International Journal of Mechatronics and Manufacturing Systems

ISSN online: 1753-1047 - ISSN print: 1753-1039
https://www.inderscience.com/ijmms

Performance improvement techniques for neural networks in tool condition monitoring

Arup Dey, Nita Yodo

DOI: 10.1504/IJMMS.2022.10048836

Article History:

Received:	14 October 2021
Last revised:	02 July 2022
Accepted:	14 February 2022
Published online:	16 August 2022

Performance improvement techniques for neural networks in tool condition monitoring

Arup Dey* and Nita Yodo

Department of Industrial and Manufacturing Engineering,
North Dakota State University,
Fargo, ND 58102, USA
Email: arup.dey@ndsu.edu
Email: nita.yodo@ndsu.edu
*Corresponding author

Abstracts: The performance of data-driven algorithms in tool condition monitoring often depends on the combinations of different factors such as data quality, input dimensions, and model architecture. Several performance improvement techniques such as data denoising, feature selection, and regularisation techniques are known to enhance prediction accuracy. Moreover, selecting model architecture and tuning hyperparameters also significantly impact the prediction performance. Although the prediction accuracy of a data-driven method can be improved using these techniques, their importance is rarely discussed for tool condition monitoring. In this paper, the importance of various performance improvement techniques is extensively analysed by applying them to a CNC milling machine dataset for tool wear prediction. The investigation results and performance measurement metrics showed data denoising techniques, feature reduction techniques, and regularisation methods improved prediction accuracy up to around 55%. The selection of techniques for improving the accuracy depends on the nature of a dataset and applied algorithms.

Keywords: tool wear; condition monitoring; neural networks; model performance; denoising; regularisation; feature extraction; feature selection.

Reference to this paper should be made as follows: Dey, A. and Yodo, N. (2022) 'Performance improvement techniques for neural networks in tool condition monitoring', *Int. J. Mechatronics and Manufacturing Systems*, Vol. 15, Nos. 2/3, pp.107–126.

Biographical notes: Arup Dey received his BSc (Engg.) in Industrial and Production Engineering at Shahjalal University of Science and Technology, in 2013 and MSc in Industrial and Production Engineering from Bangladesh University of Engineering and Technology in 2018. His MSc research area was in quantitative optimisation under uncertainty. Additionally, in Bangladesh, he worked as an industrial engineer in a renowned export-based apparel manufacturing company. Currently, he is a PhD student at North Dakota State University, and his research interest centres around quantitative optimisation, data analysis, and uncertainty quantification in additive manufacturing.

Nita Yodo is an Assistant Professor in the Department of Industrial and Manufacturing Engineering at North Dakota State University (NDSU). She received her PhD in Industrial and Manufacturing Engineering from Wichita State University in 2017. Currently, she is the research director for the Advanced System Engineering Laboratory (ASEL) and Center of Quality,

Reliability, and Maintainability Engineering (CQRME) at NDSU. Her research areas involve complex system modeling, predictive analysis for decision-making, and process optimisation towards sustainability, with a wide range of applications in critical infrastructure networks, advanced manufacturing processes, and defense industries.

1 Introduction

A tool failure may occur at any time during machining. Tool condition monitoring is necessary to predict tool failure, improve product quality, reduce the unscheduled shutdown of a machine, and avoid damage to a machining centre. One of the most common types of tool failures is tool flank wear, which affects the surface quality of a workpiece (Dhanasekaran et al., 2010). Failures due to flank wear can be monitored by using a wide range of condition monitoring sensors. To determine an optimal tool changing time, optimisation methods (Su and Wang, 2014) or data-driven prediction analysis (Mosallam et al., 2016; Wu et al., 2018; Dey and Yodo, 2021) can be applied.

Prognostics and health management (PHM) is an approach in condition monitoring through non-destructive assessment. A general PHM consists of several steps, such as sensor data collection, data processing, diagnostics for condition assessment, prognostics, and decision support (Das et al., 2011; Xu and Xu, 2011). In machining processes, PHM is widely applied to estimate tool wear as well as the remaining useful life (RUL) of a cutting tool (Cai et al., 2020; Wu et al., 2018). Generally, PHM methods can be categorised into three types: model-based method, data-driven methods, and hybrid methods (Peng et al., 2010).

The model-based method uses the physical phenomena of cutting tools to develop a mathematical prediction model. A model-based method can predict tool wear accurately if the physical phenomena of a system are modelled precisely. Therefore, an in-depth understanding of the physical behaviour of a system is essential (Li et al., 2018). In data-driven methods, different learning algorithms can be used to develop a mathematical model from historical sensor data to predict the cutting tool's condition (Li et al., 2018). The development of a precise mathematical model depends on the characteristics of the datasets. A hybrid method combines the model-based method and data-driven method and aims to get advantages of both the above-described methods. Due to the current advancements in computing capacity, the data-driven approaches have several advantages over the model-based methods (Baraldi et al., 2013; Khumprom and Yodo, 2019).

Many researchers have investigated various machine learning and deep learning algorithms for tool wear and RUL prediction. Aghazadeh et al. (2018) used a convolution neural network (CNN) for tool wear prediction of a milling machine with a hybrid feature extraction method. A support vector regression (SVR) model was approached by Benkedjouh et al. (2015) for tool wear and RUL prediction from seven sensors data of a milling machine. Gokulachandran and Mohandas (2015) compared the performance of the neuro-fuzzy logic technique and SVR for the RUL prediction of cutting tools and showed that the performance of the neuro-fuzzy logic technique is better compared to SVR. Li et al. (2017) applied v-SVR for tool condition monitoring of a turning operation. In the research, fourteen (14) time-domain statistical features were extracted from force

signals. Drouillet et al. (2016) and Mikołajczyk et al. (2018) implemented ANN for RUL estimation of end milling operation and turning operation, respectively. D'Addona et al. (2017) combined artificial neural networks (ANN) with deoxyribonucleic acid (DNA)-based computing for tool wear estimation of a turning tool. Wu et al. (2016) proposed a random forest algorithm for tool wear prediction from dry milling data. An experimental study in the dry milling process showed that the random forest algorithm could predict tool wear with high accuracy.

In addition to the above-described research, Javed et al. (2012), Mosallam et al. (2016), and Gugulothu et al. (2017), implemented different methods for cutting tool RUL prediction. Many papers, including Zhao et al. (2017), Huang et al. (2019), Wang et al. (2015), and Cai et al. (2020), applied different data-driven prognostic methods for tool wear prediction of varying CNC machine tools. From the publications of related works, it can be observed that there is a continuous trend in developing advanced tool wear prediction methods for improving prediction accuracy. For this, researchers are still exploring different data-driven methods to utilise the advantages of these methods. The accuracy of tool wear and RUL prediction remains a challenge for many maintenance practitioners because of the nonlinear and stochastic relationships of tool wear with the extracted features from sensor signals.

Both data-driven and hybrid methods used data collected by sensors or experiments for tool wear prediction. The collected data may consist of noise due to different sources of uncertainty such as measurement error, environmental instability, and natural randomness. In addition, sensor imperfections, signal wire noise, and irrelevant and redundant features are common data uncertainty sources (Dey and Yodo, 2022). As prediction accuracy largely depends on the quality of collected sensor data, it is recommended to reduce noise before using the collected data as inputs in a data-driven method (Dey and Yodo, 2020). A variety of denoising techniques has been primarily used for this purpose; for instance, Khemissi et al. (2017) and Dey and Yodo (2021) used empirical mode decomposition (EMD), Zhang et al. (2015) applied the Wavelet denoising technique, and Wu et al. (2018) applied ensemble EMD (EEMD).

In addition to the uncertainty or noise in the collected sensor data, overfitting is a common problem in using the complex architecture of data-driven learning algorithms (Van der Aalst et al., 2010). Due to overfitting, an algorithm may fit well for training data but fails to predict for the test (unseen or new) data with good accuracy. The regularisation techniques are commonly applied to deal with the overfitting problem (Mustafa et al., 2013). In addition, the regularisation technique is used to increase model stability by penalising the model parameters. The dropout, L_1 regularisation, L_2 regularisation, and dataset augmentation are some commonly used regularisation techniques (Phaisangittisagul, 2016; Jin et al., 2004; Mustafa et al., 2013).

Generally, time-domain, frequency domain, and time-frequency domain features are extracted from sensor data to make a structured dataset in applying many data-driven methods. All extracted features may not be relevant to tool wear and RUL. Besides, prediction accuracy and computational efficiency decrease due to irrelevant features. Different methods, including principal component analysis (PCA), Pearson correlation coefficient, linear discrimination analysis (LDA), chi-square test, and Fisher score, can be employed to select relevant features and reduce input dimensions. The primary purposes of feature selection are to improve model performance, avoid overfitting, and reduce model training time (Jović et al., 2015; Vergara and Estévez, 2014). In addition,

hyperparameters, model architecture, and activation functions are also significant for prediction accuracy and computational efficiency.

Model architecture, data quality, feature selection methods and regularisation are vital for improving prediction accuracy. For tool wear prediction, different model architecture, denoising techniques, and feature selection methods are applied in many published articles. In the best knowledge of the authors, the impacts of different performance improvement techniques and their combinations are rarely analysed to optimise the performance of a data-driven method. Different performance improvement techniques are applied to improve the prediction accuracy of neural networks. There are no specific rules to choose performance improvement techniques. The selection of techniques depends on data quality, model architecture, and other factors. Generally, different techniques are investigated by trial and error to choose the best combination. It is well-established that performance improvement techniques such as data denoising, feature selection, and regularisation techniques are crucial for increasing prediction accuracy. Therefore, along with applying advanced data-driven methods for tool wear prediction, it is vital to choose performance improvement techniques to improve prediction accuracy and obtain computational advantages.

In this paper, the importance of data denoising techniques, feature selection algorithms, model architectures, and regularisations techniques on the performance improvement of neural networks are analysed. For analysing the importance of data denoising techniques, a collected raw signal data of a milling machine and the data denoised by EMD are used for prediction. Two feature selection techniques, PCA and Pearson correlation coefficient, are used for feature selection. In addition, dropout, L_1 regularisation, and L_2 regularisation techniques are applied to reduce overfitting and increase model stability. Finally, different combinations of data denoising techniques, feature selection methods, and regularisation techniques are applied to train two architectures of neural networks. The performance of neural networks is compared by using three performance evaluation metrics, namely mean square error (MSE), mean absolute error (MAE), and R^2 value, and for this, the 2010 PHM data challenge dataset is used. The dataset was collected from a CNC milling machine by using sensors. A dynamometer, three accelerometers, and an acoustic emission (AE) sensor were used to collect force signals in three axes, vibration signals in three axes, and energy signals. Sensor signals may consist of noise due to data collection environment and devices uncertainty. For this, data denoising techniques can be applied. Further, as the sensor signals are used in this paper to train NN models, time-domain features are extracted from each sensor signals to generate a structured dataset.

The rest of this paper is organised as follows. Data denoising techniques, feature selection methods, regularisation methods, and neural networks are introduced in Section 2. Section 3 consists of an experimental case study with data from the 2010 PHM data challenge to demonstrate the effectiveness of the proposed methods, including detailed results and discussions. Finally, the future research direction and conclusions are given in Section 4.

2 Methods background

The different combinations of performance improvement techniques are used to train neural networks for analysing the importance of applying data-driven methods. Empirical

mode decomposition (EMD) data denoising technique and two feature selection methods, principal component analysis (PCA) and Pearson correlation coefficient are applied for investigation. In addition, L_1 regularisation, L_2 regularisation, and dropout are three regularisation techniques applied during model training to avoid overfitting. Two neural network architectures for tool wear prediction are trained with different combinations of the three types of performance improvement techniques. Finally, three performance evaluation metrics, namely mean square error (MSE), mean absolute error (MAE) and *R*-squared (R^2) value, are used to compare the prediction accuracy of neural networks for different combinations of performance improvement techniques. The methods used for analysing the importance of performance improvement techniques are discussed in this section.

2.1 Data denoising

The data denoising is a critical step in tool wear prediction as it is often known to improve the prediction accuracy. In this paper, the collected raw sensor data along with denoised data is used tool wear prediction from analysing the significance of data denoising, and the EMD approach is applied as the data denoising (preprocessing) technique. One of the advantages of EMD is it does not require any predefined mathematical model (Manjula and Sarma, 2012). The assumption of EMD is a signal that can be decomposed into different intrinsic mode functions (IMFs) by shifting process, and each IMF represents individual characteristic oscillations of a signal (Schlurmann, 2001). An IMF must satisfy the following two criteria.

1 For a given signal vector, the number of extrema and the number of zero crossings must either be equal to or differ at most by one.

2 At any point, the mean value of the envelope defined by the local maxima and the local minima is zero.

Some IMFs from all decomposed IMFs consists of noise, and the goal is to identify and subtract those IMFs from the original signal to get a noiseless signal. Before discussing the IMFs identification procedure, the signal decomposition algorithm to extract IMFs is given in Figure 1.

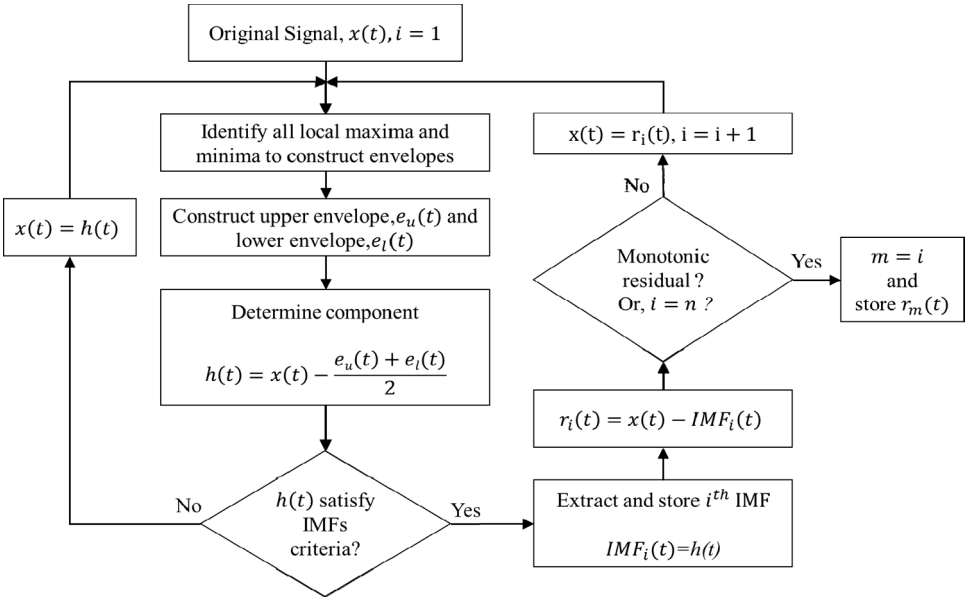
The extraction of IMFs is stopped when one of the following two conditions is satisfied: (1) a predefined number of IMFs are being extracted, or (2) the residual becomes monotonic from which no more IMF can be extracted (Huang, 2014). In this paper, it is assumed that the predefined maximum number of IMFs extracted is $n = 10$.

When all IMFs are extracted, the signal can be expressed as,

$$\boldsymbol{x}(t) = \sum_{i=1}^{m} \boldsymbol{IMF}_i(t) + \boldsymbol{r}_m(t) \tag{1}$$

where, m is the number of IMFs extracted from the original signal $\boldsymbol{x}(t)$. $\boldsymbol{r}_m(t)$ is the residual obtained after extracting the mth IMF.

Figure 1 A flow diagram of EMD approach



Consider the original signal $\boldsymbol{x}(t)$ is a collection of the noiseless signal $\tilde{\boldsymbol{x}}(t)$ and noise $\eta(t)$ as,

$$\boldsymbol{x}(t) = \tilde{\boldsymbol{x}}(t) + \boldsymbol{\eta}(t) \tag{2}$$

The target is to estimate the denoised signal $\tilde{\boldsymbol{x}}(t)$ from the original signal $\boldsymbol{x}(t)$ by removing the noise $\boldsymbol{\eta}(t)$. In the decomposed signal, the first IMF contains the high-frequency terms, and the last IMF contains the low-frequency terms (Shang-yue et al., 2015). It is also well-established and well-proven that high-frequency terms consist of more noise compared to low-frequency terms.

Consider the first *k* IMFs consist of noise. Therefore, $\tilde{\boldsymbol{x}}(t)$ can be written as

$$\tilde{\boldsymbol{x}}(t) = \boldsymbol{x}(t) - \sum_{i=1}^{k} \boldsymbol{IMF}_i(t) \tag{3}$$

The value of *k* can be determined by the correlation coefficient, σ that is defined as,

$$\sigma = \frac{\boldsymbol{x}(t)' \, \tilde{\boldsymbol{x}}(t)}{\sqrt{\boldsymbol{x}(t)' \, \boldsymbol{x}(t)} \sqrt{\tilde{\boldsymbol{x}}(t)' \, \tilde{\boldsymbol{x}}(t)}} \tag{4}$$

Assume, the threshold value of σ is ρ. Then, the value of *k* can be determined by,

$$k^* = \max\{k | \, \sigma \geq \rho\} \tag{5}$$

When, k^* has been determined, $\tilde{\boldsymbol{x}}(t)$ is further estimated from equation (3) by considering $k = k^*$. Generally, the threshold value of σ is assumed to be between 0.75 and 0.85 (Shang-yue et al., 2015). In this paper, it considered that the threshold value σ

is ρ = 0.8, and both, the denoised data by EMD and raw data are used for tool wear prediction by neural networks.

2.2 Feature selection methods

In this paper, eleven (11) time-domain features are extracted from raw and denoised sensor signals. The list of the extracted features with the mathematical formulas is given in Table 1, with $\tilde{x}$ represent the raw signal and denoised signal, and N_s is the number of sampling points.

Table 1 Time-domain features extracted from data

Features	*Formula*	*Features*	*Formula*
1. Mean	$\mu = \frac{1}{N_s}\sum_{i=1}^{N_s}\tilde{\boldsymbol{x}}_i$	7. Crest factor	$f_c = \frac{\max(\tilde{\boldsymbol{x}}_i)}{rms}$
2. Standard deviation	$\sigma = \sqrt{\frac{\sum_{i=1}^{N_s}(\tilde{\boldsymbol{x}}_i - \mu)^2}{N_s - 1}}$	8. Shape factor	$f_s = \frac{rms}{\frac{1}{N}\sum_{i=1}^{N_s}\lvert\tilde{\boldsymbol{x}}_i\rvert}$
3. Root mean square	$rms = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s}\tilde{\boldsymbol{x}}_i^2}$	9. Impulse factor	$f_i = \frac{\max(\tilde{\boldsymbol{x}}_i)}{\frac{1}{N_s}\sum_{i=1}^{N_s}\lvert\tilde{\boldsymbol{x}}_i\rvert}$
4. Square mean root	$smr = \left(\frac{1}{N_s}\sum_{i=1}^{N_s}\sqrt{\lvert\tilde{\boldsymbol{x}}_i\rvert}\right)^2$	10. Marginal factor	$f_m = \frac{\max(\tilde{\boldsymbol{x}}_i)}{smr}$
5. Skewness	$f_{sk} = \frac{\sum_{i=1}^{N_s}(\tilde{\boldsymbol{x}}_i - \mu)^3}{(N_s - 1)\sigma^3}$	11. Peak to peak	$f_{pp} = \max(\tilde{\boldsymbol{x}}_i)$ -min $(\tilde{\boldsymbol{x}}_i)$
6. Kurtosis	$f_k = \frac{\sum_{i=1}^{N_s}(\tilde{\boldsymbol{x}}_i - \mu)^4}{(N_s - 1)\sigma^4}$		

The feature selection techniques are applied in extracted features to select relevant features and reduce input dimensions. In this paper, two different feature selection techniques, PCA and Pearson correlation coefficient, are applied to investigate the importance of feature selection techniques.

2.2.1 Principal component analysis

Principal component analysis (PCA) is one of the most widely used dimension reduction techniques that capture the presence of significant variability in the dataset and minimise the loss of information. The goal of PCA is to reduce the number of features with minimum loss of information. PCA is an unsupervised statistical technique that identifies new features along with new directions, which are linear combinations of the original features and uncorrelated. In this paper, PCA is applied for two purposes: (1) to reduce the number of features, and (2) to convert the features into linearly independent features.

Reduced and uncorrelated features increase computational efficiency, model stability, and prediction accuracy. For this research, the dimension is reduced to the number of features that considered at least 95% variation.

2.2.2 Pearson correlation coefficient

Pearson correlation coefficient is a well-known feature selection technique to identify a linear correlation between two quantitative variables. Pearson correlation can be used for estimation the linear correlation between an input variable and a target variable and between two input variables. Here, the Pearson correlation coefficient is applied to estimate the correlation between input and the target variable. Pearson correlation coefficient lies between –1 to +1, where –1, 0, and +1 indicate a strong negative correlation, no correlation, and a strong positive correlation, respectively. In this paper, the used threshold value for the Pearson correlation coefficient is 0.8.

2.3 Regularisation techniques

The performance of a data-driven method may decrease due to overfitting and model instability (Reunanen, 2003; Van der Aalst et al., 2010). Overfitting occurs when a model learns well in the training data but cannot achieve good accuracy when the model is applied to test data. Small training data is one of the reasons leading to an overfitting prediction. A complex model with multiple hidden layers and a large number of neurons is also prone to overfit. Regularisation techniques are used to minimise the error in training and to reduce overfitting. In this paper, three regularisation techniques L_1 regularisation, L_2 regularisation and dropout are applied to overcome the overfitting problem and improvethe prediction performance of neural networks.

A neural network determines the relationships between inputs and output and predicts the output for new observations. One of the advantages of using data-driven algorithms is that it does not require any expert knowledge of a system for prediction. In this paper, the neural networks approach is a data-driven method used to analyse different combinations of performance improvement techniques. In neural networks, the parameters (weights) of the model are estimated by minimising the least square loss function. The least square loss function is defined as

$$L_{sq}(w) = \sum_{j=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{6}$$

where, y_j and $\hat{y}_j$ are the actual and the predicted value of j^{th} observation, and n is the number of observations in training data.

2.3.1 L_1 Regularisation

In L_1 regularisation, the loss function is a combination of the least-square loss function and a regularisation term and is defined as below.

$$L_{l_1}(w) = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 + \lambda_1 \| \boldsymbol{w} \| \tag{7}$$

where w is the parameter vector, and λ_1 is the regularisation parameter that controls the tradeoff between the least-square loss function and a regularisation term (Phaisangittisagul, 2016). The regularisation term is also called the penalty term.

2.3.2 L_2 regularisation

The loss function for L_2 regularisation can be defined as

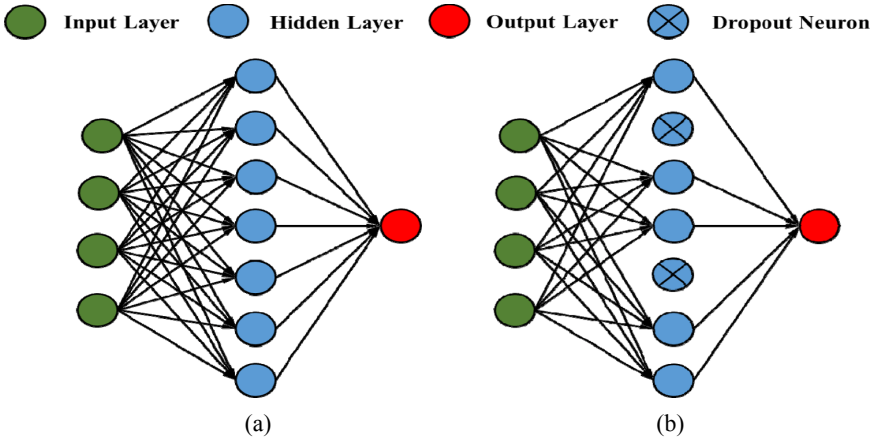
$$L_{l_2}(w) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda_2 \| w \|_2^2 \tag{8}$$

λ_2 is the regularisation parameter that controls the tradeoff between the least-square loss function and a regularisation term.

2.3.3 Dropout

Dropout is a regularisation technique applied for neural networks (NN) only. Srivastava et al. (2014) proposed a standard dropout technique to prevent overfitting and efficiently combine different NN architectures. In NN, a neuron of a layer is fully connected with all neurons of prior and post layers, as shown in Figure 2(a). The term dropout can be defined as deleting a neuron from networks along with all its associated connections (Figure 2(b)).

Figure 2 (a) A standard NN and (b) a NN after applying dropout (see online version for colours)



The dropout decision is binary. It is whether a neuron is retained in a NN or is dropped out from a NN. If the dropout probability is p, then the probability of a neuron retained in the network is $1-p$. Usually, dropout is applied to all neurons in a NN layer with the same probability. If the dropout is applied to q number of neurons, then the possible number of NN architecture is 2^q. It is recommended to apply the dropout technique only in hidden layers as neurons drop off in the input layer results in loss of features.

During training, the weights of the retained neurons are updated in each epoch, and the weights of dropped-out neurons remain unchanged. Dropout is applied randomly until a desired level of accuracy is obtained. After training NN with dropout, a fully connected network is further applied to predict for a new observation. And the final weight is

estimated as $(1-p)\,w$ for applying a fully connected network. In this paper, a total of eight $\left(2^3=8\right)$ different combinations of three described regularisation techniques are used to train the NN architectures.

2.4 Prediction accuracy measurement metrics

Evaluating the performance of any learning algorithm is an essential part of employing it in any application. Many metrics that are available to evaluate the performance of data-driven methods. In this paper, mean square error (MSE), mean absolute error (MAE) and *R*-squared (R^2) value metrics are used to quantitatively analyse the performance of the NN. The MSE is calculated as the average of the squared difference between the true and predicted values. The MAE is the average of the absolute difference between the true and predicted values. *R*-squared value is a measurement of the goodness of fit.

In this paper, all combinations of the described performance improvement techniques are applied for training neural networks. The performance is compared using the above-introduced prediction accuracy measurement metrics. Here, two (2) types of datasets, 2^3 or 8 combinations of three regularisation techniques and two (2) feature selection algorithms are used. Therefore, a total of 32 $\left(2\times8\times2\right)$ combinations of performance improvement techniques are used to train the two NN models. The impacts of the performance improvement techniques are investigated in the following section using a CNC milling machine dataset as a case study.

3 Case study

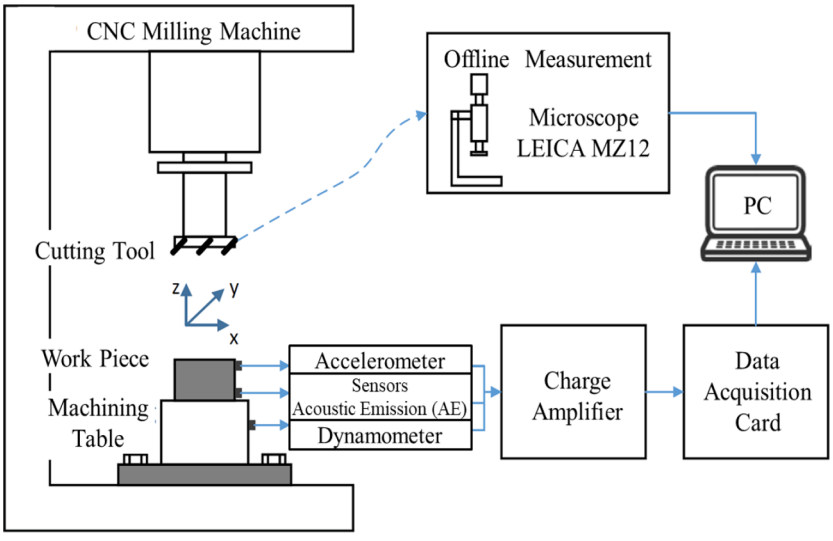
This section demonstrates the performance of the different performance improvement techniques, namely data denoising, feature selection (PCA and Pearson correlation), and regularisation techniques. The NN-based data-driven approach with a milling machine dataset from the 2010 PHM Data Challenge (Society, 2010) is employed as the case study.

3.1 Data description

The data was collected from a high-speed CNC milling machine (Röders Tech RFM760) during a down milling operation on a stainless-steel workpiece. For the machining process, a 6 mm ball nose tungsten carbide cutter with three (03) flutes were used with constant values of the following machining parameters: the spindle speed of the cutter was 10,400 RPM; the feed rate was 1555 mm/min; the *Y*-axis depth of cut (radial) was 0.125 mm, and the *Z*-axis depth of cut (axial) was 0.2 mm. The experimental setup is shown in Figure 3.

The seven sensors' signals, force_x, force_y, force_z, vibration_x, vibration_y, vibration_z, and acoustic emission (AE), were installed to collect data for tool condition monitoring and tool wear prediction (Li et al., 2009). The signals were captured by a NI DAQ PCI 1200 board with a 12 kHz frequency. Seven channels of signals were captured by the DAQ card with an accumulated sampling rate of $12\ \text{kHz}\times7=84\ \text{kHz}$.

Figure 3 Experimental setup for data collection (see online version for colours)



Source: Cai et al. (2020)

The sensor data for six cutters were collected in the experiment. The sensor data of three cutters (C_1, C_4, and C_6) were used in this paper as the values of the target variable, tool flank wear, are given for sensor signals of the three cutters. For each of three cutters, the data of 315 cycles were collected, and the tool flank wear was measured in 10^{-3} mm for each cycle. The milling cutters have three flutes. Tool flank wear was measured for each flute, and the average is considered as the target variable in this paper. Cai et al. (2020) and Zhao et al. (2017) predicted tool flank wear of cutter C_1, C_4, and C_6. Both publications have got the least prediction accuracy for the tool flank wear prediction of the cutter C_6. In this paper, tool flank wear of cutter C_6 is predicted for all combinations of performance improvement techniques based on collected historical data of C_1 and C_4.

3.2 Neural network (NN) models

A neural network, a mathematical model, has been used to develop the unknown relationship between input variables and a target variable based on several known observations. An NN is a collection of neurons (units), and the neurons are arranged in different layers, as shown in Figure 2(a). A layer is an aggregation of neurons. The simplest form of one NN has at least one input layer and one output layer. The other layers, known as hidden layers, perform different transformations on their inputs. A neuron of a layer is fully connected with all neurons of prior and post layers. The number of neurons in the input and output layers equals the number of features and the number of target variables, respectively.

After applying PCA and Pearson correlation coefficient on the dataset, the number of features will be obtained. On the other hand, the only target variable in this dataset is tool flank wear. Therefore, the output layer has only one neuron in NN models. The number of hidden layers and the number of neurons in a layer depends on data complexity. The dataset does not fit well without hidden layers as the relationship between input features,

and tool flank wear is not linear. Besides, the use of a complex NN architecture may result in overfitting.

As the dataset used in this paper is small and has fewer features, the neural network with one or two hidden layers would work well. In this paper, two NN architectures with one and two hidden layers are used. There is no rule of thumb to choose the number of neurons in a hidden layer. Generally, a trial-and-error process is applied to determine the number of neurons in hidden layers. But, as a starting point, it is considered that the number of neurons in the first hidden layer is less than the number of features. The number of neurons is recommended to keep decreasing in subsequent hidden layers. In this paper, the above-described rules are followed to choose the number of hidden layers and the number of neurons in each hidden layer.

Prior to applying the neural networks, 11 time-domain features listed in Table 1 are extracted from raw and denoised data. For data denoising, the maximum 10 IMFs are extracted from a sensor signal or less than 10 IMFs if the residual becomes monotonic before extracting 10 IMFs. For IMF selection, the threshold value of the correlation coefficient is 0.8. A total of 77 features (11 from each signal data) are extracted from the denoised signals. Similar to the denoised data, 77 features are also extracted from the collected sensor signal (raw data) without denoising. The Pearson correlation coefficient and PCA are applied to reduce the number of features from the two datasets obtained after extracting features from the raw data and denoised data.

For Pearson correlation coefficient between features and tool flank wear, the used threshold value is 0.8. Therefore, a feature will be selected for training neural networks if the absolute value of the Pearson correlation coefficient is not less than 0.8.With the threshold value, the selected features for the raw data and denoised data are 20 and 19, respectively. The selected features are used for further analysis.

Along with the Pearson correlation coefficient, PCA is also applied to reduce data dimensions and generate independent features. For both datasets, the dimension is reduced to the number of features that considered at least 95% variation. The reduced dimensions of the raw data and denoised data obtained by applying PCA are 21 and 23, respectively. The reduced datasets are used to train neural networks. Four datasets are obtained after applying Pearson correlation coefficient and PCA on raw data and denoised data.

All datasets are normalised before training the NN models. The one hidden layer NN model has one hidden layer with 18 neurons. The two hidden layers NN model consists of two hidden layers with 16 and 10 neurons. The activation function for all hidden layers is the exponential linear unit (ELU) for both NN architectures. The backpropagation method is applied to determine the gradient of loss function for the weights of NN. Adam optimiser that iteratively minimises the loss function with a learning rate of 0.001, exponential decay control parameters of 0.99 and 0.999 is used to update the weights of the NN models. The details of the Adam optimiser are not included in this paper since it has been well-documented. Readers interested in learning more about Adam optimiser are recommended to read Kingma and Ba (2014).

For training two NN models with the four datasets of the cutters C_1 and C_4, a total of 8 combinations of three regularisation techniques are applied. Therefore, 64 NN models are trained for both NN architectures. The performance of models is compared by using three evaluation metrics, MSE, MAE and R^2 value for the data of cutter C_6. For both

L_1 and L_2 regularisation techniques, the assigned regularisation parameter is 0.001 $(\lambda_1 = \lambda_2 = 0.001)$. For the dropout regularisation, the dropout is applied in only the hidden layers for both NN architectures, and the applied dropout rate is 0.1 $(p = 0.1)$. The results for the tool wear prediction by NN models are analysed in the following subsection.

3.3 Results and discussions

The MSE, MAE and R^2 value of tool wear prediction for cutter C_6 obtained from the two NN models are analysed in this subsection. The prediction accuracy of all combinations of performance improvement techniques is compared, and the best combination is identified from all analysed combinations for tool wear prediction.

The first NN model has one hidden layer with 18 neurons, and the used activation function is ELU. First, the MSE, MAE and R^2 value for cutter C_6 by the one hidden layer NN is summarised in Table 2.The smaller values of MSE and MAE and the greater value of R-squared are preferable for the NN model. For this, the combination of performance improvement techniques with the lowest values of MSE, lowest values of MAE, and the highest value of R^2 will be the best combination that may predict tool wear with high accuracy.

There is a significant difference in MSE, MAE, and R^2 values for the different combinations of data denoising, feature selection, and regularisation techniques. It indicates that the data denoising, feature selection, and regularisation techniques significantly improve the prediction accuracy. In Table 2, the ranges of MSE, MAE, and R^2 values are from 0.031 to 0.168, from 0.136 to 0.311, and from 0.832 to 0.969, respectively. The worst prediction performance is obtained when EMD and Pearson correlation coefficients are applied, respectively, for data denoising and feature selection. On the other hand, the highest prediction accuracy is achieved when EMD, Pearson correlation coefficient and L_1 & L_2 regularisation are applied to improve data quality, reduce data dimensions and minimise overfitting. It is visible that the regularisation techniques are significant as the EMD and Pearson correlation coefficient are common techniques for both the best and worst prediction accuracy. The combination of L_1 & L_2 regularisations with EMD and Pearson correlation coefficient is the best to improve the prediction accuracy.

For the one hidden layer NN, the best and worst cases estimation of tool wear for the test data (cutter C_6) are demonstrated in Figure 4. In Figure 4(a), the deviation of the predicted values from true values for the test dataset is significant. This worst-case prediction is obtained from NN with the EMD data denoising and Pearson correlation coefficient feature selection techniques. On the other hand, the deviation is comparatively low for the best case that is obtained with one hidden layer NN with denoised data, Pearson correlation coefficient, and L_1 & L_2 regularisation as shown in Figure 4(b). The deviation in both figures is an indication that the performance improvement techniques are significant for increasing the prediction accuracy.

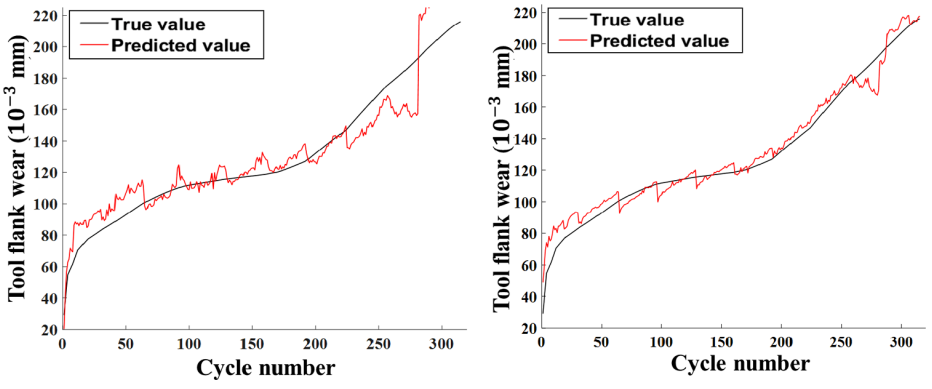
Similar to the one hidden layer NN model, the two hidden layers NN architecture is trained by using the cutter C_1 and C_4 datasets for all 32 combinations of performance improvement techniques. All trained models are used to fit the test data of the cutter C_6, and the measured MSE, MAE, and R^2 values for all combinations are given in Table 3.

Table 2 Prediction performance of the one hidden layer NN architecture

			No regularisation	L_1	L_2	*Dropout*	L_1 & L_2	L_1 & *Dropout*	L_2 & *Dropout*	L_1, L_2 & *Dropout*
Raw data	PCA	MSE	0.101	0.072	0.057	0.083	0.100	0.106	0.076	0.115
		MAE	0.225	0.216	0.200	0.220	0.239	0.262	0.223	0.267
		R^2 value	0.899	0.928	0.943	0.917	0.900	0.894	0.924	0.885
	Pearson	MSE	0.107	0.093	0.094	0.118	0.070	0.072	0.092	0.044
		MAE	0.255	0.231	0.234	0.244	0.206	0.199	0.213	0.172
		R^2 value	0.893	0.907	0.906	0.882	0.930	0.928	0.908	0.956
Denoise data	PCA	MSE	0.085	0.064	0.069	0.069	0.050	0.044	0.052	0.041
		MAE	0.207	0.175	0.182	0.167	0.165	0.152	0.158	0.156
		R^2 value	0.915	0.936	0.931	0.931	0.950	0.956	0.948	0.959
	Pearson	MSE	0.168*	0.082	0.087	0.090	0.031**	0.059	0.073	0.044
		MAE	0.311*	0.225	0.223	0.221	0.136**	0.179	0.196	0.164
		R^2 value	0.832*	0.918	0.913	0.910	0.969**	0.941	0.927	0.956

*indicates the worst performance, ** indicates the best performance.

Figure 4 The results of tool flank wear prediction by the one hidden layer NN (a) worst-case and (b) best-case (see online version for colours)



It is shown that the minimum prediction accuracy is obtained for the raw data when Pearson correlation is applied to reduce data dimensions, and L_2 regularisation is used to train the model. The best prediction performance is obtained for denoised data with Pearson correlation coefficient as feature selection technique and L_1 & L_2 regularisation method. The best and worst values for MSE are 0.028 and 0.139 and for MAE are 0.139 and 0.292, respectively. And, the best R^2 value is 0.972, and the worst R^2 value is 0.861.

The best and worst cases of tool wear prediction for the two hidden layers of NN architecture are demonstrated in Figure 5. For this NN architecture, the deviation of the predicted values from the true values is high for the worst case compared to the best case. Again, the best-case prediction is obtained with EMD, Pearson correlation coefficient, and L_1 & L_2 regularisation techniques, whereas the worst case of prediction is obtained with Pearson correlation coefficient and L_2 regularisation techniques.

In Tables 2 and 3, the combinations of performance improvement techniques are different for the worst prediction accuracy. Besides, the MSE, MAE, and R^2 values are different for the same combination of performance improvement techniques with one and two hidden layers NN architectures. Therefore, it can be concluded that NN architecture is significant along with other performance improvement techniques.

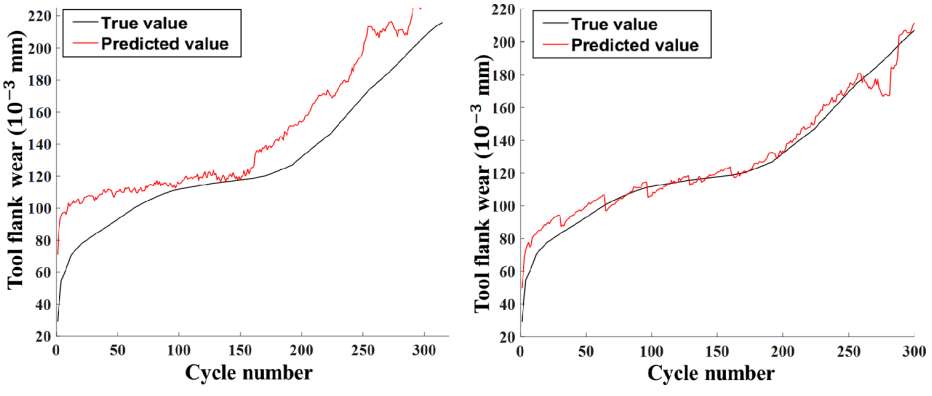
In addition, from Tables 2 and 3, it can be concluded that the data denoising technique, EMD, is vital for the dataset as the best prediction performance is achieved for both NN architectures with the denoised dataset. It is recommended to use data denoising techniques for sensor signals as they may have noise due to the data collection environment, data collection equipment, and data collecting and processing devices' uncertainty. The prediction accuracy is the highest for the combinations of EMD, Pearson correlation coefficient, and L_1 & L_2 regularisation in both NN architectures. The highest prediction accuracy is achieved for two hidden layers NN between the two NN models with the same combination of other performance improvement techniques. Although, the best values of all performance evaluation metrics are very close for both NN models. Hence, it can be said that NN model architecture is less significant compared to the other performance improvement techniques discussed.

Table 3 Prediction performance of the two hidden layers NN architecture

			No regularisation	L_1	L_2	*Dropout*	L_1 & L_2	L_1 & *Dropout*	L_2 & *Dropout*	L_1, L_2 & *Dropout*
Raw data	PCA	MSE	0.073	0.067	0.057	0.087	0.086	0.084	0.084	0.095
		MAE	0.210	0.216	0.200	0.238	0.237	0.229	0.236	0.247
		R^2 value	0.927	0.933	0.943	0.913	0.914	0.916	0.916	0.905
	Pearson	MSE	0.117	0.083	0.139*	0.109	0.080	0.084	0.117	0.051
		MAE	0.260	0.224	0.292*	0.240	0.217	0.212	0.244	0.182
		R^2 value	0.883	0.917	0.861*	0.891	0.920	0.916	0.883	0.949
Denoise data	PCA	MSE	0.103	0.064	0.061	0.073	0.090	0.115	0.124	0.098
		MAE	0.228	0.177	0.167	0.188	0.217	0.257	0.266	0.217
		R^2 value	0.897	0.936	0.939	0.927	0.910	0.885	0.876	0.902
	Pearson	MSE	0.073	0.069	0.058	0.065	0.028**	0.038	0.054	0.043
		MAE	0.204	0.210	0.181	0.206	0.131**	0.150	0.172	0.161
		R^2 value	0.927	0.931	0.942	0.935	0.972**	0.962	0.946	0.957

*indicates the worst performance, ** indicates the best performance.

Figure 5 The results of tool flank wear prediction by the two hidden layers NN (a) worst-case and (b) best case (see online version for colours)



The best-case predictions for one hidden layer NN and two layers NN are shown in Figures 4(b) and 5(b), respectively. For both best cases, the deviation of the predicted values from the actual values is high from 270th to 290th cycles. The prediction accuracy is high between 160th and 260th cycles for one hidden layer NN as shown in Figure 4(a). On the other hand, Figure 5(b) shows that the prediction deviation is minimum from 115th to 260th cycles. These results are also reflected in the values of performance measurement metrics as the MSE, MAE, and R^2 values are slightly better for two hidden layers NN.

Overall, for this dataset, among the total of 64 combinations of performance improvement techniques, the combination of EMD, Pearson correlation coefficient, L_1 & L_2 regularisations, and two hidden layers is recommended as the best values of MSE, MAE, and R^2 values are obtained for the combination. It may not always be true that a definite combination of performance improvement techniques improves performance for all datasets and model architectures. The different performance improvement techniques work for different datasets based on several factors such as data noise, model architecture, irrelevant features, and redundant features. The selection of performance improvement techniques is challenging as sometimes performance improvement techniques may deteriorate prediction accuracy. Although there is no rule of thumb to select the performance improvement techniques for a dataset, it is well-established that the techniques significantly impact on the performance of data-driven methods. Many deep learning algorithms can automatically extract features (Hao et al., 2016). For these deep learning algorithms, manual feature extraction and selection do not require. Other performance improvement techniques work well for deep learning algorithms as well.

4 Conclusions

In this paper, the impact of different performance improvement techniques and their combinations on the prediction accuracy of neural networks are analysed. For this dataset, the data denoising technique, EMD, significantly impacts neural network

performance. The minimum MSE, MAE, and maximum R^2 values are obtained in both model architectures for the denoised dataset. It indicates that the collected sensor signals have noise. It is recommended to reduce noise in a dataset to improve the performance. Besides, the Pearson correlation coefficient as feature selection techniques and the combination of L_1 & L_2 regularisation techniques are applied to achieve the maximum prediction accuracy. The combination of these techniques works well for this dataset. It should be noted that the same combination may not necessarily have the same impacts when applied to other datasets. But, it is clear that different performance improvement techniques significantly impact the prediction accuracy of NN models. The selection of performance improvement techniques is critical in applying data-driven tool wear prediction or other condition monitoring applications. Exploring different techniques based on datasets, application areas, and expert opinions are recommended as no rules of thumb for performance improvement technique selection. In the future, frequency-domain and time-frequency domain features will be explored to improve prediction accuracy. Additionally, the impacts of other performance improvement techniques on the prediction performance of deep learning algorithms will be analysed for tool condition monitoring.

References

Aghazadeh, F., Tahan, A. and Thomas, M. (2018) 'Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process', *The International Journal of Advanced Manufacturing Technology*, Vol. 98, pp.3217–3227.

Baraldi, p., Compare, M., Sauco, S. and Zio, E. (2013) 'Ensemble neural network-based particle filtering for prognostics', *Mechanical Systems and Signal Processing*, Vol. 41, pp.288–300.

Benkedjouh, T., Medjaher, K., Zerhouni, N. and Rechak, S. (2015) 'Health assessment and life prediction of cutting tools based on support vector regression', *Journal of Intelligent Manufacturing*, Vol. 26, pp.213–223.

Cai, W., Zhang, W., Hu, X. and Liu, Y. (2020) 'A hybrid information model based on long short-term memory network for tool condition monitoring', *Journal of Intelligent Manufacturing*, Vol. 31, No. 6, pp.1–14.

D'Addona, D.M., Ullah, A.S. and Matarazzo, D. (2017) 'Tool-wear prediction and pattern-recognition using artificial neural network and DNA-based computing', *Journal of Intelligent Manufacturing*, Vol. 28, pp.1285–1301.

Das, S., Hall, R., Herzog, S., Harrison, G., Bodkin, M. and Martin, L. (2011) 'Essential steps in prognostic health management', *2011 IEEE Conference on Prognostics and Health Management*, IEEE, Denver, CO, USA, pp.1–9.

Dey, A. and Yodo, N. (2022) 'Conservative confidence interval prediction in fused deposition modeling process with linear optimization approach', *ASCE-ASME J Risk and Uncert in Engrg Sys Part B Mech Engrg*, Vol. 8, No. 1, p.8.

Dey, A. and Yodo, N. (2021) 'A dropout-based neural network framework for tool wear prediction under uncertainty', *IIE Annual Conference. Proceedings*, Virtual Conference, Institute of Industrial and Systems Engineers (IISE), pp.902–907.

Dey, A. and Yodo, N. (2020) 'Interval prediction of remaining useful life of cutting tools using a neural network approach', *IIE Annual Conference. Proceedings*, Virtual Conference, Institute of Industrial and Systems Engineers (IISE), pp.145–150.

Dhanasekaran, P., Kalla, D.K. and Yodo, N. (2010) 'Recent advances in cutting tool materials and tool coatings for machining composite materials', *Tooling for Composites Conference*, Salt Lake City, UT, USA, pp.1–11.

Drouillet, C., Karandikar, J., Nath, C., Journeaux, A-C., El Mansori, M. and Kurfess, T. (2016) 'Tool life predictions in milling using spindle power with the neural network technique', *Journal of Manufacturing Processes*, Vol. 22, pp.161–168.

Gokulachandran, J. and Mohandas, K. (2015) 'Comparative study of two soft computing techniques for the prediction of remaining useful life of cutting tools', *Journal of Intelligent Manufacturing*, Vol. 26, pp.255–268.

Gugulothu, N., Tv, V., Malhotra, P., VIG, L., Agarwal, P. and Shroff, G. (2017) *Predicting Remaining Useful Life Using Time Series Embeddings based on Recurrent Neural Networks*, arXiv preprint arXiv:1709.01073.

Hao, X., Zhang, G. and Ma, S. (2016) 'Deep learning', *International Journal of Semantic Computing*, Vol. 10, No. 03, pp.417–439.

Huang, N.E. (2014) *Hilbert-Huang Transform and Its Applications*, World Scientific, Singapore.

Huang, Z., Zhu, J., Lei, J., LI, X. and Tian, F. (2019) 'Tool wear predicting based on multisensory raw signals fusion by reshaped time series convolutional neural network in manufacturing', *IEEE Access*, Vol. 7, pp.178640–178651.

Javed, K., Gouriveau, R., Zerhouni, N., Zemouri, R. and Li, X. (2012) 'Robust, reliable and applicable tool wear monitoring and prognostic: approach based on an improved-extreme learning machine', *2012 IEEE Conference on Prognostics and Health Management*, IEEE, Denver, CO, USA, pp.1–9.

Jin, Y., Okabe, T. and Sendhoff, B. (2004) 'Neural network regularization and ensembling using multi-objective evolutionary algorithms', *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No. 04TH8753)*, IEEE, Portland, OR, USA, pp.1–8.

Jović, A., Brkić, K. and Bogunović, N. (2015) 'A review of feature selection methods with applications', *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE, Opatija, Croatia, pp.1200–1205.

Khemissi, B.M., Ouelaa, N. and Djebala, A. (2017) 'Application of the empirical mode decomposition method for the prediction of the tool wear in turning operation', *Mechanics*, Vol. 23, pp.315–320.

Khumprom, P. and Yodo, N. (2019) 'A data-driven predictive prognostic model for lithium-ion batteries based on a deep learning algorithm', *Energies*, Vol. 12, p.660.

Kingma, D.P. and Ba, J. (2014) *Adam: A Method for Stochastic Optimization*, arXiv preprint arXiv:1412.6980.

Li, N., Chen, Y., Kong, D. and Tan, S. (2017) 'Force-based tool condition monitoring for turning process using v-support vector regression', *The International Journal of Advanced Manufacturing Technology*, Vol. 91, pp.351–361.

Li, X., Ding, Q. and Sun, J-Q. (2018) 'Remaining useful life estimation in prognostics using deep convolution neural networks', *Reliability Engineering & System Safety*, Vol. 172, pp.1–11.

Li, X., Lim, B., Zhou, J., Huang, S., Phua, S., Shaw, K. and Er, M. (2009) 'Fuzzy neural network modelling for tool wear estimation in dry milling operation', *Annual Conference of the Prognostics and Health Management Society*, San Diego, CA, USA, pp.1–11.

Manjula, M. and Sarma, A. (2012) 'Comparison of empirical mode decomposition and wavelet based classification of power quality events', *Energy Procedia*, Vol. 14, pp.1156–1162.

Mikołajczyk, T., Nowicki, K., Bustillo, A. and Pimenov, D.Y. (2018) 'Predicting tool life in turning operations using neural networks and image processing', *Mechanical Systems and Signal Processing*, Vol. 104, pp.503–513.

Mosallam, A., Medjaher, K. and Zerhouni, N. (2016) 'Data-driven prognostic method based on bayesian approaches for direct remaining useful life prediction', *Journal of Intelligent Manufacturing*, Vol. 27, pp.1037–1048.

Mustafa, G., Ghaffar, A. and Aslam, M. (2013) 'A subdivision-regularization framework for preventing over fitting of data by a model', *Appl. Appl. Math. Int. J.*, Vol. 8, pp.178–190.

Peng, Y., Dong, M. and Zuo, M.J. (2010) 'Current status of machine prognostics in condition-based maintenance: a review', *The International Journal of Advanced Manufacturing Technology*, Vol. 50, pp.297–313.

Phaisangittisagul, E. (2016) 'An analysis of the regularization between L2 and dropout in single hidden layer neural network', *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, IEEE, Bangkok, Thailand, pp.174–179.

Reunanen, J. (2003) 'Overfitting in making comparisons between variable selection methods', *Journal of Machine Learning Research*, Vol. 3, pp.1371–1382.

Schlurmann, T. (2001) 'The empirical mode decomposition and the Hilbert spectra to analyse embedded characteristic oscillations of extreme waves', *Rogue Waves*, pp.157–165.

Shang-Yue, Z., Yuan-Yuan, L. and Gong-Liu, Y. (2015) 'EMD interval thresholding denoising based on correlation coefficient to select relevant modes', *2015 34th Chinese Control Conference (CCC)*, IEEE, Hangzhou, China, pp.4801–4806.

Society, P. (2010) *PHM Data Chalelnge 2010* [Online], Available: https://www.phmsociety.org/competition/phm/10 (Accessed 15 March, 2021).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) 'Dropout: a simple way to prevent neural networks from overfitting', *The Journal of Machine Learning Research*, Vol. 15, pp.1929–1958.

Su, C. and Wang, X. (2014) 'Optimizing upgrade level and preventive maintenance policy for second-hand products sold with warranty', *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, Vol. 228, pp.518–528.

van der Aalst, W.M., Rubin, V., Verbeek, H., van Dongen, B.F., Kindler, E. and Günther, C.W. (2010) 'Process mining: a two-step approach to balance between underfitting and overfitting', *Software & Systems Modeling*, Vol. 9, p.87.

Vergara, J.R. and Estévez, P.A. (2014) 'A review of feature selection methods based on mutual information', *Neural Computing and Applications*, Vol. 24, pp.175–186.

Wang, J., Wang, P. and Gao, R.X. (2015) 'Enhanced particle filter for tool wear prediction', *Journal of Manufacturing Systems*, Vol. 36, pp.35–45.

Wu, D., Jennings, C., Terpenny, J. and Kumara, S. (2016) 'Cloud-based machine learning for predictive analytics: tool wear prediction in milling', *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, Washington, DC, USA, pp.2062–2069.

Wu, J., Su, Y., Cheng, Y., Shao, X., Deng, C. and Liu, C. (2018) 'Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system', *Applied Soft Computing*, Vol. 68, pp.13–23.

Xu, J. and Xu, L. (2011) 'Health management based on fusion prognostics for avionics systems', *Journal of Systems Engineering and Electronics*, Vol. 22, pp.428–436.

Zhang, C., He, Y., Yuan, L., Xiang, S. and Wang, J. (2015) 'Prognostics of lithium-ion batteries based on wavelet denoising and DE-RVM', *Computational Intelligence and Neuroscience*, Vol. 2015, p.14.

Zhao, R., Yan, R., Wang, J. and Mao, K. (2017) 'Learning to monitor machine health with convolutional bi-directional LSTM networks', *Sensors*, Vol. 17, p.273.