



# Detection of tool wear during machining by designing a novel 12-way 2-shot learning model by applying L2-regularization and image augmentation

Jawad Mahmood<sup>1</sup> · Muhammad Adil Raja<sup>1</sup> · Mudassar Rehman<sup>2</sup> · John Loane<sup>1</sup> · Sadaf Zahoor<sup>3</sup>

Received: 6 September 2022 / Accepted: 27 January 2023 / Published online: 8 March 2023  
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

## Abstract

Tool wear monitoring is regarded as an incredibly important aspect of improving the surface integrity of machined components in the manufacturing sector. This research study performed operations using twelve different types of drilling and milling tools. The worn tools ranging from grade-1 to grade-5 were categorized based on tool wear severity by measuring the flank wear land width of each tool. Advanced algorithms were designed based on short-time Fourier transform and continuous wavelet transform to convert time-series force signals' data into spectrogram and scalogram images, respectively, to increase the number of shots with which the model can work based on the methodology of 2-shot learning. An algorithm for image augmentation was developed to increase the number of images to improve the training and overall performance of the model. L2 regularization along with the optimal hyper-parameters were utilized to avoid overfitting and to improve the model's efficiency. Hyper-parameters were optimized by using the grid-search methodology. The milling and drilling data was collated into 12 classes which resulted in a 12-way learning model. Therefore, it will work for both milling and drilling operations. The model will determine whether the test tool is normal or worn. And if worn, it will determine the severity level of tool wear ranging from grade-1 to grade-5. The final results have shown that the model has worked efficiently during CNC machining and achieved 87.83% accuracy.

**Keywords** Tool wear grade · Drilling · Milling · Accuracy · 12-way 2-shot learning

## Nomenclature

ML	Machine learning	MCLSTM	Multiscale Convolutional LSTM
ANN	Artificial Neural network	RNN	Recurrent NN
SVM	Support vector machine	RF	Random forest
CNN	Convolutional NN	NB	Naïve Bayes
RUL	Remaining useful life	LSTM	Long short-term memory network
KNN	K-nearest neighbors algorithm	BLSTM	Bidirectional long short-term memory network
RUL	Remaining useful life	MCLSTM	Convolutional bidirectional LSTM
AE	Auto-encoder	XGBOOST	Extreme gradient boosting
		STFT	Short-time fast Fourier transform
		DBN	Deep belief network

✉ Muhammad Adil Raja  
Adil.Raja@dkit.ie

<sup>1</sup> Regulated Software Research Center (RSRC), Dundalk Institute of Technology, Dundalk A91K584, Co. Louth, Ireland

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

<sup>3</sup> Department of Industrial and Manufacturing Engineering, University of Engineering and Technology Lahore, Lahore 54890, Pakistan

## 1 Introduction

Tool wear is a significant concern associated with the cutting operation. Therefore, tool wear prediction utilizing AI approaches is critical for modern-day industrial processes [1]. Transfer learning is a method of ML in which a model is trained for one task, and it is repurposed for a second task. This methodology is very useful in the prediction of

tool wear. Chen et al. [2] have predicted the dynamics of the pose-dependent tooltip by using transfer learning. The assembly of the tool-holder was chosen as the source tool. By performing the impact tests, the dynamics of the pose-dependent tooltip were obtained as the source data and the same procedure was followed by selecting the target tool for obtaining the target data. Kurek et al. [3] have experimentally illustrated the methodology for the recognition of drill wear using transfer learning based on CNNs. They drilled holes into the laminated clipboard, and they manually took the image of the holes. Sun et al. [4] have used transfer learning based on the sparse auto-encoder (SAE) for the prediction of the useful life of the cutting tool. First of all, the SAE network was trained using the data of the cutting tool during the off-line process then the trained network was used to predict the new tool's data under the online cutting process using remaining useful life (RUL) information. Kim et al. [5] have done predictive modeling during the metal cutting process for machining power using multi-source transfer learning methodology. The acquired knowledge was transferred from the previous machining process to the target machining process where the machining power data was absent. Huang et al. [6] have worked on the prediction of the production progress of IoT-enabled manufacturing processes by employing the two-stage transfer learning based on the deep learning methodology. A deep belief network associated with transfer learning was used for the prediction of production progress. Li et al. [7] have predicted the state of tool wear by using transfer learning based on features. A genetic algorithm (GA) was used to choose a subset of sensor characteristics that were strongly connected to tool wear. Lee et al. [8] have used artificial neural networks (ANNs) for the prediction of tool wear using the signals of cutting force obtained by installing a dynamometer. The machining experiments were performed using the suitable force ratio for prediction. Ghosh et al. [9] have estimated the tool wear during the process of CNC milling by extracting the features from various machining zone signals and the sensor fusion model based on the use of an ANN which was designed.

The methods based on ANNs are also very useful for tool wear prediction. Kandilli et al. [10] have worked on the methodology for integrated sensor and tool wear monitoring using ANNs. They achieved good results. D'Addona et al. [11] explained the tool wear modeling by using ANN with the help of machining of Inconel 718. Machining experiments were performed for the tool wear characterization during the designing of engine parts of the Inconel 718 aircraft. Palanisamy et al. [12] experimentally investigated the prediction of tool wear during end-milling operations by using an ANN for mathematical regression. The data acquired from the milling tests were used to train an ANN. Chen et al. [13] demonstrated the real-time tool wear estimation approach using an ANN by using a 3-insert milling process.

Data obtained in milling tests were used to train an ANN by using the backpropagation methodology. Youneszadeh et al. [14] have used ANN for the prediction of project success in residential building projects (RBPs). The purpose of this research was to offer a framework for predicting the success of RBPs throughout the building phase.

Machine learning is a subfield of AI that is based on the notion that systems can learn from data, identify patterns, and make choices with little or no human intervention. Hao and Zhu [15] have devised a methodology based on an auto-encoder architecture for the monitoring of tool wear. To extract tool wear characteristics from multi-source monitoring signals, this study proposed a pyramid long short-term memory network (LSTM) based on spectral features. Huang et al. [16] have proposed a framework based on BLSTM for RUL prediction of the tool. The proposed methodology was validated experimentally using machining tool wear tests on a CNC milling machine. Xu et al. [17] have designed a model by using CNN for tool wear monitoring; 1D CNN was built for feature extraction from vibration data. Then, a fully connected ANN was developed for tool wear prediction. Chen et al. [18] have employed DBN for tool wear prediction by using multi-sensor data. To validate the DBN's superiority in forecasting tool wear, the DBN's performance is compared to ANN and support vector regression (SVR) based on the coefficient of determination ( $R^2$ ) and mean-squared error (MSE). Zhang et al. [19] have proposed a research study by designing the model based on LSTM for tool wear prediction, and the ARIMA model is also designed in time series. Qiao et al. [20] have successfully created a tool wear monitoring system based on multiscale convolutional LSTM (MCLSTM); the tool wear prediction task was accomplished by using BLSTM. Karandikar et al. [21] have monitored the tool wear by using a Naive Bayes classifier in the end-milling process. Wu et al. [22] have done a comparison of the performance of random forest (RF) with ANN and SVR for tool wear prediction, and the results show that the accuracy of RF was better than other algorithms. Wang and Huang et al. [23] have done tool wear modeling during hard turning based on RNN, and the extended Kalman filter (EKF) method is used to train the network. Qian et al. [24] have worked on a predictive tool wear model based on SVM. This study describes a new tool wear prediction model based on the integration of machined surface texture analysis and an SVM by using a GA. The proposed idea was to optimize the parameters by using a GA. Penedo et al. [25] have accomplished hybrid incremental modeling by using fuzzy KNN and least squares for the prediction of tool wear during the turning process; it is a two-step iterative method that utilizes a global model with the local model to capitalize on their complementary qualities. Alajmi and Almeshal [26] have proposed a method based on improving the hyper-parameters of the gradient boosting methodology using a

spiral dynamic optimization algorithm (XGBoost-SDA); further comparisons were done using SVM and multilayer perceptron neural network.

The novelty of this study is to determine the tool wear severity by overcoming three major shortcomings of the previous ML models for tool wear monitoring, i.e., retraining and remodifying of ANN during tool wear prediction, greater computational costs, high dimensionality, and overfitting. The rest of the paper is organized as follows: the methodology of the model based on ResNet-18 and the model's working based on similarity score is discussed in Section 2. Detailed experimental validation and discussion are in Section 3. A comparison of the developed model with existing methods is given in Section 4. Conclusions are summarized in Section 5.

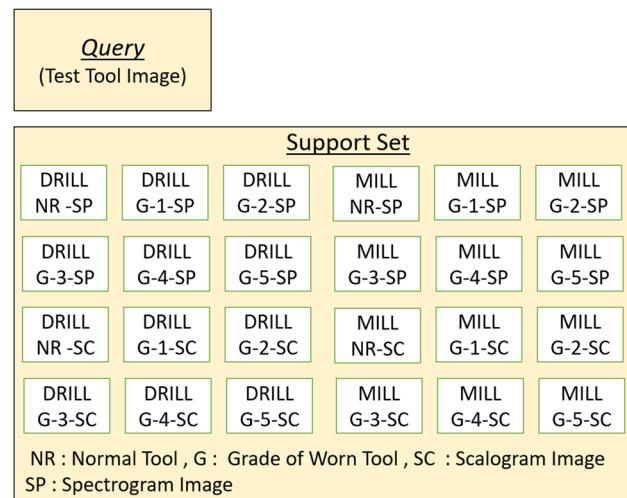
## 2 Development of a 12-way 2-shot learning model based on experimental data

The main purpose of designing the model based on 12-way 2-shot learning for the CNC milling and drilling processes is that it can work well with a very small dataset, as one-shot learning needs only one image to work. Face recognition systems are a suitable example of a one-shot learning methodology[27]. A large amount of data lead to high dimensionality and high computational costs as well as overfitting, so the model has effectively solved the major problems of existing ML models. First of all, N-way K-shot learning is defined in such a way that  $N$  represents the total number of classes based on the grades of milling and drilling tools and  $K$  represents the number of scalogram and spectrogram images present in the support set from each class. The main quality of the 12-way 2-shot learning algorithm is to solve the classification problem and to make predictions by using very small data samples. The basic working of the model is based on the query image and the image data of the support set. A total of 24 images are obtained in the primary phase from 6 milling and 6 drilling tools, as each tool will generate two images, i.e., a spectrogram and a scalogram. The image augmentation algorithm has tremendously increased the image data bank to 240 images to input into the 12-way 2-shot learning algorithm for training. The images used for training include the 24 original images of the support set plus 216 images obtained by the image augmentation methodology. Two types of conversions for the time-series data into images have been employed. For the first type, the time-series data is converted into scalogram images, whereas for the second type, the time-series force signals are converted into spectrogram images.

The working of 12-way 2-shot learning algorithm is based on a support set and a query set, and it is a type of transfer learning methodology. The methodology of the model

training is completely different from standard supervised learning. The working of 12-way 2-shot learning is not to recognize the images in the training set and then generalize the test set. Instead, its goal is to “learn to learn” which is called meta-learning [28]. The algorithm works by finding the similarities and differences between the images. The training data has 12 classes which are normal mill tool, grade-1 worn mill tool, grade-2 worn mill tool, grade-3 worn mill tool, grade-4 worn mill tool, grade-5 worn mill tool, normal drill tool, grade-1 worn drill tool, grade-2 worn drill tool, grade-3 worn drill tool, grade-4 worn drill tool, and grade-5 worn drill tool. Each class has 20 images, i.e., two original images and 18 artificially produced images using the image augmentation algorithm. (There is a total of 12 classes, so the total number of images is 240.) For example, the support set will get two images of the scalogram and spectrogram from the normal mill tool class, thus making the model 12-way 2-shot because of having 12 classes and 2 basic original images from each class. The query contains only the test image from the test tool, and this test image may be a spectrogram or scalogram. The query image along with one image out of 24 images of the support set will be input to the model, as shown in Fig. 1.

The model will compare the features extracted from the query image with extracted features of each image out of the 24 images present in the support set. The model will determine the similarity scores between the query image and all images from the support set one by one. The model will pick up the image from the support set with the greatest similarity score, and thus, it will mention the label of that image class as the final output result. For example, the test image in the query is of the G-1 drill tool scalogram. The model will match it with every 24 images in the support set one by one, and the similarity score for the G-1 drill tool



**Fig. 1** Support set and a query set of 12-way 2-shot model

scalogram image of the support set will be highest compared to the other 24 images. So, the model will mention the label of the class and give the final result that the test tool is a grade-1 worn tool.

## 2.1 12-Way 2-shot learning model operation based on ResNet-18

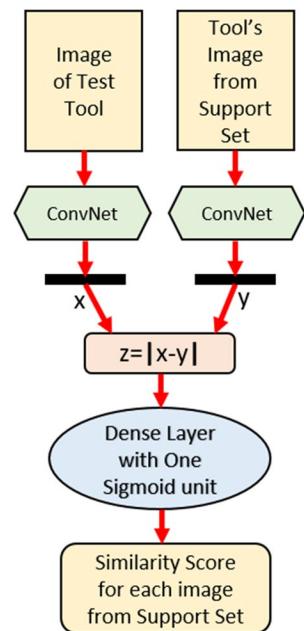
Transfer learning is an ML methodology where pre-trained models are reused to solve new problems. The ANN learns features from the data which are compiled as “weights” of the network during training on a particular task, and this ANN can be referred to as a base network. The weights of the base ANN are extracted and transferred to the target ANN which is fine-tuned. So, instead of training the target ANN from scratch, the already learned coefficients are transferred to it. In our case, the base network ResNet-18 is already trained on more than one million images, and in this case, it is called a base network. The task is the same, so when ResNet-18 is used for this research study, it is called a target network. It solves a similar task after fine-tuning the dataset based on scalogram and spectrogram images. ResNet-18 is the pre-trained CNN that is designed for this model to extract feature vectors from the images. This indicates that the deep network must not have a higher training error than its shallower equivalents. The image is inputted into the convolution layer of filter size  $7 \times 7$  having a total of 64 channels, and after that, there would be a pooling layer which would result in the stack of convolution layers having the fixed filter size of  $3 \times 3$  and 64 channels. The number of channels would get doubled in every 4-set convolution layers, and the skipping layers’ methodology is applied [29]. The final 4-set convolution layers after the application of the skipping layers’ methodology would have a total of 512 channels and have a fixed filter size of  $3 \times 3$ , and the final layer is fully connected.

## 2.2 Model’s working methodology based on similarity scores

The 12-way 2-shot learning model works by extracting the feature vectors “ $x$ ” and “ $y$ ” using a pre-trained conv (ResNet-18) from the two images, i.e., the query image and each image out of 24 images from the support set. Then, these two vectors are subtracted from each other, and the resultant difference vector “ $z$ ” is processed by the dense layers and gives the scalar output, and finally, the sigmoid activation function is applied (Fig. 2) [30].

The label is previously prepared based on the test image. For example, the target for the test image is set as 1, and the scalar output from the network is expected to be close to the target. The loss function is used to measure the difference between the target and the predicted value. Due to the loss,

**Fig. 2** Working of the 12-way 2-shot learning model



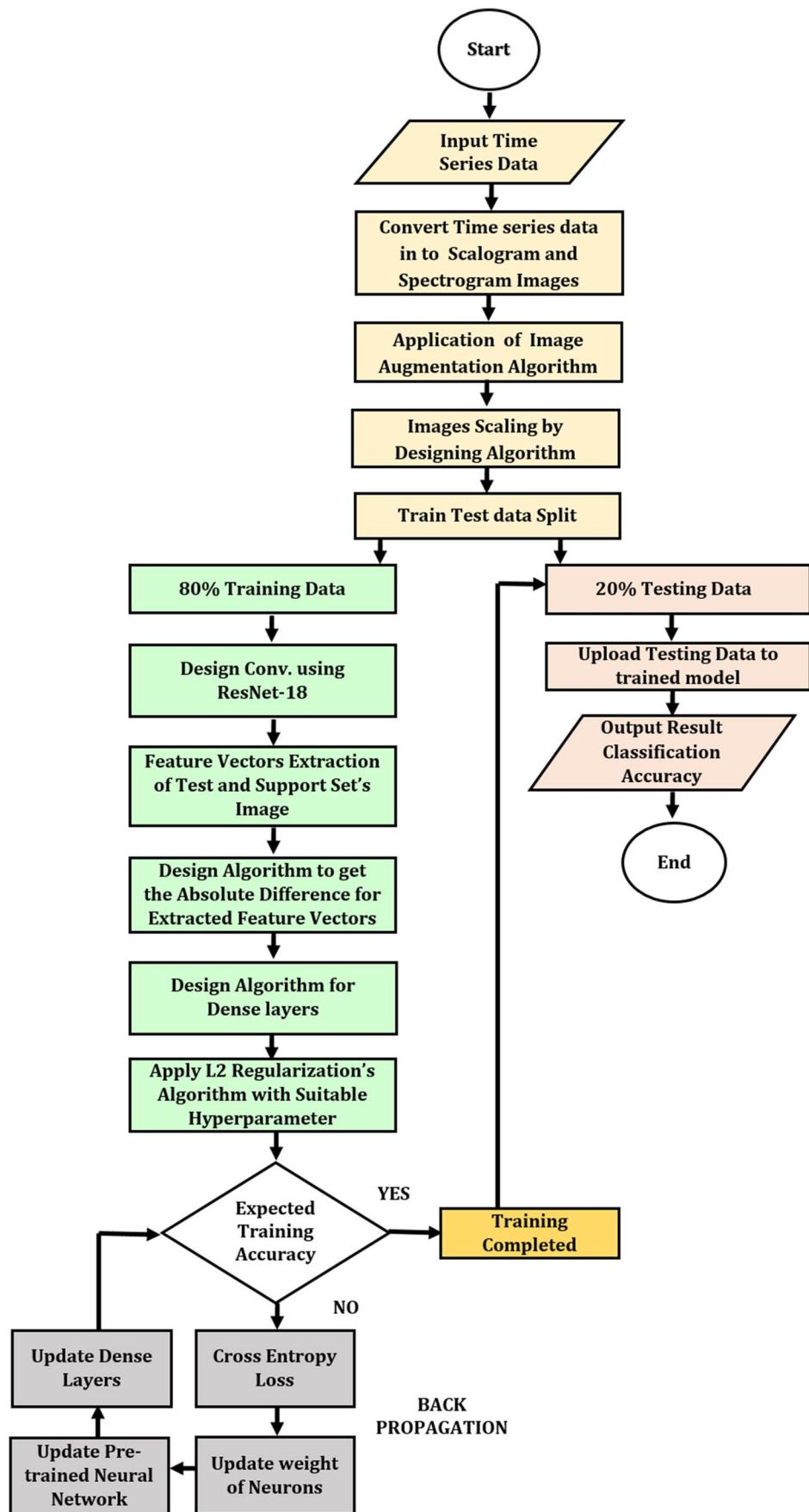
the gradients are calculated using error backpropagation, and gradient descent is utilized to update the model’s parameters. The model contains two parts. For the first one, pre-trained ResNet-18 is used only for extracting the feature vectors from the images, and the second part of the model was based on dense layers. So, to overcome the loss, both parts need to be updated. By knowing the gradient of the loss concerning the parameters of the dense layers, the parameters can be updated by gradient descent (Fig. 3).

## 3 Experimental validation and discussion

### 3.1 Experimental setup and procedure

This research study is based on performing milling and drilling operations using the CNC machine with a dynamometer with its tool magazine containing 12 different types of drilling and milling tools, thus six tools for each machining operation. The first tool with sharp edges was considered the normal tool, whereas the other five worn tools were categorized based on increasing values of  $V_B$  (flank wear land width). All tests were performed on a 3-axis CNC machine with the spindle motor having 22 kW power, and the value of maximum rotation of the tool is 12000 r/min. The workpiece is composed of Inconel 718. A KISTLER 9257BA stationary dynamometer is attached to the tool for the measurement of force signals. The milling and drilling operations were performed on the Inconel 718 to collect the algorithm’s testing/training force signals data. The National Instruments (NI) PCI-6025E digital/analogical data acquisition board and Lab view 8.5 software were used for the measurements of force

**Fig. 3** Demonstration of the designed algorithm for the model



signals. A total of 12 experimental tests were conducted. Six experimental tests were for the drilling process, whereas six experimental tests were carried out for the milling process. The main purpose of these 12 tests is to get the force signals data to train and test the 12-way 2-shot learning model. During the drilling process, the holes were drilled into the workpiece composed of Inconel 718. The first tool used to drill holes was the normal tool with sharp edges having no abnormality. The other five tools used to drill holes were worn tools with damaged edges. The five tools were categorized from grade-1 to grade-5 based on  $V_B$ . The same procedure was followed for the milling process, but unlike the drilling process, the specific patterns were designed into the workpiece instead of holes. The test tools were used for the collection of data signals to test the developed model. The tools are selected with the particular values of tool wear land width  $V_B$ . The spindle speed and feed rate would increase in ascending order for grade-1 to grade-5 tools for both drilling and milling processes. The value of the cutting length for the drilling process is smaller compared to the values of the cutting length for the milling process because only holes are drilled during drilling processes which have a smaller value of diameter. However, the values of the cutting depth of the milling process are greater than the cutting depth of the drilling process. The milling and drilling experiments were conducted in the same research group, so the same machining setup was used for this research study which is reported in [31]. The tools which have shown zero flank wear land width for both milling and drilling experiments were categorized as normal tools. The workpiece is composed of Inconel 718. End mill and twist drill tools specifications followed ISO 5103–1969 (end mill: HSS M2, 20 mm diameter, cutting length: 38 mm, overall length: 104 mm, number of flutes: 6, hardness: 750–950 HV; twist drill: HSS M2, 15 mm drill diameter, flute length: 114 mm, overall length: 212 mm, hardness: 750–950 HV). Tool wear was not predicted based on the data presented in Table 1. The tool wear prediction was accomplished based on the force signals data values in Figs. 8 and 9. The force signals data values were converted into the images of scalograms and spectrograms, as shown in Figs. 10, 11, 12, and 13. The cutting parameters in Table 1 were used only during the cutting operations based on milling and drilling processes. These cutting parameters, which are feed rate and spindle speed of milling and drilling, were used by the CNC machine to conduct the cutting operations on the workpiece.

The value of flank wear land width for the worn tool has increased gradually from a grade-1 worn tool to a grade-5 worn tool. The drilling operation was performed with 5 worn tools and 1 normal tool, and the same procedure is followed for the milling operation. To overcome abnormalities and eliminate noise in the data, we must use the same cutting conditions. Due to the input of noise-free force signals, the

**Table 1** Cutting parameters

Tools types	Feed rate (mm/rev)	Spindle speed (r/min)
Normal tool (drilling)	0.11	500
Grade-1 worn tool (drilling)	0.14	650
Grade-2 worn tool (drilling)	0.17	800
Grade-3 worn tool (drilling)	0.20	950
Grade-4 worn tool (drilling)	0.23	1000
Grade-5 worn tool (drilling)	0.26	1200
Normal tool (milling)	0.25	900
Grade-1 worn tool (milling)	0.28	1050
Grade-2 worn tool (milling)	0.31	1100
Grade-3 worn tool (milling)	0.34	1250
Grade-4 worn tool (milling)	0.37	1300
Grade-5 worn tool (milling)	0.39	1450

model worked well. The resultant force signals followed a time-series pattern (Table 2).

To determine the spindle speed and feed rate, a specific methodology is used. Every tool has its own set of SFM (surface feet per minute). The tool catalog contains the SFM range (from min. to max.). There is a formula that may be used to convert SFM to RPM (spindle speed). The formula for calculating spindle speed is given below.

$$n = \frac{SFM \times 12}{D \times \pi} \quad (1)$$

The diameter of the cutting tool is  $D$ , and the spindle speed is  $n$ . From the tool catalog given by the tool manufacturer, the specific tool number and material group (workpiece material) are utilized to find the range (from min. to max.) of SFM. If the SFM is incorrect for the tool, the spindle speed and feed rate will be incorrect as well, resulting in the tool being burned. As a result, the SFM searching strategy that uses information such as the tool's kind and specific number, as well as the material group, is critical. It is preferable to choose an SFM that is close to the minimum range because the higher the SFM, the more unstable the tool will be, and therefore, the tool's life will be shortened (Figs. 4, 5, and 6).

$$\text{feed rate} = \text{Spindle speed} \times \text{Feed} \quad (2)$$

### 3.2 Milling and drilling flank tool wear visual illustration

Flank wears usually occur at the side of the cutting edge. Figure 7, given below, has illustrated clearly the flank tool wear of drilling and milling processes. The diagram given

**Table 2** Technical data involving the 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 <sup>2</sup>
Cr	21	Ultimate tensile strength (UTS)	MPa	1.04*10 <sup>3</sup>
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 <sup>3</sup>
Co	0.21	Electrical resistivity (at 20 °C aged)	Ω·m	1.21*10 <sup>-6</sup>
Fe and others	Balance	—	—	—

below has demonstrated the approach to the measurement of flank wear. The region of the flank wear is called wear land  $V_B$ , and W is represented as the width of the outer corner of wear land. If the value of  $V_B$  becomes greater than 0.5 to 0.6 mm, then the cutting forces will cause the failure of the tool being used in the cutting operation. Flank wear is measured by using the average value of wear land size and the maximum value of wear land size. It can be defined with the help of the expectancy equation of tool life. Another important reason for the occurrence of flank wear is due to the high value of feed and larger depth of cut. Micro weld shearing is present between the work material and the tool. It increases the cutting force which is required to cut the surface. Flank wear adversely affects the dimensional accuracy of the machined product. It will increase the surface roughness of the final product. Flank wear will change the shape of the workpiece.

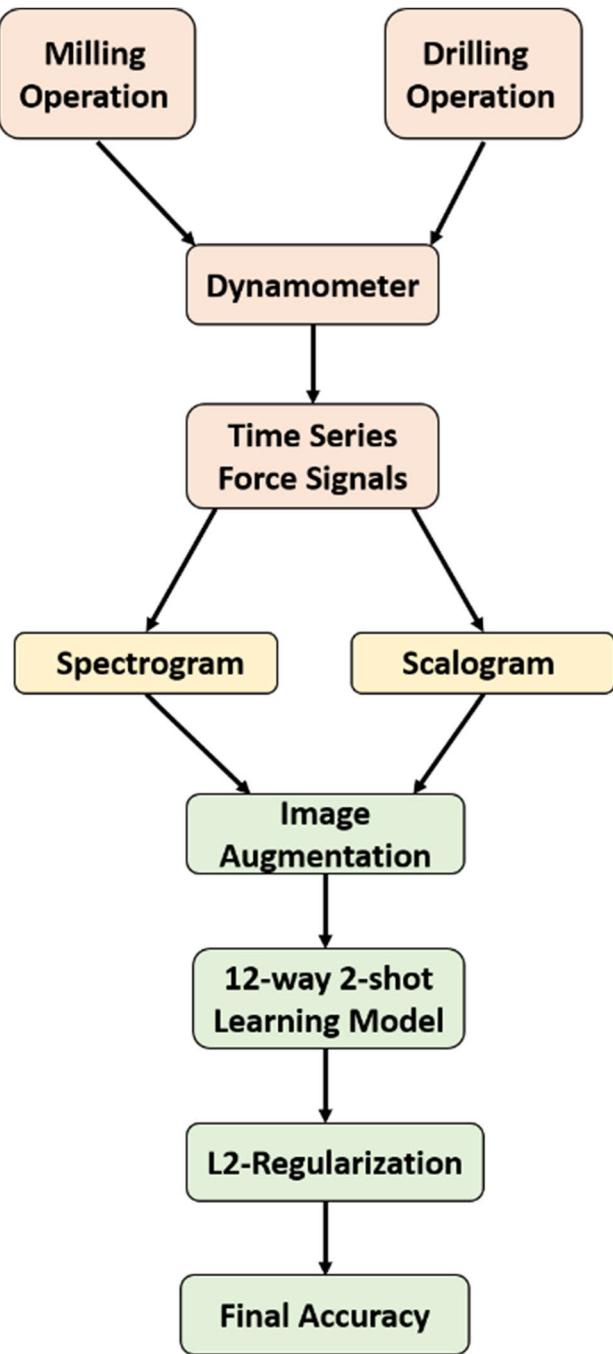
### 3.3 Drilling experiments using the normal tool

The drilling experiments were conducted successfully on Inconel 718 using a force sensor and a 3-axis CNC drilling machine. The force signals were collected for normal tool and different types of worn tools. For the first drilling test, the normal tool was selected to drill holes into the workpiece. The normal tool possesses sharp edges. The total time utilized to drill each hole into the workpiece was 30 s with a step size of 0.13 s. The dynamometer has given a total of 222 data values of force signals for each drilling experiment which is demonstrated by graphical representation, as shown in Fig. 8. This graph is even and regular. The graph represented all the top peaks on the region of force values between 275 and 300N, and this smooth representation of the given graph depicts that the corresponding tool used is fine and possesses sharp edges. Each drilling operation was divided into a total of 3 processes. The first process is called drilling in, followed by the steady drilling process, whereas the final process is named drill-out. During the drilling-in process, the preliminary force has applied to the workpiece surface and its value is increased gradually from the initial

value of 0N. The final process is named as the drilling-out process, during which the value of force is reduced gradually and constantly. The force is increased from 0 to 275N; in Fig. 8f, this depicts that during the preliminary phase of the drilling process, the maximum value of force has reached approximately 275N. The top peaks displayed in the graph are so much alike, and their in-between distance is nearly the same which depicts the smoothness of the conducted drilling process using the tool which contains sharp edges. The minor waveforms are displayed during the drill-out process which represents that the diameter of the hole and the diameter of the drill bit are nearly equal which is a reassuring result. The force is decreased gradually from approximately 210 to 0N. After the completion of the drilling experiment, raw force signal data is collected for the further processing of the conversion to spectrogram and scalogram images.

### 3.4 Drilling experiments using worn tools

The second section of the experiment comprises drilling using different grades of worn tools. Five different twist drill tools were selected carefully. These tools were divided into five classes based on the width of the tool wear  $V_B$ . The twist drill's end is called the shank. This area is strongly clutched by the chuck of the twist drill bit, whereas the chisel edge is the area present at the tip of the drill bit. The force signals of all grades of worn tools are plotted in Fig. 8. The drilling operation for each worn tool is also divided into three processes. The first process is drill-in which is illustrated in the graphs in Fig. 8, the area where the force increased from 0N and reached a certain level for constant drilling. Steady drilling is the process of the operation of drilling when the force remains constant and fluctuates between particular levels. This fluctuation depends upon the grade of the tool worn. For instance, if the tool wear of any particular tool is severe, then these fluctuations would be greater than the fluctuations of the force signals of the normal tool. The drill-out process is defined as the withdrawal of the cutting tool from the workpiece by reducing the spindle speed gradually. The most significant and obvious observation is that the force



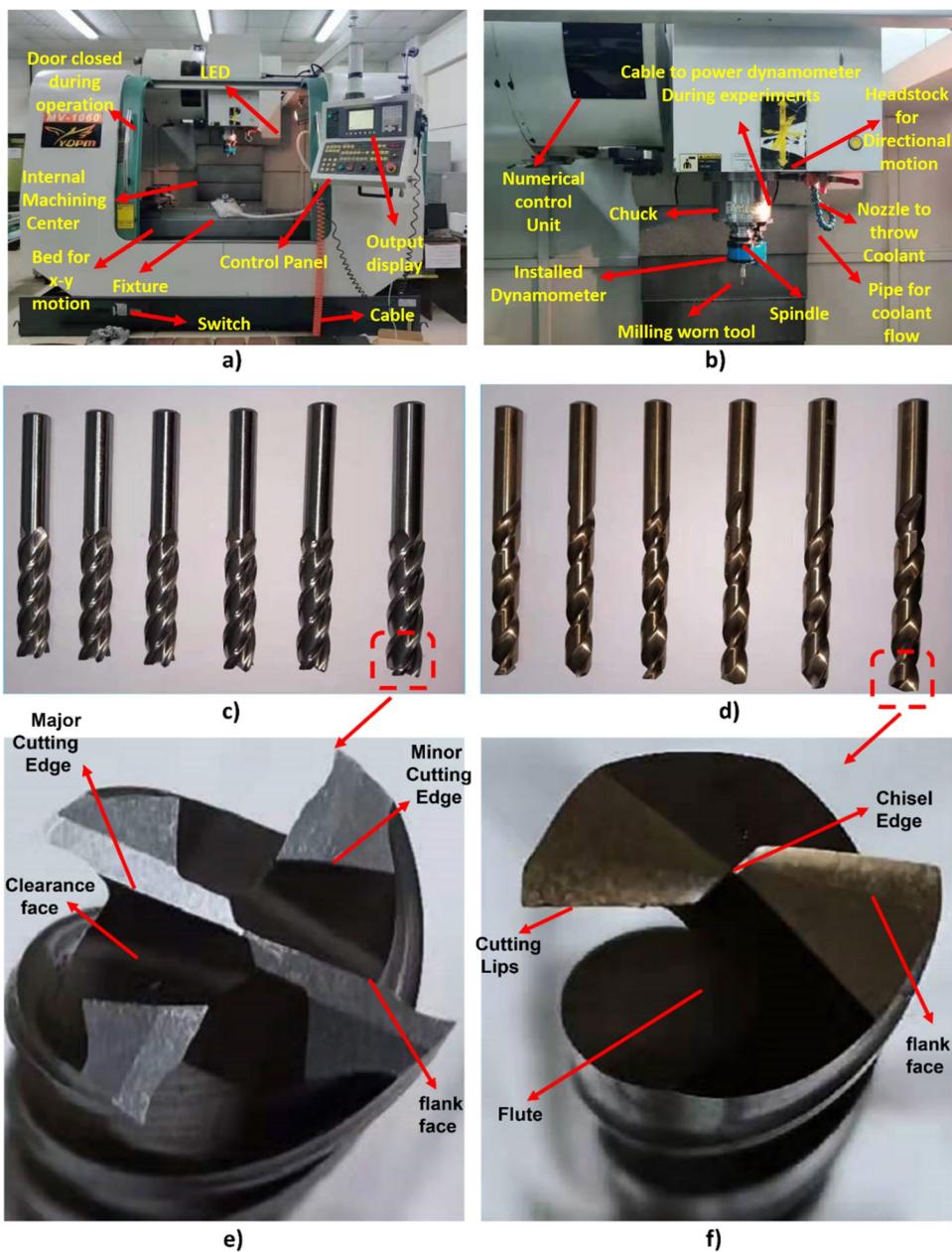
**Fig. 4** Block diagram demonstrating the data flow

in the graphs of types of worn tools has increased considerably compared to the force represented in the graph of the normal tool. The maximum force for the grade-1 tool has increased to 350N, as shown in Fig. 8a; the highest level of force for the grade-2 tool is nearly around 450N, as displayed in Fig. 8b; the force has touched the greatest level of 600N in the case of Fig. 8c. The trend of increasing force with the increase in tool wear land has also followed for grade-4 and grade-5 tools, where the force signals have reached the

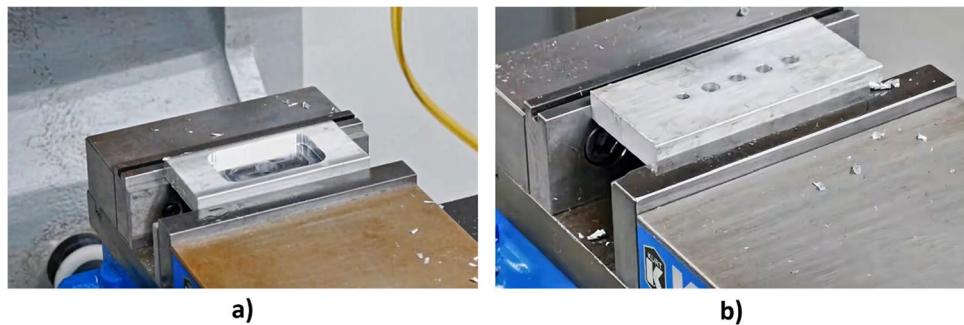
extreme value of 750N and 925N, respectively. If this region of the graph for the initial drilling of different grades of worn tools is compared with the graph of the normal tool given in Fig. 8f, during the initial phase of drilling, the force has increased to the maximum value required for constant drilling which is termed a drill-in process that has shown irregular behavior with large abnormal waveforms and sharp spikes which depict the extreme rise and fall of the force during the initial stage of drilling. That is quite noticeable for all grades of worn tools. The grading of the tool wear is categorized based on the tool wear land width values. The first tool with sharp edges was considered as the normal tool, whereas the other five worn tools were categorized based on increasing values of  $V_B$  (flank wear land width). Grade-1 was assigned to the minimal worn tool with the smallest value of  $V_B$ , and grade level increases with the increment in the value of  $V_B$ . However, grade-5 was given to the extremely worn tool with the highest  $V_B$  value. Six different tools of one type are used; for example, tools of twist drill type were used for drilling, and the tools of end mill type were used. To set to standard and common cutting parameters, the tools of the common type were used for each machining operation because there are thousands of different types of tools available in the market for cutting operations in different ways. The main aim of this research study was to focus on tool wear based on an efficient prediction algorithm. If the tools of different types were used, then the data would be extremely random and the quality of the machine learning tool wear prediction model would be very low. It is impossible to get the tools with a predefined tool wear range. The worn drilling and milling tools were chosen for performing experiments. The severity of the tool's wear is directly related to the value of  $V_B$ . For the drilling tools, the grade-1 tool has the least value of  $V_B$  equals to 0.12 mm,  $V_B$  of the grade-2 tool is greater than the grade-1 which equals 0.21 mm, whereas the width of the tool wears land of grade-3 and grade-4 tools were set as 0.34 mm and 0.41 mm, respectively. However, the grade-5 tool is extremely damaged having the highest value of  $V_B$  which was 0.57 mm. With regard to milling tools, the values of  $V_B$  for grade-1 worn tool, grade-2 worn tool, grade-3 worn tool, grade-4 worn tool, and grade-5 worn tool are 0.14 mm, 0.26 mm, 0.36 mm, 0.41 mm, and 0.49 mm, respectively. For milling operation, the tool can move in 3 axes, which are the  $x$ -axis,  $y$ -axis, and  $z$ -axis. If the operation is performed by each tool in each axis, then the 222 data values of the force signals would be generated with the step size of 0.13 s. The drilling operation can only be performed on one axis in order to drill holes in the workpiece for the generation of the force signals, so considering the equal data size of milling and drilling operations, only one axis was chosen to perform the milling operation.

The lengths of the waveforms during the steady drilling have increased with the increase in the tool wear land which

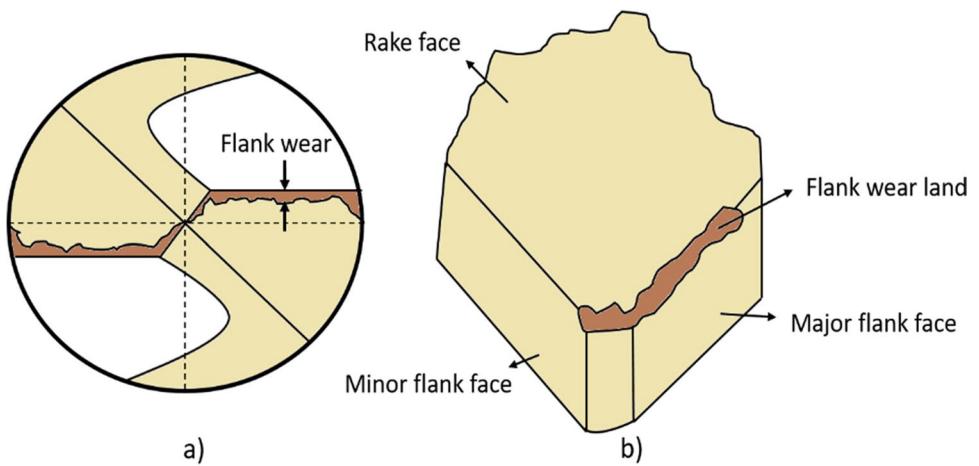
**Fig. 5** (a) CNC machine; (b) internal machining center; (c) milling tools; (d) drilling tool's tip area; (e) milling tool's tip area; (f) drilling tool's tip area



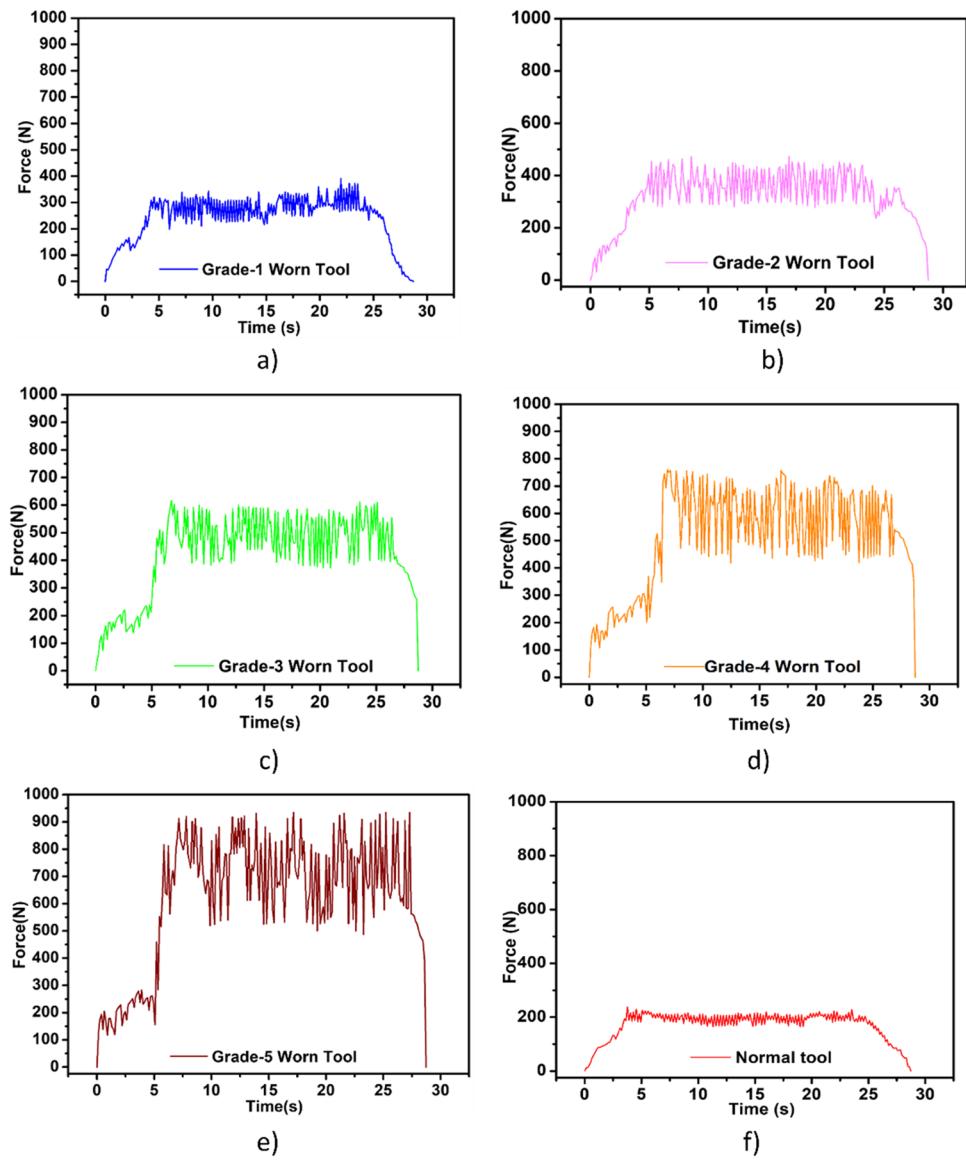
**Fig. 6** (a) Inconel 718 work-piece used for processing by milling operation; (b) Inconel 718 workpiece used for processing by the drilling operation



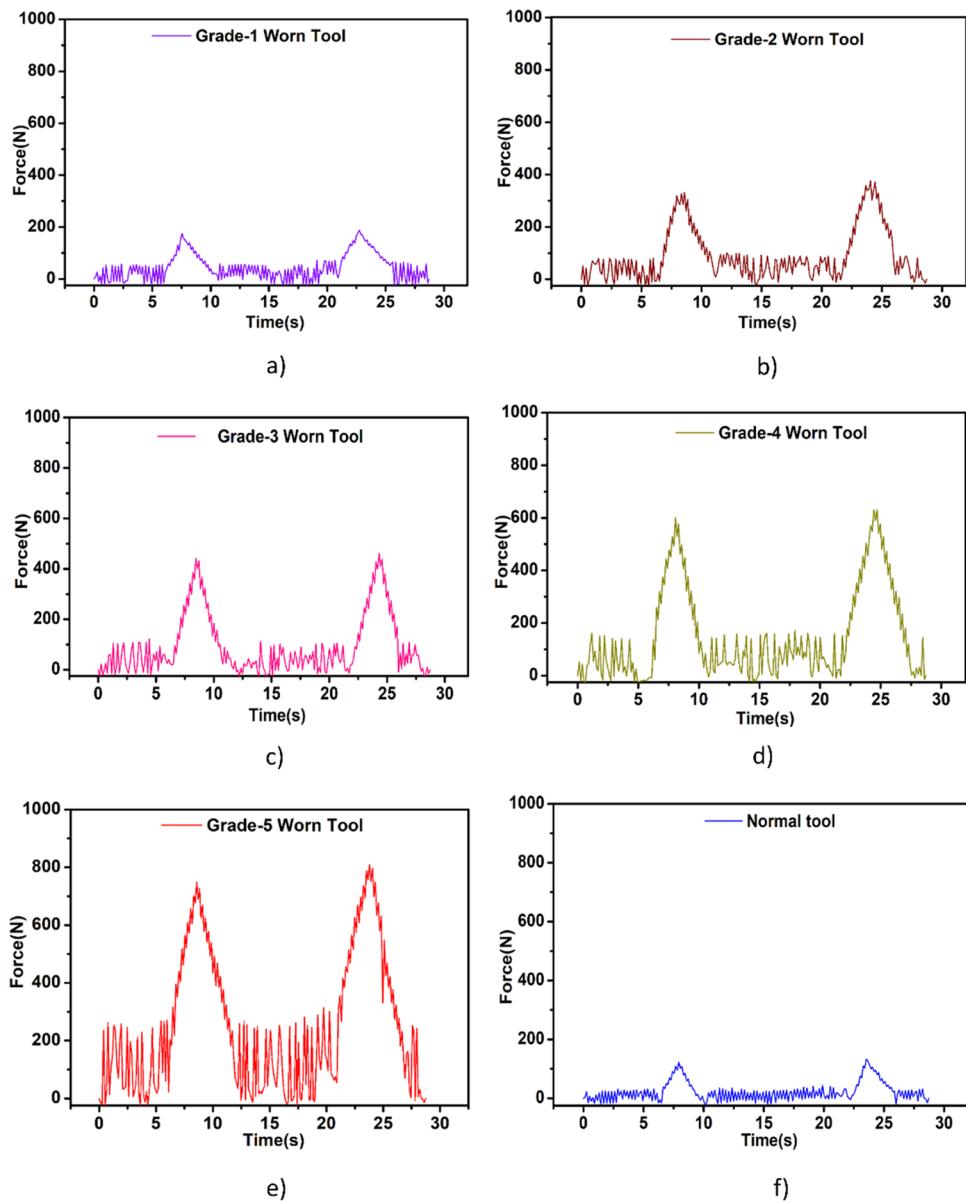
**Fig. 7** Schematic representation of flank wear tool: (a) drill worn tool and (b) mill worn tool



**Fig. 8** Graphical representation of raw force signals for drilling tools: (a) grade-1 worn tool, (b) grade-2 worn tool, (c) grade-3 worn tool, (d) grade-4 worn tool, (e) grade-5 worn tool, and (f) normal tool



**Fig. 9** Graphical representation of raw force ( $x$ -direction) signals for milling tools: (a) grade-1 worn tool, (b) grade-2 worn tool, (c) grade-3 worn tool, (d) grade-4 worn tool, (e) grade-5 worn tool, and (f) normal tool

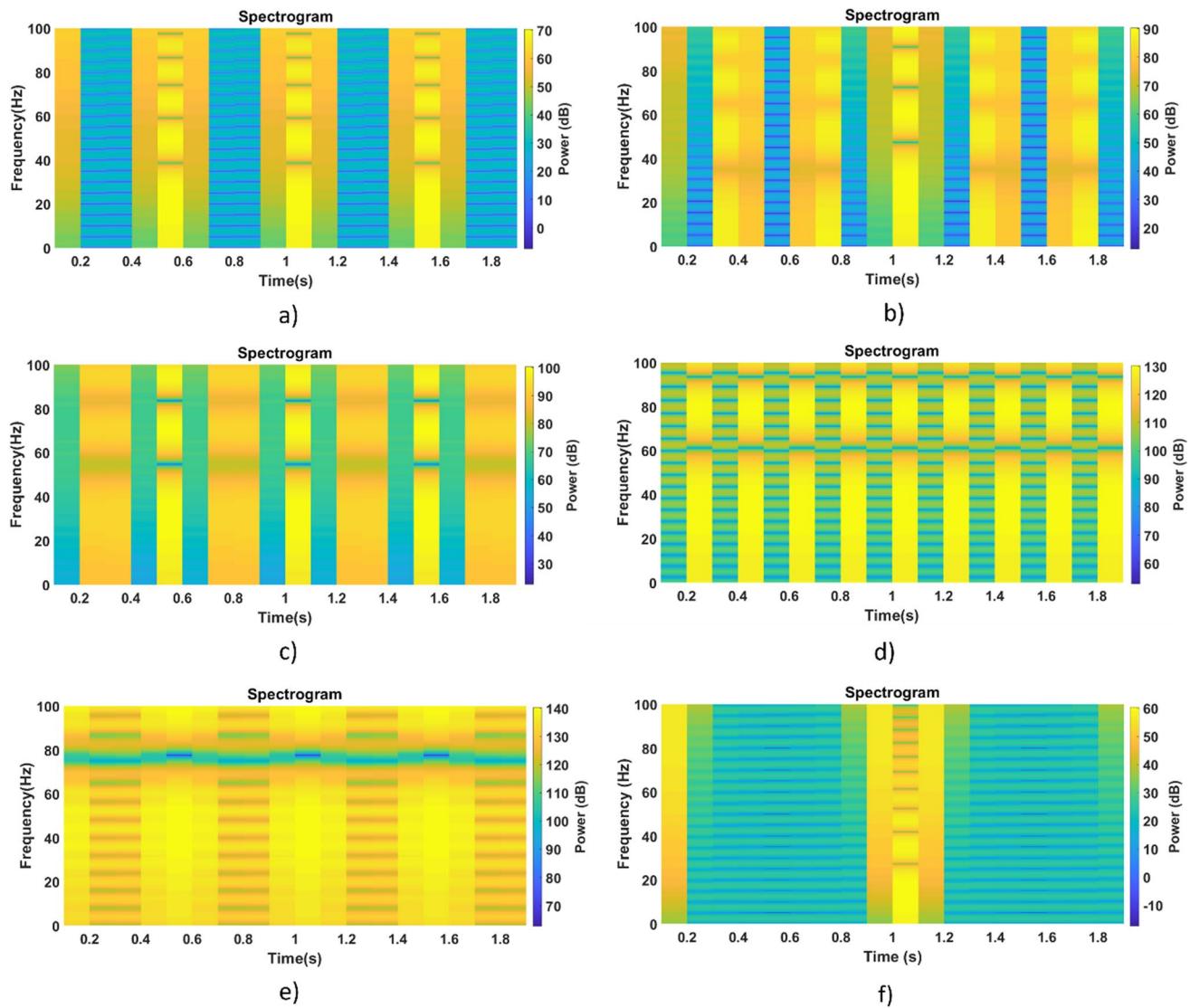


depicts that tool wear will lead to an increase in impact force exerted on the surface of the material used. With the help of comparative analysis for the graphs of all worn tools, it is very much obvious that tool wear has a great impact on the force signals generated by the dynamometer. The waveforms of the grade-1 tool have a range from 300 to 450N, the grade-2 tool range from 370 to 600N, the grade-3 tool range from 400 to 600N, the grade-4 tool range from 450 to 750N, and the grade-5 tool range from 550 to 900N.

### 3.5 Milling experiment using the normal tool

The second machining operation used in this research study is the milling process. The 6 tools include one normal tool and 5 worn tools. The worn tools were divided into 5 grades

ranging from grade-1 to grade-5 based on the flank wear land  $V_B$ . Firstly, the milling operation was performed with the normal tool which is not worn and has no flank wear land. The milling tool has performed the cutting process on the Inconel 718. The chosen workpiece made of Ni-based superalloy possesses unique properties. It has a great effect on the condition of the cutting tool due to the high value of hardness, strength, and ductility. The dynamometer was installed on the normal tool of the milling machine for the collection of the force signals. A total of 222 force signals' data values were obtained after the completion of the milling operation performed on the Ni-based superalloy. The force signals in the  $x$ -direction are plotted in Fig. 9f. The maximum force value obtained by milling with the grade-1 worn tool was 125N. The force applied by the tool on the



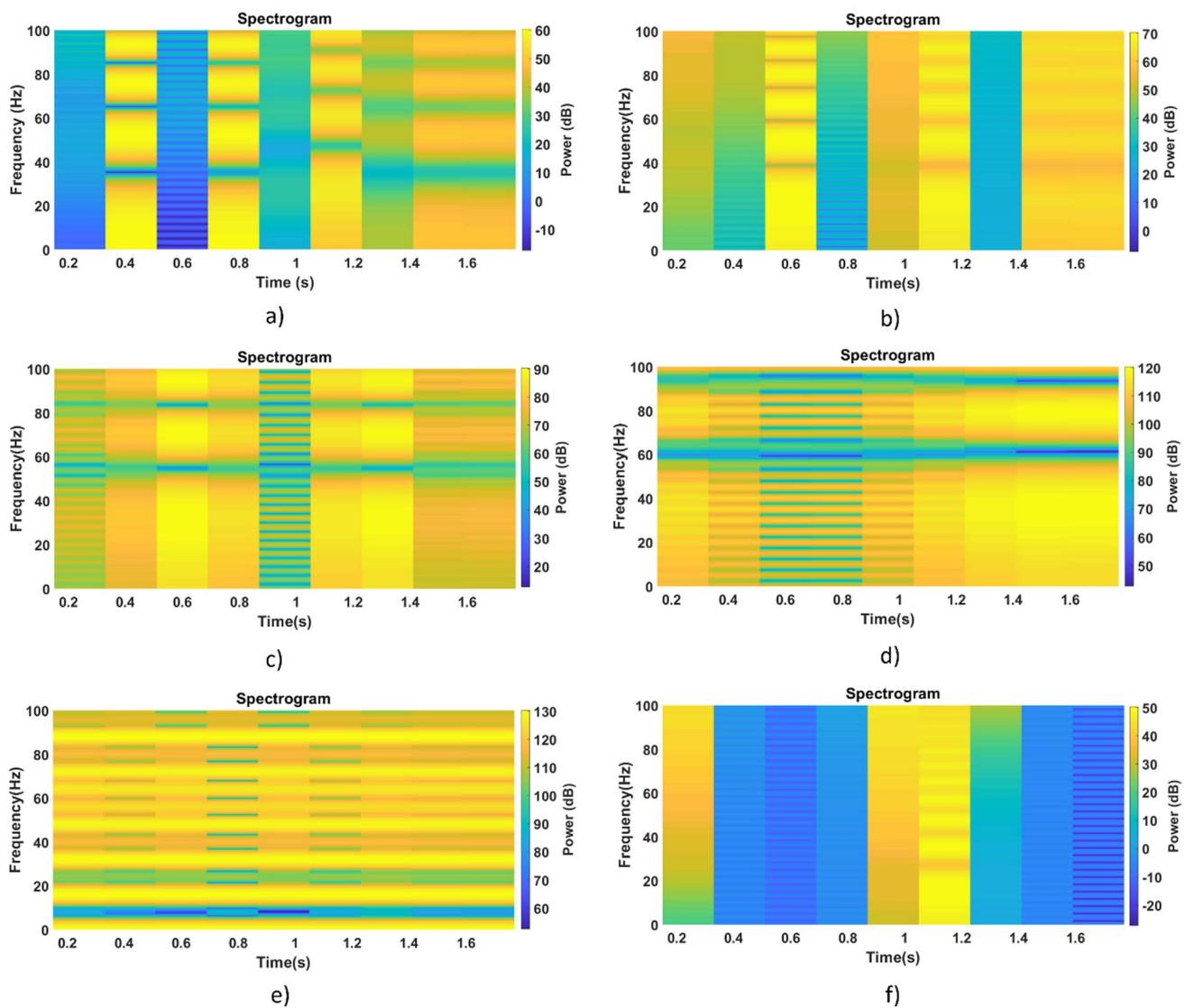
**Fig. 10** Spectrogram of drilling force signals for (a) grade-1 worn tool, (b) grade-2 worn tool, (c) grade-3 worn tool, (d) grade-4 worn tool, (e) grade-5 worn tool, and (f) normal tool

workpiece is directly related to the tool wear. In the milling experiments, the force value for the normal tool is the least compared to the various grades of the worn tools because there is no flank wear land present on the normal tool. So, during milling using a normal tool, the least cutting force was applied on the workpiece (Ni-based superalloy) parallel as well as perpendicular to the feed direction. The tool possesses fine and sharp cutting lips and chisel edges; therefore, the material was removed very smoothly by applying the least force.

### 3.6 Milling experiments using worn tools

The most important part of the experiment is to collect the force signals dataset by performing the milling operation

using the various grades of the worn tool. A total of 5 worn tools were chosen under careful examination. Each milling tool has generated one set of data based on both spectrograms as well as scalogram images, thus labeled as each class out of 6 milling classes. The quality of the ML model is that it will assign the label to the predicted class based on the nearest approximated value. For instance, if a worn tool with a value of 0.2 would be used for the milling operation, then the model will assign the nearest class based on the tool wear value. As in the case scenario, the model will assign the grade-2 worn milling tool. The force signals in the  $x$ -direction are plotted in Fig. 9. Plastic deformation of the tooltip is due to the tool wear. Cutting force value is directly related to plastic deformation as the decrease in plastic deformation would lead to a decrease in cutting force and



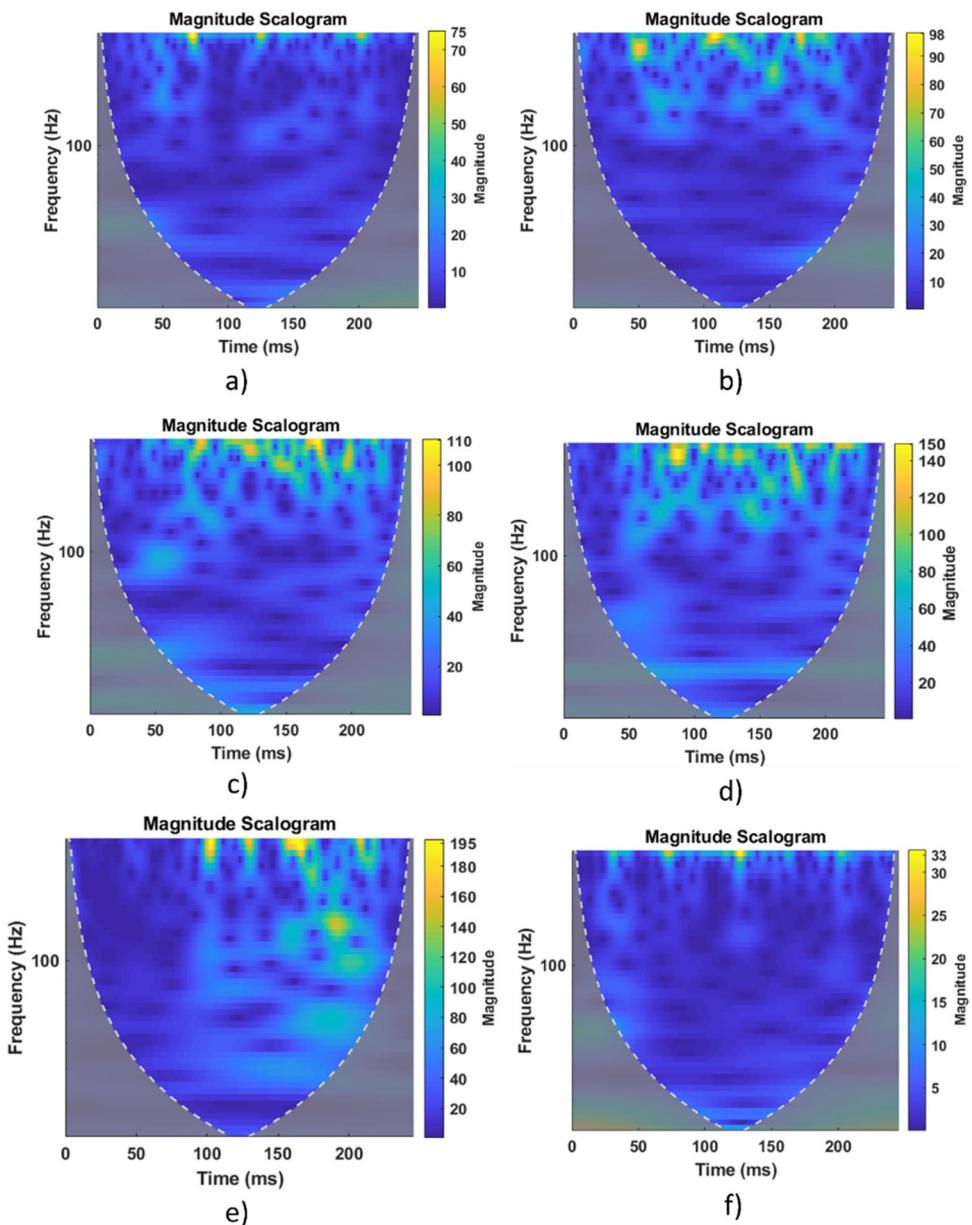
**Fig. 11** Spectrogram of milling force ( $x$ -direction) signals for (a) grade-1 worn tool, (b) grade-2 worn tool, (c) grade-3 worn tool, (d) grade-4 worn tool, (e) grade-5 worn tool, and (f) normal tool

vice versa. The force applied by the worn tool is directly proportional to flank wear land width  $V_B$ . The approximate peak value of force exerted by the grade-1 worn tool is 175N, as graphically represented in Fig. 9a. The grade-1 worn tool has applied the least peak force on the workpiece because it has the least value of flank wear width land. However, approximately 750N peak force value for a grade-5 worn tool was reported as this tool has the highest flank wear width land, as shown in Fig. 9e.

The peak force values recorded by the installed dynamometer for the grade-2 worn tool, grade-3 worn tool, and grade-4 worn tool are approximately 350N, 450N, and 625N, as graphically demonstrated in Fig. 9. All the graphs of the worn tool have shown similar patterns. Each cutting process was initiated from the instant of milling

tool touched the workpiece for cutting and ended at the instant when the tool did not have any contact with the workpiece, and during the same time, the value of the cutting force increased from zero to the maximum value and then decreased from a maximum value to zero, which is defined as the milling cycle. In each milling cycle, the cutting forces of worn tools are greater than the cutting forces of the normal tool. The cutting forces of worn tools increase with the increase in flank wear land width. All the chosen tools have plastic deformation at the cutting edges of the tool bit, and this region is also included in the tool wear land, whereas the cutting edge possesses plastic deformation at the nose radius. Another observation is that the plastic deformation in the case of an already worn tool

**Fig. 12** Scalogram of drilling force signals for (a) grade-1 worn tool, (b) grade-2 worn tool, (c) grade-3 worn tool, (d) grade-4 worn tool, (e) grade-5 worn tool, and (f) normal tool



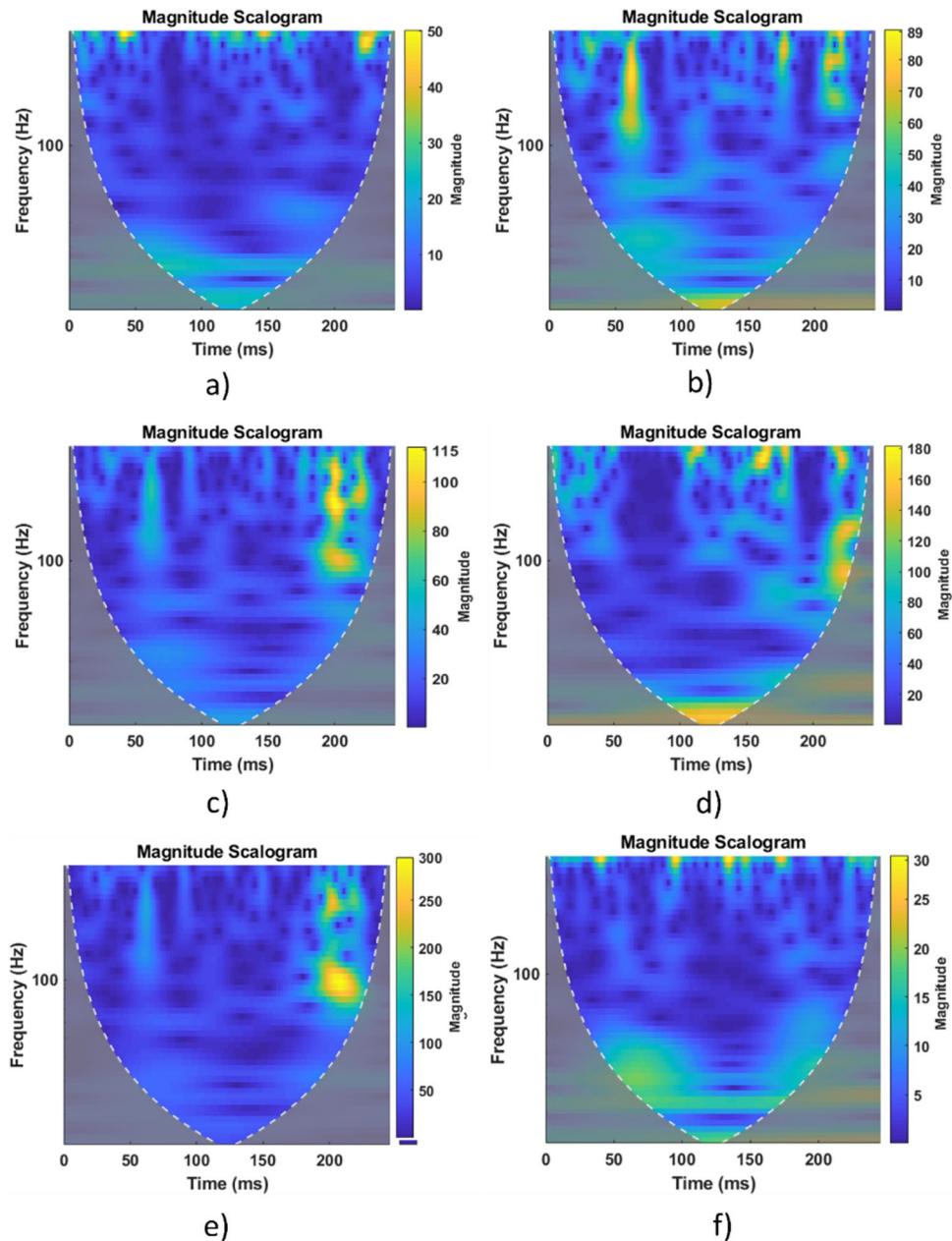
is progressive as it will increase with the increase of the impact of force during the constant milling process.

### 3.7 Conversion of raw force signals time-series data of machining processes into spectrogram images

A spectrogram is a graphical approach for demonstrating the strength of different frequencies present in a time-varying signal related to a waveform. A spectrogram is a type of graph having two geometric dimensions in which the first axis denotes the time, whereas the other axis denotes frequency. A third dimension is also depicted in a spectrogram which indicates the amplitude of a specific frequency at a

certain point in time. And this third dimension is generally represented by different colors. Spectrograms are formed by using the time-domain signal in two different ways. According to the first method, the spectrogram is approximated as the filter bank which ensues from a bank of band-pass filters. The second and most advanced methodology for creating the spectrogram is by converting the time-domain signal to the frequency domain using the well-known Fourier transform. Generating a spectrogram using Fourier transform can happen through a digital process. The data which is digitally sampled and is in the time domain is divided into small chunks that generally overlap, and the short-time Fourier transform (STFT) methodology is applied for the calculation of the frequency spectrum's magnitude for each chunk.

**Fig. 13** Scalogram of milling force ( $x$ -direction) signals for (a) grade-1 worn tool, (b) grade-2 worn tool, (c) grade-3 worn tool, (d) grade-4 worn tool, (e) grade-5 worn tool, and (f) normal tool



Thus, each chunk is represented a specific color pattern on a corresponding time segment in the spectrogram image. Time–frequency analysis is done by using the STFT. It is used to construct representations that capture both the signal's local time and frequency information. STFT visualization is frequently achieved using its spectrogram, which is the intensity plot of the magnitude of STFT over time. The force signals of drilling and milling operations are converted into the images of the spectrogram represented in Fig. 10. By using the STFT methodology, the images of spectrograms are the visual representation of STFT. The color patterns visually represent the strength of signal or power over time at different frequencies contained in a certain waveform.

Each tool used in the drilling and milling operation represents a different spectrogram image. Figure 10 shows the spectrogram for drilling worn tools. Figure 10a has displayed the spectrogram of the grade-1 worn tool which possesses mostly blue color patterns which represent low power because the force value in the grade-1 tool is 350N which is low compared to other tools graphically represented in Fig. 10a. The value of force increases for higher-grade worn tools, so the power of the spectrogram also increases which is directly related to the force value. The blue color patterns displayed in Fig. 10b are narrowed, whereas the yellow and orange color patterns are widened; thus, yellow and orange color symbolize a higher level of

power. The spectrogram has shown this variation because the force value is increased to 450N for grade-2 worn tool, as graphically represented in Fig. 10b. The same behavior is continued for the spectrograms of the grade-3 worn tool displayed in Fig. 10c, where the dark blue color has disappeared completely and was replaced by the combination of light green and light blue colors, and the force value for the grade-3 worn tool is 600N, as graphically demonstrated in Fig. 10c. In the spectrogram of a grade-4 worn tool, the blue color has completely vanished and small strips of green color have appeared, as displayed in Fig. 10d, and the value of force for the grade-4 worn tool has increased to 750N given in Fig. 10d. The spectrogram of the grade-5 worn tool has shown the highest power because it contained mostly orange and yellow colors graphically illustrated in Fig. 10e. However, the force for grade-5 worn tool has reached the maximum value of 925N, as shown in Fig. 10e. Figure 10f has demonstrated the spectrogram image of the normal tool's force signals.

The figure of the normal tool contains very few strips of yellow color as it is mostly composed of light blue color. The illustration of the low level of force value is clear in this diagram due to little amount of yellow and orange color strips compared to the spectrograms of other worn tools. A similar pattern is followed in the spectrogram of the milling normal tool, as graphically represented in Fig. 11f, which comprises mostly light blue and dark blue-colored strips and has only little amount of yellow-colored strips compared to the spectrograms of other worn tools. The same behavior of different colors represents different power levels and force values for milling tools having tool wear of various grades. The force of machining tools is also directly related to the power levels of spectrograms.

Mainly all colors discussed above in the color scheme are present in this spectrogram. The force value of this machining worn tool is 175N, as displayed in Fig. 9a. The dark blue color represented by the spectrogram of Fig. 11b and the force have increased to 350N for grade-2 worn tools. If Fig. 11c,d are considered, both spectrograms are mainly composed of light green-colored strips having very few blue-colored strips, but the yellow and orange section in the spectrogram of the grade-4 worn tool is widened which indicates the increment as the force values are also increased which are 450N for grade-3 worn tool and 625N for grade-4 worn tool and are represented graphically in Fig. 9c,d, respectively. The graphical illustration of a spectrogram of the grade-5 worn tool is different from all other spectrograms as it is mostly composed of yellow and orange colors which depict the high power levels. The main reason behind this pattern is of a high force value of 750N. The grade-5 worn tool was represented graphically in Fig. 9e.

### 3.8 Conversion of raw force signals data of machining processes into scalogram images

Scalogram is a 2D image plot as the function of frequency and time. This image is obtained by using the absolute value of the continuous wavelet transform (CWT) of the force signals data collected by the machining processes. To get the CWT, the force signal is windowed with a wavelet that has scaled and also shifted in time. The wavelet performs oscillation and may be complex-valued. The processes of shifting and scaling are applied to wavelets. The wavelet stretches and shrinks by the scaling operation which yields the high frequency and short-duration wavelets that have a good capability of detecting transient events, whereas the stretching operation yields low-frequency as well as long-duration wavelets which are capable of isolating low-frequency events. The algorithm based on continuous wavelet transform is designed to convert the time-series force signals obtained by the dynamometer into the images of a scalogram. CWT was used to build the frequency-time relationship of the force signal. The absolute value of coefficients of continuous wavelet transform (CWT) was plotted as the function of frequency and time. The scalograms given in Fig. 12. are generated using the drilling force signals obtained from the dynamometer.

Every scalogram has different and unique color patterns which represent the magnitude of the different levels of the scalogram, and another major factor that differentiates each scalogram from others is peak magnitude. There were 12 tools, so a total of 12 scalograms were plotted which are shown in the figures given below. Each tool has its individual and unique scalogram image. In the case of time-series force signals conversion to scalogram, the value of peak magnitude is directly proportional to the value of force applied by each drilling and milling tool on the workpiece. Peak magnitude is the major factor that can distinguish one scalogram from another. If the magnitude of the scalogram was higher, it would mean that a higher force was applied by the tool on the surface of the workpiece. The force applied by drilling the grade-1 worn tool on the workpiece is 350N, so the peak magnitude of the scalogram of the grade-1 worn tool is 75. As the tool wear progresses, the force applied to the workpiece by the worn tools also increases significantly.

The force values illustrated graphically in Fig. 8 of the drilling grade-2 worn tool, drilling grade-3 worn tool, drilling grade-4 worn tool, and drilling grade-5 worn tool are 450N, 600N, 750N, and 925N, respectively. So, the peak magnitude values of the scalogram of these drilling tools are 98, 110, 150, and 195, respectively, as displayed in Fig. 12. The same behavior is followed for milling worn tools. The force values applied on the workpiece by grade-1 worn tool, grade-2 worn tool, grade-3 worn tool, grade-4 worn tool, and grade-5 worn tool are 175N, 350N, 450N, 625N, and 750N,

respectively, as displayed in Fig. 8. Whereas the peak magnitude values of scalograms of these milling tools are 50, 89, 115, 180, and 300, respectively, as graphically illustrated in Fig. 13. All the obtained scalogram images were used to train the pre-trained NNs.

### 3.9 Justification for the application of image augmentation to 12-way 2-shot learning model

Image augmentation is typically needed to improve the efficiency of deep networks. The algorithm is designed and applied separately to all 24 images one by one, and 1 original image will artificially produce 9 more images. Thus, a total of 24 images will produce a total of 216 new images to train the model. An example of application to one image is given in Fig. 14.

The designed model requires a reasonable amount of training data to perform well, so image augmentation is needed which artificially produces training images using many imaging methods, such as rotation of the image, shear, change brightness shift, adding noise, changing colors, and flips. Image data augmentation is a method for making the changed copies of the images in the dataset. The most well-known method of data augmentation is image data augmentation, which includes transforming the image of the training dataset into transformed copies by using an algorithm that will be included in the same class as the original image. Image augmentation improves the model's ability to generalize the information they have learned from the new images. The greater volume of data available also increases the efficiency of deep learning ANN. When training an ML algorithm, it is essentially a process of tuning the parameters so that it can map a specific input (like an image) to a specific output (such as a label). Our optimization aim is to find the sweet spot where our model loss is the least, which occurs when parameters are calibrated correctly. Introducing the various changes to the original image resulted in producing a lot of new images. Flips, translations, and rotations are examples of minor shifts. In either case, the ANN will recognize these as distinct images.

Invariance is a property of a ResNet-18 that allows it to reliably distinguish objects even though they are positioned in different orientations. The pre-trained ANN has to extract feature vectors from the images. In short, the 12-way 2-shot model would be trained using combined 240 images of drilling as well as milling operations having 12 classes. Pre-trained CNN may learn features that are independent of the location where they appear in the picture. However, augmentation can help with this transform invariant approach to learning by assisting the model in feature learning aspects that are transformed invariant, such as left-and-right to top-to-bottom ordering and light levels in images.

### 3.10 Validation for the application of L2 regularization to 12-way 2-shot learning model

Overfitting is a major concern in an ML model when the algorithm picks the random noise and fluctuations of the training data and learns them as a concept. The model does not generalize well on the testing data, but it performs very well on the training data. L2 regularization is a crucial principle that helps to prevent overfitting. Employing the algorithm resulted in getting the lower variance associated with the training data, as well as by compressing the coefficients of predictor variables over the output variable. L2 regularization involves retaining the same number of features while decreasing the magnitude of the coefficients.

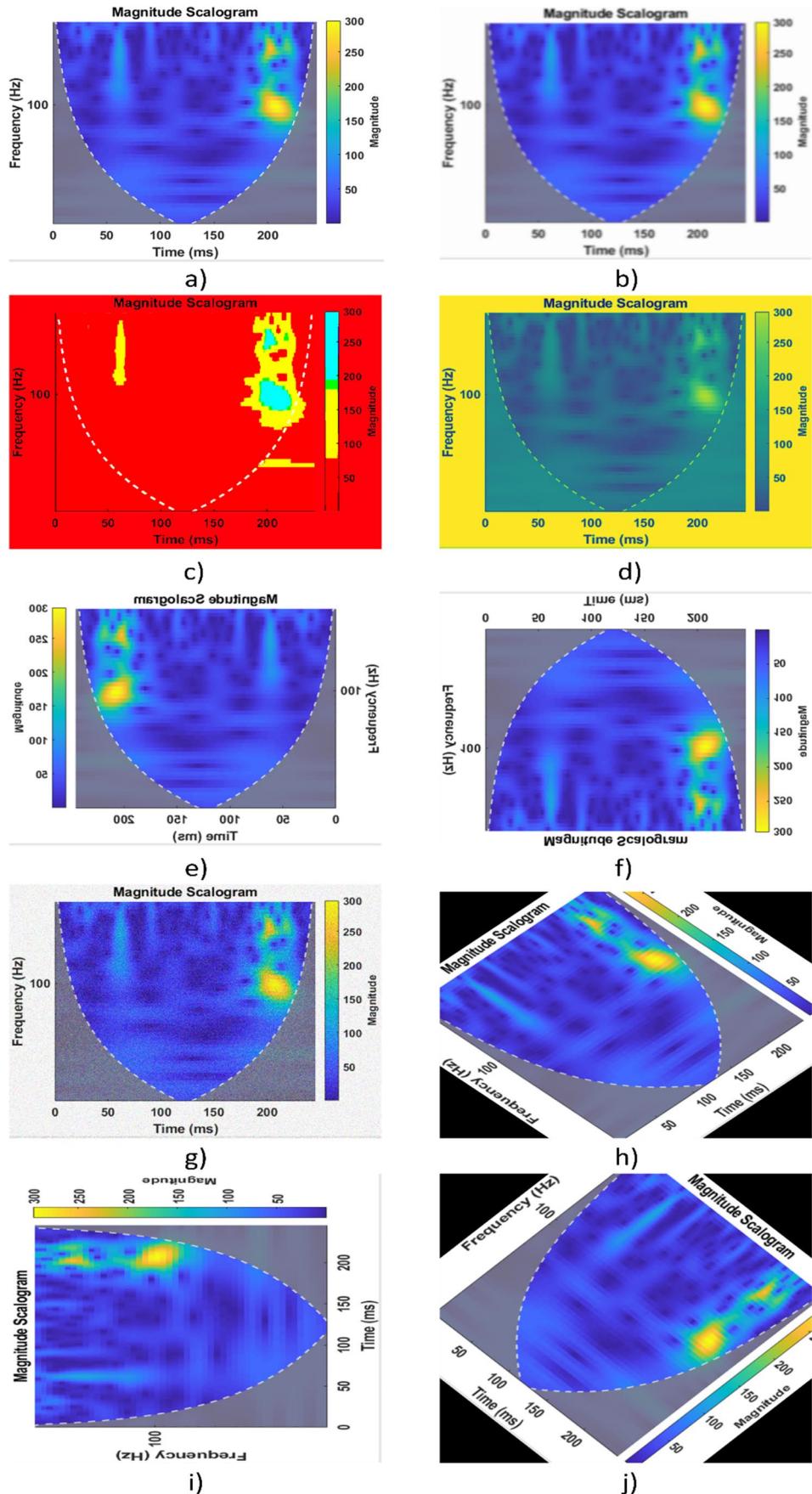
The key goal of L2 regularization is to reduce variance, sometimes at the expense of increasing bias. In this case, the model essentially learns every data point and fails to estimate accurately when evaluated on a new dataset. L2 regularization assigns minimal weight to the unimportant features. Whereas L2 regularization tends to reduce coefficients evenly, and it attempts to estimate the data's mean to avoid overfitting. L2 regularization adds the squared value of weights to the cost function. The parameter of L2 regularization was optimized by using the grid-search methodology.

$$L2 = \text{Sum}((V1 - V2) \wedge 2 + \alpha * (s1)) \quad (3)$$

where actual values and predicted values are denoted by V1 and V2. However, the sum of the absolute value of coefficients has the symbolic representation of s1.

The aim is to approximate a value that is in the middle of the data distribution to mitigate the loss function. In mathematical terms, the value would also be the median value of the data distribution. When using L2 regularization, the loss function is calculated in the gradient estimation process and attempts to minimize the loss by subtraction from the average value of the data distribution. Complexity is equal to the sum of squares of weight. The final accuracy of the 12-way 2-shot learning model is 87.83% after the application of L2 regularization. GridSearchCV is a method of tuning hyper-parameters to find the best values for a particular ML model. GridSearchCV examines the ML model for every variation of the values supplied in the training algorithm by using the cross-validation methodology. As a result of utilizing this method, we can calculate the accuracy for each combination of hyper-parameters and select the one with the greatest performance. In this research study, grid search was used to optimize the hyper-parameters of the L2-regularization algorithm. We employed the L2 regularization with these 8 values [0.00000001, 0.00000001, 0.0000001, 0.0001, 0.0001, 0.001, 0.01, 0.1], where refit=true, verbose=3, n\_jobs=-1. It is further stated that grid search

**Fig. 14** Image augmentation illustration: (a) original example image; (b) blurred image; (c) colored image; (d) RGB2gray default image; (e) horizontal flip; (f) vertical flip; (g) noised image; (h) rotated –45°; (i) rotated –90°; (j) rotated +45°



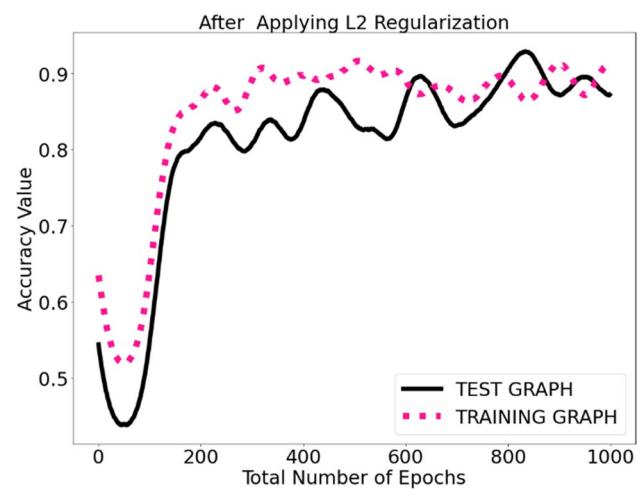
is the algorithm which has used the above parameters. For example, the GridSearchCV has chosen the value 0.0001 for the algorithm. It will evaluate the performance of L2-regularization with the chosen value, and in the same way, it will choose all the above values one by one and evaluate the performance; thus, it will ultimately select the final value with the best accuracy. The main quality of the 12-way 2-shot learning algorithm is to solve the classification problem and to make predictions by using very small data samples. The basic working of the model is based on 2 images in the support set and 20 images in total for each class. The two types of conversions for the time-series data into images have been employed. For the first type, the time-series data is converted into scalogram images, whereas for the second type, the time-series force signals have converted into spectrogram images. As a result, a total of 24 images are obtained in the primary phase for 12 milling and drilling tools (each tool will generate two images, i.e., spectrogram and scalogram), and the image augmentation algorithm has tremendously increased the image data bank to 240 images to input into the 12-way 2-shot learning algorithm for training because the training data include the 24 original images of the support set plus 216 images obtained by the image augmentation methodology.

#### 4 Comparison of model's results with existing literature's algorithms based on common force signals data

Figure 16 is the testing results based on the final accuracy percentage of previously published ML models and the current model by using the common training and testing signals. The developed model of this research study based on 12-way 2-shot learning is compared with the major ML models of previously published research studies. The comparison is built based on the principal results of the current model and the major contributions of this research study. All the major ML algorithms shown in Fig. 16 were trained and tested by the common force signals data obtained by the dynamometer after performing milling and drilling operations. CNN algorithm works better on image data, so for this purpose, the same image dataset based on spectrogram and scalogram was used, but the number of images was increased by applying an image augmentation algorithm as the CNN algorithm requires a large number of images per class for training. The major contribution of this research study is that it has overcome the shortcoming of ML models of previous research studies for the prediction of tool wear. The experimental research study has overcome three major shortcomings of the previously published ML models for tool wear monitoring, i.e., retraining and remodifying of ANN during tool wear prediction, high dimensionality, greater computational

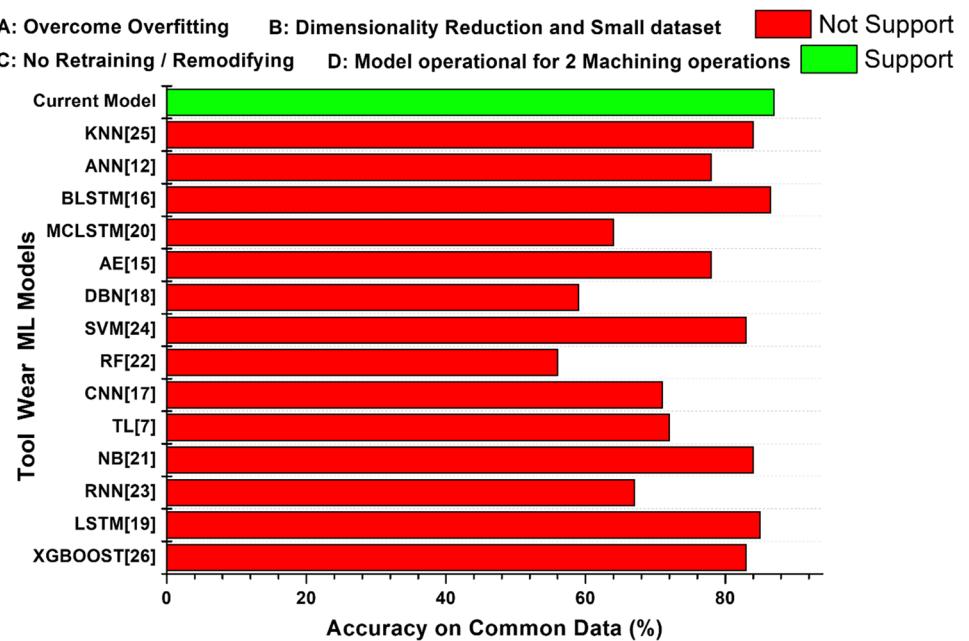
costs, and overfitting. This model will provide feedback to CNC machines in the manufacturing and aerospace industries about the tool wear and will also determine the grade of tool wear (ranging from grade-1 to grade-5). If the determined grade of the worn tool is grade-4 or grade-5, then the tool must be replaced by a new tool; otherwise, it will destroy the workpiece. However, in the case of the grade-1 worn tool, the tool can be used for some time by taking precautionary measures to avoid further tool wear progression. The model will boost the production rate and enhance the dimensional accuracy of the machined components in the industry. We hereby confirm that this manuscript is our original work and has not been published, nor has it been submitted simultaneously elsewhere. The four important parameters demonstrated in Fig. 16 are not supported by previously published ML models and are neglected. These parameters have an extremely high impact on the performance and efficiency of ML models for tool wear prediction. Overfitting is a common problem in ML models that occurs when the algorithm selects fluctuations in the training data and learns them as concepts. L2 regularization is an important methodology that aids in the prevention of overfitting. The L2-regularization methodology is applied during the training of the 12-way 2-shot learning model. The training and testing graph after the application of L2-regularization is shown in Fig. 15. L2 regularization has handled multicollinearity issues by restricting the coefficient and maintaining all variables constant. L2 regularization may be used to evaluate the relevance of predictors and penalize unimportant predictors accordingly.

The second most important parameter of the contribution of this research study is that the designed model is operational on a small dataset. Due to this quality, the model is highly capable of working on a real-time dataset during real-time milling and drilling operations. The previous research



**Fig. 15** Graphical illustration of the final accuracy of the model

**Fig. 16** Previous ML models supported or not supported with the novel parameters of this study based on grade-5 worn drill tool test force signals



studies lack this useful quality of the model. Based on the above-mentioned comparison results, all the models were trained on all 12 tools' force signals obtained by dynamometer, except the CNN model and the current model of this research article, as both were trained and tested by using an image dataset. The model works based on similarity score by extracting and comparing the features of the query image and each image of the support set, as shown in Fig. 1. The model's operation is based on a meta-learning methodology. When the test signals of the grade-5 drill worn tool were inputted into the trained models, then all models have shown different prediction results for the detection of the G-5 drill worn tool which are shown in the form of prediction accuracies in the above-mentioned comparison chart (Fig. 16).

The scalogram image of the G-5 drill worn tool, as shown in Fig. 12e, was inputted into the currently trained model for testing purposes; the model has worked in such a way that the scalogram image of the G-5 drill worn tool was selected as the query image and spectrogram as well as scalogram images of all other worn/normal tools were put into the support set. The 12-way 2-shot learning algorithm has learned the features of the test scalogram image of the G-5 drill worn tool, as well as all other images present in the support set; then, the algorithm has matched the features of the test image with all images present in the support set. The similarity scores of all the matches of all images were calculated. The similarity score of the scalogram image of the G-5 drill worn tool was the highest, so the algorithm would output the label of the G-5 drill worn tool class out of 12 classes as the final output results. The third attribute of the 12-way 2-shot learning model which distinguished it from all other previously published models is that there is no retraining and

remodifying of ANN in the current model. When the new image is inputted into the traditional ML models, as shown in the above comparison chart, all will undergo a retraining and remodifying process of the ANN which remarkably has an adverse impact on the efficiency and performance of the model, especially on working during the real-time machining operation. So, the current model has effectively overcome the shortcoming of traditional ML models. The fourth quality of the current research study is that the designed model is operational for two machining processes, i.e., drilling and milling. The data of both operations have been put in one support set so that model can work efficiently for both processes during the real-time machining operation.

## 5 Conclusion

The tool wear severity estimation system is developed based on a 12-way 2-shot learning model. The major contribution of this study is that it has overcome the three major limitations of previously published tool wear prediction models, i.e., remodifying and retraining of ANN during tool wear estimation, high computational costs of tool wear models, high dimensionality, and overfitting. The working capability of each model for two machining operations for tool wear estimation based on more levels of tool wear severity has also been neglected by already published literature. The major application of the designed models is the potential installation in the aerospace and manufacturing sectors involving the milling and drilling operations. This system will boost the production rate along with the dimensional accuracy as the worn tools negatively influence the machined surface

integrity. This system will provide feedback about a worn tool to the CNC machine, and the grade of the tool wear will also be determined. The test tool will be considered replaceable if the system will notify it as grade-4 or grade-5; otherwise, it will impart adverse effects. However, in the case of the grade-1 worn tool, the tool can be used for some time by taking precautionary measures to avoid further tool wear progression. The final results have proven the acceptability of the model with 87.83% performance accuracy.

**Acknowledgements** All authors are extremely thankful to Professor Ming Luo of Northwestern Polytechnical University for his kind contribution regarding novel ideas of experimentation as well as for supervision.

**Author contribution** J. Mahmood: conceptualization, writing – original draft, investigation, methodology, formal analysis, software, visualization, writing – review and editing. M. Adil Raja: software, project administration, investigation, supervision, software, methodology. M. Rehman: writing – review and editing, methodology, formal analysis. J. Loane: writing – review and editing, methodology, formal analysis. S. Zahoor: writing – review and editing, methodology, formal analysis.

**Funding** This research was funded by the Technological University Transfer Fund (TUTF) of the Higher Education Authority (HEA), Ireland.

**Availability of data** Complete data based on the force signals in Excel files are available which are already plotted in Figs. 8 and 9 in graphical representation.

**Availability of code** The separate flow chart is designed for the explanation of the code given in Fig. 3.

## Declarations

**Ethics approval** The research had been conducted following the rules of ethics.

**Consent to participate** All authors are willing to participate in the review of IJAMT articles.

**Consent for publication** All authors are willing to publish the manuscript of this research study.

**Conflict of interest** The authors declare no competing interests.

## References

- Shankar S, Mohanraj T, Rajasekar R (2019) Prediction of cutting tool wear during milling process using artificial intelligence methodology. *Int J Comput Integr Manuf* 32:174–182. <https://doi.org/10.1080/0951192x.2018.1550681>
- Chen G, Li Y, Liu X (2019) Pose-dependent tool tip dynamics prediction using transfer learning. *Int J Mach Tools Manuf* 137:30–41. <https://doi.org/10.1016/j.ijmachtools.2018.10.003>
- Kurek J, Wieczorek G, Kruk BSM, Jegorowa A, Osowski S (2017) Transfer learning in recognition of drill wear using convolutional NN. 2017 18th International Conference on Computational Problems of Electrical Engineering (CPEE): IEEE; p 1–4. <https://doi.org/10.1109/cpee.2017.8093087>
- Sun C, Ma M, Zhao Z, Tian S, Yan R, Chen X (2018) Deep transfer learning based on sparse autoencoder for remaining useful life prediction of the tool in manufacturing. *IEEE Trans Industr Inf* 15:2416–2425. <https://doi.org/10.1109/tii.2018.2881543>
- Kim Y-M, Shin S-J, Cho H-W (2021) Predictive modeling for machining power based on multi-source transfer learning in metal cutting. *Int J Precis Eng Manuf-Green Technol* 1–19. <https://doi.org/10.1007/s40684-021-00327-6>
- Huang S, Guo Y, Liu D, Zha S, Fang W (2019) A two-stage transfer learning-based deep learning approach for production progress prediction in IoT-enabled manufacturing. *IEEE Internet Things J* 6:10627–10638. <https://doi.org/10.1109/iotj.2019.2940131>
- Li J, Lu J, Chen C, Ma J, Liao X (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>
- Lee J, Kim D, Lee S (1996) Application of NNs to flank wear prediction. *Mech Syst Signal Process* 10:265–276. <https://doi.org/10.1006/mssp.1996.0020>
- Ghosh N, Ravi Y, Patra A, Mukhopadhyay S, Paul S, Mohanty A et al (2007) Estimation of tool wear during CNC milling using NN-based sensor fusion. *Mech Syst Signal Process* 21:466–479. <https://doi.org/10.1016/j.ymssp.2005.10.010>
- Kandilli I, Sonmez M, Ertunc HM, Cakir B (2007) Online monitoring of tool wear in drilling and milling by multi-sensor NN fusion. 2007 International Conference on Mechatronics and Automation: IEEE; p. 1388–94. <https://doi.org/10.1109/icma.2007.4303752>
- D'Addona D, Segreto T, Simeone A, Teti R (2011) NN tool wear modelling in the machining of nickel superalloy industrial products. *CIRP J Manuf Sci Technol* 4:33–37. <https://doi.org/10.1016/j.cirpj.2011.07.003>
- Palanisamy P, Rajendran I, Shanmugasundaram S (2008) Prediction of tool wear using regression and NN 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>
- Youneszadeh H, Ardeshir A, Sebt MH (2020) Predicting project success in residential building projects (RBPs) using artificial NNs (NNs). *Civil Eng J* 6:2203–19. <https://doi.org/10.28991/cej-2020-03091612>
- Hao G, Kunpeng Z (2020) Pyramid LSTM auto-encoder for tool wear monitoring. 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE): IEEE; p 190–5. <https://doi.org/10.1109/case48305.2020.9217015>
- Huang C-G, Yin X, Huang H-Z, Li Y-F (2019) An enhanced deep learning-based fusion prognostic method for RUL prediction. *IEEE Trans Reliab* 69:1097–1109. <https://doi.org/10.1109/tr.2019.2948705>
- Xu X, Wang J, Ming W, Chen M, An Q (2021) In-process tap tool wear monitoring and prediction using a novel model based on deep learning. *Int J Adv Manuf Technol* 112:453–466. <https://doi.org/10.1007/s00170-020-06354-y>
- 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>
- Zhang Z, Lu J, Zhou G, Liao X (2018) Research on tool wear prediction based on LSTM and ARIMA. Proceedings of the 2018 international conference on big data engineering and technology p 73–7. <https://doi.org/10.1145/3297730.3297732>
- 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>
- Karandikar J, McLeay T, Turner S, Schmitz T (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>

22. Wu D, Jennings C, Terpenny J, Gao RX, Kumara S (2017) A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests. *J Manuf Sci Eng* 139. <https://doi.org/10.1115/1.4036350>
23. Wang X, Huang Y (2009) Optimized recurrent NN-based tool wear modeling in hard turning. *Trans NAMRI/SME* 37:213–220
24. Qian Y, Tian J, Liu L, Zhang Y, Chen Y (2010) A tool wear predictive model based on SVM. 2010 Chinese control and decision conference: IEEE; p 1213–7. <https://doi.org/10.1109/ccdc.2010.5498161>
25. Penedo F, Haber RE, Gajate A, del Toro RM (2012) Hybrid incremental modeling based on least squares and fuzzy K \$-NN for monitoring tool wear in turning processes. *IEEE Trans Industr Inf* 8:811–818. <https://doi.org/10.1109/tii.2012.2205699>
26. Alajmi MS, Almeshal AM (2020) Predicting the tool wear of a drilling process using novel machine learning XGBoost-SDA. *Materials* 13:4952. <https://doi.org/10.3390/ma13214952>
27. Socher R, Ganjoo M, Sridhar H, Bastani O, MN Ning CD, Ng AY (2013) Zero-shot learning through cross-modal transfer. arXiv preprint arXiv:13013666
28. Xu Y, Li Y, Wang Y, Zhong D, Zhang G (2021) Improved few-shot learning method for transformer fault diagnosis based on approximation space and belief functions. *Exp Syst Appl* 167:114105. <https://doi.org/10.1016/j.eswa.2020.114105>
29. Ou X, Yan P, Zhang Y, Tu B, Zhang G, Wu J et al (2019) Moving object detection method via ResNet-18 with encoder-decoder structure in complex scenes. *IEEE Access* 7:108152–108160. <https://doi.org/10.1109/access.2019.2931922>
30. Vinyals O, Blundell C, Lillicrap T, Wierstra D (2016) Matching networks for one-shot learning. *Adv Neural Inf Process Syst* 29:3630–3638
31. 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>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.