



# Accurate estimation of tool wear levels during milling, drilling and turning operations by designing novel hyperparameter tuned models based on LightGBM and stacking

Jawad Mahmood<sup>\*</sup>, Ghulam-e Mustafa, Muhammad Ali

School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, PR China

## ARTICLE INFO

**Keywords:**  
Tool wear  
LightGBM  
Drilling  
Stacking  
Milling  
Turning

## ABSTRACT

The study is focused on executing machining operations by using 18 tools of milling, drilling, and turning thus 6 tools were used for each machining process. The VB of tool was measured for categorizing the tools ranging from level-1 to level-5 based on the severity of tool wear. The first model was designed based on LightGBM whereas the second model was developed by designing six algorithms i.e. LR, RF, CART, NB, SVM, and KNN. All algorithms were combined to develop an ensemble stacking model. Manual hyperparameter tuning was done for the LightGBM model whereas automatic hyperparameter tuning was adopted for the Stacking model by using GridSearchCV. The force signals' features extraction was done by SSA whereas dimensionality reduction was accomplished by PCA. One-hot encoding technique has transformed target variables into binary form. The application of techniques of dropout and early stopping to both models has overcome overfitting.

## 1. Introduction

The manufacturing sector has undergone many advancements throughout time due to the integration of information technology and is continually advancing toward the smart industry. Still, industries have problems in improving productivity, maintaining quality, and lowering overhead expenses. The major issue with drilling as well as other machining processes such as turning and milling is tool wear. Tool wear can occur while conducting machining operations on a hardened workpiece such as nickel-based alloys due to its high strength and hardness, which might also result in higher cutting forces, complicated kinematics of the process, and elevated heat throughout machining. Tool wear is impacted by a lot of physical factors, including the hardness of materials, feed rate, the geometry of the cutting tool, kind of coolant used, depth of cut, and cutting speed. [Table 1](#).

Marani et al. [1] investigated the expected cutting force modeling of a turning process for Al-Si-Cu cast alloys that have been modified using modifiers. The input variables were cutting speed, feed rate, and silicon spacing. Several turning tests were carried out at different cutting speeds and feed rates. The Neuro-fuzzy based model accurately estimated the cutting force in terms of speed Si spacing and feed rate, according to the predicted result. Marani et al. [2] have presented the

model-based solution of tool wear monitoring relying on an adaptive neuro-fuzzy inference system (ANFIS) for only a hardened steel bar by performing 1215 turning operations. Using the ANFIS technique, a 3-input cutting force, as well as a 1-output model, was developed and executed. A dynamometer and data collecting systems were used to evaluate the forces. A microscope was used to assess flank wear. The findings of the model estimation reveal that it is efficient enough to undertake real-time monitoring of the turning operation and can identify wear while functioning. Marani et al. [3] have worked on several adaptive network-based fuzzy inference systems (ANFISs) which were used to assess the influence of feed rate, particle size, and cutting speed on the capability of machining of an Al-20 Mg2Si matrix composites. The most exact models for estimating cutting force and surface roughness were 2 ANFIS models. The final results demonstrate that the ANFIS model can estimate the machining capability of metal matrix composites with sufficient accuracy. Marani et al. [4] investigated the fuzzy logic AI approach for forecasting the machinability of Al-Si-Cu-Fe permanent mold alloys handled with several additives to reduce surface roughness, comprising strontium, antimony, and bismuth. The Pareto-ANOVA optimizing approach was utilized to determine the best machine parameter settings. Tests were conducted with the help of CNC turning. Surface roughness data were used to maximize the parameters

\* Corresponding author.

E-mail address: [jm@mail.nwp.edu.cn](mailto:jm@mail.nwp.edu.cn) (J. Mahmood).

**Table 1**  
Nomenclature.

LR	Logistic Regression	RF	Random Forest
CART	Classification and Regression Tree	NB	Naïve Bayes
RNN	Recurrent Neural Network	ML	Machine Learning
SSA	Singular Spectrum Analysis	VB	Flank wear land width
RF	Random Forest	ANN	Artificial Neural Networks
PCA	Principal Component Analysis	SVM	Support Vector Machine
LSTM	Long Short Term Memory Network	CNN	Convolutional Neural Network
BLSTM	Bidirectional Long Short Term Memory Network	DBN	Deep Belief Network
MCLSTM	Convolutional Bi-directional LSTM	KNN	k-nearest neighbors algorithm
XGBOOST	Extreme Gradient Boosting	MCLSTM	Multiscale Convolutional LSTM
AE	Auto Encoder	TL	Transfer Learning
MCGPR	Multi-Covariance Gaussian process regression	Adaboost-DT	Adaboost Decision Tree
SBULSTM	Stacked bidirectional and unidirectional LSTM	SVR	Support Vector Regression

of cutting speed, depth of cut, and feed rate, the results also showed that Sr and Sb had a detrimental influence upon workpiece machinability. see Fig. 1.

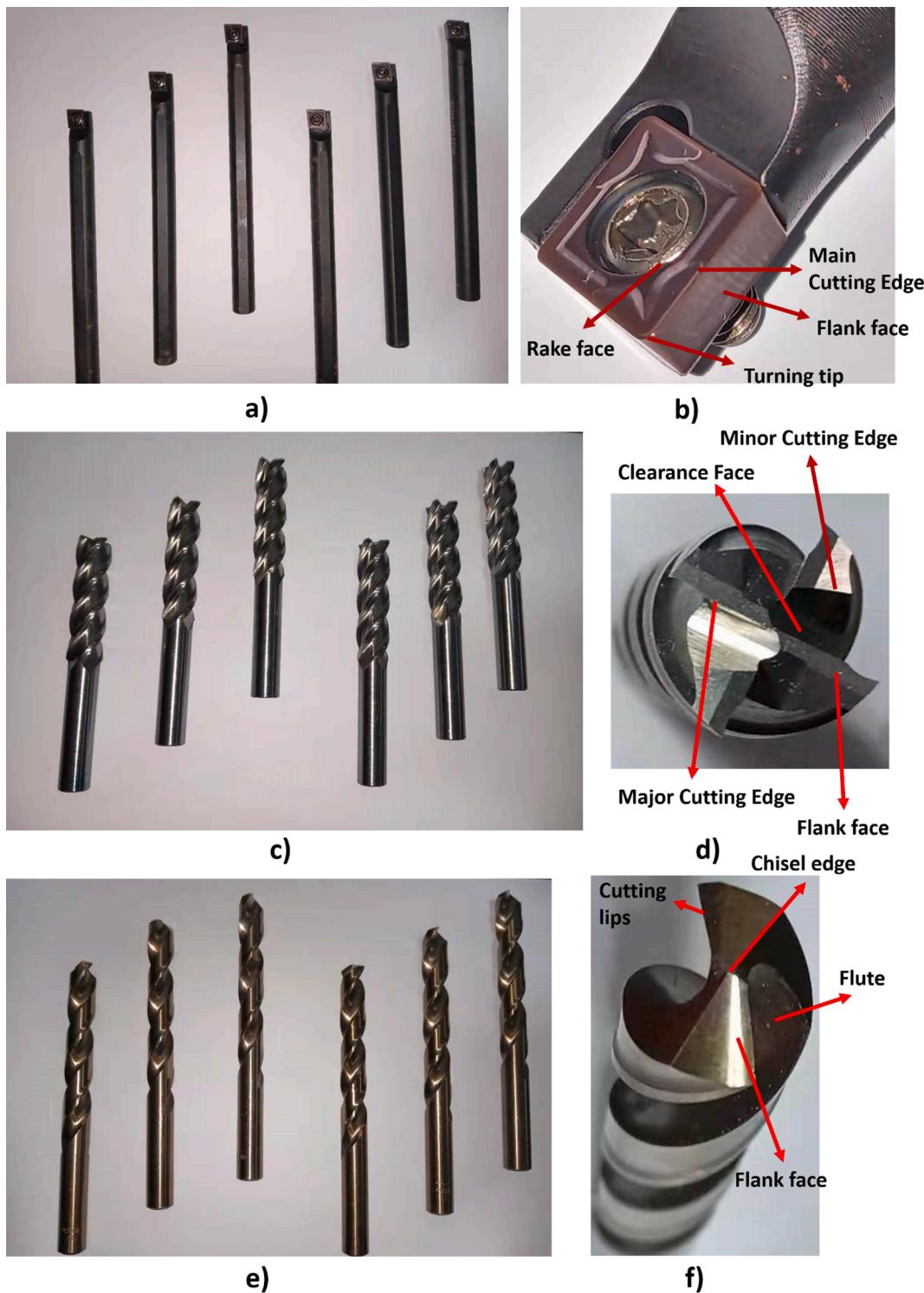
Zegarra et al. [5] have presented the current study which attempts to demonstrate that, in the wear rate prediction issue, classical approaches may get results comparable to the top of the line, as acquired utilizing deep learning methods. The data supplied here is in the terms of time series, which cannot be readily utilized by typical machine learning methods like the ones employed in this study. To connect the raw data as well as the prediction model, a collection of key features using the time series must first be extracted. Furthermore, several preprocessing strategies, like Bayesian hyperparameter optimization. Chacon et al. [6] proposed the work which provides a technique for extracting redundancy as well as non-optimal extracted features from AE signals utilizing multi-threshold counting extraction of features at the multi-resolution level employing wavelet packet transformation. The enormous number of predicted features obtained in the preceding stage is then reduced and optimized via recursive feature reduction, as well as the regression of random forests yields the predicted tool wear. Gittler et al. [7] published the paper which provides an unsupervised learning technique for machine tool element and consumable deterioration prognostics. It employs a multi-sensor-based time series of data that is translated into a data augmentation. The features are distinct characteristics of the time series that allow multiple signal assessments to be compared and clustered based on respective feature values. The density-based cluster approach developed here is being used to evaluate and forecast the deterioration statuses of parts and components in unknown settings. Twardowski et al. [8] have presented the work is about evaluating the edge state using acoustic emission data. The wear breadth on the flank face defined the overall tool geometry condition. The treated component was aluminum-ceramic alloy with 10% Sic. The tool was a tungsten mill cutter having a diamond covering. Appropriate measurements were formed based on AE signals and were associated with edge states. Depending on Signals, machine learning techniques were employed to determine the milling tool wear. Parwal et al. [9] have published the study which proposes a process supervision system based on Machine Learning that predicts tool wear. An implementation for predicting tool wear while milling is used as an instance study to demonstrate the method. The research was carried out on a manufacturing dataset comprised of many experiments with varying quantities of model parameters. Zhaopeng et al. [10] have designed the methodology for which temperature signals were acquired by an adaptive cutting tool

equipped with a slender thermocouple. Auto encoders architecture with a neural network based on backpropagation for prediction was developed as a deep learning technique to estimate wear rate based on temperature data. A revised error rate including sparse as well as weight penalty factors were employed to increase the sturdiness and generalization of the stacked sparse AE system.

The k-nearest neighbors (KNN) technique is a straightforward supervised machine learning approach that may be used to tackle classification and regression tasks. Horton and Nakai [11] have predicted the localization sites of protein cellular by using the KNN classifier. The results of stratified cross-validation have revealed that the KNN has outperformed the other techniques. Penedo et al. [12] have done hybrid incremental modeling by using KNN to monitor the tool wear during turning operation. The methodology is a two-step iterative approach that blends a global model with the developed local model to capitalize on its underlying, complementary capabilities. Shi et al. [13] have applied Reinforced KNN to identify chatter during the milling operation. They carried out several tests on a CNC milling machine with various types of sensors during high-speed milling operation, where chatter is common. Statistical techniques are used to compare signals from various sensors and extract features. Islam et al. [14] had worked on the model based on the KNN classifier for the detection of common disease symptoms. They have collected the data from the social network to input into the KNN model. Zhang [15] has worked for the detailed introduction of the architecture and basic working of KNN algorithm and then focused on the mechanism of KNN modeling to R. Zhang and Zhou [16] have developed KNN model for the multi-label classification. Experimental results on real multi-label bioinformatics data reveal that Multi labeled kNN is quite similar to other multi-label learning algorithms. The k-nearest neighbor results in saving all existing data and categorizes new data points based on the measure of similarity (e.g., distance functions). This refers to when updated data becomes available. Then, using the KNN method, it may be readily categorized into a suitable category.

Logistic regression is a statistical model that models conditional probability using the Logistic function. This is also known as conditional probability. Li [17] has designed a logistic regression methodology for the estimation of the cutting tool's operational reliability. Tools are evaluated based on several conditions and estimations to establish the optimal maintenance time. Acoustic Emission and cutting force signals show a stronger linear connection with tool wear, according to experimental analysis. Cheng and Hullermeier [18] have used logistic regression for multi-labeled classification. They developed a unique approach for multilabel classification in this study, which has built on a framework that integrates instance-based learning with logistic regression, with both techniques serving as special cases. Tomioka et al. [19] have classified the EEG signals by using the Logistic Regression Algorithm. The regression function is parameterized in different ways in the paper. Chen et al. [20] have estimated the reliability of the cutting tool based on the logistic regression model by using vibration signals. The method is validated on an NC lathe. The reliability and failure time of the cutting tools are all properly evaluated under different failure thresholds. Kurt et al. [21] have compared the performance of the Logistic regression algorithm with other neural networks for the prediction of coronary artery disease. A retrospective study of 1245 individuals was carried out. The data was divided in such a way that there were 865 cases of coronary artery disease and 380 cases of no coronary artery disease. Subasi and Ercelebi [22] have classified EEG signals by using the Logistic Regression algorithm. This study describes a unique technique for analyzing EEG data using wavelet transforms and classifying them using artificial neural networks (ANN) and logistic regression (LR). The logistic regression is a classification model that also provides a probability. This gives it a significant edge over other models that can simply deliver the final classification.

LightGBM framework is efficient and was created with distributed training in consideration. It enables large amounts of data and GPU



**Fig. 1.** Demonstration of a) Turning tools b) Turning tool's tip c) Milling tools d) Milling tool's tip e) Drilling tools f) Drilling tool's tip.

training. LightGBM has been proven to be more accurate and quicker than XGBoost in all situations. Wang et al. [23] investigated the monitoring of quantitative wear of abrasive belts considering material removal rate by using the LightGBM algorithm. The experimental results reveal that the assessment parameters of the material removal rate estimation technique are all within 5%. Furthermore, the precision of the abrasive belt's wear levels might surpass 91 percent. Minastireanu and Mesnita [24] have used the LightGBM algorithm to detect online click fraud. They analyzed click patterns on a dataset including 200

million clicks. The major aim was to evaluate a user's click path across their portfolio and identify IP addresses that generate a high volume of clicks. Tang et al. [25] have designed the LightGBM model to detect faults of wind turbine gearboxes. The implementation of selecting features for fault identification is initially accomplished by using the information gain coefficient to evaluate the correlation between features in wind turbine supervisory control and data acquisition (SCADA). Wang et al. [26] have worked on material removal rate monitoring for abrasive belt grinding utilizing an enhanced LightGBM algorithm based

on vision and sound fusion. Grinding tests were carried out using various grinding parameters, and visual and sound signals were recorded using industrial CCD cameras and microphones. LightGBM uses a unique approach known as Gradient-based One-Side Sampling (GOSS) for filtering out data instances to determine a split value, whereas XGBoost employs a pre-sorted algorithm and a Histogram-based algorithm to compute the optimum split.

Random forest is a supervised learning method. The “forest” it creates is an ensemble of decision trees that are often trained using the “bagging” approach. Dahe et al. [27] investigated the tool’s condition monitoring based on the random forest by employing statistical learning. Under various operational situations, vibration signals were recorded using both excellent and faulty tools. The vibration signals were analyzed, and the necessary statistical data was collected and achieved 93.65% accuracy. Rodriguez-Galiano et al. [28] have used multi-seasonal images and employed a random forest classification of Mediterranean land cover. The best accuracy was obtained when the RF classifier was applied to a reduced subset of input variables consisting of the most significant textural features. Wu et al. [29] predicted the tool wear on the basis of Data-driven prognostic by employing a random forest algorithm. The goal of this study is to investigate the capability of random forests (RFs) to estimate tool wear in milling operations. An experimental dataset is used to compare the performance of ANNs, SVRs, and RFs. Alam and Vuong [30] have used RF classification for android malware detection. Their objective was to evaluate Random Forest’s accuracy in identifying Android application activity to categorize apps as malicious or benign. Gailson et al. [31] have worked on RF classification of geographical data as well as multisource remote sensing. The experiments described in the study were carried out using multisource remote sensing and geographic data collection. The results from random forests were compared to the outcomes from bagging and boosting approaches. Yu et al. [32] have presented a machine vision-based measuring approach for drill chisel edge wear. threshold segmentation based on Local variance is used to extract the whole contour of the drill. Patange et al. [33] have proposed an ML-based technique for monitoring the health of multipoint inserts. During the face milling on a machine, a time-domain vibration output of four tools was recorded for defect-free and varied problematic configurations. Kuntaglu et al. [34] have presented the experimental study in which five distinct sensors were attached to the lathe to gather data on the capabilities of each sensor in detecting tool wear. During the turning of AISI 5140 employing coated carbide tool, vibration, cutting forces, temperature and acoustic emission were measured. Fong et al. [35] have evaluated the tool wear remotely from the operating workstation so, unique quantified image-based tool wear measuring method that is based on cross correlation is developed.

The developed multi-target classification model LightGBM has overcome the major shortcomings of previously published tool wear models i.e. high dimensionality and overfitting of tool wear models as well as high memory usage and slow processing speed during estimation of tool wear in machining operation. Milling, drilling, and turning operations were combined based on the features extracted from the force signals to input into the designed models. Two models based on LightGBM and Sacking were developed but the performance of LightGBM was better for tool wear severity detection during milling, drilling, and turning operations. The paper is structured in such a way that the first section is of introduction followed the longest section comprises of experimental validation and discussion. This section is composed of experimental setup and procedure, Milling, drilling, and turning experiments by using normal and worn tools, application of various algorithms for the development of Hyperparameter optimized LightGBM and Stacking Models. The third and fourth sections are of conclusion and references of paper respectively.

## 2. Experimental discussion and validation

### 2.1. Experimental procedure

This experimentally conducted research is based on turning, milling and drilling processes performed on a dynamometer mounted on the CNC machine with a tool magazine holding 12 distinct types of drilling and milling tools, resulting in six tools for each machining operation. A total of 18 tools were used for three machining operations. Milling, drilling, and turning operations were combined based on the fifteen features which were extracted from the force signals of all three machining operations and these features were input into the designed models. The first tool with sharp edges was designated as the normal tool, while the remaining 5 worn tools were classified based on increasing flank wear land width values. Grade-1 was given to the least worn tool with the smallest value of flank wear land width, and grade level increases as the value of flank wear land width increases. The severely worn tool with the maximum flank wear land width value, on the other hand, received a wear level of 5. The main objective of this experimentally performed research is to evaluate the tool wear level of machining tools. The level of tool wear is proportional to the magnitude of flank wear land width. All tests were carried out on a 3-axis CNC machine having the spindle motor rated at 22 kW and a maximum rotational speed of the cutting tool of 1250 r/min. Inconel-718 material was used as the workpiece. The tool is equipped with a KISTLER 9257BA dynamometer for measuring force signals and the sampling rate was 10000 Hz whereas the water-based mineral oil was the coolant. The Inconel-718 was milled, turned, and drilled to acquire data for the model’s testing/training. The measuring setup was based on various devices. The dynamometer is the major force measuring device used in the experimentation it also measures torque as well as power. A dynamometer’s data collection system is an essential component. The system is normally made up of two pieces, a Commander as well as a Workstation, which are linked together through an Ethernet wire. Commander is the Windows-based computer, sends orders to the Workspace, a touch-screen machine encased in a durable industrial shell. The Workspace controls its precise load as well as throttle control mechanisms, gathers data, and transmits it to the Commander for processing, storage, and analysis.PCI-6025E version 187573 J-01 Integrated Input/Output DAQ Board from National Instruments. The 6025E has 16 analog input lines, eight of which are differential. Two analog output pathways, a 100-pin connection, and 32 digital pins lines for input and output purposes. The LabVIEW software was used for measurement purposes. The eighteen experiments were conducted so there would be eighteen repetitions. Each repetition for one tool. Six tools were used for each machining operation. Each experiment of the drilling process was subdivided into three stages i.e. drilling-in, steady-drilling, and drilling-out. The initial phase is known as drilling in. The next phase is steady drilling; however, the last phase is known as drilling-out. The initial force is delivered to the surface of the workpiece during the drilling-in phase, and its value progressively increases from the beginning value of zero newtons. The next phase is steady drilling whereas The last phase is known as the drilling-out, and it involves gradually and continuously decreasing the magnitude of force. The same procedure has been followed for the milling as well as the turning process. A similar experimental strategy was used for the milling and turning process as well. The experimental final results were processed based on the machine learning methodology. All the mathematical modeling described for lightGBM and Stacking Machine learning Models were demonstrating the architectures only. The force signals were collected and after the extraction of desired features, the input data was fed into the Machine learning models for training and testing purposes. CNC machine is shown in Fig. 2.

A total of 18 experiments were performed. Six experiments were conducted for the drilling process, six tests were conducted for the milling process and six experiments were performed for turning

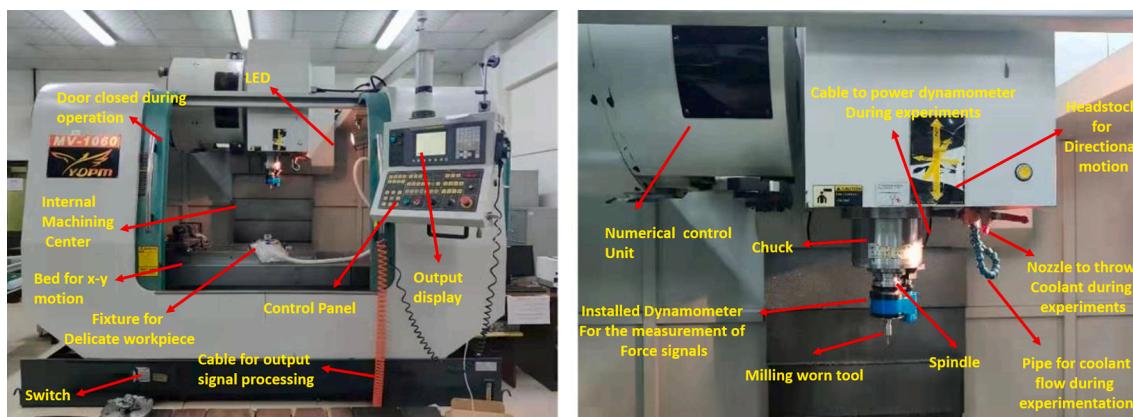


Fig. 2. CNC Machine based Experimental Setup.

operation. The primary goal of 18 experiments is to collect force signal data to extract 15 features for training and testing the LightGBM and Stacking model. The test tools were utilized to acquire data signals to test the created model. After all, slits were carved into the Inconel-718 workpiece that is not as deep as the holes drilled into the workpiece during the drilling procedure. The Model performed well because of the use of noise-free data. The force signals that generated followed the pattern of time series. Milling, drilling, and turning operations were performed on Inconel 718 material-based workpiece. This material is commonly used in the aerospace industry. Inconel 718 is considered a Ni-based superalloy which is particularly suited for greater strength applications at temperatures ranging from cryogenic to 1400°F. Inconel 718 does have high toughness strength. IN718 is still the suitable material for the bulk of aviation engine parts with ambient temperature under 650 °C. Inconel 718 is just a precipitation-hardenable Ni-chromium alloy that contains considerable quantities of iron, molybdenum, and niobium, as well as made from a mixture of aluminum as well as titanium. It features rust resistance as well as greater strength with exceptional hardenability and resistance against post-weld crack. At temperatures up to 700 °C, the alloy offers high creep-rupture durability. When subjected to heat, the superalloy does have an inherent capacity to form a consistent and resilient oxide layer. This spontaneous oxidation process shields the materials from harm. The protective oxide coating provides substantial oxidation resistance to this Inconel alloy, preventing it from rust. At ambient temperature, the exotic alloys material has a minimum yield strength of 725 MPa and tensile strength of 1035 MPa. If Inconel 718 is treated with solutions and deposition, these numbers rise to 1035 MPa and 1240 MPa, correspondingly. Inconel 718 is becoming synonymous with 3d printers because of its effectiveness in producing materials and accessories for a variety of sectors. Several research has been conducted to investigate the application of exotic alloys in the additive manufacturing sector, as well as additional improvements to maximize its properties. The purpose of conducting milling, drilling, and turning experiments was to acquire the forces signal values to train Machine learning models. The scheme (Plan) of Machining operations includes performing a total of 18 experiments. 6 experiments were performed by using 6 drilling tools. The same strategy is followed for milling as well as drilling operations. There was a total of 18 tools used for 3 machining operations. Out of 18 tools, 15 tools were worn were as 3 tools were normal means having fine cutting edges with no abnormality. Each machining operation has one normal tool and five worn tools. For five worn tools, they were distributed in such a way that level-1 worn tool has the least flank wear land width. The level of the worn tool with the increase in the value of flank wears land width. Level-5 worn tool has the greatest value of flank wear land width. Cutting parameters are described in Table 2. A dynamometer was installed on a CNC machine for the collection of force signals Table 3.

Each experiment of the drilling process was subdivided into three

**Table 2**  
Cutting parameters.

Tools Types	Feed rate (mm/rev)	Spindle Speed (r/min)
Drilling Normal tool	0.12	600
Drilling Level-1 Worn Tool	0.15	650
Drilling Level-2 Worn Tool	0.18	700
Drilling Level-3 Worn Tool	0.21	750
Drilling Level-4 Worn Tool	0.24	800
Drilling Level-5 Worn Tool	0.27	850
Turning Normal tool	0.07	200
Turning Level-1 Worn Tool	0.07	250
Turning Level-2 Worn Tool	0.09	300
Turning Level-3 Worn Tool	0.11	350
Turning Level-4 Worn Tool	0.13	400
Turning Level-5 Worn Tool	0.15	450
Milling Normal tool	0.19	1000
Milling Level-1 Worn Tool	0.21	1050
Milling Level-2 Worn Tool	0.25	1100
Milling Level-3 Worn Tool	0.29	1150
Milling Level-4 Worn Tool	0.33	1200
Milling Level-5 Worn Tool	0.37	1250

**Table 3**  
Properties and chemical composition of workpiece material IN718.

IN718 Chemical Composition		Properties of IN718		
Elements	%	Names	Units	Values
Cr	21	UTS	MPa	$1.042 \times 10^3$
Ni	53.390	Elastic Modulus	MPa	$7.253 \times 10^2$
Mo	3.02	Elongation at breaking	%	30
Nb	5.3	Specific gravity		8.191
Co	0.21	Thermal conductance	W/m-K	6.52
Ti	1.019	Electrical resistance	Ω-m	$1.211 \times 10^{-6}$
Al	0.55	Melting point	°C	$1.3753 \times 10^3$
Fe and others	Balance			

stages i.e. drilling-in, steady-drilling, and drilling-out. Only one tool can be used at a time and based on three stages to conduct each experiment. If we want to do the two experiments with one tool, then the three stages would be different and it would be impossible to combine the force signals of two experiments conducted by each tool. A similar strategy had been followed for the milling and turning process. The initial phase is known as drilling in. The next phase is steady drilling; however, the last phase is known as drilling-out. The initial force is delivered to the surface of the workpiece during the drilling-in phase, and its value progressively increases from the beginning value of zero newton. The next phase is steady drilling whereas The last phase is known as the drilling-out, and it involves gradually and continuously decreasing the magnitude of force. The same procedure has been followed for the

milling as well as the turning process. A similar experimental strategy was used for the milling and turning process as well. The data obtained by three machining processes were combined to input into the developed models.

The obtained data is very huge and contains a large number of force signal values. For turning, each tool has generated 10,000 approximately force signal values by installing a dynamometer. For milling, each tool has generated 2600 approximately force signal values by mounting a dynamometer. For drilling, each tool has generated 2900 approximately force signal values by mounting a dynamometer.  $(6 \times 10,000) + (6 \times 2600) + (6 \times 2900) = 93,000$  force signals values approximately. This is a very huge data set containing a massive amount of force signal values. This dataset is more than enough to train any machine learning model. Due to the larger size of data, LightGBM is used which can process huge data very efficiently.

The hardness of Inconel-718 is 330 MPa approximately. The hardness was measured by the Brinell hardness test. Brinell hardness test involves determining the diameter of the indentation created by a continuous focused load exerted to a test specimen using a steel sphere indenter. Steel ball indenter initially comes into contact with substrate before applying a steady force for 10 to 15 s, which is defined as the dwell period. Following the completion of the dwell period, the sphere indenter is withdrawn, creating a round-shaped indentation on the specimen. The description of the geometry of the tools is stated in such a way that the diameter of the milling and drilling tool are respectively. The milling tool has four flutes whereas the drilling tool has two flutes. The turning tool has one major cutting edge, one minor cutting edge, and one corner. The rake angle of a cutter's rake surface determined on a different plane is considered as the orientation angle of the rake surface concerning the reference plane. Clearance angle is considered as an angle of incline of the cutter's flank surface concerning velocity vector which is calculated on a different plane. The primary material of the tools was solid carbide which is much more resistant to wear as compared to high-speed steel and susceptible to chipping rather than wearing out gradually over time. As a result, this material is mostly employed in finish implementations in modern machining processes or ones with reduced spindle wear. Carbide material tools are often created by coating carbide with some other metal, such as tantalum, which gives them exceptional high thermal stability and makes them perfect for relatively high surface finishes. operate CNC machines across each axis by using the Cartesian coordinate system to turn a block of feedstock into a final product. the accompanying motions are commonly obtained from the viewpoint of an operator confronting the machine depending on the axis. The X-axis enables for "left" as well as "right" motion. The Y-axis enables for "forward" as well as "backward" motion. The Z-axis enables for "up" as well as "down" motion. Increasing the value of the X-coordinate pulls the table to a left, while from the tool's standpoint, it goes right across a workpiece. Increasing the Z-axis coordinates raises the spindle while reducing it lowers it into a product.

A particular approach is utilized to calculate the spindle speed and feed rate. Each tool comes with its own set of SFM (Surface Feet per Minute). The SFM range is available in the catalog of tools (from Minimum to Maximum). For the conversion of SFM to RPM, there is a formula that may be utilized. The spindle speed calculation formula is shown below.

$$n = 12 \times SFM / \pi D \quad (1)$$

The cutting tool diameter is D, and spindle speed is denoted by n. The exact tool number, as well as material group (material of workpiece) from the tool catalog provided by the tool manufacturer, are used to determine the range (from Minimum. to Maximum.) of SFM. If SFM for the tool is wrong, the feed rate and spindle speed would be inaccurate as well, leading to tool burn. So, the SFM searching method, which makes use of data such as the tool's kind and particular number as well as the material group, is important. It is advisable to select an SFM near to the Lowest range, because the greater the SFM, the greater unstable will

be the tool. The manufacturer's catalog has the IPR (Inch Per Revolution) value.

$$feedrate = RPM \times IPR \quad (2)$$

The most significant difficulty that could face during machining is regarding the coding. Because CNC machines are very complex machines operated by CNC computers, many of the challenges that emerge in CNC machining are frequently related to programming. These could result in a lack of analysis of various G and M codes used by the controller, incorrect setup, or entering the incorrect data parameters into the CNC controller. To correct these mistakes, new operators must receive proper training in the many ways that CNC machines may be configured. Machine vendors or professional operators upon that production floor should give comprehensive and reliable manuals and orientation, movement sequences, and machine operation to novice operators. Material selection is very important because industries demand products with highly accurate dimensions and great stability. Such circumstances will need the use of materials with high dimensional stability, i.e. a low distortion factor. Various materials react differently to the forces created by the cutting tool. As a result, the item used must be capable of meeting the specified tight tolerances. The greater the machinability of a material, the easier it is to obtain greater reliability. Corrosion resistance should be considered. Corrosion is a serious problem in many operational settings, dictating material selection. Materials can react negatively to chemical compounds in a variety of environments. High acidity is one of the chemical variables that may be found in the environment. Thermal resistance is also considered while selecting materials since the influence of excessive temperatures on materials can produce enhanced corrosion rates. The initial cost is usually a crucial consideration in the material-selection process. The technique of One-hot encoding is applied in Table 4 and Table 5.

## 2.2. Milling, turning and drilling flank worn tool's visual illustration

Flank wear is most common on the cutting edge's side. The flank wear of the tool of drilling, turning and milling procedures is clearly illustrated in Fig. 3. The method used to measure flank wear is depicted in the diagram below. Wear land in brown color denotes the zone of flank wear. Flank wear land width is shown in all drilling, milling, and turning worn tools. If the values of flank wear land width become larger than 0.5 mm to 0.6 mm, the greater cutting force values will lead a tool to fail during the cutting process. Flank wear is calculated by taking the average of wear land area and dividing it by the greatest value of wear land area. It may be defined using the tool life expectancy equation. Another major cause of flank wear is the greater level of feed and the greater depth of cut. Abrasion is caused by the discharge of tough

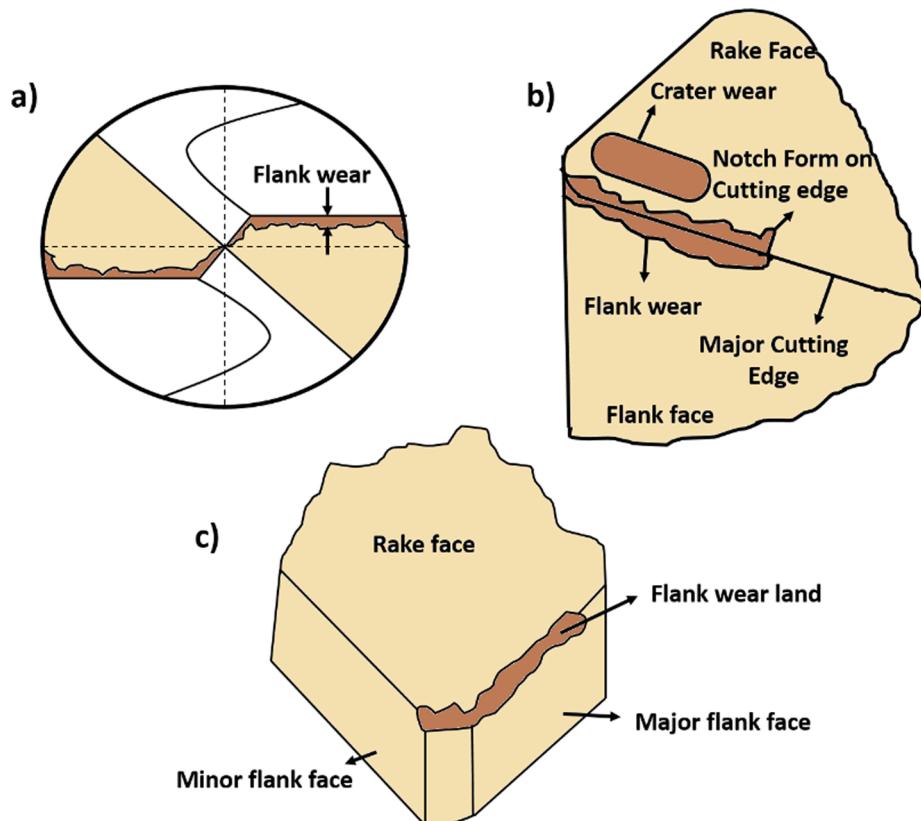
**Table 4**  
Model's target variables.

Index	Target
1	Drilling Normal Tool detected = x
2	Drilling Level-1 Worn tool detected = x1
3	Drilling Level- 2 Worn tool detected = x2
4	Drilling Level- 3 Worn tool detected = x3
5	Drilling Level- 4 Worn tool detected = x4
6	Drilling Level- 5 Worn tool detected = x5
7	Milling Normal Tool detected = y
8	Milling Level-1 Worn tool detected = y1
9	Milling Level- 2 Worn tool detected = y2
10	Milling Level- 3 Worn tool detected = y3
11	Milling Level- 4 Worn tool detected = y4
12	Milling Level- 5 Worn tool detected = y5
13	Turning Normal Tool detected = z
14	Turning Level-1 Worn tool detected = z1
15	Turning Level- 2 Worn tool detected = z2
16	Turning Level- 3 Worn tool detected = z3
17	Turning Level- 4 Worn tool detected = z4
18	Turning Level- 5 Worn tool detected = z5

**Table 5**

Application of One Hot Encoding to model's target Variables.

Num.	x	x1	x2	x3	x4	x5	y	y1	y2	y3	y4	y5	z	z1	z2	z3	z4	z5
1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
10	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	
11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	
12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
13	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
14	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
16	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
17	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	

**Fig. 3.** Visual illustration of a) Drilling worn tool b) Milling worn tool c) Turning worn tool.

particles from the surface of cutting during the procedure. There is micro weld shearing between the tool and work material. Flank wear increases the level of cutting force necessary to cut the material. It has a negative impact on the dimensional precision of the manufactured product. It will enhance the end product's surface roughness. The geometry of the workpiece will alter due to flank wear. The flank wear land width used in the study was for one flute per tool. One of the most crucial factors to consider when selecting an end mill is deciding the optimal flute count for the work at hand. In this vital stage of the tool screening process, both component and usage play an essential role. Knowing the implications of flute count as well as how a tool will react in certain scenarios, is an important aspect of the tool selection procedure. Tools with much more flutes, on average, have a bigger core and shorter flute valleys than

tools having fewer flutes. Depending on the requirements, additional flutes with a bigger core might bring benefits and challenges. Essentially the size of the core is directly related to the hardness of the tool; the greater the score, the tougher the tool. As a result, a bigger core limits the flute depths of a tool, reducing the amount of available space for chipping. The criteria used for milling was ISO 8688-2:1989 and for turning was ISO 3685:1993.

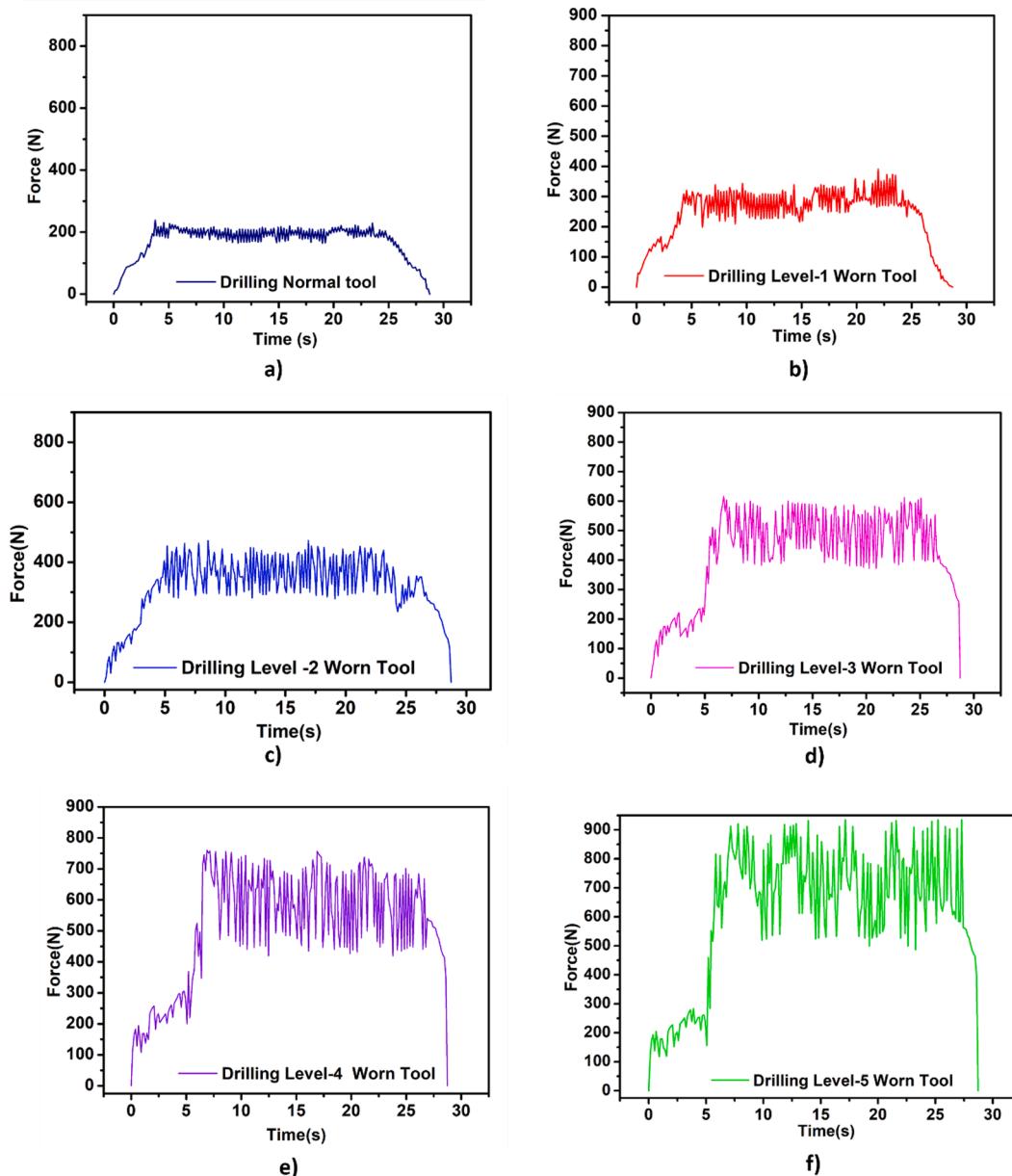
### 2.3. Drilling experiments using the normal tool and worn tools

Drilling tests on Inconel-718 were accomplished utilizing a force sensor mounted on a 3-axis CNC machine. Force signals were recorded for five levels of drilling worn tools and normal tools. During the initial

stage of experiments, the normal drill tool was used for drilling holes into the Inconel-718 workpiece. The normal tool possesses sharp edges. The length of time required to drill per hole into the material was 30 s and a step size of 0.13 secs. The force sensor dynamometer has provided 222 force signals' data values for each tool used in the drilling experiment, which can be seen in Fig. 4. This graph plotted in the figure of the normal drilling tool is considerably more smooth and more regular. The graph depicted all of the top maxima on the area of force signals' values ranging from 275 N to 300 N, and this regular and smooth depiction of the plotted graph shows that the corresponding utilized tool is sharp and has fine edges. Each drilling process was subdivided into three stages. The initial phase is known as drilling in. The next phase is steady drilling; however, the last phase is known as drill-out. The initial force is delivered to the surface of the workpiece during the drilling-in phase, and its value progressively increases from the beginning value of 0 N.

The last phase is known as the drilling-out, and it involves gradually and continuously decreasing the magnitude of force. In Fig. 4.a), the force is raised from the value of 0 N to 275 N. This shows that

throughout the early phase of the drilling operation, the highest force value reached around 275 N, but the force value ranged between 225 N and 275 N throughout the steady drilling. The top maxima in the graph are almost the same, and their in-between range is nearly the same, demonstrating the uniformity of the performed drilling operation utilizing the tool with sharp edges. During the drill-out phase, tiny waveforms are presented, indicating that the hole diameter and the drill bit diameter are approximately identical, which is a favorable outcome. The force value is considerably decreased from the force value of approximately 212 N to 0 N as demonstrated in Fig. 4.a). After it drilling test is completed, raw force signal data has been collected to process for features extraction. The second stage of the experiment involves drilling with various levels of worn tools. Five distinct drill tools had been specifically selected. Based on the breadth of the flank wear land width, the worn tools were classified into five groups. The level-1 worn tool has the smallest value of flank wear land width, which is 0.19 mm; the level-2 worn tool has a higher value of flank wear land width, which is 0.32 mm, and the flank wear land width of the level-4 and level-3 tools are

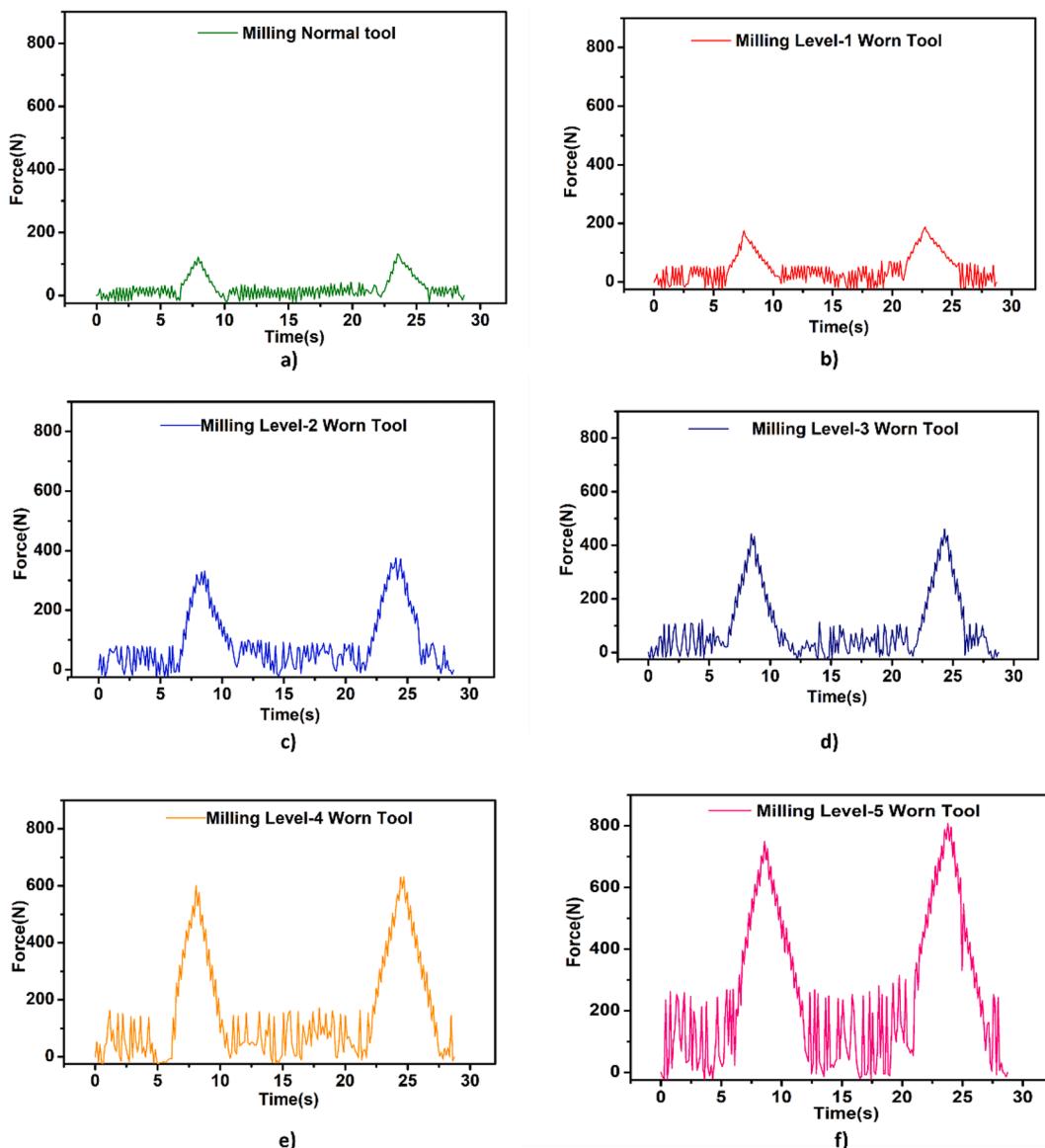


**Fig. 4.** Drilling operation based dynamometer obtained Force Signal values for a) Normal tool b) Level-1 worn tool c) Level-2 worn tool d) Level-3 worn tool e) Level-4 worn tool f) Level-5 worn tool.

0.57 mm and 0.46 mm, respectively. The level-5 tool, on the other hand, is severely damaged, with the maximum value of flank wear land width being 0.69 mm. The shank is known as the end of a twist drill. The chuck of the drill tool tightly grips this region, while the chisel edge is considered as the region located at the drill bit's tip. Fig. 4. depicts the force signals of various degrees of worn tools. Each tool's drilling experiment was similarly split into 3 phases. An initial phase is drilling-in, as seen in the graphs in Fig. 4. The region where the force value rose from 0 N to a predetermined level is continuous drilling. Steady drilling is the technique of drilling in which the force remains constant and varies between specific values. This fluctuation is correlated to the severity of tool wear, as if the tool wear of a specific tool is extreme, these oscillations will be larger than the variations of the force signals of a regular tool. The drill-out procedure is described as progressively lowering the spindle speed till the withdrawal of the cutting tool from the material.

The most notable and evident finding is that the signals displayed in the plots of diverse types of worn tools have risen significantly when compared to the force values shown in the plot of a normal tool. The highest force value for the level-1 tool has been raised to 350 N. The

maximum level of force value for the level-2 worn tool is approximately 450 N, as shown in Fig. 4., whereas the force value has reached 600 N in the scenario of the level-3 worn tool. The pattern of rising force values with increased tool wear severity has also been seen for level-4 and level-5 tools, with force signals reaching extreme values of 750 N and 925 N, respectively. If this plot region for first drilling of various degrees of all worn tools is examined in comparison to the trend of the normal tool shown in Fig. 4 a). During the initial stage of the drilling operation, the force value has risen to the peak value needed for steady drilling, which is considered as a drill-in process, displaying irregular behavior with bewildering oscillations and sharp peaks that portray the severe fall and rise of the force values during the early phase of drilling, that is well visible for all levels of worn tools. The amplitude of the oscillations during constant drilling has risen with the expansion in tool wear land, indicating that tool wear would increase by impacting force applied on the material's surface. It is clear from a comparative examination of the plots of all worn tools that the severity of the tool wear has a significant influence on the force signal produced by the dynamometer. Level-1 worn tool oscillations vary from 300 N to 450 N, Level-2 worn tool oscillations vary from 370 N to 600 N, Level-3 worn tool waves ranging



**Fig. 5.** Milling operation based dynamometer obtained Force Signal values for a) Normal tool b) Level-1 worn tool c) Level-2 worn tool d) Level-3 worn tool e) Level-4 worn tool f) Level-5 worn tool.

from 400 N to 600 N, Level-4 worn tool oscillations extend from 450 N to 750 N, and Level-5 worn tool waves ranging from 550 N to 900 N.

#### 2.4. Milling experiments using the normal tool and worn tools

The milling procedure is the second machining operation performed in this research investigation. 1 normal tool and 5 worn tools are among the total of six tools. According to the flank wear land width, all worn tools were classified into five levels ranging from level 1 to level 5. First, the milling procedure was conducted using a regular tool that was not impaired and had no value flank wear land width. The cutting operation on Inconel-718 was done by the milling tools. The selected workpiece is comprised of an Inconel-718 and has unique characteristics. Because of its toughness, hardness, and plasticity values, it has a significant impact on the cutting tool's condition. The dynamometer was mounted on the CNC milling machine to record force signals. Following the end of the milling process on Inconel-718, data values for 222 force signals were collected. Fig. 5. demonstrates the force signals in the x-direction. Milling with the level-1 worn tool yielded a high force value of 125 N. The force exerted by the worn tool on the material is proportional to tool wear. Because the standard tool has no flank wear zone, it has the lowest force value in the milling tests when compared to the other degrees of worn tools.

As a result, the minimum cutting force was delivered to the material (Inconel-718) orthogonal to feed direction while milling with a normal tool. Because the normal tool has a sharp chisel edge and fine cutting lips so, the material removal was extremely smooth and also with a minimal amount of power. One of the most essential elements of the investigation is collecting the force signals data by milling with varying levels of the worn tool. A total of 5 worn tools had been selected after a full assessment. The worn tools were classified into five levels based on the values of flank wear. V B values for level-1, level-2, level-3, level-4, and level-5 worn tools are 0.11 mm, 0.28 mm, 0.37 mm, 0.41 mm, and 0.54 mm, respectively. Fig. 5 demonstrates the force data signals plotted in the x-direction. The total force exerted by the worn tool on the Inconel 718 material is proportionate to the breadth of the flank wear area. The approximated maximal force applied by the level-1 worn tool is 175 N, as shown in plotted force signals graphs. Because it had the smallest percentage of flank wear, the level-1 worn tool exerted the minimum force on Inconel 718 material. But, a level-5 worn tool with the worst flank wear had been observed to have a maximal force of around 750 N. The mounted dynamometer measured maximum force of about 340 N, 460 N, and 630 N for the level-2 worn tool, level-3 worn tool, and level-4 worn tool, as displayed in plotted force signals graphs of the milling operation. The worn tool's graphs demonstrated the same - patterns. Each cutting phase begins when the tool interacted with the Inconel-718 material and is completed when the worn tool has no interaction with the Inconel-718 workpiece. Simultaneously, the force value increased from 0 N to the peak value and then significantly reduced from the peak value to 0 N, which is described as the milling cycle.

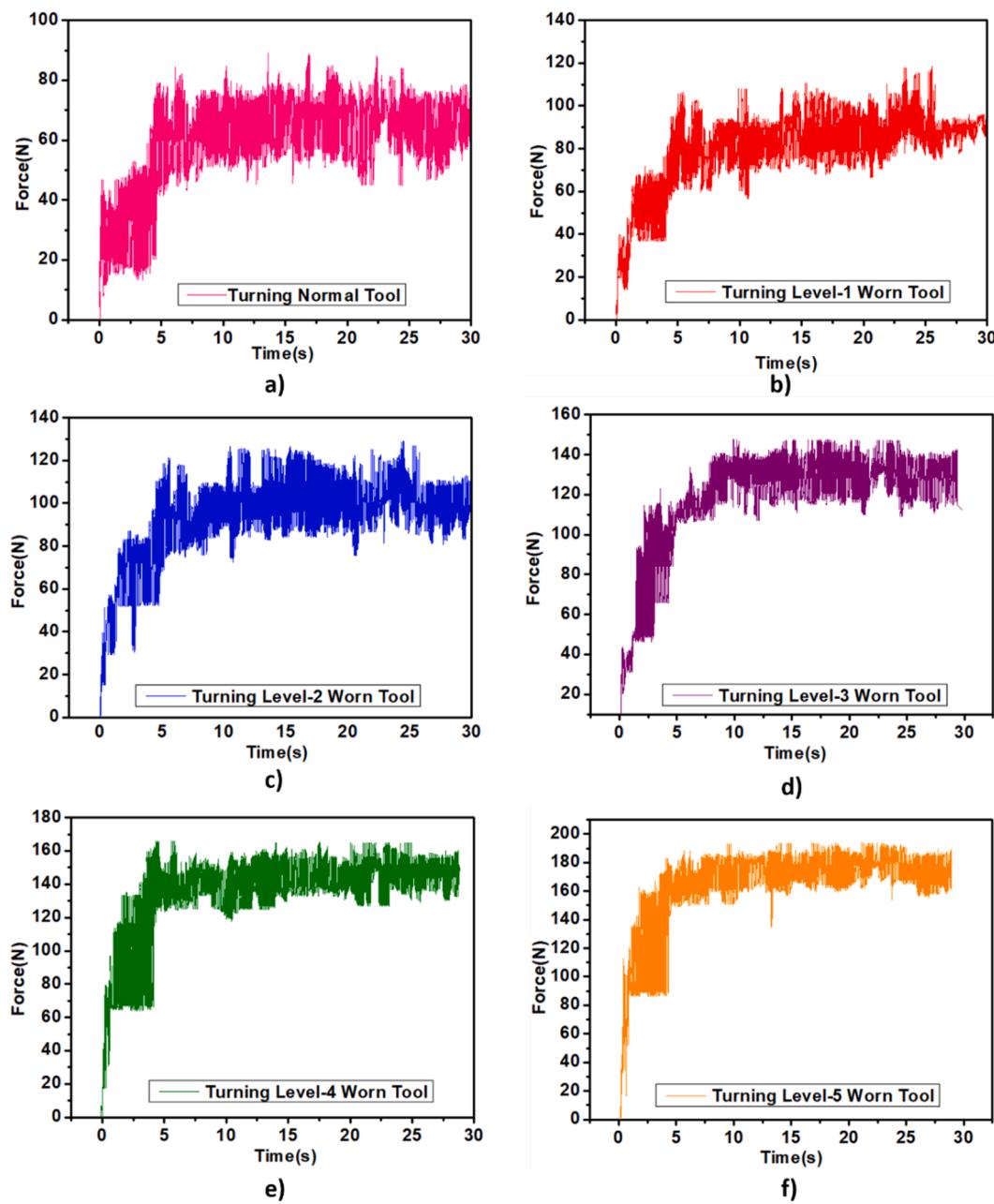
The forces exerted by the worn tools are higher than the force applied by the normal tool throughout the milling cycle. The forces of tools rise as the breadth of the flank wear zone increases. All of the selected tools contain deformations at the edge of the tool, which would be shown to reflect the tool wear area, while the rake angle has plastic deformation. Another finding is that deformations in the circumstance of the worn tool are incremental i.e. increasing with the rise as the force is exerted during the continuous milling operation.

#### 2.5. Turning experiments using the normal tool and worn tools

During the turning operation, the cutting tool involves material removal from the outside diameter of a spinning workpiece. The primary goal of turning is to decrease the diameter of the component to the required dimension. Rough turning operations seek to manufacture a

component to a predetermined thickness by cutting the largest volume of material in the quickest period, with little concern for surface integrity. Finish turning results in a clean surface finish and workpieces with final precise measurements. During CNC Turning the material bars are gripped in a chuck and then rotate while the tool works on the workpiece for material removal to achieve the required form.

A turret with tooling installed was programmed to travel to the raw material bar for material removal to get the desired output. Same as milling and drilling operations, 6 turning tools were used for the collection of force signals during the turning process. Turning tools have different levels of flank wear land width. The normal turning tool has zero value of flank wear land width whereas the tools from level-1 to level-5 have the values of 0.14 mm, 0.19 mm, 0.33 mm, 0.42 mm, and 0.51 mm respectively. The force value increased from zero to a maximum of approximately 85 N in the plotted force signals graph of the normal turning tool. The turning force oscillated between 20 N and 50 N for approximately 5 s. After that, the force significantly increased to 80 N and the waveform illustrates that the cutting force was confined to the region of 50 N to 80 N from 5 s to 30 s throughout the turning operation. The level-1 worn tool force signals are shown in plotted force signals graphs of turning operation in Fig. 6. The turning force in the plotted graph has touched the peak value of 100 N and the basic difference between the normal tool and the level-1 worn tool is the increment in the turning force whereas the pattern of the waveform is similar. The dynamometer has given more than 10,000 data values of force signals for each turning tool therefore the turning plot is so much dense as compared to the plots of milling and drilling operations. The peak force shown in the case of the level-2 worn tool was approximately 115 N. It means that cutting force increase if the tool is worn thus cutting force during the machining operation is directly related to the rate of tool wear. The maximum turning force for the level-3, level-4, and level-5 worn tools were approximately 140 N, 165 N, and 185 N as graphically illustrated by Fig. 6. The pattern plotted during the initial turning operation from 1 s to 6 s for level-4 and level-5 worn tools were pretty similar. So, the figure demonstrated that turning force has increased significantly with the increase in the flank wear width land which is easily acknowledged if the normal turning tool is compared with the turning tool having the level-5 of tool wear. The length of the bar might largely depend on a bar feeder. This has an impact on how much handling is necessary for high-volume tasks. CNC turning centers contain tooling installed on a computer-controlled turret. The more the tools that the turret can contain, the more possibilities for complexity on the component are accessible. Metal cutting forces have a significant impact on the formation of heat, complete failure, tool wear, the integrity of the workpiece surface, and the precision of the machined surface. During machining operations, a device reads and stores real-time data of the force signals. The frequency of vibration of the tool upon which the dynamometer is installed for measuring cutting force should be the same as the dynamometer's natural frequency. The natural frequency of a dynamometer must be as higher as feasible. The machine tool's vibration frequency is proportional to its spindle speed. The natural frequency of the dynamometer should be four times the machine tool's vibration frequency. When estimating tool wear during turning, the cutting force is evaluated while the feed rate is systematically changed in real-time. It is well acknowledged that when tool wear seems to be significant, the conventional cutting resistance at a lower feed rate should be substantially higher. Generally, flank wear in the turning tool develops with increasing cutting time. In each cutting condition of continual chip formation as well as broken chip creation, the dynamic force will have its distinct pattern during turning operation. As even the chips were shattered throughout the turning process, the dynamic turning force changed as a result of the chips impacting the turning tool and the workpiece. Whenever the chips are shattered into fragments, the feed force's dynamic component has the greatest amplitude of the 3 force components. Whenever during the turning process the chips are continual, however, the amplitude of the dynamic cutting force is



**Fig. 6.** Turning operation based dynamometer obtained Force Signal values for a) Normal tool b) Level-1 worn tool c) Level-2 worn tool d) Level-3 worn tool e) Level-4 worn tool f) Level-5 worn tool.

minimal. Metal turning forces have a significant impact on the formation of heat, the integrity of the workpiece, and tool wear.

#### 2.6. Dimensionality reduction and features section from the raw force signals

SSA is a sophisticated and alternative time-series approach for feature extraction. SSA (Singular Spectrum Analysis) technique is divided into two phases. The first one is called decomposition, whereas the second phase is called reconstruction; they include two different stages. The Time-series signal was deconstructed in the first step, and the original time series is rebuilt in the subsequent step. The term “singular” refers to the breakdown of a spectrum or matrix. Matrix column is decomposed using matrix eigenvalues and Eigen-decomposition. The term “spectrum” is clearly described as the combination of a collection of eigenvalues following spectral decomposition.

The time-series approach is SSA and its main goal is to decompose the initial time series into multiple linear components that are independent and completely distinct. The SSA methodology consists of many phases that include reconstruction and decomposition. SSA is a unique non-parametric technique for analyzing time series that is based on concepts and norms of multivariate statistics. SSA is a method for removing noise from a force signal that decomposes a specific time series into a group of unrelated time series. Using the principal component analysis approach, the methodology projects the raw signals onto the vector basis derived from a time series signals itself. The decomposition approach yielded a collection of series that may be inferred as a gradually shifting displaying the signal median at a certain instant. In this investigation, SSA is utilized to decompose force signals during the drilling, milling, and turning procedure. The complete method of applying SSA to time-series unprocessed force signals is split into five stages, which are explained below by using mathematical expressions.

The Trajectory method is used to compute the covariance matrix. The initial stage in SSA is to transform the force signals' time-series S represented by P into a sequence (multidimensional) of delayed vectors.  $W_l$  denotes the window's length. Column vector is Y

$$2 \leq W_l \leq \frac{N}{2} \quad (3)$$

Subseries created the window.

$$P = \{P_i, P_{i+1}, \dots, P_{i+L-1}\} \quad (4)$$

window traverses alongside with the provided time series, then Y will be created for individual subseries, and the resulting results are as follows:

$$Y_0 = (P_0, P_1, P_2, \dots, P_{L-1})^T \quad (5)$$

$$Y_1 = (P_1, P_2, P_3, \dots, P_L)^T \quad (6)$$

$$Y_2 = (P_2, P_3, P_4, \dots, P_{L+1})^T \quad (7)$$

$$Y_3 = (P_3, P_4, P_5, \dots, P_{L+2})^T \quad (8)$$

$$Y_4 = (P_{N-L}, P_{N-L+1}, P_{N-L+2}, \dots, P_{N-L})^T \quad (9)$$

The Trajectory Matrix (Hankel matrix) indicated by Y is formed by such column vectors. SVD denotes Singular Value Decomposition

$$Y = \begin{bmatrix} P_0 & P_1 & P_2 & P_3 & \dots & P_{N-L} \\ P_1 & P_2 & P_3 & P_4 & \dots & P_{N-L+1} \\ P_2 & P_3 & P_4 & P_5 & \dots & P_{N-L+2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ P_{L-1} & P_L & P_{L+1} & P_{L+2} & \dots & P_{N-1} \end{bmatrix} \quad (10)$$

The outcomes of SSA decomposition are constrained by the window's length L. SVD was implemented for the Hankel matrix once the trajectory matrix has been calculated.

$$X = Y \times Y^T \quad (11)$$

First, compute the eigenvectors and eigenvalues of matrix X having the dimension  $L \times L$ . Let  $\lambda_1, \lambda_2, \dots, \lambda_d$  be the non-zero eigenvalues of matrix X, and  $U_1, U_2, \dots, U_d$  be the coefficients of respective eigenvectors. The new vector was created by utilizing the mathematical equation and the results become

$$V_i = X^T \times U_i / \sqrt{\lambda_i} \quad (12)$$

Using the Singular value decomposition approach, the trajectory matrix was decomposed into the sum of matrices.

$$X = S_1 + S_2 + S_3 + S_d \quad (13)$$

$$S_i = \sqrt{\lambda_i} \times U_i \times V_i^T \quad (14)$$

These are the unit rank as well as mutually orthogonal elementary matrices. The singular spectrum is a methodology for plotting and visualizing the eigenvalues in decreasing order. The main components  $P_c$  would've been computed in this phase by taking the dot product matrices. whereas the trajectory matrix is denoted by Y and the eigenvector matrix is denoted by  $Z_{HO}$

$$P_c = Y \times Z_{HO} \quad (15)$$

$Z_c$  are computed by the projection of Principal Components back onto eigenvectors, and the resultant time series is referred to as reconstructed components. The fundamental approach is inverting the projection  $P_c = Y \times Z_{HO}$ . The calculation presented below yields  $P_c$ .

$$Z_c = Y \times Z_{HO} \times Z_{HO}^{-1} \quad (16)$$

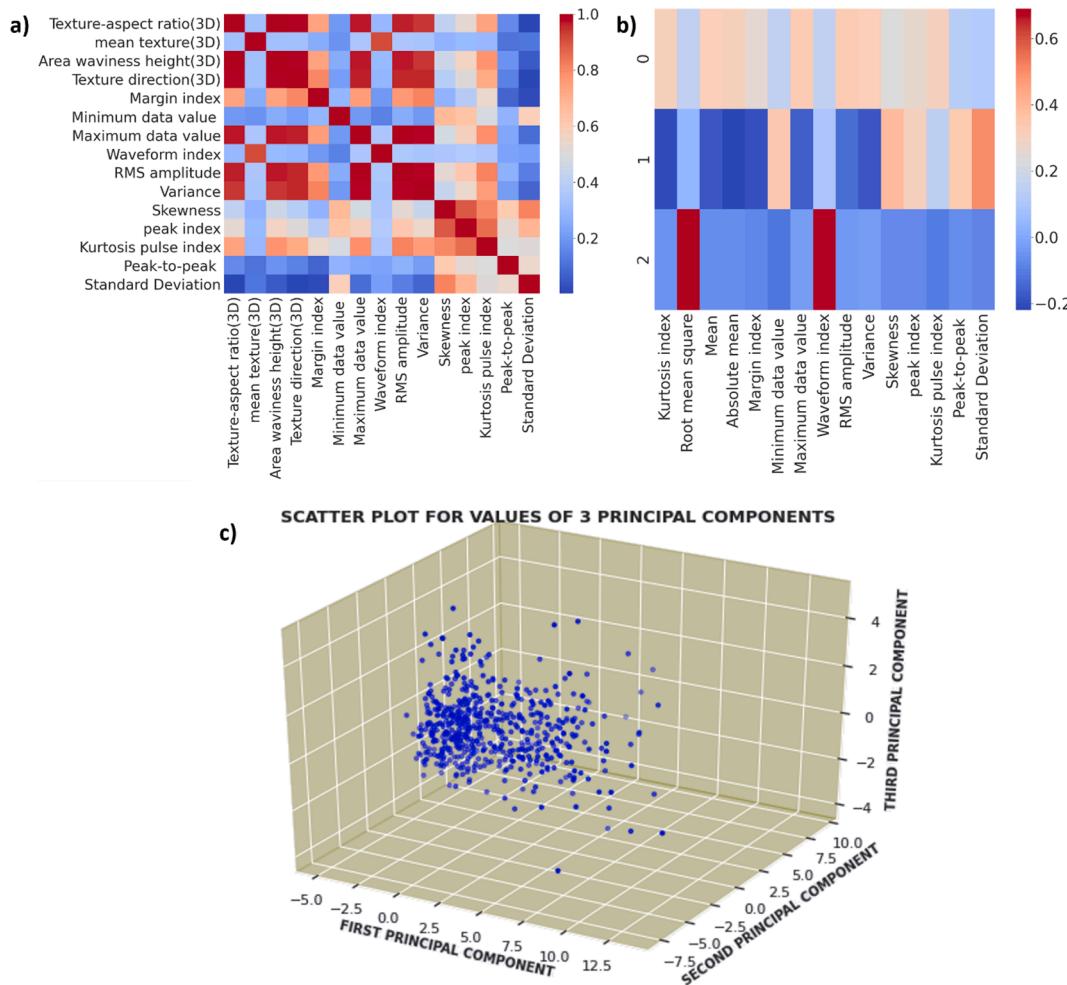
$$Z_c = P_c \times Z_{HO}^{-1} \quad (17)$$

Principal Component Analysis is the technique for decreasing the dimensionality of a big dataset composed of the force signals of drilling, milling, and turning operations while enhancing interpretability as well as limiting information loss. The SSA characteristics are subjected to PCA. The validity of the implementation of PCA on the Model's features is displayed in Fig. 7. PCA reduces a big collection of parameters to a smaller set while retaining significant information. SSA retrieved 15 features in total, and the heat map depicts the correlation of those 15 features extracted from the force signals' data. A correlation has been the most widely used statistical tool to examine associations between variables. The fifteen dimensions were transformed into three dimensions, three primary components. To build an efficient machine learning model the data need to be split into training data set and testing data set. Training data is solely used for training and making the proposed model to learn the patterns as well as hidden features present in the dataset. The same data values are input to the neural network periodically for each epoch, and the machine learning model continued to acquire the data's features. The training set should include a diverse collection of inputs so that the model can be trained in all settings and estimate any previously unknown sample data that may emerge in the process. Whereas training data set only gives an objective final performance in terms of accuracy and precision, and recall. The split ratio of data is determined by the total number of samples present in the dataset. When training data are scarce, the model exhibits a significant variation in training. More test data is beneficial since it provides a better idea of how better the model generalizes to unseen and unknown data. If there is a lack of sufficient test data, your ultimate decision of the algorithm's generalization capacity may be inaccurate.

The results of the three principal components were displayed in Fig. 7. c). The first key component possesses the most variation. The 2nd component is generated in the same way as the first, with the exception that it is orthogonal to the 1st principal component, which accounts for the most variation. Fig. 7. b) depicts the heat map and demonstrates the relation between the correlation of Fifteen features and the correlation of three main principal components. The initial stage is normalization, which is proceeded by the determination of the covariance matrix, and then the covariance matrix's eigenvectors and eigenvalues are computed to determine the major components. The three most major benefits of PCA are as follows. It eliminates the associated characteristics since they have a negative impact on the model's performance. PCA generates principal components that have no association and are completely independent of one another. Fig. 10 and Fig. 11.

PCA would make the system more effective by removing associated characteristics that do not contribute to prediction. PCA lowers the threat of overfitting, which occurs most frequently when a large number of variables are included in the data. As a result, by reducing the number of features, PCA contributes to the resolution of overfitting. PCA is also essential to reduce noise. The model's target variables were in text form. Because deep neural networks cannot perceive and interpret language, so target variables' texts have to be converted into binary format. One hot encoding method was used for the conversion of categorizing variables to binary values. The stopping criterion is mainly used to halt the training process when the efficiency of the validation data begins to deteriorate. Early stopping is a practical and extensively utilized strategy of training neural networks. The prediction value is derived based on the cross-validation which is used as the stopping criterion. This allows identifying the fewest iterations required to develop a model which generalizes well better to new data and minimizes the likelihood of over-fitting.

The developed methodology is primarily based on the experimentation conducted employing the 18 machining tools by mounting the dynamometer on the CNC machine. The eighteen tools were having various levels of flank wear land width. For instance, the flank wear land width value of the level-1 drilling tool will be the least and the value will increase with the increase in the level of the worn tool. The force signals



**Fig. 7.** Graphical demonstration of selected features a) Heat map for 15 features c) Correlation for three PCA derived principal components with fifteen selected features b) Scatter plot for values of 3 principal components.

were acquired. The force signals were in raw form and had a lot of noise in the data. Hence, different techniques were used to remove the noise from the force signals. For this purpose, the Singular Spectrum Analysis algorithm was designed for noise removal. The extraction of fifteen features was accomplished. These features were of high dimensions therefore there was the need to reduce the dimensions to have a positive effect on the efficiency of the model. The Principal Component Analysis was implemented based on the experimental data for the dimensionality reduction from fifteen to three. The dropout and early stopping techniques were implemented to avoid overfitting. After the pre-processing of the experimental data, the algorithms of the LightGBM model were designed. The manual hyperparameter tuning was done for the LightGBM model based on learning rate, number of trees, and tree depth range values. The number of trees hyperparameter has shown the best results as compared to other hyperparameter tuning parameters. The stacking model was developed by designing six different algorithms such as logistic regression, random forest, classification and regression tree, naïve Bayes, support vector machine, and k-nearest neighbors. The performance of the stacking model was not as efficient as compared to the LightGBM model. The final better choice of model for this research study was LightGBM because it was efficient has the fast processing capability of large data set and it also consumed less memory. The experimental data set was huge which was based on the force signals of three machining operations and this model has handled the huge data very effectively. The methodology has overcome the major limitations of previous research studies. The already existing tool wear models have neglected to work on the prediction of the greater number of levels of

worn tools for three machining operations and increase the efficiency of the models by reducing the memory consumption, dimensions, and overfitting.

Case 1: “RUL predictions where limited data is available”. In this condition, the LightGBM model will not work whereas other algorithms such as LR, RF, CART, NB, SVM, and KNN can work. However, their performance cannot reach up the optimum level because sufficient training data will not available for the learning of features. Case 2: “RUL prediction under the big data situation”. In this scenario, all models will work very efficiently. Case 3: “Prediction of single component involving multiple faults”. In this situation, all the models will work effectively provided that the data is available for training purposes. Even in the presence of multiple faults, the model will be able to detect a single component only if the training process has been accomplished accordingly based on enough data. Case 4: “when few data samples are available”. In this case, no machine learning model will be able to work regardless of the problem.

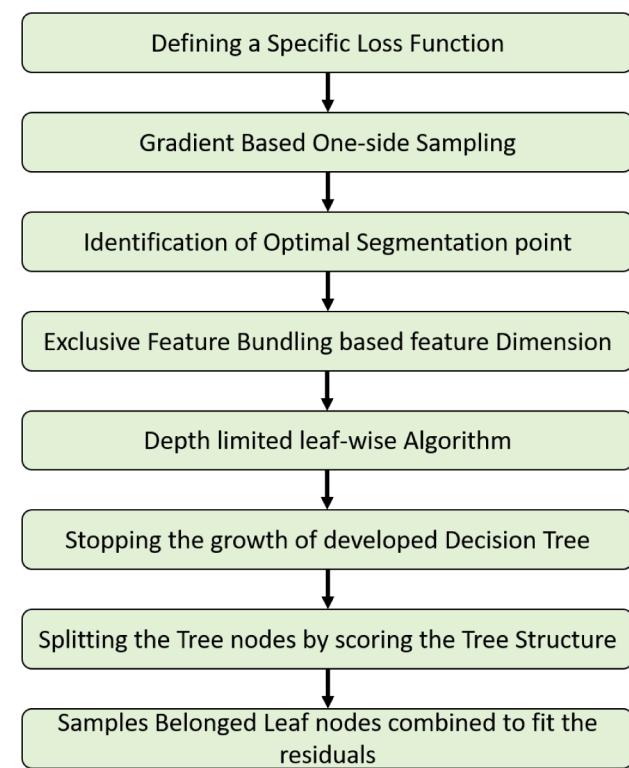
This research article has overcome the shortcoming of previous research studies based on four important parameters as mentioned in Table 7. These parameters have an extremely high impact on the performance and efficiency of ML models for tool wear prediction. Overfitting is a common problem in machine learning models that occurs when the algorithm selects fluctuations in the training data and learns them as concepts. When a designed model fits a greater amount of data than what it requires, it begins to learn noisy data as well as erroneous points in the data. As a consequence, the model’s efficiency, as well as accuracy, are reduced. The algorithm of dropout and early stopping is

applied to overcome the error of overfitting. When the model is being trained, it can monitor how it performs based on each iteration. The process can be continued until the iterations can improve the performance of the designed model. Subsequently, the developed model can overfit the data used for training as generalization worsens for each iteration. Early stopping essentially implies terminating the training phase before the designed model reaches the point when it begins to start overfitting the data used for training. In addition to the usage of early stopping, the dropout technique was also applied to eliminate overfitting. Dropout is the regularization approach that approximates concurrent training of a significant number of neurons with various architectures. Throughout the training, certain layer output is disregarded or “dropped out” at random. It has the impact of having the layer to be regarded and it does have a distinct number of nodes and connectedness to the preceding layer. Every update to the layer throughout training, in practice, is made with a distinct “view” of the specified layer. Dropout can be performed on any of the network’s hidden layers, and the viewable or input layer. For the output layer, it is not utilized. Dropout pertains to disregarding neurons throughout the training stage of a particular group of neurons. By “ignoring,” means that these neurons are not taken into account during a certain forward as well as backward processing. Another significant contribution of the article is dimensionality reduction. Principal component analysis was applied for reducing the dimensions from fifteen to three and Singular Spectrum Analysis was used to extract fifteen features from the force signals of milling, drilling, and turning operations which were obtained by mounting the dynamometer on the CNC machine. The third quality of the current research study is that the designed model is operational for three machining processes i.e. drilling, milling, and turning. Another major contribution of this article is that it has detected the tool wear severity. Total 18 tools were used and 6 for each machining operation (milling, drilling, or turning). For each machining operation, there were five levels of tool wear based on the flank wear land width of each tool. The tool with the least value of flank wear land width was given the grade-1 worn tool whereas the tool with the extreme tool wear having the greatest value of flank wear land width had been assigned the grade-5 worn tool. So this strategy was neglected by the previously published research studies.

## 2.7. Hyperparameter tuned model based on LightGBM

LightGBM is the gradient boosting architecture that employs tree-based learning algorithms, which are thought to be highly efficient in terms of computing. It is regarded as a quick processing algorithm as it can easily handle larger data of three machining operations i.e. milling, drilling, and turning. Whereas other algorithms’ trees develop horizontally, the LightGBM method advances vertically, which means it grows leaf-wise, whereas other algorithms grow level-wise. LightGBM selects the leaf with the greatest loss to grow. When expanding the same leaf, it can reduce loss more than a level-wise method. GOSS (Gradient-Based One Side Sampling) is a new sampling approach that uses gradients to down sample instances. As instances with modest gradients are well trained (low training error), but instances having big gradients are undertrained. A simple method for down sampling would be to remove cases with small gradients in favor of examples with big gradients, but this would change the data distribution. GOSS, in a nutshell, keeps instances with big gradients while randomly picking examples with tiny gradients. GOSS calculating steps include sorting all instances in descending order based on absolute gradients. The principle of LightGBM and the final results of model are shown in Fig. 8 and Fig. 9 respectively.

Choosing the top,  $a * 100$  percent of the occurrences. [ Untrained / high slopes]. Samples  $b * 100$  percent of the occurrences at random from the remainder of the data. The contribution of well-trained instances will be reduced by a factor of  $b$  ( $b < 1$ ). The sample count with tiny gradients would be  $1-a$  (currently it is  $b$ ). To preserve the exact



**Fig. 8.** The principle of the LightGBM Model for force signals’ data of milling, drilling and turning.

distribution, LightGBM multiplies the contribution of samples with modest gradients by a constant  $(1-a)/b$  to emphasize the under-trained cases. This focuses more on the under-trained cases while maintaining the distribution of data. Boosting, on the other hand, trains models consecutively, with each model learning from the mistakes of the preceding model. Models are trained repeatedly, starting with a weaker base model and adding to the predictions of the preceding model to achieve a strong overall prediction. In the case of gradient boosted decision trees, subsequent models are discovered by applying gradient descent in the directions of the leaf nodes’ average gradient of previous models, computed regarding the error residuals of the loss function.

Proceeded as follows while considering decision trees. Begin with an initial fit of our Milling, drilling, and turning data,  $T$ , which is a constant value that optimizes our loss function  $k$ .

$$T(x) = \operatorname{argmin}_r k(z_i, r) \quad (18)$$

In the scenario of mean square error optimization, the mean of the goal values is taken:

$$T(x) = \frac{1}{m} \sum_{i=1}^m z_i \quad (19)$$

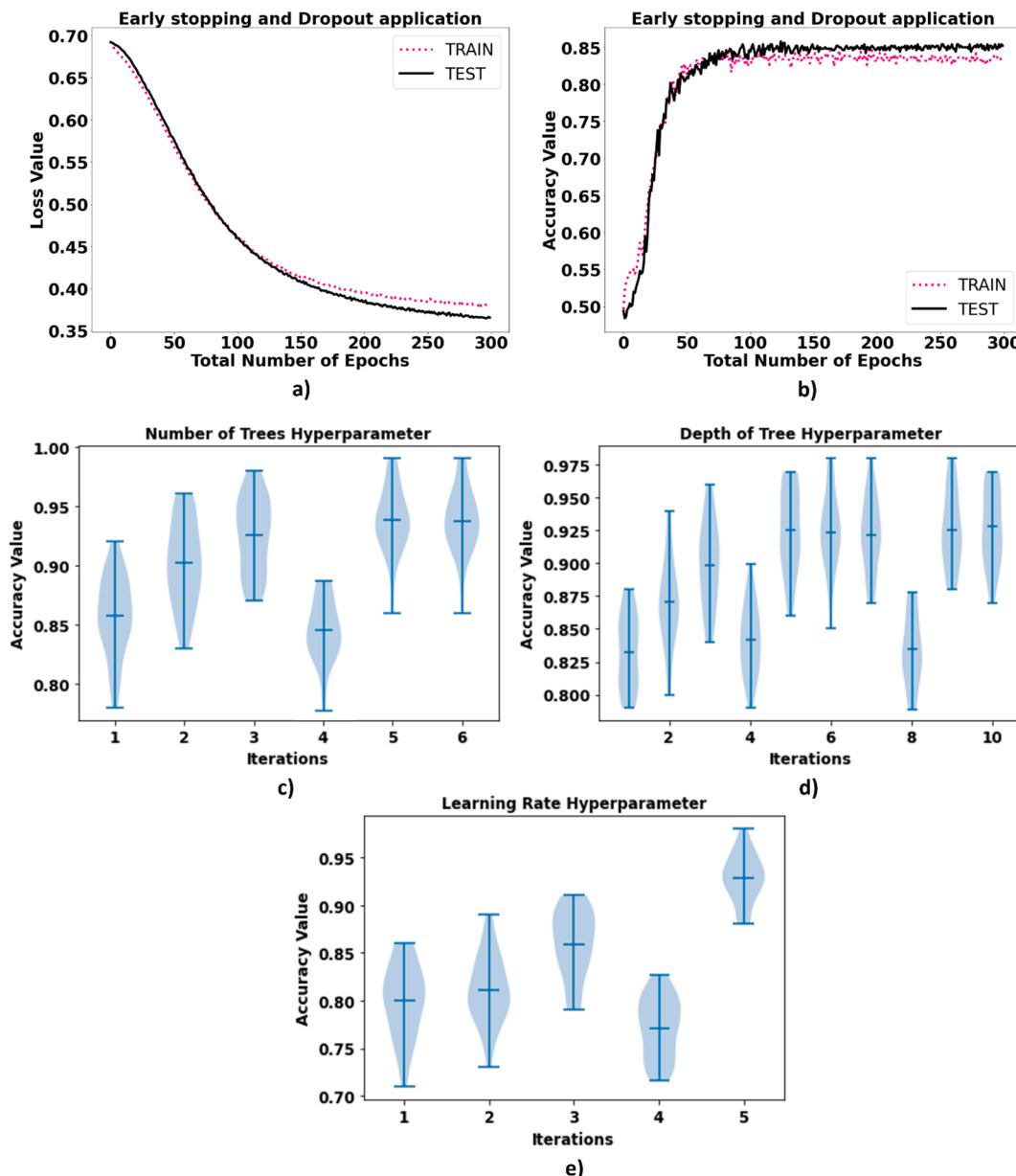
Computing the gradient or pseudo residuals of  $k$  regarding  $T$  using our original guess of  $T$ :

$$s_{i1} = -\partial k(z_i, T(x_i)) / \partial T(x_i) \quad (20)$$

To the residuals, fit the decision tree  $h_1(x)$ . This will provide the average gradient for every one of the leaf nodes by using the regression tree.

$$T_1(x) = T(x) + \lambda_1 r_1 h_1(x) \quad (21)$$

Gradient descent is used to reduce loss with each leaf by stepping in the path of the averaged gradient for the leaf nodes present in the

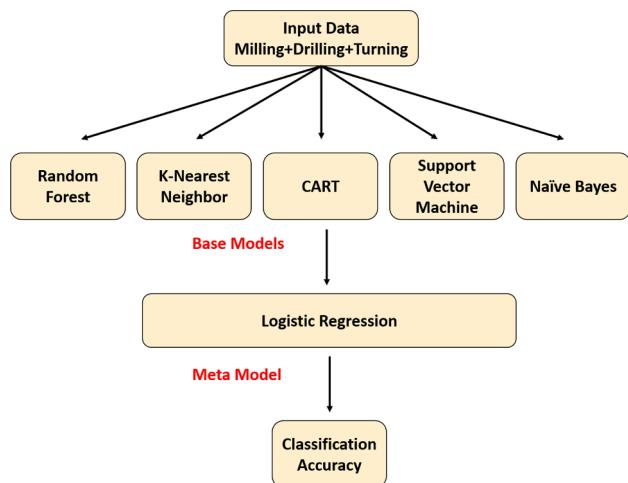


**Fig. 9.** Violin plots and accuracy curves for the LightGBM model. a) Graphical illustration of loss values b) Final accuracy curve of Model c) Accuracy values after NOT hyperparameter tuning d) Accuracy values after DOT hyperparameter Tuning e) Accuracy results after LR hyperparameter Tuning.

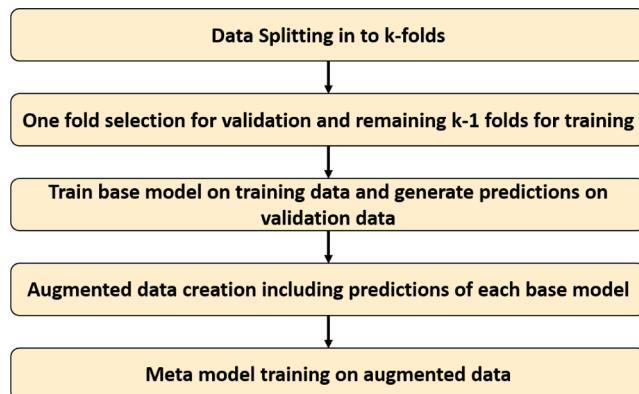
decision tree  $h_1(x)$ . A multiplier  $z_1$  determines a step size, which may be optimized via a line search. The step size is even further reduced with a learning rate of  $\lambda_1$ , resulting in a new enhanced fit of the data. The designed LightGBM Algorithm was trained and tested on the Force signals of Drilling, Milling, and Turning operations. Without hyperparameter tuning, it has given an accuracy of 85.13%. The results before and after hyperparameter optimization are shown in the figure given below. Boosting is a sequential ensemble approach in which instances that are difficult to categorize are given additional weights. This simply indicates that future learners will place a greater focus on learning misclassified data instances. The model is a weighted average of  $n$  weak learners. Following learners place a greater emphasis on untrained categorized samples. LightGBM work is based on splitting the tree with the suitable fit leaf by leaf, while other boosting techniques split the tree regarding depth or level by level rather than leaf by leaf. So, while expanding on the same leaf in Light GBM, the leaf-wise method may minimize more loss than just the level-wise technique, resulting in considerably greater accuracy as compared to the existing boosting

approaches. It is also incredibly speedy in training, thus the name “Light” is given. This model is suitable for large data set. The dataset of milling, drilling, and turning operations is very large. For example, there are more than 10,000 force values for each turning tool. Eighteen various tools are used which compiled a large amount of data.

The number of decision trees utilized in the ensemble is an essential hyperparameter for the LightGBM ensemble method. To fix and improve on the predictions produced by previous trees, decision trees are sequentially added to the model. The “n estimators” parameter specifies the number of trees. The maximum accuracy of 94.17% is achieved by the NOT hyperparameter. Another significant gradient boosting hyperparameter is changing the depth of the tree included in the ensemble model. The depth of the tree determines how customized each tree is for the training data and how generic or overfit it may be. Trees that are neither too shallow and generic nor too deep and specialized are desirable. Gradient boosting works best with trees of moderate depth, striking a compromise between skill and versatility. The “max depth” parameter controls tree depth and defaults to an undefined value since



**Fig. 10.** Block Diagram of Stacking Model.



**Fig. 11.** Basic Working of Stacking Algorithm.

the number of leaf nodes is just the default mechanism for regulating how much trees are complicated. The depth of the tree and the number of terminal nodes in the tree are the two major approaches to regulate tree complexity. Because in the mode it's exploring the number of leaves in this example, must raise the number of leaves to enable deeper trees by increasing the "num leaves" parameter. The learning rate determines how much each model contributes to the ensemble estimation. Smaller rates may need the inclusion of more decision trees in the ensemble. The "learning rate" parameter can be used to modify the learning rate. This model investigates the learning rate and examines the effect of variables ranging from 0.0001 to 1.0. It was discovered in this model that a higher learning rate leads to better performance on this dataset. We would think that adding additional trees to the ensemble for the slower learning rates will improve performance even further. With the rise in learning rate to huge values of 1.0, the overall pattern of rising model performance can be noticed. LightGBM is a histogram-based algorithm that has the capability of bucketing the continuous features by converting them into discrete bins for building histograms during the process of training. It costs the complexity of  $O(\#data * \#feature)$  for the construction of histograms and  $O(\#bin * \#feature)$  for the finding of the split point. The data can enable each set of parameters which can output different predictions hence the estimation of model parameter uncertainty is obtained by investigating the diversity of predictions. The hyperparameter tuning values are given in Table 6.

The experimentally obtained force signals were split into training and testing data. The tested results were achieved after the input of testing data into the trained model. The model was trained by using training data of milling, drilling, and turning operations. The first type of

**Table 6**  
Hyperparameter values per iteration.

Iterations	Tree depth range values	Learning Rate	Number of Trees
1	0.5	1.0	10
2	1.0	0.1	30
3	1.5	0.01	90
4	2.0	0.001	270
5	2.5	0.0001	810
6	3.0		2430
7	3.5		
8	4.0		
9	4.5		
10	5.0		

the tested results was achieved by employing the LightGBM model which is the decision tree-based gradient boosting architecture that improves model efficiency while reducing memory utilization. The GOSS (Gradient-based One Side Sampling) framework was used for the processing of the experimental results. Throughout the computing of the information gain, multiple data instances play various roles. Information gain would be greater in scenarios with greater gradients. To verify the correctness of information gain estimates, GOSS preserves these instances having substantial gradients but only drops the instances having smaller gradients as randomized. This method can result in better accurate gain estimations versus equally randomized sampling having the same desired sample rate, particularly whenever the information gain has a wide range of values. For the training dataset containing  $n$  instances ranging from  $x_1$  to  $x_n$ , and each  $x_i$  is considered as the vector of dimension  $s$  located in space  $X_s$ . The negative gradient of a loss function in terms of the designed model output is represented as  $\{g_1, g_2 \dots, g_n\}$  per iteration of gradient descent. In this GOSS technique, the instances were sorted in descending order based on their gradients' absolute values.

There are many advantages of the LightGBM model. Faster training speed and greater efficiency as this model named Light GBM employ a methodology, which speeds up the training operation. Reduced memory consumption is another advantage because LightGBM converts continuous data to discrete bins, resulting in decreased memory usage. It is better than any other boosting method in terms of accuracy. It generates significantly more complicated trees by using a leaf-wise split technique, which is the primary contributor in obtaining greater accuracy. However, early stopping and dropout techniques were applied to avoid overfitting. This overfitting problem can also be prevented by adjusting the max depth option in the LightGBM algorithm. Compatibility with Big Data is the major benefit of the LightGBM model as It is capable of performing similarly well with large datasets while requiring significantly less training time than XGBOOST. Also, parallel learning is encouraged in this model. The proposed LightGBM model is developed based on several techniques such as PCA, SSA, Early stopping, dropout, and hyperparameter tuning. The developed model has increased training efficiency and speed, reduce memory utilization, Improved accuracy, asynchronous as well as GPU learning support, and has the ability to manage massive amounts of data. LightGBM is a histogram-based technique that partitions continuous data into discrete bins, resulting in quicker training and much more economical memory consumption. It can split the tree based on the leaf-wise technique with the greatest fit since it is centered on the algorithm of the decision tree, unlike other boosting methods split the tree based on depth-wise technique instead of leaf-wise. So, while expanding on the same leaf in a model, the leaf-wise approach may minimize loss as compared to the level-wise technique, resulting in considerably superior accuracy which other boosting techniques can seldom accomplish. LightGBM can accomplish the optimization in parallel learning such as future parallel, data-parallel, and voting parallel. LightGBM employs a unique approach of Gradient-Based One Side Sampling to screen out data instances to determine a split value.

The important parameters of models are num\_leaves , min\_child\_samples , max\_depth, scale\_pos\_weight, max\_bin , min\_child\_weight, min\_child\_samples , bagging\_fraction , feature\_fraction , learning\_rate and num\_leaves. The parameters are already built-in into the algorithm and values can be set as per the requirement to get the desired output. LightGBM is a histogram-based algorithm that has the capability of bucketing the continuous features by converting them into discrete bins for building histograms during the process of training. It costs the complexity of  $O(\#data * \#feature)$  for the construction of histograms and  $O(\#bin * \#feature)$  for the finding of the split point. The data can enable each set of parameters which can output different predictions hence the estimation of model parameter uncertainty is obtained by investigating the diversity of predictions.

## 2.8. Hyperparameter tuned model based on stacking

Stacking is the process of utilizing a machine learning model to learn how to effectively aggregate predictions from various ensemble members. A weighted total prediction is produced by weighing the contributions of each ensemble member in the estimation. This permits the greater weight to be placed on algorithms that work substantially better and much less weight to be placed on algorithms that perform poorly but still have prediction skills. The weight allocated to each contributor algorithm, as well as the efficiency of each model on the training or holdout datasets, must be learned. This method is generalized by stacking, which allows any machine learning model to learn how to optimally aggregate predictions from contributing models. Stacking is intended to boost modeling efficiency. Stacking is the machine learning technique to train several base models such as LR, RF, CART, NB, SVM, and KNN to estimate the outcome variable in a deep learning task while simultaneously the *meta*-model utilizes the outputs of each base model for the better estimation of target variable's value. Moreover, stacking increases accuracy as compared to the accuracy values of the base models. The stacking model was developed by designing six different algorithms such as Logistic Regression, Random Forest, Classification and Regression Tree, Naïve Bayes, Support Vector Machine, and k-Nearest Neighbor. Stacking employs, a *meta*-learning strategy to find the optimum way to integrate estimation from six underlying machine learning methods. The advantage of it is that it may combine the skills of several high-performing algorithms on a classification job to create estimates that outperform any one model within stacking. The best method for producing the training data for the input into *meta*-model is to employ the k-fold cross-validation technique of all base algorithms, with the out-of-fold estimation serving as the foundation for the *meta*-training model's data. The *meta*-model's training set can provide inputs parameters to the base algorithms, such as training data input components. This can give more insight to a *meta*-model in terms of how to perfectly integrate the meta models' predictions.

The model that aggregates the predictions is known as the *meta*-model, and the ensemble participants are known as base models. Significantly, the *meta*-model is not trained in the same manner as the base models are. The predictions generated by the base models, not by input values from the dataset, are used as input into the *meta*-model. The predictions generated by the base models are used for training the *meta*-model which is not utilized to train the base models, therefore they are out of sample. The dataset, for instance, can be divided into validation, train as well as test datasets. After that, each base model may be fit to the training set and predicted on the validation dataset. The validation set predictions are then utilized for training the *meta*-model. The automatic hyperparameter tuning of stacking is done by using the GridSearchCV technique. With regard to the specific condition of having a greater number of features "p" than the training samples "n", there will be the risk of overfitting of the model. SVMs, like several linear models, is primarily supported by empirical risk minimization, which results in this type of optimization.

$$\min \sum l(x_i, y_i, w) + \lambda r(w) \quad (22)$$

Where  $l$  represents the loss function (in SVMs, Hinge-loss) whereas  $r$  denotes the regularization function. The SVM is a quadratic  $l_2$ -regularized linear model. This prevents "large coefficients," as they are known in regression terminology because the magnitudes of the coefficients are penalized in optimization. In the  $p > n$  case, regularization provides for one a unique solution. Where  $p$  denotes the number of features and  $n$  represents the number of training samples. When  $p \gg n$ , the difficulty is that bias created by regularization might be so greater for training samples that the proposed model underperforms. That is not to imply that SVMs cannot be employed in that case because the model will still work and give output accuracy. The kernel is the function in SVM that aids in problem-solving. They offer shortcuts to avoid doing difficult calculations. The good thing about them is that they allow to accomplish greater dimensionality and execute smooth computations. The sigmoid kernel is used in the article for SVM.

$$K(x, y) = \tanh(\gamma x^T y + r) \quad (23)$$

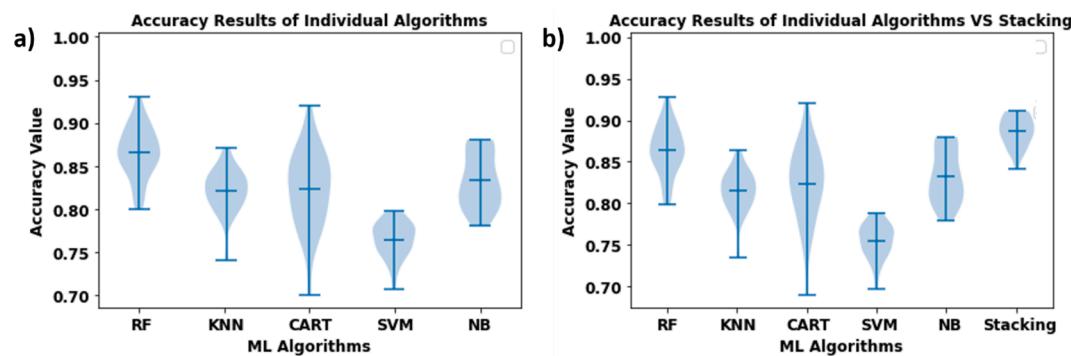
It is typically used in neural networks. This kernel is analogous to a neural network's 2-layer perceptron model that serves as the activation function of neurons. The  $C$  parameter imposes a penalty for each incorrectly categorized data value. Because the penalty of misclassified values is minimal when  $c$  is small, the decision boundary with such a big margin was chosen at the price of the higher percentage of misclassifications. If  $c$  is big, SVM attempts to reduce the number of misclassified instances as a consequence of the greater penalty, resulting in the decision boundary having a narrower margin. The penalty is not the same for all cases of misclassification. It is proportional to a distance from the decision border. When utilizing the Gaussian RBF kernel, then only Gamma is employed. A sigmoid kernel is used in the article. The values of  $0.1 < C < 100$  are utilized when choosing the  $C$  parameter and Grid Search was employed.

There is a total of five base models which include Random forest, K-Nearest Neighbor, Classification and Regression Tree, Support Vector Machine, and Naïve Bayes. However, the *meta*-model includes Logistic Regression. Stacking is the process of training several base models to estimate the target variable in a supervised learning task while simultaneously the *meta*-learning model employs the predictions of each base model to anticipate the target variable's value.

The effectiveness of stacking is found in the last phase, in which the *meta*-model can learn the weaknesses and strengths of each base model and effectively integrate their predictions to generate the final result. Stacking, like all other machine learning algorithms, has advantages and downsides. The advantage of this technique is that the model performance can be improved by stacking. By integrating the predictions of many models, stacking decreases variance and provides a more robust model. However, the limitation of the stacking model is that Stacked models can take much longer to train and consume more memory than simpler models. Predictions generated with layered models are often slower and are more computationally costly.

The results demonstrated in Fig. 12 in the form of accuracy values and comparison of results of the individual base model is done with stacking model. Base models were retrained using different training datasets in each iteration therefore the implementation of a k-fold cross-validation split guarantees that all base models were providing predictions on machining force signals' data. Each base model is trained individually, and the *meta*-learning model learns to estimate the final output using the predictions of all base models and the original data. k-fold cross-validation is a methodology in which the dataset for a machine learning task is divided into k-folds or different subsets, and the model is assessed repeatedly over all k-folds. Each iteration includes one fold for assessment and the remaining k-1 folds for training the model.

The current model of this research study based on LightGBM is compared with the major Machine learning models of previously published research studies based on the four important parameters as



**Fig. 12.** Violin Plots for a) Final accuracy of Individuals based on machining data b) Comparison of Accuracy of Individual algorithms with stacking model.

mentioned in Table 7. The comparison is built based on the principal results of the current model and the major contributions of this research study. The major contribution of this research study is that it has overcome the shortcoming of tool wear Machine learning models of previous research studies. The four important parameters demonstrated in Table 7 are not supported by previously published Machine learning (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 machine learning models that occurs when the algorithm selects fluctuations in the training data and learns them as concepts. When a designed model fits a greater amount of data than what it requires, it begins to learn noisy data as well as erroneous points in the data. As a consequence, the model's efficiency, as well as accuracy, are reduced. The algorithm of dropout and early stopping is applied to overcome the error of overfitting. When the model is being trained, it can monitor how it performs based on each iteration. The process can be continued until the iterations can improve the performance of the designed model. Subsequently, the developed model can overfit the data used for training as generalization worsens for each iteration. Early stopping essentially implies terminating the training phase before the designed model reaches the point when it begins to start overfitting the data used for training. In addition to the usage of early stopping, the dropout technique was also applied to eliminate overfitting. Dropout is

the regularization approach that approximates concurrent training of a significant number of neurons with various architectures. Throughout the training, certain layer output is disregarded or “dropped out” at random. It has the impact of having the layer to be regarded and it does have a distinct number of nodes and connectedness to the preceding layer. Every update to the layer throughout training, in practice, is made with a distinct “view” of the specified layer. Dropout can be performed on any of the network’s hidden layers, and the viewable or input layer. For the output layer, it is not utilized. Dropout pertains to disregarding neurons throughout the training stage of a particular group of neurons. By “ignoring,” means that these neurons are not taken into account during a certain forward as well as backward processing. Another significant contribution of the article is dimensionality reduction. Principal component analysis was applied for reducing the dimensions from fifteen to three and Singular Spectrum Analysis was used to extract fifteen features from the force signals of milling, drilling, and turning operations which were obtained by mounting the dynamometer on the CNC machine. The third quality of the current research study is that the designed model is operational for three machining processes i.e. drilling, milling, and turning. Another major contribution of this article is that it has detected the tool wear severity. Total 18 tools were used and 6 for each machining operation (milling, drilling, or turning). For each machining operation, there were five levels of tool wear based on the flank wear land width of each tool. The tool with the least value of flank

**Table 7**  
Comparison of the current study for Tool Wear Prediction with existing literature ✓ Support ✗ Doesn’t Support.

Models	Overcome Overfitting and Reducing Dimensions	Model Operational for three Machining Operations	Efficient, Fast Processing and use less memory	Model can determine five levels of tool wear for each Milling / Drilling / Turning tool
Shortcoming of Previous research Studies for Tool Wear Detection during Machining	XGBOOST[36] LSTM[37,38] RNN[39] NB [40] TL[41] CNN[42,43] CNN + BLSTM[44] CNN + SBULSTM[45]	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗
RF[46] SVM[47,48] DBN[49] AE[50,51] MCLSTM[52] BLSTM[53,54] ANN[55,56] KNN [12] MCGPR[57] Adaboost-DT[58] DT[59] SVR[60]	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗	✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗ ✗
Overcome Shortcoming By Current research Study	Hyperparameter Tuned model based on PCA, SSA, early stopping, and dropout methodologies of Current Study	✓	✓	✓

wear land width was given the grade-1 worn tool whereas the tool with the extreme tool wear having the greatest value of flank wear land width had been assigned the grade-5 worn tool. So this strategy was neglected by the previously published research studies.

### 3. Conclusion

The novel tool wear estimation methodology is developed based on designing several algorithms such as LightGBM, early stopping, dropout, PCA, SSA, and one-hot encoding. Two different models were designed based on LightGBM and Stacking to estimate tool wear level. LightGBM Model has performed better as compared to the stacking model thus LightGBM model is the best suitable final choice of this research study to work for three machining operations. Forces signals for the training of models were obtained by conducting experiments based on turning, milling, and drilling procedures executed on a dynamometer installed on a CNC machine with a tool magazine storing 12 various types of drilling and milling tools, yielding six tools for each machining operation. For three machining processes, a total of 18 tools were utilized. Milling, drilling, and turning procedures were integrated based on the fifteen characteristics retrieved from the force signals of the three machining operations and fed into the built models. The first sharp-edged tool was identified as the normal tool, while the next five worn tools were categorized based on increasing flank wear land width values. The least worn tool with the shortest value of flank wear land width received a grade of 1, and the grade level increases as the value of flank wear land width increases. In contrast, the severely worn tool with the highest flank wear land width value obtained a wear level of five.

1. The designed multi-target classification model LightGBM has overcome the major shortcomings of previously published tool wear models, such as high dimensionality and overfitting of tool wear models, as well as high memory usage and slow processing speed during tool wear estimation in machining operations. The prefix of Light GBM is “light” due to its high processing speed. LightGBM model can handle huge amounts of data while using less memory. The model has fast processing speed and efficiency as compared to the models of existing literature. The dimensionality reduction from fifteen dimensions to three was accomplished by designing the algorithm of Principal Component Analysis concerning experimental data. Singular Spectrum Analysis has extracted features and removed force signals’ noise. Dropout and early stopping have avoided overfitting. One-hot encoding was used to convert the target variables into binary form.
2. The developed models’ main application is the possible installation in the aerospace and manufacturing industries requiring milling, turning, and drilling processes. As worn tools have a detrimental impact on the machined surface integrity, this model will increase the production rate as well as the dimensional correctness. This system will send input to the CNC machine regarding a worn tool, as well as the level of the tool wear. The test tool will be deemed replaceable if the system classifies it as level-4 or level –5; otherwise, it will have negative consequences. However, in the case of a level –1 worn tool, the tool can be utilized for a short period while preventive steps are taken to prevent further tool wear progression
3. Future work related to this research article involves mechanistic modelling of twelve tools. i.e. all grades of worn tools of drilling and milling operations. The separate analytical model for each tool will be developed based on mathematical equations and calculations to get tangential and axial drilling coefficients of each worn tool. The coefficients for one type of used tool will differ from one another. The six important topics will be included in the mechanistic model of each tool which are drilling forces modelling for tooltip’s outer corner, drilling forces modelling for tooltip’s chisel edge, drilling forces modelling for tooltip’s cutting lips, drilling torque modelling

for tooltip’s outer corner, drilling torque modelling for tooltip’s chisel edge, drilling torque modelling for tooltip’s cutting lips

### CRediT authorship contribution statement

**Jawad Mahmood:** Conceptualization, Investigation, Methodology, Software, Writing – original draft. **Ghulam-e Mustafa:** Data curation, Writing – review & editing. **Muhammad Ali:** Visualization, Investigation, Writing – review & editing.

### Declaration of Competing Interest

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

### References

- [1] M. Marani, M. Zeinali, S. Farahany, C.K. Mechefske, Neuro-fuzzy based predictive model for cutting force in CNC turning process of Al-Si-Cu cast alloy using modifier elements, *SN Applied Sciences* 3 (2021) 1–11, <https://doi.org/10.1007/s42452-020-03980-9>.
- [2] M. Marani, M. Zeinali, J. Kouam, V. Songmene, C.K. Mechefske, Prediction of cutting tool wear during a turning process using artificial intelligence techniques, *The International Journal of Advanced Manufacturing Technology* 111 (1–2) (2020) 505–515, <https://doi.org/10.1007/s00170-020-06144-6>.
- [3] M. Marani, V. Songmene, M. Zeinali, J. Kouam, Y. Zedan, Neuro-fuzzy predictive model for surface roughness and cutting force of machined Al-20 Mg 2 Si-2Cu metal matrix composite using additives, *Neural Computing and Applications* 32 (12) (2020) 8115–8126, <https://doi.org/10.1007/s00521-019-04314-6>.
- [4] M.M. Barzani, E. Zalnezhad, A.A. Sarhan, S. Farahany, S. Ramesh, Fuzzy logic based model for predicting surface roughness of machined Al-Si-Cu-Fe die casting alloy using different additives-turning, *Measurement* 61 (2015) 150–161, <https://doi.org/10.1016/j.measurement.2014.10.003>.
- [5] F.C. Zegarra, J. Vargas-Machuca, A.M. Coronado, Tool wear and remaining useful life (RUL) prediction based on reduced feature set and Bayesian hyperparameter optimization, *Production Engineering* (2021) 1–16, <https://doi.org/10.1007/s11740-021-01086-8>.
- [6] J.L. Ferrando Chacón, T. Fernández de Barrena, A. García, M. Sáez de Buruaga, X. Badioli, J. Vicente, A Novel Machine Learning-Based Methodology for Tool Wear Prediction Using Acoustic Emission Signals, *Sensors* 21 (2021) 5984, <https://doi.org/10.3390/s21175984>.
- [7] T. Gittler, M. Glässer, E. Özürk, M. Lüthi, L. Weiss, K. Wegener, International Conference on Advanced and Competitive Manufacturing Technologies milling tool wear prediction using unsupervised machine learning, *The International Journal of Advanced Manufacturing Technology*, (2021) 1–14. <https://doi.org/10.3929/ethz-b-000494836>.
- [8] P. Twardowski, M. Tabaszewski, M. Wiciak-Pikuła, A. Felusiak-Czyryca, Identification of tool wear using acoustic emission signal and machine learning methods, *Precision Engineering* 72 (2021) 738–744, <https://doi.org/10.1016/j.precisioneng.2021.07.019>.
- [9] V. Parwal, B. Rout, Machine learning based approach for process supervision to predict tool wear during machining, *Procedia CIRP* 98 (2021) 133–138, <https://doi.org/10.1016/j.procir.2021.01.018>.
- [10] Z. He, T. Shi, J. Xuan, T. Li, Research on tool wear prediction based on temperature signals and deep learning, *Wear* 478–479 (2021) 203902, <https://doi.org/10.1016/j.wear.2021.203902>.
- [11] P. Horton, K. Nakai, Better Prediction of Protein Cellular Localization Sites with the k Nearest Neighbors Classifier, *Ismb* (1997) 147–152.
- [12] F. Penedo, R.E. Haber, A. Gajate, R.M. del Toro, Hybrid Incremental Modeling Based on Least Squares and Fuzzy k-\$NN\$ for Monitoring Tool Wear in Turning Processes, *IEEE transactions on industrial informatics* 8 (4) (2012) 811–818.
- [13] F. Shi, H. Cao, X. Zhang, X. Chen, A Reinforced k-Nearest Neighbors Method With Application to Chatter Identification in High-Speed Milling, *IEEE Transactions on Industrial Electronics* 67 (12) (2020) 10844–10855.
- [14] M.R. Islam, A.R.M. Kamal, N. Sultana, R. Islam, M.A. Moni, Detecting depression using k-nearest neighbors (knn) classification technique, 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2), IEEE, 2018, pp. 1–4. <https://doi.org/10.1109/ic4me2.2018.8465641>.
- [15] Z. Zhang, Introduction to machine learning: k-nearest neighbors, *Annals of translational medicine* 4 (11) (2016).
- [16] M.-L. Zhang, Z.-H. Zhou, A k-nearest neighbor based algorithm for multi-label classification, 2005 IEEE international conference on granular computing, IEEE, 2005, pp. 718–721. <https://doi.org/10.1109/grc.2005.1547385>.
- [17] H. Li, Y. Wang, P. Zhao, X. Zhang, P. Zhou, Cutting tool operational reliability prediction based on acoustic emission and logistic regression model, *Journal of Intelligent Manufacturing* 26 (5) (2015) 923–931, <https://doi.org/10.1007/s10845-014-0941-4>.

- [18] W. Cheng, E. Hüllermeier, Combining instance-based learning and logistic regression for multilabel classification, *Machine Learning* 76 (2009) 211–225, [https://doi.org/10.1007/978-3-642-04180-8\\_6](https://doi.org/10.1007/978-3-642-04180-8_6).
- [19] R. Tomioka, K. Aihara, K.-R. Müller, Logistic regression for single trial EEG classification, *Advances in neural information processing systems* 19 (2007) 1377, <https://doi.org/10.7551/mitpress/7503.003.0177>.
- [20] B. Chen, X. Chen, B. Li, Z. He, H. Cao, G. Cai, Reliability estimation for cutting tools based on logistic regression model using vibration signals, *Mechanical Systems and Signal Processing* 25 (7) (2011) 2526–2537, <https://doi.org/10.1016/j.ymssp.2011.03.001>.
- [21] I. Kurt, M. Ture, A.T. Kurum, Comparing performances of logistic regression, classification and regression tree, and neural networks for predicting coronary artery disease, *Expert systems with applications* 34 (1) (2008) 366–374, <https://doi.org/10.1016/j.eswa.2006.09.004>.
- [22] A. Subasi, E. Erçelebi, Classification of EEG signals using neural network and logistic regression, *Computer methods and programs in biomedicine* 78 (2) (2005) 87–99, <https://doi.org/10.1016/j.cmpb.2004.10.009>.
- [23] N. Wang, G. Zhang, W. Pang, L. Ren, Y. Wang, Novel monitoring method for material removal rate considering quantitative wear of abrasive belts based on LightGBM learning algorithm, *The International Journal of Advanced Manufacturing Technology* 114 (11–12) (2021) 3241–3253, <https://doi.org/10.1007/s00170-021-06988-6>.
- [24] E.-A. Minastireanu, G. Mesnita, Light gbm machine learning algorithm to online click fraud detection, *J. Inform. Assur. Cybersecur* (2019) 1–12, <https://doi.org/10.5171/2019.263928>.
- [25] M. Tang, Q. Zhao, S.X. Ding, H. Wu, L. Li, W. Long, B. Huang, An improved lightGBM algorithm for online fault detection of wind turbine gearboxes, *Energies* 13 (2020) 807, <https://doi.org/10.3390/en13040807>.
- [26] N. Wang, G. Zhang, L. Ren, W. Pang, Y. Wang, Vision and sound fusion-based material removal rate monitoring for abrasive belt grinding using improved LightGBM algorithm, *Journal of Manufacturing Processes* 66 (2021) 281–292, <https://doi.org/10.1016/j.jmapro.2021.04.014>.
- [27] S. Virendra Dahe, G. Sai Manikandan, R. Jegadeeshwaran, G. Sakthivel, J. Lakshmi pathi, Tool condition monitoring using Random forest and FURIA through statistical learning, *Materials Today: Proceedings* 46 (2021) 1161–1166, <https://doi.org/10.1016/j.mtproc.2021.02.059>.
- [28] V. Rodriguez-Galiano, M. Chica-Olmo, F. Abarca-Hernandez, P.M. Atkinson, C. Jeganathan, Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture, *Remote Sensing of Environment* 121 (2012) 93–107, <https://doi.org/10.1016/j.rse.2011.12.003>.
- [29] D. Wu, C. Jennings, J. Terpenny, R. Gao, S. Kumara, Data-driven prognostics using random forests: Prediction of tool wear, *International Manufacturing Science and Engineering Conference*, American Society of Mechanical Engineers, 2017, pp. V003T004A048, <https://doi.org/10.1115/MSEC2017-2679>.
- [30] M.S. Alam, S.T. Vuong, Random forest classification for detecting android malware, *2013 IEEE international conference on green computing and communications and IEEE Internet of Things and IEEE cyber, physical and social computing*, IEEE, 2013, pp. 663–669, <https://doi.org/10.1109/GreenCom-iThings-CPSCom.2013.122>.
- [31] P.O. Gislason, J.A. Benediktsson, J.R. Sveinsson, Random forest classification of multisource remote sensing and geographic data, *IGARSS 2004. 2004 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 2004, pp. 1049–1052, <https://doi.org/10.1109/IGARSS.2004.1368591>.
- [32] J. Yu, X. Cheng, L. Lu, B. Wu, A Machine Vision Method for Measurement of Machining Tool Wear, *Measurement* 182 (2021) 109683, <https://doi.org/10.1016/j.measurement.2021.109683>.
- [33] A.D. Patange, R. Jegadeeshwaran, A machine learning approach for vibration-based multipoint tool insert health prediction on vertical machining centre (VMC), *Measurement* 173 (2021) 108649, <https://doi.org/10.1016/j.measurement.2020.108649>.
- [34] M. Kuntoglu, H. Saglam, Investigation of signal behaviors for sensor fusion with tool condition monitoring system in turning, *Measurement* 173 (2021) 108582, <https://doi.org/10.1016/j.measurement.2020.108582>.
- [35] K.M. Fong, X. Wang, S. Kamaruddin, M.-Z. Ismaili, Investigation on universal tool wear measurement technique using image-based cross-correlation analysis, *Measurement* 169 (2021) 108489, <https://doi.org/10.1016/j.measurement.2020.108489>.
- [36] M.S. Alajmi, A.M. Almeshal, Predicting the tool wear of a drilling process using novel machine learning XGBoost-SDA, *Materials* 13 (2020) 4952, <https://doi.org/10.3390/ma13214952>.
- [37] Z. Zhang, J. Lu, G. Zhou, X. Liao, Research on tool wear prediction based on LSTM and ARIMA, *Proceedings of the 2018 international conference on big data engineering and technology*, 2018, pp. 73–77, <https://doi.org/10.1145/3297730.3297732>.
- [38] M. Marani, M. Zeinali, V. Songmene, C.K. Mechesfske, Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling, *Measurement* 177 (2021) 109329, <https://doi.org/10.1016/j.measurement.2021.109329>.
- [39] X. Wang, Y. Huang, Optimized recurrent neural network-based tool wear modeling in hard turning, *Trans. NAMRI/SME* 37 (2009) 213–220.
- [40] J. Karandikar, T. McLeay, S. Turner, T. Schmitz, Tool wear monitoring using naive Bayes classifiers, *The International Journal of Advanced Manufacturing Technology* 77 (2015) 1613–1626, <https://doi.org/10.1007/s00170-014-6560-6>.
- [41] J. Li, J. Lu, C. Chen, J. Ma, X. Liao, Tool wear state prediction based on feature-based transfer learning, *The International Journal of Advanced Manufacturing Technology* 113 (11–12) (2021) 3283–3301, <https://doi.org/10.1007/s00170-021-06780-6>.
- [42] X. Xu, J. Wang, W. Ming, M. Chen, Q. An, In-process tap tool wear monitoring and prediction using a novel model based on deep learning, *The International Journal of Advanced Manufacturing Technology* 112 (1–2) (2021) 453–466, <https://doi.org/10.1007/s00170-020-06354-y>.
- [43] X. Xu, J. Wang, B. Zhong, W. Ming, M. Chen, Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism, *Measurement* 177 (2021) 109254, <https://doi.org/10.1016/j.measurement.2021.109254>.
- [44] J. Ma, D. Luo, X. Liao, Z. Zhang, Y.i. Huang, J. Lu, Tool wear mechanism and prediction in milling TC18 titanium alloy using deep learning, *Measurement* 173 (2021) 108554, <https://doi.org/10.1016/j.measurement.2020.108554>.
- [45] Q. An, Z. Tao, X. Xu, M. El Mansori, M. Chen, A data-driven model for milling tool remaining useful life prediction with convolutional and stacked LSTM network, *Measurement* 154 (2020) 107461, <https://doi.org/10.1016/j.measurement.2019.107461>.
- [46] D. Wu, C. Jennings, J. Terpenny, R.X. Gao, S. Kumara, A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests, *Journal of Manufacturing Science and Engineering* 139 (2017), <https://doi.org/10.1115/1.4036350>.
- [47] Y. Qian, J. Tian, L. Liu, Y. Zhang, Y. Chen, A tool wear predictive model based on SVM, *2010 Chinese control and decision conference*, IEEE, 2010, pp. 1213–1217, <https://doi.org/10.1109/CCDC.2010.5498161>.
- [48] Y.u. Liang, S. Hu, W. Guo, H. Tang, Abrasive Tool Wear Prediction Based on an Improved Hybrid Difference Grey Wolf Algorithm for Optimizing SVM, *Measurement* 187 (2022) 110247, <https://doi.org/10.1016/j.measurement.2021.110247>.
- [49] Y. Chen, Y.i. Jin, G. Jiri, Predicting tool wear with multi-sensor data using deep belief networks, *The International Journal of Advanced Manufacturing Technology* 99 (5–8) (2018) 1917–1926, <https://doi.org/10.1007/s00170-018-2571-z>.
- [50] G. Hao, Z. Kunpeng, Pyramid LSTM auto-encoder for tool wear monitoring, *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, IEEE, 2020, pp. 190–195, <https://doi.org/10.1109/CASE48305.2020.9217015>.
- [51] J. Ou, H. Li, G. Huang, G. Yang, Intelligent analysis of tool wear state using stacked denoising autoencoder with online sequential-extreme learning machine, *Measurement* 167 (2021) 108153, <https://doi.org/10.1016/j.measurement.2020.108153>.
- [52] H. Qiao, T. Wang, P. Wang, A tool wear monitoring and prediction system based on multiscale deep learning models and fog computing, *The International Journal of Advanced Manufacturing Technology* 108 (7–8) (2020) 2367–2384, <https://doi.org/10.1007/s00170-020-05548-8>.
- [53] C.-G. Huang, X. Yin, H.-Z. Huang, Y.-F. Li, An enhanced deep learning-based fusion prognostic method for RUL prediction, *IEEE Transactions on Reliability* 69 (3) (2020) 1097–1109, <https://doi.org/10.1109/TR.2019.2948705>.
- [54] C. Liu, L. Zhu, A two-stage approach for predicting the remaining useful life of tools using bidirectional long short-term memory, *Measurement* 164 (2020) 108029, <https://doi.org/10.1016/j.measurement.2020.108029>.
- [55] P. Palanisamy, I. Rajendran, S. Shanmugasundaram, Prediction of tool wear using regression and ANN models in end-milling operation, *The International Journal of Advanced Manufacturing Technology* 37 (1–2) (2008) 29–41, <https://doi.org/10.1007/s00170-007-0948-5>.
- [56] K.V. Rao, B. Murthy, N.M. Rao, Prediction of cutting tool wear, surface roughness and vibration of work piece in boring of AISI 316 steel with artificial neural network, *Measurement* 51 (2014) 63–70, <https://doi.org/10.1016/j.measurement.2014.01.024>.
- [57] C. Zhang, W. Wang, H. Li, Tool wear prediction method based on symmetrized dot pattern and multi-covariance Gaussian process regression, *Measurement* (2021) 110466, <https://doi.org/10.1016/j.measurement.2021.110466>.
- [58] X. Li, X. Liu, C. Yue, S. Liu, B. Zhang, R. Li, S.Y. Liang, L. Wang, A data-driven approach for tool wear recognition and quantitative prediction based on radar map feature fusion, *Measurement* 185 (2021) 110072, <https://doi.org/10.1016/j.measurement.2021.110072>.
- [59] T. Mohanraj, J. Yerchuru, H. Krishnan, R.S. Nithin Aravind, R. Yameni, Development of tool condition monitoring system in end milling process using wavelet features and Hölder's exponent with machine learning algorithms, *Measurement* 173 (2021) 108671, <https://doi.org/10.1016/j.measurement.2020.108671>.
- [60] Y. Yang, Y. Guo, Z. Huang, N. Chen, L. Li, Y. Jiang, N. He, Research on the milling tool wear and life prediction by establishing an integrated predictive model, *Measurement* 145 (2019) 178–189, <https://doi.org/10.1016/j.measurement.2019.05.009>.