996;11(1):112–8. https://doi.org/10.1109/59.485992.
[42] Park J-B, Lee K-S, Shin J-R, Lee KY. A particle swarm optimization for economic dispatch with nonsmooth cost functions. IEEE Trans Power Syst 2005;20(1):34–42. https://doi.org/10.1109/TPWRS.2004.831275.
[43] D'Ambrosio A, Tutore VA. Conditional classification trees by weighting the gini impurity measure. In: New Perspectives in Statistical Modeling and Data Analysis. p. 273–80.
[44] Getting Started — scikit-learn 0.24.2 documentation. https://scikit-learn.org/stable/getting_started.html (accessed Jun. 02, 2021).
[45] The Sequential model. https://keras.io/guides/sequential_model/ (accessed Jun. 02, 2021).