

Swarm-LSTM: Condition Monitoring of Gearbox Fault Diagnosis Based on Hybrid LSTM Deep Neural Network Optimized by Swarm Intelligence Algorithms

Gopi Krishna Durbhaka¹, Barani Selvaraj¹, Mamta Mittal², Tanzila Saba^{3,*}, Amjad Rehman³ and Lalit Mohan Goyal⁴

¹School of Electrical and Electronics, Sathyabama Institute of Science and Technology, Chennai, 600119, India

²Department of Computer Science and Engineering, G.B. Pant Government Engineering College, New Delhi, 110020, India

³Artificial Intelligence and Data Analytics (AIDA) Laboratory, CCIS, Prince Sultan University, Riyadh, 11586, Saudi Arabia

⁴Department of Computer Engineering, J.C. Bose University of Science and Technology, Faridabad, 121006, India

*Corresponding Author: Tanzila Saba. Email: drstanzila@gmail.com

Received: 27 July 2020; Accepted: 11 September 2020

Abstract: Nowadays, renewable energy has been emerging as the major source of energy and is driven by its aggressive expansion and falling costs. Most of the renewable energy sources involve turbines and their operation and maintenance are vital and a difficult task. Condition monitoring and fault diagnosis have seen remarkable and revolutionary up-gradation in approaches, practices and technology during the last decade. Turbines mostly do use a rotating type of machinery and analysis of those signals has been challenging to localize the defect. This paper proposes a new hybrid model wherein multiple swarm intelligence models have been evaluated to optimize the conventional Long Short-Term Memory (LSTM) model in classifying the faults from the vibration signals data acquired from the gearbox. This helps to analyze the performance and behavioral patterns of the system more effectively and efficiently which helps to suggest for replacement of the unit with higher precision. The results have demonstrated that the proposed hybrid modeling approach is effective in classifying the faults of the gearbox from the time series data and achieve higher diagnostic accuracy in comparison to the conventional LSTM methods.

Keywords: Gearbox; long short term memory; fault classification; swarm intelligence; optimization; condition monitoring

1 Introduction

Fault diagnosis has a significance in identifying the degrading parts of the rotating machinery and replacement of the same well before a total breakdown to reduce the downtime. Especially in wind turbines, gears, shafts, blades and rolling bearings play a vital role that is widely used in the transmission of power. Any failures within them would introduce unexpected and unwarranted breakdown time, expensive maintenance, loss in production and delayed distribution of power. Hence, it is necessary to identify and predict such faults during the Operation and Maintenance (O&M) at early stages itself and increase the production of power to prevent power disruptions and catastrophic failures.



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Condition monitoring facilitates collecting the health information of the machinery through different methods such as vibration, acoustic and thermal imaging analysis. Earlier, the methodology adopted was through the processing of signals to derive deeper insights into different spectra such as time and frequency. The transition of Artificial Intelligence (AI) has been extensively investigated in the rotating machinery devices with versatile machine learning and deep neural network models. AI-led to relate the spectral insights in identifying the defects, further seeking more insights to categorize, to forecast the Remaining Useful Life (RUL) of the machinery components, and to replace the required ones.

Few research studies state how the traditional Convolution Neural Network (CNN) model has effective in the fault diagnosis for the classification of faults based on vibration analysis, by learning the features acquired from rolling bearings [1] and gearbox [2,3] as well both in time and frequency spectra. The results have been compared with a few of the peer algorithms such as Support Vector Machine and Random Forest.

New emerging methods in condition monitoring arise, to improve the reliability of gearboxes apart from the regular signal processing and applying machine learning models to categorize the faults accordingly. In recent years, Deep Neural Networks such as CNN and LSTM have widely been used to classify the faults and as well predict the RUL of the machinery too. These neural networks, used for classification of faults were Deep-Belief Network, Deep-Boltzmann Machines, Restricted Boltzmann Machines and Auto-Encoders [4] Support Vector Regression [5] and Stacked Multilevel Denoising Auto-Encoders [6] validated on vibration signals; similarly, Deep Random Forest Fusion [7] applied on both vibration and acoustic signal characteristics simultaneously for the gearbox fault diagnosis. Torsional vibration signals [8] of gear tooth have been processed in phase domain to localize the defect and derive the type of fault.

Wavelet analysis integrated with CNN model [9] was used to transform the vibration signals into time-frequency spectral images, and later capturing the features from the images to classify the faults. Cepstrum analysis [10] of the processed vibration signals with Hilbert and wavelet transform helped together to derive the gear faults and cracks in the bearings. Discrete Wavelet Transform [11] on time series vibration signals and Continuous Wavelet Transform [12] performed on acoustic signals to extract the features.

This paper has been organized as follows: Section 2 states the recent related research works regarding gearbox fault classification. Section 3 details the proposed hybrid LSTM gearbox fault diagnosis method along with the optimization techniques utilized for the classification of faults. In Section 4, the evaluated results of the custom hybrid model have been tabulated with metrics and later discussed the comparison of the proposed hybrid method with conventional LSTM methods. Finally, the conclusions followed by the references have been presented.

2 Literature Review

Some of the recent existing methodologies proposed for gearbox fault classification have been reviewed in this section.

Chen et al. [13] stated three deep neural network models Deep Belief Networks, Deep Boltzmann Machines and Stacked Auto-Encoders to assess rolling bearing fault conditions. Multiple pre-processing schemes time domain, frequency domain, and time-frequency domain have been applied for feature extraction. A single dataset of 7 fault patterns has been considered to test the efficiency of deep learning models to derive the health condition of rotating mechanical machinery. The evaluated results have demonstrated with a reliable model accuracy that was relevant for the bearing diagnostics.

Merainani et al. [14] performed an in-depth comparison of Hilbert Empirical Wavelet Transform (HEWT) and Hilbert Huang Transform (HHT) on the gearbox vibration signals. HEWT, a self-adaptive time-frequency analysis was applied to the vibration signals to obtain the instantaneous amplitude matrices. The fault feature vectors were acquired on decomposition from the vibration signals and later

classified using Self-Organizing Map model to derive the state of gear condition such as healthy gear, tooth cracking, input shaft slant crack and tooth surface pitting.

Liu et al. [15] proposed the Stacked Auto-Encoders (SAE) model for the resolution of gearbox fault diagnosis. This model was useful in extracting the salient characteristics from the frequency-domain signals. To reduce the challenge of overfitting during the training process of SAE and improve the performance with a tiny dataset, dropout technique and Rectified Linear Unit (ReLU) activation function has been introduced herewith. The effectiveness of this proposed model approach was derived with two gearbox datasets. The capability of the proposed model performed superior in comparison to the existing raw SAE.

Malik et al. [16] described fault diagnosis one of the widely used technique to identify the faults of gearbox, that would support in minimizing the operation cost and as well improve the reliability and feasibility of the wind turbine gearbox. Here the observed vibration signals of the gearbox were utilized to perform fault diagnosis by extracting the features in an Empirical Mode Decomposition technique and then the Artificial Neural Network model was utilized for the classification of faults.

Medina et al. [17] addressed the application of symbolic dynamics algorithms for the analysis of features from the vibration signals of gearbox. The key features were extracted from the vibration signals using a peak symbolic dynamics algorithm subdividing the phase space. Two different experiments were evaluated herewith with 10-fold cross-validation. In the first experiment, the dataset was partitioned randomly into 10-fold sets wherein 9 sets were considered for training and 1 set for the validation using the multi-class support vector machine model. In the second experiment, classification was performed by considering the signals at different load conditions.

Johnson et al. [18] analyzed the classification of faults using six variants of K-Nearest Neighbors (KNN) such as fine, weighted, medium, cosine, coarse and cubic. The processing of signals and analysis was performed on a MATLAB simulated high-voltage DC transmission line data.

One of the researchers, (Tang et al. [19]) introduced a custom model approach integrating K-Means with bio-inspired optimization models such as Ant Colony, Firefly, Cuckoo, Wolf, and Bat. Comparison of results enlisted by evaluating it on multiple datasets such as iris, wine, Libras, Haberman, synthetic and mouse. This research revealed how to overcome the key drawbacks of standard K-Means wherein the local optima is considered, and considering the global optima integrating with bio-inspired optimization models.

Liu et al. [20] acquired the vibration signals of multiple gears from an experimental test rig for the diagnosis of gears. This study highlights the application of ensemble empirical mode decomposition method to decompose the signals of multiple gear teeth at different levels to intrinsic mode functions. The extracted feature vectors were distinguished by a multi-class support vector machine to classify the health status of the gears. The resulting outputs with respective inputs were evaluated and compared herewith.

The existing literature has used different approaches in evaluating the gearbox fault diagnosis. Moreover, several kinds of research exist applying the vanilla deep learning models with few having in-depth customization. Hence, proposed herewith a novel approach by fine-tuning the parameters and activation functions of the conventional LSTM recurrent neural network model.

This novel approach constitutes of hybrid LSTM based model wherein optimization has been applied with swarm intelligence algorithms. Detailed evaluations and comparison of the results are performed using a gearbox condition monitoring data set to classify the faults accordingly. This article highlights the following:

- Proposal of a custom hybrid LSTM model with swarm intelligence for fault diagnosis of gearbox.
- In-depth analysis on a restricted subset of gearbox data with hybrid LSTM on 10 different load conditions.
- Evaluation on different LSTM activation function, i.e., Sigmoid, hyperbolic tangent (tanh), Rectified Linear Unit (ReLU); optimized with Particle Swarm Optimization (PSO), Firefly Algorithm (FA), Cuckoo Search Optimization (CSO), and Ant Colony Optimization (ACO) algorithms in combination with specified LSTM activation functions.

- Observations of each group of optimization and activation functions on all different 10 loads for the classification of faults.
- Tabulation of performance metrics such as accuracy, precision, recall, specificity, sensitivity, F-Score, weight, bias and activation functions such as default sigmoid, hyperbolic tangent, and ReLU activation functions along with its customized parameters.

3 Proposed Methodology

This proposed hybrid fault diagnosis methodology explains the approach of how the swarm intelligence algorithms have been combined with LSTM network model to classify the faults of gearbox. The proposed gearbox fault diagnosis method applying Hybrid-LSTM network model comprises of four steps and is illustrated in Fig. 1. Initially, the given dataset has been converted into a time series data and then given as input to the proposed Hybrid-LSTM network model.

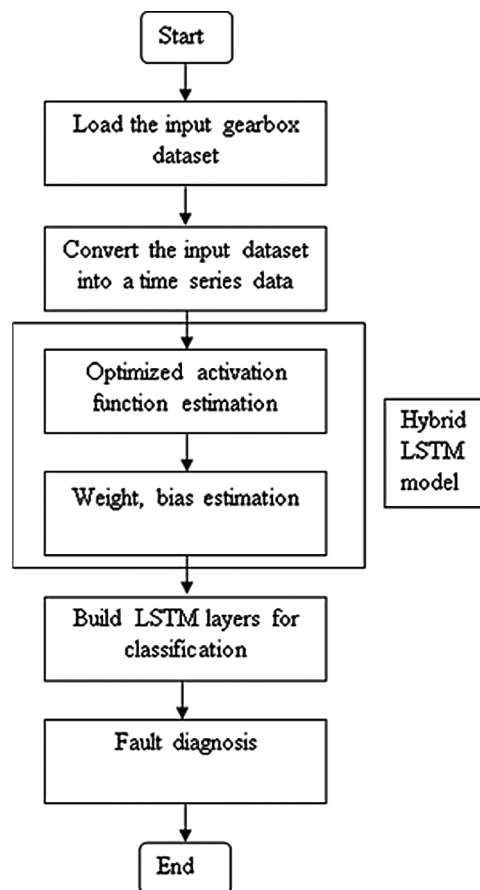


Figure 1: Flowchart of the proposed Hybrid-LSTM network model

3.1 Long Short-Term Memory Network

Long Short-Term Memory (LSTM) network is one of the variants of recurrent neural network. This has been widely applied on time series datasets recently and has been proven as a highly efficient learning neural network model among others.

An LSTM network enables to input sequence data into a network and make predictions based on the individual time steps of the sequence data. For conventional or default activation functions like sigmoid and tan h functions, the gradients decrease quickly with training error propagating to forward layers. The activation functions for sigmoid and hyperbolic tangent and its differential functions are formulated as in Eqs. (1)–(4) as:

$$f = \frac{1}{(1 + e^{-x})} \quad (1)$$

$$df = f * (1 - f) \quad (2)$$

$$f = \tan h(x) \quad (3)$$

$$df = (1 - f^2) \quad (4)$$

Recently, the ReLU activation function gained tremendous recognition, especially in the last few years. This is because its gradient will not decrease with the independent variables increasing. Hence, the network with ReLU activation function can overcome the vanishing of the gradient. The ReLU and its differential functions are mathematically formulated in the Eqs. (5) and (6).

$$f = \max(0, x) \quad (5)$$

$$df = \text{double}(x > 0) \quad (6)$$

A five-layer LSTM customized neural network is implemented with one sequence input layer, one bi-directional LSTM layer other than the default LSTM layer for standardization, one fully connected layer, one softmax layer and one output layer for classification. Each LSTM unit has an input gate, the forget gate, and the output gate along with the memory unit that is being read and updated periodically [21]. The formulated equations of LSTM have been listed as Eqs. (7)–(12):

$$f_t = q(w_{fx}x_i + w_{fh}h_{t-1} + b_f) \quad (7)$$

$$i_t = q(w_{ix}x_i + w_{ih}h_{t-1} + b_i) \quad (8)$$

$$g_t = m(w_{gx}x_i + w_{gh}h_{t-1} + b_g) \quad (9)$$

$$o_t = q(w_{ox}x_t + w_{oh}h_{t-1} + b_o) \quad (10)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes g_t \quad (11)$$

$$h_t = o_t \otimes m(c_t) \quad (12)$$

wherein x_i is the input of the memory cell, h_t is the input of the memory cell, f_t , i_t and O_t is the output of forget gate, input gate, and output gate respectively, C_t is the state of the memory cell. q is the gate activation function, \otimes denotes Hadamard product, m denotes the output activation function, w_{fx} , w_{fh} , w_{ix} , w_{ih} , w_{gx} , w_{gh} and w_{oh} represent the corresponding weight matrix, b_f , b_i , b_g and b_o are bias vectors. This is pictorially illustrated in Fig. 2.

3.2 Optimization Algorithms

In this research, few of the swarm intelligence algorithms Particle Swarm, Firefly, Cuckoo Search, and Ant Colony in combination with specified custom LSTM activation functions (in multiple combinations as well) by replacing their regular activation functions, i.e., Sigmoid, hyperbolic tangent and ReLU have been discussed and evaluated herewith. Using the parameters achieved by the respective equations mentioned

below (13,14,15,16) each for different customized activation functions, defining the layered network and the network is trained, validated, tested and executed for 5 epochs.

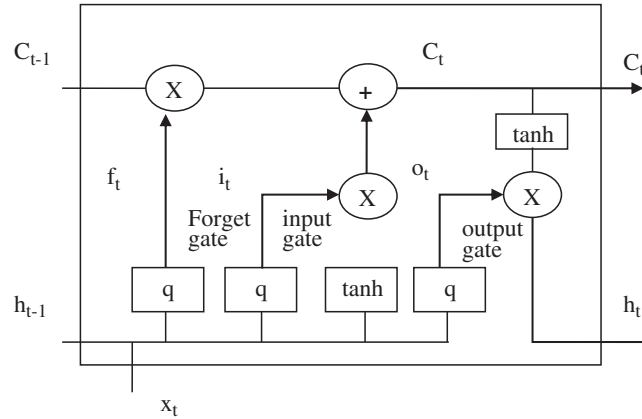


Figure 2: The LSTM Memory Cell Structure

3.2.1 Particle Swarm Optimization with LSTM

This optimization method commonly recognized as PSO [22] is an adaptive computational optimization technique that inspired from the psychological behavior of the birds flocking together. The objective function of PSO is formulated mathematically as in Eq. (13):

$$f = (x - 20)^2 \quad (13)$$

3.2.2 Cuckoo Optimization with LSTM

This optimization technique is derived from one of the bird's species known as Cuckoo following their strategy of laying eggs in the nests of different other bird species [23]. The objective function of cuckoo search is expressed mathematically as in Eq. (14):

$$f = (10 * size(x, 2) + \sum (x^2 - 10 * \cos(2\pi x))) \quad (14)$$

3.2.3 Firefly Algorithm Optimization with LSTM

Firefly Algorithm (FA) proposed here is based on the behavior of fireflies employing flashing signals to interact with light emitted by another brighter partner moving towards it. The attraction is commensurate to the brightness, i.e., both increase as their distance decreases [24,25]. Following this principle, the objective function of FA is mathematically formulated as in Eq. (15):

$$f = \sum (x - 1)^2 \quad (15)$$

3.2.4 Ant Colony Optimization with LSTM

The behavior of ants moving randomly by laying their pheromone and searing for an optimal path commonly recognized as Ant Colony Optimization. The advantage of ACO is that it shows prior success in evolving general RNNs for time series data prediction [26]. The equation for ACO utilized by the activation functions is formulated mathematically in Eq. (16):

$$f = \sum (-x * \sin(\sqrt{|x|})) \quad (16)$$

4 Simulation Results and Discussions

In this section, two fault conditional cases such as the healthy tooth and broken tooth have been considered with the gearbox fault diagnosis dataset [27] for classification. The upcoming subsections describe about the dataset details and the proposed hybrid method performance.

4.1 Dataset Description

This open source gearbox fault diagnosis dataset comprises of the vibration dataset recorded by using Spectra Quest's Gearbox Fault Diagnostics Simulator. Dataset has been recorded with the help of 4 vibration sensors placed in four different directions. Additionally, the dataset has been observed under variation of loads at a frequency of 30 Hz from '0' to '90' percent with two different scenarios: 1) Healthy condition and 2) Broken Tooth Condition.

4.2 Performance Metrics

Initially, the gearbox fault diagnosis dataset is converted to time series data to perform LSTM classification. The evaluation has been performed considering restricted data available in the dataset. To classify the two health conditions of the gearbox data, 70% samples have been employed to train the proposed hybrid network and the rest have been used for testing. That is for this work; Training and Testing dataset are of the order (4*14000) samples containing and (4*6000) samples that have been used as an input to the neural network. The learning rate is 1 and the iteration number or number of epochs is 5.

The proposed hybrid LSTM network has been trained on different conventional activation functions sigmoid, hyperbolic tangent (tanh) and ReLU. Additionally, this training has also been extended to other proposed customized activation functions such as sigmoid PSO, sigmoid Cuckoo, sigmoid FA, and sigmoid ACO; tanh PSO, tanh Cuckoo, tanh FA, and tanh ACO; ReLU PSO, ReLU Cuckoo, ReLU FA and ReLU ACO. The results have been evaluated for load, weight and bias values on a gearbox vibration data acquired at different loads 0, 10, 20, 30, 40, 50, 60 70, 80 and 90. The accuracy of each has been computed and best of the results were obtained taking into consideration every state of load. Especially, the best results were achieved with load 10 and 40. The results observed have been tabulated in [Tab. 1](#) and are also depicted in [Fig. 3](#).

Table 1: Evaluation of proposed hybrid LSTM network model

S. No.	Activation Function	Load	Bias	Weight	Accuracy
1	ReLU	0	0.2500	-0.0016	0.6250
2	Sigmoid-PSO		0.2502	-0.0019	0.6250
3	Sigmoid-Cuckoo		0.2503	-0.0035	0.7500
4	Sigmoid-Firefly		0.2502	0.0017	0.6250
5	Sigmoid-ACO		0.2502	0.0002	0.5000
6	ReLU-PSO		0.2500	0.0005	0.6250
7	ReLU-Cuckoo		0.2504	-0.0023	0.6250
8	ReLU-Firefly		0.2504	0.0006	0.6250
9	ReLU-ACO		0.2500	0.0005	0.1250
10	tanh-PSO		0.2503	-0.00268	0.6250
11	tanh-Cuckoo		0.2500	0.0015	0.5000
12	tanh-Firefly		0.2504	0.0022	0.6250
13	tanh-ACO		0.2506	0.0014	0.7500

(Continued)

Table 1 (continued).					
S. No.	Activation Function	Load	Bias	Weight	Accuracy
1	ReLU	10	0.2502	0.0005	0.6250
2	Sigmoid-PSO		2.50E-01	-2.92E-03	0.8750
3	Sigmoid-Cuckoo		2.50E-01	-7.90E-04	0.5000
4	Sigmoid-Firefly		2.50E-01	6.06E-04	0.5000
5	Sigmoid-ACO		2.50E-01	-2.83E-03	0.5000
6	ReLU-PSO		2.50E-01	-3.03E-03	0.5000
7	ReLU-Cuckoo		2.50E-01	6.81E-04	0.2500
8	ReLU-Firefly		2.50E-01	-4.26E-03	0.5000
9	ReLU-ACO		2.50E-01	1.08E-03	0.7500
10	tanh-PSO		0.2500	-0.0023	0.5000
11	tanh-Cuckoo		2.50E-01	-7.58E-05	0.3750
12	tanh-Firefly		2.50E-01	9.34E-04	0.5000
13	tanh-ACO		2.50E-01	1.66E-03	0.5000
1	ReLU	20	0.2501	-0.0019	0.6250
2	Sigmoid-PSO		2.50E-01	-1.33E-03	0.6250
3	Sigmoid-Cuckoo		2.50E-01	2.21E-03	0.6250
4	Sigmoid-Firefly		2.50E-01	-1.52E-05	0.7500
5	Sigmoid-ACO		2.50E-01	-1.33E-03	0.6250
6	ReLU-PSO		2.50E-01	1.79E-03	0.5000
7	ReLU-Cuckoo		2.50E-01	-1.49E-03	0.3750
8	ReLU-Firefly		2.50E-01	1.32E-05	0.5000
9	ReLU-ACO		2.50E-01	-9.22E-04	0.7500
10	tanh-PSO		2.50E-01	0.0027	0.7500
11	tanh-Cuckoo		2.50E-01	-1.64E-04	0.6250
12	tanh-Firefly		2.50E-01	-2.17E-03	0.7500
13	tanh-ACO		2.50E-01	1.08E-03	0.5000
1	ReLU	30	0.2499	-0.0038	0.5000
2	Sigmoid-PSO		2.50E-01	7.58E-04	0.5000
3	Sigmoid-Cuckoo		2.50E-01	-1.74E-03	0.1250
4	Sigmoid-Firefly		2.50E-01	2.83E-03	0.5000
5	Sigmoid-ACO		2.50E-01	1.58E-03	0.6250
6	ReLU-PSO		2.50E-01	-3.03E-03	0.5000
7	ReLU-Cuckoo		2.50E-01	-3.20E-03	0.5000
8	ReLU-Firefly		2.50E-01	1.35E-03	0.6250
9	ReLU-ACO		2.50E-01	2.68E-03	0.7500

Table 1 (continued).					
S. No.	Activation Function	Load	Bias	Weight	Accuracy
10	tanh-PSO		2.50E-01	-2.07E-03	0.5000
11	tanh-Cuckoo		2.50E-01	-1.35E-03	0.3750
12	tanh-Firefly		2.50E-01	-9.23E-04	0.5000
13	tanh-ACO		2.50E-01	4.30E-04	0.6250
1	ReLU	40	0.2498	-0.0019	0.5000
2	Sigmoid-PSO		2.50E-01	-5.22E-04	0.5000
3	Sigmoid-Cuckoo		2.50E-01	-2.25E-03	0.5000
4	Sigmoid-Firefly		2.50E-01	-1.61E-04	0.7500
5	Sigmoid-ACO		2.50E-01	-1.46E-03	0.6250
6	ReLU-PSO		0.2502	-5.30014	0.7500
7	ReLU-Cuckoo		2.50E-01	2.64E-03	0.8750
8	ReLU-Firefly		2.50E-01	1.97e-03	0.7500
9	ReLU-ACO		2.50E-01	7.12E-04	0.7500
10	tanh-PSO		2.50E-01	1.54E-03	0.5000
11	tanh-Cuckoo		2.50E-01	-1.35E-03	0.7500
12	tanh-Firefly		2.50E-01	-2.82E-03	0.6250
13	tanh-ACO		2.50E-01	1.35E-03	0.7500
1	ReLU	50	2.51E-01	1.00E-04	0.3750
2	Sigmoid-PSO		2.50E-01	1.52E-04	0.5000
3	Sigmoid-Cuckoo		2.50E-01	3.61E-04	0.6250
4	Sigmoid-Firefly		2.50E-01	1.17E-03	0.3750
5	Sigmoid-ACO		2.50E-01	-1.78E-03	0.2500
6	ReLU-PSO		2.50E-01	-4.84E-04	0.2500
7	ReLU-Cuckoo		2.50E-01	3.53E-03	0.6250
8	ReLU-Firefly		2.50E-01	6.33E-04	0.5000
9	ReLU-ACO		2.50E-01	4.54E-04	0.3750
10	tanh-PSO		2.50E-01	3.36E-03	0.5000
11	tanh-Cuckoo		2.50E-01	-1.15E-03	0.3750
12	tanh-Firefly		2.50E-01	-4.92E-04	0.6250
13	tanh-ACO		2.50E-01	-2.35E-03	0.2500
1	ReLU	60	2.50E-01	-3.54E-03	0.6250
2	Sigmoid-PSO		2.50E-01	-2.31E-03	0.6250
3	Sigmoid-Cuckoo		2.50E-01	2.28E-03	0.6250
4	Sigmoid-Firefly		2.50E-01	-1.46E-04	0.3750
5	Sigmoid-ACO		2.50E-01	-3.14E-03	0.1250

(Continued)

Table 1 (continued).

S. No.	Activation Function	Load	Bias	Weight	Accuracy
6	ReLU-PSO		2.50E-01	-1.30E-03	0.5000
7	ReLU-Cuckoo		2.50E-01	-1.64E-04	0.7500
8	ReLU-Firefly		2.50E-01	1.05E-04	0.7500
9	ReLU-ACO		2.50E-01	-1.23E-03	0.5000
10	tanh-PSO		2.50E-01	-1.20E-04	0.7500
11	tanh-Cuckoo		2.50E-01	-1.17E-03	0.7500
12	tanh-Firefly		2.50E-01	1.18E-03	0.5000
13	tanh-ACO		2.50E-01	-2.02E-04	0.5000
1	ReLU	70	2.50E-01	-233E-03	0.5000
2	Sigmoid-PSO		2.50E-01	1.51E-04	0.3750
3	Sigmoid-Cuckoo		2.50E-01	7.72E-04	0.5000
4	Sigmoid-Firefly		2.50E-01	-1.91E-03	0.3750
5	Sigmoid-ACO		2.50E-01	2.56E-03	0.2500
6	ReLU-PSO		2.51E-01	1.26E-03	0.6250
7	ReLU-Cuckoo		2.50E-01	2.91E-03	0.5000
8	ReLU-Firefly		2.50E-01	1.95E-03	0.5000
9	ReLU-ACO		2.50E-01	9.59E-09	0.6250
10	tanh-PSO		2.51E-01	2.79E-03	0.5000
11	tanh-Cuckoo		2.50E-01	-8.82E-04	0.6250
12	tanh-Firefly		2.50E-01	1.73E-04	0.7500
13	tanh-ACO		2.50E-01	-9.11E-04	0.5000
1	ReLU	80	2.51E-01	1.27E-03	0.3750
2	Sigmoid-PSO		2.51E-01	8.56E-04	0.5000
3	Sigmoid-Cuckoo		2.50E-01	-4.33E-03	0.7500
4	Sigmoid-Firefly		2.50E-01	2.47E-04	0.5000
5	Sigmoid-ACO		2.50E-01	3.64E-04	0.1250
6	ReLU-PSO		2.50E-01	5.32E-03	0.5000
7	ReLU-Cuckoo		2.50E-01	4.40E-04	0.5000
8	ReLU-Firefly		2.50E-01	1.95E-03	0.5000
9	ReLU-ACO		2.50E-01	-1.59E-03	0.3750
10	tanh-PSO		2.50E-01	1.99E-03	0.5000
11	tanh-Cuckoo		2.50E-01	1.12E-03	0.5000
12	tanh-Firefly		2.50E-01	-2.38E+00	0.7500
13	tanh-ACO		2.251E-01	5.57E-04	0.5000

Table 1 (continued).

S. No.	Activation Function	Load	Bias	Weight	Accuracy
1	ReLU	90	2.50E-01	-2.47E-03	0.6250
2	Sigmoid-PSO		2.50E-01	1.36E-03	0.5000
3	Sigmoid-Cuckoo		2.50E-01	1.14E-03	0.3750
4	Sigmoid-Firefly		2.50E-01	-2.81E-03	0.5000
5	Sigmoid-ACO		2.51E-01	1.06E-03	0.2500
6	ReLU-PSO		2.50E-01	5.32E-03	0.5000
7	ReLU-Cuckoo		2.50E-01	-1.63E-03	0.3750
8	ReLU-Firefly		2.50E-01	-1.58E-03	0.6250
9	ReLU-ACO		2.50E-01	-1.42E-03	0.5000
10	tanh-PSO		2.50E-01	6.20E-04	0.5000
11	tanh-Cuckoo		2.50E-01	8.72E-04	0.5000
12	tanh-Firefly		2.50E-01	-1.70E-03	0.6250
13	tanh-ACO		2.50E-01	8.56E-04	0.2500

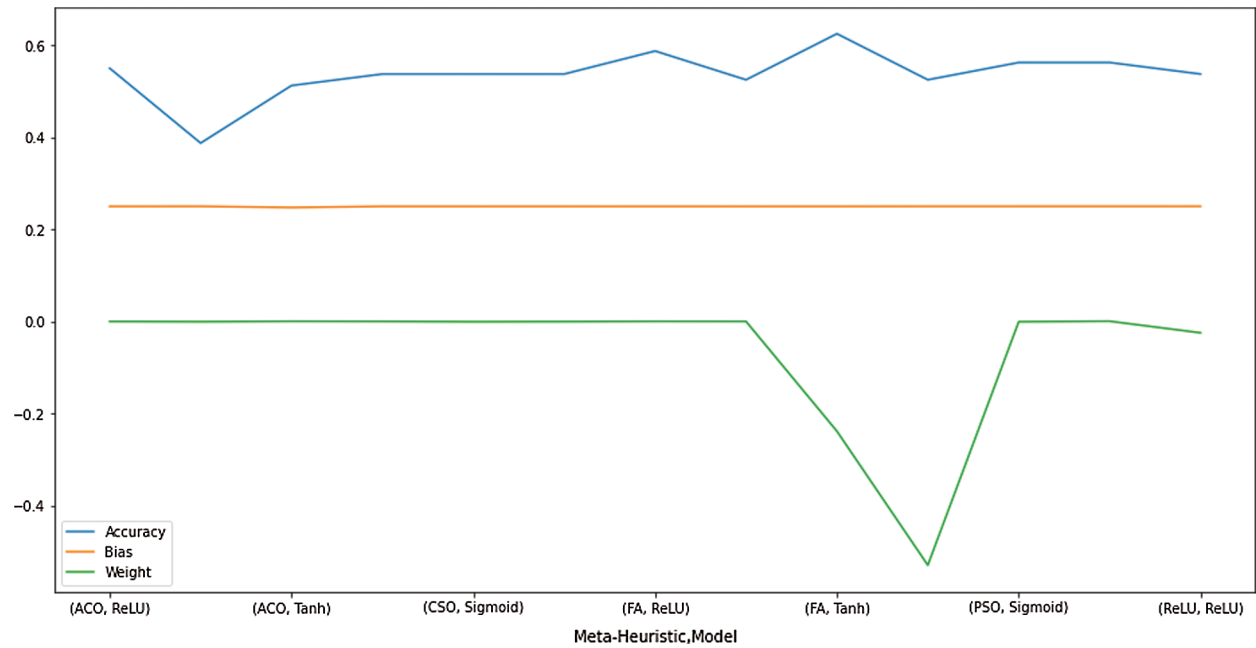


Figure 3: Performance of the proposed Hybrid LSTM Network Model

Further, the performance of the gearbox classification is examined by using various performance parameters that presents the predicted and expected/actual classifications. The result of classifying is predicted into two categories such as healthy and broken tooth conditions.

Few of the performance parameters considered here like Accuracy, true positive rate (Sensitivity), false positive rate (Specificity), Precision, and F-Score. A higher value of ‘True Positive’ detection is enviable for

vigorous gearbox classification. Accuracy is formulated as the percentage of the number of faults classified correctly versus total faults as in Eq. (17). TPR or sensitivity is defined as the ratio of correctly predicted faults to the size of the actual faults and is formulated in Eq. (18). Precision is defined as the ratio of correctly predicted faults to the predicted size of the faults and is formulated as in Eq. (20). Consequently, the classification accuracy is estimated using the following formulae:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (17)$$

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (18)$$

$$\text{Specificity} = \frac{TN}{(FP + TN)} \quad (19)$$

$$\text{Precision} = \frac{TP}{(TP + FP)} \quad (20)$$

$$F_1 \text{ Score} = 2 * \left\{ \frac{(\text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})} \right\} \quad (21)$$

where True Positives (TP) is the number of faults classified as faults, True Negatives (TN) is the number of normal classified as normal, False Positives (FP) is the number of normal classified as faults and False Negatives (FN) is the number of faults classified as normal. Tab. 2 describes the values for various performance metrics such as the sensitivity, specificity, precision, recall, F-score including the execution time for all customized activation functions.

Table 2: Performance parameters of Hybrid LSTM model

Load	Customized Activation	Execution time	Specificity	Sensitivity	Precision	Recall	F-score
0	ReLU	3 m 19 s	0.2500	1	0.5714	1	0.7273
	Sigmoid-PSO	3 m 18 s	0.2500	1	0.5714	1	0.7273
	Sigmoid-Cuckoo	3 m 19 s	0.5000	1	0.6667	1	0.8000
	Sigmoid-Firefly	3 m 22 s	0.2500	1	0.5714	1	0.7273
	Sigmoid-ACO	3 m 18 s	0	1	0.5000	1	0.6667
	ReLU-PSO	3 m 20 s	0.2500	1	0.5714	1	0.7273
	ReLU-Cuckoo	3 m 20 s	0.2500	1	0.5714	1	0.7273
	ReLU-Firefly	3 m 21 s	0.5000	0.7500	0.6000	0.7500	0.6661
	ReLU-ACO	3 m 31 s	0.2500	0	0	0	0
	tanh-PSO	3 m 14 s	0.2500	1	0.5714	1	0.7273
	tanh-Cuckoo	3 m 21 s	0	1	0.5000	1	0.6667
	tanh-Firefly	3 m 17 s	0.2500	1	0.5714	1	0.7273
	tanh-ACO	3 m 17 s	0.5000	1	0.6667	1	0.8000
10	ReLU	3 m 23 s	0.7500	0.5000	0.6667	0.5000	0.5714
	Sigmoid-PSO	3 m 10 s	0.7500	1	0.8000	1	0.8889
	Sigmoid-Cuckoo	3 m 6 s	0.2500	0.7500	0.5000	0.7500	0.6000

Table 2 (continued).							
Load	Customized Activation	Execution time	Specificity	Sensitivity	Precision	Recall	F-score
	Sigmoid-Firefly	2 m 53 s	1	0	–	0	0
	Sigmoid-ACO	3 m 13 s	0.2500	0.7500	0.5000	0.7500	0.6000
	ReLU-PSO	2 m 58 s	0	1	0.5000	1	0.6667
	ReLU-Cuckoo	3 m 8 s	0.2500	0.2500	0.2500	0.2500	0.2500
	ReLU-Firefly	3 m 9 s	1	0	–	0	0
	ReLU-ACO	3 m 1 s	0.7500	0.7500	0.7500	0.7500	0.7500
	tanh-PSO	2 m 48 s	0.2500	0.7500	0.5000	0.7500	0.6000
	tanh-Cuckoo	3 m 1 s	0.7500	0	0	0	0
	tanh-Firefly	2 m 55 s	1	0	–	0	0
	tanh-ACO	3 m 36 s	0	1	0.5000	1	0.6667
20	ReLU	3 m 34 s	1	0.2500	1	0.2500	0.4000
	Sigmoid-PSO	2 m 45 s	1	0.2500	1	0.2500	0.4000
	Sigmoid-Cuckoo	3 m 5 s	0.5000	0.7500	0.6000	0.7500	0.6667
	Sigmoid-Firefly	3 m 2 s	1	0.5000	1	0.5000	0.6667
	Sigmoid-ACO	3 m 4 s	0.2500	1	0.5714	1	0.7273
	ReLU-PSO	3 m 3 s	0.7500	0.2500	0.5000	0.2500	0.3333
	ReLU-Cuckoo	3 m 3 s	0.5000	0.2500	0.3333	0.2500	0.2857
	ReLU-Firefly	2 m 55 s	0	1	0.5000		0.6667
	ReLU-ACO	3 m 0 s	0.7500	0.7500	0.7500	0.7500	0.7500
	tanh-PSO	2 m 47 s	1	0.5000	1	0.5000	0.6667
	tanh-Cuckoo	3 m 10 s	0.7500	0.5000	0.6667	0.5000	0.5714
	tanh-Firefly	3 m 16 s	1	0	1	0	0.6667
	tanh-ACO	3 m 18 s	0	1	0.5000	1	0.6667
30	ReLU	3 m 24 s	1	0	–	0	0
	Sigmoid-PSO	3 m 1 s	0	1	0	1	0.6667
	Sigmoid-Cuckoo	3 m 2 s	0.2500	0	0	0	0
	Sigmoid-Firefly	2 m 48 s	0	1	0.5000	1	0.6667
	Sigmoid-ACO	2 m 55 s	0.5000	0.7500	0.6000	0.7500	0.6667
	ReLU-PSO	3 m 1 s	0.7500	0.2500	0.5000	0.2500	0.3333
	ReLU-Cuckoo	3 m 12 s	0.7500	0.2500	0.5000	0.2500	0.3333
	ReLU-Firefly	2 m 55 s	1	0.2500	1	0.2500	0.4000
	ReLU-ACO	3 m 5 s	0.7500	0.7500	0.7500	0.7500	0.7500
	tanh-PSO	2 m 57 s	0	1	0.5000	1	0.6667
	tanh-Cuckoo	3 m 3 s	0.7500	0	0	0	0
	tanh-Firefly	2 m 54 s	0	1	0.5000	1	0.6667
	tanh-ACO	2 m 56 s	1	0.2500	1	0.2500	0.4000

(Continued)

Table 2 (continued).							
Load	Customized Activation	Execution time	Specificity	Sensitivity	Precision	Recall	F-score
40	ReLU	3 m 21 s	0	1	0.5000	1	0.6667
	Sigmoid-PSO	3 m 18 s	1	0	–	0	0
	Sigmoid-Cuckoo	3 m 6 s	0.2500	0.7500	0.5000	0.7500	0.6000
	Sigmoid-Firefly	2 m 50 s	1	0.5000	1	0.5000	0.6667
	Sigmoid-ACO	3 m 6 s	0.5000	0.7500	0.6000	0.7500	0.6667
	ReLU-PSO	3 m 1 s	0.7500	0.2500	0.5000	0.2500	0.3333
	ReLU-Cuckoo	3 m 4 s	1	0.7500	1	0.7500	0.8751
	ReLU-Firefly	2 m 49 s	1	0.5000	1	0.5000	0.6667
	ReLU-ACO	3 m 8 s	0.7500	0.7500	0.7500	0.7500	0.7500
	tanh-PSO	3 m 1 s	1	0	–	0	0
	tanh-Cuckoo	2 m 53 s	0.7500	0.7500	0.7500	0.7500	0.7500
	tanh-Firefly	3 m 1 s	1	0	1	0.2500	0.4000
	tanh-ACO	2 m 56 s	1	0.5000	1	0.5000	0.6667
50	ReLU	3 m 22 s	0.2500	0.5000	0.4000	0.5000	0.4444
	Sigmoid-PSO	3 m 12 s	0	1	0.5000	1	0.6667
	Sigmoid-Cuckoo	3 m 3 s	0.2500	1	0.5714	1	0.7273
	Sigmoid-Firefly	2 m 49 s	0.2500	0.5000	0.4000	0.5000	0.4444
	Sigmoid-ACO	2 m 55 s	0.2500	0.2500	0.2500	0.2500	0.2500
	ReLU-PSO	3 m 6 s	0.5000	0	0	0	0
	ReLU-Cuckoo	3 m 15 s	1	0.2500	1	0.2500	0.4000
	ReLU-Firefly	2 m 57 s	1	0	–	0	0
	ReLU-ACO	3 m 2 s	0.5000	0.2500	0.3333	0.2500	0.2857
	tanh-PSO	2 m 39 s	0	1	0.5000	1	0.6667
	tanh-Cuckoo	3 m 1 s	0	0.7500	0.4286	0.7500	0.5455
	tanh-Firefly	2 m 58 s	0.2500	1	0.5714	1	0.7273
	tanh-ACO	3 m 1 s	0.5000	0	0	0	0
60	ReLU	3 m 14 s	0.2500	1	0.5714	1	0.7273
	Sigmoid-PSO	3 m 8 s	0.7500	0.5000	0.6667	0.5000	0.5714
	Sigmoid-Cuckoo	3 m 4 s	0.5000	0.7500	0.6000	0.7500	0.6667
	Sigmoid-Firefly	3 m 2 s	0.5000	0.2500	0.3333	0.2500	0.2857
	Sigmoid-ACO	3 m 4 s	0.2500	0	0	0	0
	ReLU-PSO	3 m 1 s	1	0	–	0	0
	ReLU-Cuckoo	3 m 4 s	0.7500	0.7500	0.7500	0.7500	0.7500
	ReLU-Firefly	2 m 55 s	0.5000	1	0.6667	1	0.8000
	ReLU-ACO	2 m 55 s	0.5000	0.2500	0.3333	0.2500	0.2857

Table 2 (continued).							
Load	Customized Activation	Execution time	Specificity	Sensitivity	Precision	Recall	F-score
	tanh-PSO	3 m 19 s	0.5000	1	0.6667	1	0.8000
	tanh-Cuckoo	3 m 3 s	0.7500	0.7500	0.7500	0.7500	0.7500
	tanh-Firefly	2 m 56 s	1	0	–	0	0
	tanh-ACO	2 m 55 s	0.7500	0.2500	0.5000	0.2500	0.3333
70	ReLU	3 m 22 s	0	1	0.5000	1	0.6667
	Sigmoid-PSO	3 m 12 s	0.5000	0.2500	0.3333	0.2500	0.2857
	Sigmoid-Cuckoo	2 m 57 s	0.2500	0.7500	0.5000	0.7500	0.6000
	Sigmoid-Firefly	3 m 9 s	0.2500	0.5000	0.4000	0.5000	0.4444
	Sigmoid-ACO	2 m 57 s	0.2500	0	0	0	0
	ReLU-PSO	3 m 6 s	0.2500	1	0.5714	1	0.7273
	ReLU-Cuckoo	2 m 53 s	0.5000	0.5000	0.5000	0.5000	0.5000
	ReLU-Firefly	2 m 46 s	0.2500	0.7500	0.5000	0.7500	0.6000
	ReLU-ACO	3 m 5 s	1	0.2500	1	0.2500	0.4000
	tanh-PSO	3 m 4 s	0	1	0.5000	1	0.6667
	tanh-Cuckoo	2 m 57 s	0.2500	1	0.5714	1	0.7273
	tanh-Firefly	2 m 59 s	0.5000	1	0.6667	1	0.8000
	tanh-ACO	3 m 0 s	0.7500	0.2500	0.5000	0.2500	0.3333
80	ReLU	3 m 15 s	0.5000	0.2500	0.3333	0.2500	0.2857
	Sigmoid-PSO	3 m 1 s	0	1	0.5000	1	0.6667
	Sigmoid-Cuckoo	3 m 4 s	0.5000	1	0.6667	1	0.8000
	Sigmoid-Firefly	3 m 3 s	0	1	0.5000	1	0.6667
	Sigmoid-ACO	3 m 10 s	0.2500	0	0	0	0
	ReLU-PSO	2 m 49 s	0	1	0.5000	1	0.6667
	ReLU-Cuckoo	3 m 11 s	0	1	0.5000	1	0.6667
	ReLU-Firefly	2 m 51 s	1	0	–	0	0
	ReLU-ACO	2 m 50 s	0.7500	0	0	0	0
	tanh-PSO	2 m 59 s	0.2500	0.7500	0.5000	0.7500	0.6000
	tanh-Cuckoo	3 m 8 s	0.7500	0.2500	0.5000	0.2500	0.3333
	tanh-Firefly	3 m 3 s	0.5000	1	0.6667	1	0.8000
	tanh-ACO	2 m 59 s	0.7500	0.2500	0.5000	0.2500	0.3333
90	ReLU	3 m 14 s	1	0.2500	1	0.2500	0.4000
	Sigmoid-PSO	3 m 4 s	0	1	0.5000	1	0.6667
	Sigmoid-Cuckoo	3 m 1 s	0.5000	0.2500	0.3333	0.2500	0.2857
	Sigmoid-Firefly	3 m 10 s	0	1	0.5000	1	0.6667
	Sigmoid-ACO	3 m 4 s	0.2500	0.2500	0.2500	0.2500	0.2500

(Continued)

Table 2 (continued).

Load	Customized Activation	Execution time	Specificity	Sensitivity	Precision	Recall	F-score
	ReLU-PSO	3 m 12 s	1	0.2500	1	0.2500	0.4000
	ReLU-Cuckoo	3 m 5 s	0.7500	0	0	0	0
	ReLU-Firefly	3 m 14 s	1	0.2500	1	0.2500	0.4000
	ReLU-ACO	3 m 2 s	0.7500	0.2500	0.3333	0.2500	0.2857
	tanh-PSO	3 m 7 s	0	1	0.5000	1	0.6667
	tanh-Cuckoo	3 m 4 s	1	0	–	0	0
	tanh-Firefly	3 m 0 s	0.2500	1	0	1	0.7273
	tanh-ACO	3 m 14 s	0.5000	0	0	0	0

4.3 Comparison Results

The proposed hybrid LSTM network is compared with conventional LSTM network with parameters such as bias, weight, execution time, accuracy, precision and recall. [Tab. 3](#) describes the values of the LSTM network model for conventional LSTM parameters such as the sigmoid activation functions, tan-h activation function without ReLU activation function, weight, bias and accuracy.

Table 3: Performance parameters of conventional LSTM

Load	Activation functions	Bias	Weight	Accuracy	Execution time	Precision	Recall
0	Sigmoid	0.2525	9.3737	0.6250	3m 20sec	0.5714	1
	Tan-h	0.25039	−0.00018	0.6250	4m 3s	0.6000	0.7500
10	Sigmoid	0.2503	0.0018	0.5000	3m 25s	0.5000	1
	Tan-h	0.2505	−0.0006	0.7750	3m 25s	0.3333	0.2500
20	Sigmoid	0.2508	0.0023	0.2500	3m 19s	0.3333	0.5000
	Tan-h	0.2503	−0.0031	0.5000	3m 24s	0.5000	0.5000
30	Sigmoid	0.2508	−0.0023	0.6250	3m 17s	0.6667	0.5000
	Tan-h	0.2505	−0.005	0.6250	3m 29s	0.6000	0.7500
40	Sigmoid	2.51E-01	3.12E-04	0.5000	3m 28s	–	0
	Tan-h	0.2506	−0.0001	0.5000	3m 29s	–	0
50	Sigmoid	2.50E-01	3.27E-04	0.5000	3m 21s	0.5000	1
	Tan-h	2.51E-01	−1.59E-03	0.5000	3m 39s	0.5000	0.2500
60	Sigmoid	2.51E-01	−2.29E-03	0.6250	3m 23s	0.6667	0.5000
	Tan-h	2.51E-01	1.96E-03	0.6250	3m 38s	0.6000	0.7500
70	Sigmoid	2.50E-01	7.78E-04	0.5000	3m 25s	0.5000	0.5000
	Tan-h	2.51E-01	1.52E-03	0.5000	3m 32s	0.5000	0.5000
80	Sigmoid	2.50E-01	−7.89E-03	0.6250	3m 21s	1	0.2500
	Tan-h	2.50E-01	−3.38E-04	0.2500	3m 32s	0.2500	0.2500
90	Sigmoid	2.51E-01	2.98E-03	0.2500	3m 25s	0.3333	0.5000
	Tan-h	2.51E-01	1.34E-03	0.5000	3m 32s	0.5000	0.2500

The comparison results for accuracy measurement for proposed hybrid network and conventional network are also illustrated in Fig. 4. From the graph it is evident that the proposed hybrid LSTM network shows better accuracy values of 87.5% for sigmoid PSO for load 10 and ReLU-Cuckoo for load 40. Consequently, the second highest value of 75% is identified with the customized activation functions like tanh-Firefly, tanh-Cuckoo, tanh-PSO, and tanh-ACO, Sigmoid-Cuckoo, Sigmoid-ACO, Sigmoid-Firefly, ReLU-Firefly, ReLU-Cuckoo, ReLU-ACO and ReLU-PSO. Thus, from the graph it is evident that the recommended hybrid LSTM network model shows better performance results when compared to the conventional LSTM network.

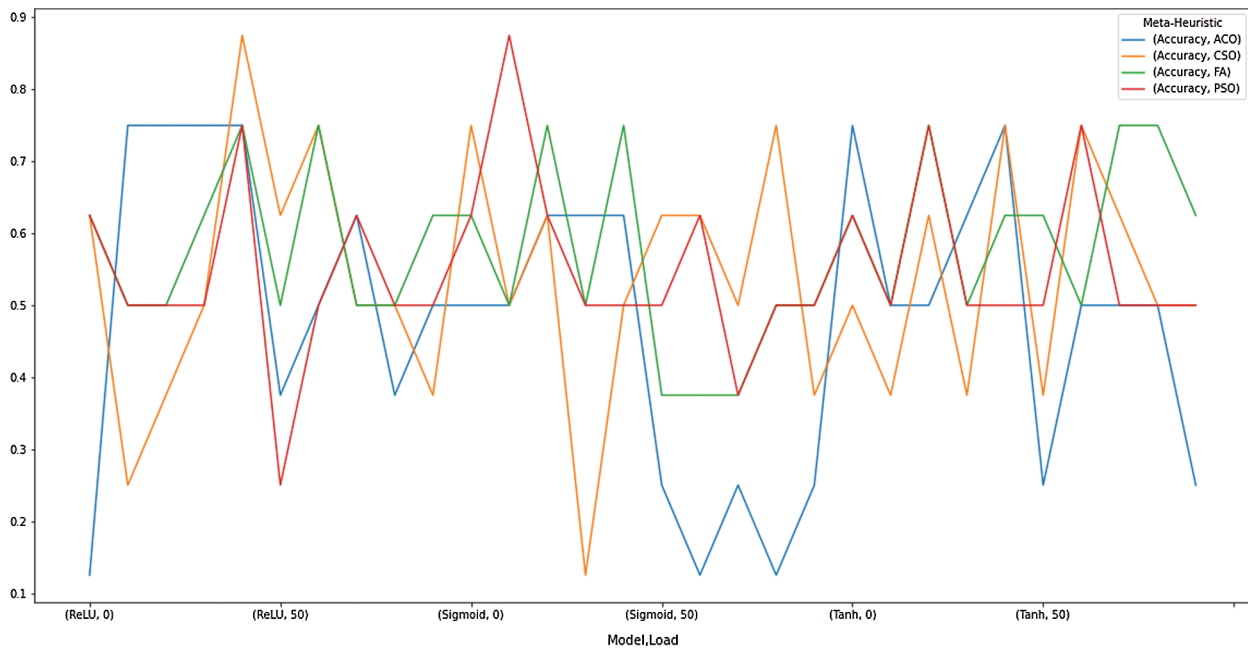


Figure 4: Accuracy comparison graph of proposed Hybrid LSTM neural network along with activation functions evaluated on different loads

5 Conclusion

In this study, the proposed hybrid LSTM network model along with different swarm intelligence algorithms has been evaluated for the fault diagnosis of the gearbox. In order, to address the challenges of over-fitting and enhancing the performance of conventional LSTM with a tiny training set, swarm intelligence optimization algorithms such as PSO, Cuckoo, Firefly and ACO along with ReLU activation function have been considered. From the evaluated results highest accuracy of 87.5% has been achieved with both Sigmoid-PSO and ReLU-Cuckoo customized activation functions. The results highlight that the proposed method would achieve higher accuracy in condition monitoring of gears for fault diagnosis. Comparative studies have also indicated that results of hybridization optimized with swarm intelligence are superior to the conventional LSTM model.

Acknowledgement: The authors would like to acknowledge the Artificial Intelligence and Data Analytics (AIDA) Laboratory, CCIS, Prince Sultan University, Riyadh, Saudi Arabia for the support.

Conflict of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

Funding Statement: The authors received no specific funding for this study.

References

- [1] C. Lu, Z. Wang and B. Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Advanced Engineering Informatics*, vol. 32, pp. 139–151, 2017.
- [2] Z. Chen, C. Li and R. V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock and Vibration*, vol. 2015, no. 2, pp. 1–10, 2015.
- [3] L. Jing, M. Zhao, P. Li and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017.
- [4] R. V. Sanchez, Z. Chen, X. Chen, C. Li and H. Qin, "Vibration-based gearbox fault diagnosis using deep neural networks," *Journal of Vibroengineering*, vol. 19, no. 4, pp. 2475–2496, 2017.
- [5] Y. Qi, C. Shen, J. Shi, W. Huang and Z. Zhu, "Adaptive morphological feature extraction and support vector regressive classification for gearbox fault diagnosis," in *Proc. of IEEE/ASME Int. Conf. on Advanced Intelligent Mechatronics*, 2019.
- [6] G. Jiang, H. He, P. Xie and Y. Tang, "Stacked multilevel-denoising autoencoders: A new representation learning approach for wind turbine gearbox fault diagnosis," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017.
- [7] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera *et al.*, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mechanical Systems and Signal Processing*, vol. 76–77, pp. 283–293, 2016.
- [8] S. Xue and I. Howard, "Torsional vibration signal analysis as a diagnostic tool for planetary gear fault detection," *Mechanical Systems and Signal Processing*, vol. 100, pp. 706–728, 2018.
- [9] P. Wang, Y. Ananya, R. Yan and R. X. Gao, "Virtualization and deep recognition for system fault classification," *Journal of Manufacturing Systems*, vol. 44, pp. 310–316, 2017.
- [10] W. Teng, X. Ding, X. Zhang, Y. Liu and Z. Ma, "Multi-fault detection and failure analysis of wind turbine gearbox using complex wavelet transform," *Renewable Energy*, vol. 93, pp. 591–598, 2016.
- [11] R. Bajric, N. Zuber, G. A. Skrimpas and N. Mijatovic, "Feature extraction using discrete wavelet transform for gear fault diagnosis of wind turbine gearbox," *Shock and Vibration*, vol. 2016, no. 1, pp. 1–10, 2016.
- [12] Y. Zhang, W. Lu and F. Chu, "Planet gear fault localization for wind turbine gearbox using acoustic emission signals," *Renewable Energy*, vol. 109, pp. 449–460, 2017.
- [13] Z. Chen, S. Deng, X. Chen, C. Li, R.V. Sanchez *et al.*, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectronics Reliability*, vol. 75, pp. 327–333, 2017.
- [14] B. Merainani, C. Rahmoune, D. Benazzouz and B. Ould-Bouamama, "A novel gearbox fault feature extraction and classification using Hilbert empirical wavelet transform, singular value decomposition, and SOM neural network," *Journal of Vibration and Control*, vol. 24, no. 12, pp. 2512–2531, 2017.
- [15] G. Liu, H. Bao and B. Han, "A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis," *Mathematical Problems in Engineering*, vol. 2018, no. 5, pp. 1–10, 2018.
- [16] H. Malik, Y. Pandya, A. Parashar and R. Sharma, "Feature extraction using EMD and classifier through artificial neural networks for gearbox fault diagnosis," in *Advances in Intelligent Systems and Computing*. Singapore: Springer, pp. 309–317, 2018.
- [17] R. Medina, J. C. Macancela, P. Lucero, D. Cabrera, M. Cerrada *et al.*, "Vibration signal analysis using symbolic dynamics for gearbox fault diagnosis," *The International Journal of Advanced Manufacturing Technology*, vol. 104, no. 5–8, pp. 2195–2214, 2019.
- [18] J. M. Johnson and A. Yadav, "Fault detection and classification technique for HVDC transmission lines using KNN," in *Information and Communication Technology for Sustainable Development*. Singapore: Springer, pp. 245–253, 2017.
- [19] R. Tang, S. Fong, X. S. Yang and S. Deb, "Integrating nature-inspired optimization algorithms to K-means clustering," in *Proc. of Seventh Int. Conf. on Digital Information Management*, 2012.

- [20] L. Liu, X. Liang and M. J. Zuo, "A dependence-based feature vector and its application on planetary gearbox fault classification," *Journal of Sound and Vibration*, vol. 431, pp. 192–211, 2018.
- [21] L. Zhang, B. Wang, B. Fang, H. Ma, Z. Yang *et al.*, "Two-stage short-term wind speed prediction based on LSTM-LSSVM-CFA," in *Proc. of 2nd IEEE Conf. on Energy Internet and Energy System Integration (EI2)*, 2018.
- [22] H. Moalla, W. Elloumi and A. M. Alimi, "H-PSO-LSTM: Hybrid LSTM trained by PSO for online handwriter identification," in *Neural Information Processing*, vol. 10637. New York, NY: Springer International Publishing, pp. 41–50, 2017.
- [23] E. Shadkam and M. Bijari, "Evaluation the efficiency of cuckoo optimization algorithm," *International Journal on Computational Science & Applications*, vol. 4, no. 2, pp. 39–47, 2014.
- [24] D. Wang, H. Luo, O. Grunder, Y. Lin and H. Guo, "Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm," *Applied Energy*, vol. 190, pp. 390–407, 2017.
- [25] X. S. Yang, "Firefly algorithm, stochastic test functions and design optimisation," *International Journal of Bio-Inspired Computation*, vol. 2, no. 2, pp. 78, 2010.
- [26] A. ElSaid, B. Wild, F. E. Jamiy, J. Higgins and T. Desell, "Optimizing LSTM RNNs using ACO to predict turbine engine vibration," in *Proc. of GECCO 2017: Genetic and Evolutionary Computation Conf.*, 2017.
- [27] OpenEI, "Gear Box Fault Diagnosis Data Set," 2019. [Online]. Available: <https://openei.org/datasets/dataset/gearbox-fault-diagnosis-data/resource/affa53da-cae6-42f2-b898-ad018ff91641>.