#### **Research Article**

P B Mallikarjuna\*, Sreenatha M, Manjunath S, and Niranjan C Kundur

# Aircraft Gearbox Fault Diagnosis System: An Approach based on Deep Learning Techniques

https://doi.org/10.1515/jisys-2019-0237 Received Oct 11, 2019; accepted May 27, 2020

**Abstract:** Gearbox is one of the vital components in aircraft engines. If any small damage to gearbox, it can cause the breakdown of aircraft engine. Thus it is significant to study fault diagnosis in gearbox system. In this paper, two deep learning models (Long short term memory (LSTM) and Bi-directional long short term memory (BLSTM)) are proposed to classify the condition of gearbox into good or bad. These models are applied on aircraft gearbox vibration data in both time and frequency domain. A publicly available aircraft gearbox vibration dataset is used to evaluate the performance of proposed models. The results proved that accuracy achieved by LSTM and BLSTM are highly reliable and applicable in health monitoring of aircraft gearbox system in time domain as compared to frequency domain. Also, to show the superiority of proposed models for aircraft gearbox fault diagnosis, performance is compared with classical machine learning models.

Keywords: Gearbox, Vibration data, Long short term memory, Bi-directional long short term memory

## 1 Introduction

Gears are the vital part of the mechanical transmission systems. It is used in rotating machinery and in the design of transmission systems for automobiles. In the field of rotating machinery and mechanical transmission systems, the application of gears is essential. So, the proper monitoring of the gear system is crucial to ensure the performance of the mechanical transmission systems. In any production industries, the breakdown of such vital components leads to production losses and increases the cost of maintenance. Therefore, it is very essential to detect the faults of gears proactively to prevent breakdowns, accidents and to ensure the operation of mechanical transmission systems with no faults.

Impacting and friction can occur in gearboxes because of a crack in a gear and that can also causes a slight change in the speed. Cracks in a gearbox system will occur due to the continuous usage over a period of time. This will leads to defects in a gearbox. Hence, periodic maintenance and feedback are necessary to prevent defects in gearboxes. Defects in gearbox are the source of vibration in machinery. Mechanical transmission systems will exhibit high level of vibration at some point in their lifetime. The defects in the mechanical transmission system can be detected by analysing vibration data of mechanical devices. The analysis of vibration data is extensively used to diagnose mechanical machine's health condition. Vibration data analysis is a process of looking for deviations from the standard condition of the mechanical devices. Hence, these defects can be identified by analysing the vibration data of gearbox system collected through electronic sensors. In the following section, some of the attempts have been traced on diagnosis of gearbox system using machine learning and deep learning techniques.

**Sreenatha M, Niranjan C Kundur:** JSS Academy of Technical Education, Bengaluru-560060, Karnataka, India **Manjunath S:** Freelance Academician, Mysore-570016, Karnataka, India

<sup>\*</sup>Corresponding Author: P B Mallikarjuna: JSS Academy of Technical Education, Bengaluru-560060, Karnataka, India; Email: putaplar@gmail.com

### 2 Related work

In connection with this, few attempts have been made on classification of gearbox system's health condition using classical machine learning algorithms. Gearbox health diagnosis model was proposed based on Continues Wavelet Transform (CWT) coefficients, which were extracted from vibration data of gear box system. Further, Gaussian Mixture Model (GMM) and KNN (K- Nearest Neighbor) classifier are used separately for the purpose of classification [1]. The statistical features were extracted from vibration data of gearbox system and support vector machine (SVM) classifier was used [2]. Spectral correlation density was estimated from signals of vibration data of gear system and used as an input to the SVM classifier to classify the gear's health condition into good or bad [3]. Statistical features extracted out of vibration data of gearbox system were used to build a SVM classifier for gearbox diagnosis [4]. However, the limitation of KNN and GMM is that these will not perform well when the dimensionality of data is high. Also, KNN does not perform well when dataset is large. The limitation in building gearbox health monitoring system using SVM classifier is that it is not suitable for large datasets [5]. A rotating machinery fault diagnosis method based on local mean decomposition was proposed to analyze vibration signal in time-frequency domain [6]. A PCA based model was proposed to diagnose rotatory machine faults using statistical features extracted out of vibration data [7]. An adaptive neuro-fuzzy inference system was proposed for identification and classification of gear's health condition [8]. In this model, discrete wavelet transform (DWT) method was used to extract the features from the spectrum of vibration signals. A back propagation neural network (BPNN) model was designed and implemented to diagnose the gearbox systems. The model was trained using selected FFT features extracted from vibration data [9]. Ensemble Empirical mode decomposition (EEMD) based Deep Briefs Network (DBN) was proposed. In this, the vibration data is decomposed in to a set of IMFs (intrinsic mode functions) using EEMD and signals are reconstructed from main IMFs. These reconstructed signals are used as input to the DBN [10].

Explicit feature extraction from raw data is the first step in building classification model in the above said classical machine learning techniques. But, there is an option to eliminate explicit feature extraction step by using deep learning techniques i.e., deep learning techniques automatically extract required learning parameters independent of number of training samples. A Convolution Neural Network (CNN) was designed to learn features automatically from vibration data of gearbox system [11]. Statistical features (standard deviation, skewness, and kurtosis) were extracted in the time domain of vibration data and CNN was used for classification [12]. An adaptive multi-sensor data fusion method based on Deep Convolutional Neural Networks (DCNN) was proposed for fault diagnosis in gearbox systems [13]. Hybrid-deep model was proposed to diagnose faults in rotary machines [14]. This model consists of multi-channel CNN followed by stack of denoising encoders. The model was validated on benchmark vibration dataset. An intelligent mechanical fault diagnoses method based on DTCWT (Dual-Tree Complex Wavelet Transform) and CNN was proposed to identify mechanical faults [15]. CNN architectures are designed to process spatial data such as image data whereas recurrent neural network (RNN) architectures are designed to process time series and sequence data [16]. However, RNN exhibits vanishing gradient and long term dependency problems. These problems are overcome by forming LSTM network (modified version of RNN) [17]. LSTM cells are connected in chain like structure and each cell has the ability to remove or add information to cell state and passes the updated cell state to the next cell. Therefore, LSTM network is more suitable for problems involving long sequential data processing. Therefore, in this paper, a LSTM based aircraft gear-box diagnosis model is proposed to classify health condition of gearbox system into good or bad based on the analysis of vibration data (time series data). We have also proposed a BLSTM model which is LSTM version Bidirectional RNN (BRNN) structure. Unlike the standard LSTM structure, two different LSTM networks (forward and backward) are trained for sequential inputs in the BLSTM architecture [18]. This BRNN version of LSTM (BLSTM) improved the performance of classification of Gearbox health condition.

For both LSTM and BLSTM based model, the input can be raw data signal. The raw data can be in either time domain or frequency domain. Hence, in this work, we explore the suitability of input data to LSTM or BLSTM in both time and frequency domain.

With this, the following objectives are addressed in this paper.

- 1. We design LSTM and BLSTM based architecture for the analysis of gearbox vibration data.
- 2. We study the suitability of input data in time and frequency domain to the proposed architectures.
- 3. Comparing the performance of proposed models with classical machine learning models.

# 3 Proposed Model

LSTM and BLSTM based gear-box diagnosis model is proposed to classify the health condition of gearbox into good or bad. Section 3.1 describes the architecture of LSTM cell. Section 3.2 presents the design of general LSTM based learning model and proposed LSTM and BLSTM based models in both time and frequency domain.

#### 3.1 Architecture of LSTM Cell

LSTM network is comprised of different memory blocks called LSTM cells and each cell has hidden state. These states are responsible for transferring previous time-step information from one cell to the next cell in the LSTM network. This is achieved by three different gates viz, input, forget and output gates. These gates are embedded in each cell as shown in the Figure 1. The function of each gate is illustrated as follows.

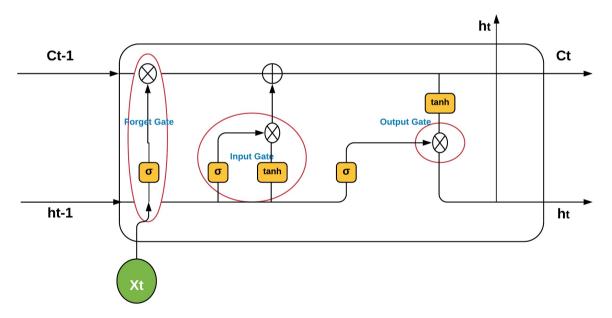


Figure 1: LSTM Cell Structure

The forget gate is responsible for removing insignificant information from the cell state received from the previous cell. The input gate is responsible for addition of information to the cell state. The output gate is responsible for selecting useful information from the current cell state and showing it gives as an output. The respective function of each gate is achieved by the following equations.

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

$$i_t = \sigma(W_i . h_{t-1} + W_i . x_t + b_i)$$
 (2)

$$o_t = \tanh(W_o.h_{t-1} + W_o.x_t + b_o)$$
 (3)

$$c_t = f_t * c_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$
(4)

$$h_t = o_t * \tanh(c_t) \tag{5}$$

261

Where,

 $f_t$  = Forget gate at time step t

 $i_t$  = Input gate at time step t

 $o_t$  = Output gate attime step t

 $c_{t-1}$  = Previous cell state

 $h_{t-1}$  = Previous hidden state

 $c_t$  = Current cell state

 $h_t$  = Current hidden state

 $\sigma$  = Sigmoid function

tanh = Hyperbolic tangent function

 $b_f$ ,  $b_i$ ,  $b_o$  = bias of foerget gate, input gate and output gate respectively

 $W_f$ ,  $W_i$ ,  $W_o$  = Weight matrices of forget, input and output gate respectively

## 3.2 Design

Given any vibration of time step T with d output labels, a LSTM based model can be built as shown in the Figure 2.

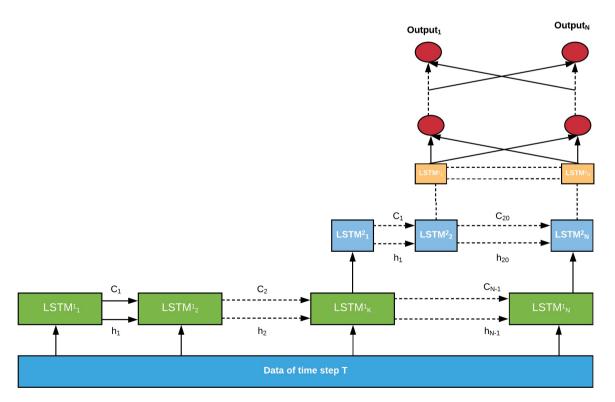


Figure 2: General LSTM based learning model

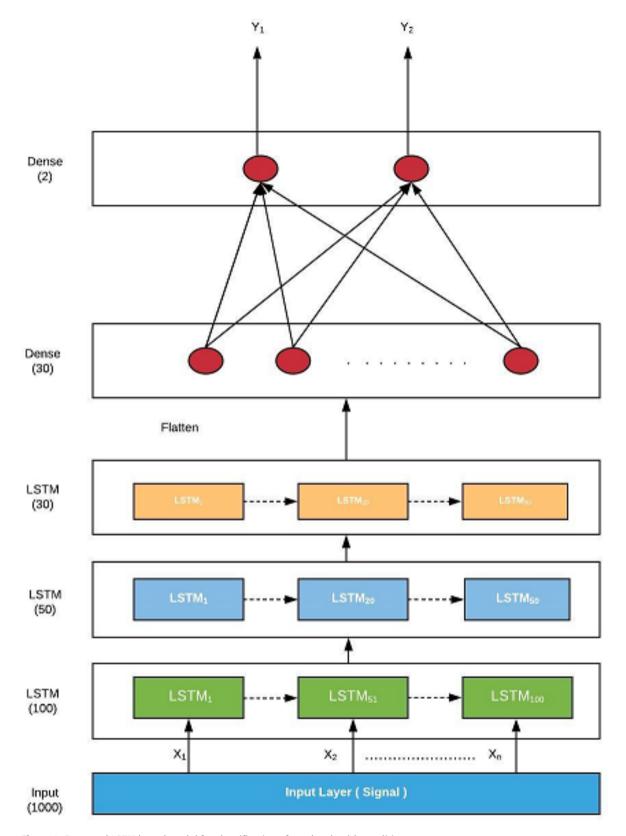


Figure 3: Proposed LSTM based model for classification of gearbox health conditions.

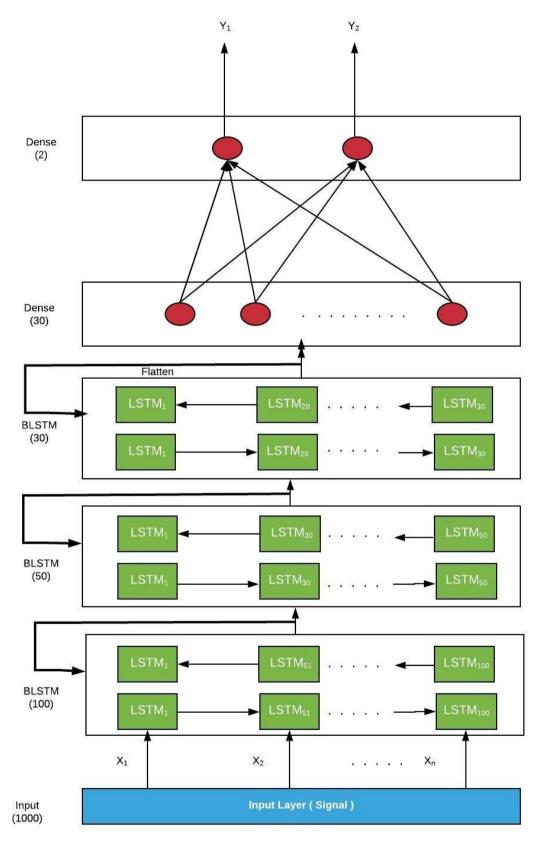


Figure 4: Proposed BLSTM based model for classification of gearbox health conditions.

Given any LSTM architecture, we can have any number of LSTM layers. Let's say we have ' $\gamma$ ' number of layers. In each layer, we have 'N' number of LSTM cells. From each layer, we can take 'K' number of LSTM cells output as input to the next layer. The output of the last layer can be used as a compressed representation of the input data. The obtained representation of the data can be feed to any learning algorithm, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN), or Neural Network (NN). We proposed to use simple fully connected NN to have better representation. The last layer of the LSTM network is considered as input layer for the fully connected neural network. The number of neurons in the output layer of fully connected neural network is equal to the number of classes need to be predicted.

Based on above general architecture, for the considered aircraft gearbox data vibration, we have designed the model in time domain as follows:

- (1) Number of LSTM layers =  $3 (\gamma = 3)$
- (2) Number of LSTM cells in
  - Layer 1 = 100 (N1 = 100)
  - Layer 2 = 50 (N2 = 50)
  - Layer 3 = 30 (N3 = 30)
- (3) We have taken two layer fully connected layers with  $L_1 = 30$  and  $L_2 = 2$ , as our problem is a two class problem.

The block diagram of the proposed model is as shown in the Figure 3. It consists of 3-layer LSTM network followed by fully connected two layer neural network. The values  $\gamma$ , N1, N2, N3 and K are obtained empirically (Refer experiment section 4).

The design of proposed BLSTM model is same as proposed LSTM model. The only changes in BLSTM are that it tries to understand the data in both directions (forward and backward) to have better representation. The proposed BLSTM based model for classification of gearbox health conditions is as shown in the Figure 4.

In order to design proposed models in frequency domain, the raw vibration data (time series vibration data) of aircraft gearbox is converted in to frequency domain data using Fast Fourier Transform [19]. The representation of good and bad conditioned sample signals of aircraft gearbox system in frequency domain is depicted in Figure 5 and Figure 6 respectively. By observing signals of good and bad conditioned samples in frequency domain, discrimination between good and bad samples are found in terms of amplitude levels in the frequency range 1hz - 50hz. Therefore, each sample is considered of vector length 50 with corresponding

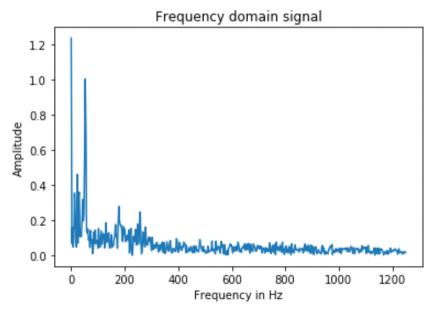


Figure 5: A sample signal representation of bad conditioned aircraft gearbox system in frequency domain

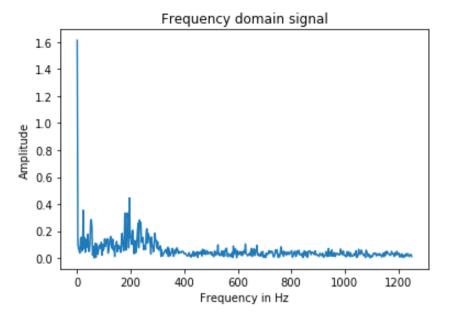


Figure 6: A sample signal representation of good conditioned aircraft gearbox system in frequency domain

amplitude levels against to the frequency range 1hz - 50hz. The following changes are made for proposed LSTM and BLSTM models in the frequency domain.

- (1) Number of LSTM layers =  $3 (\gamma = 3)$
- (2) Number of LSTM cells in
  - Layer 1 = 50 (N1 = 50)
  - Layer 2 = 50 (N2 = 50)
  - Layer 3 = 50 (N3 = 50)
- (3) We have taken two layer fully connected layers with  $L_1 = 50$  and  $L_2 = 2$ , as our problem is a two class problem.

The values  $\gamma$ , N1, N2, N3 and K are obtained empirically (Refer experiment section 4).

# 4 Experimentation and Results

In order to conduct the experimentation, we have used publicly available aircraft gearbox vibration dataset [20]. Vibration data taken on the exterior of an aircraft as it climbed from 23,000 feet to 40,000 feet. The dataset consists of 13 samples of vibration data of good condition and 11 samples of vibration data of bad condition aircraft gear. Each sample is captured at a rate of 2500 time steps per second with duration of 38 seconds. Dataset is tabulated in the following Table 1.

Table 1: Aircraft gearbox vibration dataset

Air craft gear box condition	Number of samples	Time steps per second	Time steps per sample	
Good	13	2500	07000	
Bad	11	2500	97000	

From the above Table 1, each sample consists of 97000 time steps. As the length of the sequence is too large we split the data into subsamples, where each subsample is of 1000 time steps. Hence, from each original sample of length 97000 time steps, we derive 97 samples by considering 1000 time steps for each individual sample. From this we have a total of 1261 (97  $\times$  13) good condition gearbox samples and a total of 1067 (97  $\times$  11) bad condition gearbox samples. Hence, the derived dataset consists of 2328 samples as tabulated in the following Table 2. The reason for selecting the time step of subsample as 1000 is by convenience for experimentation; however the length of the subsample can be varied. We have chosen 2000 samples randomly out of 2328 samples for experimentation.

Table 2: Derived Aircraft gearbox vibration dataset

Air craft gear box condition	Number of samples	Total Number of samples
Good	1261	2220
Bad	1067	2328

Proposed models are evaluated in both time and frequency domain using performance measures such as Precision, Recall, F-measure and classification accuracy. In order to conduct the experimentation, we have done in two folds. In the first fold, dataset is split into 70% training (1400 samples) and 30% testing (600 samples). In the second fold dataset is split into 80% training (1600 samples) and 20% testing (400 samples). The number of epochs used is 100.

In order to select the number of cells in three hidden layers for the proposed BLSTM architecture (Figure 4), we carry out experimentation by varying the number of cells in each layer and classification accuracy has been estimated for 80% training. The experimentation results under varying cells are tabulated in the Table 3 and the same is graphically shown in Figure 7.

Table 3: Classification Accuracy under varying number of cells in three hidden layer for the proposed BLSTM architecture

Test Case	BLSTM Cells (N1, N2, N3)	Classification Accuracy
1	100,100, 100	80.37
2	100, 75, 55	96.25
3	100, 50, 30	99.75
4	90,45,25	97.75
5	80,40,20	83.5

From the Table 3 and Figure 7, it is observed that the best accuracy is obtained for N1 = 100, N2 = 50, N3 = 30 cells (Test Case 3) and hence we have used the same in the article. We have adopted a simple technique of using only 10% of actual data in the first layer and in further layers we have reduced the number of cells approximately by 50%. However, one can think of using grid search algorithm for selecting these parameters.

Similarly, we have also selected the number of cells in three hidden layers in the proposed LSTM architecture using above said empirical method. It is found that the best accuracy is obtained for N1 = 100, N2 = 50, N3 = 30 cells. In frequency domain, best accuracy is obtained for N1 = 50, N2 = 50, N3 = 50 cells in the both the proposed architectures.

Performance results of the proposed models are tabulated in Table 4. The ROC curves of proposed models are depicted in Figure 8 to Figure 15. From Table 4, it is evident that BLSTM in time domain performs better than LSTM and we have achieved an accuracy of 99.75% and error rate is about 0.25%.

We also build classical machine learning models such as K-Nearest Neighbour (KNN), Linear SVM, RBF (Radial Basis Function) SVM, Gaussian Process (GP), Decision Tree (DT), Random Forest (RF), Multilayer Perceptron (MLP), AdaBoost, Gaussian Naive Bayesian (NB) and Quadratic Discriminant Analysis (QDA) for

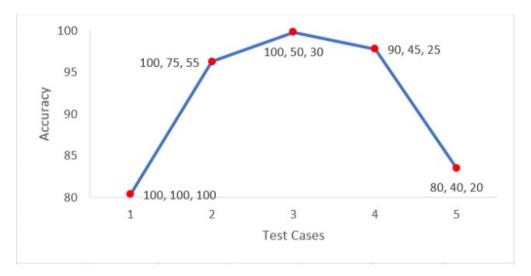


Figure 7: Classification accuracy versus varying number of cells in three hidden layers for the proposed BLSTM architecture

Table 4: Performance results of proposed models

Models	Training Samples	Precision	Recall	F-measure	Classification accuracy
LSTM in time domain	70%	0.978	0.977	0.977	97.83%
	80%	0.983	0.980	0.981	98.38%
BLSTM in time domain	70%	0.982	0.981	0.981	98.25%
	80%	0.997	0.997	0.997	99.75%
LSTM in frequency domain	70%	0.923	0.924	0.923	92.25%
	80%	0.934	0.934	0.934	93.25%
BLSTM in frequency domain	70%	0.935	0.930	0.932	93.08%
	80%	0.937	0.938	0.937	93.75%

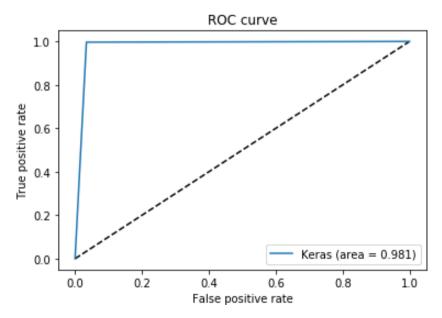


Figure 8: ROC curve for 70% training using LSTM in time domain

 Table 5: Comparison of performance of Proposed models with classical machine learning models

Models	<b>Training Samples</b>	Classification accuracy
V Nearest Neighbour	70%	71.83%
K-Nearest Neighbour	80%	75.75%
Linear SVM	70%	60.16%
Linear SVM	80%	59.75%
DDE (Dadial Basis Function) SVM	70%	77.83%
RBF (Radial Basis Function) SVM	80%	78.00%
Gaussian Process	70%	78.50%
Gaussian Process	80%	79.75%
Decision Tree	70%	74.83%
Decision Tree	80%	76.25%
Dan dam Fayaat	70%	77.16%
Random Forest	80%	77.50%
A4 4:	70%	77.33%
Multilayer Perceptron	80%	78.25%
A J. D +	70%	76.50%
AdaBoost	80%	78.75%
Coursian Naiva Paussian	70%	70.33%
Gaussian Naive Bayesian	80%	70.75%
Overductic Discriminant Analysis	70%	76.33%
Quadratic Discriminant Analysis	80%	75.00%
Dranged Model (LCTM in time domain)	70%	97.83%
Proposed Model (LSTM in time domain)	80%	98.38%
Duna and Mardal (DI CTM in time adamain)	70%	98.25%
Proposed Model (BLSTM in time domain)	80%	99.75%

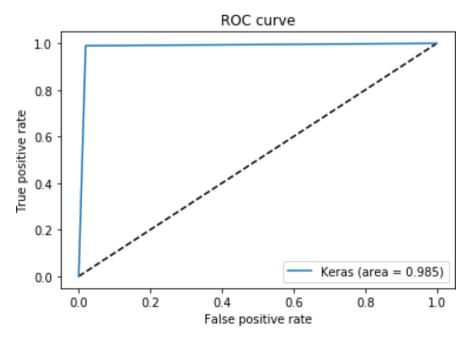


Figure 9: ROC curve for 80% training using LSTM in time domain

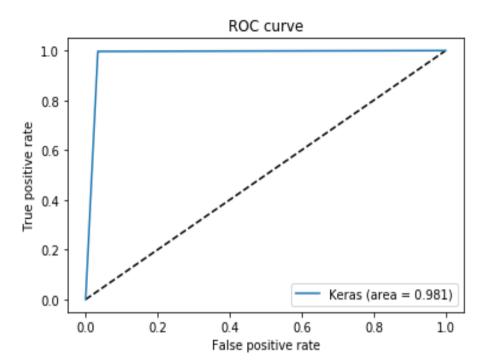


Figure 10: ROC curve for 70% training using BLSTM in time domain

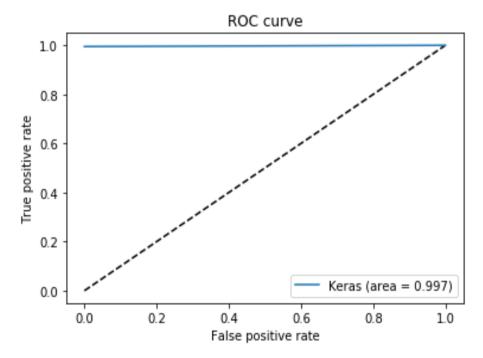


Figure 11: ROC curve for 80% training using BLSTM in time domain

gearbox diagnosis and compared the performance of proposed models with these classical machine learning models. Hence, Statistical features such as min, max, mean, standard deviation, variance, autocorrelation, quantile, skewness and entropy are extracted from the raw vibration data (time series vibration data) of bad and good conditioned samples of aircraft gearbox system. The classical machine learning classifiers are used separately for the purpose of classification of samples of aircraft gearbox vibration data into good or bad. We made a same experimental setup as in the proposed models. Comparison of performance of the proposed

models (LSTM and BLSTM) with classical machine learning models is tabulated in Table 5. From Table 5, it is evident that best classification accuracy is achieved using proposed models in time domain as compared to classical machine learning models.

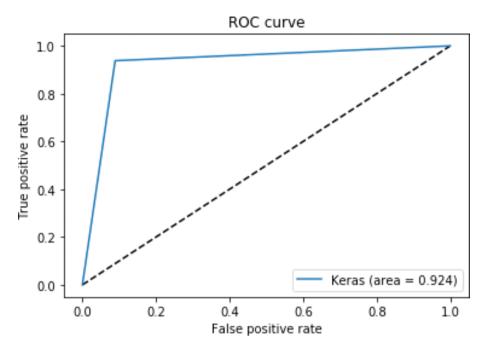


Figure 12: ROC curve for 70% training using LSTM in frequency domain

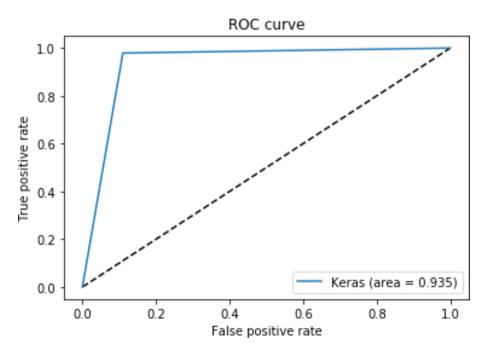


Figure 13: ROC curve for 80% training using LSTM in frequency domain

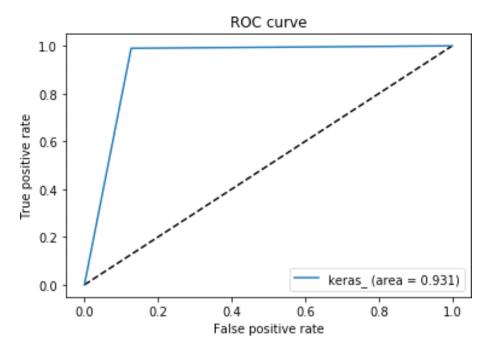


Figure 14: ROC curve for 70% training using BLSTM in frequency domain

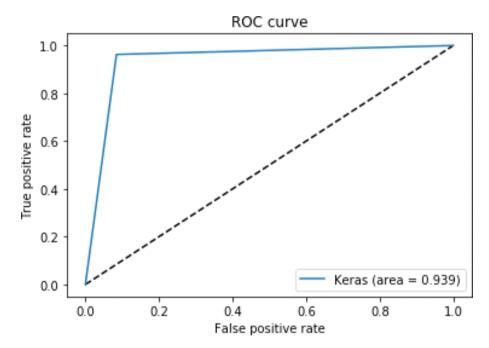


Figure 15: ROC curve for 80% training using BLSTM in frequency domain

# 5 Conclusion

In this paper, we have proposed LSTM and BLSTM based model for analysis of aircraft gearbox vibration data to diagnose healthiness of gearbox. We conducted experimentation on publicly available aircraft gearbox data set. The experimental results show that BLSTM model is superior to LSTM model in both time domain and frequency domain. Proposed models outperform the classical machine learning models in diagnosing the gearbox system's health conditions in both time and frequency domain.

## References

- [1] Gharavian, M.H., Ganj, F.A., Ohadi, A.R. and Bafroui, H.H., 2013. Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes. Neurocomputing, 121, pp.150-159.
- [2] Praveenkumar, T., Saimurugan, M., Krishnakumar, P. and Ramachandran, K.I., 2014. Fault diagnosis of automobile gearbox based on machine learning techniques. Procedia Engineering, 97, pp. 2092-2098.
- [3] Moster, P.C., 2004. Gear fault detection and classification using learning machines. Sound and vibration, 38(3), pp.22-26.
- [4] Tiwari, R., Bordoloi, D. J., Bansal, S. and Sahu, S., 2017. Multi-class Fault Diagnosis in Gears Using Machine Learning Algorithms Based on Time Domain Data. International Journal of COMADEM, 20(1).
- [5] Burges, C. J. C., 1998. A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery,2, pp. 121-167
- 6] Cheng, J., Yang, Y. and Yang, Y., 2012. A rotating machinery fault diagnosis method based on local mean decomposition. Digital Signal Processing, 22(2), pp.356-366.
- [7] He, Q., Yan, R., Kong, F. and Du, R., 2009. Machine condition monitoring using principal component representations. Mechanical Systems and Signal Processing, 23(2), pp. 446-466.
- [8] Wu, J.D., Hsu, C.C. and Wu, G.Z., 2009. Fault gear identification and classification using discrete wavelet transform and adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(3), pp.6244-6255.
- [9] Waqar, T.W., Demetgul, M. and Kelesoglu, C., 2015. Fault diagnosis on bevel Gearbox with neural networks and feature extraction. Elektronika ir Elektrotechnika, 21(5),pp.69-74.
- [10] Chen, K., Zhou, X.C., Fang, J.Q., Zheng, P.F. and Wang, J., 2017. Fault feature extraction and diagnosis of gearbox based on EEMD and deep briefs network. International Journal of Rotating Machinery, 2017, pp 1-10.
- [11] Jing, L., Zhao, M., Li, P. and Xu, X., 2017. A convolution neural network based and fault diagnosis method for the condition monitoring of gearbox. Measurement, 111, pp. 1-10.
- [12] Chen, Z., Li, C. and Sanchez, R.V., 2015. Gearbox fault identification and Classification with Convolutional Neural Networks. Shock and Vibration, 2015.
- [13] Luyang Jing, Taiyong Wang, Ming Zbao and Peng Wang., 2017. An Adaptive Multi-Sensor Data Fusion method Based on Deep Convolutional Neural networks for Fault Diagnosis of Planetary Gearbox, 2017, pp.414-428
- [14] Shaheryar, A., Yin, X.C. and Ramay, W.Y., 2017. Robust Feature Extraction on Vibration Data under Deep-Learning Framework: An Application for Fault Identification in Rotary Machines. International Journal of Computer Applications, 167(4), pp. 975-
- [15] Sun, W., Yao, B., Zeng, N., Chen, B., He, Y., Cao, X. and He, W., 2017. An Intelligent gear fault diagnosis methodology using a complex wavelet enhanced convolutional neural network. Materials, 10(7), pp. 1-18.
- [16] Alom, M. J., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., hasan, M., essen, B. C. V., Awwal, A. A. S and Asari, V. K., 2019. A State-ofArt Survey on deep learning Theory and Architectures. Electronics, 8(292), pp. 1-67
- [17] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.
- [18] Yildirim, Ö., 2018. A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in biology and medicine, 96, pp.189-202.
- [19] Baek, W., Baek, S. and Kim, D. Y., 2018. Characterization of System Status Signals for Multivariate Time Series Discretization Based on Frequency and Amplitude variation. Sensors, 18(1), pp. 1-13.
- [20] https://info.mide.com/data-loggers/vibration-analysis-files-download-page-thank-you?submissionGuid=f91dd92a-7bc4-4698-9db7-860e52fd20e4