A Hybrid 1D-CNN-Bi-LSTM based Model with Spatial Dropout for Multiple Fault Diagnosis of Roller Bearing

Gangavva Choudakkanavar , J. Alamelu Mangai

Department of Computer Science and Engineering
Presidency University, Bangalore, India

Abstract—**Fault diagnosis of roller bearings is a crucial and challenging task to ensure the smooth functioning of modern industrial machinery under varying load conditions. Traditional fault diagnosis methods involve preprocessing of the vibration signals and manual feature extraction. This requires domain expertise and experience in extracting relevant features to accurately detect the fault. Hence, it is of great significance to implement an intelligent fault diagnosis method that involves appropriate automatic feature learning and fault identification. Recent research has shown that deep learning is an effective technique for fault diagnosis. In this paper, a hybrid model based on 1D-CNN (One-Dimensional Convolution Neural Networks) with Bi-LSTM (Bi-directional Long-Short Term Memory) is proposed to classify 12 different fault types. Firstly, vibration signals are given as input to 1D-CNN to extract intrinsic features from the input signals. Then, the extracted features are fed into a Bi-LSTM model to identify the faults. The performance of the proposed method is enhanced by applying Softsign activation function in the Bi-LSTM layer and Spatial Dropout in the neural network. To analyze the effectiveness of the proposed method, Case Western Reserve University (CWRU) bearing data is considered for experimentation. The results demonstrated that the proposed model has attained an accuracy of 99.84% in classifying the various faults. The superiority of the proposed method is verified by comparing the predictive accuracy of the proposed method with the existing fault diagnosis methods.**

Keywords—Fault diagnosis; roller bearing; deep learning; 1D-CNN; Bi-LSTM; spatial dropout

I. INTRODUCTION

Roller Bearing (RB) is a key component of any rotating machinery where rotation is involved. It is widely used in various industries such as transportation, agriculture, aerospace, medical domain and so on. RB is more susceptible to damage due to its continuous rotation with varying loads and pressure. Due to which there's a break-down of the entire machine which results in magnificent economic loss and severe safety accidents [1]. Therefore, it is very much essential to diagnose the roller bearing fault accurately because each fault type exhibits distinct characteristics and the fault may exist in any of the components such as Inner Race (IR), Outer Race (OR) and Ball.

Traditional vibration-based bearing fault diagnosis methods involved mainly three steps as data pre-processing, feature extraction, and fault classification. The vibration signals collected from sensors represents the information about bearing condition. In order to classify and detect the faults, many signal processing techniques have been discussed through analysis of signal characteristics in various domains such as time, frequency and time-frequency domain [2]. Due to the non-stationary nature of vibration signal, various feature extraction techniques such as Short-Time Fourier Transform (STFT), Wavelet Analysis (WA), Empirical Mode Decomposition (EMD), etc. were applied to extract the features [3]. Once the features are extracted and selected then those features are fed into the network model for classification.

Recently, Deep Learning (DL) technology has gained more importance in various domains such as image processing, natural language processing, speech recognition and so on. It uses multiple layers of the network to learn and extract relevant features from raw data and identifies the pattern for classification or recognition problems. Roller bearing's vibration data has similar dimensionality as that of image or speech. Hence, DL architecture can be used to diagnose roller bearing fault by transforming vibration signal into the framework of pattern recognition problem. DL model has an ability in automatic feature learning and classification that involves automatic feature extraction and identification of the faults accurately [4-5].

In this research, a hybrid method based on 1D-CNN-Bi-LSTM with Spatial Dropout is proposed for multiple fault diagnosis of roller bearing. Initially, one-dimensional raw vibration signal is collected and input into CNN model. Then, CNN extracts feature information from the signals and these extracted features are provided to Bi-LSTM network model to acquire the failure information to identify 12 types of bearing faults. For experimentation, CWRU dataset is being used to analyze the effectiveness of the method.

The rest of this paper is organized as follows: In Section II related work is discussed; Section III describes proposed methodology architecture which includes one-dimensional CNN and Bi-LSTM models. Section IV illustrates an experimental setup of bearing data collection; and Section V shows the discussion of results and its analysis.

II. RELATED WORK

Many researchers have applied various deep learning models for fault diagnosis such as Deep Neural Networks [6], Long Short-Term Memory [7], Deep Belief Networks [8], Deep Auto-encoders [9], Gated Recurrent Unit Networks [10],

and Convolutional Neural Networks [11] and so on. Among these, CNN has received more importance in the study of roller bearing defect diagnosis. Abed et al. [12] proposed a robust approach for fault diagnosis of Brushless DC Motors through feature extraction and reduction using discrete wavelet transform (DWT) and orthogonal fuzzy neighborhood discriminant analysis (OFNDA) from vibration and current signals and RNN model was used for classification of faults. P. Zou et al. [13] focused on empirical mode decomposition (EMD) method which was combined with LSTM to obtain kurtosis value by extracting intrinsic mode functions (IMF) components and long-term dependencies from vibration signals to monitor the health status of an electrical machine. Cao, Lixiao et al. [14] constructed a fault diagnosis framework by extracting ten time-domain statistical features from vibration signals under varying load conditions and these features were fed into deep Bi-directional LSTM to identify the faults of Wind Turbine Gearbox. In [15], Mel Frequency Cepstral Coefficient features are obtained from vibration signals and given these features as input to Random Forest and eXtreme Gradient Boosting algorithms for diagnosis of roller bearing fault.

Zheng Wang et al. [16] discussed an architecture to obtain unsupervised H-statistic value from sensor time-series data based on deep LSTM and CNN for performance degradation valuation of roller bearing. Shichao and Haibin proposed a bearing fault diagnosis model in which 1D-CNN with LSTM is implemented, which adaptively extracted potential features from the original vibration signal and ensured the validity of the features through merging of pooling layers of max and average values to down sample the features. Then, LSTM was employed to acquire the dependencies among features of time-domain signals to perform fault classification [17]. Zhe Yuan et al. [18] presented a fault recognition approach for roller bearing using Multiscale CNN and Gated Recurrent Unit Network (GRUN) by providing multiple time scaled vibration data into the CNN to train the model and added the gated recurrent unit network to make the model predictive with an attention mechanism. In [19], the proposed adaptive anti-noise neural network architecture employed random sampling approach and boosted CNN with the exponential linear activation function to enhance the adaptability of the network without manual feature selection. GRUN was implemented to learn the features processed by CNN and classify the faults. This approach solved the problem of bearing fault diagnosis under changing load conditions and heavy noise.

Wenbing Yu et al. [20] discussed an intelligent fault diagnosis method for identifying ten different bearing faults based on lightweight MobileNet CNN by considering Western Reserve University dataset for evaluating the model and also computed average precision, recall and F1 score which resulted into 96%, 82% and 88%, respectively. Kai Gu et al. [21] discussed a novel diagnostic method to accurately identify the fault status of bearing based on LSTM and DWT for multi-sensors by obtaining fault details in both frequency and time scales through DWT and LSTM algorithm was used to characterize the long-term dependency information hidden in the time series data of a signal. In [22], a combined wavelet regional correlation threshold denoising (WRCTD) algorithm

with CNN-LSTM was proposed for fault detection. WRCTD algorithm utilized the regional association of the wavelet decomposition coefficients and 3σ criterion to reduce noise in the raw sensor data and CNN-LSTM model reduced the hidden features of the pre-processed signal data to identify the fault type of the harmonic reducer under multiple working conditions. A novel fault diagnosis method was presented through application of sliding window processing to integrate the feature and time delay information from multivariate time series samples and then, the samples obtained were fed into the CNN-LSTM model to perform feature learning and capture time delay information to diagnose the fault of Tennessee Eastman chemical process [23]. A robust approach was proposed in [24] to predict the Remaining Useful Life (RUL) of roller bearing with combination of Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and LSTM to detect the damage state and identify the abnormal state of bearing to estimate the RUL through feature extraction from signals. A new convolution-based bidirectional long and short-term memory network method was proposed to predict RUL, in which CNN was used to obtain feature information and BI-LSTM to acquire time-frequency information from the signals to construct health indicators (HI) and the experiments conducted on the PRONOSTIA bearing dataset showed that the proposed method performed better compared to other methods [25].

This work uses deep learning technique for fault diagnosis and it's motivated by the fact that deep learning involve automatic feature extraction whereas Machine learning needs manual feature extraction, in which prior domain knowledge and expertise is required.

III. PROPOSED METHODOLOGY

In this research, a multi-class fault diagnosis method is proposed based on 1D-CNN with Bi-LSTM to classify various faults. The advantage of CNN lies in automatic feature extraction and Bi-LSTM in handling gradient loss and explosion. The main goal is to diagnose the 12 different fault types using 1D-CNN with Bi-LSTM model by collecting vibration signals from CWRU dataset which contains set of ball bearings having localized faults. The proposed 1D-CNN-Bi-LSTM model consists of four convolutional and pooling layers, a Bi-LSTM layer, a LSTM layer and one fully connected layer as shown in Fig. 1.
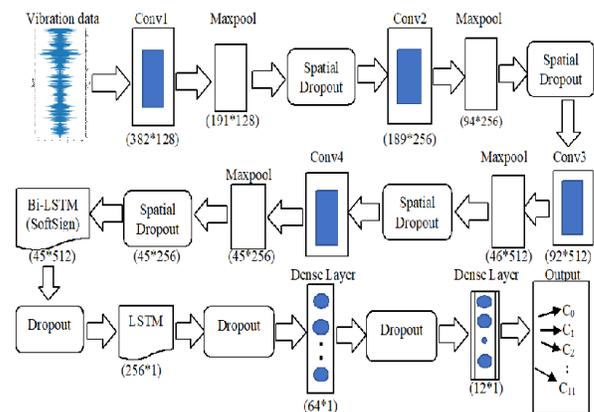
Fig. 1.   Block Diagram of the Multi-class Fault Diagnosis Method.

Firstly, the raw input signals are input to the model, then convolution layers and pooling layers helps in automatic feature extraction. Next, these features are passed to Bi-LSTM and LSTM layers to highlight the features and finally, dense layer is used to classify the various faults. Bi-LSTM layer uses softsign activation function to improve the performance of the model. For verifying the effective performance of the proposed method, publicly available bearing dataset is considered [26]. The vibration data is measured for four operational conditions such as:

1) Normal Bearing- No fault, sampling frequency of 12kHz with 1797rpm (rotations/min).

2) Fault in Outer race - Sampling frequency of 12kHz with 1797rpm.

3) Fault in Inner race - Sampling frequency of 12kHz with

4) 1797rpm.

5) Ball fault- Sampling frequency of 12kHz with 1797rpm.

Details of each layer used in the proposed method is explained in the following subsections.

A. 1D-CNN

CNN is a deep learning algorithm which was originally proposed for processing of visual data. It is more effective in identifying image patterns in a stratified way from simple to complex features because of the two important properties such as weight sharing and spatial pooling. CNN consists of 3 layers namely convolutional layer, pooling layer, and fully connected layer. The convolution layer converts the input data into smaller feature maps through convolutional kernels by performing a summation of multiplications between the vectors of input data and weight coefficients [27]. In this paper, 1D-CNN is constructed, whose convolutional kernels and feature maps are all one-dimensional because of the one-dimensional characteristics of mechanical vibration signals.

Suppose 'x' is an input to 1D-CNN, then the output of the convolutional layer is computed as given in (1):

$$y_{i, j, k} = f\left(\sum_{i=1}^{m} x_{i, k} * w_{j, i} + b_i\right) \tag{1}$$

In equation (1), 'f' represents an activation function, which is typically a hyperbolic tangent, ReLu (Rectified Linear Unit), or sigmoid function; 'm' is number of samples ($1 \leq i \leq m$); 'p' is length of the convolutional kernels ($1 \leq j \leq p$); 'n' is length of the input data ($1 \leq k \leq n$); * represents convolution operation; $w_{j,i}$ is the weight and b_i is the bias.

The pooling layer is the sub-sampling layer to compress the size of feature maps. Down sampling is performed to minimize the dimensionality of the output from the previous convolution layer by moving the filter window from starting point to the end of feature map. Then a maximum or average of each part of the feature map is considered to represent each corresponding area. The role of pooling layer is to reduce the number of parameters and the computation in the network, so that it prevents overfitting and improves the generalization ability of the model. Max pooling is frequently used in the pooling layer which is computed as maximum of the previous feature maps. It is expressed as given in (2).

$$z_{i, j, k} = \max\left(x_{2i-1, j, k}, x_{2i, j, k}\right) \tag{2}$$

where, $1 \leq l \leq m/2$

Fully connected or Dense layer plays the role of classifier in CNN. For a multi-class classification problem, usually softmax is applied in the dense layer to ensure the range of output value lies between 0 and 1, and sum equals to 1. The predicted output represents the probability and value with the highest probability is considered as the final predicted result. The output of 1D-CNN is an input to the Bi-LSTM model to reduce variance in time series.

B. Batch Normalization

It is a regularization technique, which avoids model overfitting. In the training process of the deep neural networks, the distribution of inputs to the layers deep in the network which keeps changing for each mini batch as the weights are updated. This problem is known as "internal covariate shift". It delays the network to converge during the training phase. To avoid this problem, Batch normalization standardizes the input to each layer after every mini-batch and hence accelerates the network training. It is usually applied either before or after the activation functions of each hidden layer.

The process of batch normalization is shown in Fig. 2 [28]. It shifts the values of the input distribution to a hidden layer, such that the mean of these values is zero (zero centered) and then normalizes the inputs. It creates two parameter vectors for each layer, one with the scaled values and the second vector with the shifted values of the inputs to the layers.

The output scale vector 'γ' and the output offset vector 'β' are learnt through backpropagation. The final input mean vector 'μ' and the final input standard deviation vector 'σ' are estimated using exponential moving average during training.

C. Dropout

In a fully connected neural network, the probability of co-adaptation among the neurons is likely higher. As a result, the features extracted by the neurons for learning is more or less similar. This co-adaptation makes the model to overfit the training data and generalize poorly on unseen test data. Dropout is a technique to overcome this problem. It chooses a specified percentage of neurons randomly to be dropped during training by making their connection weights to zeros. Repeated application of this technique creates an ensemble of network architecture with a different set of neurons and their weights are dropped in each architecture as shown in Fig. 3(a) and (b) [29].
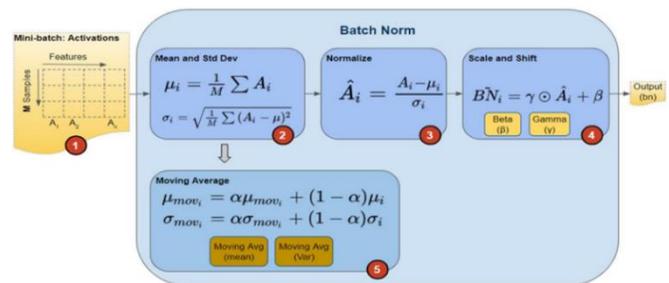
Fig. 2.   Batch Normalization.
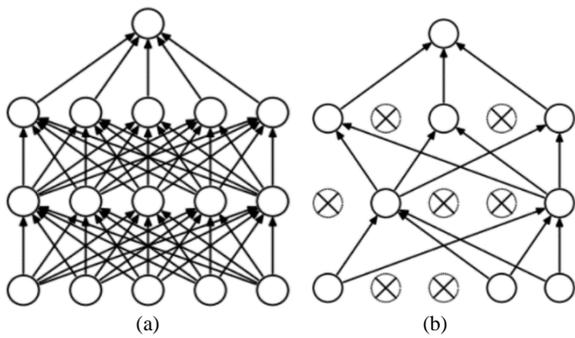
Fig. 3. (a). Standard Neural Network. (b). Neural Netw or with Dropouts.

The weights of these dropped neurons are not updated during backpropagation. When some of the neurons are dropped, the other neurons take the responsibility of propagating the features to the subsequent layers of the network in the forward pass. Hence it prevents a sort of co-adaptation among the neurons and makes the network less reliable to the learning units, their weights and existence. All these factors help to generalize the model well on the test samples. Dropouts in convolutional layers are applied to the individual cells of the feature map/kernel and are called spatial dropouts. Dropouts applied to the hidden layers are regular dropouts.

D. Bi-LSTM (Bi-Directional Long-Short Term Memory)

Bi-Directional Long-Short Term Memory is a type of Recurrent Neural Network (RNN), which is a deep learning technique that is used to categorize and regress timeseries data such as audio, text forecasting and so on. Bi-LSTM combines LSTM layers from both directions. Hence, it captures long-term dependencies between signal patterns by making the flow of information in both forward and backward directions. There are 3 components in LSTM, namely i) forget gate, ii) update gate, and iii) output gate. The forget gate eliminates the irrelevant information which is received from the preceding unit. The update gate performs addition of information to the cell state, and the output gate selects the relevant information from the present cell state and gives the output [30]. The LSTM gating structure manages the information by enabling the memory cells to preserve long-term dependencies through selective passage. It avoids the problems of gradient loss and gradient explosion by strengthening the weight of relevant information and weakening the weight of irrelevant ones. The structure of the LSTM cell is shown in Fig. 4(a).

The LSTM network cannot make use of the full data while processing the time series signals because it processes the data only in one direction. Hence, the Bi-LSTM network is implemented, which contains LSTM layers overlaid on each other in reverse direction. It improves the performance by enabling the model to make efficient use of the main features. The unit structure of the Bi-LSTM network is shown in Fig. 4(b).

The internal processing of the LSTM cell is shown in Fig. 4(c).
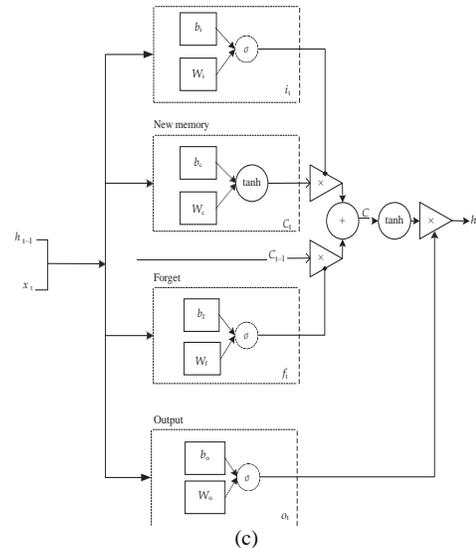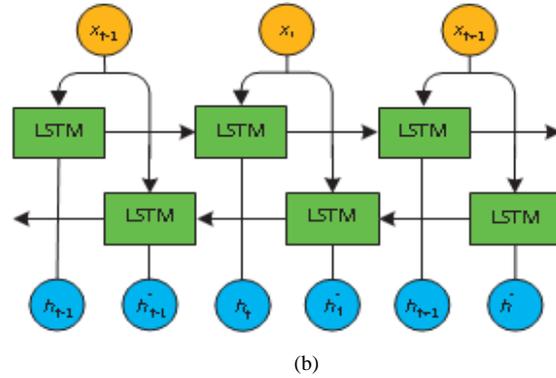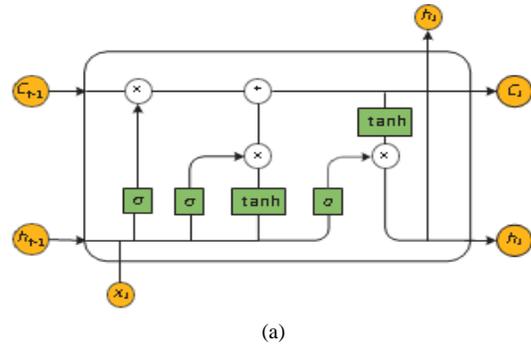
Fig. 4. (a). LSTM Cell Structure, (b). Bi-LSTM Unit Structure, (c). The Internal Process of the LSTM Cell.

The inputs W_i, W_o, W_f, W_c represents weights and b_i, b_o, b_f, b_c represents bias vectors of input gate, output gate, forget gate and cell state respectively. The input of current state and output of the previous state is represented as x_t and h_{t-1}. The input value C_t' at moment 't' is calculated by applying the tanh activation function on the result obtained by computing the matrix product of vector $[h_{t-1}, x_t]$ with W_c and b_c as given in (3).

$$C_t' = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{3}$$

The parametric value for each gate i.e., f_t-forget gate, i_t-input gate, and o_t-output gate at moment 't' is calculated by applying the activation function as shown in (4), (5) and (6).

$$f_t = \sigma \left(W_f \cdot [h_t\text{-}1, x_t] + b_f \right) \qquad (4)$$

$$i_t = \sigma \left(W_i \cdot [h_t\text{-}1, x_t] + b_i \right) \qquad (5)$$

$$o_t = \sigma \left(W_o \cdot [h_t\text{-}1, x_t] \right) + bo \qquad (6)$$

An element-wise product of '*ft*' with the last cell state 'C_{t-1}' determines the info that is to be forgotten and remembered by realizing the control on C_{t-1} and the element-wise product of 'i_t' with the current input cell state C_t' determines the info in C_t' that needs to be stored and used. The state value 'C_t' of the hidden node at time '*t*' is calculated as given in (7).

$$C_t = f_t * C_{t-1} + i_t * C_t' \qquad (7)$$

The output value 'h_t' at time '*t*' is computed as product of tanh function applied on unit state C_t, and output gate o_t, as given in (8).

$$h_t = o_t * \tanh(C_t) \qquad (8)$$

E. Softsign Activation Function

The softsign function squishes its input to a range of -1 to +1 as like tanh. The function and its derivative are defined as given in (9) and (10).

$$\text{softsign } f(x) = \frac{x}{1 + |x|} \qquad (9)$$

$$f'(x) = \frac{1}{(1 + |x|)^2} \qquad (10)$$

Unlike tanh, this function has a flatter curve, its derivative descends slowly, and is less saturated. Functions that are more saturated, have their gradients vanishing quickly before reaching the initial layers of the network during backpropagation [31]. Hence, softsign solves this vanishing gradient problem better than tanh. Softsign converges in polynomial time whereas tanh converges in exponential time. Since softsign transforms the inputs between -1 to +1, the negative values enable the LSTM gates to delete the information when required.

IV. EXPERIMENTAL SETUP

As the benchmark study, CWRU bearing dataset has been widely considered by many researchers for condition monitoring and fault diagnosis. An experimental setup of CWRU is shown in Fig. 5. It consists of a 2- hp (horsepower) motor, encoder, torque transducer, dynamometer, electric motor and so on. The deep groove ball bearing was mounted on the drive end of the motor to support the shaft which needs to be tested. An accelerometer was positioned above the bearing base of the drive end to measure the vibration signals. The load considered was about 1HP with 1772 rpm, and sampling rate of 12 kHz. Fault was induced in each component of rolling bearing by electro discharge with varying diameters of 0.007, 0.014, and 0.021 inches (1 inch = 25.4 mm).
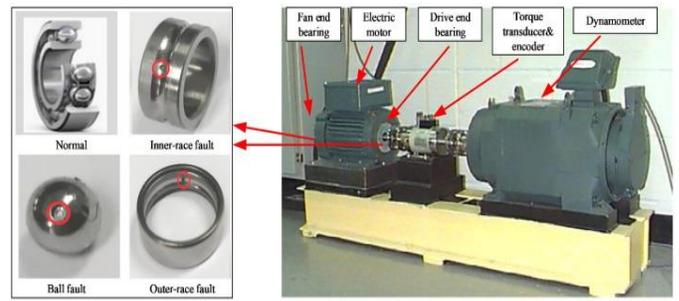
Fig. 5. Experimental Set-up of CWRU Data Collection.

In total, 12 bearing fault types with respect to BF, IR, OR and normal bearing were considered in this work as given in Table I. To label the fault types, One-Hot coding technique was used. In this experiment, to confirm adequate training size, 80% of the data was randomly chosen as training set and 20% as test set. For validation of the model, 10% of the training set was selected randomly to adjust model parameters.

V. RESULTS AND DISCUSSION

In this proposed work, the vibration data is collected for 12 different bearing conditions that is provided by CWRU. The description of various fault types and count of samples considered for each fault class from the experimental setup is given in Table I.

Time-domain features for normal and faulty bearings are shown in Fig. 6(a)-(d).

The proposed hybrid model is implemented using the Python's deep learning modules i.e. Tensorflow and Keras [32].

TABLE I. SUMMARY OF 12 DRIVE END FAULT DATA OF CWRU

Class	Types of Bearing Faults	No. of samples	Data description
C0	0.007-Ball	319	Ball fault level =0.007
C1	0.007-InnerRace	315	Inner Race fault level=0.007
C2	0.007-OuterRace12	318	Outer Race fault level at 12'o clock position = 0.007
C3	0.014-Ball	317	Ball fault level =0.014
C4	0.014-InnerRace	317	Inner Race fault level=0.014
C5	0.014-OuterRace6	317	Outer Race fault level at 6'o clock position = 0.014
C6	0.021-Ball	317	Ball fault level =0.021
C7	0.021-InnerRace	318	Inner Race fault level=0.021
C8	0.021-OuterRace12	317	Outer Race fault level at 12'o clock position = 0.021
C9	0.028-Ball	314	Ball fault level =0.028
C10	0.028-InnerRace	314	Inner Race fault level=0.028
C11	Normal	634	Normal Bearing

(a)

(b)

(c)

(d)
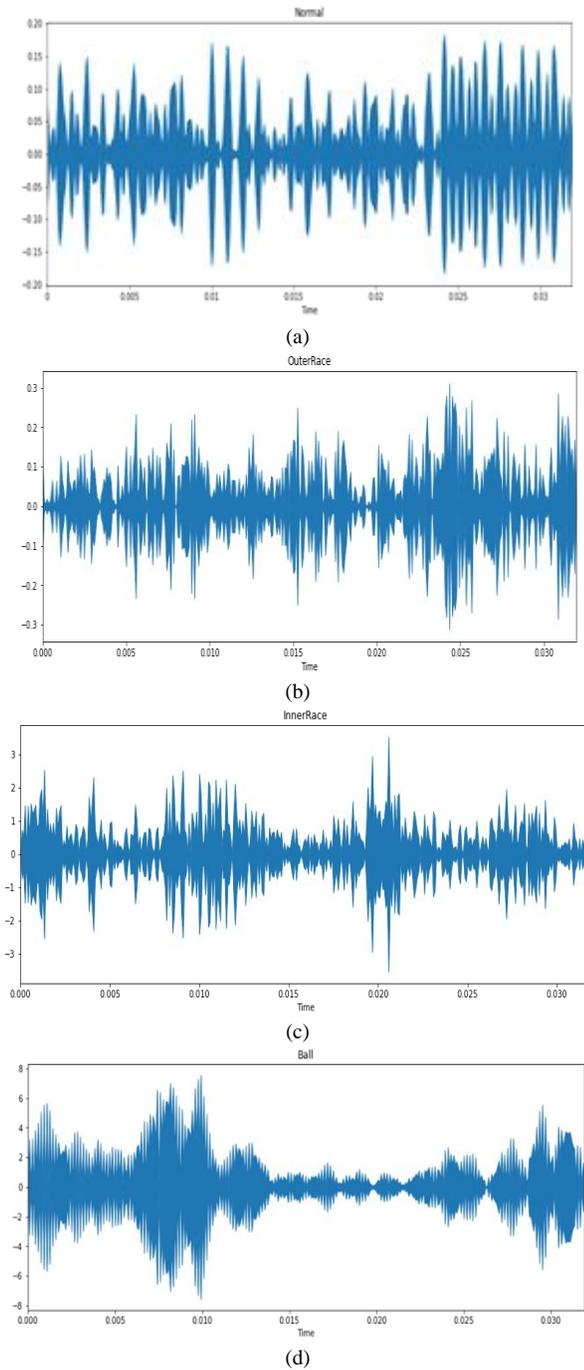
Fig. 6. (a). Time-domain Feature Representation for Normal Bearing, (b). Time-Domain Feature Representation for Outer Race Fault, (c). Time-Domain Feature Representation for Inner Race Fault, (d). Time-Domain Feature Representation for Ball Fault.

F. Parameter Settings for CNN-Bi-LSTM Model

The summary of the model's parameters set for the proposed hybrid CNN-Bi-LSTM architecture is shown in Fig. 7.

The first four convolutional layers used batch normalization and spatial dropout with a value of 0.25, which could effectively improve the performance of the network by preventing the overfitting problem. Softsign activation function

was used as a classifier in Bi-LSTM with filter size of 256 and Adam optimizer for compilation. The loss function categorical cross-entropy and batch size of 32 was set to identify the fault state by setting a short time of 100 epochs.

The trainable parameters are those which are learnt by the model during the feature learning from the classification layers namely convolution, LSTM, and the fully connected layers. The non-trainable parameters are learnt by the model from the batch normalization layers.

Layer (type)	Output Shape	Param #
conv1d_4 (Conv1D)	(None, 382, 128)	512
leaky_re_lu_7 (LeakyReLU)	(None, 382, 128)	0
max_pooling1d_4 (MaxPooling1D)	(None, 191, 128)	0
batch_normalization_7 (BatchNormalization)	(None, 191, 128)	512
spatial_dropout1d_4 (SpatialDropout1D)	(None, 191, 128)	0
conv1d_5 (Conv1D)	(None, 189, 256)	98560
leaky_re_lu_8 (LeakyReLU)	(None, 189, 256)	0
max_pooling1d_5 (MaxPooling1D)	(None, 94, 256)	0
batch_normalization_8 (BatchNormalization)	(None, 94, 256)	1024
spatial_dropout1d_5 (SpatialDropout1D)	(None, 94, 256)	0
conv1d_6 (Conv1D)	(None, 92, 512)	393728
leaky_re_lu_9 (LeakyReLU)	(None, 92, 512)	0
max_pooling1d_6 (MaxPooling1D)	(None, 46, 512)	0
batch_normalization_9 (BatchNormalization)	(None, 46, 512)	2048
spatial_dropout1d_6 (SpatialDropout1D)	(None, 46, 512)	0
conv1d_7 (Conv1D)	(None, 45, 256)	262400
leaky_re_lu_10 (LeakyReLU)	(None, 45, 256)	0
max_pooling1d_7 (MaxPooling1D)	(None, 45, 256)	0
batch_normalization_10 (BatchNormalization)	(None, 45, 256)	1024
spatial_dropout1d_7 (SpatialDropout1D)	(None, 45, 256)	0
bidirectional_1 (Bidirectional)	(None, 45, 512)	1050624
leaky_re_lu_11 (LeakyReLU)	(None, 45, 512)	0
batch_normalization_11 (BatchNormalization)	(None, 45, 512)	2048
dropout_3 (Dropout)	(None, 45, 512)	0
lstm_3 (LSTM)	(None, 256)	787456
leaky_re_lu_12 (LeakyReLU)	(None, 256)	0
batch_normalization_12 (BatchNormalization)	(None, 256)	1024
dropout_4 (Dropout)	(None, 256)	0
dense_2 (Dense)	(None, 64)	16448
leaky_re_lu_13 (LeakyReLU)	(None, 64)	0
batch_normalization_13 (BatchNormalization)	(None, 64)	256
dropout_5 (Dropout)	(None, 64)	0
dense_3 (Dense)	(None, 12)	780

```
Total params: 2,618,444
Trainable params: 2,614,476
Non-trainable params: 3,968
```

Fig. 7. Summary of Model's Parameters.

G. Confusion Matrix

The confusion matrix is the representation of matching degree between the actual and predicted labels in the form of matrix. The confusion matrix for the proposed 1DCNN-Bi-LSTM model is shown in Fig. 8. The model has correctly classified 4111 samples out of 4118 by demonstrating an accuracy of 99.84%.
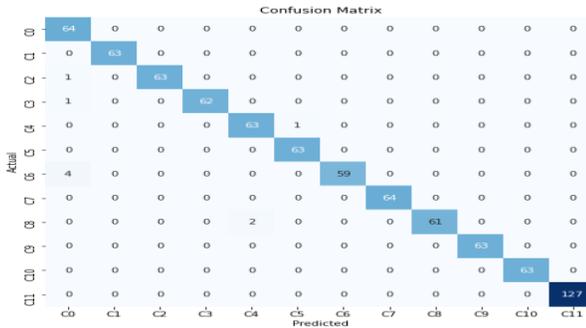
Fig. 8.   Confusion Matrix.

H. Learning Curve

A learning curve is a plot of model's learning performance over experience or time. The model is evaluated during training phase after each update based on the training and validation dataset. It gives an idea of how well the model is learning and generalizing. The learning curve for the proposed method is shown in Fig. 9. It demonstrates the training and validation accuracy versus number of epochs.
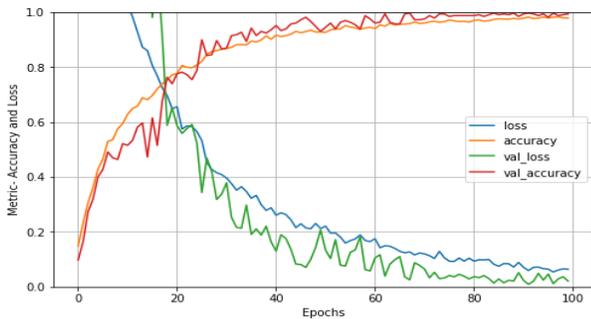
Fig. 9.   Learning Curve for Hybrid CNN-Bi-LSTM Model.

A comparative analysis of the proposed hybrid model is made with other existing DL models as shown in Table II. The performance indicates that the proposed model accomplishes better results in classifying multiple faults as compared to other models.

TABLE II.   COMPARATIVE ANALYSIS OF THE PROPOSED METHOD WITH OTHER DL MODELS

Model	Training Accuracy	Test Accuracy	Validation Accuracy
1D-CNN only	98.18	95.99	95.45
1D-CNN with LSTM	98.18	99.51	99.09
1D-CNN-Bi-LSTM	98.01	98.78	99.09
1D-CNN-Bi-LSTM (Softsign and Spatial Dropout)	99.84	98.17	99.69

VI. CONCLUSION

In this research, A Hybrid 1D-CNN-Bi-LSTM model with Spatial Dropout for Multiple Fault Diagnosis of Roller Bearing is proposed. Usage of Spatial Dropout technique and Softsign activation function in the proposed hybrid fault diagnosis method has shown an improvement in the accuracy by performing automatic feature extraction and preventing the problem of overfitting. 1D-CNN extracts the features from the raw signal and Bi-LSTM layer fuses the feature information to enhance stability of the model and classify the faults. The efficiency of the proposed model is analysed by considering CWRU bearing vibration data for experimentation. A comparative analysis of the proposed method is made with other existing models. The model has shown the performance accuracy of 99.84% in classifying 12 different fault types. Therefore, the proposed hybrid 1D-CNN-Bi-LSTM (Softsign and Spatial Dropout) is an effective multi-class fault diagnosis method with the prevention of model overfitting problem.

REFERENCES

[1] Yuan, Laohu, Dongshan Lian, Xue Kang, Yuanqiang Chen, and Kejia Zhai. "Rolling bearing fault diagnosis based on convolutional neural network and support vector machine." IEEE Access 8 (2020): 137395-137406.

[2] Gupta, Pankaj, and M. K. Pradhan. "Fault detection analysis in rolling element bearing: A review." Materials Today: Proceedings 4, no. 2 (2017): 2085-2094.

[3] Neupane, Dhiraj, and Jongwon Seok. "Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review." IEEE Access 8 (2020): 93155-93178.

[4] Zhang, Shen, Shibo Zhang, Bingnan Wang, and Thomas G. Habetler. "Deep learning algorithms for bearing fault diagnostics—A comprehensive review." IEEE Access 8 (2020): 29857-29881.

[5] Tang, Shengnan, Shouqi Yuan, and Yong Zhu. "Deep learning-based intelligent fault diagnosis methods toward rotating machinery." Ieee Access 8 (2019): 9335-9346.

[6] Chen, Zhiqiang, Shengcai Deng, Xudong Chen, Chuan Li, René-Vinicio Sanchez, and Huafeng Qin. "Deep neural networks-based rolling bearing fault diagnosis." Microelectronics Reliability 75 (2017): 327-333.

[7] Lei, Jinhao, Chao Liu, and Dongxiang Jiang. "Fault diagnosis of wind turbine based on Long Short-term memory networks." Renewable energy 133 (2019): 422-432.

[8] Chen, Zhuyun, and Weihua Li. "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network." IEEE Transactions on Instrumentation and Measurement 66, no. 7 (2017): 1693-1702.

[9] Zhang, Yuyan, Xinyu Li, Liang Gao, Wen Chen, and Peigen Li. "Intelligent fault diagnosis of rotating machinery using a new ensemble deep auto-encoder method." Measurement 151 (2020): 107232.

[10] Zhao, Rui, Dongzhe Wang, Ruqiang Yan, Kezhi Mao, Fei Shen, and Jinjiang Wang. "Machine health monitoring using local feature-based gated recurrent unit networks." IEEE Transactions on Industrial Electronics 65, no. 2 (2017): 1539-1548.

[11] Ince, Turker, Serkan Kiranyaz, Levent Eren, Murat Askar, and Moncef Gabbouj. "Real-time motor fault detection by 1-D convolutional neural networks." IEEE Transactions on Industrial Electronics 63, no. 11 (2016): 7067-7075.

[12] Abed, Wathiq, Sanjay Sharma, Robert Sutton, and Amit 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 26, no. 3 (2015): 241-254.

[13] Zou, Ping, Baocun Hou, Jiang Lei, and Zhenji Zhang. "Bearing fault diagnosis method based on EEMD and LSTM." International Journal of Computers Communications & Control 15, no. 1 (2020).

[14] Cao, Lixiao, Zheng Qian, Hamidreza Zareipour, Zhenkai Huang, and Fanghong Zhang. "Fault diagnosis of wind turbine gearbox based on

deep bi-directional long short-term memory under time-varying non-stationary operating conditions." IEEE Access 7 (2019): 155219-155228.

[15] Choudakkanavar, Gangavva, J. Alamelu Mangai, and Mohit Bansal. "MFCC based ensemble learning method for multiple fault diagnosis of roller bearing." International Journal of Information Technology (2022).

[16] Wang, Zheng, Hongzhan Ma, Hansi Chen, Bo Yan, and Xuening Chu. "Performance degradation assessment of rolling bearing based on convolutional neural network and deep long-short term memory network." International Journal of Production Research 58, no. 13 (2020): 3931-3943.

[17] Sun, Haibin, and Shichao Zhao. "Fault diagnosis for bearing based on 1DCNN and LSTM." Shock and Vibration 2021 (2021).

[18] Zhang, Xiaochen, Yiwen Cong, Zhe Yuan, Tian Zhang, and Xiaotian Bai. "Early fault detection method of rolling bearing based on MCNN and GRU network with an attention mechanism." Shock and Vibration 2021 (2021).

[19] Jin, Guoqiang, Tianyi Zhu, Muhammad Waqar Akram, Yi Jin, and Changan Zhu. "An adaptive anti-noise neural network for bearing fault diagnosis under noise and varying load conditions." IEEE Access 8 (2020): 74793-74807.

[20] Yu, Wenbing, and Pin Lv. "An end-to-end intelligent fault diagnosis application for rolling bearing based on MobileNet." IEEE Access 9 (2021): 41925-41933.

[21] Gu, Kai, Yu Zhang, Xiaobo Liu, Heng Li, and Mifeng Ren. "DWT-LSTM-based fault diagnosis of rolling bearings with multi-sensors." Electronics 10, no. 17 (2021): 2076.

[22] Zhi, Zhuo, Liansheng Liu, Datong Liu, and Cong Hu. "Fault detection of the harmonic reducer based on CNN-LSTM with a novel denoising algorithm." IEEE Sensors Journal 22, no. 3 (2021): 2572-2581.

[23] Huang, Ting, Qiang Zhang, Xiaoan Tang, Shuangyao Zhao, and Xiaonong Lu. "A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems." Artificial Intelligence Review 55, no. 2 (2022): 1289-1315.

[24] Hotait, Hassane, Xavier Chiementin, and Lanto Rasolofondraibe. "Intelligent Online Monitoring of Rolling Bearing: Diagnosis and Prognosis." Entropy 23, no. 7 (2021): 791.

[25] Luo, Jiahang, and Xu Zhang. "Convolutional neural network based on attention mechanism and Bi-LSTM for bearing remaining life prediction." Applied Intelligence 52, no. 1 (2022): 1076-1091.

[26] https://engineering.case.edu/bearingdatacenter

[27] Li, Mingyong, Qingmin Wei, Hongya Wang, and Xuekang Zhang. "Research on fault diagnosis of time-domain vibration signal based on convolutional neural networks." Systems Science & Control Engineering 7, no. 3 (2019): 73-81.

[28] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." In International conference on machine learning, pp. 448-456. PMLR, 2015.

[29] Srivastava, Nitish, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15, no. 1 (2014): 1929-1958.

[30] Pan, Honghu, Xingxi He, Sai Tang, and Fanming Meng. "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM." Strojniski Vestnik/Journal of Mechanical Engineering 64 (2018).

[31] Qiu, Dawei, Zichen Liu, Yiqing Zhou, and Jinglin Shi. "Modified Bi-directional LSTM neural networks for rolling bearing fault diagnosis." In ICC 2019-2019 IEEE international conference on communications (ICC), pp. 1-6. IEEE, 2019.

[32] https://www.tensorflow.org/api_docs/python/tf/keras

AUTHORS' AFFILIATIONS

First Author received her B.E. and M.Tech degree in Computer Science and Engineering from Visvesvaraya Technological University (VTU), Belagavi, in 2009 and 2013 respectively. Currently pursuing Ph.D. degree from Presidency University, Bangalore. Since from 2013, she is working as Assistant Professor in various recognized institutions. Her research interests include Machine Learning, Deep Learning and Fault Diagnosis.

Second Author received her Ph.D from BITS Pilani, Dubai Campus in 2015. She is working in the department of Computer Science and Engineering at Presidency University, Bangalore. Her research interests include Data Mining and Machine Learning Algorithms and Applications.