**DOI 10.52209/2706-977X_2024_3_96** **IRSTI 55.03.33** **UDC 62-233.2**

# Remaining Useful Life Prediction Method For Different Types Of Rolling Bearings Based On Bi-Lstm Quantification

**Kondhalkar Ganesh Eknath*, Dr. G. Diwakar**
Koneru Lakshmaiah Education Foundation (Deemed to be University),Vaddeswaram, India
*corresponding author

**Abstract.** The remaining Useful Life (RUL) forecast for rolling bearings is still the crucial part of condition-based maintenance (CBM) for mechanical systems. To predict the RUL, the existing research utilized traditional Deep Learning techniques, however, it has trouble quantifying uncertainty. Therefore, this research suggested a novel DL model to improve RUL prediction. Initially, to define the degree of rolling bearing deterioration and comprehend the non-linear qualities, time domain features, frequency domain features, & time-frequency domain features are removed. Then, this study suggested using a Bi-LSTM - RF framework to predict the RUL, this framework has an LSTM layer in a combination of forward and backward motion, a fully connected layer, an RF classifier, & a dropout layer. As a result, our proposed deep learning-based RUL prediction obtains the Accuracy of 0.9845, Precision of 0.93, Recall of 1.0, & F1-score of 0.9656.

**Keywords:** bidirectional long short term memory (Bi-LSTM), deep learning, degradation features extraction, normalization, remaining useful life (RUL) prediction

## Introduction

Due to their significant impact on safety, production, and financial efficiency, equipment stability, and reliability are essential in many industrial domains. Reduced maintenance costs and unneeded downtime are advantages of Prognostic and Health Management (PHM), which is further known as terms like Condition-Based Maintenance (CBM) & predictive maintenance (PdM) [1-2]. In rotating equipment, rolling bearings are a common mechanical component that must frequently sustain a variety of mechanical and thermal pressures. Bearing problems account for more than 40% of motor failures. RUL forecasting, fault locating, and anomaly detection are all components of PHM for rolling bearings [3]. Utilizing historical trajectory data, one may evaluate and project roller bearings' RUL, it is crucial for maintaining mechanical material effectively during its service life. It is suggested to develop structure accessibility, dependability, & less expensive equipment maintenance. The Remaining Service Life (RSL) of the linked equipment before failure is considered to have happened is precisely specified as RUL [4]. Each of the maximum essential elements of spinning machinery, rolling bearings frequently has a straight effect on the security of the mechanical system's overall operation. As a result, production safety may be ensured and significant economic benefits can be realized through actual & precise RUL prediction of rolling bearings [5].

Three general categories may be used to classify RUL prediction approaches: physics model-based techniques, data-driven techniques, & hybrid techniques. To create a physical model, methodologies based on physical models investigate the components' damage mechanisms & the rules of degradation for certain fault modes. This typically calls for a variety of previous information, making it challenging to exactly create a deterioration model in difficult situations [6–9]. Data-driven techniques goal to change the information given by IoT into dependent models, either parametric or non-parametric, which only use sufficient previous data and do not require knowledge of specific deterioration processes [10,11]. They can show the fundamental relationships and causes between RUL and the raw sensor data. Data-driven methods are frequently utilized in industrial applications as a result of the quick development of intelligent technology [12].

The RUL is predicted using machine learning techniques that identify features from raw sensor data using signal processing methods and expert knowledge [13–16]. As a result, deep learning technology offers a potentially effective way to raise RUL prediction accuracy. To extract features from high-dimensional data, deep learning techniques like Recurrent Neural Networks (RNN), Long Short-Term Memories (LSTM), Gated Recurrent Units (GRU) for time series modeling, and convolutional neural networks (CNN) have been extensively used. Even though the deep learning neural network has demonstrated tremendous potential for RUL prediction, there are a few data- and practically-related difficulties that merit more thought. Since physical equipment operates under complicated settings, noise fluctuations and measurement mistakes must be present in the sensor streaming data. Therefore, the RUL prediction and feature extraction are greatly impacted by the changing variations.

The studies mentioned above have confirmed the potential of DL-based approaches for RUL prediction. But, the majority of these techniques use deterministic neural networks as their implementation, which eventually produces RUL point estimation. There are several different kinds of forecast ambiguities, including quantification uncertainty caused by interference from noise, model uncertainty relating to the forecast model, and unreliability environments caused by operation irrationality, which have an impact on RUL prediction in real applications [17]. Many important judgments are founded on the quantification of uncertainty. The anticipated cost of the RUL prediction point struggles to provide adequate direction for the maintenance strategy in real-world situations if the

uncertainty is not quantified. Therefore, this research proposed a bearing RUL prediction approach to address these issues. The following is the primary contribution of this study:

For RUL prediction, this research proposed a BiLSTM – RF framework, which consists of the following steps:

- the first step of this framework is the extraction of degradation features in rolling bearing. Here, we extract 13 frequency domain features to analyze the vibration signal, 12 time-domain features to represent the degree of rolling bearing degeneration, and 5 time-frequency domain features to comprehend its nonlinear characteristics;
- then, the extracted features are get normalized, then we create the sample data by using the sliding time window approach. Then, for prediction, this work employs a Bi-LSTM model, which employs LSTM layers in both forward and backward directions, a fully connected layer, an RF classifier, & a dropout layer.

As a result, when compared to the current methodologies, this suggested RUL prediction offers greater accuracy.

This research project is organized as follows: Segment 2 investigates an artificial intelligence-based method to predict the RUL. Following that, Segment 3 describes a proposed deep learning model to predict the RUL, and Segment 4 contains the simulation findings and comparison analysis of the suggested method. This research study is concluded in Segment 5.

**1. Literature survey**

Statistical models or Artificial Intelligence are typically used in data-driven ways to predict RUL. The construction of statistical models using empirical knowledge is the basis of statistical model-based approaches. A Wiener process-based real-time prognostic strategy for wind turbine bearing was introduced by Hu et al. [18]. The parameters of a Wiener process model are determined using maximum likelihood estimation. When combined the reverse Gaussian distribution, the RUL of a wind turbine bearing may determine. It is difficult to choose the right parameters for a given scenario because the statistical models' parameters vary depending on the scenario. An adaptive network-based fuzzy inference system (ANFIS)-based prognostic technique was presented by Wu et al. [19]. After the data pre-processing is complete, statistical features from the multi-sensor data are extracted. ANFIS and polynomial curve fitting is then used to fuse data from many sensors and estimate the RUL. For bearing under various operational circumstances, an RUL prediction technique was suggested by Kundu et al. [20]. This method establishes a Weibull accelerated failure time regression (WAFTR) model that acknowledges together operational state variables & monitoring signals. In conclusion, the outcomes demonstrate this strategy takes better forecast presentation. While allowing them to examine enormous amounts of data, the standard AI model's shallow designs restrict their capacity to learn complicated non-linear correlations.

Deep learning is already becoming a powerful technique for pattern detection and data prediction. DL is an addition of ANN that can extract data features from input data by performing deep analysis and mining using several hidden layers & nonlinear transformations [21]. Due to its remarkable representative feature capture capacity, it has established extensive use in the area of roller bearings RUL prediction. Hu et al. [22] utilized the DBN to construct the rolling bearing health index extractor, combining the network because of the diffusion process technique to forecast RUL. Deep convolution neural networks (DCNN), which can use several convolutional and pooling layers to extract hidden features from input data, were employed by Cheng et al. [23] to accomplish RUL prediction of rolling bearing. The RUL estimation issue was seen by the RNN network as a time series regression problem, which makes it a prime choice for DL techniques to analyze time series data and solve. According to CNN & the Bi-LSTM network, Zhao C. et al. [24] built the hybrid two-channel prediction model. Since multivariate degradation equipment stores information about various elements of degradation, degradation laws, life contribution rates, and coupling relationships between the various aspects are all diverse. The aforementioned model will inevitably have limitations when used to anticipate data from a single network. As a normal RNN, Bi-LSTM may successfully overcome the Temporal Convolution network's limitations, However, compared to convolution operation, its short-term memory is slower and less precise. The LSTM model & the AdaBoost regression model were combined to create a hybrid data-driven RUL prediction technique by Zhu et al. [25] depending on data trajectory expansion. The studies mentioned above have confirmed the potential of DL-based approaches for RUL prediction. RUL prediction in practical applications is impacted by a variety of prediction uncertainties, including measurement uncertainty caused by noise interference, model uncertainties relevant to the prediction model, & uncertainty conditions affected by operation haphazardness. The predicted RUL prediction point value struggles to provide adequate guidance for the preservation approach in real-world presentations if the uncertainty is not defined [26]. According to Li et al. [27], to deal with uncertainty in deep learning, Bayesian inference can be utilized as a learning method. A possible way for measuring uncertainty is the Bayesian approach.

Eknath et al. [28] DCNN & Gated Recurrent Units (GRU) are used in a unique prediction approach that uses feature extraction to derive vibration signal properties. Simultaneously, to estimate a rolling bearing's remaining usable lifespan, the Gated recurrent unit (BiGRU) has been used. However, its sluggish convergence rate and learning efficiency lead to an excessively extended training period.

According to the study and description above, it can be said that both traditional deep learning techniques are hindering RUL prediction at the moment and common uncertainty measurement techniques have a hard time

adapting to RUL prediction methods. These challenges must be overcome; this research suggests a bearing RUL prediction technique depends on RNN & unreliability quantification.

## 2. Proposed approach

RUL prediction is a crucial component to increase the dependability & availability of machines. The bearing deterioration rule was mined from operational data in the current study using DL to get the RUL. However, building a suitable DL model for precise RUL prediction is typically challenging due to the complexity of operating data. Also, external uncertainties have a big impact on bearing degradation, to overcome that existing research utilized traditional deep learning techniques, however, it has trouble quantifying uncertainty, and conventional approaches to measuring uncertainty have limited ability to accommodate RUL prediction techniques. In addition, the Bi-GRU model is used in the existing RUL prediction, due to difficulties including a poor convergence rate and low learning efficiency, training times are excessively long [28]. Therefore, this research proposed a novel DL-based RUL prediction which was described in the following section.

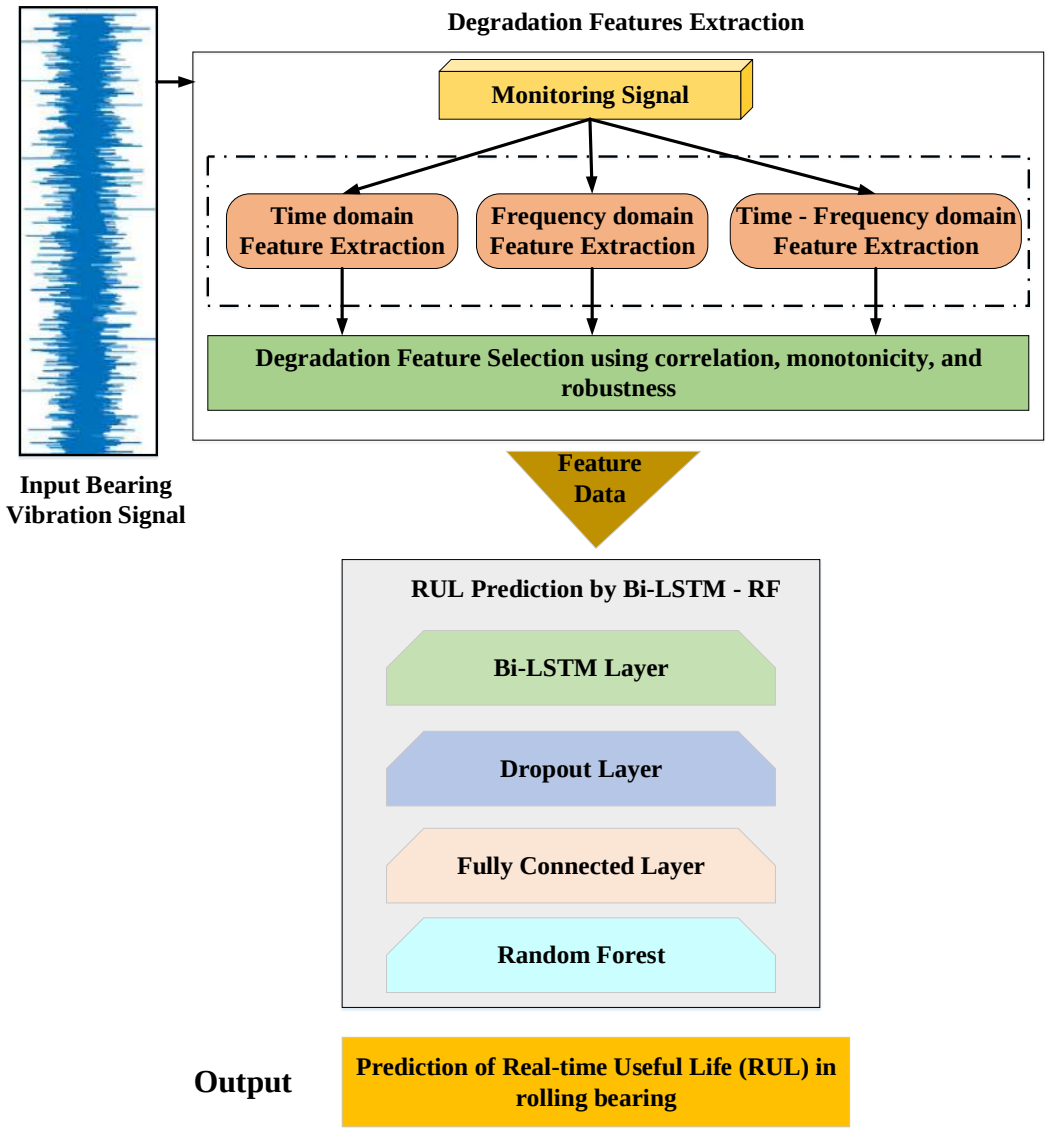


**Fig. 1**. - Architecture of the proposed RUL prediction approach

### 2.1 Building Feature Set

Since vibration signals may include a wealth of information, they are frequently chosen for rolling bearing deterioration process monitoring. Since the vibration signal cannot automatically portray the bearing's deteriorating method, the degradation characteristics must be removed from the raw vibration data. To replicate the degrading method of rolling bearings from various dimensions, time-domain features, frequency-domain features, & time-frequency domain features were retrieved in this research.

Different techniques may be used to obtain the deterioration characteristics of bearings in different domains. Even though frequency domain characteristics were typically statistical features depending on the Fast Fourier transform, which is useless for explaining how a single frequency changes over time, bearing deterioration is a non-

stationary dynamic method with substantial time correlation. Time-frequency domain features were chosen to take the transitory peculiarities in the bearing deterioration.

In contrast to its role in fault detection, RUL prediction must be compatible with the degradation process, which is impacted by failure types. As a result, some other characteristics that are only appropriate for certain failure modes remain not appropriate for RUL prediction. The characteristics that may accurately depict the deterioration process and have predictability are screened using three feature assessment indices: a correlation indicator, a monotonicity indicator, and a robustness indicator.

2.1.1 Time domain feature extraction

One of the easiest and most efficient techniques for analyzing vibration signals is the time domain deterioration characteristic. The vibration signal was statistically analyzed to determine the time-domain deterioration features. To represent the degree of rolling bearing degradation, 12 time-domain degradation characteristics are removed and displayed.

**Table 1.** Time domain degradation features

| S.No | Name | Formula |
|---|---|---|
| 1 | Absolute Mean | $T_1 = \frac{1}{N}\sum_{t=1}^{N}\lvert x(t)\rvert$ |
| 2 | Root Mean Square | $T_2 = \sqrt{\frac{1}{N}\sum_{t=1}^{N} x(t)^2}$ |
| 3 | Peak | $T_3 = max\ (x(t))$ |
| 4 | Square Root Amplitude | $T_4 = \left[\frac{1}{N}\sum_{t=1}^{N}\sqrt{\lvert x(t)\rvert}\right]^2$ |
| 5 | Skewness | $T_5 = \frac{1}{N}\sum_{t=1}^{N}(x(t) - -x)^3$ |
| 6 | Kurtosis | $T_6 = \frac{1}{N}\sum_{t=1}^{N}(x(t) - -x)^4$ |
| 7 | Waveform Indicator | $T_7 = \frac{T_2}{\lvert -x\rvert}$ |
| 8 | Peak Indicator | $T_8 = \frac{T_3}{T_2}$ |
| 9 | Impulse Indicator | $T_9 = \frac{T_2}{\lvert -x\rvert}$ |
| 10 | Margin Factor | $T_{10} = \frac{T_3}{T_4}$ |
| 11 | Kurtosis Factor | $T_{11} = \frac{T_6}{T_2}$ |
| 12 | Skew Factor | $T_{12} = \frac{T_5}{T_2}$ |

Table 1. T1–T12 in the table stand for time domain deterioration characteristics, $x(t)$, t =1,2,..N. N, where N is the total number of signal data values, stands for the signal data points in vibration signals. This characteristic capture the altering regulations of various bearing health states. For instance, the RMS value primarily imitates the amplitude of the overall energy of the entire bearing monitoring signal, whereas the maximum value depicts the effect force on the bearing at the specific value of failure, the peak factor can be used to define the signal's peak level for bearing monitoring.

2.1.2 Frequency domain feature extraction

It is required to do frequency domain analysis on the vibration signal to acquire the frequency domain deterioration aspect since the error features of rolling bearings were typically concealed in frequency domain information. Here, the monitoring signal is transformed into a frequency spectrum using the fast Fourier transform, & 13 frequency domain characteristics were then retrieved from the monitoring signal's frequency spectrum.

**Table 2.** Frequency domain degradation features

| S.No | Formula |
|---|---|
| 1 | $F_1 = \frac{\sum_{i=1}^{K} s(i)}{K}$ |
| 2 | $F_2 = \frac{\sum_{i=1}^{K} (s(i) - F_1)^2}{K - 1}$ |
| 3 | $F_3 = \frac{\sum_{i=1}^{K} (s(i) - F_1)^3}{K\left(\sqrt{F_2}\right)^3}$ |
| 4 | $F_4 = \frac{\sum_{i=1}^{K} (s(i) - F_1)^4}{K(F_2)^2}$ |
| 5 | $F_5 = \frac{\sum_{i=1}^{K} f_i s(i)}{\sum_{i=1}^{K} s(i)}$ |
| 6 | $F_6 = \sqrt{\frac{\sum_{i=1}^{K} {f_i}^4 s(i)}{\sum_{i=1}^{K} {f_i}^2 s(i)}}$ |
| 7 | $F_7 = \sqrt{\frac{\sum_{i=1}^{K} {f_i}^2 s(i)}{\sum_{i=1}^{K} s(i)}}$ |
| 8 | $F_8 = \sqrt{\frac{\sum_{i=1}^{K} (f_i - F_5)^2 s(i)}{K}}$ |
| 9 | $F_9 = \frac{\sum_{i=1}^{K} {f_i}^2 s(i)}{\sqrt{\sum_{i=1}^{K} s(i) \ \sum_{i=1}^{K} {f_i}^4 s(i)}}$ |
| 10 | $F_{10} = \frac{F_6}{F_5}$ |
| 11 | $F_{11} = \frac{\sum_{i=1}^{K} (f_i - F_5)^3 s(i)}{KFF_6^3}$ |
| 12 | $F_{12} = \frac{\sum_{i=1}^{K} (f_i - F_5)^4 s(i)}{KFF_6^4}$ |
| 13 | $F_{13} = \frac{\sum_{i=1}^{K} (f_i - F_5)^{\frac{1}{2}} s(i)}{KFF_6}$ |

These frequency domain properties are listed in Table 2, in which the vibration signal's frequency spectrum is depicted by s(i). $x(t)$, i = 1,2,3,.., K, the amount of spectral lines is K, & the frequency of the i$^{th}$ spectral line is represented by $f_i$. Features $F_1$- $F_5$ characterize alterations in the major frequency band location of the recorded signal in the frequency domain. The aspect domain spectral energy distribution's dispersion level is described by the values of $F_6$- $F_{13}$.

2.1.3 Time - Frequency domain feature extraction

Rolling bearing monitoring signals are often non-linear and non-stationary during real operation. Due to the complex interaction between time, frequency, and amplitude, the Time-Frequency (TF) domain analysis of bearing monitoring signals was used to identify the features of changes in bearing health status. In this work, a three-level wavelet packet decomposition technique called the Haar wavelet was utilized to break down the vibration signal into a collection of wavelet nodes.

Entropy is suggested as a way to gauge both the complexity of the data and the likelihood that a new signal model will emerge. The aim of assessing the sparsity is to provide a purpose, which expresses the sparse distribution of energy. The KL divergence of the multivariate joint probability density & its marginal probability density product was used to characterize the mutual information of several variables. A physical quantity called kurtosis has been presented as a way to assess how much a random variable has a Gaussian distribution.

Where $TF_1$ is represented as energy entropy, the percentage of each value in the time-frequency energy histogram is called $p_i$ , the energy entropy can also show the degree of uncertainty in the energy distribution. The time-frequency energy histogram is divided into m parts according to frequency, & TF2 is shown as the energy correlation coefficient. Where $corcoef(.)$ is the cross-correlation function. $TF_3$ is represented as energy Sparsity, where the sparsity is determined using x. $TF_4$ is represented as Energy Mutual Information, p(x) and q(x) are two distinguish probability density functions of the random vector x. $TF_5$ is represented as Energy Kurtosis, where $\mu$ means data mean, & $\sigma$ means standard deviation, and $x_i$ is the random variable with N observations (i=1,2,..., N).

**Table 3.** Time-Frequency domain degradation features

| S.No | Name | Formula |
|---|---|---|
| 1 | Energy Entropy | $TF_1 = -\sum_{i=1}^{n} p_i \log(p_i)$ |
| 2 | Energy Correlation Coefficient | $TF_2 = [corcoef(1), corcoef(2), \dots, corcoef(m)\ ]^T$ |
| 3 | Energy Sparsity | $TF_3 = \frac{\|X\|_p}{n^{1/p-1/2}.\|X\|_2}$ |
| 4 | Energy Mutual Information | $TF_4 = \int P(x) \log \frac{P(x)}{q(x)} dx$ |
| 5 | Energy Kurtosis | $TF_5 = \frac{\sum_{i=1}^{N}(x_i - \mu)/N}{\sigma^4} - 3$ |

2.1.4 Degradation feature selection

RUL prediction, in contrast to its function in fault detection, must be compatible with the deterioration process, which failure modes affect. As a result, a few different characteristics that are only appropriate for certain failure modes were not appropriate for RUL prediction. Using a correlation indicator and three feature evaluation indices Corr( f, t), monotonicity indicator, Mon( f ), and robustness indicator, Rob(f ), were employed to identify characteristics that can accurately and predictably reflect the deterioration process [29].

First, when assessing the monotonic growing or reducing the development of the degradation characteristics, the monotonicity is stated as:

$$Mon = \left|\frac{no.of\ dF>0}{L-1} - \frac{no.of\ dF<0}{L-1}\right| \quad (1)$$

L denotes the amount of signal models, & the derivative of the aspect sequence is called dF. $Mon$ denotes a quantity between 0 and 1, where 1 means the character is fully monotonous and 0 means it is constant.

The connection between the rolling bearing degradation characteristics and the degradation period is then evaluated using the Pearson correlation coefficient (PCC), which is written as:

$$PCC = \frac{|\sum_{t=1}^{T}(F_t - \bar{F})(L_t - \bar{L})|}{\sqrt{\sum_{t=1}^{T}(F_t - \bar{F})^2 \sum_{t=1}^{T}(L_t - \bar{L})^2}} \quad (2)$$

Where $F_t$represents the current value of the deterioration feature $t$ & $L_t$ displays the time value associated with the signal sample. The amount of signal samples is T. The PCC also has a 0–1 range. 1 indicates that the degradation characteristic and the degradation time are perfectly connected:

$$Robustness = Robustness = \frac{1}{T}\sum_{t=1}^{T} \exp\left(-\left|\frac{F_t}{\bar{F}}\right|\right) \quad (3)$$

To identify deterioration characteristics that can precisely depict the deterioration procedure of rolling bearings, three pointers are linearly integrated to create selection criteria (SC). The SC formula is shown as:

$$SC = \omega_1 Mon + \omega_2 PCC + \omega_3 Robustness \quad (4)$$

Where $\omega_i$ is the weighting coefficient.

**2.2 Standardizing Features and Constructing Samples**

The comparative size of various aspect measures has a significant impact on training. The characteristics with big magnitudes would be crucial in model training if screened features weren't regular. The characteristics with modest magnitudes, however, have trouble making it easier to update model parameters. As a result, since some crucial properties were not included in the model training process, it is problematic to train the model to its optimum form. Additionally, this would result in frequent gradient direction oscillations in the process of model optimization, which would hinder convergence and lengthen training. As a routine pre-processing step, normalizing the input characteristics into a machine learning model is done most frequently by employing the mean and standard deviation. The Z-score standardization criteria are requested to handle partitioned characteristics. The formulation of the Z-score standardization criterion is:

$$X_{stand} = \frac{X_{orig} - \mu}{\sigma} \quad (5)$$

where $X_{stand}$ and $X_{orig}$ are the original signal & the standardized signal, and the original signal's variance & mean, respectively.

In this research, model information that satisfies the input specifications of Bi-LSTM is created using the sliding time window approach in addition to normalizing the aspect signals. The input size and length of the sequence must be specified in the model input for the Bi-LSTM prediction model. After processing in the preceding stages, the sample data constitute a two-dimensional array of size $N_{sample\_point} \times N_{feature}$, where $N_{sample\_point} and\ N_{feature}$ is the quantity of features and sampling points, respectively. It is necessary to first identify the sample length that corresponds to the sequence length. The input sample is the distance of the sample, $N_{length}$, the information related to the model positions 0-$N_{length}$, and the sample label is the RUL of the $N_{length}$-th sampling point. To get a sequence of input samples, after that, a single unit is added to the sampling time axis to advance the interception window, & the previous action is frequent. Since neighboring samples built using sliding time windows overlap, the training set of data can be perfectly fitted using Bi-LSTM.

### 2.3 RUL prediction based on Bi-LSTM – RF Model

An LSTM layer, a fully connected layer, an RF classifier, & a dropout layer make up the proposed Bi-LSTM-RF-based prediction model in this research. The Bi-LSTM - RF's suggested design is seen in Figure 2.

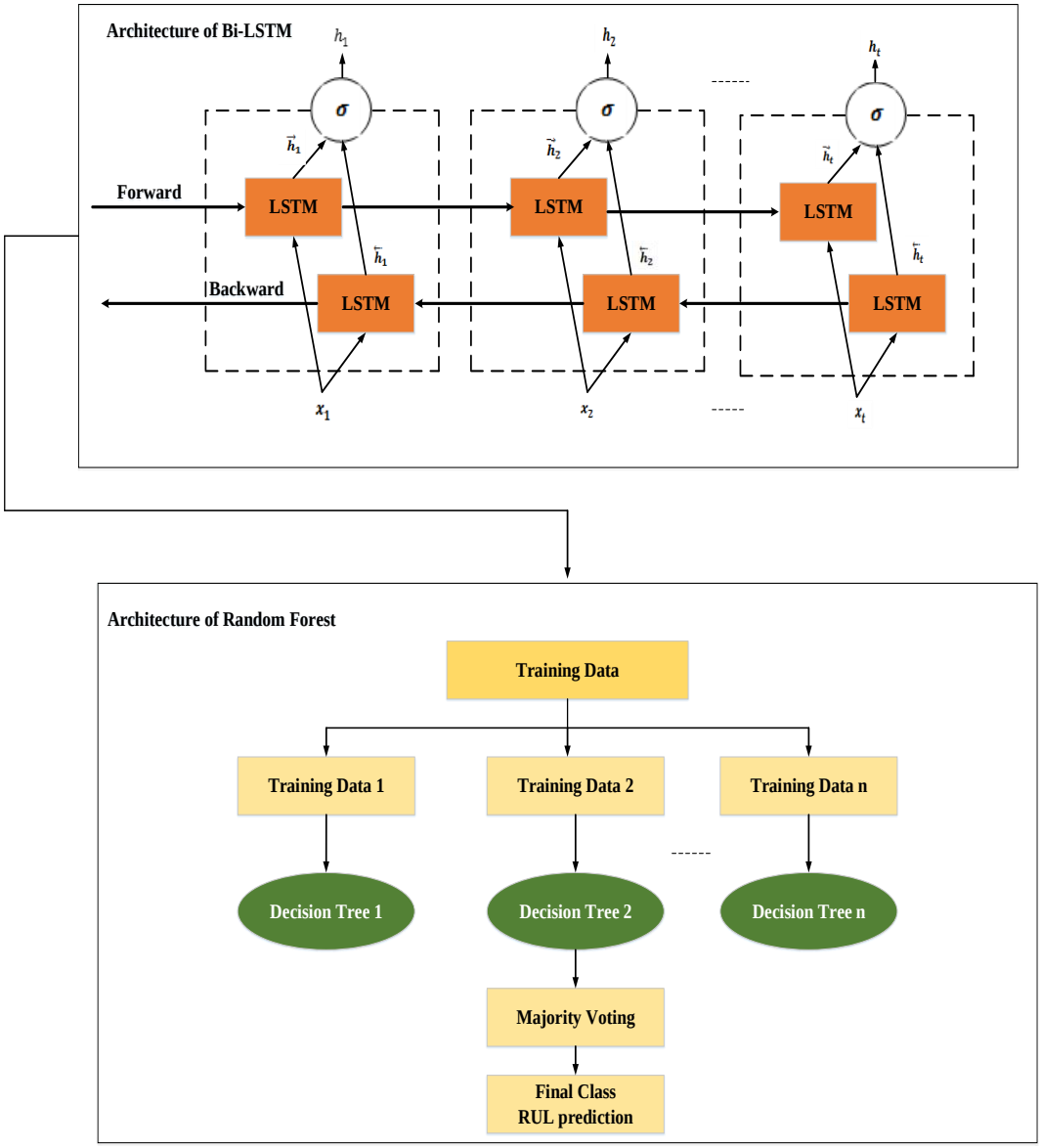


**Fig. 2. -** Architecture of the Bi-LSTM – RF

A Bi-LSTM is a bidirectional variation of the LSTM that can analyze lengthy data sequences by learning in each direction, including forward and backward. To efficiently learn the long-term data sequences in both forward and backward orientations, Bi-LSTM uses two distinct hidden layers. To access long-range information, it is desirable to

record two-direction contextual relationships. It is made up of two LSTMs, the first of which feeds the learning process in an onward manner, and the second of which is utilized to learn from the inputs in a backward (reverse) way.

Here, $\vec{h}$ and $\overleftarrow{h}$ are utilized to signify the onward and reverse hidden layers' respective outputs separately. The training method for both LSTM units makes use of the ordered input data sequences. The forward $\vec{h}_t$ and $\overleftarrow{h}_t$ LSTM layers' output is estimated using a recursive technique. The mode property, which includes the following potential merging approaches: average, sum, multiplied, and concat, is used to combine the output of both LSTM layers.

The design we've suggested calls for an average mode to combine results from onward and reverse LSTM units. In conclusion, the combined output is obtained using a flattened layer $f_{layer}$ which is then converted into a one-dimensional vector v and passed to the softmax function to get the desired result.

Equation 6 shows how the layer of Bi-LSTM concatenates the output sequences of both LSTM units using the merge mode approach to produce a two-dimensional output vector, Y, which is used to build bi-directional sequences.

$$\boldsymbol{y_t} = \propto \left(\vec{h}_t, \overleftarrow{h}_t\right) \tag{6}$$

Where $\propto$ denotes the merge mode approach applied to the $\vec{h}_t$ and $\overleftarrow{h}_t$ output sequences, respectively. To combine the output sequences of both the onward and reverse LSTM units, the symbol $\propto$ denotes an average mode method. The multiplication function, summation function, averaging function, or concatenating function can be used as the merge mode technique. The final result of the two LSTM units is shown as a one-dimensional vector, $Y = [y_1, y_2, \ldots, y_t]$, where the last component, $y_t$, is the best-predicted value for the following time iteration.

A tanh function serves as the activation function after the LSTM layer, followed by two linked layers on top of each other. The dropout layer was employed in the testing step to measure the prediction uncertainty and avoid over-fitting. LSTM between two layers, and two completely connected layers in this investigation, a dropout layer is placed.

Then, the RFC was utilized for the final RUL prediction. The ensemble learning idea, which is useful for prediction issues, is the foundation of RF. The ensemble learning idea, which is useful for prediction issues, is the foundation of RF. As the name suggests, RF is a classifier that improves classification accuracy by using multiple decision trees on different dataset subsets. Instead of relying just on one decision tree, the random forest together forecasts from all decision trees & predicts the eventual decision made based on the majority of estimates. As the number of trees rises, so does the precision and danger of over-fitting. As the number of trees rises, so does the precision and danger of over-fitting. Finally, to reduce the cost function, the Adam optimizer is utilized, & associated learning rate is set to 0.001.

**Table 4.** Results of the proposed method with different epochs

| Epoch | Learning rate | Dropout | MSE` | MAE |
|---|---|---|---|---|
| 10 | 0.001 | 0.2 | 2.3850 | 1.3009 |
| 20 | 0.001 | 0.2 | 2.3850 | 1.3009 |
| 30 | 0.001 | 0.2 | 2.12 | 1.2950 |
| 40 | 0.001 | 0.2 | 1.28 | 1.10 |
| 50 | 0.001 | 0.2 | 1.170 | 0.95 |

Additionally, for the suggested prediction model, the RUL point estimates & kernel distributions are obtained by calculating the uncertainty using nonparametric kernel density estimation and dropout. The dropout procedure resembles parallel network training with a predetermined network topology. As a consequence, when the same data are repeatedly input into a Bi-LSTM-based prediction model with operational dropout, approximately alternative forecast outcomes may be obtained, indicating the quantification of uncertainty. The mean value may be thought of as the RUL's point estimation. These prediction findings may be processed using the nonparametric kernel density estimation to get the kernel density distributions of the RUL at various model locations, which could have been utilized to provide judgments with a strong basis in uncertainty.

**3. Result and discussion**

This segment discusses the performance of our suggested explanation as well as the implementation results. Also mentioned are the comparison outcomes from the baseline method.

Tool: PYTHON 3
OS: Windows 7 (64-bit)
Processor: Intel Premium
RAM : 8GB RAM

**3.1 Dataset Description**

The IMS Bearing dataset, which was produced by the NSF 1/UCR Centre for Intelligent Maintenance System (IMS), was used in this study.

3.1.1 Experimental Setup

A shaft has four bearings placed on it. An AC motor connected to the shaft by rub belts controlled the revolution speed to remain constant at 2000 RPM. A spring mechanism provides a radial force of 6000 lbs on the shaft & bearing. All bearings were greased by force. As seen in Figure 3, Rexnord ZA-2115 double-row bearings were mounted on the shaft. PCB 353B33 The bearing housing was equipped with High Sensitivity Quartz ICP accelerometers (two accelerometers for all bearing (x- & y-axises) for data set 1, and one accelerometer for every bearing for data sets 2 & 3). Figure 1 illustrations display the positioning of the sensors. All failures happened next to the bearing had completed more than its intended lifetime above 100 million revolutions.

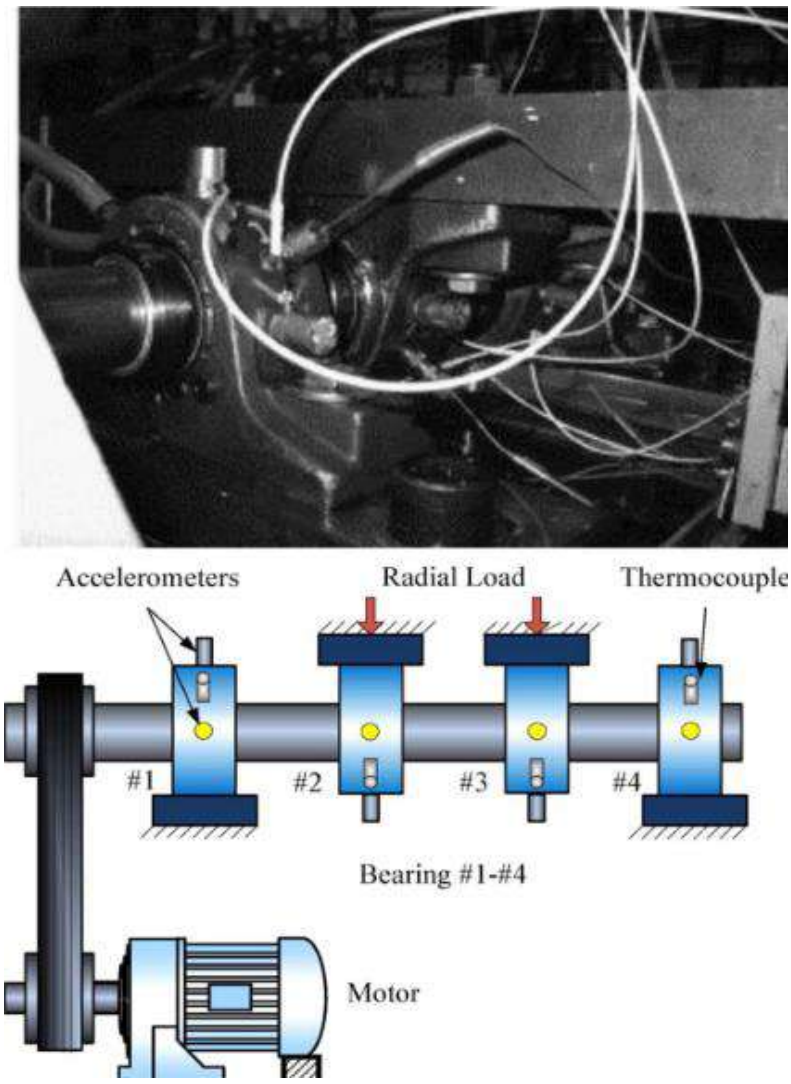


**Fig. 3.** Experimental Setup

**3.1.2 Structure of IMS data**

The data packet contains three (3) data sets. A test-to-failure experiment is described in each piece of data. Every data set was composed of separate files that are snapshots of a 1-second vibration signal taken at predetermined intervals. All file has a sampling rate of 20 kHz & a total point count of 20,480. The document name displays the date of data gathering. A record (row) in a file that contains information is known as a data point. NI DAQ Card 6062E enabled data collecting. The research was restarted the following working day, as shown by longer time stamp intervals (seen in file names).

**Table 5.** Dataset Details

| Parameters | Set No.1 | Set No.2 | Set No.3 |
|---|---|---|---|
| Recording Duration | October 22, 2003 12:06:24 to November 25, 2003 23:39:56 | February 12, 2004 10:32:39 to February 19, 2004 06:22:39 | March 4, 2004 09:27:46 to April 4, 2004 19:01:5 |
| No. of files | 2,156 | 984 | 4448 |
| No. of channels | 8 | 4 | 4 |
| Channel Arrangement | Bearing 1 – Ch 1&2; Bearing 2 – Ch 3&4;<br>Bearing 3 – Ch 5&6;<br>Bearing 4 – Ch 7&8. | Bearing 1 – Ch 1;<br>Bearing 2 – Ch 2;<br>Bearing 3 – Ch 3;<br>Bearing 4 – Ch 4; | Bearing 1 – Ch 1;<br>Bearing 2 – Ch 2;<br>Bearing 3 – Ch 3;<br>Bearing 4 – Ch 4; |
| File Recording Interval | Except for the first 43 files, which were captured every 5 minutes, every 10 minutes. | Every 10 minutes | Every 10 minutes |
| File Format | ASCII | ASCII | ASCII |
| Description | An inner race flaw and a roller element defect both appeared in bearings 3 and 4 after the test-to-failure experiment. | After the test-to-fail experiment, bearing 1 experienced outer race failure. | After the test-to-fail experiment, bearing 3 experienced outer race failure. |

3.1.3 Training and Testing Loss of Dataset

The training loss & testing loss of dataset 1 is 0.0278, and 0.0275 at epoch 34, dataset 2 is 0.0038 and 0.0037 at epoch 42, and dataset 3 is 0.0255 and 0.0252 at epoch 41 by using our proposed Bi-LSTM – RF which is presented in figure 4 (a), (b) and (c). As a result, dataset 2 has less training and testing loss when compared to the existing dataset.

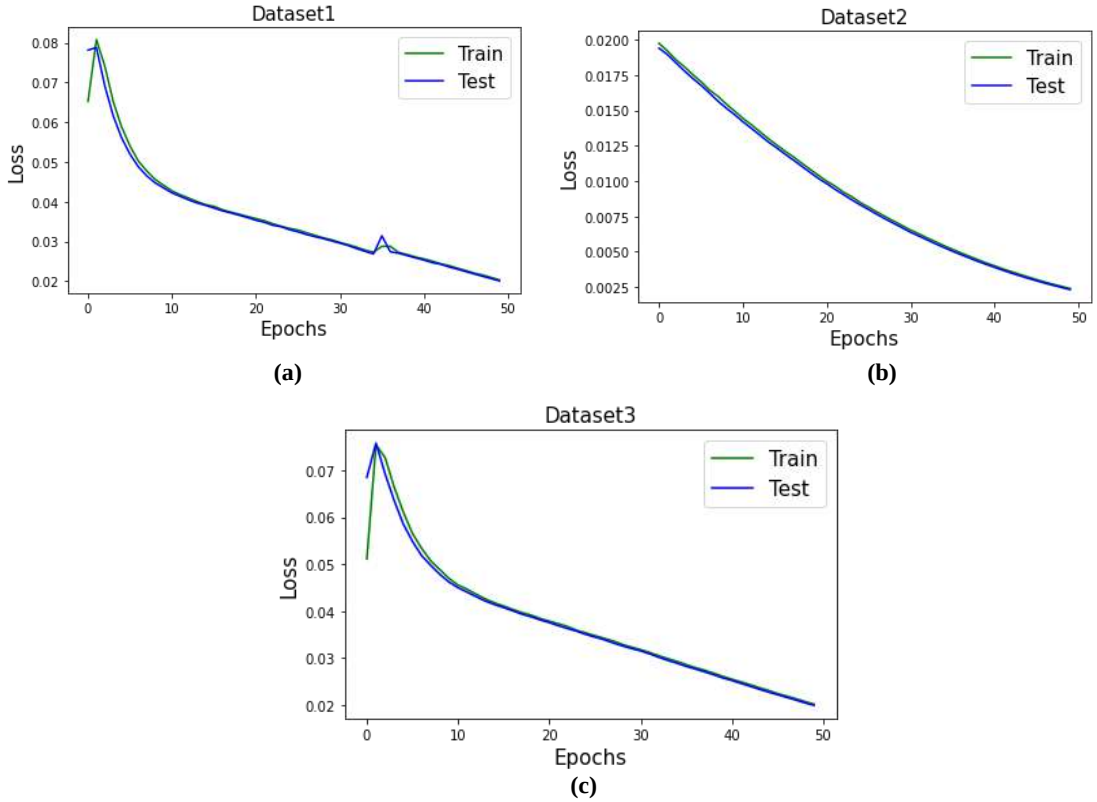


**Fig. 4.** Train and Test Loss of the dataset

### 3.2 Parameter Settings

**Table 6.** Parameter Settings of the proposed approach

| Algorithm | Parameters | Value |
|---|---|---|
| | Learning rate | 0.001 |
| | Batch Size | 64 |
| | Dropout | 0.2 |
| | Testing Size | 0.25 |
| | Training Size | 0.75 |
| Bi-LSTM | Input Size | 64×25×9 |
| | LSTM | (9,9) |
| | Dropout | 0.2 |
| | LSTM | (9,9) |
| | Dense | (9× 25 × 100), linear |
| | Dropout | 0.2 |
| | Dense | (100,1) |
| | Output | (64,1) |
| | Optimizer | Adam |
| RF | Random state | 1 |
| | Max_depth | 2 |
| | Max_features | 30 |

### 3.3 Experimental Results

3.3.1 Results of degradation features

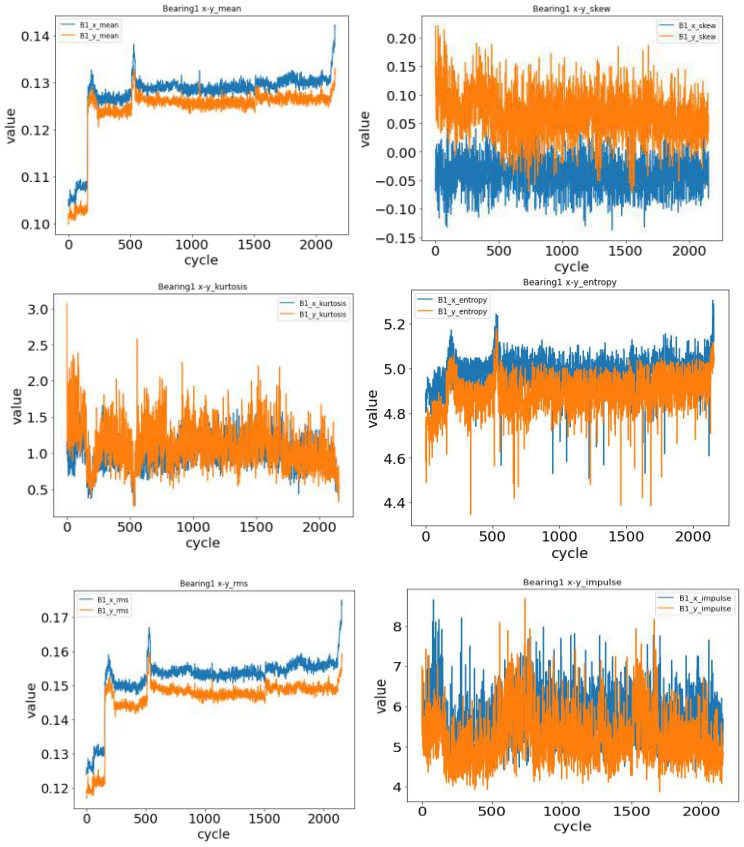


**Fig. 5.** Results of degradation features

This research presented some of the degradation features which are depicted in Figure 5. To define the degree of rolling bearing degeneration and comprehend its nonlinear characteristics, these features are retrieved. The chosen degradation features can adequately capture the deterioration of rolling bearings. Hence, the identified features are accurate and predictable depicting the deterioration process.

3.3.2 RUL Prediction Results

**Table 7.** RUL prediction results

| S.No | Bearing No | Cycle | Prediction | Is valid | (Normal / Suspected) |
|---|---|---|---|---|---|
| (a) | Bearing 1 | 2206 | 1950 | False | Suspected |
| (c) | Bearing 2 | 2156 | 1800 | False | Suspected |
| (b) | Bearing 3 | 2231 | 2281 | False | Imminent failure |
| (d) | Bearing 4 | 2256 | 2256 | False | Suspected |

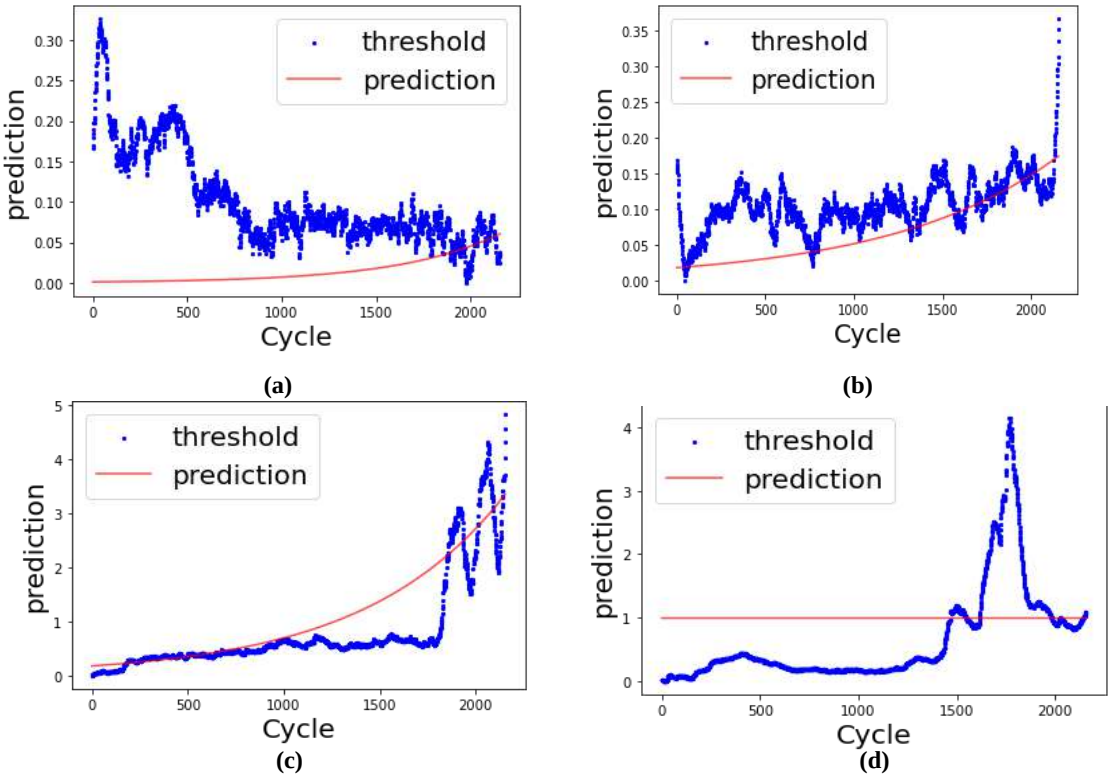


**Fig. 6.** Results of RUL Prediction

The RUL prediction consequences are exposed in Figure 6. To predict the RUL this research utilized a Bi-LSTM model, which employs LSTM layers in both forward and backward directions, a dropout layer, an RF classifier, & a fully connected layer. The Adam Optimizer is utilized to modify the Bi-LSTM network's parameters. Additionally, dropout & nonparametric kernel density estimation is combined for the proposed prediction model to quantify the uncertainty, results in the kernel distributions, and RUL point predictions.

### 3.4 Performance Parameters

Stated figure 7 is a performance measurement for the proposed Bi-LSTM - RF. The following equations are used to obtain the f1-score, which is very helpful for assessing accuracy, recall, precision, and other metrics.

$$\text{Accuracy} = \frac{\text{True Positive}+\text{True Negative}}{\text{True Positive}+\text{True Negative}+\text{False Positive}+\text{False Negative}} \quad (7)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive}+\text{False Positive}} \quad (8)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive}+\text{False Negative}} \quad (9)$$

$$\text{F1 Score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision}+\text{Recall}} \quad (10)$$

Several metrics, including Accuracy, F1 score, precision, & recall have been used to evaluate the effectiveness of our suggested method's performance in forecasting RUL.

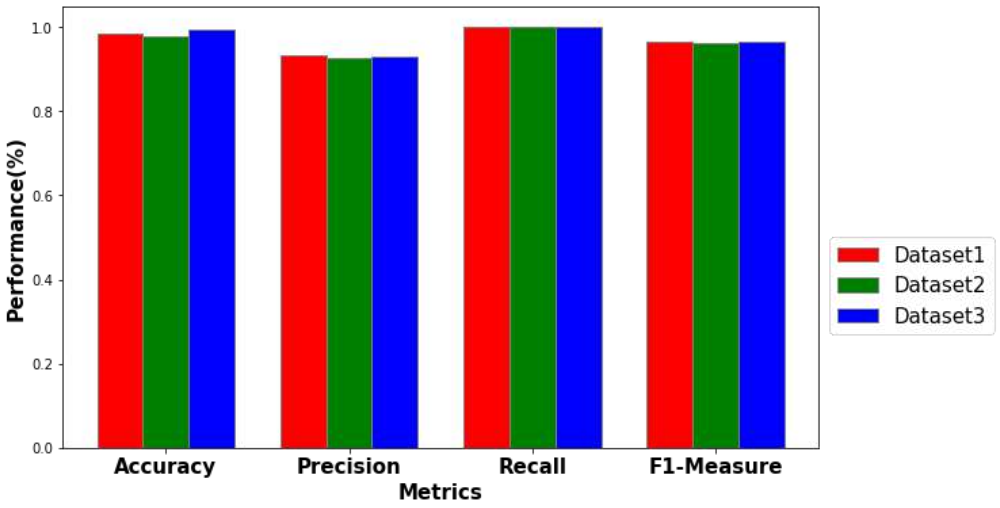


**Fig. 7. -** Performance metrics of the proposed approach

The suggested method's presentation evaluation measures are shown in Figure 7. Accuracy, Precision, Recall, and F1-score were measured and found to be respectively 0.9845, 0.93, 1.0, and 0.9656 (dataset 1). The obtained values (dataset 2) of Accuracy, Precision, Recall, and F1-score are 0.9786, 0.912, 1.0, and 0.952. The obtained values (dataset 3) of accuracy, precision, recall, and f1-score are 0.9953, 0.93, 1.0, and 0.965. By utilizing a unique Bi-LSTM - RF methodology for RUL prediction in rolling bearings, our suggested method performs better in terms of Accuracy, F1 score, Precision, & Recall.

The following formula is used to assess index model parameters like MAE and RMSE:

$$\text{MAE} = \frac{\sum_{i_z=1}^{m_z} \left| y_{zi_z,f_z} - \overline{y_{zf_z}} \right|}{m_z} \quad (11)$$

where $y_{zi_z,f_z}$ = prediction, $\overline{y_{zf_z}}$= true value, $m_z$ = there are a total of N data points.

$$\text{RMSE} = \sqrt{\frac{\sum_{i_z=1}^{m_z} (x_{zi_z,f_z} - \overline{x_{zf_z}})^2}{m_z}} \quad (12)$$

where $x_{zi_z,f_z}$ = actual observation time series,$m_z$ = number of non-missing data points, $\overline{x_{zf_z}}$ = estimated time series, $i_z$= variable $i_z$

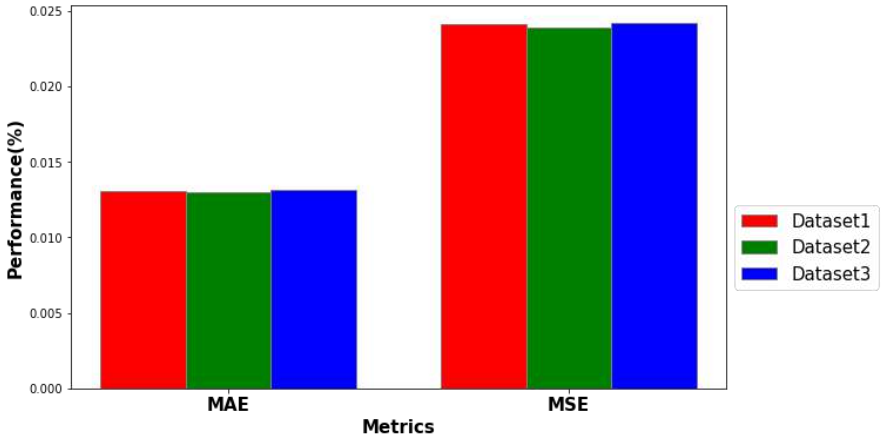


**Fig. 8.** Error analysis of the proposed approach

Figure 8 illustrates the comparison using sample assessment indices such as MAE. The suggested method reduces the error by incorporating Bidirectional LSTM with RFC. For dataset 1, the MAE and RMSE errors are 0.012 and 0.0243, for dataset 2, 0.011 and 0.0239, and dataset 3, 0.013 and 0.0245.

**3.5 Comparison analysis**

This segment compares the current methodologies, putting the suggested method up to standard methods like LSTM with random features [29], Convolutional Neural Network [29], Adaptive Kalman Filter (AKF) [29], LSTM with uncertainty quantification [29], and DCNN-Bi-GRU [28].

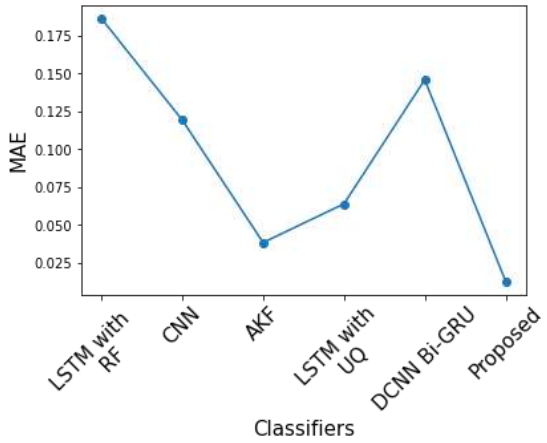


**Fig. 9.** Comparison analysis of MAE

Figure 9 illustrates the MAE of the suggested approach. By combining Bi-LSTM with a Random Forest classifier, a suggested method lowers the error. Our suggested method related to the baseline LSTM with random features, Convolutional Neural Network, AKF, LSTM, and DCNN-Bi-GRU with uncertainty quantification such as 0.1864, 0.1613, 0.0640, 0.0567, and 0.146. The MAE of the proposed approach is 0.013. As a consequence, our suggested strategy makes improved than the current methods.

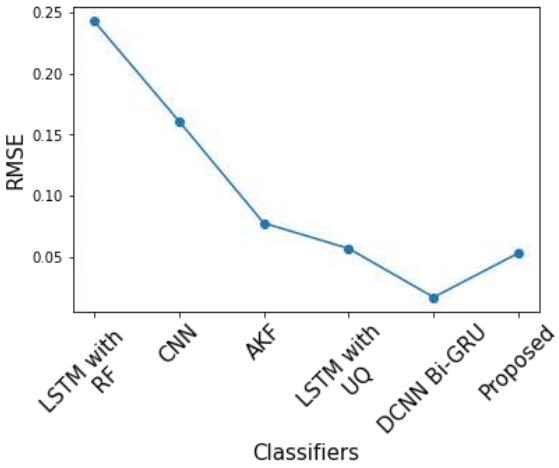


**Fig. 10. -** Comparison analysis of RMSE

Figure 10 illustrates the RMSE of the suggested approach. By combining Bi-LSTM with a Random Forest classifier, the suggested method lowers the error. Our suggested method related to the baseline LSTM with random features, Convolutional Neural Network, AKF, LSTM with uncertainty quantification, and DCNN-Bi-GRU such as 0.2463, 0.1613, 0.0775, 0.0567, and 0.167. The RMSE of the proposed approach is 0.053. As a result, our suggested strategy performs better than the current methods.

**Conclusion**

This research suggested a unique dL model develop the RUL prediction model's accuracy. To predict the RUL, Bi-LSTM – RF and uncertainty quantification is proposed. Initially, to define the level of rolling bearing degradation, in this study, 12 time-domain features, 13 frequency-domain features, & 5 time-frequency domain features were retrieved. Moreover, to identify features that may accurately depict the deterioration process, three

feature assessment indices correlation, monotonicity, and robustness are used. The proposed Bi-LSTM - RF model is then fed the normalized input characteristics to most accurately forecast the RUL. As a result, this proposed approach obtained a greater accuracy, precision, recall, & f1-score of 0.9845, 0.93, 1.0, & 0.9656. Moreover, this research obtained a lesser error such as MAE of 0.013, and RMSE of 0.053 when compared to the existing approaches. In the future, this research may concentrate on a hybrid deep learning-based framework to predict the RUL with the highest accuracy.

**References**

[1] de Pater, I., Reijns, A., & Mitici, M. Alarm-based predictive maintenance scheduling for aircraft engines with imperfect Remaining Useful Life prognostics. Reliability Engineering & System Safety, 221, 2022, 108341.
[2] Kondhalkar, G. E., & Diwakar, G. Crest factor measurement by experimental vibration analysis for preventive maintenance of bearing. In ICRRM 2019–System Reliability, Quality Control, Safety, Maintenance, and Management: Applications to Civil, Mechanical and Chemical Engineering, 2020, 133-138
[3] Kondhalka, G., & Diwakar, E. Effect of various defects in roller bearings and ball bearings on vibration. Int. J. Innov. Technol. Explor. Eng, 8(12), 2019, 5137-5141.
[4] Zhang, P., Gao, Z., Cao, L., Dong, F., Zou, Y., Wang, K. & Sun, P. Marine systems and equipment prognostics and health management: a systematic review from health condition monitoring to maintenance strategy //Machines, 10(2), 2022, 72.
[5] Cao, Y., Ding, Y., Jia, M., & Tian, R. A novel temporal convolutional network with residual self-attention mechanism for remaining useful life prediction of rolling bearings //Reliability Engineering & System Safety, 2021, 215, 107813.
[6] Cao, Y., Jia, M., Ding, P., & Ding, Y. Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network //Measurement, 178, 2021, 109287.
[7] Chen C, Liu Y, Wang S, Sun X, Di Cairano-Gilfedder C, Titmus S, et al. Predictive maintenance using Cox proportional hazard deep learning //Adv Eng Informatics, Volume 44, April 2020, 101054
[8] Aremu O.O., Hyland-Wood D, McAree PR. A Relative Entropy Weibull-SAX framework for health indices construction and health stage division in degradation modeling of multivariate time series asset data //Adv Eng Informatics 2019.
[9] Caravaca CF, Flamant Q, Anglada M, Gremillard L, Chevalier J. Impact of sandblasting on the mechanical properties and aging resistance of alumina and zirconia-based ceramics //J Eur Ceram Soc, 2017, 38(3)
[10] Ding P, Jia M, Yan X. Stationary subspaces-vector autoregressive with exogenous terms methodology for degradation trend estimation of rolling and slewing bearings //Mech Syst Signal Process, 202, 150.
[11] Ding P, Jia M, Wang H. A dynamic structure-adaptive symbolic approach for slewing bearings' life prediction under variable working conditions //Struct Heal Monit 2020, Volume 20, Issue 1
[12] Meissner, R., Rahn, A., Wicke, K. Developing prescriptive maintenance strategies in the aviation industry based on a discrete-event simulation framework for post-prognostics decision-making. Reliability Engineering & System Safety, 214, 2021, 107812.
[13] C. Ordo´nez, F. S. Lasheras, J. Roca-Pardinas, F. J. de Cos Juez, A hybrid arima–svm model for the study of the ˜ remaining useful life of aircraft engines //Journal of Computational and Applied Mathematics 346, 2019, 184–191.
[14] Z. Chen, Y. Li, T. Xia, E. Pan, Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy //Reliability Engineering & System Safety 184, 2019, 123–136.
[15] Yang, B., Liu, R., & Zio, E. Remaining useful life prediction based on a double-convolutional neural network architecture //IEEE Transactions on Industrial Electronics, 66(12), 2019, 9521-9530.
[16] Z. Liu, Y. Cheng, P. Wang, Y. Yu, Y. Long, A method for remaining useful life prediction of crystal oscillators using the Bayesian approach and extreme learning machine under uncertainty //Neurocomputing 305,2018, 27–38.
[17] Liu, Z.; Cheng, Y.; Wang, P.; Yu, Y.; Long, Y. A Method for Remaining Useful Life Prediction of Crystal Oscillators Using the Bayesian Approach and Extreme Learning Machine under Uncertainty. Neurocomputing 2018, 305, 27–38.
[18] Hu Y., Li H., Shi P., Chai Z., Wang K., Xie X., Chen Z. A prediction method for the real-time remaining useful life of wind turbine bearings based on the Wiener process //Renew Energ 127, 2018, 452–460.
[19] Wu J., Su Y., Cheng Y., Shao X., Deng C., Liu C. Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network-based fuzzy inference system //Appl Soft Comput 68, 2018,12–23.
[20] Kundu P, Darpe AK, Kulkarni MS Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions //Mech Syst Signal Process, 2019, 134:106302.
[21] Cheng Y, Hu K, Wu J, Zhu H, Shao X, Auto-encoder quasi-recurrent neural networks for remaining useful life prediction of engineering systems // IEEE/ASME Trans Mechatron,
[22] Hu C., Pei H., Si X., Du D., Pang Z., Wang X.A. prognostic model based on DBN and diffusion process for degrading bearing. IEEE Trans Ind Electron 67(10), 2020, 8767–8777.
[23] Cheng H., Kong X., Chen G., Wang Q., Wang R. Transferable convolutional neural network based remaining useful life prediction of bearing under multiple failure behaviors //Measurement, 168, 2020, 108286.

[24] Zhao, C., Huang X., Li, Y., Yousaf Iqbal M. A. double-channel hybrid deep neural network based on CNN and Bi-LSTM for remaining useful life prediction //Sensors 2020, 20, 7109.
[25] Zhu X., Zhang P., Xie M. A Joint Long Short-Term Memory and AdaBoost Regression Approach with Application to Remaining Useful Life Estimation //Measurement 2020, 170, 108707.
[26] Biggio, L., Wieland, A., Chao M.A., Kastanis, I., Fink O. Uncertainty-aware Remaining Useful Life predictor. arXiv 2021, arXiv:2104.03613.
[27] Li G., Yang L., Lee C.G., Wang X., Rong M. A Bayesian Deep Learning RUL Framework Integrating Epistemic and Aleatoric Uncertainties //IEEE Trans. Ind. Electron. 2020, 68, 8829–8841
[28] Eknath K. G., Diwakar G. Prediction of Remaining useful life of Rolling Bearing using Hybrid DCNN-BiGRU Model //Journal of Vibration Engineering & Technologies, 2022, 1-14.
[29] Yang J., Peng Y., Xie J., Wang P. Remaining Useful Life Prediction Method for Bearings Based on LSTM with Uncertainty Quantification //Sensors, 22(12), 2022, 4549.

**Information of the authors**

**Kondhalkar Ganesh Eknath**, PhD, professor, Koneru Lakshmaiah Education Foundation (Deemed to be University)
e-mail: kondalkarganesheknath@gmail.com

**Garikapati Diwakar,** PhD, professor, Koneru Lakshmaiah Education Foundation (Deemed to be University)