Faulty gear diagnosis using weighted PCA with swish activated BLSTM classifier

Rohit Ghulanavar1,2 · A. Jagadeesh1 · Kiran Kumar Dama1

Abstract
The early faulty gear diagnosis is most necessary in the industry. In the current decade, with the tremendous growth of ANN (Artificial Neural Network), the researcher planned to use DL (Deep Learning) methods to sketch out faults in gear in an early stage. Traditional gear fault diagnosis method mostly utilizes deep NN (Neural Network) related to time sequence of gathered signals. In this instance, feature extraction in the direction of inverse time domain signal is commonly ignored. To overcome this issue, here in this paper, proposed Weighted Principal Component Analysis (WPCA) and BLSTM (Bi-Directional Long Short Term Memory) along with Swish Activation function for faulty gear diagnosis from the vibration signals. WPCA is utilized to extract multi-scale features related to faulty gear from the vibration signal. Likewise, BLSTM is used to classify the extracted features to diagnose the fault in an earlier stage. Several experiments were conducted to evaluate the proposed work of categorizing the defects in gear from the vibrating signal. Experiments were conducted on three kinds of the dataset to classify the type of faulty gear accurately. The proposed work proves its superiority in organizing the gear faults in a most efficient way than existing methods.

Keywords Deep learning · BLSTM · WPCA · ANN · NN · Gear · Vibration signal

1 Introduction

Gearboxes are broadly utilized in rotary Machines like artifacts engines; industrial gearboxes, power station, etc. are the essential elements of the different mechanical system. However, gearbox is subjected to tedious and harsh working conditions, which create spare parts like...
gear, prone to failure [18]. Transverse vibration is developed because of fatigue, in several rotor elements of dynamic rotor system which causes various catastrophic failure and damages of machinery. Vibration signals are produced due to defects in machine components like a pulley, belt, looseness, rotor unbalance, cracks, misalignment, shaft bow, coupling and rotor hub defects. Gear failure affects mechanical equipment’s regular operation and in some cases, damage the entire machine, which leads to threatening people’s property and life. Therefore, it is necessary to extract and separate faulty gear’s faulty feature to manage the working efficiency of mechanical equipment to safeguard people’s property and life. Related survey express that localizing gear fault produces a periodic transient pulse in standard speed mode [22]. Because of these characteristics, vibration signal analysis technique is most effective and reasonable. Apart from the transient component with larger amplitude, gear fault signal consists of harmonic components produced by gear pair’s meshing vibration [29]. Meanwhile, two segments were sub-merged quickly with the use of intense noise [30, 31] creating mining of transient component was very tedious, so accurate extraction of faults from faulty gear is key to research.

Figure 1a expresses that, to investigates the relationship among system and vibration mode of the gearbox using multi-body dynamic simulation.

There are several efficient methods like EMD [25], spectral kurtosis [16, 28], wavelet transform [30, 31], stochastic resonance [13], DL [5], sparse representation [7, 8] and time-frequency analysis [34]. However, vibration signals are computed from faulty gear which includes fault transient, harmonic component and interference noise. Current methods can’t accurately predict fault from the vibration of fault gear.

To overcome the aforementioned issues, the researcher decides to use weighted principal components Analysis (WPCA) to extract multi-scale features regarding fault in gear. For classification BLSTM (Bidirectional Long Short Term Memory) along with swish activation function is used. BLSTM supports in the classification of responsibility in the gearbox.

Figure 1b illustrates that the vibration experiment was carried out to evaluate modal optimization, decreasing gear vibration.

1.1 Objective of paper

- To classify faulty gear from vibration signal of the gearbox

![Fig. 1](image_url) a Captures faulty vibration Signal. b Captures Faulty Gear Signal
To extract multi-scale features with the use of weighted principal component analysis (WPCA).
To classify faults of gear, BLSTM method along with Swish activation function is utilized.

1.2 Organization of Paper

The paper’s remaining section is organized as Section 2 discusses the related works regarding WPC, BLSTM and Swish activation function. Section 3 discuss the proposed work to extract and classify faults occurred in gear. Section 4 discusses the performance analysis of the proposed work, and its conclusion is expressed in Section 5.

2 Related works

Karim et al. [12] proposed LSTM-FCN (Long Short Term Memory-Fully Convolutional Neural Network) with multivariate time series classification model with augmenting convolutional blocks and excitation and squeeze block to enhance the accuracy. The proposed work outperforms existing models in the way of reduced preprocessing steps. Proposed architecture more efficiently on different tedious multivariate time series categorization tasks like action recognition. Additionally, proposed work is sufficient at test time and small in size to be easily implemented in the restricted memory system.

Greff et al. [6] introduce 1st large scale analysis of 8 LSTM variant over three representative tasks like handwriting recognition, speech recognition and polyphonic music designing. The LSTM variant’s hyperparameter for every job was optimized by utilizing random search, and their significance was assessed using the powerful fA NOVA framework. The result shows that the proposed work shows superior output than existing work. Additionally, individual usage of hyperparameter offers a guideline for effective adjustment.

Karim et al. [11] proposed augmented CNN (Convolutional Neural Network) with LSTM-RNN (Recurrent Neural Network) to classify time series. Proposed work prominently improves CNN performance with minor improvement in model size and requests for reduced processing of dataset. The proposed LSSTM-FCN (Fully Convolutional Network) attains better understanding compared to others. Additionally, attention mechanism is utilized to enhance time series categorization with Attention (LSTM-FCN). Then, refinement is proposed as a method to better the trained model’s performance. The overall performance of the proposed work is evaluated and compared with other existing techniques.

Akram et al. [2] presents proposed EMD (Empirical Mode Decomposition), and traditional PSD (Power Spectrum Density) and time waveform to predict localized spur gear. The test model is primarily created for analyzing vibration of gear at various RPM, and then a particular fault was presented in driven gear for different damage conditions. Data recorded by a wireless tri-axial accelerometer was analyzed utilizing PSD and EMD method and result express that EMD is better than traditional time waveform and PSD method.

Mishra et al. [19] introduce the novel fault diagnosis technique that integrates three methods like EMD, PSO-SVM (Particle Swarm Optimization- Support Vector Machine) and fractal box dimension. Primarily, fault gear’s vibration signal is decomposed to various IMF (Intrinsic Mode Function) by EMD technique. Then, energy, frequency box dimension and time are computed from frequency, time, fractal, and energy domain. Then characteristic of gear fault
under various load excitations was acquired. At last, mined features are given to the PSO-SVM model for classification of gear fault. Result express that proposed work discovers the kind of gear fault effectively under various load excitation.

Huang et al. [9] proposed to utilize minimal concave penalty function to develop the objective function. Among other process proposed function is superior in protecting high amplitude components. The entire dictionary was then launched depending on Fourier bases, where every two elements of every dictionary from the tight frame facilitate easy calculation and eliminate calculating the higher dimensional inverse matrix while doing repeated optimization. By constructing numerous dictionary, numerous components related to gear vibration signal was mined. Sparse depiction coefficient was computed utilizing split augmented Lagrangian shrinkage algorithm and completed dictionary and depicted transient could be mined. The experimental result expresses that the harmonic component and fault transient was extracted accurately without underrating higher amplitude components. Comparison result illustrates that proposed work outperforms existing work.

Li et al. [14] presents novel feature extraction technique by integrating EMD and ALC (Autocorrelation Local Cepstrum) to diagnose fault multistage gearbox. Primarily, in the preprocessing step, signal reconstruction was implemented to address the oversampled problem caused by higher resolution of angular sensor and test speed. Then adaptive EMD was utilized to obtain numerous IMF (Intrinsic Mode Function). Various IMF had various sensitivity to a fault. So all IMF is not used for further analysis. For this purpose, the cosine similarity metric is utilized to choose the most sensitive IMF. Only IMF is not sufficient for fault diagnosis, so ALC is utilized for feature extraction and signal denoising. Robustness of proposed work was evaluated experimentally using 2 sets of gear test rigs and gear in various working conditions. Result proves that proposed work was efficient in the diagnosis of fault in gear.

Li et al. [15] for diagnosing a fault in gear, VMD (Variation Model Decomposition) and DNN (Deep Neural Network) was proposed. Three-axial vibration signals of gear were gathered and decomposed to narrowband components with various frequency centre sand bandwidth depends on VMD. PSE (Power Spectral Entropy) is utilized as an original feature to denote distribution and the spectral amplitude of every component. DNN depending on AE (Automatic Encoder) and backpropagation NN is being used for signal feature reduction and gear states categorization. Results express that proposed work had the ability of mining sensitive features and diagnosing of fault.

Athiwaratkun and Stokes [3] proposed various novel malware categorization architecture which consists of LSTM and GRU (Gated Recurrent Unit) language models. Additionally, attention mechanism is presented. At last, introduced single-stage malware classifier depending on character level CNN. The experimental result expresses that LSTM with logistic regression, and max-pooling layer offers better enhancement than the current system.

Yildirim [35] proposed (Deep Bidirectional) DBLSTM-WS (Wavelet Sequence) for categorizing ECG signals. For this reason, this wavelet depended layer was employed to produce a signal sequence for ECG. This ECG signal was decomposed to sub-bands at the various scale of the coating. These sub-bands were utilized as input for the LSTM network. Here for comparison, unidirectional and bidirectional LSTM were used. Experiments were conducted on five kinds of heartbeats from datasets, its noticed that proposed work outperforms existing work in terms of improving the performance of recognition of CN.

Jinsakul et al. [10] the aim was to offer experimental alteration of Deep Learning (DL) of exception Swish and review the chance to create a primary colorectal polyp screening system
by training the proposed work with colorectal program dataset. Result express that submitted work achieves better accuracy of prediction compared to the traditional model.

Cerrada et al. [4] deals with feature selection issues over attribute clustering. The proposed algorithm was encouraged by current methods, where relative dependency among attributes was utilized to compute dissimilarities values. The centroid of the developed cluster was chosen as representative attributes. Selection algorithm utilize the random process for offering centroid candidates, in a way, random search in inherent exploration was included. For this algorithm, the hierarchical procedure was proposed. In every level of the hierarchy, the whole set attributes were divided into disjoint sets and selection process is utilized with every subset. Once a prominent characteristic was proposed for every subset, a new set of features is developed, and once again selection process continues in the next stage. Hierarchical execution aims to refine every level of search space in minimized set of chosen attributes. Result express better diagnosis precision than current methods.

Park et al. [20] classification of gear teeth is done by utilizing Ensemble (EEMD) to TE (Transmission Error Rate) computed by encoders of output and input shafts. Gear with two faults was modelled, and TE was acquired by simulating faulty gear overloaded contact to discover various characteristics. Testbed for evaluating proposed work is created, where TE was measured for gear. EEMD was utilized to mine false feature of gear over noise from computed TE. The experimental result shows that proposed work gives a better measurement of gear fault directly with least noise, assures successful diagnosis.

Sun et al. [26] to mine transient features from vibration signal of gear, fault diagnosis technique depending on SSTFA (Sparsity Time-Frequency Analysis) was proposed. Impulsive and steady modulation components of faulty gear vibration signal were mined parallel by selecting a various time-frequency neighbourhood and standard thresholding operator. The diagnostic conclusion was made regarding the envelope spectrum of impulsive components or by the periodicity of impulse. The proposed work’s method was evaluated to know its efficiency; proposed work express it’s more superior than existing work.

Tripathi et al. [27] proposed DNN digital pre-distortion (DPD) utilize swish or sigmoid weight linear unit activation function instead of ReLU and sigmoid to eliminate dead neuron and gradient vanishing issues. A comparison was made for various activation functions for real value concentrated time delay NN had been done. Proposed work attains better enhancement than existing work.

Xie et al. [33] proposed stacking method based on WPCA. The principal component of data matrix-like sound signals was mined depending on the lower rank-decomposition method by resolving optimization issues with lower rank limitations. Optimization issue is determined through the standard singular value decomposition algorithm. Data matrix’s lower rank-decomposition would improve the impact of non-Gaussian random noise, erratic and abnormal trace, and it’s stronger than current alternatives. Here the performance of proposed work is evaluated on-field and synthetic data, and it proves its success in terms of its performance.

Qi et al. [21] investigates correlation among QoS criteria of multimedia services and also analyze criteria correlation’s the bad effect for service selection for multimedia in the cloud. Then, W-PCA_MSSM () was proposed to remove criterion correlation and simplify a piece of the service process. Several experiments were conducted to estimate the possibility of proposed work in terms of efficiency and effectiveness, proving its success. In future, this proposed work several applications like service replacement and composition.

Shao et al. [23] developed a new DL framework to diagnose highly-accurate machine faults by utilizing transfer learning to accelerate and enable in-depth NN training. Compared to
current techniques, the proposed work was more accurate and much faster to train. Primarily, original sensor information was altered to images by directing wavelet transformation to acquire time-frequency distribution. Then low-level features were extracted by the pre-trained network. NN’s higher level is fine-tuned by labelled time-frequency idea. Experiments were carried out to evaluate the performance of proposed work, and it shows that it works is superior to all other work.

Wen et al. [32] proposed TCNN (ResNet-50) with 51 convolutional layers for diagnosing the fault. By merging with transfer learning, proposed work use ResNet-50 on ImageNet to extract features for diagnosis. Primarily, method signal-image is created to transfer time-domain fault signal to the image format of RGB as input data kind of ResNet-50. Experiments were conducted on three datasets, and results show it a success for accurate prediction.

3 Proposed work

BLSTM-WPCA based fault gear diagnosis method is proposed to classify the kind of fault in gear with support vibration signal acquired from machine or gearbox (Fig. 2).

3.1 Weighted principal component analysis (WPCA)

Since PCA doesn’t include time-varying process, a weighted criterion in the dimension minimization process is created to mine feature information. WPCA merges SFA (Slow Feature Analysis) and PCA (Principal Component Analysis) which includes time-varying features and a slow varying feature hidden in the process. WPCA can be done by resolving

![Fig. 2 BLSTM-WPCA Based Faulty Gear Diagnosis](image-url)
constrained minimization issues with SFA and PCA’s joint objective function; the objective function is distinct as,

\[ I_1 = \max \ o^D Y Y^D o \]  

Where \( o \in \mathbb{R}^{t \times t} \) is a transformation matrix. And \( Y \in \mathbb{R}^{t \times t} \) is a matrix of high dimension data. Unlike SFA and PCA reduce the variation of hidden features. Optimization equation is given below:

\[ \min \Delta x_i (d) = \min (x_i^2) d \]  

Where \( x_i \) is i-th 1st order derivate of \( x \) and \( x \) is a t-dimensional slow feature. (\( \cdot \)) this sample means of whole available time. X satisfy subsequent constraints:

\[ \langle x_i \rangle_d = o \]
\[ \langle x_i^2 \rangle_d = 1 \]
\[ \langle x_j x_i \rangle_d = o \]

Where \( \Delta Y = [\Delta Y_1 \Delta Y_2 \ldots \Delta Y_t] \) and \( \Delta Y (d) = Y(d)-Y(d-1) \). Objective function \( I_1 \) and \( I_2 \) have the same structure. By merging two objective functions, SFA and PCA’s following can be taken simultaneously. After merging the function, WPCA can be done by computing \( o \) satisfying subsequent objective function:

\[ I = \max \alpha o^D Y Y^D o - (1-\alpha) o^D \Delta Y \Delta Y^D o \]
\[ = \max o^D (\alpha Y Y^D - (1-\alpha) \Delta Y \Delta Y^D) o \]
\[ \text{s.t.} o^D o = 1, \]
\[ = \max o^D Y' o \]  

Where \( Y' = \alpha Y Y^D - (1-\alpha) (\Delta Y \Delta Y^D) \), and \( \alpha \) is weighting indicate \( \alpha \leq 1 \). PCA’s relative influence grThe optimal detectability of WPCA, \( \alpha \) should be selected correctly because it indicates SFA and PCA’s relative influence. Optimal solution of Eq. (3) corresponds to eigenvalue and optimal transformation vector as below:

\[ Y' Q = \lambda Q \]  

\( Q = [q_1 q_2 \ldots q_t] \) is optimal transformation vectors, also referred to as the loading vector. WPCA aims to mine lower dimensional linear features; the research uses few PCs that brief their variance. Then, the loading vector becomes \( Q_{\text{new}} = [q_1 q_2 \ldots q_t] \) where \( p << t \). Now PC can be computed as

\[ d = Q_{\text{new}}^D Y. \]  

Like, PCA, hotellings\( D^2 \) and SPE (Square Prediction Error) are utilized to monitor the WPCA depended on process monitoring method. \( D^2 \) denotes distance among current variable’s value and its PC subspace’s average, taking the variance of data as below:

\[ D^2 = \sum_{j=1}^{p} \frac{d_j^2}{\sigma_j^2} \]  

Where \( DJ \) is the score on j – the PC and \( \sigma_j \) is the standard deviation of scores on j – the PC. Score on j – the PC is acquired from WPCA. SPE (Square Prediction Error) denotes error among WPCA and process data in residual space below:
\[
\text{SPE} = (y - Q_{\text{new}}d)^D (y - Q_{\text{new}}d)
\]  

(7)

Provided process data has non-Gaussian distribution, then utilizing KDE (Kernel Density Estimation) SPE and control limit of \(D^2\) is computed. When SPE or \(D^2\) exceeds control limits, it’s taken as an alarm. KDE is done as

\[
e(y) = \frac{1}{mg} \sum_{j=1}^{m} K \frac{y - y_j}{g}
\]  

(8)

where \(e\) denotes density function, \(m\) represents the count of the sample, \(g\) refers to smoothing parameter, \(y_j\) is noticed value and \(K\) denotes kernel function. \(D^2\) is summation denotes error and PC variation as a signal statistic, if the value of PC goes above the control limit, then detection of \(D^2\) is not easy. Based on KDE, multi-scale feature is predicted.

**Convolutional layer** Convolutional layer performs the process of feature extraction from the input feature map over the convolution kernel. Downsampling is performed with the support of pooling layer. One of NN (Neural Network) ‘s main features is weight sharing, local perception, and pooling. In our architecture, the primary layer is convolution layer the formula for computation is as follows:

\[
y^l_j = e \left( \sum_{j=1}^{N} y^{l-1}_j * k^l_{ji} + a^l_i \right)
\]  

(9)

Where, \(y^l_j\) is \(i\) – the feature map of \(l\) – th layer \(e(\cdot)\) denotes activation function, \(N\) denotes the count of the input feature map, \(y^{l-1}_j\) denotes \(j\) – the feature map of \(l\)-1 layer, \(*\) denotes convolution operation, \(k^l_{ji}\) denotes trainable convolution kernel, and \(a^l_i\) denotes \(i\) – the bias of \(l\) – the layer.

**3.2 Pooling layer**

Then pooling layer is linked to the convolution layer, where the feature map was downsampled related to few pooling plans to acquire low-resolution feature map. The most frequently utilized pooling plan is max pooling. This pooling layer minimizes the count of output nodes and improves the network’s robustness to input characteristics. \(L + 1\) – the layer is pooling layer. It’s computed as

\[
y^{l+1}_j = e \left[ \text{down}(y^l_j) + y^{l+1}_j \right]
\]  

(10)

Where \(\cdot\) denotes downsampling function.

By considering overfitting issues and convergence speed, here the researcher decides to use Swish activation function.

**3.3 Swish activation function**

Most broadly and widely used to activation function in NN is ReLU, which is \(e(y) = \max(0,y)\). Though there are several alternatives to ReLU has been emerged, but none can be
able to replace it. To overcome this, Google Brain Team had introduced new activation function named swish activation function. Which is \( e(y) = y \cdot \text{sigmoid}(y) \).

\[
\text{swish}(y) = y \cdot \left(1 + e^{-y}\right)^{-1}
\]

(11)

This function is similar to ReLU; it is different; domain around 0 differs from ReLU. It is a smooth function that doesn’t alter the direction frequently as ReLU does nearer \( y = 0 \). Instead, it bends smoothly from 0 and move towards value <0 and then again move upwards (Fig. 3).

This observation indicates that it’s non-monotonic. Similarly, it doesn’t move in a single direction like ReLU. These properties make it more superior than other activation function. Experiments express that it works better than ReLU for the various challenging dataset. Swish is non-monotonic, smooth operation achieves better test accuracy, which is steadily superior to ReLU on NN, which supports most challenging domains like machine translation and image classification.

### 3.4 BLSTM unit

BLSTM network improves convolutional neurons from RNN (Recurrent Neural Network) with a series of the gate, enforces standard error flow over internal network state. Current sequence value \( y_d \) is considered input of memory cell, whereas previous internal state is denoted as \( E_{d-1} \) and the earlier detection is denoted as \( g_{d-1} \). Then current prediction \( g_d \) is computed using input gate \( \delta_d \), forget gate \( o_d \) and output gate \( i_d \), while internal state \( E_d \) is updated utilizing input and forget gate. More than one memory cell is merged in the form of series to form a single BLSTM-NN layer.

![Swish Activation Function](Fig. 3)
BLSTM network is utilized for forecasting time series by using sequence to sequence regression. At the same time, responses \( x_d \) are input time series shifted by one-time step, \( y_d = y_{d+1} \).

For the gearbox, the vibration signal is considered as input time series \( y_d \). BLSTM is trained on a dataset acquired from the gearbox, which is in a normal state. If the gearbox is in the normal state then determination would match the response well that is \( g_d \approx x_d \), and prediction error \( r_d = x_d - g_d \), would be small. In case when a fault occurs in the gearbox, then response time series consist of health vibration \( x_d \) and \( a_d \) as defect induced vibration. Fault state response is denoted as \( X_d \), then \( X_d = x_d + a_d \). If \( X_d \) is fed to trained network, then since \( g_d \approx x_d \), prediction error would approximate bearing fault signal, \( r_d = X_d - g_d \approx x_d \). Here, BLSTM network act as a filter to expose anomalies.

### 3.5 Softmax layer

The final layer of the network is a softmax layer which non-linear. This layer supports in handling numerous classes. The softmax function can squeeze output of every level among 0 and 1 and divides by sum of results. In short, this layer determines the probability of output. Mostly softmax function is utilized in classifier’s output layer.

### 4 Performance analysis

Here for evaluating the performance of proposed work, three kinds of the dataset were utilized. 1st dataset, a random number were generated to pick data set regarding Local fault with noise and distribution fault [17]. 2nd dataset is gearbox fault diagnosis dataset which includes dataset regarding vibration recorded by utilizing spectraquest gearbox fault diagnostics simulator. This dataset was created by recording four vibration sensors positioned in 4 various direction. Records had been done with loads varying from 0% to 90%. Here for our purpose dataset related to Broken tooth condition is taken [24]. 3rd dataset is GFDS2 which group data related to faulty gear diagnosis from the sound signal. Here load normal, load pitting, load tooth fracture, load wear, normal, pitting, tooth fracture and wear datasets were utilized.

Figure 4 express various faulty signals of gear, these signal is utilized for training classifiers. So that faults occurred in gear can be analyzed easily.

Table 1 compares the proposed method with and without multi-scale features and with VGG16 and ResNet-59 methods. The graph illustrates that the proposed method with multi-scale feature outperforms all other ways.

Figure 5 illustrates the proposed method’s performance analysis in terms of accuracy, sensitivity, specificity, precision, recall, and F-measures for various testing sets.

Learning rate is hyperparameter which controls weight adjustment of a network concerning loss gradient. Table 2 expresses the proposed method’s performance analysis in terms of accuracy, sensitivity, specificity, precision, recall, and F-measures for various cross-validation set of learning rate.

### 5 Conclusion

Fault diagnosis plays a crucial role in maintaining equipment. This paper presents the gear fault diagnosis method depending on multi-layer BLSTM models. Additionally, for feature
Fig. 4 Various faulty signals of gear
extraction, the researcher introduces WPCA to extract multi-scale features related to gear faults from the vibration signal. Here, BLSTM is used along with Swish activation function to

### Table 1  Comparison of proposed work with and without multi-scale feature [23, 32]

| Learning Rate | 0.1   | 0.01  | 0.001 | 0.0001 |
|---------------|-------|-------|-------|--------|
| Accuracy      | 98.4211 | 95.2381 | 96.9737 | 96.1842 |
| Sensitivity   | 88.8889 | 97.8349 | 90.9091 | 95.2381 |
| Specificity   | 98.7119 | 55.5556 | 97.1545 | 96.2111 |
| Precision     | 67.7966 | 95.2381 | 48.7805 | 41.6667 |
| Recall        | 88.8889 | 0.7018  | 90.9091 | 95.2381 |
| F-Measure     | 0.7692  | 0.9216  | 0.6349  | 0.5797  |

![Performance Analysis of Proposed Work on the Cross-validation set](image)

**Fig. 5** Performance Analysis of Proposed Work on the Cross-validation set

### Table 2  Performance Analysis of Proposed Work on Learning Rate

| Learning Rate | 0.1   | 97.7632 | 0.001  | 0.0001 |
|---------------|-------|---------|--------|--------|
| Accuracy      | 98.4211 | 95.2381 | 96.9737 | 96.1842 |
| Sensitivity   | 88.8889 | 97.8349 | 90.9091 | 95.2381 |
| Specificity   | 98.7119 | 55.5556 | 97.1545 | 96.2111 |
| Precision     | 67.7966 | 95.2381 | 48.7805 | 41.6667 |
| Recall        | 88.8889 | 0.7018  | 90.9091 | 95.2381 |
| F-Measure     | 0.7692  | 0.9216  | 0.6349  | 0.5797  |
classify the extracted feature and diagnose or predict the kind of responsibility in gear. The performance of the proposed work is evaluated to know its efficiency with challenging datasets. Proposed work proved its superiority compared to existing techniques.

**Declarations**

**Conflicts of interest/competing interests**
No financial/personal interest.

**References**

1. Abood OG, Elsadd MA, Guirguis SK (2017) Investigation of cryptography algorithms used for security and privacy protection in smart grid. In 2017 Nineteenth International Middle East Power Systems Conference (MEPCON) (pp. 644–649). IEEE. https://doi.org/10.1109/MEPCON.2017.8301249
2. Akram MA et al (2019) Vibration based gear fault diagnosis under empirical mode decomposition and power Spectrum density analysis. Advances in Science and Technology Research Journal 13:3
3. Athiwaratkun B, Stokes JW (2017) Malware classification with LSTM and GRU language models and a character-level CNN. In 2017 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 2482–2486). IEEE. https://doi.org/10.1109/ICASSP.2017.7952603
4. Cerrada M, Sánchez RV, Pacheco F, Cabrera D, Zurita G, Li C (2016) Hierarchical feature selection based on relative dependency for gear fault diagnosis. Appl Intell 44(3):687–703
5. Ding X, He Q (2017) Energy-fluctuated multiscale feature learning with deep convekt for intelligent spindle bearing fault diagnosis. IEEE Trans Instrum Meas 66(8):1926–1935
6. Greff K, Srivastava RK, Koutnik J, Steunebrink BR, Schmidhuber J (2016) LSTM: a search space odyssey. IEEE transactions on neural networks and learning systems 28(10):2222–2232
7. Han T, Jiang D, Sun Y, Wang N, Yang Y (2018) Intelligent fault diagnosis method for rotating machinery via dictionary learning and sparse representation-based classification. Measurement 118:181–193
8. He G, Ding K, Lin H (2016) Fault feature extraction of rolling element bearings using sparse representation. J Sound Vib 366:514–527
9. Huang W, Li S, Fu X, Zhang C, Shi J, Zhu Z (2020) Transient extraction based on minimax concave regularized sparse representation for gear fault diagnosis. Measurement 151:107273
10. Jinsakul N, Tsai CF, Tsai CE, Wu P (2019) Enhancement of deep learning in image classification performance using Xception with the swish activation function for colorectal polyp preliminary screening. Mathematics 7(12):1170
11. Karim F et al (2017) LSTM fully convolutional networks for time series classification. IEEE Access 6:1662–1669
12. Karim F, Majumdar S, Darabi H, Harford S (2019) Multivariate lstm-fcns for time series classification. Neural Netw 116:237–245
13. Lei Y, Han D, Lin J, He Z (2013) Planetary gearbox fault diagnosis using an adaptive stochastic resonance method. Mech Syst Signal Process 38(1):113–124
14. Li B, Zhang X, Wu J (2017) New procedure for gear fault detection and diagnosis using instantaneous angular speed. Mech Syst Signal Process 85:415–428
15. Li Y, Cheng G, Liu C, Chen X (2018) Study on planetary gear fault diagnosis based on variational mode decomposition and deep neural networks. Measurement 130:94–104
16. Liu H, Huang W, Wang S, Zhu Z (2014) Adaptive spectral kurtosis filtering based on Morlet wavelet and its application for signal transients detection. Signal Process 96:118–124
17. López-Nicolás C, Meroño-Cerdán AL (2011) Strategic knowledge management, innovation and performance. Int J Inf Manag 31(6):502–509
18. Lu S, He Q, Wang J (2019) A review of stochastic resonance in rotating machine fault detection. Mech Syst Signal Process 116:230–260
19. Mishra S et al (2019) A comparative analysis of centralized and distributed spectrum sharing techniques in cognitive radio. Springer, Computational Intelligence in Sensor Networks, pp 455–472
20. Park S, Kim S, Choi JH (2018) Gear fault diagnosis using transmission error and ensemble empirical mode decomposition. Mech Syst Signal Process 108:262–275
21. Qi L, Dou W, Chen J (2016) Weighted principal component analysis-based service selection method for multimedia services in cloud. Computing 98(1–2):195–214
22. Qin Y, Xing J, Mao Y (2016) Weak transient fault feature extraction based on an optimized Morlet wavelet and kurtosis. Meas Sci Technol 27(8):085003
23. Shao S, McAleer S, Yan R, Baldi P (2018) Highly accurate machine fault diagnosis using deep transfer learning. IEEE Transactions on Industrial Informatics 15(4):2446–2455
24. Shi C, Li Y, Zhang J, Sun Y, Yu PS (2016) A survey of heterogeneous information network analysis. IEEE Trans Knowl Data Eng 29(1):17–37
25. Singh J, Darpe AK, Singh SP (2017) Bearing damage assessment using Jensen-Rényi divergence based on EEMD. Mech Syst Signal Process 87:307–339
26. Sun R, Yang Z, Chen X, Tian S, Xie Y (2018) Gear fault diagnosis based on the structured sparsity time-frequency analysis. Mech Syst Signal Process 102:346–363
27. Tripathi GC, Rawat M, Rawat K (2019) Swish activation based deep neural network predistorter for RF-PA. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) (pp. 1239–1242). IEEE. https://doi.org/10.1109/TENCON.2019.8929500
28. Wang T, Chu F, Han Q, Kong Y (2017) Compound faults detection in gearbox via meshing resonance and spectral kurtosis methods. J Sound Vib 392:367–381
29. Wang L, Cai G, Wang J, Jiang X, Zhu Z (2018) Dual-enhanced sparse decomposition for wind turbine gearbox fault diagnosis. IEEE Trans Instrum Meas 68(2):450–461
30. Wang D, Tsui KL, Qin Y (2019a) Optimization of segmentation fragments in empirical wavelet transform and its applications to extracting industrial bearing fault features. Measurement 133:328–340
31. Wang Z, du W, Wang J, Zhou J, Han X, Zhang Z, Huang L (2019b) Research and application of improved adaptive MOMEDA fault diagnosis method. Measurement 140:63–75
32. Wen L, Li X, Gao L (2020) A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications 32:6111–6124. https://doi.org/10.1007/s00521-019-04097-w
33. Xie J, Chen W, Zhang D, Zu S, Chen Y (2017) Application of principal component analysis in weighted stacking of seismic data. IEEE Geosci Remote Sens Lett 14(8):1213–1217
34. Yang Y, Dong XJ, Peng ZK, Zhang WM, Meng G (2015) Vibration signal analysis using parameterized time–frequency method for features extraction of varying-speed rotary machinery. J Sound Vib 335:350–366
35. Yildirim Ö (2018) A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Comput Biol Med 96:189–202

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.