



An accurate detection of tool wear type in drilling process by applying PCA and one-hot encoding to SSA-BLSTM model

Jawad Mahmood^{1,2} · Ming Luo^{1,2} · Mudassar Rehman²

Received: 15 March 2021 / Accepted: 10 October 2021 / Published online: 20 October 2021
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Tool condition monitoring (TCM) is significant in advanced manufacturing systems for achieving high productivity in the manufacturing industries. The main objective of the research study is to design a high-quality tool wear detection system. This methodology was experimentally executed by installing the dynamometer to the CNC drilling machine to perform drilling operations using the normal tool, crater wear tool, chisel wear tool, flank wear tool, and outer corner wear tool. The removal of noise from the raw force signal data and extraction of features was done by using the singular spectrum analysis (SSA) algorithm. The algorithm based on the technique of principal component analysis (PCA) is designed for dimensionality reduction and for improving performance, accuracy, and efficiency by avoiding the overfitting of the model. Early stopping and dropout algorithms are designed to effectively overcome the overfitting problem. Both techniques have made the training process efficient by automatic selection of the most suitable number of epochs. The textual form of target variables of the model was converted into binary numerical form by using one-hot encoding as the deep learning algorithm can read numerical data only. The model would determine whether the test tool is worn or not and would also predict the type of tool wear and achieved an accuracy of 97.94%.

Keywords Drilling · Tool wear · SSA-BLSTM · PCA

Nomenclature

AI	Artificial intelligence
ANN	Artificial neural networks
ARIMA	Autoregressive integrated moving average
BLSTM	Bidirectional long short-term memory network
BGRU	Bidirectional gated recurrent unit
CNN	Convolutional neural network
LSTM	Long short-term memory network
MLP	Multilayer perceptron
PCA	Principal component analysis
RF	Random forest
RNN	Recurrent neural network
RT	Regression trees

RUL	Remaining useful life
SMOTE	Synthetic minority over-sampling technique
SVM	Support vector machine
SSA	Singular spectrum analysis
TCM	Tool condition monitoring

1 Introduction

The manufacturing industry has gone through several developments over time and is constantly moving toward the smart industry because of the integration of information technology. Still, the industries are facing challenges to improve their methods of increasing productivity, sustainability quality, and reduction in overhead costs [1–3]. Drilling is considered a complex manufacturing process in the manufacturing industry. The twist drills are mostly used in around 72% of drilling operations [4]. Tool wear is the main problem associated with drilling and other machining operations like turning and milling. Tool wear may occur by performing machining operations on hard materials like nickel (Ni)-based alloys because of their high strength and toughness which may lead to generate greater cutting

✉ Ming Luo
luoming@nwpu.edu.cn

¹ Key Laboratory of High-Performance Manufacturing for Aero Engine, Northwestern Polytechnical University, Ministry of Industry and Information Technology, Xi'an 710072, People's Republic of China

² School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China

forces, complex kinematics of the operation, and also high temperature during machining. The increase in temperature during machining causes the loss of edge geometry of the cutting tool which is termed as the “tool wear.” Tool wear depends upon several physical parameters such as the material, shape of the cutting tool, type of coolant used, feed rate, cutting speed, and depth of cut. An increase in the tool wear reduces the tool life which also imparts severe impacts on the machined workpiece in terms of drill quality, chip clogging, and many other similar issues. Therefore, early prediction of tool degradation assists in avoiding the failures and also helps in alleviating the manufacturing costs and improves machining productivity.

Interactions along cutting tool-workpiece are crucial for the tool wear mechanism. In drilling, the cutting speed (surface speed) and feed rate are the key process variables affecting the tool wear. Furthermore, the material and tool factors like geometry, hardness, and machining complexity play a vital role in tool wear. A tool material should possess three key properties including wear resistance, toughness, and hot hardness [5]. Various tool materials like high-speed steels, ceramics, carbides, and diamonds are used depending on the hardness of the workpiece. Celik et al. [6] employed the Tungsten carbide and high-speed steel tool to perform drilling of Ti-6Al-4V, their results show that feed rate and cutting forces have a major impact on tool wear. It was also concluded that the tungsten carbide drill tool was superior to machine the high strength alloy (Ti-6Al-4 V) than high-speed steel tool in terms of increased tool life. Iliescu et al. [7] performed drilling of CFRP using tungsten carbide tool that resulted in the higher influence of cutting speed and tool wear on the thrust force. Dehghan et al. [8] performed a comparative study of three materials, namely IN718, Ti-6Al-4 V, and AISI 304 steel, for the investigation of tool performance in drilling. It was inferred that spindle speed is a highly significant variable for all the said materials compared to the feed rate. Wang et al. [9] evaluated the wear mechanism in the drilling of Mg alloys using a high-speed steel tool and inferred that three wear mechanisms, namely adhesive wear, abrasive wear, and diffusion wear, were produced after the machining.

Artificial neural network (ANN)-based methods are effective for the prediction of tool wear. Karandikar et al. [10] briefly described the sensor integration technique and employed the ANN algorithm for the monitoring of the tool wear, and it has given 97% accuracy. D'Addona et al. [11] have worked on the prediction of tool wear and pattern recognition by employing ANN algorithm and DNA-based computing. ANN has detected the tool wear from the worn tool region in images, whereas DNA-based computing has identified the similarity as well as dissimilarity of processed images. Palanisamy et al. [12] investigated the tool wear prediction using the ANN model and regression mathematical

model during the end-milling operations. The data values obtained from the experiment were used to train the ANN feed forward propagation. The output values predicted by both models were compared with experimental values. Chen et al. [13] demonstrated the in-process tool wear prediction methodology using the ANN model during a 3-insert milling operation based on a real-time approach. Experimental data was used to train the ANN back propagation model. ANN is considered to have the good capability of learning and modeling complex and nonlinear relationships and sorting information over the entire network. ANN has some limitations as the network duration is not defined. During training, the network is reduced to the particular value of error, and it does not give the desired output results.

Deep learning methods use architectures of neural networks which consist of a training phase and an interfering phase. Li et al. [14] have developed the model to predict the tool wear state by using the feature-based transfer learning technique. The group of highly correlated sensor features was selected by using the generic algorithm. After transferring the features, the similarity score of features was calculated. SVM coupled with PSO was used for the prediction of tool wear states. Mun and Jeong [15] have designed the model for the prediction of the RUL of the tool based on tool wear estimation of CNC machine cutting tools by using CNN and BLSTM algorithms. Wu et al. [16] have introduced the prognostic method based on the RF for tool wear prediction and compared the results with other machine learning algorithms such as ANN, SVM, and feed forward back propagation. The designed model has given better results as compared to other algorithms. Qian et al. [17] have presented the new predictive model for tool wear by combining the techniques of SVM and surface texture analysis of workpiece with the designed generic algorithm. The obtained experimental results of turning results have shown the effectiveness of the model. Martínez-Arellano et al. [18] have presented the big data approach for the classification of tool wear based on deep learning and signal imaging. The approach has worked directly with raw data after the combination of these two techniques. The model has identified the intrinsic features of raw data and given 90% classification accuracy. Chen et al. [19] have employed the technique of deep belief network for the prediction of tool wear by using multi-sensor data. The model has given better performance as compared to ANN and support vector regression (SVR). Glass and Colbaugh [20] developed the strategy to estimate the tool wear during the metal cutting process. The purposed algorithm consists of two components such as RNN and robust layer. CNN possesses the ability to effectively extract and learn the relevant features from the particular image and to perform the mapping of image data into the output variable. However, CNN methods face some limitations to perform encoding of the orientation and the object. RNN

is considered suitable to predict the time series data and can memorize the information throughout the entire time period. RNN also finds complications of gradient vanishing and exploding and also faces limitations in processing very long data sequences.

The aim of TCM is to use sensor signal processing to monitor the tool wear state for minimizing the losses caused by tool damage. Mikolajczyk et al. [21] have worked on the two-step method to automatically predict the tool life during the turning operation. Bustillo et al. [22] have used various machine learning techniques for the automatic prediction of flatness deviation with appropriate consideration to the wear amount of insert's edge of the cutting tool. Different algorithms of machine learning were examined. Synthetic minority oversampling technique (SMOTE) combined with random forest showed good performance. Pimenov et al. [23] have used the methods of artificial intelligence to predict automatically surface roughness deviations by taking tool wear into account. A total of three parameters are estimated which are cutting power, machining time, and maximum tool wear. A series of algorithms were applied such as RT, multilayer perceptron (MLP), radial-based functions, and random forest (RF). RF has given the highest accuracy, whereas RT has also shown better results than MLP and radial-based functions. Kuntoglu et al. [24] have reviewed the systems based on indirect TCM and decision-making methodologies during the turning operation. The main advantage of tool condition monitoring is that it improves the quality of the machined surface by avoiding tool wear. The major drawback of TCM is that it is not cost-effective because high-quality sensors are required and a lot of algorithms are needed to be designed to implement TCM.

Long short-term memory network (LSTM) possesses the capability to learn the long sequences and can do the processing of the entire sequence of data. LSTM cell can remember the data values for an arbitrary time period, so this network is considered effective for tool wear prediction. Zhang et al. [25] have developed the methodology based on gated RNN including LSTM with long-term dependencies for prognosis and diagnosis of tool wear. Multiple experiments were conducted using the milling cutting tool for online diagnosis. The model has achieved better accuracy with minimum mean square error. Zhang et al. [26] have designed an ARIMA model for processing time series data and employed an LSTM algorithm for the prediction of the tool wear. Hao and Kunpeng [27] have worked on the LSTM auto-encoder for tool wear monitoring. LSTM was used to extract the tool wear features from the monitoring signals, and unlabeled signals were trained by using an auto-encoder to avoid information loss. High-speed milling validation has verified the effectiveness of the methodology. Ma et al. [28] have analyzed three stages of tool wear according to chemical elements and micro-topography and designed the tool wear model by a combination of deep learning algorithms

such as CNN was combined with BLSTM and then CNN was combined with BiGRU, and the results have shown that error of predicted values was less than 8%; thus, model has achieved better results. Qiao et al. [29] have designed the tool wear prediction system based on fog computing and deep learning algorithms. The task of tool wear monitoring was accomplished by multi-scaled LSTM, whereas BLSTM has worked for the prediction of tool wear. Both deep learning models were employed in fog architecture for the reduction of response latency. The main advantages of the LSTM network include keeping the past contextual information from the previous intervals and can learn the long sequences of data without having the problems of vanishing gradient. However, this algorithm requires more memory to train.

The main purpose of this study is to determine the tool wear type by using the novel bidirectional multi-target classification technique with high accuracy. Previous studies have not addressed the prediction of tool wear type, dimensionality reduction in tool wear model, and elimination of overfitting. The bidirectional multi-target classification model is developed by designing the advanced algorithms w.r.t. experimental data to detect tool wear type. The SSA-BLSTM model can preserve and process both past and future contextual information of the worn tool and can learn the long sequences of data. A model can propagate forward as well as backward to update weights for achieving better performance and possess a greater processing speed and has the great ability to get trained in the least time on the experimentally obtained dataset. SSA has removed noise from the signal and multiple features extracted. PCA has reduced dimensions and improved the model's performance and efficiency. The early stopping and dropout techniques are applied to eliminate the problem of overfitting. Drilling tests were performed on a nickel-based superalloy using a normal tool and different types of worn tools for the collection of force signal data. The following contents are divided into five sections. The first section is based on an introduction followed by a discussion of different types of tool wear in Sect. 2. The third section contains an analytical investigation using mathematical formulas. An experimental investigation was performed in the fourth section followed by the conclusion.

2 Classification of tool wear types in drilling operation

The standard twist drill has various cutting edges like chisel edge, cutting lips (main cutting edge), margin edges, and outer corner [30].

- Flank wear: An uneven wear land that appears on the flank face of the tool after drilling is called flank wear. Flank wear is produced as a result of friction present at

the contact interface of the tool and workpiece during machining. Flank wear is the common type of tool wear that occurs more frequently.

- Crater wear: This type of tool wear emerges on a rake face and uneven width appears. The crater wear gets widened at a position close to the drill axis. Crater wear occurs due to the application of a large amount of pressure on a tool chip interface, spiral chips generated during the drilling process are thicker at the position close to the drill axis. Thus, it resulted in high deformation resistance near the drill axis which produces a large amount of normal and shear stress applied on a rake face to curl a chip leading to the wider wear land.
- Chisel wear: This type of wear normally occurs around the center of a tool. Chisel wear is usually caused by huge compressive stresses in a flow zone between the workpiece material and the chisel. This effect is particularly prominent at higher feed rates resulting in the erosion of the chisel edge.
- Outer corner wear: This type of wear is formed on the outer corners of the drill tool due to rubbing between the machined hole wall and the outer corner. The highest cutting speed at the marginal point of cutting lips results in high friction generated on outer corners due to increased temperature at the contact area.

3 BLSTM bidirectional processing-based force signal classification

LSTM network contains both hidden state and cell state. The purpose of forget gate is to delete the data from the cell state C_{t-1} . The data which is useless for the LSTM is removed by the multiplication filter. There are two inputs for the forget gate; the first one is x_t , and another one is h_{t-1}

where x_t symbolized as the input at a specific time step, as shown in Fig. 1.

The bidirectional LSTM network has two LSTM networks. The first and second networks can propagate in forward as well as backward directions, respectively. Both networks have the access to future as well as past contextual information. The resulting output is generated based on the future and past contexts. BLSTM network can give high accuracy by the classification of the time series sequences:

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

$$f_t^{\rightarrow} = \sigma(W_f^{\rightarrow}[h_{t-1}^{\rightarrow}, x_t^{\rightarrow}] + b_f^{\rightarrow}) \quad (2)$$

$$C_t^{\rightarrow} = \tanh(W_c^{\rightarrow}[h_{t-1}^{\rightarrow}, x_t^{\rightarrow}] + b_c^{\rightarrow}) \quad (3)$$

$$C_t^{\leftarrow} = i_t^{\leftarrow} C_t^{\rightarrow} + f_t^{\leftarrow} C_{t-1}^{\leftarrow} \quad (4)$$

$$o_t^{\rightarrow} = \sigma(W_o^{\rightarrow}[h_{t-1}^{\rightarrow}, x_t^{\rightarrow}] + b_o^{\rightarrow}) \quad (5)$$

$$h_t^{\rightarrow} = (o_t^{\rightarrow})\tanh(C_t^{\rightarrow}) \quad (6)$$

where i_t^{\rightarrow} , f_t^{\rightarrow} , and o_t^{\rightarrow} are the input gate, forget gate, and output gate for the forward direction and h_{t-1}^{\rightarrow} and x_t^{\rightarrow} are the two inputs for the forward input gate and σ is the sigmoid function and b_i^{\rightarrow} is the bias signal. This is the forward direction input gate. The main function of the gates in the LSTM network is to control the information to let it pass or block it. The symbol x_t^{\rightarrow} denotes the input of new force signals, whereas h_{t-1}^{\rightarrow} is the input of previous force signals obtained by using a dynamometer. The two inputs are concatenated together, and weight W_i^{\rightarrow} and bias signals b_i^{\rightarrow} are applied to

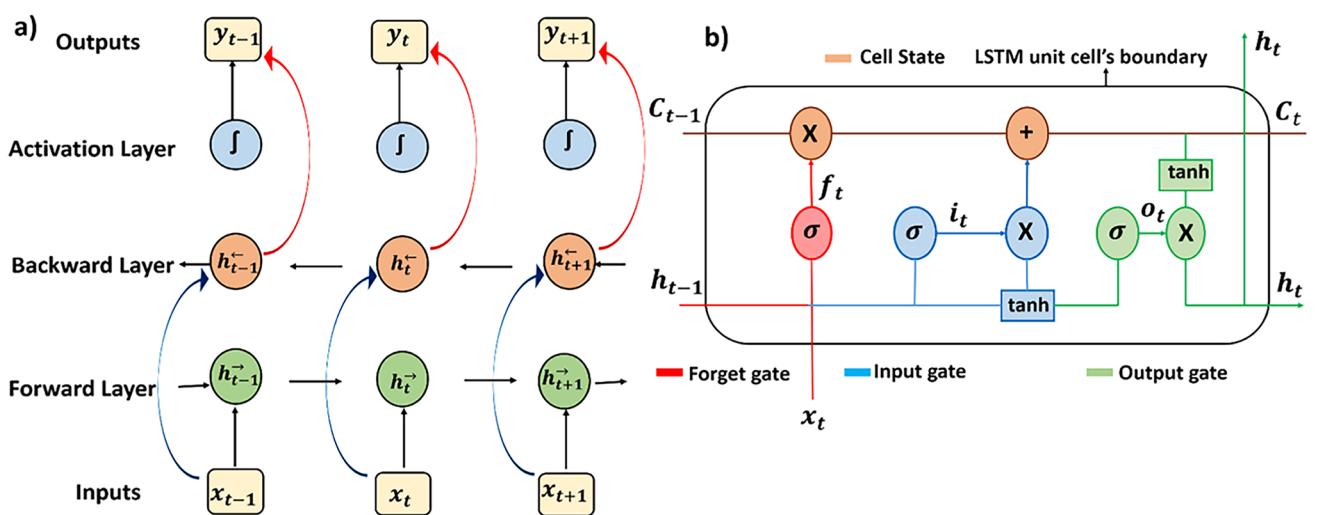


Fig. 1 a Demonstration of BLSTM backward processing. b Schematic representation of the internal architecture of LSTM

them, and the resultant force signal of the tool is controlled by a sigmoid activation function whose values range from 0 to 1. If the value of the sigmoid function is 0, then the force signal is blocked and will not pass on, and if the value of the sigmoid activation function is 1, then the force signals will pass on. h_t^{\rightarrow} is the output value for the forward gate whose value is obtained when the tanh function is applied to the forward cell state multiplying the resulting value with the forward output gate. The complete output result will be in the form of classification accuracy.

After the process of forward propagation and obtaining the output value that is the predicted value, the calculation of loss is needed. For the calculation of error or loss, the predicted value needs to be compared with the actual output result. The loss function is used for the calculation of the error. The derivative of this error needs to be calculated with respect to weight in the neural network. Stochastic gradient descent will minimize the loss by calculating the gradient of the loss function and by updating the weight of the neural network. A gradient is calculated by back propagation. The bidirectional algorithm will process the input in bidirectional ways. The most important processing is from past information of worn tool and doing the prediction of the future outcomes related to that specific tool. It also saves the contextual information of past and future related to the worn tool. The algorithm is trained on the force signals of normal tools and different types of worn tools. BLSTM neural network retains the information of the tool's condition in terms of past and future:

$$i_t^{\leftarrow} = \sigma(W_i^{\leftarrow}[h_{t-1}^{\leftarrow}, x_t^{\leftarrow}] + b_i^{\leftarrow}) \quad (7)$$

$$f_t^{\leftarrow} = \sigma(W_f^{\leftarrow}[h_{t-1}^{\leftarrow}, x_t^{\leftarrow}] + b_f^{\leftarrow}) \quad (8)$$

$$C_t^{\leftarrow} = \tanh(W_c^{\leftarrow}[h_{t-1}^{\leftarrow}, x_t^{\leftarrow}] + b_c^{\leftarrow}) \quad (9)$$

$$C_t^{\leftarrow} = i_t^{\leftarrow} C_{t-1}^{\leftarrow} + f_t^{\leftarrow} C_{t-1}^{\leftarrow} \quad (10)$$

$$o_t^{\leftarrow} = \sigma(W_o^{\leftarrow}[h_{t-1}^{\leftarrow}, x_t^{\leftarrow}] + b_o^{\leftarrow}) \quad (11)$$

$$h_t^{\leftarrow} = (o_t^{\leftarrow}) \tanh(C_t^{\leftarrow}) \quad (12)$$

where i_t^{\leftarrow} , f_t^{\leftarrow} , and o_t^{\leftarrow} are the input gate, forget gate, and output gate for backward direction, respectively. The above equations are for BLSTM backward processing, h_t^{\rightarrow} is the output value for the backward gate, and h_{t-1}^{\leftarrow} and x_t^{\leftarrow} are the two inputs for the backward input gate. The input test tool's force signals x_t^{\leftarrow} would pass through the BLSTM neural network in the backward direction. If the features of the test tool's force signals match the features of signals during the backward processing of the trained BLSTM neural network, then the test tool would be classified with respect to that match. When the

algorithm will have the information of the tool's condition of past and future, then it will be easy for the BLSTM algorithm to classify the test tool's force signals into the category of normal tool or different types of worn tools. The mathematical expression reported in Eq. (3) indicated the total output value of the forward and backward network gate as:

$$h_t = h_t^{\rightarrow} + h_t^{\leftarrow} \quad (13)$$

The test tool's force signals would pass through the entire trained BLSTM neural network in forward processing as well as backward processing. During the forward processing, the final classification accuracy would be equal to h_t^{\rightarrow} if features of the test tool's force signals matched to the particular features of the signals on which the BLSTM neural network is trained and vice versa for backward processing. The final output value would be obtained by the sum of output values got from the forward network and backward network. BLSTM neural network is better as compared to the traditional LSTM network as it can give high output accuracy and can do both forward and backward processing. Because it has two networks, forward network and backward network, it can be used to predict how much tool wear can occur in the future. The flow chart for the algorithm is shown in Fig. 2.

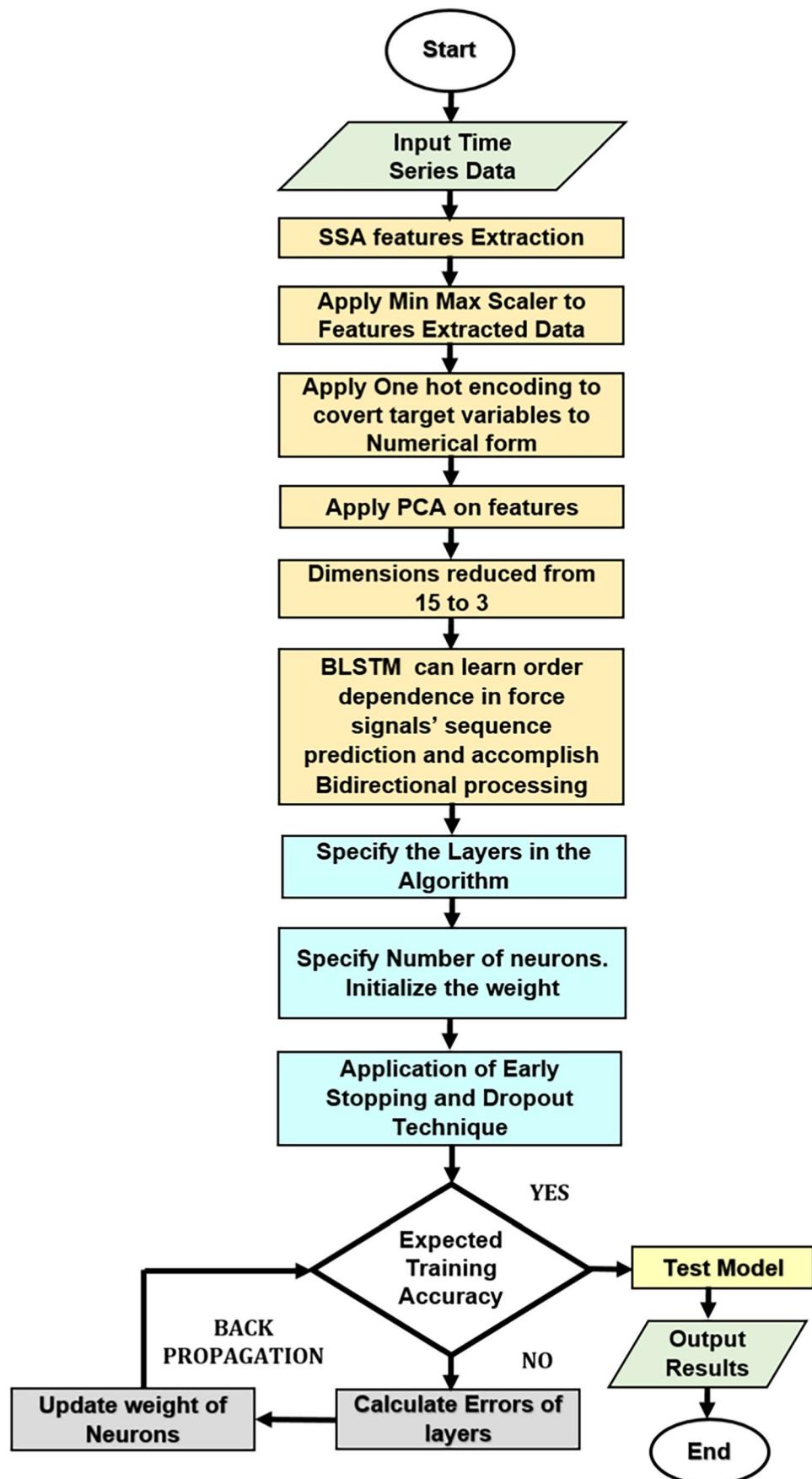
The flowchart has demonstrated the important steps of designing the algorithm of the model. The raw time series force signals were processed by the SSA algorithm for noise removal and useful feature extraction. The numerical data values of extracted features were scaled by using the tool of Min. Max. scaler. The textual target variables of the scaled data were converted into binary form by using the one-hot encoding technique as the machine learning algorithms can interpret only numerical data values. The resultant data has higher dimensions which lead to have adverse effects on the efficiency of the model and cause overfitting. The PCA algorithm was designed to reduce dimensions. The final processed data was input into the BLSTM algorithm for classification. The most suitable hyperparameters were selected to optimize the results. If the expected accuracy is achieved, then the test signals would input into the algorithm for classification; otherwise, hyperparameters would be adjusted and weights would be updated to improve the training accuracy and the loop continues until the best results are achieved.

4 Experimental validation and discussion

4.1 Experimental setup and procedure

A series of hole drilling experiments were performed by using new twist drills and different types of worn twist drills for the calibration and validity of the model. Inconel 718 was used as the workpiece material for performing

Fig. 2 The flow chart for the designed algorithm of model



experimentation. It is a nickel-based superalloy with strong thermal stability, high temperature, corrosion resistance in oxidization, and high mechanical strength; its chemical composition is shown in Table 1. Inconel 718 is a material that is considered difficult to cut because of severe plastic deformation in the shear zone. During drilling, this material possesses an extreme level of friction at the workpiece-tool interface which is the primary cause of tool wear. The purpose of performing drilling experiments is to acquire the drilling force signal data. The experimental setup is shown in Fig. 3. The drilling experiments were performed by using a 3-axis CNC machine. The carbide twist drills GD03-1000 and ZCC CT with 10 mm diameter, 47 mm flute length, and 89 mm length were used in all conducted drilling experiments, and these twist drills are also commercially available. Inconel 718 plates were used to drill the holes, and these plates were 16 mm thick.

The coolant used is mineral oil in water. The required drilling thrust and torque values were obtained using the Kistler 9123C rotary dynamometer. Different parameters of drill wear were successfully measured with the help of a 3D optical surface measurement device (IFM-G4) shown in Fig. 3. The control variables include spindle speed, feed rate, and depth of cut. These levels were of different control variables, and the response measures are shown in Table 2. Under the careful examination of the Alicona optical 3D surface measurement device (IFM-G4), the different types of worn tools were chosen to perform drilling operations for the collection of force signals for each type of worn tool. During the experiment, five different types of tools were selected to perform the drilling operation. These five tools include normal tool, flank wear tool, chisel wear tool, crater wear tool, and outer corner wear tool. Each tool is differentiated by the type of tool wear on the twist drill which was identified by IFM-G4. Cutting conditions are the same for all tools because the force signals of all tools are needed to input the algorithms for classification. The generated force signals were in the time series sequence.

The IFM-G4 is used to measure the flank wear, chisel wear, crater wear, and outer corner wear of the drill. The

application package of IFM-G4 includes IF-Measure Suite. Every tool has given different output results in the terms of force signals because every tool has a different type of tool wear.

4.2 Drilling experiment using the normal and worn tools

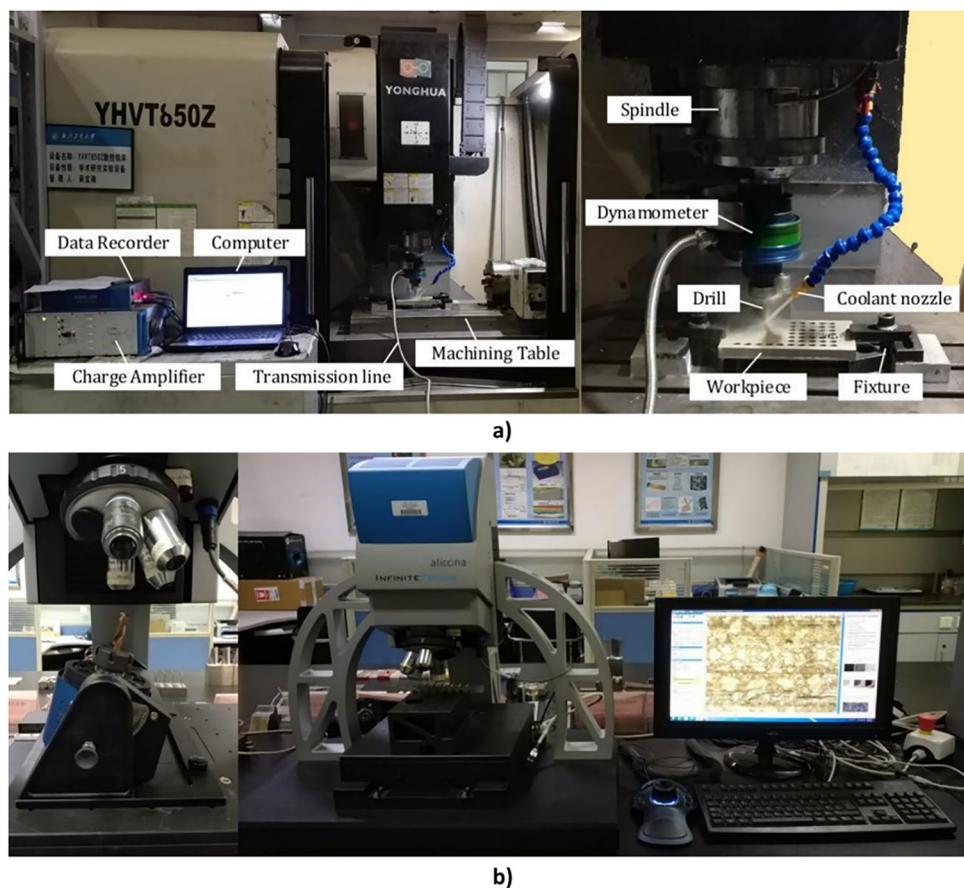
The drilling experiments were performed successfully on Inconel 718 using a dynamometer and CNC machine tool. The force signal dataset was collected for the normal tool and types of worn tools. A total of five datasets based on raw force signals had obtained after the successful completion of drilling experiments. The drilling experiments were conducted using the normal tool having fine edges, and the force signals of the normal tool were plotted as shown in Fig. 4a). The graph shows all the upper peaks around the region between the 275 and 300 N of force values. It is depicted that the cutting tool has sharp and fine edges.

The drilling operation is divided into three processes. The first process is drill-in. In this process, the initial force was applied on the surface of the workpiece and the applied force was increased slowly. The second process is the steady drilling process; the drilling tool in this process continues to remove the material from the workpiece. Drilling-out is the third process during which the force will decrease slowly and continuously. As far as Fig. 4a is concerned, during the drilling-in process, the force increased from 0 to around 275 N. The second part of the experimentation consists of drilling operations using the worn tools. The four worn twist drill tools were chosen, including chisel wear, crater wear, outer corner wear, and flank wear. The damaged chisel's edge and cutting lips of a twist drill bit have a great influence on the force signals generated by the rotary dynamometer. Figure 4 b, c, d, and e have shown the force signal graphs for the chisel wear drill, crater wear drill, flank wear drill, and outer corner wear drill, respectively. The drilling operations using different types of worn tools were also divided into three parts including drill-in, steady drilling, and drill-out as

Table 1 Technical data involving chemical composition and properties of IN718

IN718 composition		IN718 properties (physical, mechanical, thermal)		
Element	%	Name	Unit	Value
Ni	53.39	Modulus of elasticity	MPa	7.25×10^2
Cr	21	Ultimate tensile strength (UTS)	MPa	1.04×10^3
Nb	5.3	Elongation at break	%	30
Mo	3.02	Specific gravity	—	8.19
Ti	1.019	Thermal conductivity	W/m·K	6.52
Al	0.55	Melting point	°C	1.375×10^3
Co	0.21	Electrical resistivity (at 20 °C aged)	Ω·m	1.21×10^{-6}
Fe and others	Balance	—	—	—

Fig. 3 A detailed demonstration of experimental setup. **a** Drill-ing operation. **b** Alicona optical 3D surface measurement device (IFM-G4)



shown in the above graphs. The edge of the cutting tool gets deteriorated rapidly because of the high temperature. If the tool is more severely worn, then the irregular contact area between the tool and workpiece also increases. The main conditions that account for the rapid deterioration of the tool are the greater value of hardness of the workpiece, high cutting forces, and high temperature during machining. Higher cutting forces have rapidly deteriorated the tool. Chinchanikar and Choudhury [31] have explained the relationship between cutting forces and tool

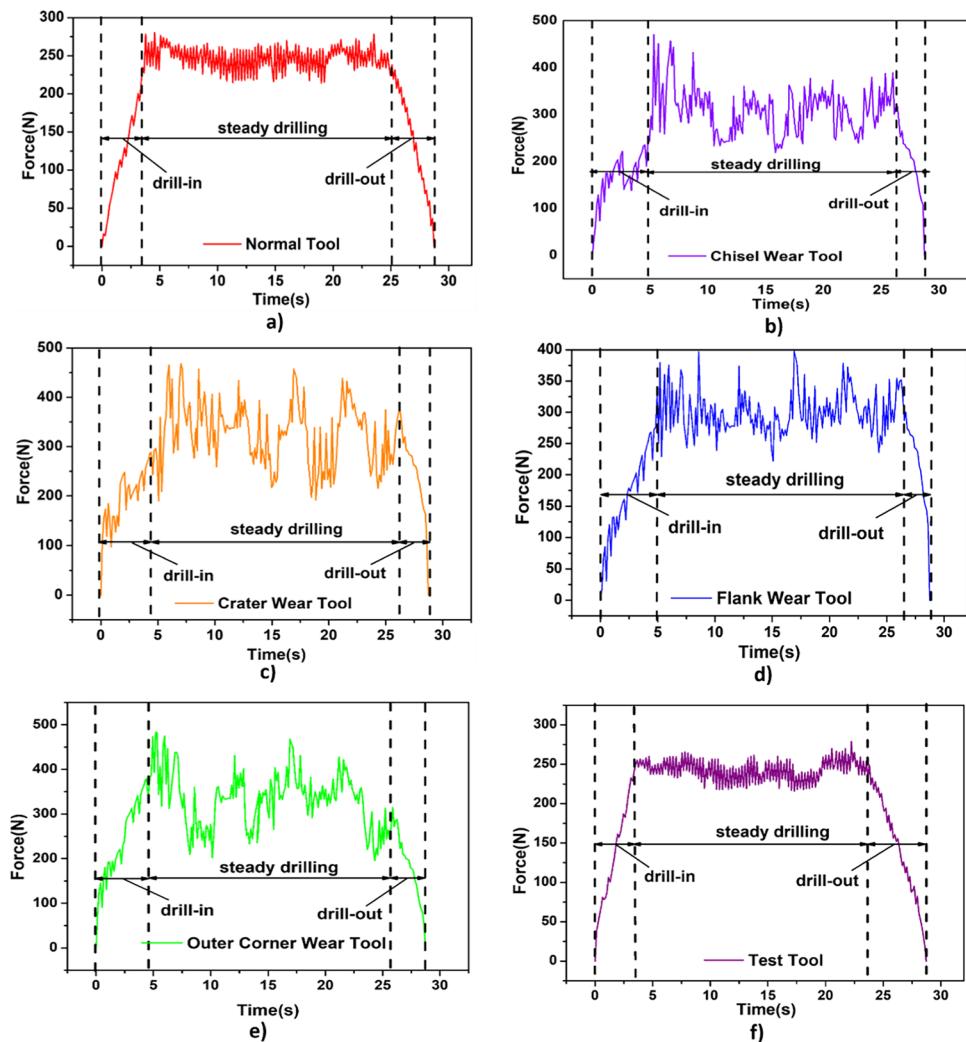
wear area by using mathematical equations and performing experiments.

The most important and well noticeable observation is that the value of force in the graph of different types of worn tools is increased drastically. The maximum value of force increased approximately from 400 to 500 N. If the drill-in process of the graphs of various types of worn tools is compared with the graphs of normal, then it is concluded that the drill-in operation shown in the graphs of types of worn tools has shown abnormal behavior. The

Table 2 Cutting parameters

Variables	Spindle speed	Feed rate	Output response	Constant factors
Symbol	n	f_r	Tool wear (crater wear, flank wear, chisel wear, outer corner wear)	Workpiece thickness 16 mm
Units	“rpm”	“mm/rev”	Classification using BLSTM algorithm	Type of coolant Flood coolant
Level 1	450	0.09		Type of tool Twist drill
Level 2	600	0.12		Type of drill hole Through
Level 3	750	0.15		Working angle 90°
Level 4	900	0.18		Working temperature Room temperature
Level 5	1050	0.21		Type of jig used Universal jig used for the drilling tests

Fig. 4 Graphical representation of force signals generated by drilling tools. **a** Normal tool. **b** Chisel wear tool. **c** Crater wear tool. **d** Flank wear tool. **e** Outer corner wear tool. **f** Test tool



force value was increased from 0 N after touching the twist drill bit on the surface of plates of nickel-based superalloys, and then the graph has shown abnormal large waveforms with sharp spikes which depict the extreme rise and fall in the value of force during the drill-in process. This type of behavior is well expected and obvious because the tool used for the operation was worn out.

Crater wear tool's graph has shown irregular large waves having abnormal behavior after 100 N, whereas small irregular abnormal waves are displayed in the outer corner wear tool's graph throughout the drill-in process. The drill-in process of the chisel wear graph is different from all other five graphs as after 100 N, the force value increases to around 225 N and then suddenly declines to 150 N which depicts that fine edges and damaged tool edges have touched the workpiece followed by each other. The second operation after the drill-in process was the steady drilling process. During this process, the maximum values of peaks in all worn tool types of graphs have touched approximately 400

to 500 N. When the tool wear occurs, then the applied force increases drastically.

The force profiles in the steady drilling process are irregular and are not concentrated at the specific region because the cutting edges of the twist drill bit used during the drilling operation are worn out. Therefore, the maximum value and the average value of force during the steady drilling process with worn tools were much greater than those during the steady drilling with the normal tool. As a whole, forces are not uniform and they are highly irregular. Moreover, they are not concentrated in the fixed regions. Some forces in the flank wear graph are concentrated around the region of 350 to 400 N, and some forces are concentrated on 300 to 350 N because when the worn edges of the worn tools touch the workpiece, they cut in a highly irregular manner. Forces of crater wear tool during the steady drilling are large as compared to the other types of worn tool graphs. There is one similarity in all the types of worn tool graphs that is the profiles are not uniform and the spikes are highly irregular. The

drill-out operation would start from the force value between 250 and 350 N approximately. There is a simple and almost tilted continuous line shown in the drill-out operation, which is common in all the graphs of worn tool types.

4.3 Feature extraction of time series force signals by SSA

The singular spectrum analysis is a contemporary and alternate time series technique for the extraction of the features. SSA method contains two corresponding stages. The first is decomposition and the second is a reconstruction; both contain two distinct steps. In the first stage, the time series signal is decomposed, whereas the original series is reconstructed in the second stage [32]. The expression “singular” signifies the decomposition of spectral or the matrix. Eigen decomposition and matrix eigenvalues are utilized for the decomposition of each matrix column. The expression “spectrum” is well-defined as the addition of a set of eigenvalues after the process of spectral decomposition. SSA is a time series technique; its key objective is the decomposition of the original time series into various independent linear components that are entirely different.

SSA algorithm comprises various steps containing decomposition and reconstruction; it is a novel non-parametric method for the analysis of time series based on the principles and rules of multivariate statistics [33]. SSA decomposes the particular time series into a set of additive independent time series, and it is an effective technique used to remove noise from the signal. The technique projects the original time series onto a vector foundation acquired from the time series itself by following the method of PCA. The set of series obtained from the decomposition technique can be deduced as a gradually changing trend, demonstrating the mean of the signal at a particular instant. In this study, SSA is used for the decomposition of force signals in a drilling process. The entire procedure of application of SSA on time series raw force signals is divided into five steps given below.

- Step # 1

Calculation of the covariance matrix by using the trajectory approach. The first step of SSA is mapping the time series sequence of force signals denoted by S to a multidimensional sequence of lagged vectors. Let W be the length of the window.

$$2 \leq W \leq \frac{N}{2} \quad (14)$$

The window is formed by using the subseries given by:

$$S = \{S_i, S_{i+1}, \dots, \dots, S_{i+L-1}\} \text{ for } i = 0, \dots, N-L \quad (15)$$

If the window W is sliding along the given time series S , then the column vector denoted by Y_i would be formed for each subseries, and then the resultant equations become:

$$Y_0 = (S_0, S_1, S_2, \dots, S_{L-1})^T \quad (16)$$

$$Y_1 = (S_1, S_2, S_3, \dots, S_L)^T \quad (17)$$

$$Y_2 = (S_2, S_3, S_4, \dots, S_{L+1})^T \quad (18)$$

$$Y_3 = (S_3, S_4, S_5, \dots, S_{L+2})^T \quad (19)$$

$$Y_{N-L} = (S_{N-L}, S_{N-L+1}, S_{N-L+2}, \dots, S_{N-L})^T \quad (20)$$

These column vectors form the trajectory matrix denoted by Y is:

$$Y = \begin{bmatrix} S_0 & S_1 & S_2 & S_3 & \dots & S_{N-L} \\ S_1 & S_2 & S_3 & S_4 & \dots & S_{N-L+1} \\ S_2 & S_3 & S_4 & S_5 & \dots & S_{N-L+2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ S_{L-1} & S_L & S_{L+1} & S_{L+2} & \dots & S_{N-1} \end{bmatrix} \quad (21)$$

The trajectory matrix Y given above is also called the Hankel matrix. The length of window L constrains the results of SSA decomposition. Its suitable selection depends on the structure of the given time series to decompose. Results for step 1 are shown in Fig. 5. The color patterns illustrate the correlation values of the trajectory matrix Y . The darker color represents the negative correlation; however, the lighter color demonstrates the positive correlation.

- Step # 2

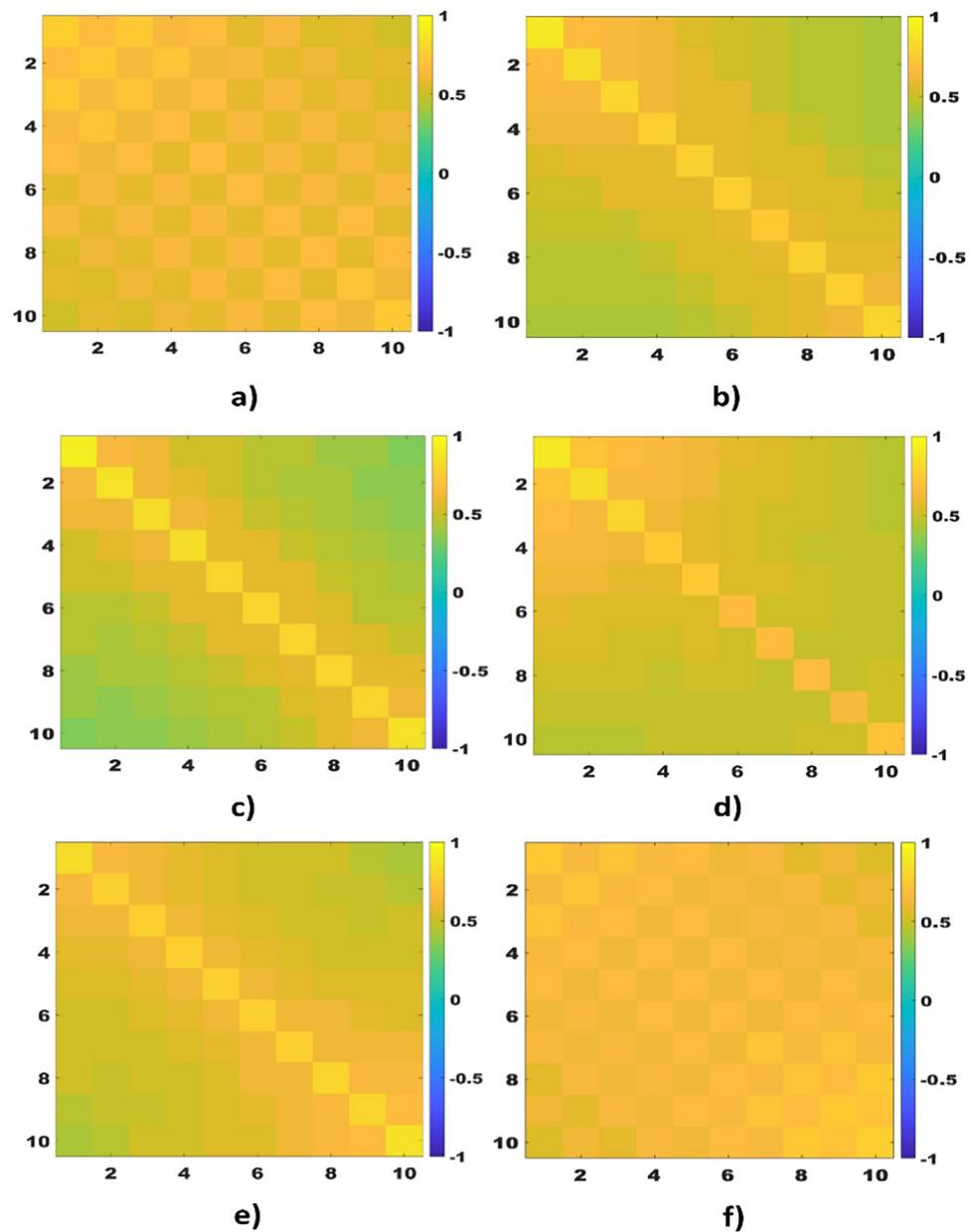
After the calculation of the covariance matrix (trajectory matrix), the technique of singular value decomposition would be applied to the trajectory matrix denoted by Y .

$$Z = Y \times Y^T \quad (22)$$

Firstly, calculate the eigenvalues and eigenvectors of the matrix Z of dimension $L \times L$. Let $\lambda_1, \lambda_2, \dots, \lambda_d$ be the eigenvalues of matrix Z which are non-zero, whereas U_1, U_2, \dots, U_d are the values of corresponding eigenvectors. The new vector formed by using the mathematical expression is given below:

$$V_i = Z^T \times \frac{U_i}{\sqrt{\lambda_i}} \quad (23)$$

Fig. 5 Covariance matrix results for **a** normal tool, **b** chisel wear tool, **c** crater wear tool, **d** flank wear tool, **e** outer corner wear tool, and **f** test tool



The trajectory matrix decomposed into a sum of matrices by using the singular value decomposition technique:

$$Z = P_1 + P_2 + P_3 + \dots + P_d \quad (24)$$

$$P_i = \sqrt{\lambda_i} \times U_i \times V_i^T \quad (25)$$

These are the elementary matrices which are unit rank and also mutually orthogonal. The method of plotting the eigenvalues in the decreasing order is the singular spectrum. Figure 6. demonstrates the calculation of eigenvalues (λ) and eigenvectors (EVI).

- Step # 3

In this step, the principal components (PC) would be calculated by using the scalar product of two matrices. The calculation of PC is shown in Fig. 7:

$$PC = Y \times RHO \quad (26)$$

where Y is the trajectory matrix and RHO is the matrix of eigenvectors.

- Step # 4

Reconstructed components (RC) are calculated by projecting principal components back onto eigenvectors.

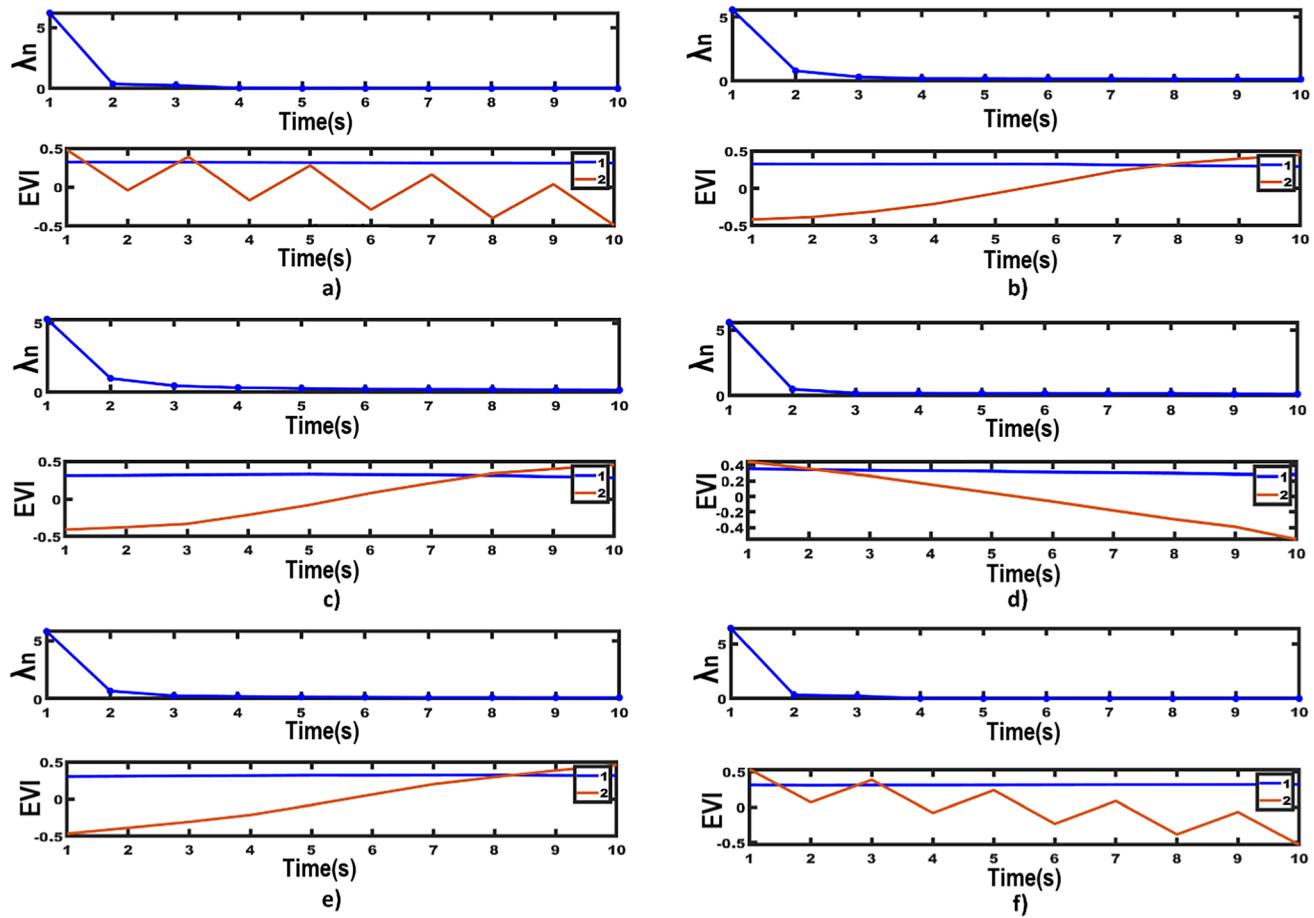


Fig. 6 Graphical trends to indicate the calculation of eigenvalues and eigenvectors for **a** normal tool, **b** chisel wear tool, **c** crater wear tool, **d** flank wear tool, **e** outer corner wear tool, and **f** test tool

tors; thus, the resulting time series obtained is called reconstructed components. The basic methodology is to invert the projection $PC = Y \times RHO$ and averaging the anti-diagonals for the original time series. PC is obtained by the mathematical expression given below.

$$RC = Y \times RHO \times RHO' \quad (27)$$

$$RC = PC \times RHO' \quad (28)$$

The results of step 4 are shown in Fig. 8.

- Step # 5

In the final step, both reconstructed component (RC) and original time series of raw force signals were plotted on the same graph for the comparison of original time series raw force signals obtained by using a dynamometer and the reconstructed component obtained by SSA. As shown in Fig. 9, the graphs of RC possess patterns that

are similar to original force signals, but the main quality is that RC do not have noise and all the fifteen features were derived based on reconstructed components by applying different mathematical formulas.

Fifteen independent features were extracted. The extracted features were used to train the deep learning algorithms. These features are kurtosis index, root mean square, mean, absolute mean, margin index, minimum data value in signal (raw data), maximum data value in signal (raw data), waveform index, RMS amplitude, variance, skewness, peak index, kurtosis pulse index, peak-to-peak, and standard deviation.

The major concern with the raw data obtained from the dynamometer is of having noise in the signal. Noise destroys the quality of the signals and makes the signal unstable. It also has an adverse impact on the algorithm's performance as well as accuracy. The SSA has successfully removed the noise from the raw signal data of the force sensor. The raw data was in the form of a 1-D signal which was decomposed

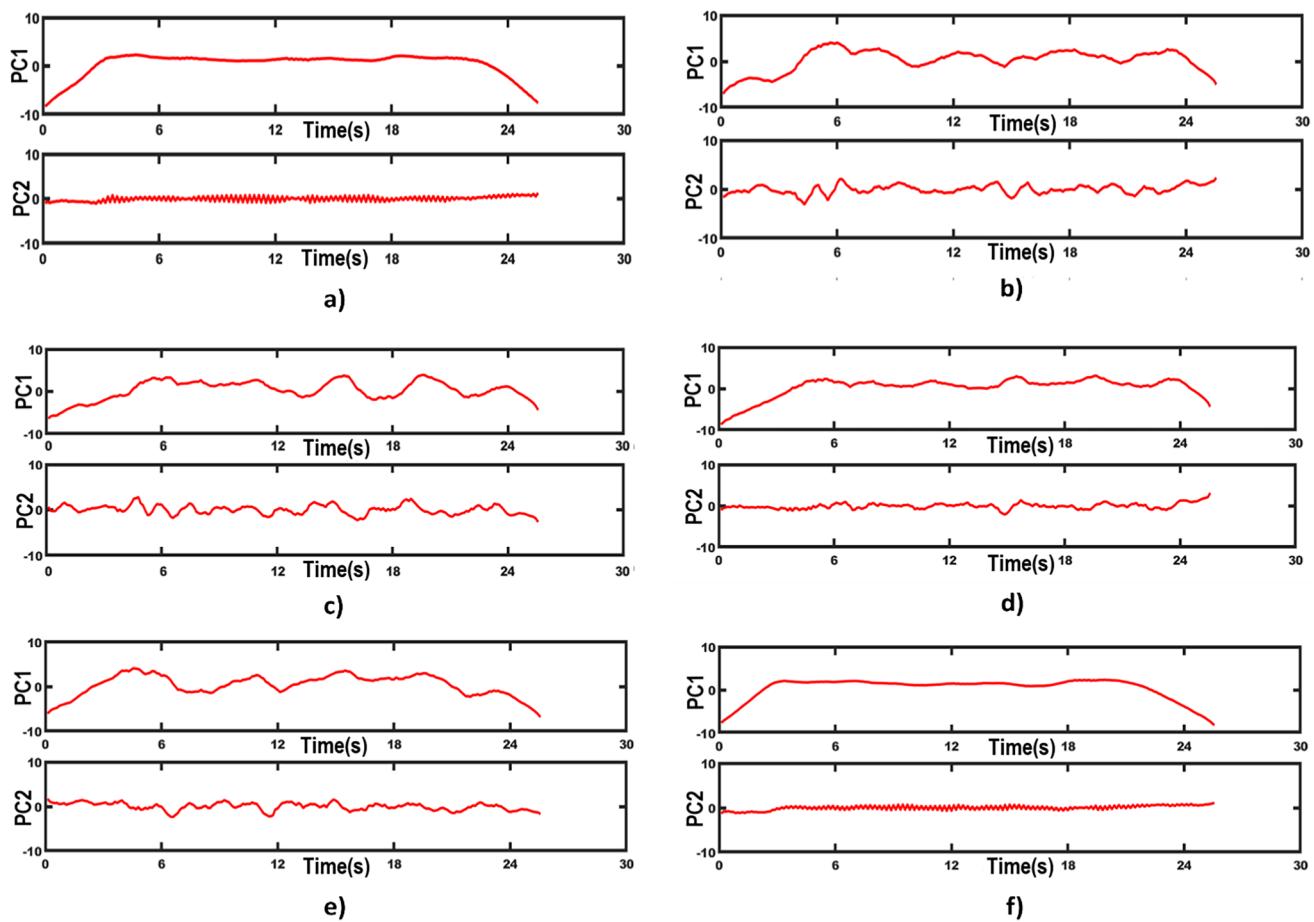


Fig. 7 Graphs to explore the calculation of PC for **a** normal tool, **b** chisel wear tool, **c** crater wear tool, **d** flank wear tool, **e** outer corner wear tool, and **f** test tool

into components comprising different trends, oscillations, and noise. The discriminating capability of extracted features is enhanced after the removal of the noise component from the raw force signal. The above-mentioned features were extracted separately from all the raw force signals of the normal tool, crater wear tool, chisel wear tool, outer corner wear tool, and flank wear tool. The fifteen distinct features were used as an input to the BLSTM algorithm. The SSA feature extraction technique has increased the classification accuracy of the model by enhancing the performance of the BLSTM algorithm due to its noise removal capability.

4.4 Application of PCA and one-hot encoding

PCA is a methodology for reducing the dimensionality of such a large dataset, maximizing interpretability, and mitigating the information loss simultaneously. PCA is applied to the SSA obtained features. Figure 10 has shown the validation of the application of PCA on the features of the model. PCA transforms a large set of variables into a smaller one that still retains important information. A total of 15

features were extracted by SSA, and the correlation of 15 extracted features is shown in the heat map given in Fig. 10a.

The most commonly used statistical measure to test relationships between variables is a correlation. The 15 dimensions (features) are converted into 3 dimensions which are 3 principal components, and the values of 3 principal components are plotted in Fig. 10c. PCA has generated 3 uncorrelated components which have maximum variance. The main components are built in such a way that the first main component has the greatest possible variance. The second component is computed in the same manner on the condition that it is perpendicular to the first principal component which is accounted for the maximum variance. The heat map given in Fig. 10b shows the relationship between the correlation of 15 features and 3 principal components. On the x-axis of the heat map, the 15 features are given, whereas on the y-axis, three principal components, i.e., 0, 1, and 2, are given. There is a total of 5 steps for the calculation of principal components. The first step is standardization followed by the computation of the covariance matrix, and then eigenvalues and eigenvectors of the covariance

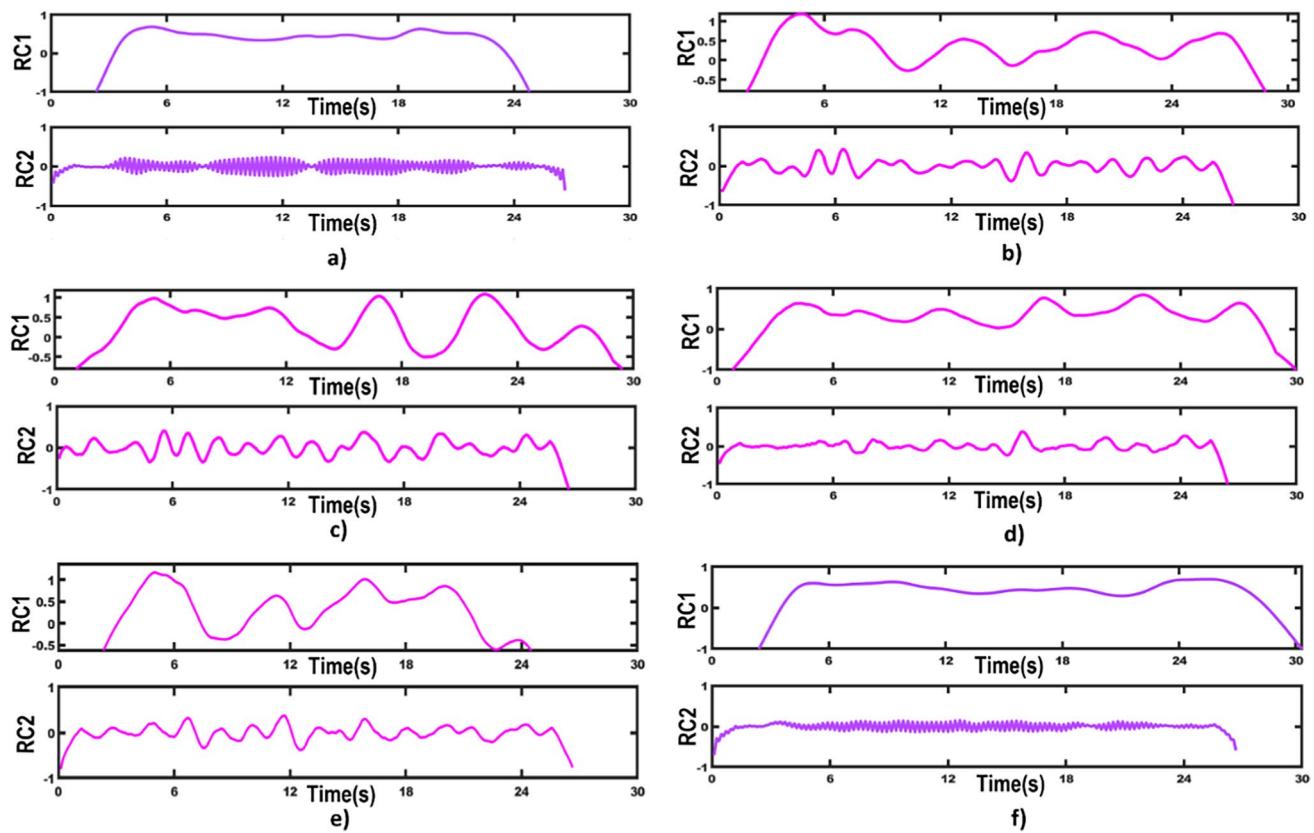


Fig. 8 Graphs to illustrate the calculation of reconstructed components (RC) for **a** normal tool, **b** chisel wear tool, **c** crater wear tool, **d** flank wear tool, **e** outer corner wear tool, and **f** test tool

matrix are calculated to identify the principal components. In the next step, the new feature vectors are created, and the final step is the recasting of data points along the principal axis. There are three most important advantages of PCA. It removes the correlated features as such features adversely affect the performance of the model. PCA creates the principal components which have no correlation among them and are independent of each other. PCA will make the model efficient by getting rid of correlated features which do not lead to any decision-making, and PCA is the most common way to speed up the deep learning algorithm. With fewer dimensions, the training time of the algorithms decreases considerably. PCA reduces the risk of overfitting which happens predominantly when a lot of variables are present in the dataset. Therefore, by decreasing the number of variables, PCA helps to resolve the overfitting problem. PCA also plays an important role in noise reduction. The target variables of the model were in textual form. Deep learning models cannot interpret and process the texts, so there was the need to convert the textual form of target variables into numerical form. One-hot encoding is a technique that transforms categorical variables into binary numbers. The technique of one-hot encoding is applied in Table 3. All the 5 target variables will be stored in the y_train of the model,

and one-hot encoding will make the model a multi-labeled classifier (Table 4).

4.5 Prevention of overfitting of the model

Overfitting of the model is the major problem associated with artificial intelligence algorithms. During overfitting, the model picks up the noise and random fluctuations of training data and learns them as concepts. Overfitting will increase the generalizing error, and it will reduce the model's ability to perform on the new dataset. Overfitting adversely affects the model performance and accuracy. Selecting a suitable number of training epochs is the most important part of neural network training. If the number of epochs is high, then it will lead to overfitting. However, the low numbers of epochs would cause the model to underfit. Therefore, the early stopping technique, which allows to set an arbitrary greater number of epochs and then stop training when the model's accuracy on the validation dataset stops improving and initiates to learn the statistical noise of training data, can be applied here. The training of the neural network should be stopped at the specific point when the model's performance starts degrading on the validation data. The model's performance on the present training epochs will be compared to

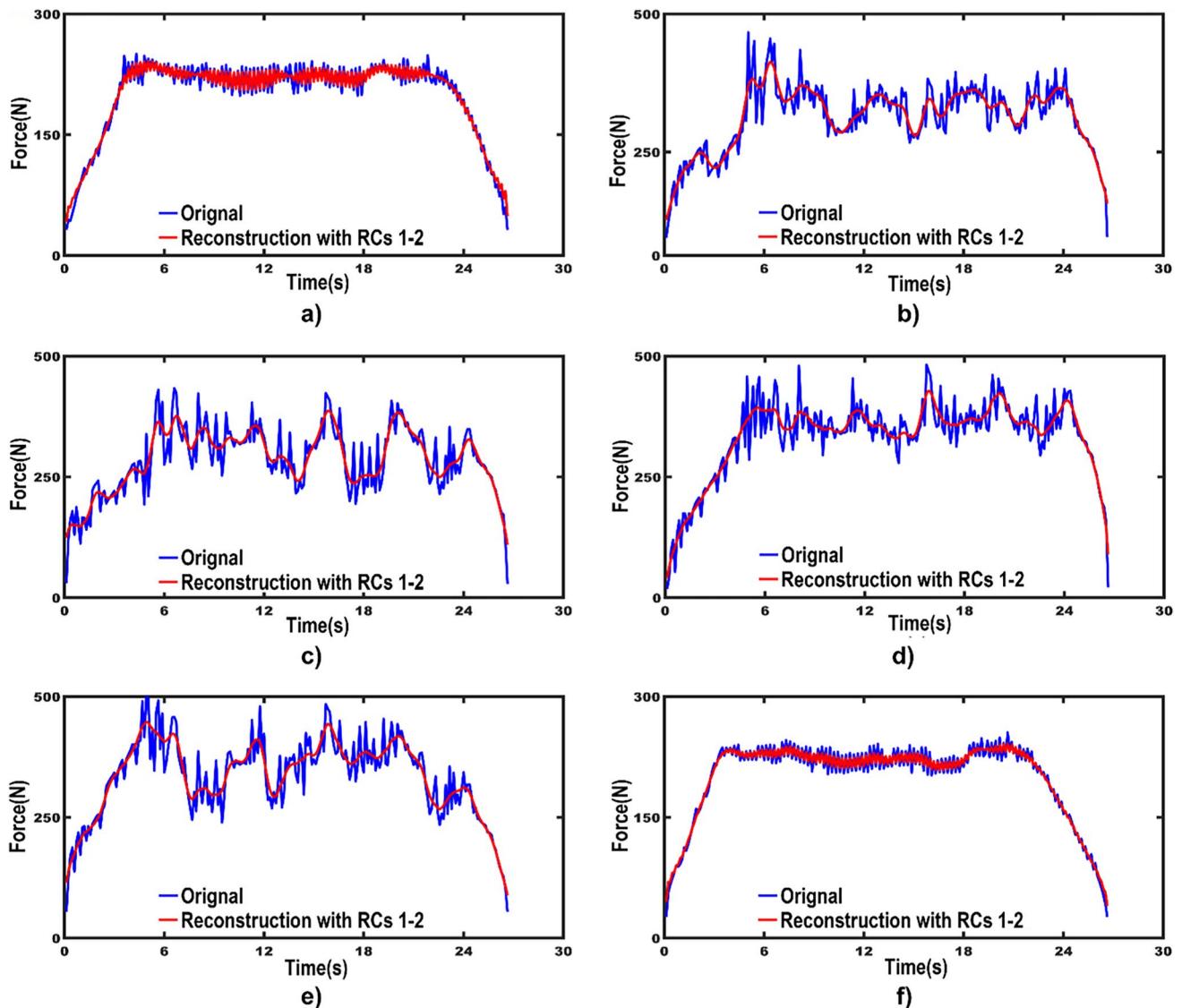


Fig. 9 Generated graphical trends to illustrate the comparison of RC and SSA for **a** normal tool, **b** chisel wear tool, **c** crater wear tool, **d** flank wear tool, **e** outer corner wear tool, and **f** test tool

the performance on the previous training epochs. Dropout is also a very useful and effective technique to avoid the overfitting of the model. This technique has the ability to modify the BLSTM neural network. During each iteration of processing, it removes neurons from the network randomly. During training, certain layers' outputs are dropped out randomly.

4.6 Justification of solving overfitting problem

In the above sections, SSA has removed the noise from the raw force signals and extracted fifteen features. When these features were input to the BLSTM for training, sometimes results were not satisfactory because the over-fitting problem may occur.

The graph given in Fig. 11a has shown the overfitting problem during the training of the BLSTM algorithm using fifteen features extracted by SSA. The validation loss is higher than the training loss which is a clear sign of overfitting of the algorithm. Overfitting occurred in the above case when the appropriate and suitable fit for the model has been achieved but not generalizing well using the unseen new test data so the model has learned the patterns which are unique to training data. PCA has reduced the dimensions of features from fifteen to three. Early stopping and dropout algorithms are designed by specifying the most suitable parameters. The monitor was set to “val_loss,” the mode was set to “min,” and patience was set equal to the value of 25. Verbose was set to 1 to show the progress during the training process. The batch size was equal to 10. The “adam”

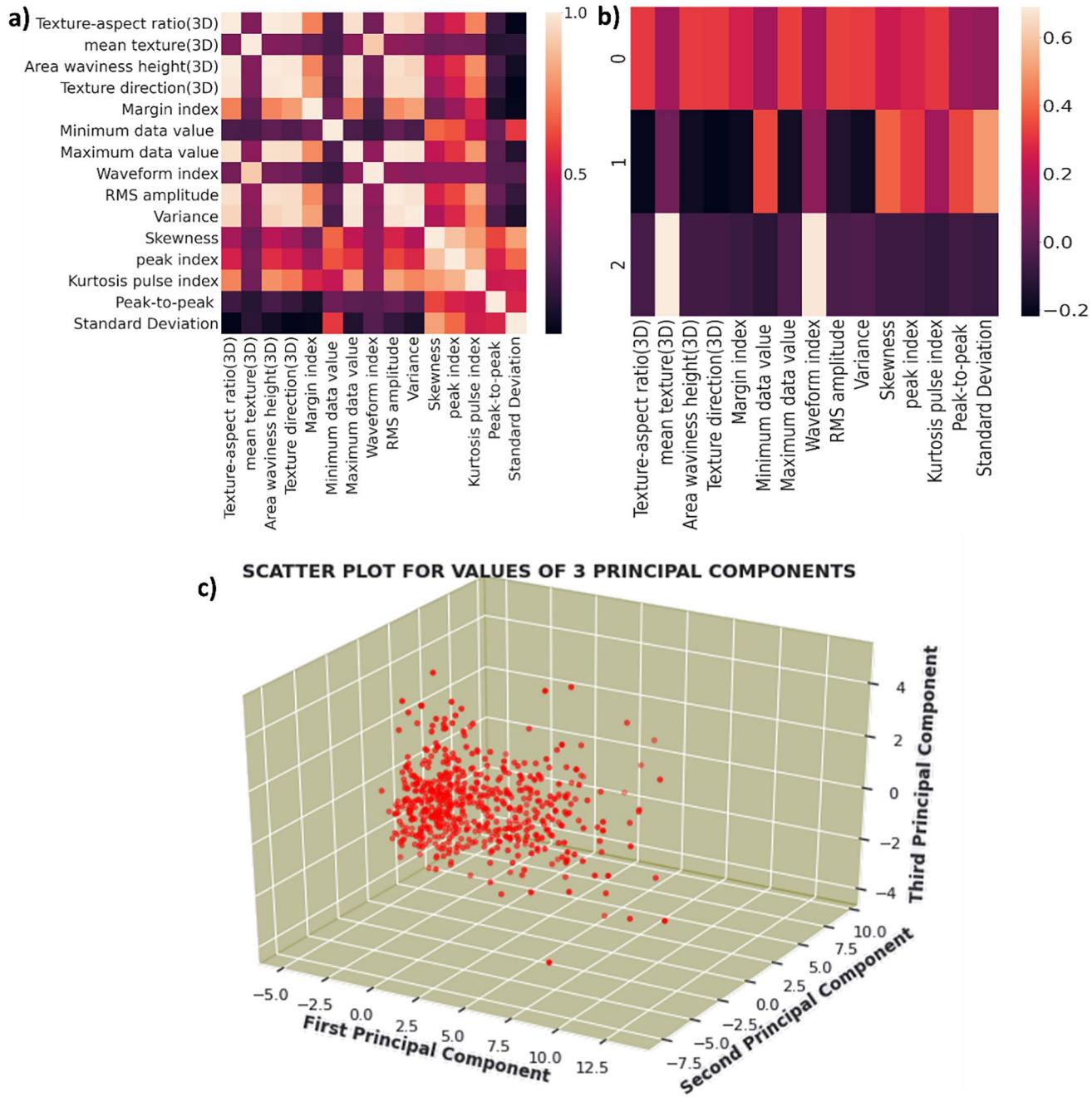


Fig. 10 **a** Heat map shows the correlation of features. **b** Heat map shows a relationship between the correlation of features and principal components. **c** 3D scatter plot for derived 3 principal components

Table 3 Target variables of model

Index	Target
1	Crater wear detected
2	Chisel wear detected
3	Flank wear detected
4	Outer corner wear detected
5	Normal tool

optimizer was used. The value of dropout was 0.5. The validation split was equaled to 0.2. During the training, the numbers of epochs were set to 700, but after the application of PCA, by applying dropout and early stopping, the training process was automatically stopped at the epoch value chosen by algorithms, and this value is the performance peak point of the model, after which the performance of the model will start to decline. The automatically chosen epoch value by the algorithm was much smaller than the manually set value

Table 4 One-hot encoding applied to the target variables of model

Index	Target _ Chisel wear detected	Target _ Crater wear detected	Target _ Flank wear detected	Target _ Outer corner wear detected	Target _ Normal tool
1	0	1	0	0	0
2	1	0	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	1

which demonstrates that the training time has also decreased tremendously and the model has become efficient. The validation loss and actual training loss are nearly equal after the application of designed algorithms, and both graphs have touched each other after decreasing the loss value at the end of the training process. Figure 11c has shown that model has done the classification of force signals of the test tool by achieving an accuracy of 97.94%. Therefore, the new methodology has been designed in this research study, which possesses the experimental data to determine the different types of worn tools and also solved the overfitting problems.

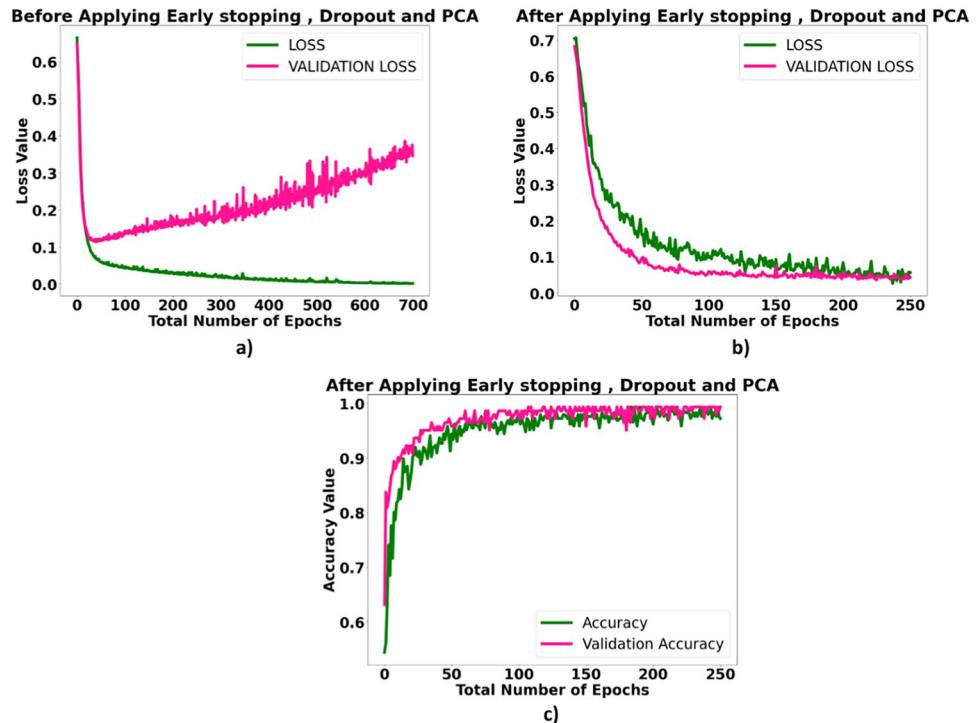
4.7 Comparison of model's results with existing algorithms based on common real-time data

The drilling operations were conducted by using the flank wear tool and chisel wear tool after the deployment of the

designed model to get the new real-time data for testing purposes. In order to construct good comparison results, all major artificial intelligence algorithms of existing literature were fully trained separately on the original data (features of all worn tools and normal tool) of the current research study, and the data was completely processed into the final state after the extraction of all features from the force signals in order to input into other machine learning and deep learning algorithms. The developed model of the current research study and models of previous research studies were then tested on the common data (testing data signals of flank wear tool and chisel wear tool) and have shown different results based on final accuracy. The working of some models of the existing literature was based on the images; therefore, the same data in the form of force signals was converted into images by using a spectrogram algorithm. As shown in Fig. 12, in the case of convolutional neural network (CNN), the final processed data was converted into images to input into this algorithm because the basic working of CNN depends on the feature extraction from images. The force signal data was converted into another state in the form of images, and the number of spectrogram images per class was increased by using the technique of image augmentation, so the CNN algorithm performed worst as compared to all other machine learning algorithms.

The performance of the random forest algorithm was unsatisfactory on current data. The main reason for the limitation is that the large number of trees constructed for the multiclass data was unstable. Since the transfer

Fig. 11 Final results for training and testing of model. **a** Training graph before dropout, PCA, and early stopping application. **b** Training graph after dropout, PCA, and early stopping application. **c** Graphical demonstration of test and validation accuracy



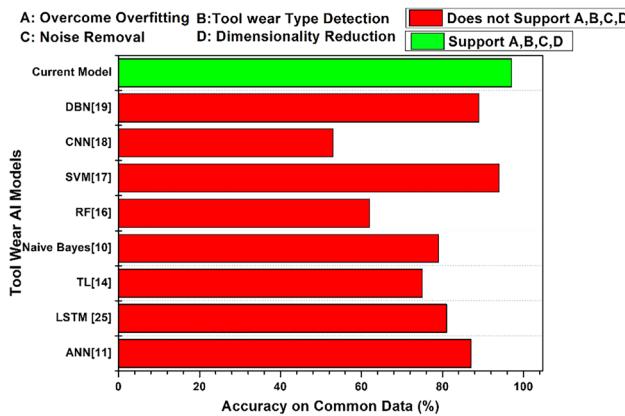


Fig. 12 Comparison with existing models based on testing signals of flank wear tool

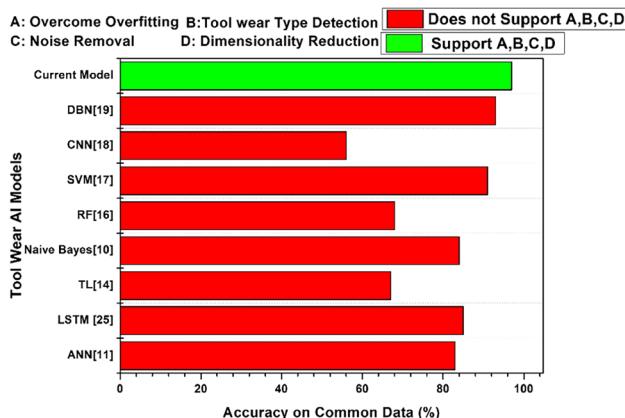


Fig. 13 Comparison with existing models based on testing signals of chisel wear tool

learning algorithm has the limitation of negative transfer, which adversely affects the final accuracy, two tests were conducted on all models by using the data features of the flank wear tool and chisel wear tool as input testing signals. Figure 13 has shown the comparison of the performance of all models on the common flank tool wear testing data. The accuracy of the current model was better than existing models of previous research studies. The comparison was built based on the five important parameters. The first parameter is accuracy, and other parameters were denoted by alphabets A, B, C, and D. The performance and efficiency of tool wear models most likely depend upon these parameters, but these important parameters are neglected by existing research studies by compromising on the accuracy and efficiency of their model. Overfitting is the major problem for artificial intelligence algorithms in tool wear AI models. All algorithms most often have overfitting during tool wear prediction as some AI algorithms show a high level of overfitting, whereas others experience a low level in which

previous research studies have ignored this important aspect and compromised the final accuracy. Noise in the signals can have an adverse impact on the performance of the tool wear model. The algorithm of singular spectrum analysis is employed to reduce noise from the force signals in order to achieve high accuracy. The bidirectional algorithm will process the input in bidirectional ways. The most important processing is from past information of worn tool and doing the prediction of the future outcomes related to that specific tool. Principal component analysis (PCA) is applied on tool wear models for dimensionality reduction from 15 to 3 for maximizing interpretability and mitigating the information loss simultaneously and has made the model more efficient. The current model has worked in the detection of the type of tool wear which has not been addressed previously by any research study.

In both test cases, the accuracy of the current model is greater than all other existing models. The final results of the machine learning model solely depend on how accurately it predicts the target variables. The five different types of tools have to be predicted by the model, so there would be five target variables. The main quality of the current model is that it is multi-target and all target variable is in binary form. One-hot encoding technique is applied to all models to get the comparison based on accuracy and binary output values. When the models were tested for the chisel wear tool, the predicted binary value was 10,000. When the features of the crater wear tool were input into the model, the predicted binary value was 01,000. When the model was tested for flank wear tool and outer corner wear tool, the predicted binary values were 00,100 and 00,010, respectively. However, if the input test signals contain the features of the normal tool, then the predicted binary value will be 00,001. The current model has predicted the target variables with an overall accuracy of 97% by using 250 epochs. The algorithm will pass through the data 250 times. The accuracy is the ratio of total positive predictions of all target variables to the total number of predictions of all target variables. The validation accuracy of the model was better than the testing accuracy which means the model has successfully overcome the problem of overfitting in the tool wear model.

5 Conclusion

This research study has successfully developed the methodology based on practical experiments to detect the type of tool wear in the drilling operations. In this paper, different types of tool wear were discussed. SSA was used for noise removal and to extract the features from the raw force signals. PCA was used to reduce dimensions and improve the efficiency and performance of the model. One-hot encoding

is applied to target variables of the model to convert the textual data into binary form. Early stopping and dropout algorithms have solved the overfitting problem. A deep learning BLSTM algorithm was used for the multi-target classification. This experimentally illustrated research study has developed an efficient and cost-effective system that has shown a test accuracy of 97.94%.

The future work corresponding to this research study includes mechanistic modeling of five tools, i.e., flank wear tool, outer corner wear tool, chisel wear tool, normal tool, and crater wear tool. And the estimation of tool wear values for different tool wear type will be conducted to help to identify the tool wear status.

Acknowledgements The authors also would like to thank Dr. Ce Han for providing data and help in the experiment.

Author contribution J. Mahmood and M. Luo have generated the new research idea, designed methodology, done investigations, and conducted experiments; J. Mahmood and M. Rehman have designed algorithms. All authors have actively participated in the revision and approved the manuscript.

Funding This study was co-supported by the National Natural Science Foundation of China (Grant No. 91860137) and the 111 project (Grant No. B13044).

Data availability Complete data values for all tools are plotted in cutting force graphs.

Code availability The algorithm (code) designed for the model is clearly explained in the flow chart.

Declarations

Conflict of interest The authors declare no competing interests.

References

- Kusiak A (2017) Smart manufacturing must embrace big data. *Nature* 544:23–25. <https://doi.org/10.1038/544023a>
- Zhong RY et al (2017) Intelligent manufacturing in the context of industry 4.0: a review. *Engineering* 3:616–630. <https://doi.org/10.1016/j.eng.2017.05.015>
- Kusiak A (2017) Smart manufacturing must embrace big data. *Nature News* 544:23. <https://doi.org/10.1038/544023a>
- Zhang Z et al (2020) Multi-objective parameter optimization to support energy-efficient peck deep-hole drilling processes with twist drills. *Int J Adv Manuf Technol* 106:4913–4932. <https://doi.org/10.1007/s00170-020-04967-x>
- Groover MP (2020) Fundamentals of modern manufacturing: materials, processes, and systems. Wiley
- Çelik YH, Yıldız H, Özük C (2016) Effect of cutting parameters on workpiece and tool properties during drilling of Ti-6Al-4V. *Mater Test* 58:519–525. <https://doi.org/10.3139/120.110886>
- Iliescu D et al (2010) Modeling and tool wear in drilling of CFRP. *Int J Mach Tools Manuf* 50:204–213. <https://doi.org/10.1016/j.ijmachtools.2009.10.004>
- Dehghan S, Soury E (2021) A comparative study on machining and tool performance in friction drilling of difficult-to-machine materials AISI304, Ti-6Al-4V, Inconel718. *J Manuf Process* 61:128–152. <https://doi.org/10.1016/j.jmapro.2020.10.078>
- Wang J et al (2008) Wear mechanism map of uncoated HSS tools during drilling die-cast magnesium alloy. *Wear* 265:685–691. <https://doi.org/10.1016/j.wear.2007.12.009>
- Karandikar J et al (2015) Tool wear monitoring using naive Bayes classifiers. *Int J Adv Manuf Technol* 77:1613–1626. <https://doi.org/10.1007/s00170-014-6560-6>
- D'Addona DM, Ullah AS, Matarazzo D (2017) Tool-wear prediction and pattern-recognition using artificial neural network and DNA-based computing. *J Intell Manuf* 28:1285–1301. <https://doi.org/10.1007/s10845-015-1155-0>
- Palanisamy P, Rajendran I, Shanmugasundaram S (2008) Prediction of tool wear using regression and ANN models in end-milling operation. *Int J Adv Manuf Technol* 37:29–41. <https://doi.org/10.1007/s00170-007-0948-5>
- Chen JC, Chen JC (2005) An artificial-neural-networks-based in-process tool wear prediction system in milling operations. *Int J Adv Manuf Technol* 25:427–434. <https://doi.org/10.1007/s00170-003-1848-y>
- Li J et al (2021) Tool wear state prediction based on feature-based transfer learning. *Int J Adv Manuf Technol* 113:3283–3301. <https://doi.org/10.1007/s00170-021-06780-6>
- Mun J, Jeong J. Design and analysis of RUL prediction algorithm based on CABLSTM for CNC machine tools. 2020 2020 7th International Conference on Soft Computing & Machine Intelligence (ISCM). IEEE. <https://doi.org/10.1109/iscmi51676.2020.9311582>
- Wu D et al (2017) A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. *Journal of Manufacturing Science and Engineering* 139. <https://doi.org/10.1115/1.4036350>
- Qian Y, A tool wear predictive model based on SVM. et al (2010) 2010 Chinese control and decision conference. IEEE. <https://doi.org/10.1109/ccdc.2010.5498161>
- Martínez-Arellano G, Terrazas G, Ratchev S (2019) Tool wear classification using time series imaging and deep learning. *Int J Adv Manuf Technol* 104:3647–3662. <https://doi.org/10.1007/s00170-019-04090-6>
- Chen Y, Jin Y, Jiri G (2018) Predicting tool wear with multi-sensor data using deep belief networks. *Int J Adv Manuf Technol* 99:1917–1926. <https://doi.org/10.1007/s00170-018-2571-z>
- Glass K, Colbaugh R. Real-time tool wear estimation using cutting force measurements. 1996 Proceedings of IEEE International Conference on Robotics and Automation. IEEE. <https://doi.org/10.1109/robot.1996.509178>
- Mikołajczyk T et al (2018) Predicting tool life in turning operations using neural networks and image processing. *Mech Syst Signal Process* 104:503–513. <https://doi.org/10.1016/j.ymssp.2017.11.022>
- Bustillo A et al (2021) Machine-learning for automatic prediction of flatness deviation considering the wear of the face mill teeth. *J Intell Manuf* 32:895–912. <https://doi.org/10.1007/s10845-020-01645-3>
- Pimenov DY, Bustillo A, Mikolajczyk T (2018) Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth. *J Intell Manuf* 29:1045–1061. <https://doi.org/10.1007/s10845-017-1381-8>
- Kuntoğlu M et al (2021) A review of indirect tool condition monitoring systems and decision-making methods in turning: critical analysis and trends. *Sensors* 21:108. <https://doi.org/10.3390/s21010108>
- Zhang J, Zeng Y, Starly B (2021) Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis

- and prognosis. *SN Appl Sci* 3:1–13. <https://doi.org/10.1007/s42452-021-04427-5>
- 26. Zhang Z, et al. Research on tool wear prediction based on LSTM and ARIMA. 2018 Proceedings of the 2018 international conference on big data engineering and technology. <https://doi.org/10.1145/3297730.3297732>
 - 27. Hao G, Kunpeng Z. Pyramid LSTM auto-encoder for tool wear monitoring. 2020 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). IEEE. <https://doi.org/10.1109/case48305.2020.9217015>
 - 28. Ma J et al (2021) Tool wear mechanism and prediction in milling TC18 titanium alloy using deep learning. *Measurement* 173:108554. <https://doi.org/10.1016/j.measurement.2020.108554>
 - 29. Qiao H, Wang T, Wang P (2020) A tool wear monitoring and prediction system based on multiscale deep learning models and fog computing. *Int J Adv Manuf Technol* 108:2367–2384. <https://doi.org/10.1007/s00170-020-05548-8>
 - 30. Han C, Luo M, Zhang D (2020) Optimization of varying-parameter drilling for multi-hole parts using metaheuristic algorithm coupled with self-adaptive penalty method. *Appl Soft Comput* 95:106489. <https://doi.org/10.1016/j.asoc.2020.106489>
 - 31. Chinchanikar S, Choudhury S (2014) Characteristic of wear, force, and their inter-relationship: in-process monitoring of tool within different phases of the tool life. *Procedia Mater Sci* 5:1424–1433. <https://doi.org/10.1016/j.mspro.2014.07.461>
 - 32. Trovero MA, Leonard MJ. Time series feature extraction. 2018 SAS
 - 33. Poskitt DS (2020) On singular spectrum analysis and stepwise time series reconstruction. *J Time Ser Anal* 41:67–94. <https://doi.org/10.1111/jtsa.12479>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.