AUT Journal of Electrical Engineering

AUT J. Elec. Eng., 54(1) (2022) 107-120
DOI: 10.22060/eej.2021.20438.5430

# Bearing Fault Detection and Classification Based on Temporal Convolutions and LSTM Network in Induction Machine

M. Hoseintabar Marzebali[1], S. Hasani Borzadaran[2], H. Mashayekhi[2], V. Mashayekhi[1*]

[1] Department of Electrical Engineering, Shahrood University of Technology, Shahrood, Iran
[2]Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran

**ABSTRACT:** One of the critical components in most electromechanical systems are the bearing system. Therefore, a proper condition monitoring method that can classify the type and the severity of electrical machine faults in different load levels is crucial to avoid unwanted downtime and loss of operation. Non-invasive condition monitoring methods based on electrical signatures of machine in an electromechanical system, are considered as simple and cost-effective approaches for the fault detection process. In this paper, a deep learning approach based on a combination of temporal convolutions and Long Short Term Memory (LSTM) network is used for fault diagnosis. The two architectures are both shown to be effective for time-series classification and sequence modeling. Temporal convolutions are shown to be competent in feature extraction for time-series classification; however, they are rarely studied in bearing fault detection and classification in an electromechanical system. The presented method does not need any preprocessing or predetermined signal transformation, and uses the raw time-series sensor data. In this regard, three different faults, as inner race, outer race, and balls are considered for validity of the proposed method. The results show that healthy cases can be separated from faulty cases in different load levels with high accuracy (95.8%).

**Review History:**

Received: Aug. 21, 2021
Revised: Nov. 27, 2021
Accepted: Nov. 27, 202
Available Online: Jun. 01, 2022

**Keywords:**

Fault diagnosis
induction motor
ball bearing
condition monitoring
deep learning

## 1- Introduction

Three-phase induction machines have been widely used in industry due to low cost, simplicity, and working in different rotational speeds. It is necessary to note that detecting mechanical and electrical faults in the nascent stage is crucial, especially for high power rating machine, where their services are costly and time-consuming. Moreover, the cost of line production loss also has to be considered in total cost. It has been measured that 40% - 50% of total failures of the squirrel cage induction machine are related to the bearing fault [1]. In other words, bearing faults are one of the major reasons for downtime in rotating machines. The main cause for this rate of failure is related to the mechanical fatigue that occurred due to normal operation. Therefore, detecting bearing fault in the early stage can prevent unexpected downtime of the machine. The bearing fault may be originated from localized faults that occurred in the inner raceway, outer raceway, and ball. Torque oscillation may be caused by passing the ball through the location of fault which depend on rotational speed, diameter of ball, and bearing, as well as the number of balls in the bearing system. The amplitude of faults depends on the type of faults and the defect extent. Thus, diagnosing localized fault in the bearing of the machine can prevent the extension of fault.

Condition monitoring for mechanical and electrical faults in electrical machines can be carried out through electrical and mechanical signatures of machines [2, 3]. Generally, these methods can be classified into two different classes as invasive and non-invasive methods. Since non-invasive methods such as fault detection using stator current do not need any additional external sensors, these methods are categorized as simple and cost-effective for condition monitoring of electrical machines in electromechanical systems. It is worth mentioning that the effects of vibration originated from localized fault in the bearing of the machine are also reflected in the current of the machine. Therefore, Motor Current Signature Analysis (MCSA) can be used for bearing fault detection process of the machine due to reliability and ease of use [4].

Recently, advancements in the technologies of sensing and wireless communication, provide huge amount of data from operational machines. Consequently, data-driven condition monitoring of manufacturing systems, which can detect faults in the mechanical and electrical components of system, is being actively studied [5]. The data is usually high dimensional, and the target systems involve multiple classes of interest. Many existing methods use manual extraction of the discriminative features, which require prior knowledge of signal processing techniques from the time domain, frequency domain, and time–frequency domain, based on electrical and mechanical signatures [6]. Additionally, shallow learning architectures may not lead to detection systems with the desired accuracy.

*Corresponding author's email: vmashayekhi@shahroodut.ac.ir

Deep Learning (DL) is increasingly used for automatic bearing fault diagnosis. Compared to the conventional machine learning methods, DL is superior in terms of feature extraction, diagnosis performances, and transferability [7]. Different deep architectures are used for machine health monitoring, such as Convolutional Neural Networks (CNNs) [8], auto-encoders [9], Deep Belief Networks (DBNs) [10], Recurrent Neural Networks (RNNs) [11], and Generative Adversarial Networks (GANs) [12]. A comparison of different intelligent fault diagnosis systems, including traditional ML algorithms and deep architectures, is performed by Zhao et al. [13]. Despite the extensive literature on application of DL in bearing fault diagnosis, some recent deep learning schemes, which are shown to be effective in time series classification [14, 15] are rarely studied in this context.

CNNs are a popular deep architecture capable of extracting features at different abstraction levels. Multiple studies use CNNs and their variations for the bearing fault diagnosis, commonly using up to four layers of convolution and pooling. Many studies apply different preprocessing steps in the time and frequency domain, and convert the signal to a two-dimensional format before giving it as input to the deep network. Chen et al. preprocess the vibration signals using different statistical measures in the time domain [8]. They also apply FFT to obtain the spectrum which is divided into multiple bands, further calculating the Root Mean Square (RMS) value to maintain the energy shape at the spectrum peaks. The preprocessed signal is then classified using a CNN architecture. Lu et al. adopt a four layer CNN structure for fault classification [16]. Wen et al. convert signals into two-dimensional images and apply CNN based on LeNet-5 for fault diagnosis [17]. Their architecture has two alternating convolutional-pooling layers and two fully-connected layers. Padding is used to adapt the size of features. Zhang et al. transform data into spectrograms and use a deep fully convolutional neural network with four convolution-pooling layer pairs [18]. Zhang et al. use a very deep CNN of 14 layers to perform in noisy environments [19]. Nevertheless, this architecture can increase the risk of overfitting.

Avoiding the conversion of data to a two-dimensional signal can simplify the diagnosis process. Qian et al. use an adaptive overlapping CNN, which directly processes the raw vibration signal and avoids the shift variant property of the signals [20]. Similarly, Eren et al. use an adaptive one-dimensional CNN classifier for bearing fault diagnosis [21]. Although we also use one-dimensional CNNs, we apply them as temporal units in a Temporal Convolutional Network (TCN), which are shown to be superior for time series classification [14].

A Recurrent Neural Network (RNN) is trained considering a recurrent behavior and can capture sequential relationships in the input sequence. As a special architecture of RNNs, Long Short Term Memory (LSTM) networks have shown an outstanding performance in modeling the long-term dependency in data. Abed et al. apply RNN for bearing fault diagnosis [11]. They use discrete wavelet transform to extract features and further reduce them by applying orthogonal fuzzy neighborhood discriminative analysis. Pan et al. combine a one-dimensional CNN, LSTM, and Softmax classifier to classify bearing fault types [22]. They use the raw signal without preprocessing. Yu et al. apply a hierarchical structure based on stacked LSTM to extract features from the raw temporal signals [23].

In this paper, we benefit from both architectures of temporal convolutions and LSTMs in a combined architecture for condition monitoring of bearing fault in induction machine. The use of temporal convolutions in health monitoring of manufacturing systems is rarely studied. The current signals obtained from stator winding of induction machine in different load levels of healthy and faulty conditions are used as system input. In contrast to most previous works, which utilize extra preprocessing steps for feature extraction, we use the raw time-series sensor data as input. The effective features are learned by the deep architecture, and data is classified through a Softmax classifier. Therefore, the proposed system can work without any transformations of the signal or manual feature extraction. In this regard, three different types of faults, commonly occurring in the bearing of the induction machine, are considered for confirmation of the presented model. The goal of the proposed method is to separate healthy and faulty signals under different loads and speeds. The experimental results show more than 95% accuracy in classifying the bearing condition. We evaluate the system with different metrics and also visually demonstrate the capability of the deep network in discriminating different classes in the input data.

## 2- Localized Defect in Roller Bearing

As it is shown in Fig. 1, the roller bearing comprise of four components; inner ring, outer ring, ball, and cage which holds each ball in specific spaces form the others [1].

Mechanical faults on bearing system cause variation on the surface of mechanical components of roller bearing, such as ball, outer and inner ring. The occurred mechanical components generate periodic variation on the vibration signal of motor. The amplitude of fault on motor vibration signals depends on level of load, extension of fault, the type and location of fault. Therefore, all parameters, which effects on the amplitude of fault, need to be considered in order to avoid condition monitoring system from unwanted false alarm. Due to mechanical fault in the motor vibration signal, the location of generated frequencies depends on geometry of bearing, type of fault, and shaft speed. Consequently, each fault can be detected by specific frequency in vibration signal. Vibration frequency generated by the mechanical fault in ball ($f_b$), inner and outer rings of roller bearing ($f_i$ and $f_o$) are calculated as follows [24]:

$$f_b = \frac{D_c}{D_b} f_r \left(1 - \frac{D_b^2}{D_c^2} \cos^2 \theta\right) \quad (1)$$

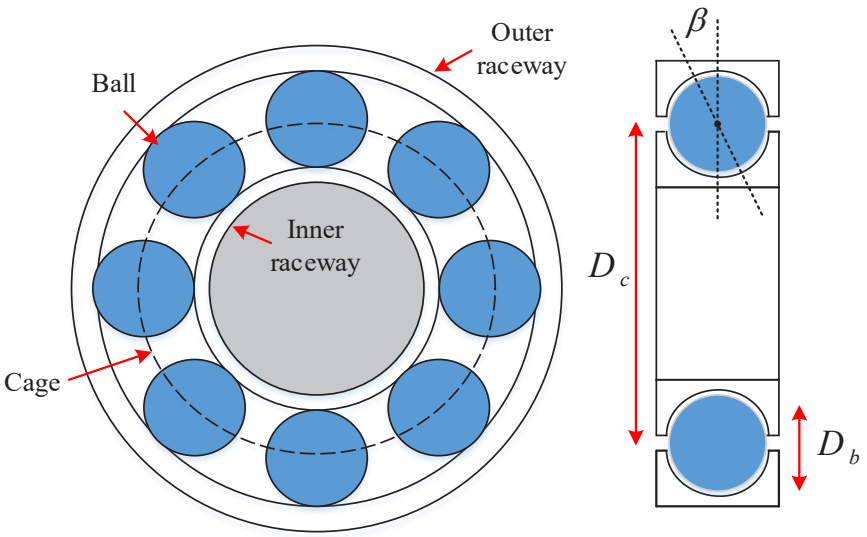


**Fig. 1. Geometry of roller bearing**

$$f_b = \frac{D_c}{D_b} f_r \left(1 - \frac{D_b^2}{D_c^2} \cos^2 \theta\right) \tag{2}$$

$$f_i = \frac{N_b}{2} f_r \left(1 + \frac{D_b}{D_c} \cos \theta\right) \tag{3}$$

Where:

$f_r$: frequency speed of shaft
$D_b$: ball diameter
$D_c$: cage diameter
$D_c$: ball number
$\beta$: contact angle of ball

The vibration generated by roller bearing fault create harmonics components around the fundamental frequency in the line current of motor as follows:

$$f_h = \left| f_s \pm k f_i \right| \qquad k = 1, 2, 3, \ldots \qquad i = \text{o, r, b} \tag{4}$$

Where:

$f_s$: fundamental frequency

Therefore, due to some important characteristic such as simplicity and being cost-effective, electrical signatures have been considered for fault detection in electrical machine.

## 3- Test-rig Description

A three-phase 250W, 380V, 50Hz Squirrel Cage Induction Motor (SCIM) for bearing fault detection and classification is considered. In order to analyse the effects of load and speed on fault identification and classification, digitally brake system is linked to the SCIM through coupling. Consequently, machine can be tested under different load levels (Fig. 2). Bearing faults are created by drilling holes on 6202 healthy bearing (Fig. 3(a), Fig. 3(b)) with diameter of 1mm on inner raceway (Fig. 3(g) and Fig. 3h), outer raceway (Fig. 3(e), Fig. 3(f) and Fig. 3(i)) and on the ball (Fig. 3(c) and Fig. 3(d)).

The stator current of SCIM is measured through three-phase current sensor, and recorded with a sampling frequency of 2.5kHz with time duration of 10S for 50 times in each load. Since load variation affects the amplitude of fault index and leads to incorrect fault detection, the motor is tested in four different slips which cover extent range of motor operation (0.060, 0.047, 0.033 and 0.020). For sake of simplicity, all the collected data are classified in 16 classes in which A11, A21, A31 and A41 are healthy data collected in 0.060, 0.047, 0.033 and 0.020 slips, respectively. Similarly, other classes are named as shown in Table 1.

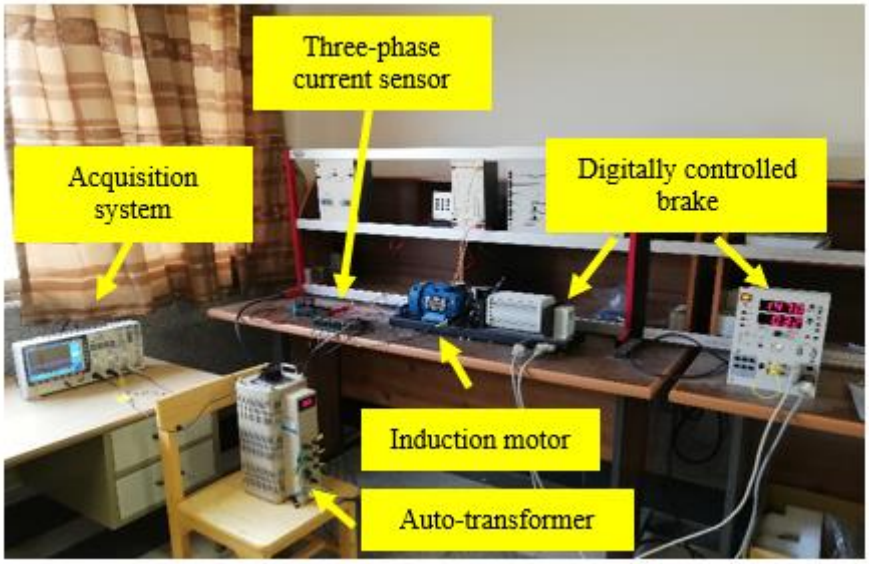


**Fig. 2. Bearing fault test bench**

**Table 1. Classification of collected data.**

| Slip | Healthy | Inner raceway fault | Outer raceway fault | Ball fault |
|---|---|---|---|---|
| 0.060 | $A_{11}$ | $A_{12}$ | $A_{13}$ | $A_{14}$ |
| 0.047 | $A_{21}$ | $A_{22}$ | $A_{23}$ | $A_{24}$ |
| 0.033 | $A_{31}$ | $A_{32}$ | $A_{33}$ | $A_{34}$ |
| 0.020 | $A_{41}$ | $A_{42}$ | $A_{43}$ | $A_{44}$ |

## 4- Computational Model

The process of the bearing fault detection is shown in Fig. 4. From the designed test rig, the training and test datasets are collected. The training dataset is used to build the classification model, which is then used to predict the class label of the test data. In this section, we first describe the building blocks of the adopted architecture, then discuss the architecture in detail. Induction machine fault detection can be considered as a sequence classification task. Such task is a predictive modeling problem where given a sequence of inputs over time; a category should be predicted for the sequence. This task is challenging as the sequences may vary in length and may require learning the dependencies among the input elements. Therefore, we consider two deep models which are suitable for such a modeling context, namely temporal convolutions and recurrent networks. We apply them in a conjoined manner for sequence modeling. This combined model is shown to be effective in time series classification [22].

### 4- 1- Preliminaries

Convolutional Networks have shown to be effective for feature extraction from structural data through different abstraction levels. The Temporal Convolutional Network (TCN) is recently used for the sequence modelling tasks [15]. TCN is a framework employing casual convolutions and dilations suitable in modeling sequential data. TCNs have two main principles. First, the output has an identical length to the input (or shorter if desired), and second, there is no information exposure from future to past. TCN uses a dimensional Fully Convolutional Network (FCN) architecture, and casual convolutions to satisfy the second principle.

RNNs [25] are high-level temporal classifiers, which are able to handle variable-length sequence inputs. They are capable of detecting modeling dependencies over different time scales in sequence of data. As electrical signatures have temporal inter-dependence, RNNs are natural candidates to model these sequences. An RNN computes the output de-

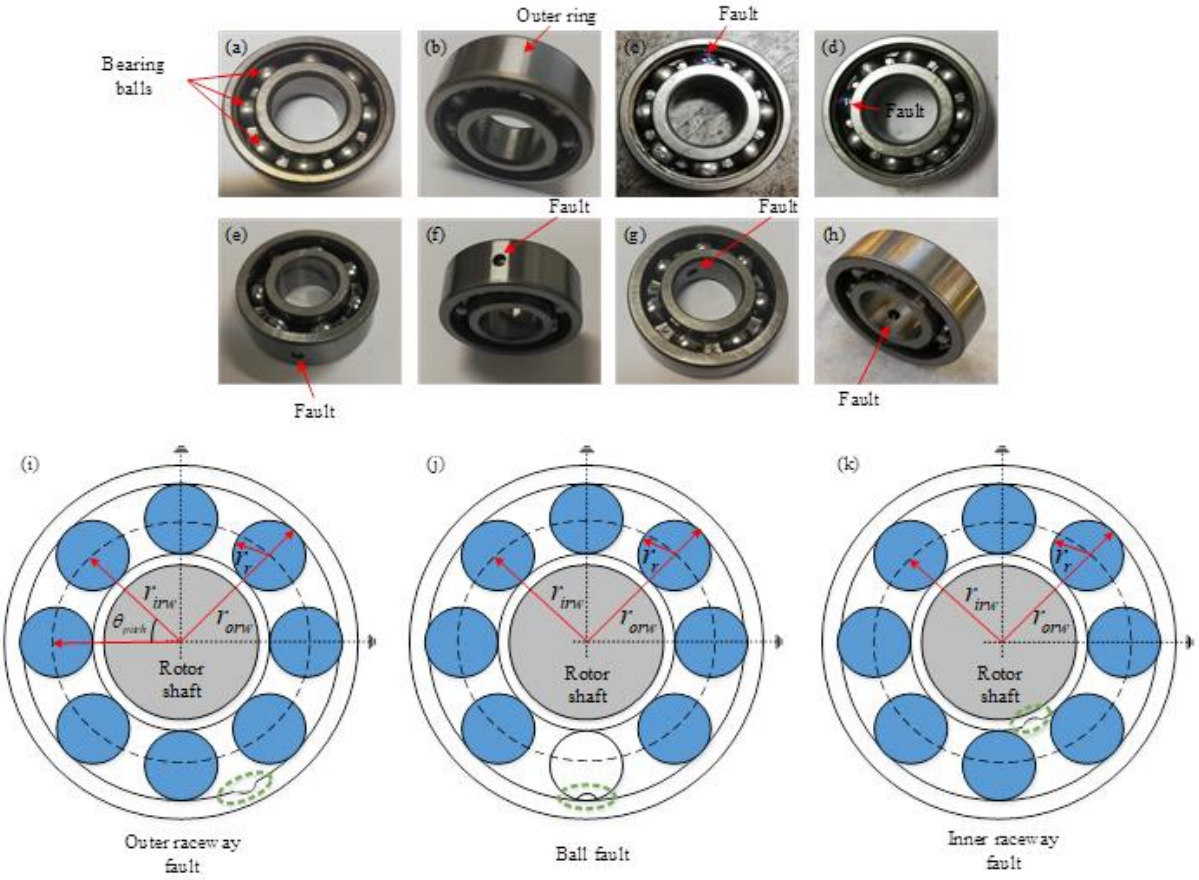


**Fig. 3. Photograph and schematic of bearing in healthy (a, b) and faulty conditions with hole on inner raceway (g, h, k), outer raceway (e, f, i) and ball (c, d, j).**

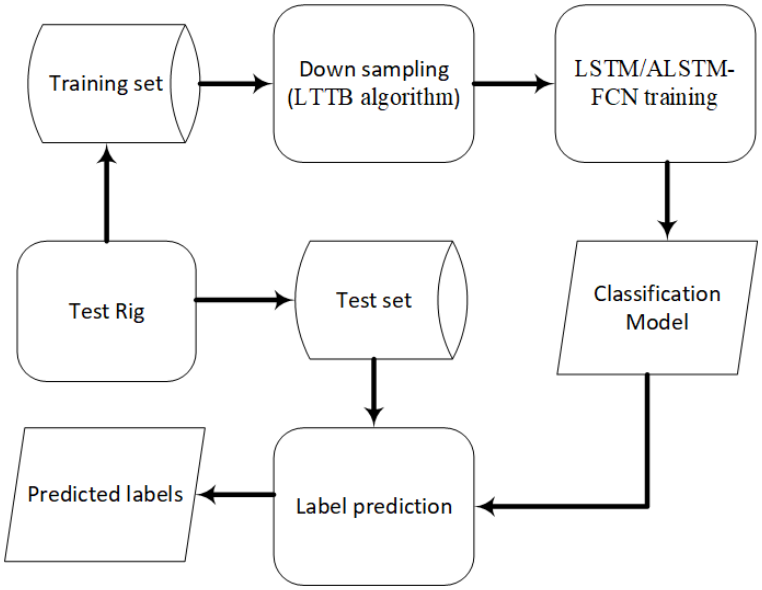


**Fig. 4. The process of bearing fault classification**

pending of the current state on the outputs of the previous states. Therefore, the network outcome is influenced by what it has learnt from the past. In recent years, a main variant of RNN, named Long Short-Term Memory (LSTM) [26], has been developed and increasingly applied in different applications. LSTM is an RNN variation, which can model hidden temporal states by internal gating mechanisms and addresses the vanishing gradient problem in RNNs. LSTM networks excel at classifying and processing time series data, as such data may contain variable-length dependencies.

In an RNN, the input and hidden states are simply passed through a single tanh layer. LSTM networks improve on this simple transformation and introduce additional gates and a cell state. The cell state is used to remember values over arbitrary time intervals. The gates can adjust the flow of information in and out of the cell. The forget gate controls persistence of a value in the cell, the input gate restrains entrance of a new value into the cell, and the output gate determines how much effect the cell value has on the cell output.

The hidden layer in an LSTM network consists of a set of recurrently connected units. At each time t, the vector $x_t$ is given as input to the network. The equations (5-7) define the forget, input, and output gates of a block in the hidden layer, respectively. In an LSTM network with multiple hidden layers, the input vector $x_t$ is the output of the previous layer. $C_t$, defined by equation (9), is the cell state at time t, and is computed from the previous cell state and the block input ($\tilde{C}_t$) at time t. $\tilde{C}_t$ is a tanh layer depending on the input (or the output of the previous layers), and the output of the previous time step. $h_t$, defined by equation (10), is the block output at time t.

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{5}$$

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{6}$$

$$o_t = \sigma\left(W_o x_t + U_o h_{t-1} + b_o\right) \tag{7}$$

$$\tilde{C}_t = tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \tag{8}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{9}$$

$$h_t = o_t \odot \tanh\left(C_t\right) \tag{10}$$

$W$ and $U$ are weight matrices, and b is the bias vector. Point-wise multiplication is shown by $\odot$ sign, and $\sigma$ is point-wise non-linear logistic sigmoid function.

The Attention mechanism can also be applied to the LSTM architecture. Although LSTMs can learn long-range dependencies, their performance may degrade when the length of the input sequence increases, as long-range dependencies may not be learned properly. The Attention mechanism allows to give relative importance to elements in the input sequence. Thus, along using all the input, the network may focus on certain parts of the input sequence when trying to predict particular elements of the output sequence. This mechanism, which is now widely used in different context, was initially designed for the sequence to sequence models. It can lift the limitation of fixed-length internal representation [27], and help improve the network performance for sequences of longer lengths.

### 4- 2- Network Architecture

The architecture is composed of two parts: LSTM block and FCN block. The LSTM network as previously discussed, is suitable for sequence classification. This block also contains a dimension shuffle before the LSTM network. As LSTM networks commonly have the problem of overfitting, a dropout layer is used to mitigate this problem.

The FCN block consists of three convolutional layers as shown in Fig.5. LSTM-FCNs are able to augment FCN models, appreciably increasing their performance with a nominal increase in the number of parameters. LSTM cells can be replaced by Attention LSTM cells to construct the ALSTM-FCN.

The fully convolutional block views a time series of length N as a univariate time series with N time steps. In the LSTM block, on the other hand, the dimension shuffle layer transposes the input temporal dimension that an input univariate time series of length N is turned into a multivariate time series (with N variables). This transformation improves the performance of LSTM block [14].

The Fully Convolutional Network (FCNs), comprised of three stacked temporal convolutions is shown in Fig. 6. Each convolution element consists of a temporal convolution followed by batch normalization and ReLU activation, adopted from the CNN architecture proposed by Wang et al. [28]. This network can extract efficient features from the input sequence. The dimension is reduced at the end of the block with global average pooling. The output from the LSMT/ALSTM and FCN blocks is passed to a Softmax layer after concatenation. The Softmax layer performs the classification and assigns the signal to one of the 16 classes.

### 4- 3- Input Sequence

As mentioned earlier, we use the raw signal values in the time domain as input sequence to the network. However, reducing the time complexity of training the network is possible by down sampling. For this purpose, we use the Largest Triangle Three Buckets (LTTB) algorithm [29]. This algo-

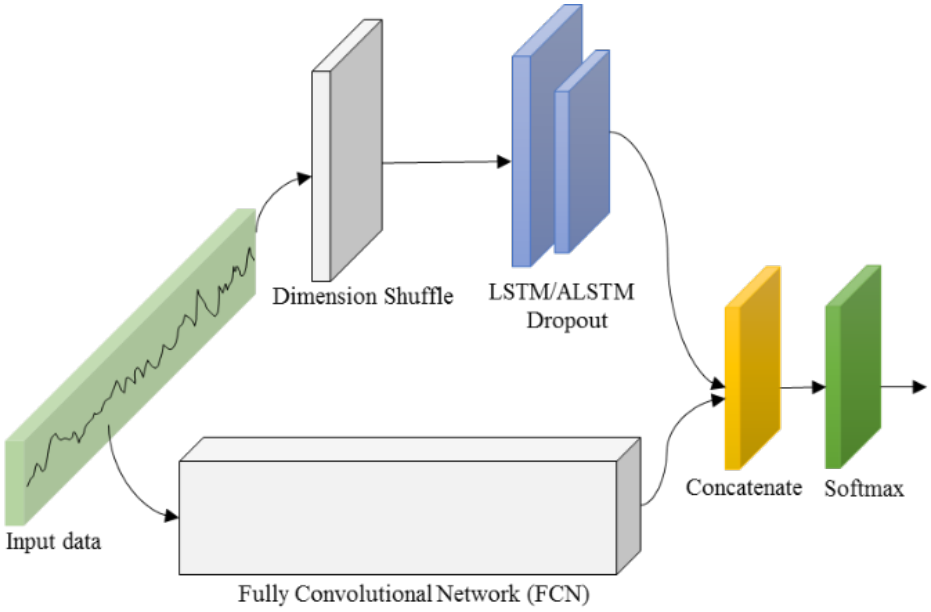


**Fig. 5. LSTM-FCN architecture**

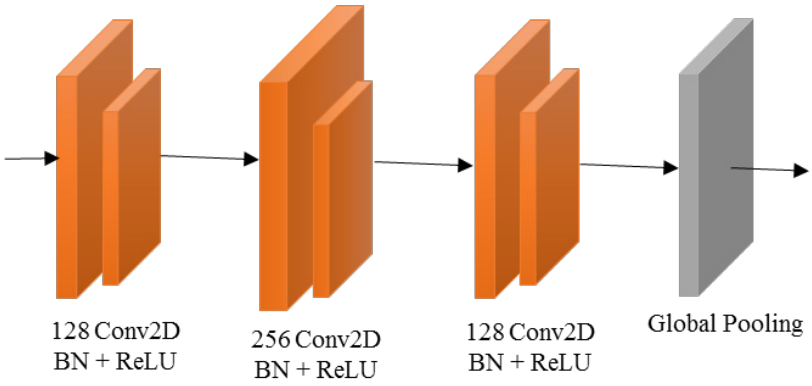


**Fig. 6. The FCN structure of the LSTM/ALSTM-FCN network architecture**

rithm partitions the signal to a number of buckets equal to the desired approximation samples. The first and last buckets contain only the first and last data signal points, while other signal values are divided evenly into the remaining buckets. In each bucket, the data point that forms the largest triangle (in terms of area), with the average point of the next bucket and the point in the last bucket, is selected. The first and last data points along with these selected points constitute the down sampled signal.

## 5- Simulation Results

In this section we first describe the setup of the network and the evaluation measures. The evaluation results come next.

### 5- 1- Simulation Setup

The deep architecture was implemented using the Keras library with TensorFlow [29], and we trained the model using the Adam optimizer with initial and final learning rates of 1e−3 and 1e−4, respectively. The number of training epochs was set to 1500. The initial batch size was set to 128, and a dropout rate of 80% was used. The number of LSTM blocks is set to 64. The dataset is divided to two train (80%) and test (20%) subsets to be used in the experiments.

### 5- 2- Evaluation Measures

We use several metrics to assess the algorithm performance from different perspectives. As we are performing classification to predict the induction machine fault and its severity, we use the well-established measures in this context.

**Table 2. Performance of the models on the test dataset**

| Model | LSTM-FCN | ALSTM-FCN |
|---|---|---|
| Accuracy (%) | 92.00 | **95.80** |

The accuracy measures the percent of correctly classified instances among all classes. In a two-class scenario, the target class can be seen as positive and the other as negative. After classifying the instances, the number of positive/negative samples correctly classified, are referred to as True Positive (TP), and True Negative (TN) respectively. The number of positive/negative samples incorrectly classified, are referred to as False Negative (FN), and False Positive (FP) respectively. In the case of more than two classes, a one-vs-all scheme may be used, where the class of interest is considered positive and the other classes are negative.

The precision and recall metrics as defined by the equations (11) and (12), measure the proportion of correct identifications, and the proportion of actual positive samples which were identified, respectively.

$$\text{Precision} = \frac{TP}{\text{TP} + \text{FP}} \tag{11}$$

$$\text{Recall} = \frac{TP}{\text{TP} + \text{FN}} \tag{12}$$

Higher precision and recall values indicate better performance of the algorithm in terms of correct classifications. The F1-score combines these two measures to obtain a single measure of performance.

$$\text{F1-Score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{13}$$

An AUC-ROC curve demonstrates the capability of the model in discriminating the classes. To plot the ROC curve, the discrimination threshold of the classifier is varied and per each threshold value, TPR (True Positive Rate or Sensitivity), and FPR (False Positive Rate or 1- Specificity) are measured and plotted. TPR is similar to recall and FPR is defined in equation (14). AUC is the Area Under the Curve of ROC and its larger values indicate better classification performance.

$$FPR = 1 - Specificity = \frac{FP}{FP + TN} \tag{14}$$

Moreover, as we have a multi-class problem, ROC should be extended to such context. Instead of plotting ROC for each class, the micro and macro averages may be plotted. Macro-averaging computes the TPR and FPR metrics independently for each class and takes the average by giving equal weight to all classes. In the micro-average method, the individual true positive, false positive, and false negative values are computed for each class and aggregated to get the statistics [21].

### 5- 3- Experiments

We first train the network and compare the two models, LSTM-FCN and ALSTM-FCN in terms of accuracy. As demonstrated from Table 2, both networks achieve high accuracy values with ALSTM-FCN being superior. This result shows that learning-based fault detection of the induction machine is effective.

The precision, recall, and F1-score of LSTM-FCN and ALSTM-FCN models are presented in Tables 3 and 4, respectively. The average F1-score is 0.92 and 0.96 for the two models, respectively. This shows that the model is performing near optimal in discriminating healthy and faulty functions of the machine. This result is further demonstrated by the ROC curves of Figs. 7 and 8. The AUC of both curves is almost one, with ALSTM/FCN performing slightly (1e-4) better. In Fig. 9, the precision, recall and F1-scores are shown for healthy and fault categories averaged over different slip values. As illustrated, the ALSTM-FCN is superior in nearly all cases. However, according to precision, it majorly outperforms the LSTM-FCN in the outer raceway and ball faults, meaning that it has relatively higher true positives or less false positives in these classes. According to recall, it outperforms LSTM-FCN majorly in the healthy and inner raceway fault classes, meaning that it has relatively higher true positives or less false negatives. Combining the two measures, ALSTM-FCN outperforms LSTM-FCN in all categories in a nearly uniform manner. Overall, the results show that both networks have high performance in predicting the machine condition with the ALSTM/FCN network being superior.

**Table 3. Individual precision, recall and f1-score measurements of 16 classes for LSTM-FCN network.**

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| H_1410 | 0.91 | 0.88 | 0.89 |
| H_1430 | 0.93 | 0.93 | 0.93 |
| H_1450 | 0.91 | 0.91 | 0.91 |
| H_1470 | 0.97 | 0.71 | 0.82 |
| F_OR_1410 | 0.95 | 0.95 | 0.95 |
| F_OR_1430 | 0.88 | 0.92 | 0.90 |
| F_OR_1450 | 0.90 | 0.97 | 0.93 |
| F_OR_1470 | 0.90 | 0.96 | 0.93 |
| F_IR_1410 | 1.00 | 1.00 | 1.00 |
| F_IR_1430 | 1.00 | 0.95 | 0.98 |
| F_IR_1450 | 0.97 | 1.00 | 0.98 |
| F_IR_1470 | 0.90 | 0.96 | 0.93 |
| F_BA_1410 | 0.88 | 0.88 | 0.88 |
| F_BA_1430 | 0.85 | 0.97 | 0.91 |
| F_BA_1450 | 0.76 | 0.97 | 0.85 |
| F_BA_1470 | 0.91 | 0.96 | 0.93 |
| | | **Accuracy** | **92.00%** |

**Table 4. Individual precision, recall and f1-score measurements of 16 classes for ALSTM-FCN network.**

| Class | Precision | Recall | F1-Score |
|---|---|---|---|
| H_1410 | 0.94 | 0.94 | 0.94 |
| H_1430 | 0.96 | 0.93 | 0.95 |
| H_1450 | 0.96 | 0.96 | 0.96 |
| H_1470 | 0.97 | 0.94 | 0.95 |
| F_OR_1410 | 0.95 | 1.00 | 0.97 |
| F_OR_1430 | 0.96 | 1.00 | 0.98 |
| F_OR_1450 | 1.00 | 0.89 | 0.94 |
| F_OR_1470 | 0.97 | 0.97 | 0.95 |
| F_IR_1410 | 0.89 | 1.00 | 1.00 |
| F_IR_1430 | 1.00 | 1.00 | 1.00 |
| F_IR_1450 | 1.00 | 1.00 | 1.00 |
| F_IR_1470 | 1.00 | 1.00 | 1.00 |
| F_BA_1410 | 0.97 | 0.97 | 0.97 |
| F_BA_1430 | 0.92 | 0.88 | 0.90 |
| F_BA_1450 | 0.91 | 0.94 | 0.93 |
| F_BA_1470 | 0.89 | 0.97 | 0.93 |
| | | **Accuracy** | **95.80%** |

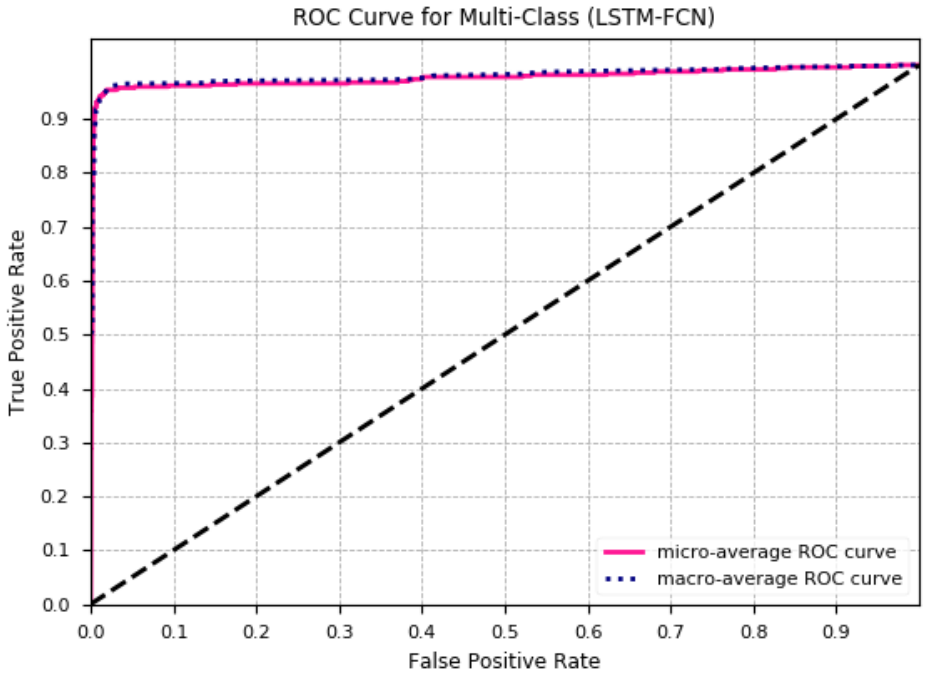


**Fig. 7. ROC curve for Multi-Class LSTM-FCN**

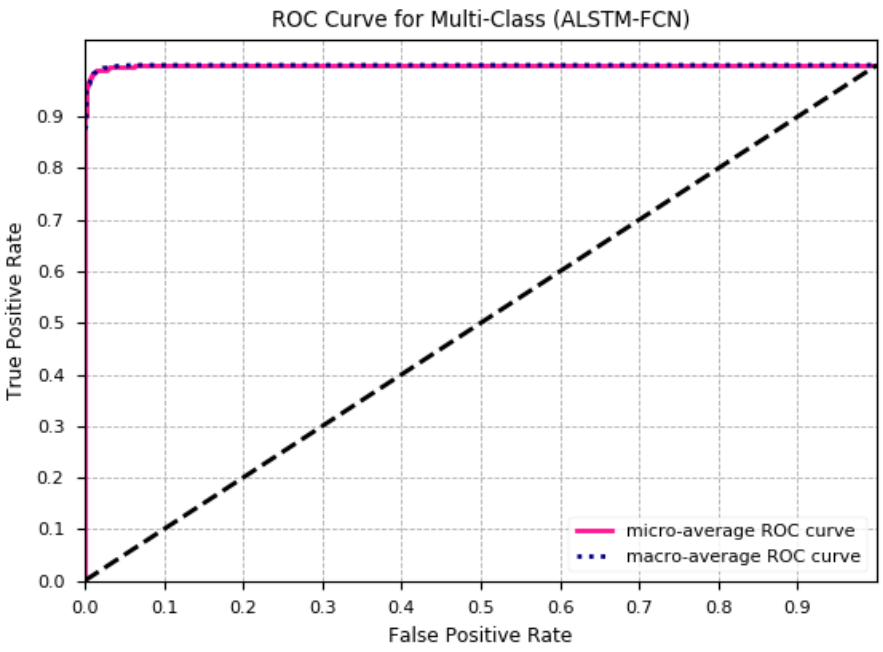


**Fig. 8. ROC curve for Multi-Class ALSTM-FCN**

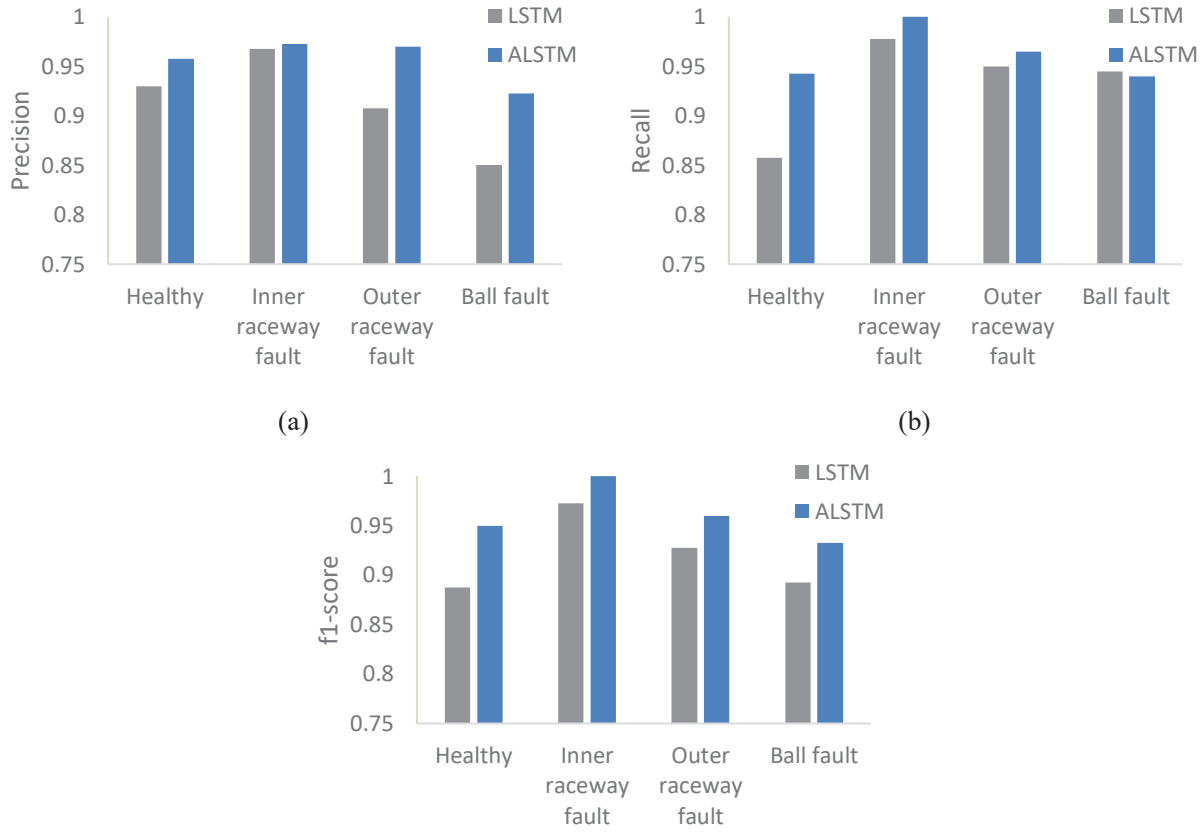


**Fig. 9. Average (a) precision, (b) recall and (c) f1-score measurements of healthy and different fault categories for LSTM/ALSTM-FCN networks**

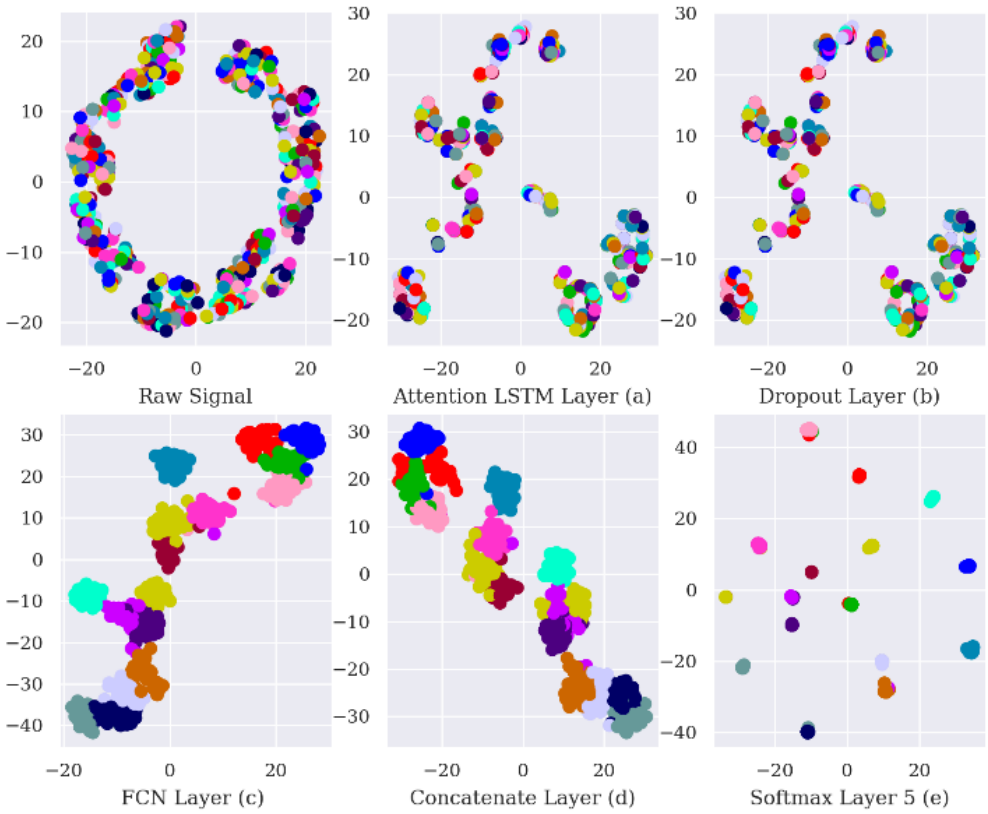


**Fig. 10. Feature visualization via t-SNE for the test dataset: t-SNE for (a,b) the ALSTM structure of the ALSTM-FCN architecture, (c,d) the FCN structure, and (e) the Softmax layer**

To visually inspect the classes in the data and performance of network in discriminating different conditions, the t–SNE (t- Distributed Stochastic Neighbor Embedding) can be used. This algorithm provides a means of visualizing complex high-dimensional datasets [32]. Hidden patterns in the data, sketched in two- or three-dimensional spaces based on this algorithm, can be inspected. It transposes similarities between data points to joint probabilities, and tends to minimize the Kullback–Leibler divergence between the joint probabilities of low dimensional embedding and high dimensional data. Fig. 10 shows t-SNE for the ALTSM-FCN network. As illustrated, the network is successful in discriminating data of different classes into distinguished groups.

## 6- Conclusion

In this paper, a new approach of bearing fault detection and classification for induction machines is proposed based on the stator current signature. In this regard, two deep architectures which are both shown to be effective in classifying sequences are jointly used in a combined architecture. Temporal convolutions use 1D CNNs in a temporal manner and are effective in extracting features from time-series data. The LSTM network can detect dependencies in the input sequence elements. The combined architecture along with the Softmax classifier is effective for bearing fault diagnosis as shown in the experiments. In comparison to the traditional fault detection approaches, the proposed method minimizes the need for expert supervision, manual feature extraction, and predetermined signal transformations. The proposed method does not need any pre-processing phase and directly uses raw time-series of electrical signals for fault detection and classification. The results show that the faulty and healthy cases in different load levels can be separated with high accuracy of 95.8%.

## References

[1] M. Ojaghi, M. Sabouri and J. Faiz, “Analytic Model for Induction Motors Under Localized Bearing Faults,” IEEE Transactions on Energy Conversion, vol. 33, no. 2, pp. 617-626, June 2018.

[2] Xin Zhang, Zhiwen Liu, Jiaxu Wang, Jinglin Wang, “Time–frequency analysis for bearing fault diagnosis using multiple Q-factor Gabor wavelets,” ISA Transactions, Volume 87, pp. 225-234, 2019,

[3] Zhibin Zhao, Tianfu Li, Jingyao Wu, Chuang Sun, Shibin Wang, Ruqiang Yan, Xuefeng Chen, “Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study,” ISA Transactions, Volume 107, pp. 224-255, 2020.

[4] M. Yang, N. Chai, Z. Liu, B. Ren and D. Xu, “Motor

Speed Signature Analysis for Local Bearing Fault Detection With Noise Cancellation Based on Improved Drive Algorithm," IEEE Transactions on Industrial Electronics, vol. 67, no. 5, pp. 4172-4182, May 2020.

[5] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen and J. Wang, "Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks," IEEE Transactions on Industrial Electronics, vol. 65, no. 2, pp. 1539-1548, Feb. 2018.

[6] W. You, C. Shen, D. Wang, L. Chen, X. Jiang and Z. Zhu, "An Intelligent Deep Feature Learning Method With Improved Activation Functions for Machine Fault Diagnosis," IEEE Access, vol. 8, pp. 1975-1985, 2020.

[7] S. Zhang, S. Zhang, B. Wang, T.G. Habetler. "Deep Learning Algorithms for Bearing Fault Diagnosticsx—A Comprehensive Review," IEEE Access, vol. 8, pp.29857-29881, Feb. 2020.

[8] Z. Chen, C. Li, R.V. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," Shock and Vibration, 2015.

[9] S. Tao, T. Zhang, J. Yang, X. Wang, W. Lu, "Bearing Fault Diagnosis Method bBased on Stacked Autoencoder and Softmax Regression," 34th Chinese Control Conference (CCC), IEEE, pp. 6331-6335, Jul. 2015.

[10] Z. Chen, W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 7, pp. 1693-1702, Mar. 2017.

[11] W. Abed, S. Sharma, R. Sutton, A. Motwani, "A Robust Bearing Fault Detection and Diagnosis Technique for Brushless DC Motors Under non-Stationary Operating Conditions," Journal of Control, Automation and Electrical Systems, vol. 26, no. 3, pp. 241-254, Jun. 2015.

[12] W. Mao, Y. Liu, L. Ding, y. Li, "Imbalanced Fault Diagnosis of Rolling Bearing Based on Generative Adversarial Network: A Comparative Study," IEEE Access, vol. 7, pp. 9515-9530, Jan. 2019.

[13] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, "Deep Learning and its Applications to Machine Health Monitoring," Mechanical Systems and Signal Processing, pp.213-237. Jan. 2019.

[14] F. Karim, S. Majumdar, H. Darabi and S. Chen, "LSTM Fully Convolutional Networks for Time Series Classification," IEEE Access, vol. 6, pp. pp. 1662-1669, 2018.

[15] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, "Temporal Convolutional Networks: A Unified Approach Tto Action Segmentation," European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, pp. 47–54, Oct. 2016.

[16] C. Lu, Z. Wang, B. Zhou, " Intelligent Fault Diagnosis of Rolling Bearing using Hierarchical Convolutional Network Based Health State Classification," Advanced Engineering Informatics, pp. 139-151, Apr. 2017.

[17] L. Wen, X. Li, L. Gao, Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," IEEE Transactions on Industrial Electronics. vol. 65, no. 7, pp. 5990-5998, Nov. 2017.

[18] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, "A Deep Convolutional Neural Network with New Training Methods For Bearing Fault Diagnosis Under Noisy Environment and Different Working Load," Mechanical Systems and Signal Processing, pp. 439-453, Feb. 2018.

[19] W. Zhang, F. Zhang, W. Chen, Y. Jiang, D. Song, "Fault State Recognition of Rolling Bearing Based Fully Convolutional Network," Computing in Science & Engineering. Vol. 21, no. 5, pp. 55-63, Jan. 2018.

[20] W. Qian, S. Li, J. Wang, Z. An, X. Jiang, "An Intelligent Fault Diagnosis Framework for Raw Vibration Signals: Adaptive Overlapping Convolutional Neural Network," Measurement Science and Technology, vol. 29, no. 9, p. 095009, Aug. 2018.

[21] L. Eren, T. Ince, S. Kiranyaz, "A Generic Intelligent Bearing Fault Diagnosis System Uusing Compact Adaptive 1D CNN Classifier," Journal of Signal Processing Systems, vol. 91, no. 2, pp. 179-189, Feb. 2019.

[22] H. Pan, X. He, S. Tang, F. Meng, "An Improved Bearing Fault Diagnosis Method using One-Dimensional CNN and LSTM," J. Mech. Eng. vol. 64, no. 7-8, pp. 443-452 May. 2018.

[23] L. Yu, J. Qu, F. Gao, Y. Tian, "A Novel Hierarchical Algorithm for Bearing Fault Diagnosis Based on Stacked LSTM," Shock and Vibration, 2019.

[24] M. Blodt, P. Granjon, B. Raison, and G. Rostaing, "Models for Bearing Damage Detection in Induction Motors Using Stator Current Monitoring," IEEE Transaction on Industrial Electronics, vol. 55, no. 4, pp. 1813–1822, Apr. 2008.

[25] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent Neural Network Regularization," arXiv preprint arXiv:1409.2329, 2014.

[26] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[27] D. Bahdanau, K. Cho, and Y. Bengio, ''Neural Machine Translation by Jointly Learning to Align and Translate,'' arXiv preprint arXiv: 1409.0473, 2019. [Online]. Available: https://arxiv.org/abs/1409.0473.

[28] Z. Wang, W. Yan and T. Oates, "Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, pp. 1578-1585, 2017.

[29] S. Steinarsson, "Downsampling Time Series for Visual Representation," PhD diss., 2013.

[30] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, ... & M. Kudlur, "TensorFlow: A System for Large-Scale Machine Learning," in Proc. OSDI, Savannah, GA, USA, pp. 265-284, Nov. 2016.

[31] D. J. Hand and R. J. Till, "A Simple Generalisation of The Area Under The ROC Curve For Multiple Class Classification Problems," Machine Learning, vol. 45, no. 2, pp. 171–186, 2001.

[32] L. V. D. Maaten and G. Hinton, "Visualizing Data using t-SNE," Journal of Machine Learning Research , vol. 9, pp. 2579-2605, Nov. 2008.

**HOW TO CITE THIS ARTICLE**

*M. Hoseintabar Marzebali, S. Hasani Borzadaran, H. Mashayekhi, V. Mashayekhi, Bearing Fault Detection and Classification Based on Temporal Convolutions and LSTM Network in Induction Machine, AUT J. Elec. Eng., 54(1) (2022) 107-120.*

**DOI:** 110.22060/eej.2021.20438.5430