Results in Engineering 24 (2024) 103420

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Research paper

An efficient deep learning prognostic model for remaining useful life estimation of high speed CNC milling machine cutters

Hamdy K. Elminir [a], Mohamed A. El-Brawany [b], Dina Adel Ibrahim [a,*], Hatem M. Elattar [c], E.A. Ramadan [b]

[a] Department of Electrical Engineering, Faculty of Engineering, Kafr Elshiekh University, Kafr Elshiekh, Egypt
[b] Industrial Electronic and Control Engineering, Faculty of Electronic Engineering, Menoufia University, Egypt
[c] Faculty of Computer Science, October University for Modern Sciences & Arts, Egypt

ARTICLE INFO

Keywords:
CNC milling machine
Deep Learning DL
LSTM
AutoEncoder and RUL

ABSTRACT

CNC machines are engaged in numerous industries, including critical ones like the aerospace, automotive, and military sectors, among others. Sensor data are time-series that may suffer from complex interconnections between variables and dynamic features. Long Short Term Memory LSTM excels in dynamic feature extraction, and Autoencoder AE has great capabilities in nonlinear deep knowledge of time-series data variables. In this work, we propose a model for tool wear prediction of CNC milling machine cutters as a type of time-series data taking advantage of the LSTM and AE capabilities. The framework consists of many steps, including extracting multi-domain features and a correlation analysis to select the most correlated features to the tool wear. New features are added, such as entropy and interquartile range IQR, which proved to be highly correlated to the cutter tool wear. An LSTM`-AE model is then trained, validated, and tested on this feature map to predict the target tool wear value. The model is provided with degradation or Run-To-Failure data for CNC machine cutters, the PHM10 dataset, to predict the tool wear values. The predicted tool wear value is compared against the wear curve to estimate RUL values. The predicted RUL values mostly underestimate the real values, which helps schedule for maintenance or equipment replacement before failure. The experimental results show that the proposed framework outperforms state-of-the-art DL methods in tool wear prediction accuracy approaching %98, as well as an enhancement of MAE and RMSE in the test set by reaching 2.6 ± 0.3222E-3 and 3.1 ± 0.6146 E-3, respectively.

1. Introduction

The target of the tool condition monitoring (TCM) process can be categorized into fault detection, fault type determination, and system Remaining Useful Life (RUL) estimation, which can be identified by the term "prognostics." Prognostics is a growing field of research nowadays to help prevent rather than detect tool breakage. The term "prognostics" is "anything that foretells." Prognostics is mainly a health statement problem that takes as input sensory data and outputs RUL. This is done through many steps: detection of failure indicators, generating a current state estimate and health index constitution [1,2]. RUL prediction or estimation is the main target of prognostics, which is the time the machine can safely work before failure [3]. Tool wear can be considered a health index that helps RUL estimation for the machine.

Data-driven prognostics require no knowledge about system physics but demand Run-To-Failure RTF data regarding system performance. Artificial Intelligence AI is very popular nowadays in predictive maintenance models, either for diagnosis, fault classification, or RUL prediction applications [4].These methods, which use classical AI methods, can be machine learning ML-based, such as Support Vector Machine SVM and Random Forest RFs or Deep Learning DL approaches. SVM and RFs techniques were widely used in research to forecast tool wear and cutter RUL prediction. Utilizing Random Forest RF, the authors demonstrated a technique for tool wear prediction in milling operations [5]. After that, they conducted a comparison with prior ML algorithms in [6]. The XGboost algorithm is a ML algorithm based on gradient boosting, which is used for ML model optimization problems. XGboost was employed for RUL estimation of LIB with fine tuning of its hyper-parametrs [7]. XGboost algorithm was emphasized for LIB state of charge estimation and the model exceled in terms of MSE and RMSE

* Corresponding author.
E-mail address: dinaadelibrahim@gmail.com (D.A. Ibrahim).

https://doi.org/10.1016/j.rineng.2024.103420
Received 18 September 2024; Received in revised form 25 October 2024; Accepted 13 November 2024
Available online 16 November 2024

over traditional regression models [8]. Other ML methods were used for battery health assessment [9].SVM was used with ANN and XGboosting algorithm for predictive maintenance PdM of active chilled beam air conditioning systems [10]. In the age of Industry 5.0, ML methods were highly implemented for PdM and CM [11]. The performance of RFs was compared by the authors to feed-forward back propagation (FFBP). The result showed that RF performed better than other preceding techniques. Other studies utilized SVMs in tool condition monitoring [12]. Artificial neural networks ANNs are also widely used in TCM. Different ML approaches including SVM, RF and multi-layer perceptron were implemented for tool wear prediction and classification for additively manufactured 316 stainless steel [13].

ANN, as the dominant research era in AI, has attracted attention for enhancing its capabilities of prediction, failure analysis, and diagnosis. Sindhu, Tabassum Naz, et al. proposed a model for disease analysis with ANN structure enhancement [14]. Çolak, Andaç Batur, et al. concentrated their study on how ANN and maximum likelihood estimation can predict electrical component reliability [15].

The previous model was emphasized to explore and model the breast carcinoma to stress on the fact that ANN is very effective at predicting various parameters, including patient survival in this model [16]. Shafiq, Anum, et al. held a comparative study to investigate using ANN and maximum likelihood estimation in COVID-19 dataset analysis [17]. In [18], the authors concentrated on using Rayleigh distribution to develop a multi-layer ANN with Bayesian optimization used in reliability parameter estimation. ANN has been used to investigate fluids flow under certain conditions and predict its controllable parameters like Ree–Eyring fluid [19] and nanofluids [20,21]. To conclude, ANN has proved its superiority in reliability analysis of lifetime models [22,23].

DL techniques are heavily used in failure analysis and prognostics of industrial systems [24] .DL-based models have invaded the Prognostic field as a result of the development of sensors and Big Data [25]. DL architectures, especially Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTMs), have been widely utilized in the TCM process. CNN has many applications in prognostics including fault diagnosis of robotic fuses [26]. CNN was applied for tool wear prediction in many research studies on milling data. In [27], a deep learning model was developed for tool wear prediction utilizing different-domain features. They extracted these features for force and vibration sensors only. Then, they used raw sensory data to develop a new model for tool wear prediction, which was named the reshaped time series convolutional neural network RTSCNN [28], where they used CNN as a feature extractor. A dense layer with a Rectified Linear Unit (Relu) activation function and then a regression layer are added for tool wear prediction. The researchers themselves concluded that there were no distinguishable improvements over their earlier research. CNN was used in a hybrid model with LSTM for tool wear prediction in many research studies [29].

In [30], Convolutional Bi-directional Long Short-Term Memory (CBLSTM) networks are intended to handle input composed of raw sensor readings. CNN is first used for local feature extraction. To forecast the target value, bi-directional LSTMs are combined with stacked, fully-connected layers and a linear regression layer. A real-life tool wear test was introduced using raw sensor readings. CNN is used again with LSTM to constitute a health index HI then RUL prognosis for C-MAPSS data set [31].

LSTM is a branch of Recurrent Neural Network (RNN) that is popular in sensor data or time-series data applications. Many researchers used LSTM or its variants for prognostic applications. LSTM was recently used in other applications such as Lithium-ion battery LIB health prediction and tool wear prediction. In [32], the LIB experimental data is used for health indicator HIs extraction, and then LSTM was used with GPR to construct a degradation model of these HIs. This model was used for battery pack prediction, which performed well in terms of MAE and RMSE. A model of three steps for battery capacity degradation prediction using XGboosting, Stacked bidirectional LSTM, and Bayesian optimization was proposed in [33]. Wang, Jiujian, et al. constructed a model using LSTM to solve the Prognostics and Health Management PHM competition PHM08 [34]. Transfer learning bi-directional LSTM was used for RUL prediction for rolling bearings under different operating situations [35]. LSTM was used in prediction in different applications like road traffic flow prediction along with GRU and wavelet transform [36]. A model using LSTM and GRU was built for energy consumption of water treatment plants [37]. LSTM has been used for tool wear prediction applications in many studies. LSTM was used again with singular spectrum analysis for feature extraction and PCA for dimensionality reduction to construct a model for tool wear prediction in [38].

LSTM is highly combined with CNN to benefit from the advantages of both architectures in enhancing the prediction accuracy of the prognostic model [29]. In [39], a TCM model was created using CNN and BiLSTM, and then the measured tool wear value is used for tool wear prediction based on ResNetD. A model for RUL prediction of milling cutters was proposed in [40] based on CNN and BiLSTM together with an attention mechanism to select the highly relative features. The model was evaluated against different datasets to prove its accuracy. A 1D-CNN with a residual structure is used to extract features in a TCM model, and then a BiLSTM is used for prediction [41].

AE has advantages similar to CNN in extracting long-term dependencies and highly representative features, as long as it has the capability of data-denoising and dynamic feature extraction. Thus being used in fault diagnosis of industrial processes. In [42], a model was used for fault diagnosis of rotating machinery using AE for feature learning and an artificial fish swarm algorithm for optimization. Denoising AE was enhanced in [43] to improve fault diagnosis of rolling bearings. A sparse stacked AE was used in simultaneous fault analysis of solid oxide fuel cell system and proved the superiority of the method [44]. The AE also has the capability of reconstructing the input signals with the best accuracy, so it can perform well in sequence prediction applications or sensor-based data such as PHM10. We proposed a hybrid organization between LSTM and AE for tool wear prediction, hence RUL estimation of CNC milling machine cutters.

RUL estimation depending on RTF data is a big concern in prognostics. We built our model on RUL estimation depending on the predicted value of tool wear and the limit of the wear curve. The efficacy of the proposed prediction method is a key factor in the success of RUL estimation. The proposed method proved perfect for estimating the RUL of CNC milling machine cutters.

The sensor data is multi-variable time-series data, which may suffer from dynamic variables with auto-correlation or different transitory correlations among different variables. This makes dealing with sensor data very challenging, especially in the case of big data like the PHM10 dataset. Since AE has the capability of dimensionality reduction and dynamic feature extraction of the input data [45], a property that makes it very suitable for big sensor datasets. LSTM has proved its superiority in time-series data prediction applications.

A composite model LSTM-AE model is built for prediction of sensor data that is quite big based on LSTM and AE neural networks. The model was evaluated on the PHM10 data set and proved its accuracy in tool wear prediction and RUL estimation of the cutters. The model can be easily generalized to other datasets that are composed of sensor data with its raw state and predefined target variable or a large set of features to be reduced also with a target variable.

From the preceding literature, the main contributions of this research are:

A framework was proposed for a TCM process of the cutters of a CNC machine, including tool wear prediction and RUL estimation. The PHM10 original dataset was used using C1, C4, and C6 cutter data for model training, testing and validation. An approach for extracting features from different domains is used in the suggested framework and selecting the most correlated features to the tool wear. A DL-based model is used with a hybrid organization to fit the selected features and the target wear. The model consists of an LSTM-AUTOENCODER architecture. The model is trained, tested, validated, and used for flank tool wear prediction of PHM10 data. Utilizing the wear limit and the wear curve, the predicted wear value is utilized to estimate RUL. The model has outperformed state-of-the-art DL methods in tool wear prediction accuracy, approaching 98 % in predicting tool wear, which leads to very accurate estimation to RUL, thus enabling to schedule maintenance before it is needed. The accurate prediction of tool wear allowed underestimation of the RUL meaning to predict a value of RUL that is a small bit less than the actual value which is the correct case for the predictive maintenance.

The rest of the paper is organized as follows. *Section 2* gives a short description of basic DL architectures used to build the model LSTM network and AE network. Then, the proposed model is explained in Section 3. An experimental study is proposed in Section 4 that introduces the PHM10 dataset with pre-processing of the data before model application. The model training and RUL estimation algorithm are proposed in Section 5. The model accuracy and effectiveness are evaluated through experimental results in Section 6. The concluding remarks are given in Section 7.

2. Theoretical methodology

2.1. Basic LSTM architecture

Recurrent Neural Networks (RNNs) store the state of previous cells, making them ideal for datasets that have the type of sequences or time steps. To train the network, the state of the hidden unit is changed based on the preceding cell state and the result from the current input passing by an activation function. RNNs can detect long-term and transitory relationships in sequence-like and time-series like datasets, but suffer from significant limitations, such as gradient exploding or vanishing problems. To address this problem, new versions of RNN were released: Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM). LSTM captures long-term dependencies using a set of gates controlling which information enters the memory and which is forgotten. The LSTM neural network is a long string of LSTM basic cells shown in Fig. 1. Unlike standard RNN models, LSTMs has four network layers shown by rectangles in Fig. 1 and three control gates (forget gate, input gate, and output gate) that function together. The three gates are: (1) the forget gate f^tselects which part of information in the previous instant of time to be forgotten by the cell state. (2) The input gate i^tselects which of the input data to update the cell state. (3) The output gate determines what information about the cell's state is kept via the output.

In LSTM, the hidden state h^t is updated at each time step t according to the following equations Eq. (1) [46]:

$$\begin{aligned} i^t &= \sigma\left(W^i x^t + V^i h^{t-1} + b^i\right), \\ f^t &= \sigma\left(W^f x^t + V^f h^{t-1} + b^f\right), \\ o^t &= \sigma\left(W^o x^t + V^o h^{t-1} + b^o\right), \end{aligned} \tag{1}$$

$$\boldsymbol{c}^t = f^t \odot \boldsymbol{c}^{t-1} + i^t \odot \tanh\left(W^c x^t + V^c h^{t-1} + b^c\right)$$

$$h^t = o^t \odot \tanh(\boldsymbol{c}^t)$$

x^t is the current data at the same time step, h^{t-1} is the preceding instant hidden state, the input i^t, the forget gate f^t, 5-the output gate o^t and a memory cell c^t. Where parameters including all $\in \mathbb{R}^{d\times k}$, $V \in \mathbb{R}^{d\times d}$ and $b \in \mathbb{R}^d$ are all learned during network training and are common to all time steps. The last equation gives the hidden layer function $\mathbb{H}$.

σ () represents the logistic activation function and $\odot$ represents the element by element product formula and a hyper-parameter that represents hidden vector dimensions is called k.

2.2. Auto-Encoder

An Autoencoder is a specific type of self-supervised learning model that has the ability to learn a compressed version of the input data and proved its superiority in de-noising applications, fault diagnosis, and feature learning tasks. The Auto-Encoder AE is composed of two components: a block for encoding and another one for decoding as shown in Fig. 2. The encoder transforms input data into latent descriptions of features. The decoder uses concealed information as input and reconstructs the data at the output layer. Both the input and the output layers have equal neurons. For a particular input $x \in \mathbb{R}^m$, the mathematical equations for the encoder and the decoder can be written as follows [47]:

$$E(X) = \sigma(W_{xh} + b_{xh}) \tag{2}$$

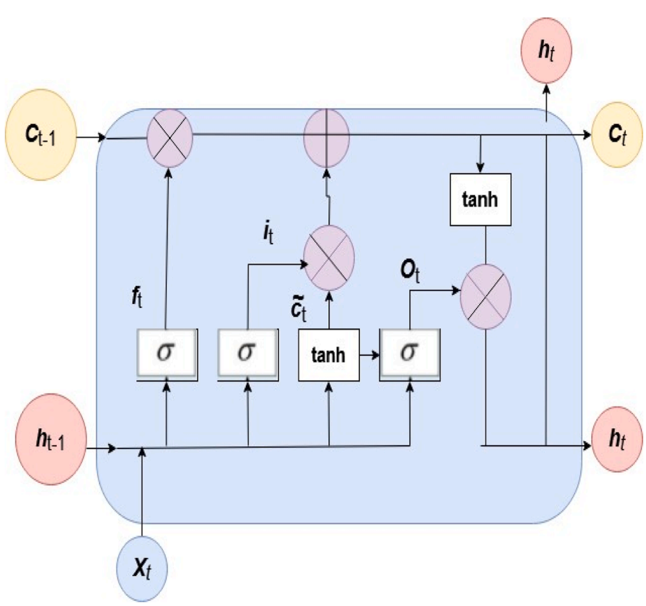

Fig. 1. Basic LSTM cell.

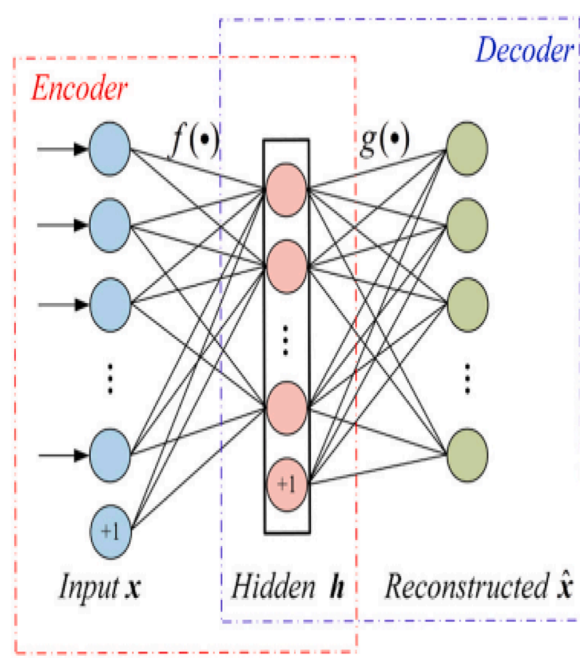

Fig. 2. Structure Diagram of AE network.

$$D(h) = \sigma(W_{h\hat{x}}\ h + b_{h\hat{x}}) \quad (3)$$

Where $\sigma$the sigmoid activation function and h is the hidden layer data after the encoding conversion. The gradient descent is used for error minimization between input and reconstructed output and the loss function is given as follows:

$$\text{Loss} = \arg\min\| X - \widehat{X} \|^2 \quad (4)$$

where $\widehat{X}$ is the data reconstructed by the decoder.

3. Methodology

3.1. Research motivation

Data in time-series form has more complicated interconnections between their features than single variable datasets. This is a challengeable task in feature engineering. LSTM may efficiently capture the non-static properties and nonlinear interactions among parameters. However, when it comes to learning deep nonlinear information between time series variables, the AE neural network excels. The LSTM-AE architecture is a composite structure of LSTM and AE networks to benefit from their advantages. In this model, the input data can be encoded as constant-length vectors and then decoded into the objective sequences. Combining the two networks, this model has the ability to capture the input data dynamic properties, making it suitable for industrial processes, including industrial datasets such as PHM10.

3.2. Model construction

An LSTM-AE is based on a hybrid organization between LSTM and AE. The LSTM-AE has two components: an encoder and a decoder as illustrated by Fig. 3. The LSTM encoder has many layers of concatenated LSTM cells to capture long-term representations between the input features. An encoded vector is the output of the encoding part, which is repeated by a RepeatVector part no of times that is equal to the timesteps of LSTM. This repeated vector is fed into the decoder LSTM, which is layers of LSTMs in the reverse order of the encoder part, then a dense layer is added and used with a Timedistributed function to get the reconstructed features. The model's performance is assessed according to its capability of the input pattern reconstruction. The decoder component of the model may be removed after the trained model achieves an acceptable performance measure in reproducing the input data. In this situation, when the model is fitted, the reconstruction portion can be dropped, and the model can only be used for the prediction task. The stacked LSTM is responsible for the prediction process, but this ensemble model helps increase wear prediction accuracy. This is due to its capability of capturing long-term dependencies and most representative features in the encoder part.

4. Experimental study

The general framework of the model can be shown in Fig. 4. It has many stages, including feature extraction, selection, model training, and tool wear prediction. The description of the dataset is given first in the next section. Feature engineering is discussed after highlighting data analysis and feature selection. Model training and RUL estimation are presented after that.

4.1. Data description

We used dataset from the PHM10 data competition that contain high speed CNC milling machine cutters data [48]. The dataset contain files for six cutters with three flutes (C1, C2, C3, C4, C5, and C6). Dynamometer, accelerometer, and acoustic emission sensors are set up in a proper place to capture data as shown in Fig. 5. Readings for these seven

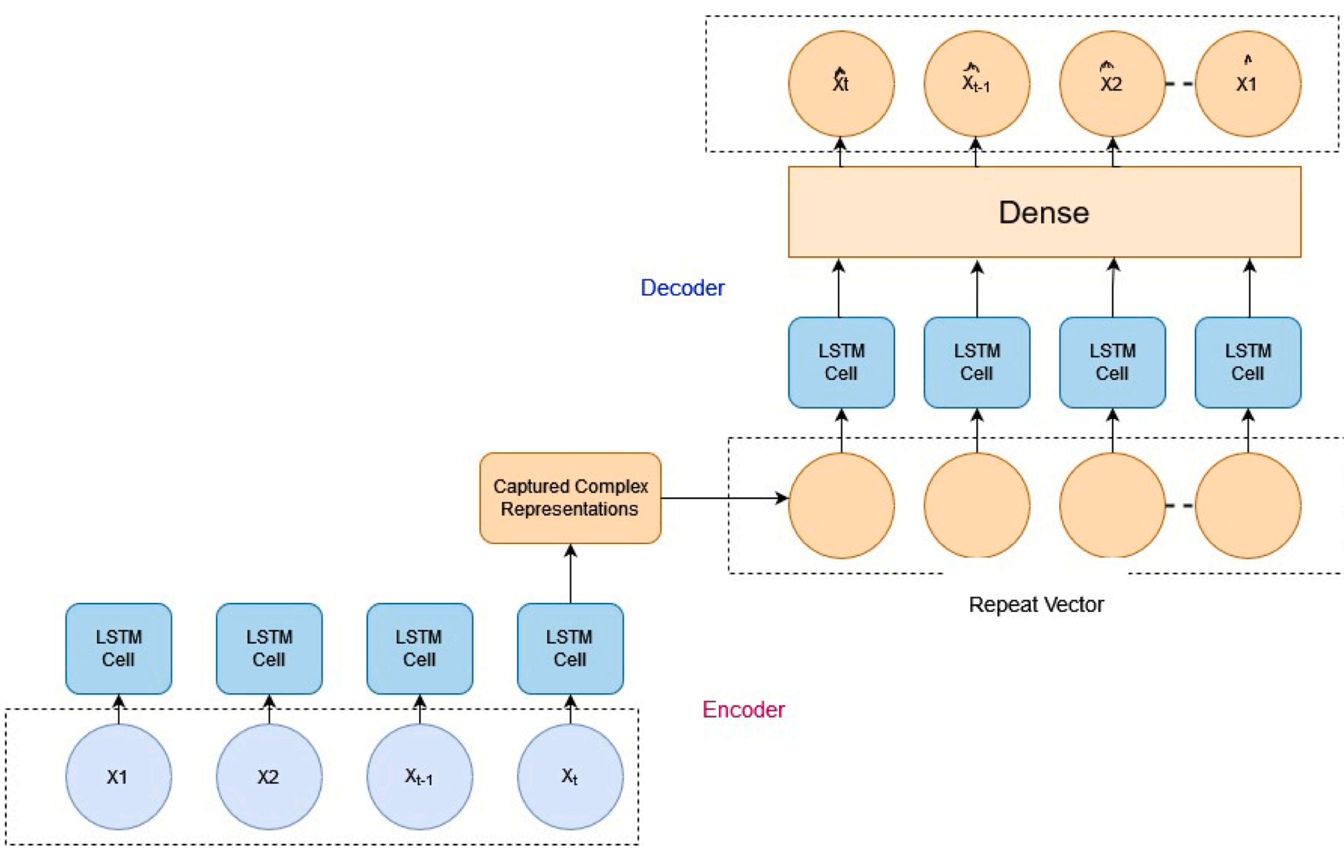

Fig. 3. LSTM-AUTOENCODER architecture.

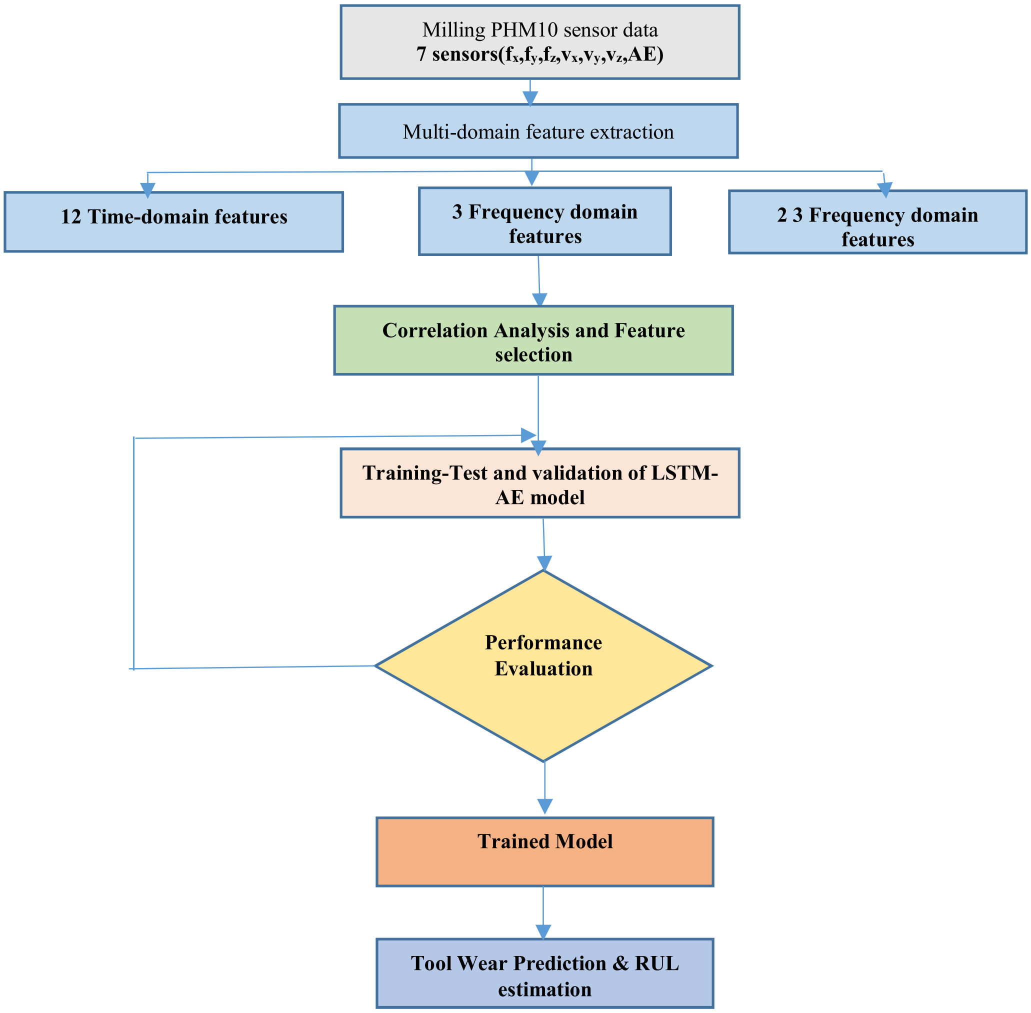

Fig. 4. Tool wear prediction proposed framework.

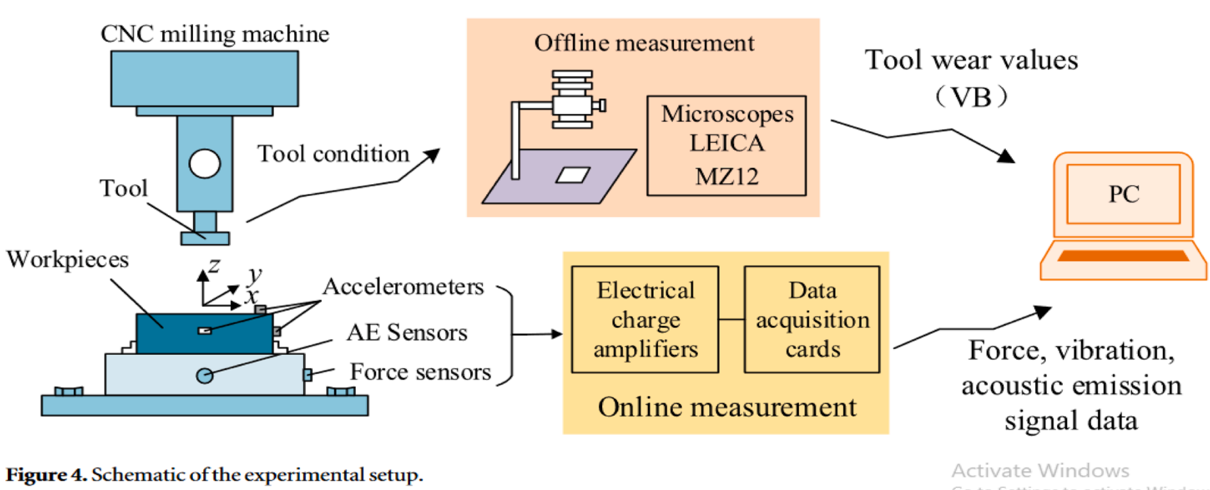

Fig. 5. Sensor setup for CNC machine.

sensors were obtained for each of the 315 cuts performed by the cutter. At 50,000 Hz per channel, these readings were collected. Each cutter data has 315 independent file for the 315 runs and the data files are structured in seven columns, representing: three-dimensional force (f_x, f_y, f_z) and three-dimensional vibration (v_x, v_y, v_z)), and AE-RMS (V), which is acoustic emission (AE) signal in rms value. The operational features of the high speed CNC milling machine under examination are listed in Table 1. C1, C4 and C6 data constitutes the training, testing, and validation dataset part since their wear file is given in the dataset files. Each run file is framed into 50 frames to have >15,000 records for each

Table 1
Variables of the Experimental setup.

Parameters	Value
The running speed of the spindle	10,400 rpm
The feed rate in x direction	1555 mm/min
The depth of cut (radial) in y direction	0.125 mm
The depth of cut (axial) in z direction	0.2 mm

cutter.

4.2. Feature engineering

Researchers frequently set up a dynamometer, an accelerometer, and a microphone on the machine in the proper location to capture cutting force, vibration, and acoustic emission data. According to earlier studies, features from different domains captured in time and frequency can precisely evaluate tool wear status. Different-domain features for the force and vibration sensor readings were used in [49] for the purpose of detecting virtual tool wear. A health index was developed to keep track of the tool status using time-frequency domain features for all the sensors in [50]. These features were extracted by wavelet packet decomposition. Wu et al. [51] extracted different domain features, and then an analysis bof the input dataased on Pearson Correlation Coefficient (PCC) was done to determine how far these features are correlated together. The selected features were input into an ANFIS that was used for RUL prediction of machining tools. PCC was used also in [52] to select the key features for tool wear regression and RUL estimation.

Through literature review, it can be found that multi-domain feature extraction is a mandatory step in our feature engineering. However, we studied adding different features, such as interquartile range IQR and entropy in different domains. Interquartile Range can be added to the features extracted through the review because it is a good measure of data spread and can identify outliers and skewness of the dataset. Another important feature is the entropy, which is defined as the average amount of "information" or "uncertainty" associated with the variable's potential results. These features have been investigated, and they proved a high correlation to the tool wear.

An important step in the PHM cycle used to be feature extraction as a part of exploratory data analysis (EDA). These features are condition indicators that reflect the health of the machinery being monitored. Features can also be combined from different domains to constitute a HI that express the degradation process. Redundant data can cause noise and lead to bad performance of the model. One way to overcome this is choosing features that have monotonic, trendable, and predictable behavior, as will be presented next. We have added minimum value, mean value, and interquartile range as parts of the five-number statistics [53]. IQR and entropy of the dataset. IQR is added to the features extracted through the review. It is a good measure of data spread and can identify outliers and skewness of the dataset. Another important feature is the entropy, which is defined as the average level of "information" or "uncertainty" inherent to the variable's possible outcomes. Entropy was studied in the time and frequency domains. We have extracted the multi-domain features shown in Table 2 and added entropy and IQR using TSFEL [54]. We extracted multi-domain features for the seven signals f_x, f_y, f_z,v_x,v_y,v_z and AE of each cutter and studied their correlation with the tool wear to best constitute the health index of the tool wear for accurate tool wear prediction. Fig. 6 shows the extracted features for sensor f_x for the whole lifecycle from RTF as an example on the features extracted.

4.3. Data analysis and feature selection

After feature extraction, a data analysis is used to select features that are highly correlated to wear values. These are considered as health indicators for the target wear value to be predicted. The Pearson correlation coefficient quantifies the degree of relationship between the two parameters, which determines how one parameter is altered by the change in the other one. A correlation value of −0.1, for instance, indicates a modest negative relationship between variables X and Y, but They would be considered to have a high negative relationship if their correlation coefficient was about −0.9. Principal Component Analysis PCA is one way of to reduce dimensions of the dataset and shrink the feature space to best express our data. Many researchers applied PCA for dataset reduction expressing target variable [55–56]. In [57], it was concluded that about 24 features can cover around 98 % percent of original dataset variance. This Data is first standardized using standard scalar expressed by Eq.2, where x_{mean} is the mean value and σ is the standard deviation. Correlation between features of different signals and the target variable has been studied according to Pearson Correlation Criteria PCC to select most correlated features to the wear. As shown in Fig. 7, the heatmap shows the correlation between wear and features in order to select the most correlated ones. It can be noticed that the added features by the proposed framework has significant correlation to the wear value as for entropy and IQR. The study was done against the other sensors as f_y, f_z etc., and they all had similar behavior according to the selected features. Some of the features had inter-correlation between each other, this was also take into account when selecting the feature space. We fitted our feature space to 6 features for each signal to contain most correlated ones to the wear signal. Acoustic Emission signal has shown low correlation with the target wear value so we excluded it from the feature space used in training the model.

Table 2
Multi-domain features [49].

Feature	Expression
RMS	$z_{rm}\sqrt{\frac{1}{n}\sum_{i=1}^{n} {z_i}^2}$
Variance	$z_{var}=\frac{1}{n}\sum_{i=1}^{n}(z_i-\overline{z})^2$ $\overline{z}\ =\mu$ mean value
Maximum	$z_{max}=\max(z)$
Skewness	$z_{skew}=E[\left(\frac{Z-\mu}{\sigma}\right)^3]$
Kurtosis	$z_{kurt}=E\left[\left(\frac{Z-\mu}{\sigma}\right)^4\right]$
Peak to Peak	$z_{p\text{-}p}=\max(z)-\min(z)$
Spectral skewness	$f_{skew}=\sum_{i=1}k\left(\frac{f_i-\overline{f}}{\sigma}\right)^3 S(f_i)$ *where S(fi) is the power spectrum density obtained using the Welch method.*
Spectral Kurtosis	$f_{kurt}=\sum_{i=1}k\left(\frac{f_i-\overline{f}}{\sigma}\right)^4 S(f_i)$
Wavelet Energy	$E_{WT}=\sum_{i=1}^{N} {wt_\phi}^2(i)/N$

$$x_{new}=(x-x_{mean})/\sigma \tag{5}$$

Where x_{mean} is the mean value and σ is the standard deviation value.

5. Model training

In the training process, C1 and C6 data is used for training, validation and C4 is used for testing of the network. Vanilla LSTM model with a tensorflow keras backend is trained for tool wear prediction and RUL estimation. Number of epoaches is 700, two layers LSTM with 2 dropout layers and one dense layer with relu activation function. Training is done using google colab T4 GPU backend to decrease time per epoch to 12 s. The value of the tool wear is predicted and due to the wear limit, the maximum number of cuts that can be safely made i.e. RUL can be estimated. Table 3 shows the trained network parameters. The layers of the model are shown in Fig. 8.

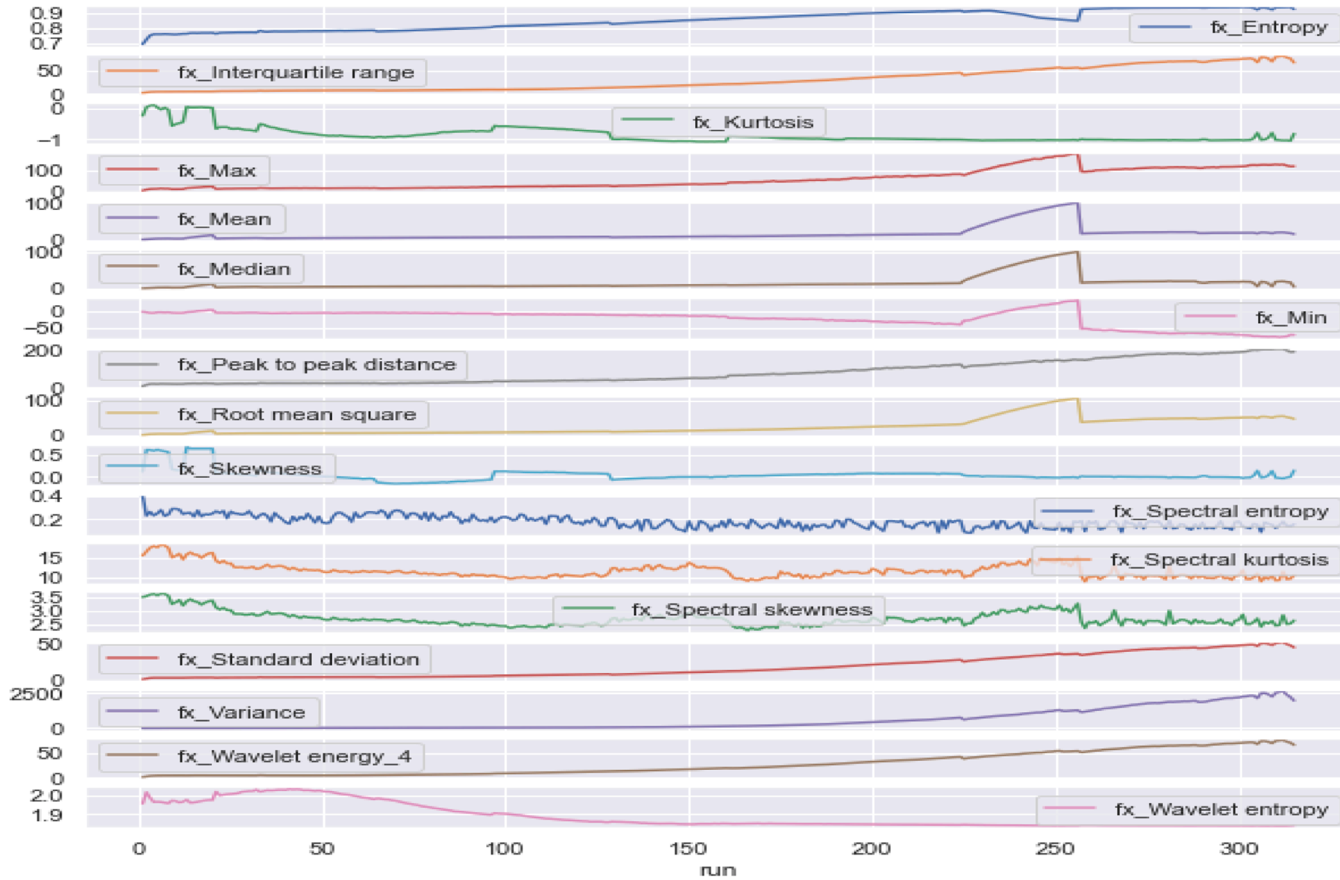

Fig. 6. Extracted features for f_x sensor against run#.

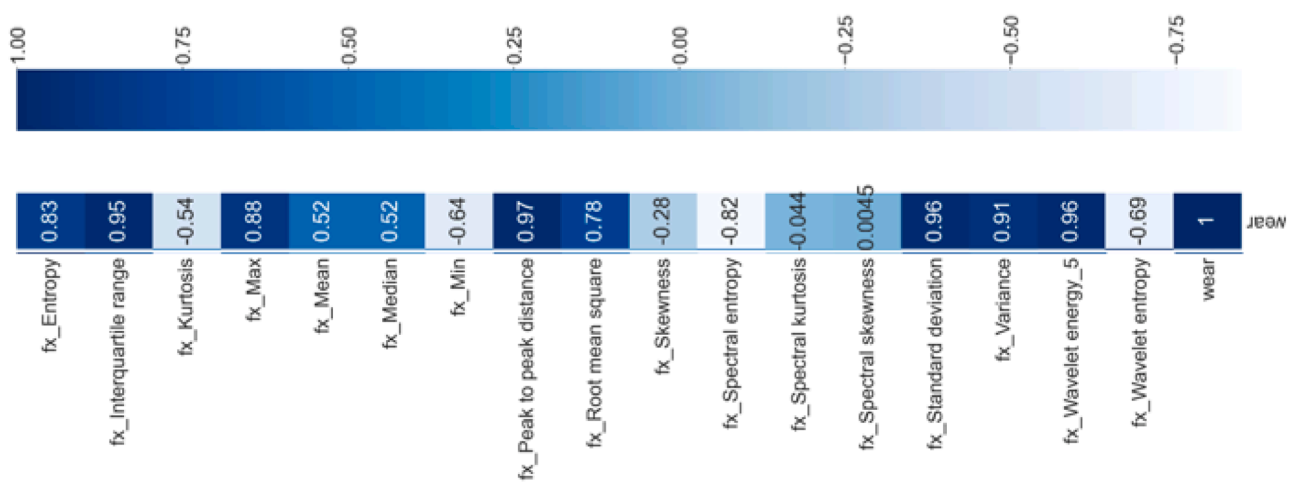

Fig. 7. Heatmap showing correlation between extracted features for f_x sensor and wear value.

Table 3
Network parameters.

Gradient Descent Algorithm	Batch size	Number of epoch	Ratio of Dropout	Activation function
Nestrov	32	700	0.3	Relu

5.1. RUL estimation

RUL is the number of cuts the cutter's tool can safely make before complete failure. With the availability of run-to-failure RTF data, i.e., from beginning to failure data, as shown in Fig. 9, predicted tool wear values can be used with RTF data for RUL estimation using algorithm shown in Table 4. Example of RUL estimation using predicted tool wear and the wear curve It can be noticed during the study that the first and last few runs contain noisy measurements, so they can be removed, and the wear limit is until run #300 only. The PHM society provided this data set in the PHM10 competition.

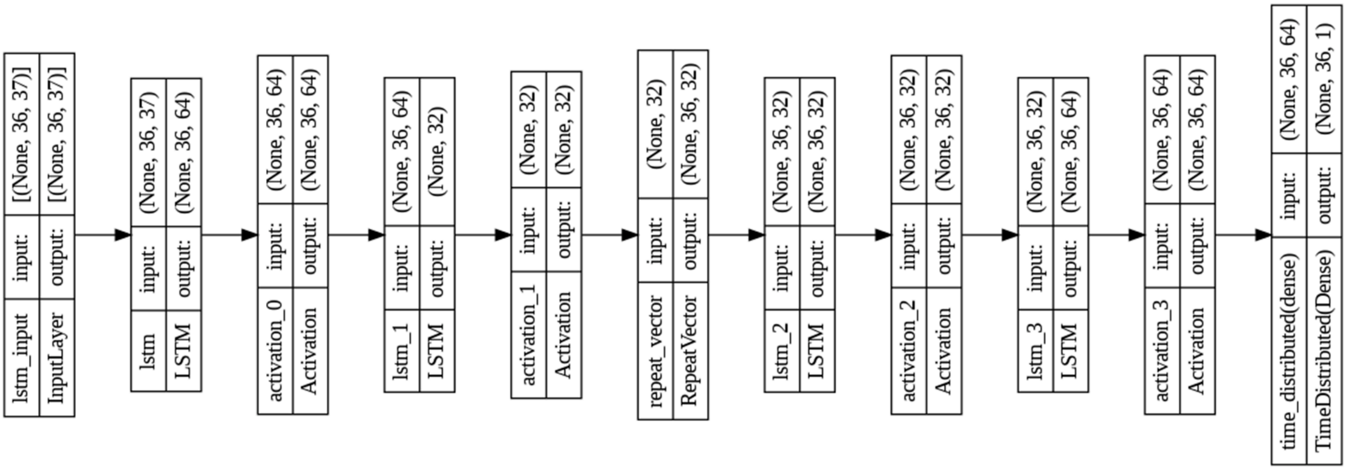

Fig. 8. Layers of LSTM-AutoEncoder proposed model.

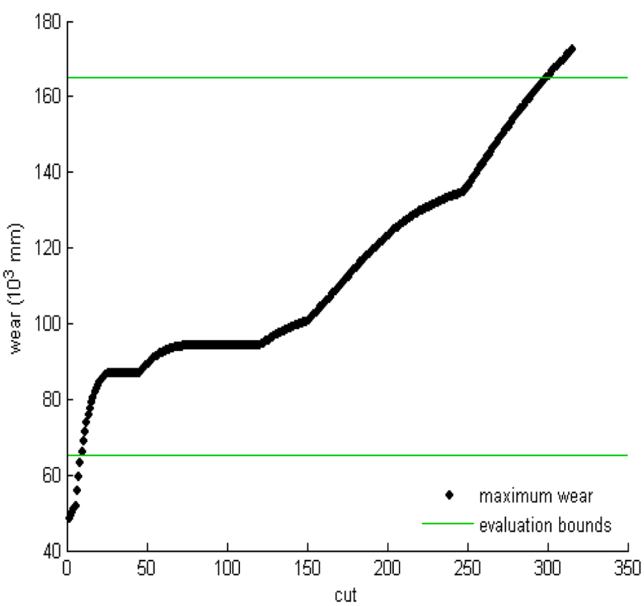

Fig. 9. Maximum wear for Cutter and Wear Limit.

The leaderboard of this competition can be found in [58]. After tool wear prediction. The RTF data can be used to estimate RUL i.e. the number of cuts that can be safely made before failure as it is shown in Fig. 10. The scoring function used for evaluation of the proposed model and that was used in PHM10 competition can be expressed by Eq. (3). This formula was proposed in [59]. The score value of our RUL estimation algorithm was reduced to about 40, while the winner in PHM10 competition score was 5500.

$$S_i(ifd < 0) = \left(e^{-d_{i/13}}\right) - 1$$
$$S_i(ifd \geq 0) = \left(e^{d/4.5}\right) - 1 \qquad (6)$$
$$Score = \sum_{1}^{m} s_i$$

Where m is the total number of data points, and di denotes the RUL´$_i$ – RUL (i.e., the calculated RUL value compared to the real value for each cycle). The lower the value of this score function, the better the prediction algorithm is. The accuracy of the RUL estimation approaches 99 % and low score value comparable to leaderboard of this competition.

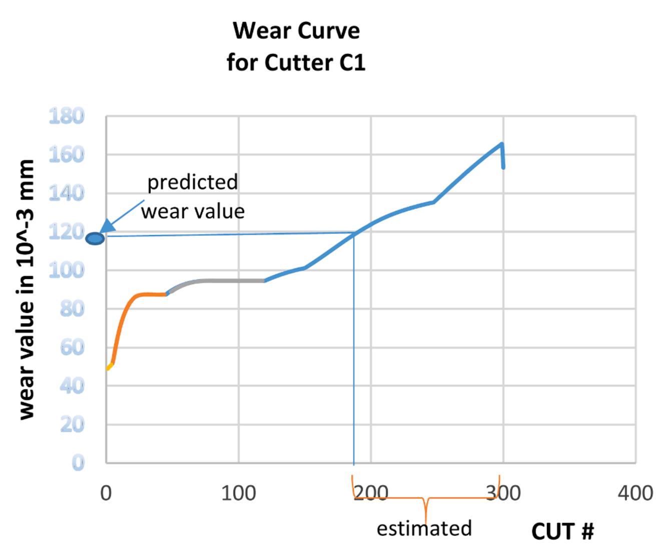

Fig. 10. RUL estimation method.

Table 4
RUL Estimation Algorithm.

Input predicted tool wear
Output RUL i.e. number of cuts the machine can safely made.
RUL estimation: Using predicted tool wear value with the nearest value against wear curve.
The intersection between predicted wear value and wear curve gives the current cut#.
RUL can be estimated as: $RUL_{est} = 300 - \boldsymbol{cut}\#$. as shown in Fig 10.

Table 5
Formula of applied criteria.

Pearson Correlation Coefficient	$PCC = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}}$ $\hat{y}_i$ ***is the predicted value***
Mean Absolute Error	$f_{MAE} = \sum_{i=1}^{N} \lvert \hat{y}_i - y_i \rvert / N$
Root Mean squared error	$f_{RMSE} = \sqrt{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2 / N}$

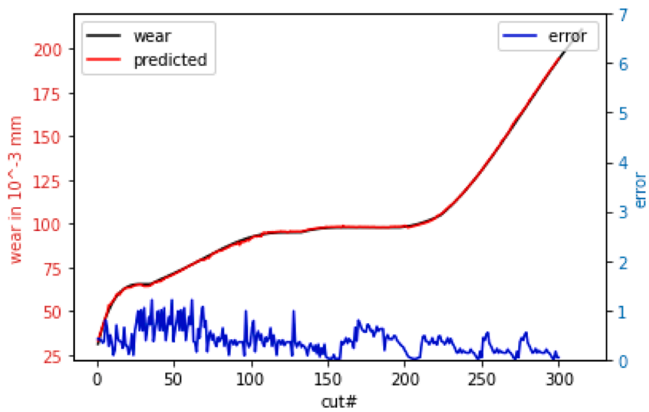

Fig. 11. Error between Predicted and Actual Wear for Cutter C4.

6. Experimental results

6.1. Performance evaluation criteria

Mean absolute error (MAE) and root mean squared error (RMSE) are the performance criteria used to address the total efficiency of the proposed LSTM-AE tool wear prediction model. Their mathematical formulas are shown in Table 5. Fig. 11 shows the predicted and actual value of tool wear for C4 that was used as the test set. After testing the method against the C4 data, the wear curve for other tools/cutters can be anticipated, and then the RUL for this cutter can be estimated. The model can then be leveraged for other cutters RUL estimation. The prediction error is reduced as the predicted value of wear approaches the actual value, and the LSTM-AE efficiency is better within the dimensional criteria listed above for performance assessment.

Table 6
Result Comparison.

Proposed mehod	MAE(10^{-3} mm)	RMSE (10^{-3} mm)
SVR+PCA [49]	3.9583±0.9371	5.4428±1.5894
RNN [30]	12.1667±6.2292	15.7333±6.2164
SVR [49]	9.3770±2.0422	11.968±3.3337
LSTM [30]	10.7333±3.8734	13.7333±4.5742
CBLSTM [30]	7.2333±1.0263	9.2333±1.9140
CNN [60]	11.0000±1.3000	14.0428±5.5588
Proposed (LSTM-AE)	2.6 ± 0.3222	3.1 ± 0.6146

Results from method in [49].

6.2. Method comparison

Fig. 12 shows the MAE and RMSE values of the cutter wear value for various methods. The LSTM-AE-based approach is contrasted with other approaches to show its efficacy and advancement. For forecasting tool wear, we specifically used conventional intelligence techniques like SVR and SVR+KPCA [49]. Additionally, for performance comparison, DL

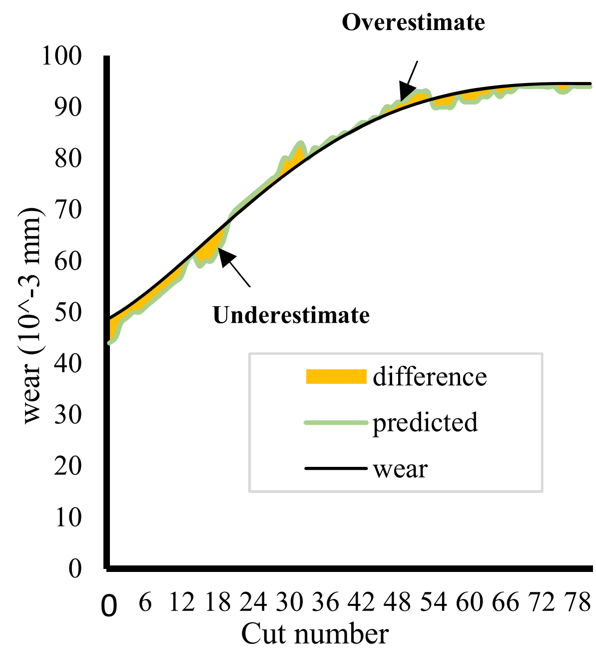

Fig. 13. Overestimation and Underestimation calculation of RTF curve.

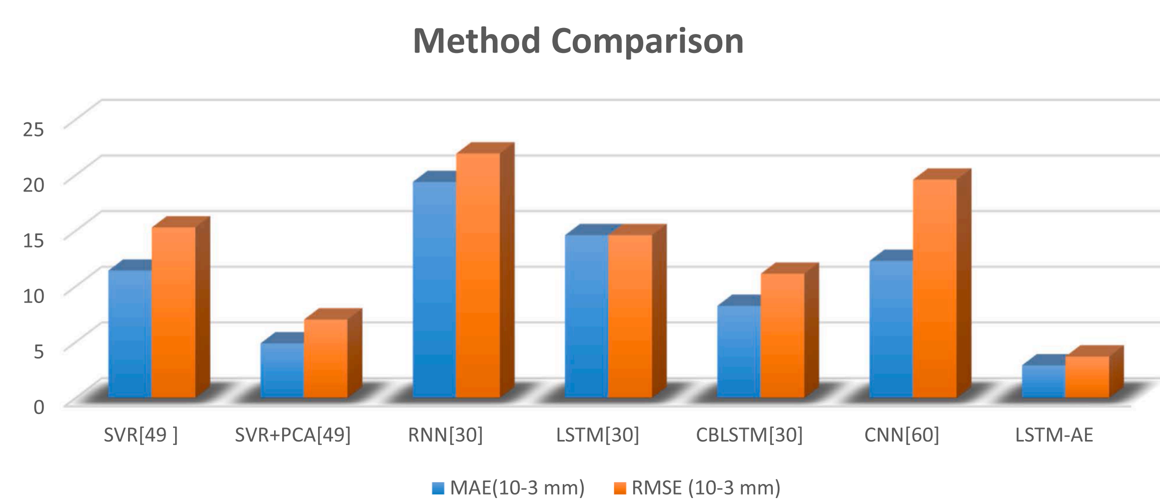

Fig. 12. Method Comparison between our LSTM-AE and earlier methods.

techniques, including RNN, LSTM, CBLSTM [30], and CNN [60], are also used. Table 6 provides the outcomes of the comparisons made using the proposed framework. As illustrated in Table 6, the SVR algorithm performs poorly when dealing with large-scale nonlinear regression problems. Even though deep learning techniques like RNN, LSTM, CBLSTM, and CNN may apply nonlinear regression without reducing the feature space, their projected accuracy is still inferior to the proposed framework.

6.3. Result discussion

Fig. 13 shows a comparison of expected and actual tool wear values for various cutter cuts during their life cycle. The figure also depicts the contrast between these two values in the darkened areas. The distinction is that some sections underestimate, while others overestimate. Underestimation occurs when anticipated values are lower than actual values, and the score function performs better in this scenario. Overestimation is defined as parts that exceed the actual tool wear curve. Returning to Fig. 11, based on the expected values for the entire curve, the overall predicted curve is underestimated. This suggests that the efficiency of the suggested framework to predict the RUL is less than the actual values to schedule for maintenance or component replacement before the failure occurs.

The score function is illustrated by Fig. 14 [59]. One special characteristic for this function is that in the case of underestimation, the penalty of the function is almost the same, then the error is lower and the performance is better, as the case is in the results for the proposed model as shown in Fig. 15.

6.4. Model generalization

The model can be generalized to any dataset with sensor data or a feature map that needs reduction or even any feature map to predict the target variable with commendable accuracy. If any operating condition exists, the operating condition can be treated as one component of the feature map or the values of this condition can be examined with its different cases to add new columns or features to the dataset.

The model may suffer with very long data from computational complexities and redundant information, thus lowering its performance. This problem may be solved by adding a weighting mechanism in the case of extremely long datasets. This weighting mechanism assigns different weights depending on how valuable the information is.

7. Conclusion

TCM is very crucial in industrial processes to help prevent rather than detect failure, thus reducing replacement costs and saving human lives. Time-series data have multi-variables and complex interconnections between variables along with dynamic features. This made traditional TCM very challenging particularly in the case of big sensor data. However, DL networks like LSTM and AE have promising characteristics to solve the complexity of such datasets. Hence, a hybrid DL model LSTM-AE is constructed using an encoder LSTM and an LSTM decoder for tool wear prediction. This model was evaluated for tool wear prediction of the cutters of a CNC milling machine. The PHM10 original dataset was used using C1, C4, and C6 cutters data for model training, testing, and validation.

An overall framework, including multiple steps: 1. Extracting features from different domains from the raw sensor data 2. A PCC correlation analysis to construct the feature map that is input to the LSTM model. The study focused on extracting distinguishing features of the cutter data, including adding new ones such as IQR and entropy. These features proved to have a high correlation to the tool wear using PCC, which helped improve the model performance. 3. The model is fed with the selected features to predict the target wear.

The suggested framework beats previous methods in tool wear prediction accuracy, achieving %98, as well as a distinguishable improvement in terms of RMSE and MAE for the test set compared to earlier methods. Using both the wear threshold and the tool's degradation curve, the expected wear value is used to determine RUL, ensuring underestimation of RUL values for most of the dataset. This helps prevent machine failure before it occurs. The model can be generalized to any dataset with sensor data or a feature map that needs reduction or even any feature map to predict the target variable with commendable accuracy. The model may suffer with very long data from computational complexities and redundant information, thus lowering its performance. This problem may be solved by adding a weighting mechanism in the case of extremely long datasets. This weighting mechanism assigns different weights depending on how valuable the information is. This may open new eras of research and applications on other datasets. Future research may also focus on providing a multi-objective optimization algorithm for best selecting the hyper-parameters of the model.

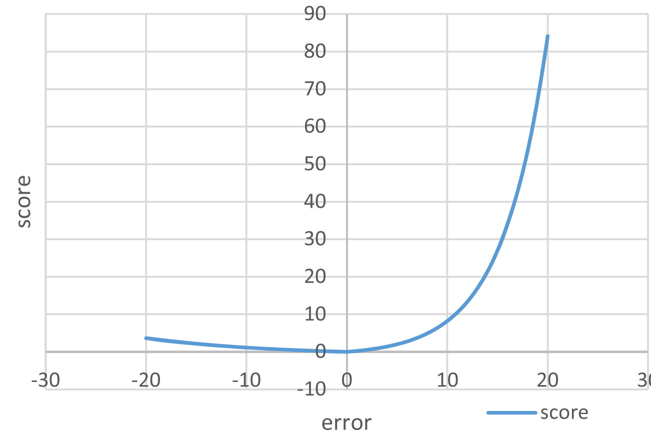

Fig. 14. Score fubnction against cut error.

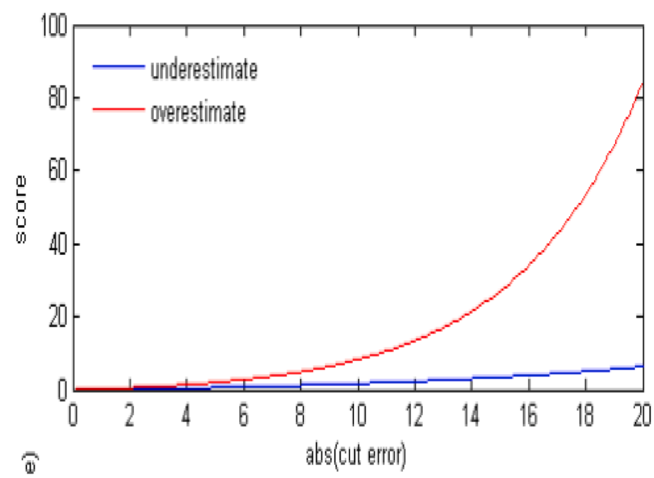

Fig. 15. Scoring function penalty for an over or under estimate as a function of the error [59].

CRediT authorship contribution statement

Hamdy K. Elminir: Writing – review & editing, Visualization, Conceptualization. **Mohamed A. El-Brawany:** Validation, Supervision, Investigation, Data curation. **Dina Adel Ibrahim:** Writing – original draft. **Hatem M. Elattar:** Software, Methodology, Resources, Visualization. **E.A. Ramadan:** Visualization, Validation, Supervision.

Declaration of competing 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.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this **manuscript.**

Data availability

Data will be made available on request.

References

[1] H.-T. Yau, C.-C. Wang, J.-Y. Chang, X.-Y. Su, A Study On the Application of Synchronized Chaotic Systems of Different Fractional Orders For Cutting Tool Wear Diagnosis and Identification, 7, IEEE, 2019, pp. 15903–15911. Access.
[2] K. Javed, R. Gouriveau, X. Li, N. Zerhouni, 'Tool wear monitoring and prognostics challenges: a comparison of connectionist methods toward an adaptive ensemble model, J. Intell. Manuf. 29 (8) (2016) 1873–1890.
[3] N. Clements, Introduction to prognostics," tutorial, in: Proc. Annu. Conf, Prognostics Health Manage. Soc., 2011.
[4] Anthony D. Scaife, Improve predictive maintenance through the application of artificial intelligence: a systematic review, Results. Eng. (2023) 101645.
[5] Dazhong Wu, et al., Data-driven prognostics using random forests: prediction of tool wear, in: International Manufacturing Science and Engineering Conference 50749, American Society of Mechanical Engineers, 2017, https://doi.org/10.1115/MSEC2017-2679.
[6] Dazhong Wu, et al., A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests, J. Manuf. Sci. Eng. 139 (7) (2017).
[7] Chico Hermanu Brillianto Apribowo, et al., Early prediction of battery degradation in grid-scale battery energy storage system using extreme gradient boosting algorithm, Results. Eng. 21 (2024) 101709.
[8] Abdullah. Alwabli, From Data to Durability: evaluating Conventional and Optimized Machine Learning Techniques for Battery Health Assessment, Results. Eng. (2024) 102445.
[9] Sadiqa Jafari, Ji-Hyeok Yang, Yung-Cheol Byun, Optimized XGBoost modeling for accurate battery capacity degradation prediction, Results. Eng. (2024) 102786.
[10] Nima Hajimirza Amin, Alireza Etemad, Ashkan Abdalisousan, Data-driven performance analysis of an active chilled beam air conditioning system: a machine learning approach for energy efficiency and predictive maintenance, Results. Eng. 23 (2024) 102747.
[11] Aitzaz Ahmed Murtaza, et al., Paradigm Shift for Predictive Maintenance and Condition Monitoring from Industry 4.0 to Industry 5.0: a Systematic Review, Challenges and Case Study, Results. Eng. (2024) 102935.
[12] Achmad Widodo, Bo-Suk Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal. Process. 21 (6) (2007) 2560–2574.
[13] Mohd Danish, et al., Machine learning models for prediction and classification of tool wear in sustainable milling of additively manufactured 316 stainless steel, Results. Eng. 22 (2024) 102015.
[14] Tabassum Naz Sindhu, et al., A decreasing failure rate model with a novel approach to enhance the artificial neural network's structure for engineering and disease data analysis, Tribol. Int. 192 (2024) 109231.
[15] Andaç Batur Çolak, et al., A comparative analysis of maximum likelihood estimation and artificial neural network modeling to assess electrical component reliability, Qual. Reliab. Eng. Int. 40 (1) (2024) 91–114.
[16] Anum Shafiq, et al., Modeling and survival exploration of breast carcinoma: a statistical, maximum likelihood estimation, and artificial neural network perspective, Artif. Intell. Life Sci. 4 (2023) 100082.
[17] Anum Shafiq, et al., "Comparative study of artificial neural network versus parametric method in COVID-19 data analysis", Results. Phys. 38 (2022) 105613.
[18] Andaç Batur Çolak, et al., "Reliability study of generalized Rayleigh distribution based on inverse power law using artificial neural network with Bayesian regularization", Tribol. Int. 185 (2023) 108544.
[19] Anum Shafiq, Andaç Batur Çolak, Tabassum Naz Sindhu, "Optimization of the numerical treatment of the Darcy–Forchheimer flow of Ree–Eyring fluid with chemical reaction by using artificial neural networks", Int. J. Numer. Methods Fluids. 95 (1) (2023) 176–192.
[20] Anum Shafiq, Andaç Batur Çolak, Tabassum Naz Sindhu, "Significance of bioconvective flow of MHD thixotropic nanofluid passing through a vertical surface by machine learning algorithm", Chin. J. Phys. 80 (2022) 427–444.
[21] Anum Shafiq, Andaç Batur Çolak, Tabassum Naz Sindhu, "Optimization of bioconvective magnetized Walter's B nanofluid flow towards a cylindrical disk with artificial neural networks", Lubricants. 10 (9) (2022) 209.
[22] Tabassum Naz Sindhu, et al., "Reliability study of generalized exponential distribution based on inverse power law using artificial neural network with Bayesian regularization", Qual. Reliab. Eng. Int. 39 (6) (2023) 2398–2421.
[23] Anum Shafiq, et al., Reliability analysis based on mixture of lindley distributions with artificial neural network, Adv. Theory. Simul. 5 (8) (2022) 2200100.
[24] Siguo Bi, et al., "A comprehensive survey on applications of AI technologies to failure analysis of industrial systems", Eng. Fail. Anal. 148 (2023) 107172.
[25] Mohamed A. El-Brawany, et al., "Artificial intelligence-based data-driven prognostics in industry: a survey", Comput. Ind. Eng. 184 (2023) 109605.
[26] Hosameldin Eltayeb A. Adam, James K. Kimotho, Jackson G. Njiri, Multiple faults diagnosis for an industrial robot fuse quality test bench using deep-learning, Results. Eng. 17 (2023) 101007.
[27] Zhiwen Huang, et al., "Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations", J. Intell. Manuf. 31 (4) (2020) 953–966, https://doi.org/10.1007/s10845-019-01488-7.
[28] Zhiwen Huang, et al., Tool wear predicting based on multisensory raw signals fusion by reshaped time series convolutional neural network in manufacturing, IEEE Access. 7 (2019) 178640–178651, https://doi.org/10.1109/ACCESS.2019.2958330.
[29] Qinglong An, et al., A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network, Measurement 154 (2020) 107461, https://doi.org/10.1016/j.measurement.2019.107461.
[30] Rui Zhao, et al., Learning to monitor machine health with convolutional bi-directional LSTM networks, Sensors 17 (2) (2017) 273, https://doi.org/10.3390/s17020273.
[31] Zhengmin Kong, et al., Convolution and long short-term memory hybrid deep neural networks for remaining useful life prognostics, Appl. Sci. 9 (19) (2019) 4156, https://doi.org/10.3390/app9194156.
[32] Jiwei Wang, et al., Early Prognostics of Lithium-Ion Battery Pack Health, Sustainability. 14 (4) (2022) 2313.
[33] Reza Rouhi Ardeshiri, Ming Liu, Chengbin Ma, Multivariate stacked bidirectional long short term memory for lithium-ion battery health management, Reliab. Eng. Syst. Saf. 224 (2022) 108481.
[34] Jiujian Wang, et al., Remaining useful life estimation in prognostics using deep bidirectional lstm neural network, in: 2018 Prognostics and system health management conference (PHM-Chongqing), IEEE, 2018, https://doi.org/10.1109/PHM-Chongqing.2018.00184.
[35] Maan Singh Rathore, S.P. Harsha, Rolling bearing prognostic analysis for domain adaptation under different operating conditions, Eng. Fail. Anal. 139 (2022) 106414.
[36] Fouzi Harrou, et al., Enhancing Road Traffic Flow Prediction with Improved Deep Learning using Wavelet Transforms, Results. Eng. (2024) 102342.
[37] Fouzi Harrou, et al., Energy consumption prediction in water treatment plants using deep learning with data augmentation, Results. Eng. 20 (2023) 101428.
[38] Jawad Mahmood, Ming Luo, Mudassar Rehman, An accurate detection of tool wear type in drilling process by applying PCA and one-hot encoding to SSA-BLSTM model, Int. J. Adv. Manufact. Technol. 118 (11) (2022) 3897–3916.
[39] Minghui Cheng, et al., Intelligent tool wear monitoring and multi-step prediction based on deep learning model, J. Manuf. Syst. 62 (2022) 286–300.
[40] Lei Nie, et al., Remaining Useful Life Prediction of Milling Cutters Based on CNN-BiLSTM and Attention Mechanism, Symmetry. (Basel) 14 (11) (2022) 2243.
[41] Ning Zhang, et al., A novel hybrid model integrating residual structure and bi-directional long short-term memory network for tool wear monitoring, Int. J. Adv. Manufact. Technol. 120 (9) (2022) 6707–6722.
[42] Haidong Shao, et al., A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Signal. Process. 95 (2017) 187–204, https://doi.org/10.1016/j.ymssp.2017.03.034.
[43] Zong Meng, et al., An enhancement denoising autoencoder for rolling bearing fault diagnosis, Measurement 130 (2018) 448–454, https://doi.org/10.1016/j.measurement.2018.08.010.
[44] Z. Zhang, S. Li, Y. Xiao, Y. Yang, Intelligent simultaneous fault diagnosis for solid oxide fuel cell system based on deep learning, Appl. Energy 233-234 (October) (2019) 930–942, https://doi.org/10.1016/j.apenergy.2018.10.113.
[45] Kamal Berahmand, et al., Autoencoders and their applications in machine learning: a survey, Artif. Intell. Rev. 57 (2) (2024) 28.
[46] Yishun Liu, et al., Non-ferrous metals price forecasting based on variational mode decomposition and LSTM network, Knowl. Based. Syst. 188 (2020) 105006.
[47] Fuqiang Gu, et al., Accurate step length estimation for pedestrian dead reckoning localization using stacked autoencoders, IEEE Trans. Instrum. Meas. 68 (8) (2018) 2705–2713.
[48] Agogino, K. Goebel, Mill data set. NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007. http://ti.arc.nasa.gov/project/prognostic-datarepository.
[49] Jinjiang Wang, et al., Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing, Robot. Comput. Integr. Manuf. 45 (2017) 47–58.
[50] Tarak Benkedjouh, et al., Health assessment and life prediction of cutting tools based on support vector regression, J. Intell. Manuf. 26 (2) (2015) 213–223.
[51] Jun Wu, et al., Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system, Appl. Soft. Comput. 68 (2018) 13–23.
[52] Cunji Zhang, et al., Tool condition monitoring and remaining useful life prognostic based on a wireless sensor in dry milling operations, Sensors 16 (6) (2016) 795, https://doi.org/10.3390/s16060795.
[53] Mirjana. Čižmešija, Five-number summaries. International Encyclopedia of Statistical Science, Springer, Berlin, Heidelberg, 2011, pp. 526–527.

[54] Marília Barandas, Duarte Folgado, et al., TSFEL: time Series Feature Extraction Library, SoftwareX. 11 (2020), https://doi.org/10.1016/j.softx.2020.100456.
[55] Shaojiang Dong, Tianhong Luo, Bearing degradation process prediction based on the PCA and optimized LS-SVM model, Measurement 46 (9) (2013) 3143–3152.
[56] Alkan Alkaya, İlyas Eker, Variance sensitive adaptive threshold-based PCA method for fault detection with experimental application, ISA Trans. 50 (2) (2011) 287–302.
[57] Mohamed El-Barawany, et al., Computer Numerical Control CNC Machine Health Prediction using Multi-domain Feature Extraction and Deep Neural Network Regression, J. Eng. Res. (Ponta Grossa) 6 (5) (2022) 7–12.
[58] https://www.phmsociety.org/competition/phm/10/leaderboard. 2010.
[59] A. Saxena, et al., Metrics for evaluating performance of prognostic techniques, in: Proc. Int. Conf. Prognostics Health Manage. (PHM), Oct. 2008, pp. 1–17.
[60] Rui Zhao, et al., Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213–237, https://doi.org/10.1016/j.ymssp.2018.05.050.