Koc. J. Sci. Eng., 7(2): (2024) 137-150 https://doi.org/10.34088/kojose.1426751

Kocaeli University

Kocaeli Journal of Science and Engineering

http://dergipark.org.tr/kojose

A Remaining Useful Life Approach using an Ensemble Regressor enhanced with Hilbert Transform and CNN-LSTM Model

F. Azeez ADEBAYO [1,*], Kaplan KAPLAN [2], H. Metin ERTUNC [3]

[1] *Department of Mechatronics Engineering, Kocaeli University, Kocaeli, 41001, Turkey,* ***ORCID:*** *0000-0002-8160-6949*
[2] *Department of Software Engineering, Kocaeli University, Kocaeli, 41001, Turkey,* ***ORCID:*** *0000-0001-8036-1145*
[3] *Department of Mechatronics Engineering, Kocaeli University, Kocaeli, 41001, Turkey,* ***ORCID:*** *0000-0003-1874-3104*

Article Info

Research paper

Received : *January 27, 2024*
Accepted : *April 25, 2024*

Keywords

Predictive Maintenance
CNN-LSTM
Feature Extraction
RUL
Hilbert Transform

Abstract

In this study, we propose a hybrid approach that integrates signal-driven and knowledge-based techniques to estimate the Remaining Useful Life (RUL) of bearings. The experimental data for this research is sourced from the FEMTO-ST Institute. Firstly, the horizontal and vertical acceleration data is ordered chronologically by time, and a band-pass filter is used for early-stage preprocessing of the vibration signals below 20 kHz. Then, the overall behavior of the signal is characterized by Hilbert-Transform. For the feature extraction scheme, a model that integrates Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks is implemented. These features form historical data on health indexes describing fault stages and are as such used to fit a voting regressor yielding an extrapolated future. The voting regressor is based on support vector regression (SVR) and linear regressor methods and a fault threshold is determined as 0.8 based on prior experiments. Finally, the proposed methodology distinguishes itself by recording the smallest average percentage error on the FEMTO dataset. This method proves that early-stage predictions are possible with run-to-failure data provision ranging from 60% and above, averaging some 1400 seconds into the future implying its suitability and effectiveness for real industrial applications.

1. Introduction

Various Modern industrial machines predominantly operate using rotational motion, where bearings play a pivotal role. Bearings are important components of these machines, and their reliable operation is essential for the overall performance and longevity of the systems. Despite being relatively low-cost elements, any failure in these components can lead to catastrophic damage to the entire system. The health and uninterrupted performance of mechanical systems depend on identifying potential malfunctions at an early stage, enabling predictive maintenance. To achieve this, substantial efforts have focused on precisely estimating the Remaining Useful Life (RUL) of rolling bearings, which is fundamental for efficient predictive maintenance [1-4].

Methods for predicting RUL can be generally divided into two categories: those based on knowledge or analytical approaches [1,2] and those that rely on data [3,4]. By combining physical and dynamic phenomena with system measurements such as parameter estimation, state observer, and system identification, analytical-based techniques use mathematics to simulate the actual behavior of the system. For instance, in a study by Lei et al., a novel health indicator named the weighted minimum quantization error was introduced, and the maximum-likelihood prediction algorithm helped to initialize the model parameters. They used finally a particle filtering-based algorithm for predicting RUL [9]. Analytical-based approaches are the least used techniques as they involve heavy mathematical computation and as such involve a lot of linear approximations of non-linear real systems which in turn introduces some errors in their results making them not suitable for large industrial plants. Also, defining a precise and convenient model is a hard process and requires prior knowledge and equipment of the history of the mechanical system. Defining a Health Indicator (HI) and predicting model parameters from real-time information are the two main problems of model-based

[1] Corresponding Author: hazeezadebayo@gmail.com

solutions [10-12].

On the other hand, the data-based approach uses online monitored sensor data or signals and tries to imitate the reasoning of humans in fault diagnostics rather than the mathematical modeling of the system. Increasing learning capabilities, the development of sensor technologies, and the development of signal processing techniques have made data-based methods more applicable and studies on them have increased. It includes methods such as fuzzy systems [13], neural networks [14], and signal processing methods. This approach can be used in various industry applications. Data-driven approaches are typically bifurcated into two steps: feature extraction and model development. Obtaining characteristic and representative features is a challenging problem in this approach. Features derived from time-domain statistics, frequency-domain analysis, and time-frequency domain analysis are commonly used in contemporary studies. Chen et al. utilized five energy values from specific frequency bands as features. They developed a recurrent neural network model based on an encoder-decoder architecture, supplemented with an attention mechanism, to predict the Health Index (HI). According to the result, their study achieved the best performance among the studies they compared [15]. Yan et al. proposed two measures based on the Root Mean Square (RMS): the Inertial Relative Root Mean Square (IRRMS) and the Relative Root Mean Square (RRMS). They employed a machine learning technique known as the support vector machine to categorize the degradation stage of the bearing. According to their finding, they obtained effective results for classifying the degradation process [16]. Wu et al. derived features from the time and frequency domains of acoustic, current, vibration, and temperature sensor data to estimate the RUL of the shaft production system. They utilized a multi-stage approach, which was based on models for classification and regression. They noted an improvement in accuracy by 6.5% [17]. Li et al. utilized the Root Mean Square (RMS) and kurtosis as features to characterize the state of degradation. They implemented a multilayer artificial neural network (ANN) and a regression model to forecast the RUL of rolling element bearings [18]. Wu et al. obtained statistical features, intrinsic energy features, and frequency features of fault using Hilbert–Huang transforms (HHT) methods and ensemble empirical mode decomposition (EMD). They employed dynamic principal component analysis and Mahalanobis distance methods to select important features. They predicted the degradation process using exponential regression-based and empirical Bayesian algorithms [19]. Meng et al. analyzed degradation trends using gray markov model and empirical mode decomposition based on differential (DEMD). They reported that by using mathematical morphology fractal spectrum parameters they obtained a 4% decrease in the root mean square error (RMSE) value of degradation trend [20]. Ahmad et al. used Relative Root Mean Square (RRMS) as the HI indicator, and they predicted RUL using a dynamic regression model [21]. Soualhi et al. combined the Hilbert-Huang transform (HHT) and the Support Vector Regression (SVR) for bearing RUL prediction. [22].

The models mentioned so far use manual computation or signal processing to obtain the features. The computation of these features is very time-consuming and needs theoretical background. Today, deep learning algorithms that do not contain these problems have been developed. One of these model implementation fields is the detection, diagnostic, and prognostic of the bearing degradation process. Ren et al. proposed a hybrid algorithm utilizing deep neural networks (DNN) and autoencoders. To illustrate the degradation process, they utilized a unique eigenvector that was derived from joint features in the time-frequency-wavelet domain. They compressed features by using a deep autoencoder. They employed DNN for RUL prediction [23]. Lu et al. extracted bearing features based on the Autoencoder (AE). Subsequently, they employed a Gated Recurrent Unit (GRU) for the RUL prediction. [24]. Peng et al. presented an unsupervised health indicator using a deep belief network with a particle filter for the prediction of the RUL [25]. Zhao et al. developed a unique two-channel hybrid model for RUL prediction, which was based on a Capsule Neural Network and a Long Short-Term Memory Network (Cap-LSTM). They extracted spatial feature data from multivariate time-series sensory data. They observed superior success in their study [26]. Li et al. used information from the time-frequency domain and CNN to predict the degradation process of bearings [27]. Wang et al. Introduced a novel deep learning algorithm named deep separable convolutional network (DSCN) [28].

In the context of this analysis, the Hilbert Transform is employed to derive an envelope signal from the vibration data with a CNN-LSTM sub-model learning the signal features and forming a fault index representation. The learned features are fitted into a voting regressor which then helps to extrapolate into the future, based on the fitted data.

The key contributions of this research can be outlined as follows:

A novel hybrid approach for remaining useful life estimation is introduced, aiming to predict the onset of degradation in a precise and prompt manner.

. The health indicator of the study can illustrate the phase of regular operation and the phase of hastened degradation of the bearings based on the degradation initial

time.

The proposed method is benchmarked against other cutting-edge methods using the dataset to validate its effectiveness and fitness for the industry.

The flexibility of the algorithm suggests its potential for application to other types of machinery and under varying operating conditions. ultimately, it ensures a possible application to other domains.

The proposed method harnesses the strengths of its constituent models: it employs CNN for effective feature extraction, leverages LSTM for capturing temporal dependencies, and utilizes voting regressors to achieve a balance between overestimation and underestimation through consensus-based extrapolation.

The rest of this research paper is as follows: Section 2 provide overview of the theoretical background. Section 3, the dataset is described. Section 4 illustrates the proposed methodology. Section 5 gives equations of performance metrics. Section 6 gives the results of experimental studies and comparative findings. Section 7 draws some discussions, Sections 8 and 9 present the conclusions and future work, respectively.

2. Theoretical Background

2.1. Band Pass Filtering via Finite Impulse Response (FIR) Method

The periodic signal, x(t), can be expressed as a linear combination of complex exponential functions, each multiplied by its frequency response. This concept, represented in Eq. (1)-(6), underscores the idea of filtering, where we convolute the signal with an impulse response function to selectively attenuate or pass specific signal frequencies by setting certain frequency components to “0” or “1”.

$$x(t) = \sum_{k=-\infty}^{\infty} c_k\, e^{jk\omega_0 t} \quad (1)$$

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau \quad (2)$$

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, e^{j\omega_0 (t-\tau)}\, d\tau \quad (3)$$

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, e^{j\omega_0 t} e^{-j\omega_0 \tau}\, d\tau \quad (4)$$

$$y(t) = e^{j\omega_0 t} \int_{-\infty}^{\infty} h(\tau)\, e^{-j\omega_0 \tau}\, d\tau \quad (5)$$

which can be also re-written as a sum involving Fourier series coefficients:

$$y(t) = \sum_{k=-\infty}^{\infty} c_k\, e^{jk\omega_0 t} H(jk\omega_0) \quad (6)$$

where c_k is the Fourier series co-efficient and can be found using the Fourier analysis equation, ω_0 is the fundamental frequency, $H(jk\omega_0)$ accounts for the frequency response at each harmonic, and where $y(t)$ represents the filtered output signal. FIR bandpass filters, which are typically below 20 kHz, are used in bearing life prediction to obtain the frequency components of the vibration signal that are most informative for predicting bearing failure. The ideal filter response amplitude is a rectangular function with a passband centered on the bearing's natural frequency. In this study, the ideal filter response is defined as in Figure 1 with the Nyquist frequency, $f_{Nyquist}$ normalized to "1".

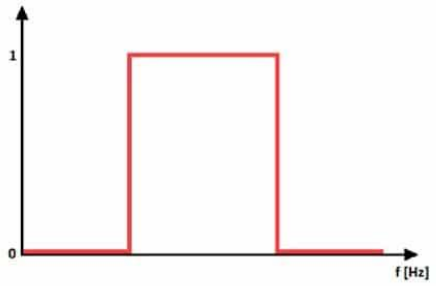

Figure 1. Ideal filter response.

However, to avoid ringing in the output signal, it is important to have a smooth shift between the pass band and stop band of the filter. In this study, we use a transition zone of 20% of the cutoff frequency between the points representing upper and lower bounds. The number of time points in the filter kernel, also referred to as filter order, is calculated using Eq. (8):

$$f_{Nyquist} = \frac{f_s}{2} \quad (7)$$

$$N = \frac{3 x f_s}{f_{LB}} \quad (8)$$

$$f_{filter} = \frac{1}{f_{Nyquist}} \begin{bmatrix} 0 \\ (1-w) x f_{LB} \\ f_{LB} \\ f_{UB} \\ (1+w) x f_{UB} \\ f_{Nyquist} \end{bmatrix} \quad (9)$$

where the f_s is sampling rate defined as the frequency at which the vibration signal is sampled. N is filter order, f_{LB} is lower filter bound, f_{UB} is upper filter bound, and w is transition width.

2.2. The Hilbert Transform

The use of the FIR bandpass filter facilitates narrow band isolation, enabling the extraction of the analytic-complex (real and imaginary) valued time series- signal. The resulting envelope signal has extreme values for phase and amplitude. Generally, for a given time series signal

$x(t)$, the Hilbert transform, $\hat{x}(t)$ is given as in Eq. (10).

$$\hat{x}(t) = H[x(t)] = \frac{\int_{-\infty}^{\infty} \frac{x(t)}{t-\tau} d\tau}{\pi} = x(t) * \frac{1}{\pi t} \quad (10)$$

Here, $x(t)$ and $\hat{x}(t)$ constitute a pair of complex conjugates. Hence, we may define the analytic signal as in Eq. (11).

$$z(t) = x(t) + j\hat{x}(t) = a(t)exp\,[j\theta(t)] \quad (11)$$

The polar coordinate expression can be described as the optimal local approximation of a trigonometric function, $x(t)$ as in Eq. (12)-(13) which exhibits variations in amplitude and phase.

$$a(t) = \left[x^2(t) + \hat{x}^2(t)\right]^{\frac{1}{2}} \quad (12)$$

$$\theta(t) = arctan\frac{\hat{x}(t)}{X(t)} \quad (13)$$

Here, $a(t)$ represents the instantaneous amplitude and θ signifies instantaneous phase. The definition of the instantaneous frequency is provided in Eq. (14) and the Hilbert envelope spectrum is given by Eq. (15) [30].

$$\omega = \frac{d\theta}{dt} \quad (14)$$

$$h(f) = \int_{-\infty}^{\infty} \sqrt{x^2(t) + \hat{x}^2(t)}\, e^{-j2\pi ft}\, dt \quad (15)$$

2.3. Development of a model architecture combining Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM).

It is helpful to think of the CNN + LSTM as defining two sub-models. CNNs are good at feature extraction and hence are naturally used as image encoders. The 2D Convolution network as used in this study comprises of the application of 2D kernel, or stacks of kernels as in filter as seen in Eq. (16), to capture certain features, and followed by MaxPooling2D layer organized into a layered structure of the necessary depth. This ensures that the information being abstracted is greatly reduced in size and yields an efficient interpretation as the internal matrix or vector representation.

$$l_t = tanhx_t * k_t + b_t \quad (16)$$

where l_t represents the output value after convolution, tanh is the activation function, x_t is the vector of the input values, k_t is the weight of the convolution kernel, and b_t is the bias of the convolution kernel.

In this bearing RUL prediction algorithm, portions of the processed 1D vector representing vibration measurements (i.e. acceleration records) are transformed into interpretable 2D images using spectrograms. A spectrogram serves as a graphical illustration of a signal's frequency content over a period of time. The x-axis denotes time (in seconds), while the y-axis signifies frequency (in Hz). Each pixel in the spectrogram represents the energy or power at a specific frequency and time. The resulting spectrogram is normalized to ensure consistent intensity levels across different signals.

The output generated by the CNN encoder is channeled into the LSTM decoder. The LSTM serves to interpret the features across time steps and generates a fault index. The fault index represents values between "0" and "1", indicating the probability of the bearing being faulty. However, we can only achieve this transfer from CNN to LSTM by encapsulating the entire CNN input model within a time-distributed layer, enabling the desired outcome to be achieved.

. This layer functions as if the same layer is applied multiple times, which is necessary due to the time step variable required by the LSTM. Or we may reshape the output of the CNN model to the required or desired number of time steps as we have done in this study. The LSTM memory cell consists of four parts: the forget gate, Eq. (17), the input gate, Eq. (18), cell state, Eq. (20), and the output gate, Eq. (21) assembled as depicted in Figure 2 [31].

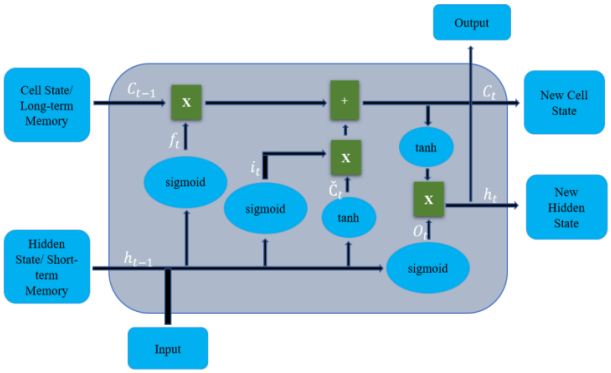

Figure 2. The structural design of the LSTM model.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (17)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (18)$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (19)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (20)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (21)$$

$$h_t = o_t * tanh\,(C_t) \quad (22)$$

Here the value range of f_t, i_t, C_t, o_t is "$(0,1)$", with values closer to zero implying to be unimportant or dropped and closer to one implying to be important or kept. The C_t represents candidature in which values are regulated between "$(-1,1)$" to help regulate the network. The weight matrix is represented by W_f, W_i, W_c, W_o and the bias vectors are denoted by b_f, b_i, b_c, b_o. The current time's input value is represented x_t, and h_{t-1} is the output value of the previous time or previous hidden state. While h_t is the output value of the current time or current hidden state. The subscripts f, i, C and o represent forget gate, input gate, cell state, and output gate.

2.4. The voting regressor model

The primary objective of employing a voting regressor model is to facilitate the extrapolation of future fault indices. The Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are utilized for feature extraction from the vibration signals, which are subsequently transformed into spectral images. These images, representing varying stages of bearing faults, serve as training data for the CNN-LSTM model to generate a fault index, a vector mapping that reflects the degradation state of the bearing based on the spectral image analysis.

However, the CNN-LSTM model, while adept at generating a fault index, does not inherently possess predictive capabilities. Consequently, when the objective is to predict future states, particularly when the available data does not encompass a run-to-failure sequence, extrapolation becomes necessary. This is achieved through the voting regressor, which, based on SVR and linear regression methods, extrapolates and reaches a consensus on the potential failure point of the bearing. This is accomplished by fitting the CNN-LSTM output data (the fault indices) and the respective index of the data files.

Each base regressor within the voting regressor model independently fits the dataset, and the final prediction is computed using a weighted average. This approach enables the prediction of future fault indices, thereby facilitating the estimation of the bearing's failure point.

2.4.1. Support Vector Regression (SVR)

SVR is a powerful regression technique rooted in the theory of Support Vector Machines (SVMs). SVR operates by mapping input data into a high-dimensional feature space using a kernel function and then finding a hyperplane [32] that fits the data in this space. The objective of SVR is to find a function $f(x)$, defined in Eq. (23) that has at most ϵ deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible.

$$f(x) = \langle w,x \rangle + b \quad (23)$$

Here, ⟨w,x⟩ denotes the dot product of the weight vector w and the predictor or feature vector x, and b is a bias term. The "flatness" of the function f(x) which describes the hyperplane, as measured by the norm of the weight vector w, is crucial in this context because it determines the generalization capability of the model. In the context of this study, the weight vector w corresponds to the coefficients of the SVR model that are learned from the features extracted from the health index and data mapping. The constraints for the SVR problem are defined in Eq. (24) as:

$$|f(x_i) - y_i| \leq \epsilon, \forall i \quad (24)$$

Where ϵ is deviation or threshold value, and y_i represents actual targets.

2.4.2. Linear Regressor (LR)

Linear Regression (LR) is a statistical technique that involves fitting a linear equation to minimize the overall distance measured as the sum of squares between the observed target values and the predicted values of the approximation. If we denote the dependent variable as y and the independent variable as x, the linear regression model [33] can be expressed in Eq. (25) as

$$y = b_o + b_1 x + \epsilon \quad (25)$$

Where $b_o$is the y-intercept, b_1 is the slope of the regression line (regression coefficient), and ϵ is the error term.

3. Dataset

To validate our hypothesis and facilitate performance comparison, we utilized the FEMTO-bearing public dataset, obtained from a test rig, illustrated in Figure 3. The experiment encompassed three distinct operating conditions, detailed in Table 1.

Table 1. An overview of the dataset from the IEEE 2012 PHM prognostic challenge.

Operation conditions	Conditions 1	Condition 2	Condition 3
Rotation speed	1800 rpm	1650 rpm	1500 rpm
Loading force	4000 N	4200 N	5000 N
Training set	Bearing1_1	Bearing2_1	Bearing3_1
	Bearing1_2	Bearing2_2	Bearing3_2
Test set	Bearing1_3	Bearing2_3	Bearing3_3
	Bearing1_4	Bearing2_4	
	Bearing1_5	Bearing2_5	
	Bearing1_6	Bearing2_6	
	Bearing1_7	Bearing2_7	

This dataset contains run-to-failure tests of 17 bearings, 6 for the purpose of training, and 11 sets for testing. The vibration data is then captured by accelerometers installed on one or more orthogonal axes. The signal was sampled at 25.6 kHz enabling it to be digitally reconstructed for vibration analysis. Vibration samples were collected every 10 seconds, for a period of 0.1 seconds. More detailed information experiments are given in [34].

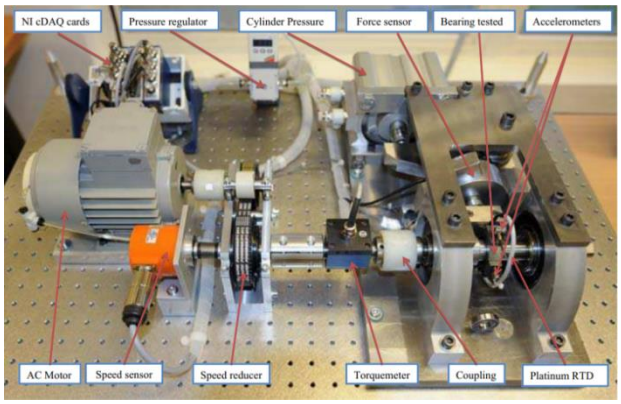

Figure 3. The PRONOSTIA experimental setup

4. Proposed Methodology

In the literature, there are many monitoring condition techniques, including vibration, temperature, chemical, and current monitoring. In the case of vibration, the analysis is typically performed directly on the time waveforms of the vibration signal, or alternatively, on the frequency spectrum., or sometimes a combination of both time and frequency spectrum. In chronologically recorded vibration waveform analysis in the time domain, parameters such as RMS, standard deviation, kurtosis, skewness, and others are analyzed to detect abnormal vibration events. This approach can generally assess the fault condition. In contrast, especially in rotating machinery, a waveform does not display the individual frequencies at which vibration occurs. It is thus not as easy to diagnose bearing problems using waveforms. The usual step is to perform a Fast-Fourier Transform (FFT), or other different advanced signal processing techniques such as Short-Time-Fourier-Transform (STFT), Wigner–Ville distribution (WVD), Hilbert Transform (HT), Hilbert–Huang Transform (HHT) and Wavelet Packet Transform (WPT) algorithm to convert the time waveform into a vibration frequency spectrum i.e., the energy distribution of the signal over the frequency domain at every instant of time describing the repetitiveness of vibration patterns [35].

The proposed technique to predict bearings RUL is a hybrid approach that combines both signal-based and knowledge-based techniques to analyze vibration patterns. Chronologically ordered acceleration data undergoes preprocessing, including bandpass filtering and Hilbert Transform for enveloping signal derivation. The resulting signal feeds into a CNN-LSTM model, predicting fault indices based on indicative degraded condition features. In Figure 4 and Figure 5, some sample of run-to-failure horizontal vibration signals of bearing is illustrated.

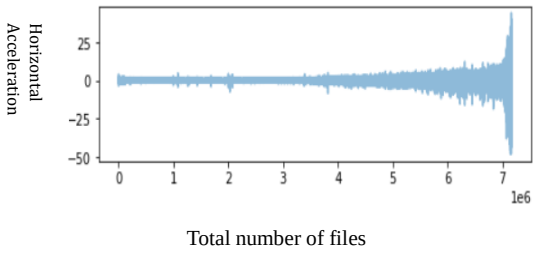

Figure 4. Run-to-failure horizontal vibration signals of bearing B1_1 in FEMTO dataset.

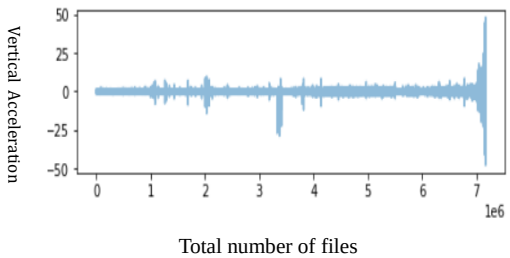

Figure 5. Run-to-failure vertical vibration signals of bearing B1_1 in FEMTO dataset.

However, since our data is in the time domain, we first output a real-valued narrowband signal to serve as the expected HT input through the computation of filter coefficients for the least square linear-phase finite impulse response (FIR) filter, we achieve an optimal approximation to the anticipated frequency response, as delineated by our

desired band in the temporal domain. In essence, the frequency and amplitude attributes of the resultant filter align with the specifications of our original design. The resulting filter, its frequency, and time-domain representation plot and the generated narrowband data are given in Figure 6.

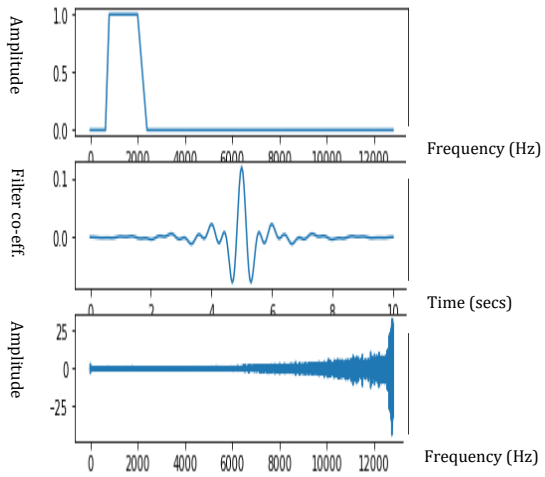

Figure 6. (a). Filter kernel representation frequency-domain, (b). Filter kernel representation in time-domain (c). Filtered Signal.

The Filter-Hilbert approach is next and is selected for its superior control over frequency filtering, shape flexibility, and high stability. With HT being used here for signal demodulation, a section of the modulated fault signals of bearing B1_1 shown in Figure 7(d) (solid orange curve) which are separated, and a carrier envelope obtained by HT plotted in blue curve better summarizes the signal.

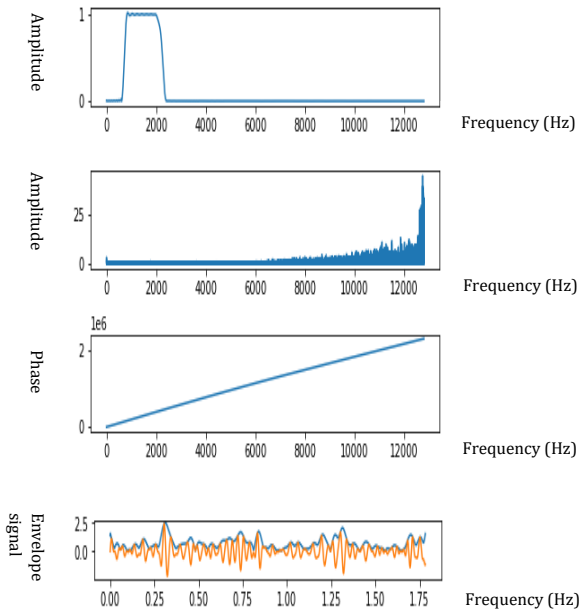

Figure 7. (a). Filter frequency response (b). Hilbert analytic signal real (c.) instantaneous phase plot (d). Envelope function and narrowband signal -cut section-.

A frequency spectral image of the nature given in Figure 8 is also generated per data file i.e., data collected in one single sampling. The FEMTO data files are 10 seconds apart and a file is recorded for 0.1 second. The red/blue color varies relative to the damage intensity.

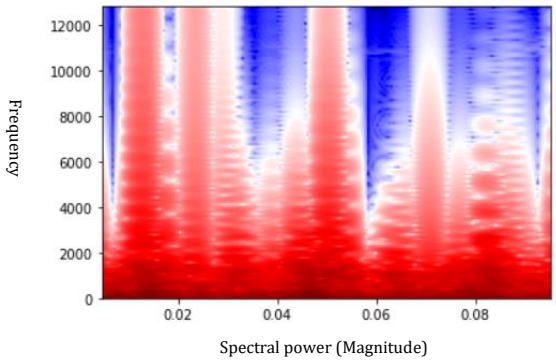

Figure 8. Frequency spectral image.

Table 2 delineates the structural composition of the CNN-LSTM model. Batch normalization is also used to maintain the distribution of the data and as such improves model performance, and mitigates internal co-variate shift, while also applying a regularization effect in the CNN sub-model. Conversely, random unit dropout is implemented in the LSTM sub-model to prevent over-fitting by selectively deactivating a predetermined proportion of network units. The final output shape of the CNN-LSTM model is "1" representing the fault index.

Table 2. The hybrid CNN-LSTM model architecture

Layer Type	Output Shape	Number of Parameters
Convolution	124×191×8	152
Batch normalization	124×191×8	32
Max Pooling	62×95×8	0
Convolution	60×93×16	1168
Batch normalization	60×93×16	64
Max Pooling	30×46×16	0
Convolution	28×44×24	3480
Batch normalization	28×44×24	96
Max Pooling	14×22×24	0
Convolution	12×20×24	5208
Batch normalization	12×20×24	96
Max Pooling	6×10×24	0
Flatten	1440	0
Dense Layer	256	368896
Dropout Layer	256	0

Table 2. (Cont.) The hybrid CNN-LSTM model architecture

Layer Type	Output Shape	Number of Parameters
Reshape Layer	256×1	0
LSTM Layer	256×60	14880
Dropout Layer	256×60	0
LSTM Layer	60	29040
Dropout Layer	60	0
Dense Layer	30	1830
Dropout Layer	30	0
Dense Layer	10	310
Dense	1	11

The proposed model was trained using 3000 epochs upon which a training loss value of 1.7906e-04 was obtained. The model, when evaluated against the training data, again yielded the results depicted as blue for predicted values and orange for actual values in Figure 9:

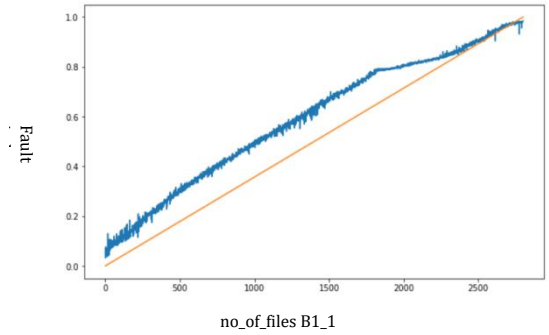

Figure 9. Model's performance on training data

An algorithm breakdown is given in Table 3 showing how the input, being data files of real-time measured horizontal and vertical frequencies, is processed, learned and converted to fault indexes before being extrapolated to any desired time instance in the future.

Table 3. Algorithm breakdown of the bearing RUL prediction

Algorithm step	Explanation
Initialize:	Set data points per one data file, sampling frequency, root file director, and FIR bandpass filter parameters.
Input:	Get current training data.
Compute:	Determine the data order by date/time and sort, apply FIR band pass filter and calculate the filter Hilbert envelope of the narrowband signal.
For	i in data_files:

Table 3. (Cont.) Algorithm breakdown of the bearing RUL prediction

Algorithm step	Explanation
Input:	Obtain the horizontal and vertical envelope signal.
Compute:	Generate the frequency spectral image matrix.
	Create train data from image matrix as X and assign a label i/data files as Y.
End For	
Repeat	
Start	Train CNN LSTM model with X and corresponding Y
Until	A satisfactory number of epochs
Examine	Make predictions with CNN LSTM model with Test Set.
Examine	Fit (predictions, data file index) to the Voting Regressor model and make new predictions on any length. i.e., informed extrapolation.
Visualize	Prediction graph: CNN LSTM in red, Voting Regressor model in blue, Highest point on Voting Regressor model in green (vertical), and fixed fault threshold in green (horizontal).

5. Performance Metrics

To evaluate the forecasting effect of CNN-LSTM, the Huber loss is selected as it is a combination of the mean square error (MSE) and absolute value loss. It tries to highlight the best of both losses and combines them in a differentiable way, making for a good evaluation criterion. The Huber calculation formula is shown in Eq. (26), for each error value z.

$$L_i = \begin{cases} \frac{1}{2}z^2 & for|z^{(i)}| \le \delta \\ \delta|z| - \frac{1}{2}\delta^2 & for|z^{(i)}| > \delta \end{cases} \quad (26)$$

where δ is a float chosen as the point where the Huber loss function changes from quadratic to linear.

First, the percentage error in prediction is defined in Eq. (27).

$$Er_i = 100x\frac{ActRUL_i - \hat{RUL}_i}{ActRUL_i} \quad (27)$$

where $ActRUL_i$ is the actual remaining useful life of the bearing to be predicted, and $\hat{RUL}_i$ is the algorithm's estimated or predicted value.

Then a score, A_i which is a measure of the early prediction capability of the prediction is defined in Eq.

(28). It ensures that underestimates are considerably weighted more and as such represent good performance in RUL prediction (i.e., where Er_i>0), while overestimates are weighted less and are regarded as late predictions. (i.e., where Er_i<0). As such, the total score for the experiment is an average and is also given in Eq. (29).

$$A_i = \begin{cases} exp\left(-ln(0.5)\cdot(Er_i/5)\right) & if\, Er_i \le 0 \\ exp\left(+ln(0.5)\cdot(Er_i/20)\right) & if\, Er_i > 0 \end{cases} \quad (28)$$

$$Score = \frac{1}{N}\sum_{i=1}^{N} A_i \quad (29)$$

where N, represents the aggregate count of test instances.

However, the score defined above is not a measure of the algorithm's prediction accuracy as it only favors cases of early prediction. For this reason, two other measures are further defined; the mean error and the mean absolute error as a measure of the prognosis method's accuracy, Eq. (30) and precision, Eq. (31) respectively as defined in [15].

$$\overline{Er} = \frac{1}{N}\sum_{i=1}^{N} Er_i \quad (30)$$

$$\overline{|Er|} = \frac{1}{N}\sum_{i=1}^{N} |Er_i| \quad (31)$$

where N defined previously, is the total number of test cases. Other important performance metrics [36] that have been used on FEMTO dataset are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Symmetric Mean Absolute Percentage Error (SMAPE), and Relative Accuracy (RA).

6. Experimental Results

Overall, results obtained from the experiment on the test data in the first operating condition are also depicted in Figures 10-14. In the figures, the CNN LSTM model outputs are shown in red, the Voting Regressor model output is shown in blue, the Highest point on the voting regressor is shown as a green vertical line, and the fixed fault threshold is shown in a green horizontal line.

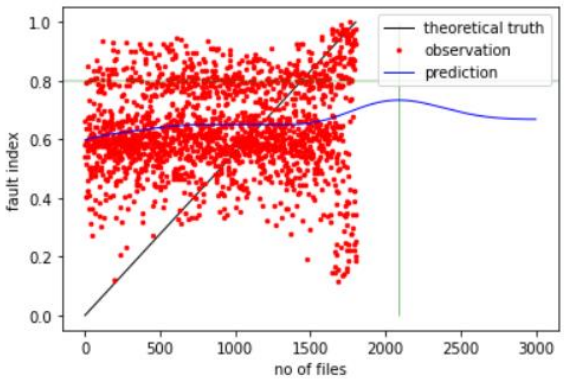

Figure 10. B1_3 (pred 2088, actual 1802)

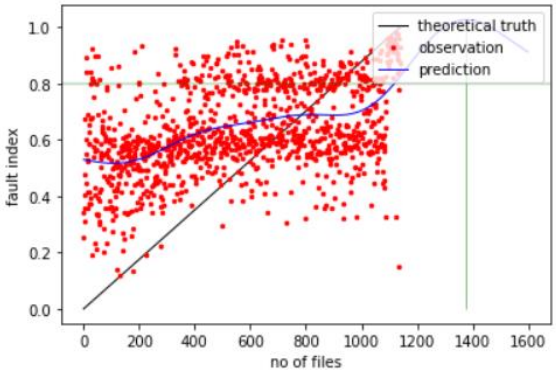

Figure 11. B1_4 (pred 1376, actual 1139)

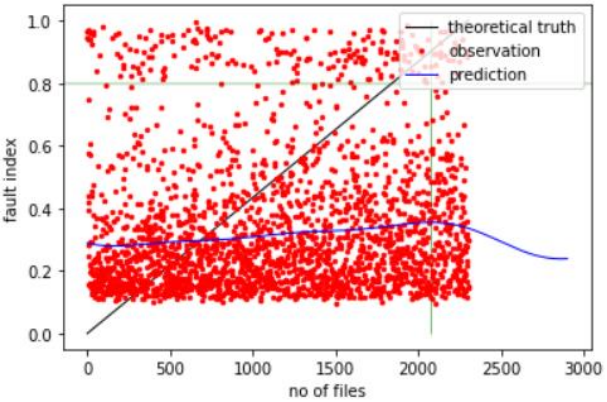

Figure 12. B1_5 (pred 2073, actual 2302)

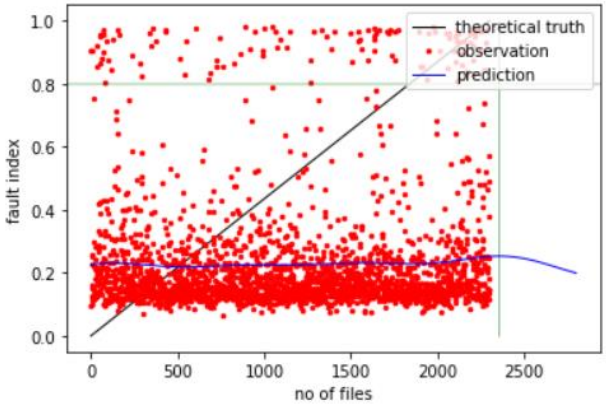

Figure 13. B1_6 (pred 2354, actual 2302)

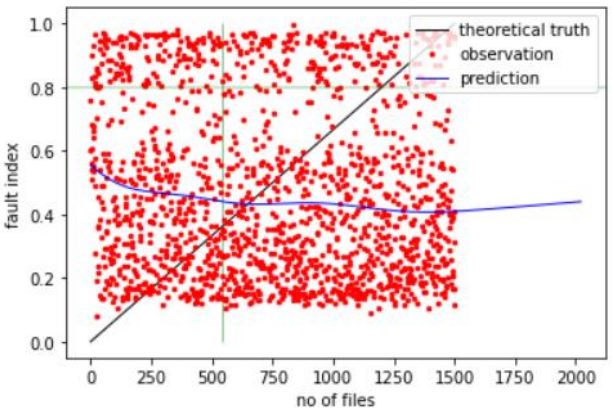

Figure 14. B1_7 (pred 545, actual 1502)

The results obtained from the experiment on test data in the second operating condition can be seen in Figures 15-19.

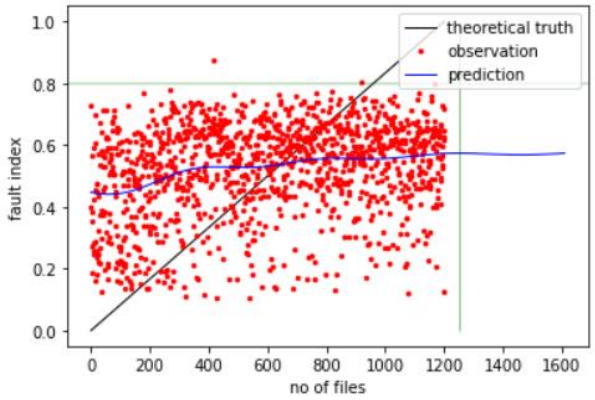

Figure 15. B2_3 (pred 1250, actual 1202)

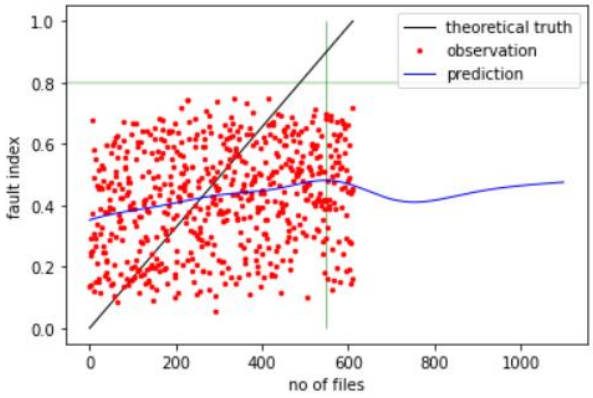

Figure 16. B2_4 (pred 547, actual 612)

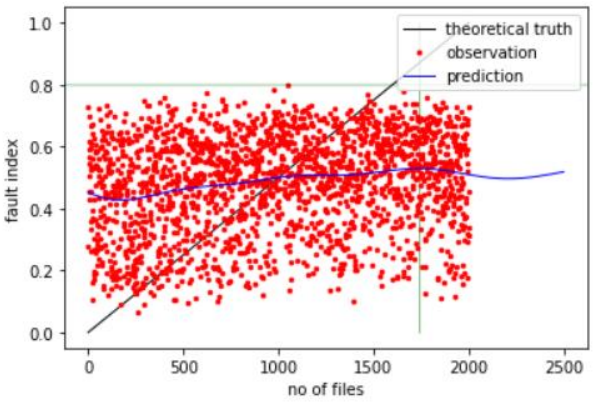

Figure 17. B2_5 (pred 1737, actual 2002)

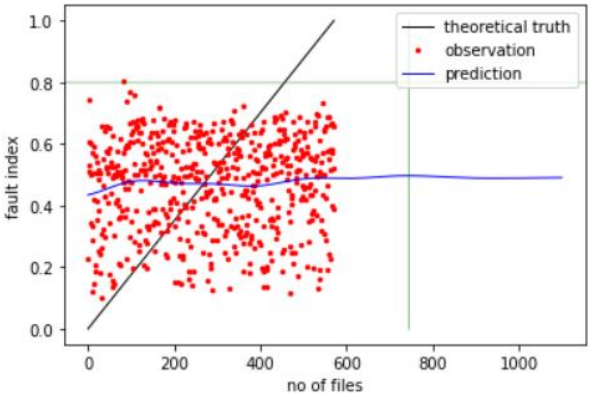

Figure 18. B2_6 (pred 743, actual 572)

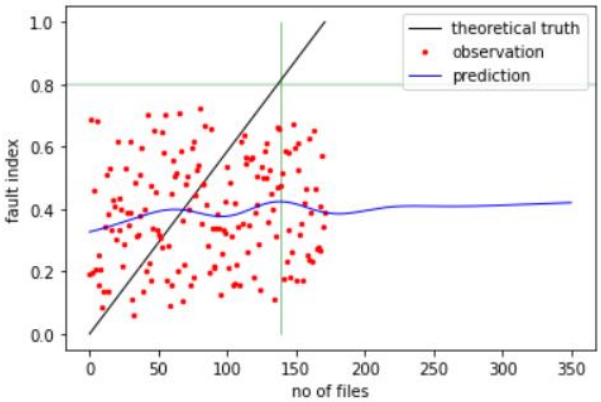

Figure 19. B2_7 (pred 139, actual 172)

Lastly, the results obtained from the experiment on test data in the third operating condition is depicted in Figure 20.

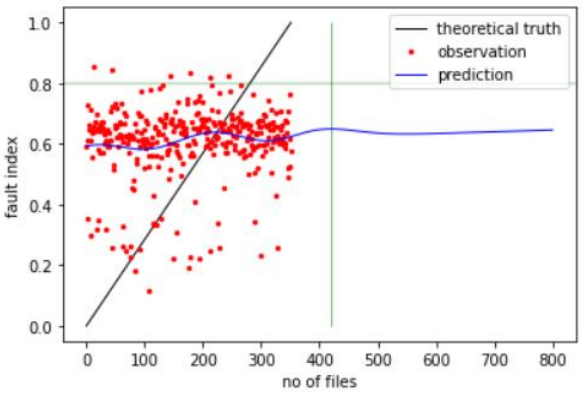

Figure 20. B3_3 (pred 419, actual 352)

The algorithm's performance across all test sets is presented in Table 4. While it can be seen that this study provides a more accurate and precise RUL prediction solution, it scores lower than [15] and [37] for early predictions. This is so because, it achieves a reasonable balance between overestimation and underestimation, thereby making it more reproducible and trustworthy.

Table 4. The comparative finding of the study.

Error	This paper	[15]	[37]	[38]	[39]	[40]	[41]
B1_3	-15.8	7.62	43.28	37.0	54.73	-0.35	-1.04
B1_4	-20.8	-157.71	67.55	80.0	38.69	5.6	-20.94
B1_5	9.94	-72.57	-22.98	9.0	-99.4	100	-278.26
B1_6	-2.25	0.93	21.23	-5.0	-120.07	28.08	19.18
B1_7	63.7	85.99	17.83	-2.0	70.65	-19.55	-7.13
B2_3	-3.9	81.24	37.84	64.0	75.53	-20.19	10.49
B2_4	10.62	9.04	-19.42	10.0	19.81	8.63	51.8
B2_5	13.2	28.19	54.37	-440.0	8.2	23.3	28.8
B2_6	-29.8	24.92	-13.95	49.0	17.87	58.91	-20.93
B2_7	19.18	19.06	-55.17	-317.0	1.69	5.17	44.83
B3_3	-19.03	2.09	3.66	90	2.93	40.24	-3.66
Er	**2.278**	2.62	12.20	-38.64	6.42	20.89	-16.08
|Er|	**18.929**	44.49	32.48	100.27	46.32	28.18	44.28
Score	0.3871	**0.4384**	0.2631	0.3066	0.3829	0.4285	0.3550

7. Discussions

To check to what length into the future the algorithm will be accurate. We provide 1400 data files, as shown in Figure 21, from B1_3 representing over 70% of its run-to-failure data, to the model, which predicted 1704 as its failure point. This is relatively close, as the run-to-failure data is 1802. Similarly, B1_5 depicted in Figure 22, was also tested with 1550 data files, i.e., over 60% of its run-to-failure data, wherein 2050 was predicted, for an actual failure at 2302. Further experimentation showed that the algorithm is accurate for approximately 1400 seconds into the future, indicating that 140 data files were predicted accurately. Recall that each data file is 10 seconds apart.

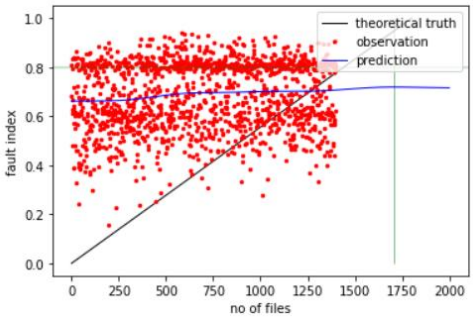

Figure 21. B1_3 with over 70% of its run-to-failure data (980Secs)

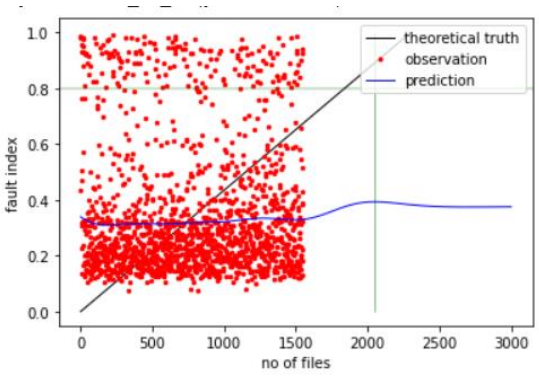

Figure 22. B1_5 with over 60% of its run-to-failure data (2520 seconds)

The effectiveness of this approach is substantiated through the results as they validate the key contributions of this research:

The approach's success in safely predicting an average of 1400 seconds into the future underscores its potential for industrial applications. Even more generally, the results above show that; early-stage predictions are possible with run-to-failure dataset provision ranging from 60% and above.

Also, while the current study focuses on the RUL of bearings, the algorithm's reliance on historical data suggests its applicability to similar predictive maintenance problems in other domains.

Lastly, the balance between overestimation and

underestimation achieved by the Voting Regressors is particularly noteworthy, as it contributes to the algorithm's minimal error in its predictions.

8. Conclusions

For an accurate RUL prediction, the dependency on historical information or data cannot be overestimated. This research has made significant contributions to the literature by introducing a novel Hilbert Transform data-driven hybrid approach using deep learning for the prediction of the Remaining Useful Life (RUL) of bearings. The proposed method integrates signal-based and knowledge-based techniques, employing the Hilbert Transform for initial signal processing, a CNN-LSTM model for feature extraction and sequence memorization, and a voting regressor for extrapolation into the future.

The unique combination of these techniques allows for a more accurate and precise RUL prediction, as validated against other cutting-edge methods using the FEMTO dataset. The proposed method achieves a balance between overestimation and underestimation, enhancing its reproducibility and reliability. A threshold value of 0.8, determined based on prior knowledge, enables us to validate the practical knowledge with contemporary data-driven processes, illustrating the close alignment between theory and practice. Furthermore, the algorithm demonstrates that early-stage predictions are feasible with run-to-failure data provision ranging from 60% and above, averaging 1400 seconds into the future. Ultimately, during the validation phase, the method put forth attains the minimal mean percentage discrepancy on the FEMTO dataset when compared with other methods. This implies that in an actual industry, the usage of the proposed method is feasible.

9. Future Work

The proposed method has shown promising results. However, further research is necessary to enhance its performance and expand its applicability.

Anomaly Detection: Future work could explore the integration of anomaly detection techniques to improve feature extraction. Anomaly detection could help identify unusual patterns in the vibration signals that may indicate the onset of bearing faults, thereby enhancing the predictive capabilities of the model.

Language Models: The use of language models could be explored to generate proposed behaviors based on trends in the data. Language models have been successful in capturing temporal dependencies in time-series data, and their application to vibration signals could provide new insights into the degradation process of bearings.

Adaptive Thresholding: The concept of thresholds could be further developed to be more adaptive and data-dependent not set to a fixed value as it was noted that some bearings failed having never crossed it. This could potentially improve the accuracy of early-stage predictions.

Applications to other domains: Future research could explore these applications, further validating the effectiveness of the method and broadening its scope of application.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kaplan K., Kaya Y., Kuncan M., Minaz M. R., Ertunç H. M., 2020. An improved feature extraction method using texture analysis with LBP for bearing fault diagnosis, Applied Soft Computing, 87, pp. 1-13.

[2] Kuncan M., Kaplan K., Minaz M. R., Kaya Y., Ertunç H. M., 2020. A novel feature extraction method for bearing fault classification with one dimensional ternary pattern, ISA Transactions, 100, pp. 346-357.

[3] Kaya Y., Kuncan M., Kaplan K., Minaz M. R., Ertunç H. M., 2020. Classification of bearing vibration speeds under 1D-LBP based on eight local directional filters, Soft Computing, 24, pp. 12175–12186.

[4] Kaya Y., Kuncan M., Kaplan K., Minaz M. R., Ertunç H. M., 2021. A new feature extraction approach based on one dimensional gray level co-occurrence matrices for bearing fault classification, Journal of Experimental & Theoretical Artificial Intelligence, 33(1), pp. 161-178.

[5] Simani S et al., 2003. Model-based fault diagnosis in dynamic systems using identification techniques. Springer-Verlag, Secaucus, NJ, USA.

[6] Isermann R., 2005. Model-based fault-detection and diagnosis – status and applications, Annual Reviews in Control, 29, pp. 71-85.

[7] Bensaadi R., Mouss H., Mouss M.D., Benbouzid M., 2005. Fuzzy Pattern Recognition Based Fault Diagnosis. International Review on Modelling and Simulations, 4, pp. 347-356.

[8] Do V. T., Chong U., 2011. Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain. Strojniški vestnik - Journal of Mechanical Engineering, 57, pp. 655–666.

[9] Lei, Y., Li, N., Gontarz, S., Lin, J., Radkowski, S., & Dybala, J., 2016. A model-based method for remaining useful life prediction of machinery. IEEE Transactions on reliability, 65(3), pp. 1314-1326.

[10] M. Benbouzid et al., 1999. Induction motors' faults detection and localization using stator current advanced signal processing techniques. IEEE Trans Power Electron, 14, pp. 14–22.

[11] A. Widodo et al., 2009. Intelligent fault diagnosis system of induction motor based on transient current signal. Mechatronics, 19, pp. 680–689.

[12] Wang R., Zhan X., Bai H., Dong E., Cheng Z., & Jia X., 2022. A Review of Fault Diagnosis Methods for Rotating Machinery Using Infrared Thermography. Micromachines, 13(10), pp. 1644.

[13] Witczak, M., Lipiec, B., Mrugalski, M., & Stetter, R., 2020. A fuzzy logic approach to remaining useful life estimation of ball bearings. In Advanced, Contemporary Control: Proceedings of KKA 2020—The 20th Polish Control Conference, Łódź, Poland, 22-25 June, pp. 1411-1423.

[14] Saon, S., & Hiyama, T., 2010. Predicting remaining useful life of rotating machinery based artificial neural network. Computers & Mathematics with Applications, 60(4), pp. 1078-1087.

[15] Chen Y., Peng G., Zhu Z., Li S., 2020. A novel deep learning method based on attention mechanism for bearing remaining useful life prediction, Applied Soft Computing, 86, pp. 1568-4946.

[16] Yan, M., Wang, X., Wang, B., Chang, M., & Muhammad, I., 2020. Bearing remaining useful life prediction using support vector machine and hybrid degradation tracking model. ISA transactions, 98, pp. 471-482.

[17] Wu J. Y., Wu M., Chen Z. H., Li X. L., & Yan R. Q., 2021. A joint classification-regression method for multi-stage remaining useful life prediction J. Manuf. Syst., 58, pp. 109-119.

[18] Li X., Elasha F., Shanbr S., and Mba D., 2019. Remaining useful life prediction of rolling element bearings using supervised machine learning Energies 12(14), pp. 2705.

[19] Wu J., Wu C. Y., Cao S., Or S. W., Deng C., and Shao X. Y., 2019. Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines IEEE Trans. Ind. Electron, 66, pp. 529-539.

[20] Meng, Z., Li, J., Yin, N., & Pan, Z., 2020. Remaining useful life prediction of rolling bearing using fractal theory. Measurement, 156, pp. 0263-2241.

[21] Ahmad, W., Khan, S. A., Islam, M. M., & Kim, J. M. 2019. A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models. Reliability Engineering & System Safety, 184, pp. 67-76.

[22] Soualhi A., Medjaher K., & Zerhouni N., 2014. Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression. IEEE Transactions on instrumentation and measurement, 64(1), pp. 52-62.

[23] Ren, L., Sun, Y., Cui, J., & Zhang, L., 2018. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. Journal of Manufacturing Systems, 48, pp. 71-77.

[24] Lu Y. W., Hsu C. Y., and Huang K. C., 2020, An autoencoder gated recurrent unit for remaining useful life prediction Processes, 8, pp. 1155-1159.

[25] Peng K. X., Jiao R. H., Dong J., and Pi Y. T., 2019. A deep belief network based health indicator construction and remaining useful life prediction using improved particle filter. Neurocomputing, 361, pp. 19-28.

[26] Zhao C., Huang X., Li Y., & Li S., 2021. A novel cap-LSTM model for remaining useful life prediction. IEEE Sensors Journal, 21(20), 23498-23509.

[27] Li X., Zhang W., and Ding Q., 2019. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. Reliability Engineering & System Safety, 182, pp. 208-218.

[28] Wang B., Lei Y. G., Li N. P., and Yan T., 2019. Deep separable convolutional network for remaining useful life prediction of machinery. Mechanical Systems and Signal Processing, 134, pp. 106330.

[29] Lee K., Man Z., Wang D., & Cao Z., 2013. Classification of bioinformatics dataset using finite

impulse response extreme learning machine for cancer diagnosis. Neural Computing and Applications, 22, pp. 457-468.

[30] Feldman, M., 2011. Hilbert transform applications in mechanical vibration. John Wiley & Sons, West Sussex, England.

[31] Yu Y., Si X., Hu C., & Zhang J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation, 31(7), pp. 1235-1270.

[32] Awad M., Khanna R., Awad M., & Khanna R., 2015. Efficient learning machines: Theories, concepts, and applications for engineers and system designers. New York.

[33] Maulud D., & Abdulazeez A. M., 2020. A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4), pp. 140-147.

[34] Nectoux P., Gouriveau R., Medjaher K., Ramasso E., Chebel-Morello B., Zerhouni N., Varnier C., 2012. PRONOSTIA: An experimental platform for bearings accelerated degradation tests, IEEE International Conference on Prognostics and Health Management, June, pp. 1-8.

[35] Zhou W., Habetler T. G., & Harley R. G., 200, September). Bearing condition monitoring methods for electric machines: A general review. IEEE international symposium on diagnostics for electric machines, power electronics and drives, pp. 3-6.

[36] Akcan E., Kaya Y., 2023. A new approach for remaining useful life prediction of bearings using 1D-ternary patterns with LSTM. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 45, pp. 378.

[37] Rumelhart D. E., Hinton G. E., Williams R. J., 1986. Learning representations by back-propagating errors. Nature, 323, pp. 533-536.

[38] Sutrisno E., Oh H., Vasan A., Pecht M., 2012. Estimation of remaining useful life of ball bearings using data driven methodologies. IEEE Conference on Prognostics and Health Management, pp. 1-7.

[39] Hinchi Z., Tkiouat M., 2018. Rolling element bearing remaining useful life estimation based on a convolutional long-short-term memory network. Procedia Computer Science, 127, pp. 123–132.

[40] Lei Y., Li N., Gontarz S., Lin J., Radkowski S., Dybala J., 2016. A model-based method for remaining useful life prediction of machinery. IEEE Transactions on Reliability, 65(3), pp. 1314–1326.

[41] Hong S., Zhou Z., Zio E., Hong K., 2014. Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method. Digital Signal Process, 27, pp. 159–166.