Wasit Journal of Engineering Sciences
Journal homepage: https://ejuow.uowasit.edu.iq
Vol. 14 No. 2 June 2026, pp.80-90
https://doi.org/10.31185/ejuow.v14i2.768

ISSN 2305-6932
Original Research

Intelligent Diagnosis of Bearing Faults Using Feature Selection Based on Minimum Redundancy Maximum Relevance and Neural Network Optimization with Particle Swarm Method

Mohammed M. Thajeel[1], Hussein Al-Bugharbee[1]
[1]Department of Mechanical Engineering, Wasit University, Wasit, Iraq

Corresponding Author Email: Std2023203.M.M@uowasit.edu.iq

Received Jul.21, 2025 Revised Sept.25, 2025 Accepted. Oct.16, 2025 Online Jun.1, 2026

ABSTRACT

Rolling element bearings play a crucial supporting role in rotating machines in many manufacturing processes. The failure of rolling bearings can result in catastrophic damage and human casualties, and their condition has a direct impact on the safe functioning of rotating machinery. Numerous bearing detection techniques have been designed to prevent such outcomes and avoid unplanned downtime. Techniques that have no dependence on the manual selection of attributes or the classifier parameters are still required. In order to meet this requirement, the current method incorporates the following algorithms: Max Relevance Min Redundancy (mRMR) for automated feature ranking, Artificial Neural Network (ANN) for bearing condition categorization, and Particle Swarm Optimization (PSO) for tuning the ANN hyperparameters, including the hidden layer size, the number of neurons in each layer, and the learning rate. The performance of the hybrid model is investigated using different scenarios covering noisy samples and variable rotational speed. In addition, the performance of the present hybrid technique is compared with a standalone ANN as well as a hybrid ANN-Bayesian to highlight the strength points in the methodology and identify the scenarios where the present model performs better. The outcomes demonstrate that this approach can, in a number of testing scenarios, achieve a very high accuracy percentage and outperforms other comparable models.

Keywords: Condition monitoring, metaheuristic, feature ranking, PSO

1. Introduction

Rotating machinery is employed in modern industrial systems in the chemical industry, electric power, mining, aviation, etc. The rotating machinery system gets more complicated as science and technology advance [1]. Rolling element bearings cause almost 40% of the mechanical problems in rotating machines [2], [3]. As a critical component, the rolling bearing is exposed to failure due to constant harshness and complicated loads [4], [5]. The element rolling bearing failures affect the dependability and stability of rotating equipment systems and might cause high maintenance costs, safety risk, or tragic events [6]. Thus, rolling element bearing defect diagnosis is necessary. Many problems in fault diagnostics and classification have been solved by utilizing machine learning (ML) in the past decades [7]. The quantity and quality of the input characteristics determine the resilience of the trained model [8]. Feature selection is crucial for machine learning efficacy, diagnostic speed, and accuracy [9], [10]. Bearing defect diagnostics utilized many feature selection techniques with good results. Sharma et al. [11] utilized attribute filters to reduce dimensionality and obtain accurate classification results, while the Laplace score was employed in [12] to select critical features from multiple domains. Imane et al. [13] combined standard deviation and Random Forest to improve diagnostic robustness. Inadequate or redundant features can lead to poor classification, increased complexity, and inefficient computation. Van et al.[14] employed the Maximum Relevance Minimum Redundancy (mRMR) method to eliminate the unnecessary and duplicate features, allowing for the selection of an ideal subset. This subset significantly improved the classification accuracy in bearing fault diagnosis. In [15], the combination of the mRMR

algorithm and the Random Forest algorithm for feature selection enhances the model performance with the lowest standard deviation value. The results reveal that this integration improved the accuracy and confirmed its ability in intelligent fault diagnosis. Thus, effective feature selection is required for building reliable diagnostic models. Artificial Neural Networks (ANNs) are an effective machine learning approach for handling complicated issues [16]. The hyperparameters in neural networks play a key role in simulating performance [17]. It takes a lot of time and effort, mostly based on empirical techniques, to tune these parameters. Moreover, because there is no theoretical support for this endeavor, the optimization step is neither as visible nor as interpretable. As a result, tuning and optimization procedures need to be more methodical and efficient to fully utilize the neural networks in fault diagnosis [18]. A group of researchers utilized heuristic optimization algorithms for training the neural networks. Unal et al. [19] employed the Genetic Algorithm (GA) to optimize the structure of the ANN, and the latter is used for classifying the bearing conditions. In [20], the researchers utilized Bayesian Optimization to optimize the hyperparameters and architecture of the convolutional neural network. This process enhanced the performance and achieved a classification accuracy of 99.4% in rotating machinery without human intervention in hyperparameter adjustment. Tian et al.[21] optimized a CNN-LSTM model for bearing defect identification using hybrid particle swarm optimization (HPSO). This optimization allows the model to modify parameters to solve nonlinear and complex multivariate optimization problems, maximizing diagnostic performance. Chen et al [22] have used the Improved Genetic Algorithm and back propagation neural network (IGA-BPNN) model for bearing defect diagnosis was proposed to tune the weights, biases, and hyperparameters, and confirmed the effectiveness on the CWRU dataset.

The current technique presents a methodology that combines the mRMR and hybrid ANN for fault diagnosis in rolling element bearings. The authors aim to achieve several contributions through the following:

1. Proposing a framework including extracting 15 temporal features and ranking them according to their importance in the classification using mRMR, and submitting either a subset or all of them to the hybrid PSO-ANN classifier.
2. Investigate the use of three different training algorithms (Scaled Conjugate Gradient algorithm, Resilient Backpropagation algorithm, and One Step Secant algorithm) for training the ANN to enhance the model's accuracy and stability.
3. Validating the methodology by employing two vibration signal sources (CWRU and Wasit University test rig) bearing datasets to reveal the generalization of the technique across different scenarios of bearing conditions.
4. The performance of the present methodology has also been compared with a standalone ANN and a hybrid Bayesian-ANN model, and two other standalone ANN models from the literature for CRWU datasets.
5. The present technique has also been tested under the use of a noisy testing sample to investigate its robustness against real-world noise.

The remaining sections of this document are organized below. In Section II, the materials and methods are illustrated. Section III explains the performance metrics. The results and discussion are presented in Section IV. In the last section, Section V, the conclusion is given.

2. Materials and Methods

2.1. Datasets

For the data collection stage, a widely used experimental platform and our test rig are provided. The specifications of the bearing dataset utilized in the two scenarios are clarified below:

Case 1) The well-known datasets of bearing failure from the Case Western Reserve University (CWRU) [23] were utilized to evaluate our proposed approach. Figure 1 illustrates the experimental platform's structure, which includes a 2 HP motor, control electronics, a dynamometer, and an encoder. A 6205-2RSJEMSKF deep-groove ball bearing was employed in the system. Four kinds of vibration signals, including Healthy (H), inner race fault (IRF), ball fault (BF), and outer race fault (ORF), were recorded utilizing the acceleration sensors. For every fault type, three separate fault sizes (0.007, 0.014, 0.021) inches were adopted to reflect varying degrees of fault severity. Different rotational speed was conducted (1730, 1750, 1772, 1797) rpm. Datasets are splitted into shorter, non-overlapped segments of 2048 points.

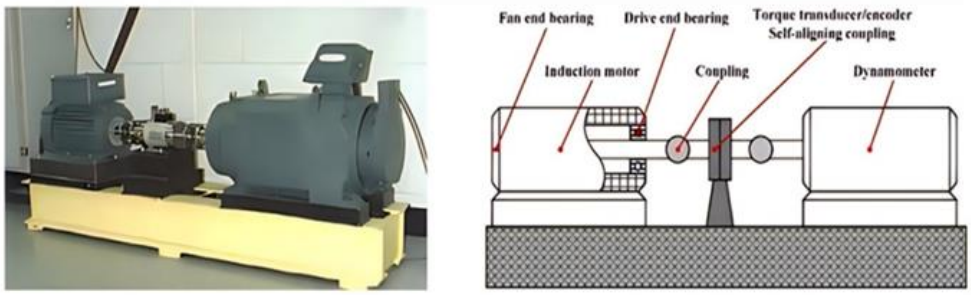
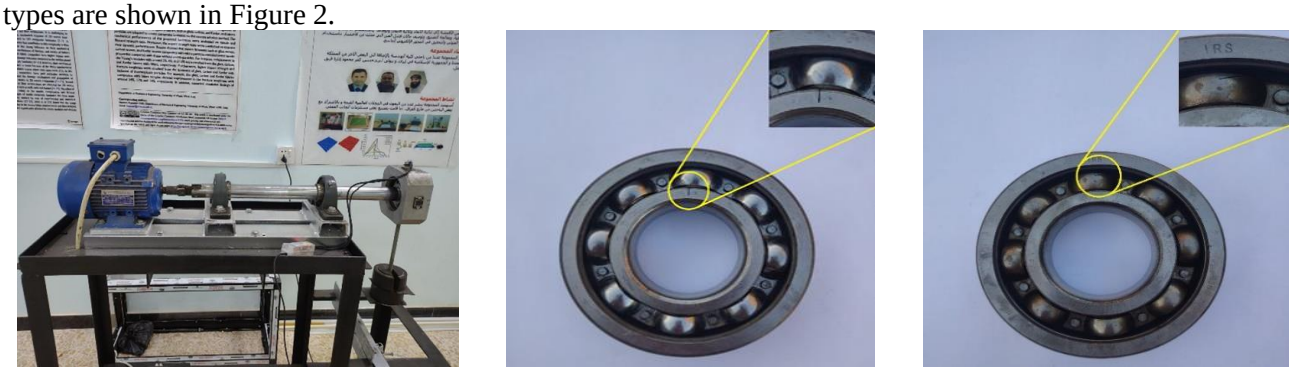


Figure 1. The CWRU experimental apparatus [24]

Case 2) A test rig has been built in the laboratory of mechanical engineering at Wasit University for the purpose of conducting the experiments. The test apparatus incorporates a coupling connected bearing shaft to a 1 HP induction motor with two poles and three phases. In this configuration, the 6308-ball bearing is employed as the test bearing. Measurements are taken by utilizing an ADXL355 logger accelerometer. The experiments are carried out at a range of rotation speeds of 605, 1020, and 1500 rpm for four bearing conditions, namely H, IRF, BF, and ORF. A 0.2 mm defect size was introduced by a wire electrical discharge machine (EDM). The variable frequency drive is employed to regulate the speed. The datasets were collected utilizing a sampling frequency of 4 kHz and a data sample of 2048 points. All the experimental apparatus and some of the produced bearing types are shown in Figure 2.



(a) (b) (c)

Figure 2. (a) Experimental Test Rig, (b) bearing with IRF of 0.2 mm, (c) bearing with BF of 0.2 mm.

In the present approach, twenty percent of the data is employed as an independent testing part, while the rest is utilized as a training set. The remaining 80% training set is further partitioned into a 60% training part and a 20% validation part using the 4-fold cross-validation scenario. This scenario is commonly utilized to validate the models, and it involves folding training data into four almost equal portions. In each of the four iterations of this scenario, one fold is employed as validation data and the other three folds as training data. K-fold cross-validation creates several training and data sets to ensure the model can handle all the data during multiple assessment rounds. Thus, the model's generalizability may be calculated more precisely without bias from a single data split [25].

2.2. Feature extraction and Feature selection

At first, the collected vibration data are divided into training and testing samples. Besides the original data, additional datasets are generated after adding Gaussian noise with different signal-to-noise ratios to the clean data. This is useful when testing the method on a noisy sample, as it will be mentioned later. From these clean and noisy samples, fifteen temporal features are extracted as in Table 1. Features are normalized using the Min-Max technique to ensure that all features have the same scale [26]. These features are then ranked according to their importance in classification using the mRMR algorithm. This strategy sorts the obtained features based on how relevant they are to the categorization step. By plugging a given feature vector f_{ij} into equation 1, its feature relevance could be determined [27].

$$f^{MRMR}(f_{ij}) = I(Y_i, f_{ij}) - \frac{1}{|S|}\sum_{F_S \in S} I(F_S, f_{ij}) \quad (1)$$

Where the j^{th} feature of the i^{th} is represented by f_{ij}, the mutual information is represented by $I(Y_i, f_{ij})$. S is a set of selected characteristics. $|S|$ number of attributes in S.

Table 1. Typical Time domain features

Features	Equation	Features	Equation
Standard Deviation	$Std = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}$	Peak to Peak	$P - to - P = x_{max} - x_{min}$
Mean	$Mean = \frac{1}{N}\sum_{i=1}^{N} x_i$	Peak value	$Peak\ value = max\|x\|$
Max	$Max = max\ (x_i)$	Crest Factor	$CF = \frac{max\|x\|}{RMS}$
Min	$Min = min\ (x_i)$	Shape Factor	$Shape\ factor = \frac{RMS}{\frac{1}{N}\sum_{i=1}^{N} \mid x_i \mid}$
Root Mean Square	$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$	Impulse Factor	$Impulse\ Factor = \frac{max \mid x \mid}{\frac{1}{N}\sum_{i=1}^{N} \mid x_i \mid}$
Skewness	$Skewness = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^3}{(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2)^{3/2}}$	Mobility	$Mobility = \sqrt{\frac{var(x'(t))}{var(x(t))}}$
Variance	$Variance = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$	Complexity	$Complixity = \sqrt{\frac{Mobility(x'(t))}{Mobility(x(t))}}$
Kurtosis	$Kurtosis = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}{(\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2)^2}$		

2.3. ANN architecture

ANNs are a subset of ML algorithms inspired by the way the human brain works. In the human brain, a wide variety of neurons can be found. Throughout the artificial neural networks' (ANNs) development process, nodes connect these neurons. Each node processes data of a certain kind and executes a unique function. Adjusting the weights of network connections to achieve a specific objective is the responsibility of the learning process [28]. A basic artificial neural network (ANN) consists of three layers, namely the input layer, which takes in data as input, the hidden layers, which process that data, and the output layer, which gives the results. A specific decimal weight is learned by each layer of the neural network to be employed during the learning phase [29]. Before being added to the bias term b_j, the inputs to the j^{th} neuron, denoted as $X = [x_1, x_2, \ldots, x_n]$, will be multiplied by the weights vector $W_j = [w_{j1}, w_{j2}, \ldots, w_{jn}]$.

$$Y_j = \sum_{i=1}^{n} w_{ji}\, x_i + b_j \quad (2)$$

Where Y_j stands for the total weight of the input [30]. Figure 3 (a) presents the structure of the proposed ANN model. In neural networks, the values of hyperparameters significantly affect the performance. The optimal hyperparameter set applicable for all datasets and models does not exist. Therefore, optimization methods can be employed to address the hyperparameter optimization problem [31]. In this study, the following hyperparameters of the neural network (learning rate, number of hidden layers, and neurons in each layer) are considered to be optimized by utilizing the PSO algorithm. In Table 2, the hyperparameters and their ranges in the optimization stage are presented.

Table 2. Hyperparameters and their ranges for optimization

Hyperparameters	The Ranges
Learning rate	(0.001-0.1)
No. of hidden layers	(1-3)
No. of neurons	(2-100)

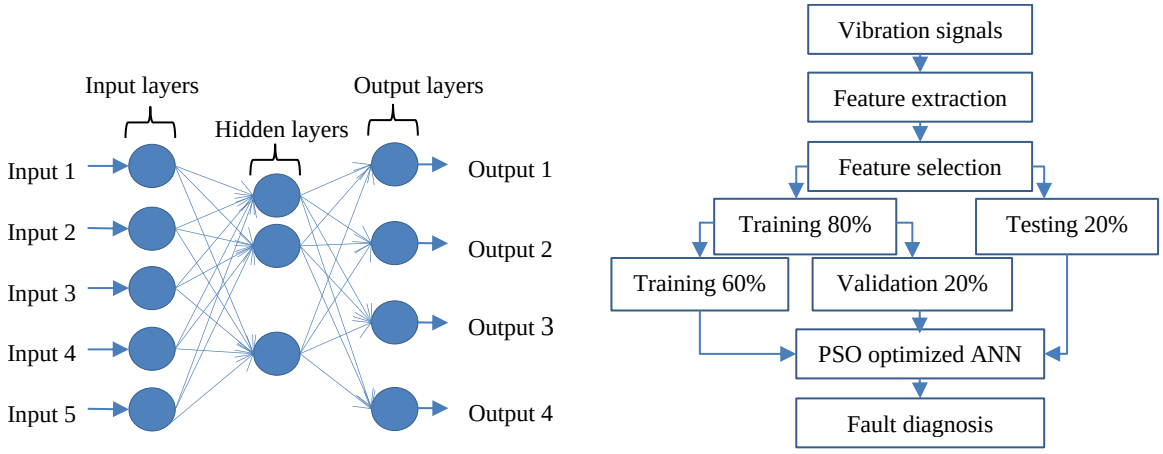
2.4. Particle Swarm Optimization (PSO) algorithm

The Particle Swarm Optimization (PSO) is an algorithm for evolution that was designed based on the social behavior of flocking birds as its primary source of inspiration [32]. The PSO, like the other evolutionary algorithms, seeks the best global optimum by starting with initial solutions. It specifically looks at a bunch of particles as potential answers. Each particle navigates the search space with a certain speed in pursuit of the optimal solution. A particle's best solution up to this point is donated as *pbest*. *"gbest"* refers to the particle that performs the best out of all *pbest*. Every particle needs to consider its present location, velocity, and distance to *pbest* and *gbest* if it wants to change its position. The following equation was provided to represent this adjustment [33]:

$$v_i^{t+1} = w \times v_i^t + c_1 \times rand \times (pbest_i - x_i^t) + c_2 \times rand \times (gbest - x_i^t) \quad (3)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (4)$$

Here, v^t_i is the particle *i's* velocity at *t* iteration, *w* stands for the inertia weight, the acceleration coefficients are c_1 and c_2, rand is an arbitrary number between 0 and 1, x^t_i is the particle *i's* current position at *t* iteration, $pbest_i$ is the particle *i's pbest* at *t* iteration, and *gbest* is the optimized solution thus far. In the suggested approach, the values of the parameters *w*, c_1, and c_2 are selected to be (0.7, 1.5, 1.5), respectively, and the number of particles is set to [50, 40, 30, 20]. Equation (3) is utilized to obtain the velocities of the article in each iteration. After that, equation (4) is employed to determine the particle locations. Continuously shifting the positions of the particles will continue until the halting condition is satisfied. The diagram of the suggested model is presented in Figure 3 (b).



(a) (b)

Figure 3. (a) The architecture of the ANN model, (b) the methodology flowchart.

3. Performance metrics

Four common ML assessment metrics were utilized to evaluate the classification results. These metrics assessed ML algorithm performance. The assessment metrics are accuracy, precision, recall, and F1-score [34]. The following four equations provide the mathematical notation of these metrics:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (8)$$

TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. The ratio of the algorithm's correct predictions to its total predictions is called accuracy. By dividing the total number of positive samples by the number of successfully recognized ones, recall measures the model's ability to detect positive cases. The accuracy rate of positive class predictions is measured by precision, another evaluation parameter. Lastly, the F1 score is a measure of the algorithm's accuracy that combines recall and precision [35].

4. Results and Discussion

4.1. CWRU bearing datasets

4.1.1. Data preprocessing and feature extraction

The first bearing datasets in this study are provided by CWRU. In Figure 4, different types of faults are not clearly distinguishable in time-domain vibration signal diagrams. Therefore, this data must be subjected to the classifier during the training process.

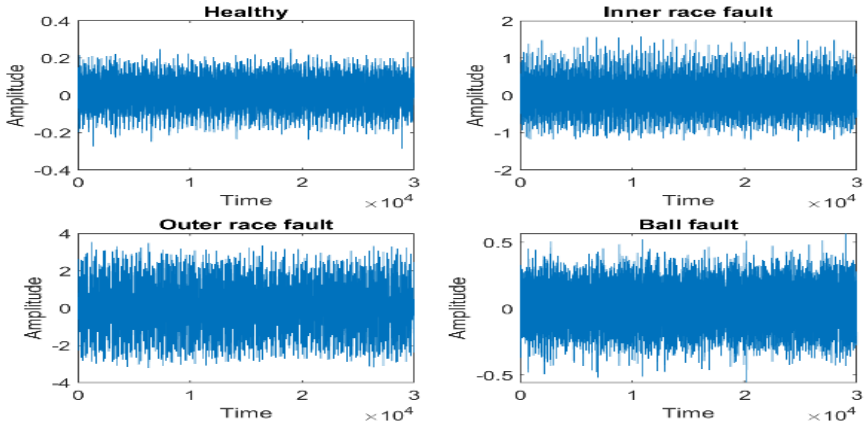


Figure 4. Acceleration bearing vibration signals of the four different conditions.

4.1.2. Feature selection

Table 3 shows the highest-ranked features that appear frequently in datasets belonging to various rotational speeds. The mobility, complexity, and mean were frequently selected across all rotational speeds.

Table 3. The common features subsets: CWRU case study.

Features	Speed (rpm)			
	1730	1750	1772	1797
Std		•	•	•
Mean	•	•	•	•
Max	•	•		
Min			•	
Crest Factor	•			
Shape Factor				•
Mobility	•	•	•	•
Complexity	•	•	•	•

*The features in bold were selected across all speeds

4.1.3. ANN Hyperparameters tuning based on PSO

Figure 5 shows the convergence curve during the tuning of ANN hyperparameters using PSO for different particle swarms (20, 30, 40, and 50). The swarm 30 was selected as it converges faster compared to swarm 40 and 50. The maximum iteration for PSO optimization was set as 60; however, the process can reach the target loss function in fewer iterations. All results are obtained using Matlab software version 2020a.

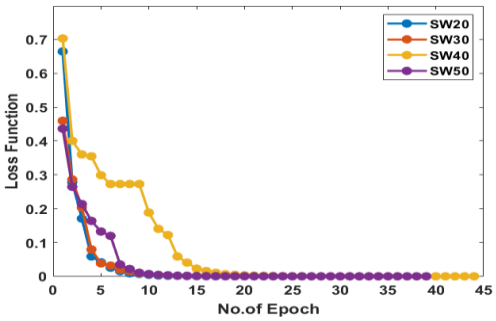


Figure 5. The convergence curve of PSO with different swarm sizes at 1730 rpm.

The optimized parameters for all the hybrid PSO-ANN models at different motor speed cases are shown in Table 4. It can be seen that the classification accuracy is very high in general, and this is applicable as the training and testing samples were free of additional noise. This agrees with many results in the literature. The high accuracy provided by the model across different rotational speeds, feature vector lengths, and training functions reveals the effectiveness of the proposed model for accurate classification of laboratory data.

Table 4. Classification results of CWRU bearing datasets (Swarm 30), top five features.

Speed (rpm)	Optimum Hyper-parameters			Evaluation Metrics of ANN Classifier			
	No. of Layers	No. of Neurons	LR	Accuracy	Precision	Recall	F1score
1730	3	100	0.017	98.94	98.95	98.95	98.95
1750	2	92	0.077	100	100	100	100
1772	2	68	0.072	100	100	100	100
1797	2	88	0.053	100	100	100	100

The performance of the present technique is also evaluated under the presence of noise, where it is trained on a mixed sample including clean and noisy datasets. These noisy datasets are generated by adding Gaussian noise at different signal-to-noise ratio (SNR). For the hardest noisy condition (i.e., SNR is 0 and -5 dB), the present model showed better accuracy in comparison to that obtained from the ANN optimized with Bayesian in a number of testing scenarios, as in Table 5. It is worth noting that both models were trained and tested on the same data samples to secure a fair comparison.

Table 5. Classification results of CWRU bearing datasets (noisy)

Scenario	(mRMR+Hybrid Bayesian-ANN)	Present Model
RPM=1750, No=10, SNR=0, trainscg	78.52%	**84.97%**
RPM=1730, No=10, SNR=-5, trainscg	75.05%	**78.63%**
RPM=1750, No=15, SNR=0, trainscg	95.45%	**89.84%**
RPM=1797, No=10, SNR=-5, trainscg	69.68%	**89.05%**

In addition, the present model has outperformed other model configurations in several scenarios, as shown in Table 6. The percentage accuracy values of the present model were compared to those obtained from Model A (i.e., mRMR+Hybrid Bayesian-ANN), Model B (i.e., ANN (200,200)) [36], and Model C (i.e., ANN (18,20)) [37].

Table 6. Scenarios where Present models outperform other models -CWRU bearing datasets

Scenario (RPM, No. of features, SNR (dB), training function)	Model A	Model B	Model C	Present Model
RPM=1730, No=10, SNR=0, trainscg	74.95%	75.47%	75.37%	**85.47%**
RPM=1750, No=10, SNR=-5, trainscg	75.13%	69.84%	63.39%	**75.77%**
RPM=1772, No=10, SNR=0, trainrp	74.81%	73.76%	53.76%	**93.23%**
RPM=1797, No=10, SNR=-5, trainscg	72.84%	69.68%	88.21%	**89.05%**

4.2. Test rig bearing datasets

4.2.1. Feature selection

The test rig bearing dataset is utilized to extract fifteen time domain features from vibration signals. The most frequently selected top features are shown in Table 7. Root mean square (RMS) was the most popular attribute across all rotational speeds, demonstrating its usefulness in bearing defect diagnosis. The experimental datasets are also segmented identically to the CWRU datasets.

Table 7. The common features subsets: Wasit University case study.

Features	Speed (rpm)		
	605	1020	1500
Mean		•	•
Max	•	•	
Min	•		•
RMS	•	•	•
Skewness	•		
Kurtosis		•	
Shape Factor			•
Mobility	•	•	
Complexity			•

4.2.2. ANN Hyperparameters tuning based on PSO

Table 8 shows that the hybrid PSO-ANN has performed exceptionally well in classifying vibration data collected from the Wasit University test rig at speeds of (605, 1020, and 1500) rpm. Even very shallow to moderately deep designs may successfully capture fault patterns, as most configurations achieved 100% accuracy, especially with networks with 1 to 3 hidden layers and 35 to 92 neurons. Confirming the necessity for adaptive tuning based on network topology, the learning rate fluctuated substantially from 0.001 to 0.1 while maintaining a consistently high level of accuracy. Most settings were 100% accurate, while some dropped to 98–99%. These decreases may be attributable to inadequate learning rates or network topologies, but still indicate model reliability.

Table 8. Classification results of experimental bearing datasets (Swarm 30), top five features

Speed (rpm)	Optimum Hyper-parameters			Evaluation Metrics of ANN Classifier			
	No. of Layers	No. of Neurons	LR	Accuracy	Precision	Recall	F1score
605	1	35	0.022	100	100	100	100
1020	2	70	0.001	100	100	100	100
1500	3	15	0.006	100	100	100	100

Table 9 provides a comparison summary where the present model outperforms the mRMR+Hybrid Bayesian-ANN in a number of scenarios, including very noisy operating conditions.

Table 9. Classification results of Wasit University bearing datasets (noisy).

Scenario(RPM, No. of features, SNR (dB), training function)	(mRMR+Hybrid Bayesian-ANN)	present model
RPM=605, No=5, SNR=-5, trainscg	48%	49.16%
RPM=605, No=10, SNR=0, trainscg	75.16%	76.95%
RPM=605, No=10, SNR=-5, trainscg	64.32%	67.58%
RPM=605, No=15, SNR=-5, trainscg	51.68%	55.37%
RPM=1020, No=10, SNR=0, trainscg	87.77%	91.49%
RPM=1020, No=15, SNR=0, trainscg	88.94%	95.21%
RPM=1500, No=5, SNR=5, trainscg	99.47%	100%
RPM=1500, No=5, SNR=-5, trainscg	66.49%	72.98%
RPM=1500, No=10, SNR=10, trainscg	99.47%	100%
RPM=1500, No=10, SNR=0, trainscg	90.11%	90.64%

Although the present model struggles in some scenarios of heavy noise levels, it still outperforms the performance where the Bayesian algorithm has been used for tuning the ANN hyperparameters. Table 10 captures the optimal hyperparameters for two case studies.

Table 10. The optimal hyperparameters of the ANN for both datasets.

Data source	No. of hidden layers	No. of neurons	Learning rate
CWRU	2	17-100	0.02-0.1
Test rig	1-3	15-92	0.05-0.1

5. Conclusions

This paper proposes a rolling-element bearing fault diagnosis model utilizing the mRMR technique for feature selection and an Artificial Neural Network (ANN) optimized by Particle Swarm Optimization (PSO). The accuracy of the present model is tested at different scenarios covering various training functions, the signal cleanliness by using various signal-to-noise ratios, a variety of machine rotational speeds, and a variety of feature numbers. It was taken into consideration the use of cross-validation with 4-folds and an independent testing set in order to enhance the reliability of the model and to minimize the overfitting effect. The results showed that the training function of the model does not provide much variation in the accuracy. However, the Scaled Conjugate Gradient algorithm (trainscg) was found to be the best in general. For the case study of the CRWU datasets, the present model was compared to three other models. It shows better performance in a number of testing scenarios, including scenarios where the noise level is high. For the case study of the Wasit University Test rig datasets, the current model provides higher accuracy compared to the accuracy of another model that uses mRMR with hybridized Bayesian–ANN in a number of scenarios. Despite those scenarios where the present model shows better accuracy, it still has some limitations and requires further investigation, which can be conducted in another work by the authors.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Funding Information

No funding was received from any financial organization to conduct this research.

Author Contributions

Hussein Razzaq Al-Bugharbee (conceptualization, methodology, supervision, and final review). Mohammed Mutar (literature review, experimental work, software, analysis, results, and conducting the first draft of the manuscript). Both authors discussed the results and revised the final version of the manuscript.

References

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," Oct. 2006. doi: 10.1016/j.ymssp.2005.09.012.

[2] L. Ciabattoni, F. Ferracuti, A. Freddi, and A. Monteriu, "Statistical Spectral Analysis for Fault Diagnosis of Rotating Machines," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4301–4310, May 2018, doi: 10.1109/TIE.2017.2762623.

[3] V. Tra, J. Kim, S. A. Khan, and J. M. Kim, "Bearing fault diagnosis under variable speed using convolutional neural networks and the stochastic diagonal levenberg-marquardt algorithm," *Sensors (Switzerland)*, vol. 17, no. 12, Dec. 2017, doi: 10.3390/s17122834.

[4] X. Zhang, Q. Miao, H. Zhang, and L. Wang, "A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery," *Mech Syst Signal Process*, vol. 108, pp. 58–72, Aug. 2018, doi: 10.1016/j.ymssp.2017.11.029.

[5] M. Cerrada *et al.*, "A review on data-driven fault severity assessment in rolling bearings," Jan. 15, 2018, *Academic Press*. doi: 10.1016/j.ymssp.2017.06.012.

[6] D. Wang and K. L. Tsui, "Statistical modeling of bearing degradation signals," *IEEE Trans Reliab*, vol. 66, no. 4, pp. 1331–1344, Dec. 2017, doi: 10.1109/TR.2017.2739126.

[7] S. E. Pandarakone, Y. Mizuno, and H. Nakamura, "A comparative study between machine learning algorithm and artificial intelligence neural network in detecting minor bearing fault of induction motors," *Energies (Basel)*, vol. 12, no. 11, 2019, doi: 10.3390/en12112105.

[8] K. H. Hui, C. S. Ooi, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "An improved wrapper-based feature selection method for machinery fault diagnosis," *PLoS One*, vol. 12, no. 12, Dec. 2017, doi: 10.1371/journal.pone.0189143.

[9] M. Peña *et al.*, "Feature engineering based on ANOVA, cluster validity assessment and KNN for fault diagnosis in bearings," in *Journal of Intelligent and Fuzzy Systems*, IOS Press, 2018, pp. 3451–3462. doi: 10.3233/JIFS-169525.

[10] P. Shunmugapriya and S. Kanmani, "A hybrid algorithm using ant and bee colony optimization for feature selection and classification (AC-ABC Hybrid)," *Swarm Evol Comput*, vol. 36, pp. 27–36, Oct. 2017, doi: 10.1016/j.swevo.2017.04.002.

[11] A. Sharma, M. Amarnath, and P. K. Kankar, "Feature extraction and fault severity classification in ball bearings," *JVC/Journal of Vibration and Control*, vol. 22, no. 1, pp. 176–192, Jan. 2016, doi: 10.1177/1077546314528021.

[12] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, Nov. 2018, doi: 10.1016/j.neucom.2018.05.002.

[13] M. Imane, C. Rahmoune, and D. Benazzouz, "Rolling bearing fault feature selection based on standard deviation and random forest classifier using vibration signals," *Advances in Mechanical Engineering*, vol. 15, no. 4, Apr. 2023, doi: 10.1177/16878132231168503.

[14] M. Van, D. T. Hoang, and H. J. Kang, "Bearing fault diagnosis using a particle swarm optimization-least squares wavelet support vector machine classifier," *Sensors (Switzerland)*, vol. 20, no. 12, pp. 1–19, Jun. 2020, doi: 10.3390/s20123422.

[15] Z. Xue, Y. Huang, W. Zhang, J. Shi, and H. Luo, "Intelligent Fault Diagnosis of Rolling Bearings Based on a Complete Frequency Range Feature Extraction and Combined Feature Selection Methodology," *Sensors (Basel)*, vol. 23, no. 21, Oct. 2023, doi: 10.3390/s23218767.

[16] X. Liang, J. Yao, W. Zhang, and Y. Wang, "A Novel Fault Diagnosis of a Rolling Bearing Method Based on Variational Mode Decomposition and an Artificial Neural Network," *Applied Sciences (Switzerland)*, vol. 13, no. 6, Mar. 2023, doi: 10.3390/app13063413.

[17] Y. Keshun, Q. Guangqi, and G. Yingkui, "Optimizing prior distribution parameters for probabilistic prediction of remaining useful life using deep learning," *Reliab Eng Syst Saf*, vol. 242, p. 109793, 2024.

[18] J. Zhou, K. Zhou, G. Zhang, F. Neri, W. Shen, and W. Jin, "A data-driven optimisation method for a class of problems with redundant variables and indefinite objective functions," *Inf Sci (N Y)*, vol. 656, p. 119899, 2024.

[19] M. Unal, M. Onat, M. Demetgul, and H. Kucuk, "Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network," *Measurement (Lond)*, vol. 58, pp. 187–196, Dec. 2014, doi: 10.1016/j.measurement.2014.08.041.

[20] D. Kolar, D. Lisjak, M. Pajak?, and M. Gudlin, "Intelligent fault diagnosis of rotary machinery by convolutional neural network with automatic hyper-parameters tuning using bayesian optimization," *Sensors*, vol. 21, no. 7, Apr. 2021, doi: 10.3390/s21072411.

[21] H. Tian, H. Fan, M. Feng, R. Cao, and D. Li, "Fault Diagnosis of Rolling Bearing Based on HPSO Algorithm Optimized CNN-LSTM Neural Network," *Sensors*, vol. 23, no. 14, Jul. 2023, doi: 10.3390/s23146508.

[22] Z. Chen *et al.*, "Research on bearing fault diagnosis based on improved genetic algorithm and BP neural network," *Sci Rep*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-66318-0.

[23] B. D. Center, "Case western reserve university bearing data," 2013.

[24] W. Shen, M. Xiao, Z. Wang, and X. Song, "Rolling Bearing Fault Diagnosis Based on Support Vector Machine Optimized by Improved Grey Wolf Algorithm," *Sensors*, vol. 23, no. 14, Jul. 2023, doi: 10.3390/s23146645.

[25] V. K. Verma, K. Saxena, and U. Banodha, "Analysis Effect of K Values Used in K Fold Cross Validation for Enhancing Performance of Machine Learning Model with Decision Tree," in *International Advanced Computing Conference*, Springer, 2023, pp. 374–396.

[26] A. Alhams, A. Abdelhadi, Y. Badri, S. Sassi, and J. Renno, "Enhanced Bearing Fault Diagnosis Through Trees Ensemble Method and Feature Importance Analysis," *Journal of Vibration Engineering and Technologies*, Dec. 2024, doi: 10.1007/s42417-024-01405-0.

[27] H. Al Bugharbee, H. Samaka, and S. L. Zubaidi, "Diagnosis of air compressor condition using minimum redunduncy maximum relevance (mrmr) algorithim and distance metric based classification," *Diagnostyka*, vol. 22, no. 4, pp. 25–32, 2021, doi: 10.29354/diag/143762.

[28] U. Parmar and D. H. Pandya, "Experimental investigation of cylindrical bearing fault diagnosis with SVM," in *Materials Today: Proceedings*, Elsevier Ltd, 2021, pp. 1286–1290. doi: 10.1016/j.matpr.2020.11.327.

[29] N. Ketkar, J. Moolayil, N. Ketkar, and J. Moolayil, "Convolutional neural networks," *Deep learning with Python: learn best practices of deep learning models with PyTorch*, pp. 197–242, 2021.

[30] X. Wang, Y. Liu, and H. Xin, "Bond strength prediction of concrete-encased steel structures using hybrid machine learning method," *Structures*, vol. 32, pp. 2279–2292, Aug. 2021, doi: 10.1016/j.istruc.2021.04.018.

[31] D. Kolar, D. Lisjak, M. Pajak?, and M. Gudlin, "Intelligent fault diagnosis of rotary machinery by convolutional neural network with automatic hyper-parameters tuning using bayesian optimization," *Sensors*, vol. 21, no. 7, Apr. 2021, doi: 10.3390/s21072411.

[32] J. Kennedy, R. Eberhart, and bls gov, "Particle Swarm Optimization."

[33] J. F. Chen, Q. H. Do, and H. N. Hsieh, "Training artificial neural networks by a hybrid PSO-CS Algorithm," *Algorithms*, vol. 8, no. 2, pp. 292–308, 2015, doi: 10.3390/a8020292.

[34] M. Süpürtülü, A. Hatipoğlu, and E. Yılmaz, "An Analytical Benchmark of Feature Selection Techniques for Industrial Fault Classification Leveraging Time-Domain Features," *Applied Sciences (Switzerland)*, vol. 15, no. 3, Feb. 2025, doi: 10.3390/app15031457.

[35] F. Althobiani, "A Novel Framework for Robust Bearing Fault Diagnosis: Preprocessing, Model Selection, and Performance Evaluation," *IEEE Access*, vol. 12, pp. 59018–59036, 2024, doi: 10.1109/ACCESS.2024.3390234.

[36] N. F. Fadzail, S. M. Zali, E. C. Mid, and N. N. A. Bakar, "Hyperparameter Optimization using Automated Machine Learning based Artificial Neural Network for Fault Detection in Wind Turbine Generator," *Next Research*, p. 100785, 2025.

[37] J. Ben Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Applied Acoustics*, vol. 89, pp. 16–27, 2015, doi: 10.1016/j.apacoust.2014.08.016.