J Intell Manuf
DOI 10.1007/s10845-016-1221-2

Tool wear monitoring and prognostics challenges: a comparison of connectionist methods toward an adaptive ensemble model

Kamran Javed[1] · Rafael Gouriveau[1] · Xiang Li[2] · Noureddine Zerhouni[1]

Received: 6 October 2015 / Accepted: 12 April 2016

Abstract In a high speed milling operation the cutting tool acts as a backbone of machining process, which requires timely replacement to avoid loss of costly workpiece or machine downtime. To this aim, prognostics is applied for predicting tool wear and estimating its life span to replace the cutting tool before failure. However, the life span of cutting tools varies between minutes or hours, therefore time is critical for tool condition monitoring. Moreover, complex nature of manufacturing process requires models that can accurately predict tool degradation and provide confidence for decisions. In this context, a data-driven connectionist approach is proposed for tool condition monitoring application. In brief, an ensemble of Summation Wavelet-Extreme Learning Machine models is proposed with incremental learning scheme. The proposed approach is validated on cutting force measurements data from Computer Numerical Control machine. Results clearly show the significance of our proposition.

Keywords Applicability · Data-driven · Ensemble · Monitoring · Prognostics · Robustness · Reliability

Introduction

The high speed milling process has become the most important and cost-effective means in manufacturing industry, to produce parts with high surface quality due to advantages like: high productivity, reliable and repeatable accuracies, good surface finish, etc., (Zhai et al. 2010; Saikumar and Shunmugam 2012; Rizal et al. 2013). This process is performed in a dynamic environment and under diverse conditions, where a cutting tool acts as a backbone of machining process. The high speed milling manufacturing is of complex nature, and requires special care since its performances are closely related to the conditions like cutting tool wear, hardness variations, and abrupt breakage of cutter. Moreover, life span of the cutting tools varies between minutes or hours, and failure of cutting tool can affect product quality and cause machine down-time (Wu et al. 2015; Zhou et al. 2011). Therefore, ensuring high surface quality of the workpiece and avoiding machine down-time, requires the cutting tool to be replaced before the tool wear passes failure threshold. This task can be achieved through condition monitoring and prognostics (Benkedjouh et al. 2013). In fact, prognostics has been investigated for several applications like: epidemiology prediction, weather forecasting, stock market prediction, etc., Fig. 1.

Prognostics for manufacturing refers to tool wear prediction and estimation of its life span for timely replacement. More precisely, for tool condition monitoring application, the prognostics model uses monitoring data from sensors (e.g. vibration, force or acoustic emission) to predict tool wear after each cut and to determine the number of cuts that could be made safely before failure.

In recent years, research on prognostics for manufacturing has grown rapidly, and a vast number of prognostics algorithms are introduced to enable short-term or long-term decisions, particularly from data-driven category. According to literature, for prediction in milling operations the Artificial neural networks (ANNs) are the most widely used connectionist methods among data-driven prognostics approaches

✉ Kamran Javed
Kamran.javed@femto-st.fr

1 FEMTO-ST Institute (AS2M Department), UMR CNRS 6174, UBFC/ UFC/ ENSMM / UTBM, 24 Rue Alain Savary, 25000 Besançon, France

2 Singapore Institute of Manufacturing Technology, 71 Nayang Drive, 638075 Singapore, Singapore

Published online: 27 April 2016

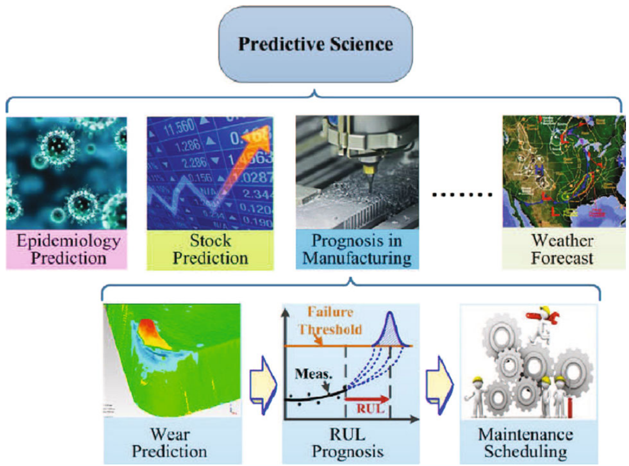

Fig. 1 Predictive science and manufacturing application (Gao et al. 2015)

(Grzenda and Bustillo 2013). As for some examples from the recent publications: Pal et al. (2011) used a standard back-propagation neural network and a Radial Basis Function network for predicting tool condition. This work also evaluated the robustness of ANNs against uncertainty of input data. Das et al. (2011) used a ANN approach to learn relationship of extracted features and the wear magnitude of cutting tool. In Wang and Cui (2013), a Levenberg Marquardt algorithm is introduced to improve the accuracy of auto associative neural network for tool wear monitoring. Wu et al. (2015) proposed bayesian-multilayer perceptron approach to estimate tool wear. Cojbasic et al. (2015) proposed one-pass Extreme Learning Machine (ELM) algorithm to estimate roughness of machined surface.
Although several works have been proposed for tool condition monitoring application, however, the issues with existing connectionist approaches are as follows.

- Cutting tools life varies between minutes/hours, therefore time for tool condition monitoring is critical, which requires rapid connectionist approaches.
- The common drawbacks of classical connectionist approaches are model complexity, slow iterative tuning, imprecise learning rate, presence of local minima and overfitting.
- Due to uncertainties from different sources like tool degradation process, data, operating condition and model, it is essential to manage and quantify uncertainty to enable decisions.
- It is difficult to generalize tool wear prediction model on cutting tools data that are not included in the learning phase.

To address those issues, this paper contributes relatively a new data-driven connectionist approach for tool condition monitoring application. More precisely, an ensemble of Summation Wavelet-Extreme Learning Machine (SW-ELM) models is proposed with incremental learning scheme to update model parameters on-line, predict tool wear, estimate tool life span and give confidence for decision making. The proposed SW-ELM ensemble (SW-ELME) is validated on cutting force measurements data from Computer Numerical Control machine (CNC). This contribution is achieved through following objectives.

1. Define prognostics modeling challenges.
2. Compare SW-ELM with rapid learning approaches.
3. Build SW-ELME with incremental learning scheme.
4. Validate SW-ELME on unknown cutting tools data.

The remaining paper is organized as follows. "Towards an enhanced data-driven prognostics" section gives the background of data-driven framework for tool condition monitoring, defines prognostics modeling challenges and discusses the choice of data-driven connectionist approach according to those challenges. "Proposed data-driven approach" section dedicated to our basic SW-ELM algorithm and SW-ELM ensemble with incremental learning scheme for tool condition monitoring. "Case study: tool condition monitoring" section presents a comparison of basic SW-ELM with ELM and Echo State Network (ESN) to encounter prognostics challenges and demonstrates the performances of SW-ELM ensemble on real data of cutting tools from CNC machine. Finally, "Conclusion" section concludes this work and gives future perspectives.

Towards an enhanced data-driven prognostics

Data-driven tool wear monitoring framework

To transform raw monitoring data into relevant behavior models, the framework of data-driven tool wear monitoring is based on the following steps (Fig. 2).

Data acquisition

During the cutting treatment of metal workpiece, the cutter wear can increase due to varying loads on its flutes that are always engaged and disengaged with the surface of workpiece (Das et al. 2011). This may result in increased imperfections in the workpiece surface i.e., dimensional accuracy of finished parts. Most CNC machines are unable to detect tool wear on-line, which is measured by using optical or electrical resistance sensors. Therefore, the estimate of cutting performance is usually performed through indirect method of tool condition monitoring (without shutting down the machine), by acquiring data that can be related to suit-

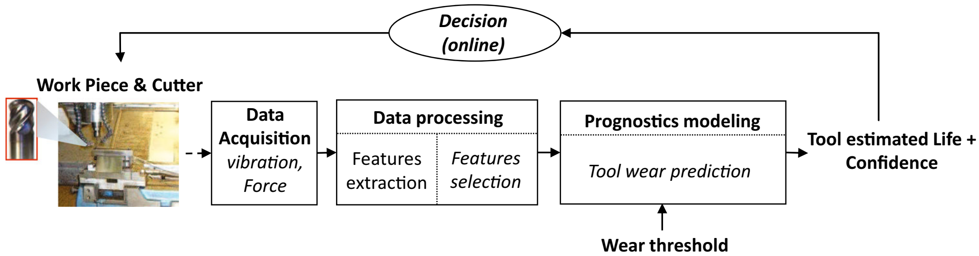

Fig. 2 Data-driven tool wear monitoring for on-line decisions

able wear models (Zhou et al. 2011). The most commonly employed are: cutting vibration (Haddadi et al. 2008), and force signals (Zhai et al. 2010). Such data are collected at regular intervals under given operating conditions.

Data processing

The cutting vibration measurements benefit from wide frequency range and are easy to implement (Ding and He 2011). Whereas, cutting force signals are more sensitive to tool wear than vibration (Ghasempoor et al. 1998), and preferred for modeling due to good measurement accuracy (Zhou et al. 2006). Also they are easy to manipulate and considered as the most useful to predict cutting performance (Zhai et al. 2010; Zhou et al. 2009).
The raw monitoring data acquired from the cutting tools are redundant and noisy, which can not be used directly for tool wear prediction. The data processing step enables extracting and selecting features from vibration/ force measurements, preferably having monotonic trends (Javed et al. 2015). The selection of features can be done by transforming them to another space or based on highest information content (Benkedjouh et al. 2013; Javed et al. 2015).

Prognostics modeling

This step aims at building an effective model that is capable of predicting the tool wear during machining process and estimating its life span to enable short-term or long-term decisions. The data-driven tool wear modeling is achieved in two steps: learning and testing. In the learning phase, data are used to establish model which learns a relation between input features and target measured wear. The learning step is directly linked to tool wear prediction performances in the test phase. For example lack of data, uncertainty of data collection/ processing, and varying context, etc., can strongly impact model performances. Moreover, in the learning phase model complexity, parameter initialization and computational time are the factors which should be properly addressed to build a right model.
In the testing phase, the learned model is used to predict the tool condition online and to estimate its life span, when the predicted wear intersects the failure threshold. However, in this step it is essential to provide confidence to the predictions, without that prognostics is not useful (Fig. 2).

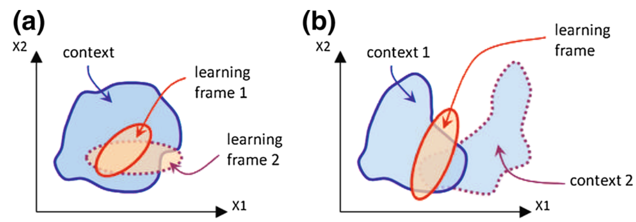

Fig. 3 Illustration of challenges: **a** robustness, **b** reliability

Open challenges of prognostics modeling

According to literature, various approaches for prognostics exist, i.e., physics based, data-driven and hybrid approaches (Javed 2014). However, real prognostics systems to meet industrial challenges are still scarce. This can be due to inherent uncertainties associated to deterioration process, lack of sufficient data quantities, sensor noise, unknown operating conditions, and engineering variations, which prevents building prognostics models that can accurately capture the evolution of degradation. In other words, highly complex and non-linear operational environment of industrial equipment makes it hard to establish efficient prognostics models, that are robust enough to tolerate uncertainty of data, and reliable enough to show acceptable performances under diverse conditions (Javed et al. 2012; Hu et al. 2012; An et al. 2015). Robustness of prognostics models appears to be an important aspect (Liao 2010), and still remains an important issue to be resolved (Javed et al. 2012; Camci and Chinnam 2010). Besides that, reliability performance is also crucial to prognostics. A reliable prognostics model should be capable of dealing with variations in data, that are directly associated to the context (e.g. for machining its variable geometry/ dimensions of cutters, materials differences of components, etc.). It is found that robustness and reliability of a prognostics model are closely related (Peng et al. 2010), and both should be considered as essential to ensure accuracy of estimates (Javed et al. 2012). Moreover, prognostics model has to be chosen

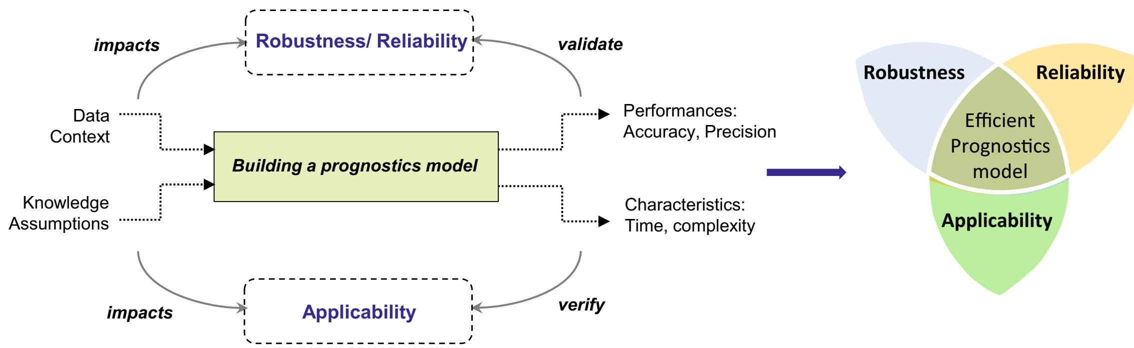

Fig. 4 Enhancing prognostics—frame and expected performances

according to implementation requirements and constraints that can limit its applicability (Javed et al. 2012; Sikorska and Hodkiewicz 2011).

Finally, prognostics model should be enhanced by handling simultaneously all three challenges, robustness, reliability[1] and applicability, which are still open areas. However, practitioners still encounter difficulties to identify their relationships and to define them. On the basis of our previous work we define them as follows (Javed et al. 2012).

- Robustness of prognostics—it can be defined as "the ability of a prognostics approach to be insensitive to inherent variations of the input data".

It means that, whatever the subset from the entire learning frame is used, the performances of a robust prognostics model should not impair (i.e., steady performance). In other words, robustness validates prognostics performance when exposed to variations in learning data having same context, i.e., operating conditions, geometrical scale, material, etc. An illustration is given in Fig. 3.

- Reliability of prognostics—it can be defined as "the ability of a prognostics approach to be consistent in situations when new/ unknown data are presented".

It means that, reliability validates prognostics performances when data with different context are presented to the model i.e., geometrical scale, material, operating conditions, etc. In other words, a reliable prognostics model can adapt variations related to context and can deal with uncertainty when exposed to new data with small deviation from learned cases (i.e., context is partially known), or when totally unknown data with large deviations are presented (i.e., unknown context). An illustration is given in Fig. 3.

- Applicability of prognostics—it can be defined as "the ability of a prognostics approach to be practically applied under industrial constraints".

The applicability verifies suitability or ease of implementation of a prognostics model for a particular application, i.e., requirements like failure definition, human intervention, model complexity, computation time, and theoretical limits of the approach or any hypothesis. A synthetic scheme of robust, reliable, applicable prognostics is shown in Fig. 4. Thus, validating the robustness, reliability performances and verifying the applicability of prognostics will enable practitioners to build an efficient prognostics model.

Choice of data-driven prognostics approach

The data-driven approaches are considered model free, as they do not need mathematical formulation of the process and solely depend on the data. To capture complex nonlinear relationships among data (e.g. features and tool wear) they learn from examples. In general, data-driven approaches have better applicability as compared to other prognostics approaches i.e., physics based or hybrid, due to their following advantages.

1. Better generality and system wide scope.
2. Do not require degradation process model.
3. Easy to implement and have low complexity.
4. Require few knowledge of the equipment.
5. Usually have low computation time.

With the advance of modern sensor, data storage and processing technologies, data-driven prognostics is becoming popular (Hu et al. 2012). According to literature, among data-driven approaches the Artificial neural networks (ANNs) are a special case of connectionist networks that are most com-

[1] Note: classical definition of reliability "the ability of an item to perform a required function under given conditions for a given time interval" (NF EN 13306 2010) is not retained here. Actually, the acception used in this paper is according to application of machine learning approaches in PHM, that do not consider reliability as dependability measure (Bosnić and Kononenko 2009).

monly used for prognostics (Zemouri et al. 2010; Ren et al. 2015), and for prediction in milling operations (Grzenda and Bustillo 2013). However, according to discussions in "Open challenges of prognostics modeling" section, to select a prognostics approach for a particular application, the applicability requirements must be verified. In addition, the robustness and reliability of the prognostics model must improved and validated, to ensure its effectiveness.

Brief overview of ANN architectures

Constructing a good neural network model is non-trivial task and practitioners still have to encounter several issues that may affect the performance of ANNs and limit their applicability (Singh and Balasundaram 2007). As for examples, such problems involve: parameter initialization, complexity of hidden layer, activation functions, slow iterative tuning, local minima, over-fitting, generalization ability, etc., (Javed et al. 2012).

In general, ANNs are classified into two types of architectures: a feed-forward network (FFNN) and a recurrent neural network (RNN). A FFNN has connections in forward direction, and RNN has cyclic connections Fig. 5. It is mentioned, that around 95 % of literature is on FFNNs (Feng et al. 2009). However, such systems must be tuned to learn parameters like weights and bias, in order to fit the studied problem. According to literature, the most popular learning scheme for FFNN is Extreme Learning Machine (ELM) (Huang et al. 2004), and for RNN its Echo State Network (ESN) (Jaeger 2001). Unlike classical techniques for ANNs, the ELM and ESN avoid slow iterative learning and they are based on random projection. In brief, with ELM/ ESN algorithms, input-hidden layer/reservoir parameters are randomly initialized, and learning is only achieved by solving the inverse of a least square problem. In addition, both are sensitive to the number of neurons in the hidden layer/ reservoir. The main differences of ELM and ESN are depicted in Fig. 5.

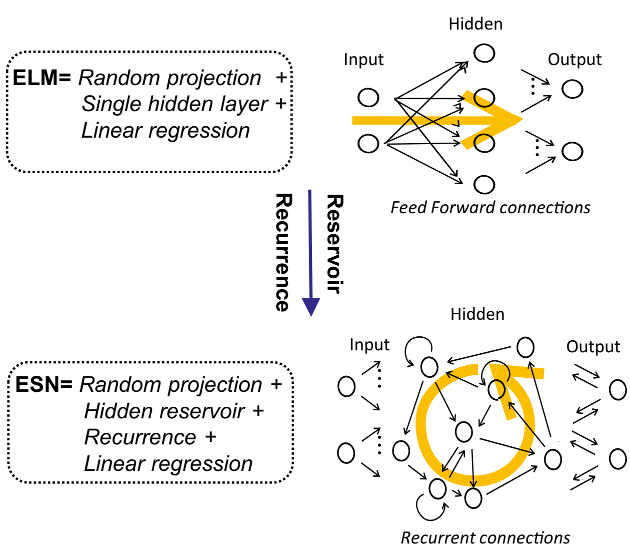

Fig. 5 ELM for FFNN and ESN for RNN (Jaeger 2002)

To the best of our knowledge, ELM has been proved for its universal approximation capability (Huang and Chen 2007, 2008; Huang et al. 2006), whereas for ESN its not the case. In addition, recent survey shows the advantages of ELM over conventional methods to train ANNs (Huang et al. 2011). As a matter of fact, ELM is an effective algorithm with several advantages like: ease of use, quick learning speed and capability for nonlinear activation (Shamshirband et al. 2015). Such findings highlight the importance of ELM as a suitable candidate for prognostics.

The Extreme Learning Machine

Basically, ELM is a batch learning scheme for single hidden layer feed forward neural networks (SLFNs). A slight difference in architecture of ELM and typical SLFN is that, there is no bias for the neurons in the output layer. To initiate rapid learning operation, the input weights and hidden neurons biases are chosen randomly without any prior knowledge of hidden to output layers weights. The random parameters are also independent from the learning data. Consequently, ELM transforms into a system of linear equations and the unknown weights between the hidden layer and the output layer nodes can be determined analytically by applying Moore–Penrose generalized inverse method (Rao and Mitra 1971; Petkovi et al. 2016).

Let note n and m the numbers of inputs and outputs (i.e., targets), N the number of learning data samples (x_i, t_i), where $i \in [1 \ldots N]$, $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]^T \in \Re^n$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in \Re^m$, and $\tilde{N}$ the number of hidden nodes, each one with an activation function $f(x)$. To minimize the difference between network output o_j and given target t_j, $\sum_{j=1}^{\tilde{N}} \left\| o_j - t_j \right\| = 0$, there exist β_k, w_k and b_k such that:

$$\sum_{k=1}^{\tilde{N}} \beta_k . f(w_k . x_j + b_k) = t_j \ , \ \ j = 1, 2, \ldots, N \tag{1}$$

where $w_k = [w_{k1}, w_{k2}, \ldots, w_{kn}]^T \in \Re^n$, is an input weight vector connecting the kth hidden neuron to the input layer neurons, $(w_k \cdot x_j)$ is the inner product of weights and inputs, and $b_k \in \Re$ is the bias of kth neuron of hidden layer. Also, $\beta_k = [\beta_{k1}, \beta_{k2}, \ldots, \beta_{km}]^T \in \Re^m$, is the weight vector to connect the kth neuron of hidden layer and output neurons. Eq. 1 can be expressed in matrix form as,

$$H\beta = T \tag{2}$$

$$H\left(w_1,\ldots,w_{\tilde{N}},x_1,\ldots,x_{\tilde{N}},b_1,\ldots,b_{\tilde{N}}\right) = \begin{bmatrix} f(w_1.x_1+b_1) & \ldots & f(w_{\tilde{N}}.x_1+b_{\tilde{N}}) \\ \vdots & \ldots & \vdots \\ f(w_1.x_N+b_1) & \ldots & f(w_{\tilde{N}}.x_N+b_{\tilde{N}}) \end{bmatrix}_{N\times\tilde{N}} \quad (3)$$

$$\beta = \left[\beta_1^T \cdots \beta_{\tilde{N}}^T\right]_{\tilde{N}\times m} and \ T = \left[t_1^T \cdots t_N^T\right]_{N\times m} \quad (4)$$

The least square solution of the linear system defined in Eq. (2),

$$\beta = H^{\dagger}T = \left(H^T H\right)^{-1} H^T T \quad (5)$$

where $H^{\dagger}$ is the Moore–Penrose generalized inverse.
In view of expected performances of a prognostics model highlighted in "Open challenges of prognostics modeling" section, practical considerations related to model accuracy and implementation issues should be addressed for real applications. In context to that, benefits, issues and requirements of ELM algorithm are given as follows.

Benefits

- ELM does not require slow iterative learning and it is one-pass algorithm.
- ELM has only one control parameter to be manually tuned, i.e. number of hidden neurons.

In general, rapid learning ability and less human intervention shows the better applicability of ELM, which makes it suitable for real applications (Huang et al. 2006; Bhat et al. 2008). Also, recent study confirms the advantages of ELM over earlier approaches for ANN (Huang et al. 2011).

Issues and requirements

- Due to random initialization of parameters (weights and bias), ELM model may require a complex hidden layer (Rajesh and Prakash 2011). This may cause ill-condition, and reduce robustness of ELM to encounter variations in the input data, and the expected output of the model may not be close to the real output (Zhao et al. 2009). The variance of randomly initialized weights can affect model generalization ability which should also be considered. Moreover, random initialization of parameters results poor consistency of the algorithm. In other words, the algorithm gives different solution for each run, which makes it less reliable.
- It is required to carefully choose hidden neuron activation functions that can participate in better convergence of the algorithm, ability to handle nonlinear inputs, and also result to a compact structure of network for a suitable level of accuracy (Javed et al. 2012; Jalab and Ibrahim 2011; Huang and Chen 2008).
- ELM does not quantify uncertainty of model like any ANN. Therefore, in terms of prognostics, a single ELM model lacks in real tangible foresight. Thus, it is required to bracket unknown future to show reliability of estimates and to enable timely decisions.

Obviously no model is perfect, following topic presents the proposition of an improved data-driven connectionist approach to encounter robustness and reliability challenges of prognostics modeling. The proposed approach is based on our improved variant of ELM namely, the Summation Wavelet-Extreme Learning Machine with new incremental learning scheme.

Proposed data-driven approach

Summation Wavelet-Extreme Learning Machine

SW-ELM combines ANN and wavelet theory for estimation or predictions problems, which appears to be an effective tool for different applications in industry (Javed et al. 2014). SW-ELM also represents one-pass learning for $SLFN$. It benefits from an improved parameter initialization phase to minimize the impact of random weights and bias (of input-hidden layer), structure with dual activation functions that can handle nonlinearity in a better manner and it also works on actual scales of the data.

Structure and parameters

To address the issues and to meet the requirements highlighted in "The Extreme Learning Machine" section, the differences with ELM are as follows:

- Structure: each hidden node holds a parallel conjunction of two different activation functions (f_1 and f_2) rather than a single activation function. Output from a hidden neuron is the average value from dual activations ($\bar{f} = (f_1 + f_2)/2$) (see Fig. 6).
- Activation function: convergence of algorithm is improved by an inverse hyperbolic sine (Eq. 6) and a Morlet wavelet (Eq. 7) as dual activation functions, which operate on array (X) element-wise ($x_j,\ j = 1, 2, \ldots, n$).

$$f_1 = \theta(X) = log\left[x + \left(x^2+1\right)^{1/2}\right] \quad (6)$$

$$f_2 = \psi(X) = cos(5x)\, e^{(-0.5x^2)} \quad (7)$$

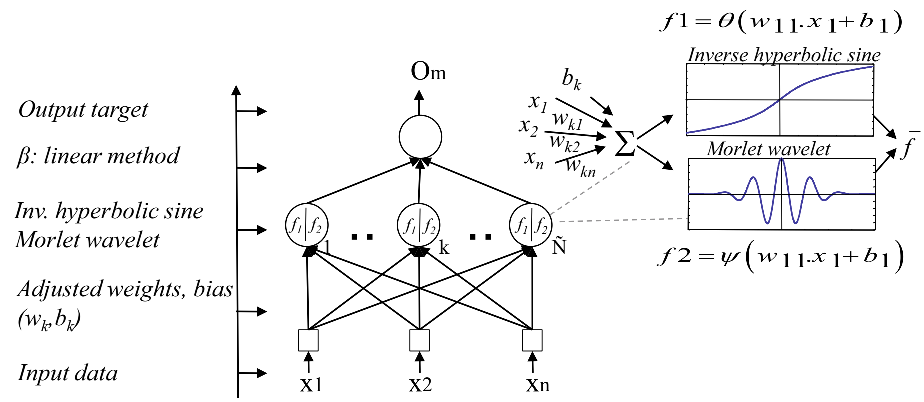

Fig. 6 Machine learning view of SW-ELM

- Parameter initialization: to provide a better starting point to the algorithm, two types of parameters have to be considered: those from the wavelets (dilation and translation) adapted by a heuristic procedure (Oussar and Dreyfus 2000), and those from the $SLFN$ (weights and bias for input to hidden layer nodes), initialized by Nguyen Widrow (NW) procedure (Nguyen and Widrow 1990).

Learning scheme

Given N learning data samples (x_i, t_i) and the number of hidden nodes $\tilde{N}$, each with activation functions f_1 and f_2. To minimize the difference between network output o_j and target t_j, $\sum_{j=1}^{\tilde{N}} \|o_j - t_j\| = 0$, there exist β_k, w_k and b_k such that:

$$\sum_{k=1}^{\tilde{N}} \hat{\beta}_k \bar{f}\left[(\theta, \psi)\left(w_k.x_j + b_k\right)\right] = t_j, \quad j = 1, 2, \ldots, N \tag{8}$$

Equation (8) can be expressed in matrix form as,

$$H_{avg}\hat{\beta} = T \tag{9}$$

$$H_{avg}\left(w_1, \ldots, w_{\tilde{N}}, x_1, \ldots, x_{\tilde{N}}, b_1, \ldots, b_{\tilde{N}}\right) \tag{10}$$

$$= \bar{f}(\theta, \psi)\begin{bmatrix} (w_1.x_1 + b_1) & \ldots & (w_{\tilde{N}}.x_1 + b_{\tilde{N}}) \\ \vdots & \ldots & \vdots \\ (w_1.x_N + b_1) & \ldots & (w_{\tilde{N}}.x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \tag{11}$$

$$\hat{\beta} = \left[\beta_1^T \cdots \beta_{\tilde{N}}^T\right]_{\tilde{N} \times m} \; and \; T = \left[t_1^T \cdots t_N^T\right]_{N \times m} \tag{12}$$

The least square solution of the linear system defined in Eq. (9).

$$\hat{\beta} = H_{avg}^{\dagger} T = \left(H_{avg}^T H_{avg}\right)^{-1} H_{avg}^T T \tag{13}$$

where $H_{avg}^{\dagger}$ represents the Moore–Penrose generalized inverse. The SW-ELM algorithm is given in algorithm 1.

SW-ELM ensemble with incremental learning

Although, ELM based algorithms have several advantages over traditional methods for SLFN, but the main shortcoming can be that their solution vary for each run due to random parameters initialization, which can result poor reliability performances. This issue is also for classical ANNs. Also, such methods do not furnish any indication about the quality of outcomes in order to facilitate practitioner with decision making. That is, considering the uncertainties which arise either due to model misspecification or either due to variations of input data by probabilistic events (Khosravi et al. 2011). Although there is no single algorithm or model for prognostics that works for all sorts of situations, the ensemble of multiple models appears to be less likely in error than an individual model (Khosravi et al. 2011; Hu et al. 2012). Due to such issues, in literature it is preferred to apply an ensemble of multiple models to improve robustness and to show reliability of estimates (Hu et al. 2012). Therefore, the combined estimate obtained from an ensemble of models is more accurate as compared to a single model. That also indicate the presence of uncertainty and can facilitate decision making for further plan of actions. A detailed review about ELM ensembles can be found in Huang et al. (2011). In this paper, the ensemble strategy is achieved by integrating several SW-ELM models, where each individual model is initialized with different parameters (Fig. 7). Following that, the desired output $\overline{O}$ can be obtained by averaging outputs from multiple SW-ELM models.

$$\overline{O}_j = \frac{1}{M}\sum_{m=1}^{M} \hat{o}_j^m \tag{14}$$

where $\hat{o}_j^m$ is the predicted output of mth model against the jth input sample. Tool wear prediction task continues with each input sample (i.e., after a cut) along with confidence bounds. The life span of cutting tool is estimated when the predicted value $\overline{O}_j$ intersects the failure threshold (FT), as given in Eq. (15).

Algorithm 1 Learning scheme of the SW-ELM

Require - N learning data samples (x_i, t_i), n inputs $(j = 1 \dots n)$, $\tilde{N}$ hidden nodes $(k = 1 \dots \tilde{N})$.
- An inverse hyperbolic sine and a Morlet activation functions (θ and ψ).

Ensure - Initialize weights and bias from SLFN, initialize Morlet parameters.
- Find output weights matrix β to minimize the difference between the network outputs and the targets.

SW-ELM learning procedure

Initialization of wavelet parameters
- Define the input space domain intervals
  - Compute $[x_{jmin}\ ;\ x_{jmax}]$: {domain containing the input item x_j in all observed samples}
- Define dilation and translation parameters per domain
  - Compute $d_{kj} = 0.2 \times [x_{jmax} - x_{jmin}]$: {temporal dilatation parameter for input item x_j}
  - Compute $m_{kj} = [x_{jmin} + x_{jmax}]/2$: {temporal translation parameter for input item x_j}
- Initialize Morlet parameters (a_k and b_k)
  - Compute $a_k = mean(d_{kj})_{j=1 \dots n}$: {dilatation factor}
  - Compute $b_k = mean(m_{kj})_{j=1 \dots n}$: {translation factor}

Initialization of weights and bias parameters by Nguyen Widrow (NW) approach
- Initialize small (random) input weights $w_{k(old)}$ in $[-0.5\ ;\ +0.5]$: {weights from input nodes to hidden nodes}
- Adjust weights parameters by applying NW approach
  - Compute $\beta_{factor} = C \times \tilde{N}^{\frac{1}{n}}$: {$C$ is a constant ≤ 0.7}
  - Compute $w_{k(new)} = \beta_{factor} \times \frac{w_{k(old)}}{\| w_{k(old)} \|}$: {normalized weights}
- Initialize bias values b_k
  - b_k = random number between $-\beta_{factor}$ and $+\beta_{factor}$

Adjust linear parameters: the ones from the hidden to the output layers
- Obtain hidden layer output matrix H_{avg} using Eq. 11
- Find the output weight matrix $\hat{\beta}$ in Eq. 13 by applying Moore–Penrose generalized inverse procedure

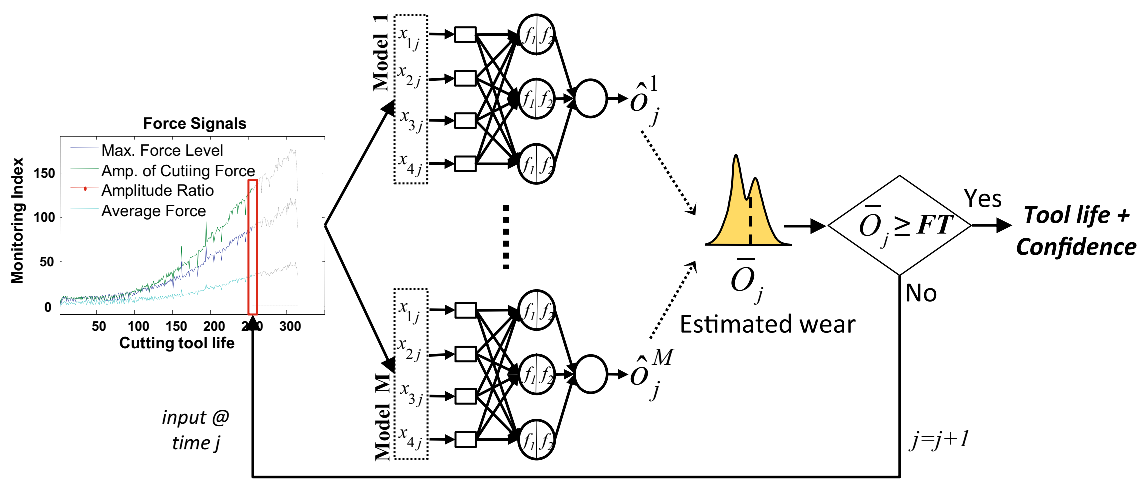

Fig. 7 Structure of SW-ELM ensemble

$$\overline{O}_j \geq FT \tag{15}$$

According to data-driven framework ("Data-driven tool wear monitoring framework" section), trained model in used to predict tool wear online. However, throughout the prediction process the model parameters remain static, which can lead to poor predictions. To address this issue a new incremental learning procedure is proposed, which uses the input features data and re-simulated data from predictions (up to current time), to retrain the data-driven model online.

In order to elaborate incremental learning procedure for the ensemble structure, consider learning data record of 630 samples (inputs and targets) from two cutting tools. During an online application on a new cutting tool, the input features data sample (after a cut) and their corresponding predicted tool wear value from the SW-ELME are stored sequentially in the learning data record (which becomes 631 samples). Following that, the SW-ELME is retrained with that data and model parameters (i.e., weights and bias) are updated before the next input. The learning procedure continues after each cut until the FT is reached. This proposition allows performing incremental learning without actual tool wear values and using artificial data from predictions, which enables improving the adaptability of prognostics model and managing its uncertainty.

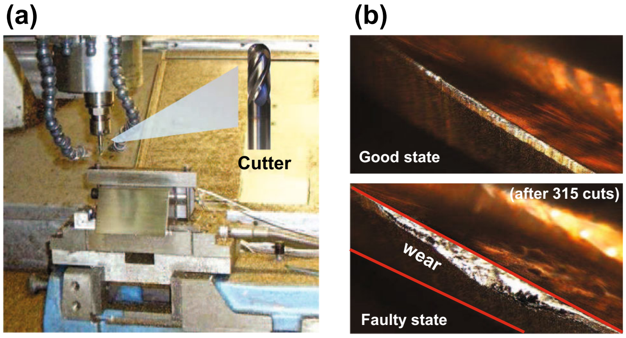

Fig. 8 **a** Work piece and cutter, **b** cutter berfore and after

Note that, due to rapid learning ability of SW-ELM algorithm, the proposed incremental learning can be computationally efficient. However, the computational time can increase with the complexity of ensemble structure.

Case study: tool condition monitoring

Experimental arrangements

To investigate suitability of the proposed approach, real data from a high speed CNC machine are used to monitor condition of the cutting tools. The experimental data related to wear of cutting tools are provided by SIMTECH Institute in Singapore, where a high-speed CNC milling machine (Roders Tech RFM 760) was used as a testbed under constant operating conditions. In the machining treatment, the spindle speed was set to 10,360 rpm. The material of workpiece used was Inconel 718. Also, 3 tungsten carbide cutters with 6 mm ball-nose/ 3-flutes were used in the experiments. To achieve high quality surface finish, the metal workpiece was made via face milling to remove its original skin layer of hard particles Fig. 8a. During milling, the feed rate was 1.555 mm/min, the Y depth of generated cuts was 0.125 mm and the Z depth of cuts was 0.2 mm (Massol et al. 2010).

Data acquisition and processing

During the cutting operation, authors of the experiments recorded data from cutting force and vibration measurements Fig. 9. The cutting operation was stopped after each cut and tool wear was measured via Olympic microscope, Fig. 8b. The acquired data are composed of 315 cuts made by three different cutters, namely C33, C18 and C09. In this paper, the cutting force data are used for tool condition monitoring ("Data processing" secion). A total of 16 main features are derived from force signals and a subset of four features are selected to train models (Table 1). Figure 9c, d gives an illustration of force features and the tool wear. The details of feature extraction and selection are given in (Li et al. 2009; Zhou et al. 2006; Massol et al. 2010).

Most importantly, even if the operating conditions are constant, cutting force is affected by: cutter geometry, coating and properties of workpiece, which impacts the reliability of tool wear estimation models. Considering complications of tool wear modeling, it is important to highlight the characteristics of all cutters that were used. That is, cutting tools C33 and C18 had the same geometry but different coatings, while cutting tool C09 has its own geometry and coating, Table 2.

Tool wear model settings and performance metrics

Simulations in following sections are given in two parts.

1. A comparison of tool wear prediction models to encounter prognostics challenges ("Open challenges of prognostics modeling" section).

Given a learning dataset for a model (SW-ELM, ELM or ESN), 100 trials are performed and test performances are averaged for different complexities of hidden layer/ reservoir (i.e., 4–20 hidden nodes). It should be noted that, either for model robustness or reliability analysis ("Robustness and applicability: results discussion" and "Reliability and applicability: results discussion" sections), the model performance for a single trial is equal to an accumulated performance from ten different simulations for which data subsets (of cutters) are learned after random permutation (to introduce data uncertainty). The tests are performed on remaining data in chronological order (Fig. 10).
Besides that, to improve the generalization performance of ELM small random weights of input-hidden nodes are initialized i.e., $[-0.1, +0.1]$ rather than $[-1, +1]$. The ESN is used according to its default settings given in (Echo state network).

2. Adaptive ensemble to predict tool wear, estimate tool life span and give confidence for decision making.

The simulations aim to show improved reliability of the proposed SW-ELME model and its applicability for an online application. Learning and testing are performed using *leave-one-out* strategy with three different cutters data. For example, learn data from cutters C33, C18 and test cutter C09 to predict tool wear until the failure threshold (FT) is reached, which is the max value of tool wear. However, it can be difficult to generalize tool wear prediction model on cutting tools data that are not included in the learning phase. Therefore, complexity of tests can be clearly understood by comparing wear patterns from all three cutters in Fig. 11. Further details on simulation setting are given in "Adaptive SW-ELME and its reliability" section.

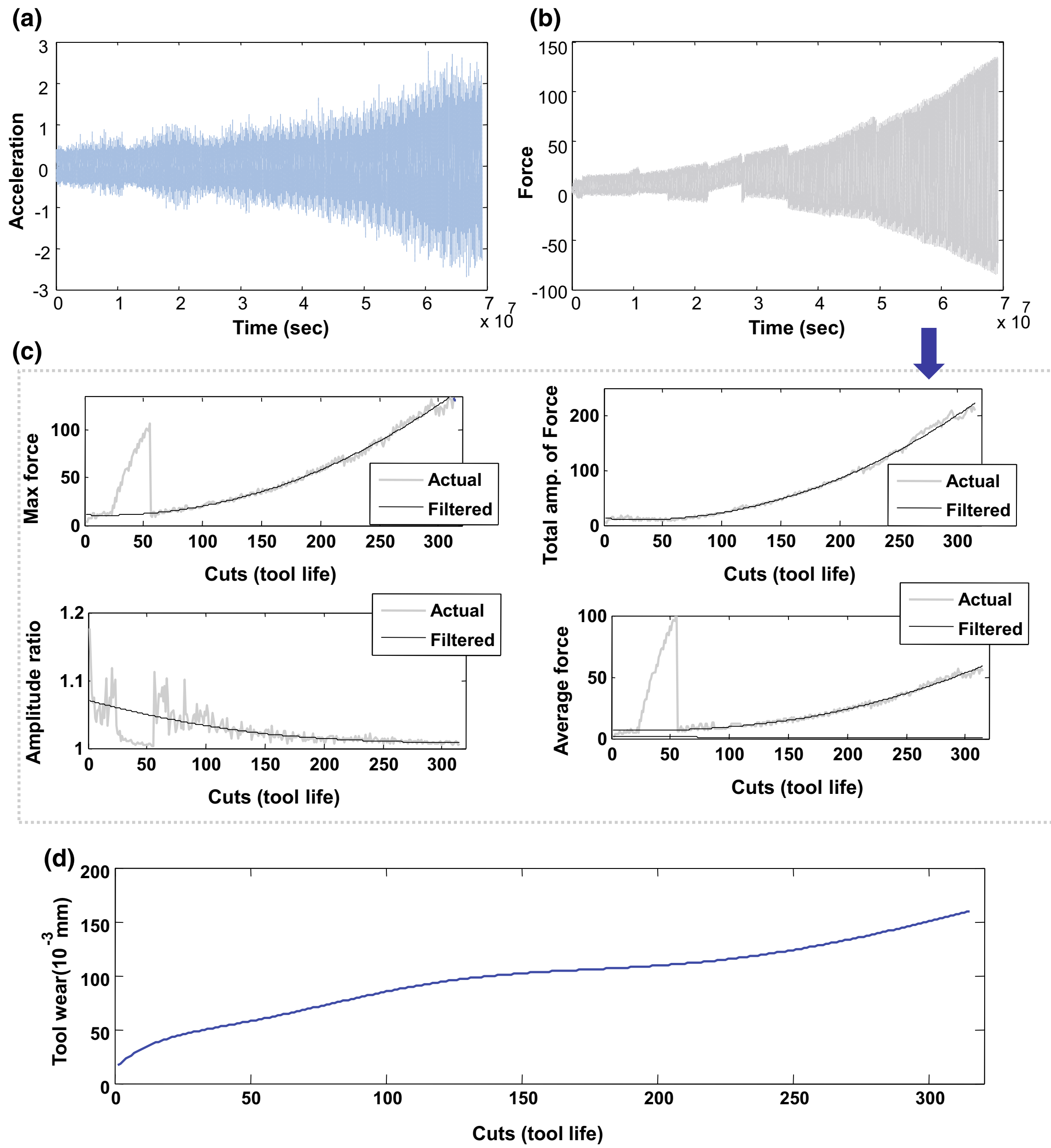

Fig. 9 Cutter C33 **a** Acceleration, **b** force signals, **c** force features and **d** tool wear (C33)

Table 1 Selected force features

No	Force feature
1	Maximum force level
2	Total amplitude of cutting force
3	Amplitude ratio
4	Average force

Table 2 Type of cutting tools used during experiments

Cutters	Geometry	Coating
C33	Geom1	Coat1
C18		Coat2
C09	Geom2	Coat3

To discuss the robustness, reliability, and applicability of the wear estimation ("Open challenges of prognostics modeling" section), model performances are assessed in terms of *accuracy*, *network complexity* and *computation time*. More precisely, metrics for performance evaluation are: coefficient of determination ($R2$) that should be close to 1, complexity of hidden layer, and learning/testing time in seconds (s).

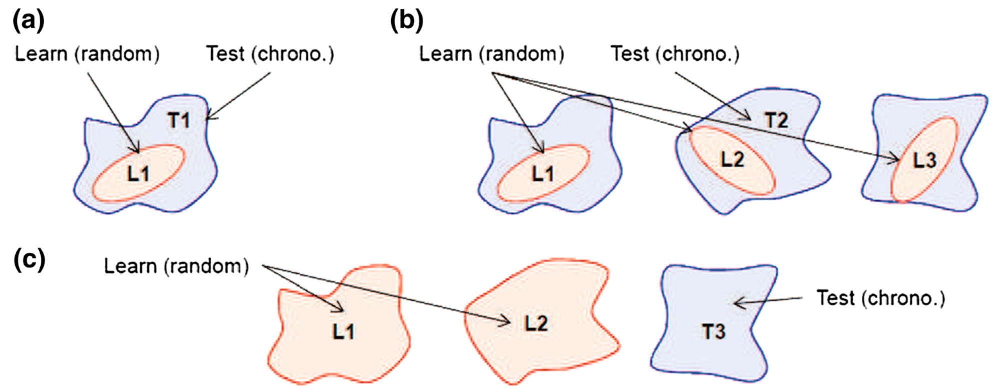

Fig. 10 Tests for robustness and reliability

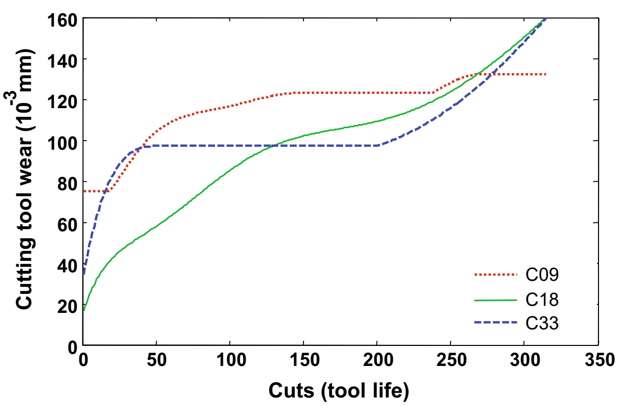

Fig. 11 Wear patterns from all cutters

Table 3 Robustness and applicability for a single cutter model

Cutter 09	SW-ELM	ELM	ESN
Hidden nodes	16	16	16
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.0009	**0.0005**	0.014
*R*2	**0.824**	0.796	0.542
Cutter 18	SW-ELM	ELM	ESN
Hidden nodes	12	12	12
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.0007	**0.0004**	0.013
*R*2	**0.955**	0.946	0.59

Bold values indicate the better results

Comparison of connectionist approaches

Robustness and applicability: results discussion

This case aims at evaluating the robustness of tool wear model when exposed to variations in the learning data having same context. Therefore for learning, the dataset from a single cutting tool is created by randomly selecting 150 data samples, while rest of the data of 165 samples are presented in chronological order to test the accuracy of the learned model (Fig. 10a). This procedure is repeated ten times for the model-cutting tool couple and considered as a "single trial". That is creating random training input datasets and assessing model accuracies on test sets to evaluate robustness. A comparative analysis on robust tool wear prediction performances is given in Table 3.

Among model-cutting tool couples and even with small learning data, the SW-ELM model showed better robustness for all tests as compared to ELM and ESN. However, the average learning time of ELM is faster than other approaches. The detailed simulations results are presented in Fig. 12. In brief, Fig. 12a, shows an average accuracy (*R*2) performances for five different network complexities, where the best results are achieved by SW-EM for tests on cutter C09 and C18 (with no. of hidden nodes 16 and 12). Figure 12b compares the steadiness of all models (SW-ELM, ELM and ESN) for 100 trials. One can see that SW-ELM is more robust to input variations as its accuracy (*R*2) is stable for 100 trials, for the tests on both cutters. Figure 12c compares average results of tool wear prediction (from 100 trials) on cutter C09.

Reliability and applicability: results discussion

Reliability on partially known data This case aims at evaluating the reliability of tool wear model when exposed to variations in the data from multiple cutters having different attributes (geometrical scale and coating). In order to build a "multi-cutters" model, a partial dataset of 450 samples from all cutters are presented in random order for learning and data of 165 samples in chronological order from any of these cutters are used for the test (Fig. 10b). Like the previous case "Robustness and applicability: results discussion" section, this procedure is repeated 10 times for each multi-cutters model and considered as "single trial". A comparative analysis on reliable tool wear prediction performances is given in Table 4.

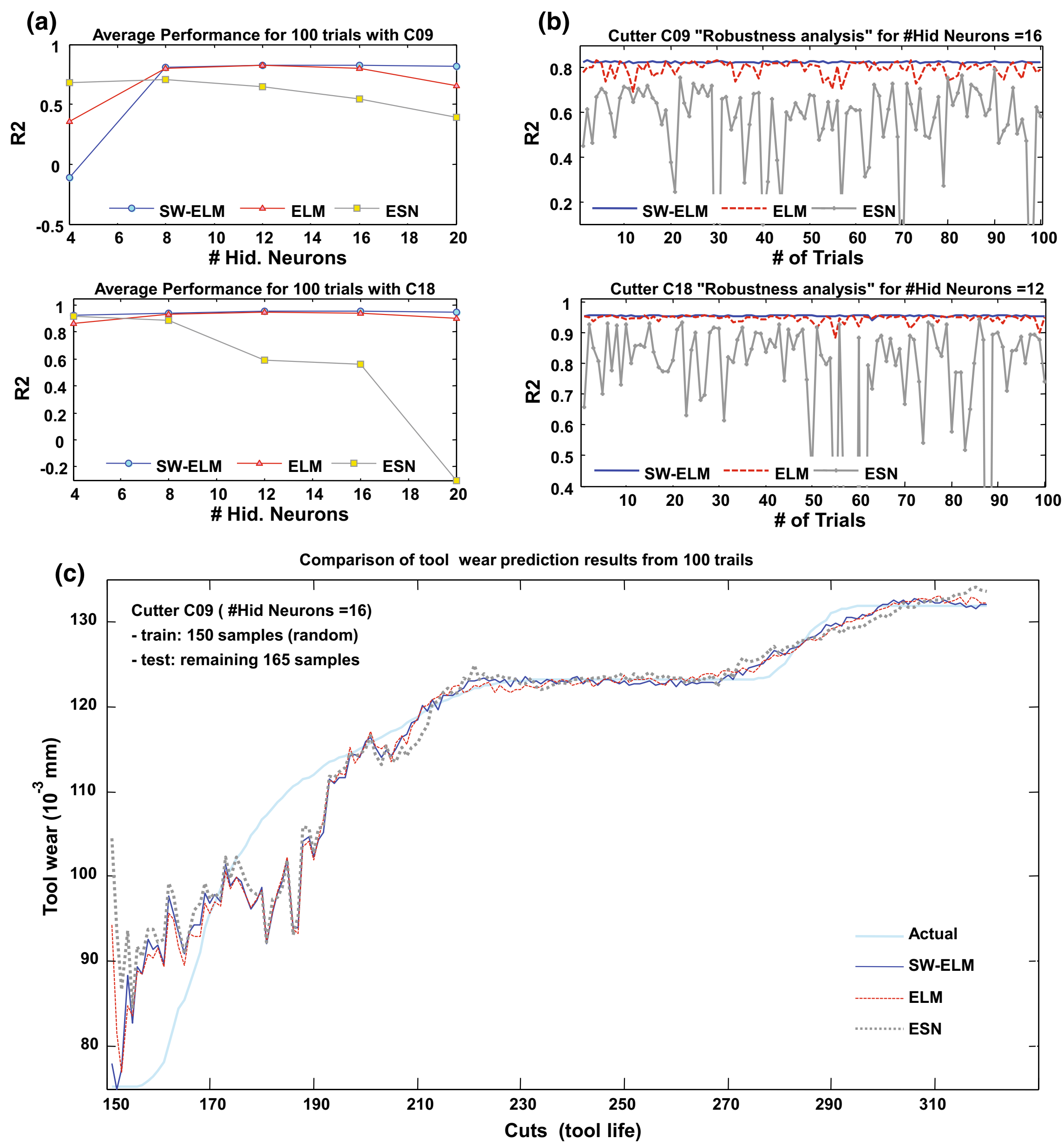

Fig. 12 Robustness analysis with partially unknown data for a "single cutter" model

It can be observed from results that, even if the learning data are increased, ELM based methods are still faster than ESN. The average learning times for both tests show that, ELM is less time consuming for the same complexity of models. As far as accuracy ($R2$) is concerned, SW-ELM showed better reliability performances on cutters data with different attributes. The detailed simulations results are presented in Fig. 13. In brief, Fig. 13a, shows an average accuracy ($R2$) performances for 5 different network complexities, where best results are achieved by SW-EM for tests on cutter C18 and C33 (with no. of hidden nodes 20 and 16). Considering these results, Fig. 13b compares the steadiness of all models (SW-ELM, ELM and ESN) for 100 trials. One can see that SW-ELM is more stable to input variations, as its test accuracy ($R2$) is consistent for 100 trials on cutters i.e., C18 and C33. Finally, Fig. 13c compares average results of tool wear prediction (from 100 trials) on cutter C33.

Reliability on totally unknown data This case aim at evaluating the reliability performances of wear estimation models (SW-ELM, ELM and ESN) when unknown cutters data with different attributes (geometrical scale and coating) are presented for tests. For this purpose, test performances are assessed by *leave-one-out* strategy. That is, by establishing a reference model from learning complete data of two different

Table 4 Reliability and applicability for three cutters models

Train: C33, C18, C09 Test: C18	SW-ELM	ELM	ESN
Hidden nodes	20	20	20
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.002	**0.001**	0.04
$R2$	**0.837**	0.836	0.817
Train: C33, C18, C09 Test: C33	SW-ELM	ELM	ESN
Hidden nodes	16	16	16
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.002	**0.0009**	0.04
$R2$	**0.847**	0.80	0.75

Bold values indicate the better results

cutting tools and testing its tool wear prediction capability on data from another cutter that was totally unknown. Therefore, learning data of 630 samples from two different cutters are presented randomly to train the model, whereas 315 data samples from a third cutter are presented in a chronological order for testing i.e., performed once for a trial unlike previous cases (see Fig. 10c). This procedure is repeated for 100 trails for different model complexities (i.e., no. of hidden neuron 4–20). Averaged performances from each case of multi-tool model are given in Table 5.

It can be observed form all tests that SW-ELM has better reliability performance as compared to ELM and ESN. The averaged accuracy performance of SW-ELM for the tests is also improved from our previous results in (Javed et al. 2012). Note that, for the tests on cutters C33 and C09 the accuracy of each approach decreased to a poor level i.e., $R2 < 0$. Therefore, the model reliability still needs to be improved when totally unknown data of different attributes are used, which is the aim of the following proposition. The detailed simulations results from the best tests case (for SW-ELM, ELM and ESN models) is illustrated in Figs. 14 and 15.

According to results in Fig. 14, the SW-ELM has better prediction performances as indicated by the stability of $R2$ values for 100 trials. The prediction results in Fig. 15 show that except SW-ELM all other models are unable to estimate tool initial wear (i.e., ELM and ESN). Moreover, all models are unable to estimate tool worn out state from the data of unknown cutter C18.

Synthesis

The main points from the results comparison are summarized as follows.

- All connectionist algorithms discussed above (SW-ELM, ELM, ESN), are based on random projection.
- For all tests on robustness and reliability performances, the SW-ELM outperform ELM and ESN algorithms.
- SW-ELM has better performances due to improved parameter initialization and structure with dual activation functions.
- SW-ELM algorithm requires two parameter to be set by the user, i.e., hidden neurons and parameter initialization constant C.
- SW-ELM takes twice the learning time than ELM.
- ELM algorithm has better applicability with only one parameter to manually tune and fastest training time.
- ESN requires several parameters to be set by the user and more training time as compared to ELM based methods.
- For some cases on reliability performance, ELM and ESN showed close accuracy performances.
- ESN is much more sensitive to input variations as compared to ELM based methods.
- Like any ANN, the SW-ELM, ELM and ESN can not quantify or manage prediction uncertainty.

Adaptive SW-ELME and its reliability

Considering the better performances of SW-ELM over ELM and ESN. This topic presents the reliability of SW-ELM ensemble with incremental learning scheme.

Simulation settings

The initial step is to determine complexity of hidden layer for a single SW-ELM model, which results satisfactory performances. Following that, multiple SW-ELM models of same complexity are integrated to produce averaged output. The complexity of hidden layer of each SW-ELM model is set to 7 neurons and the number of SW-ELM models for an ensemble is set to 50 (Fig. 7).

To reduce the uncertainty of estimates, features from each cutter data are filtered to obtain smooth trends by applying *rloess* filter with span value 0.9 (Fig. 9). Basically, rloess is a robust local regression filter that allocates lower weight to outliers, see Mathworks (2010). Each individual model is learned with same dataset, but initialized with different parameters, i.e., weights and bias. The parameter initialization constant is set to $C = 0.0001$.

The tests are performed on cutters data using leave-one-out strategy, e.g., learning C33, C18 and testing C09. The cutting tool life span determined when the predicted wear intersects FT (Eq. 15), which is set to the maximum tool wear at 315 cuts. For each test, the lower and upper confidence of tool wear predictions and the evolution of *probability density function* are given to quantify the uncertainty (Fig. 16). Also, the total time to learn-test SW-ELME

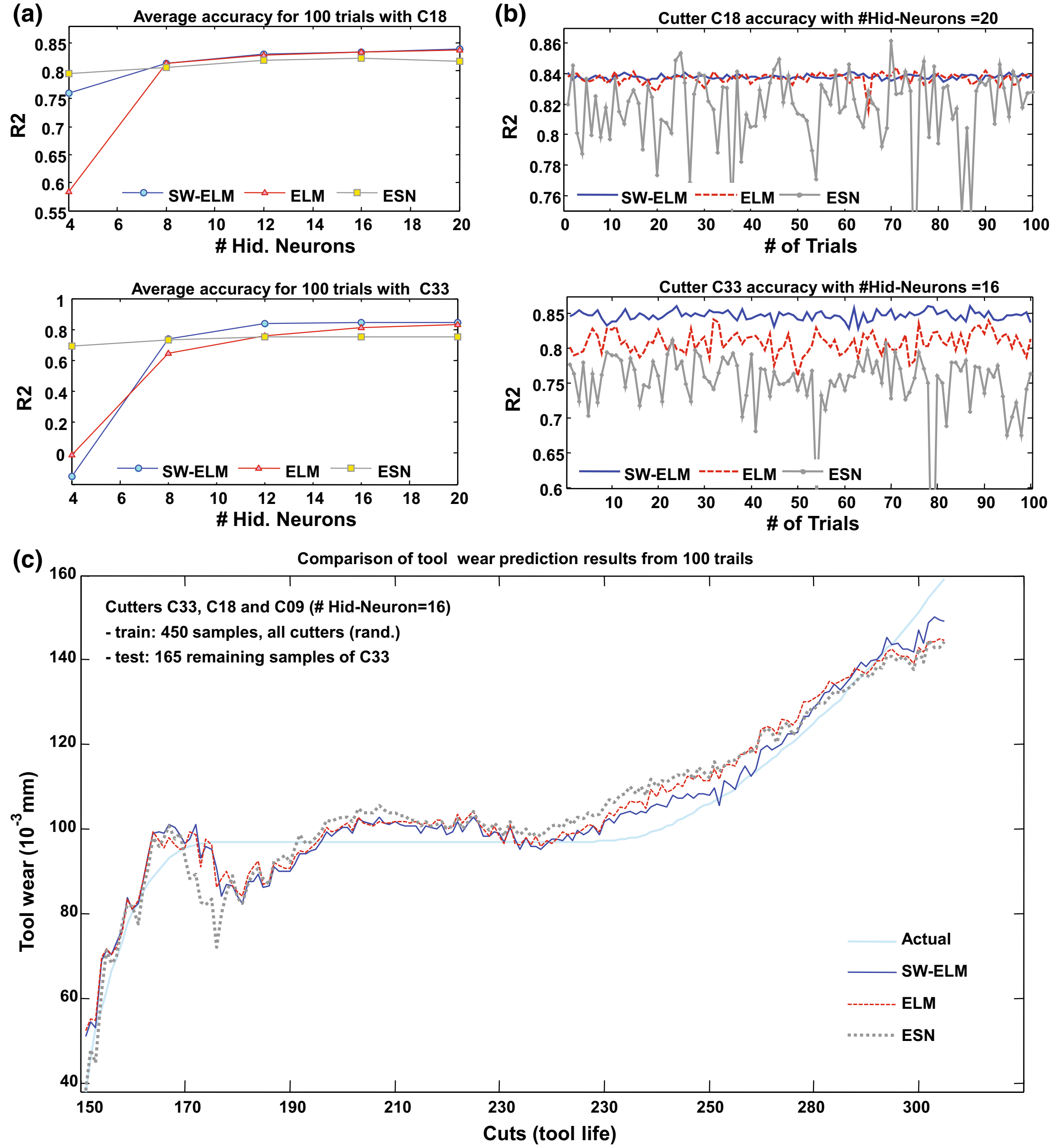

Fig. 13 Reliability analysis with partially unknown data for "multi-cutter" model

online is given to show its suitability for a real application.

SW-ELME: results discussion

Results from all test cases (i.e., C33, C18 and C09) using SW-ELME model are summarized in Table 6. According to those results, the SW-ELME has superior accuracy than a single SW-ELM model, which is indicated by low error values of estimated life span (i.e., number of cuts) in comparison to actual 315 cuts. Here, we compare the previous results on reliability of SW-ELM on unknown data from Table 5 and the results of SW-ELME from Table 6, one by one. In case of test cutter C18, the $R2$ is improved from 0.701 to 0.745. For test cutter C33 the $R2$ is improved from -0.5 to 0.89 and for cutter C09 the $R2$ is improved from -0.9 to 0.52, which is a significant improvement.

Table 5 Reliability and applicability for unknown data

Train: C33 & C09 Test: C18	SW-ELM	ELM	ESN
Hidden nodes	4	4	4
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.0009	**0.0004**	0.055
*R*2	**0.701**	0.44	0.6
Train: C09 & C18 Test: C33	SW-ELM	ELM	**ESN**
Hidden nodes	4	4	4
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.0008	**0.0004**	0.054
*R*2	**−0.5**	−1.3	−1.9
Train: C33 & C18 Test: C09	SW-ELM	ELM	ESN
Hidden nodes	16	16	16
Activation function	asinh & Morlet	sigmoid	tanh
Training time (s)	0.0026	**0.0013**	0.058
*R*2	**−0.73**	−1.2	−0.98

Bold values indicate the better results

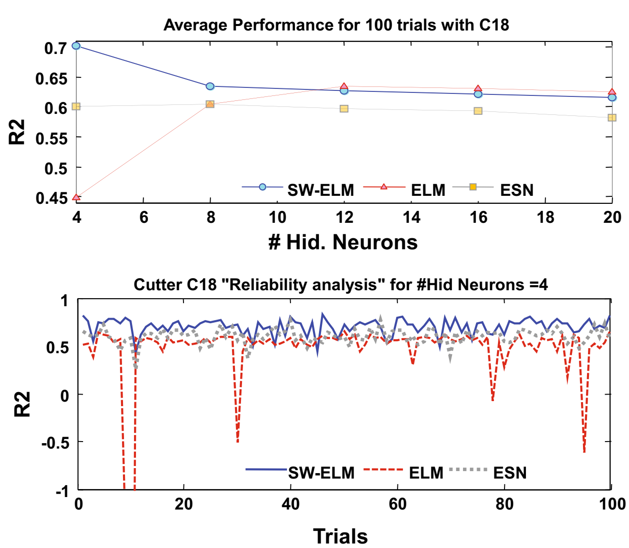

Fig. 14 Reliability analysis with totally unknown data

Moreover, for each test case the lower and upper confidence bounds indicate that the final target values are within the confidence level (Fig. 16). Finally, due to ensemble strategy and increased data, for each test case the total time for learning and testing (online) is around 2 minutes, which is quite satisfactory from practical point of view.

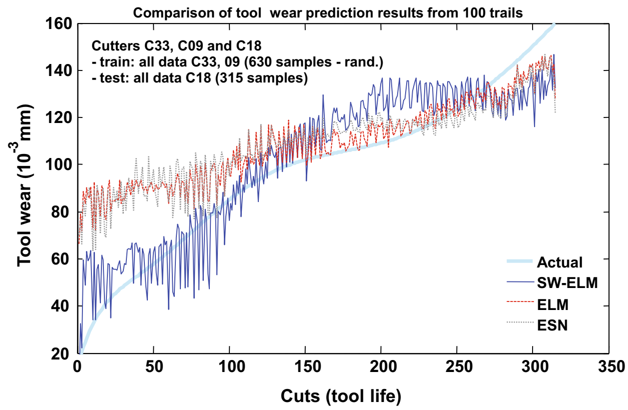

Fig. 15 Prediction results on unknown data

Conclusion

In this paper a data-driven prognostics approach is proposed for tool condition monitoring during high-speed milling operation. The proposed approach aims at transforming the monitoring data (from the cutting tool) into relevant models for predicting tool wear and estimating life span prior to costly failure. Considering highly complex and nonlinear nature of real industrial equipment, building accurate prognostics models is not a trivial task. Therefore, open challenges for prognostics modeling are defined for building an efficient prognostics approach, namely "robustness", "reliability", and "applicability". The data-driven models are established using rapid learning connectionist algorithms that are, Extreme Learning Machine (ELM), Summation Wavelet-Extreme Learning Machine (SW-ELM) and Echo State Network (ESN). The performances of connectionist algorithms are compared to encounter prognostics challenges using data from cutting tools with different geometric scale, coating and under constant operating conditions. Experimental results show that, SW-ELM algorithm outperforms ELM and ESN in terms of robustness and reliability performances for estimating tool condition, without compromising rapid learning ability. This shows better applicability of the SW-ELM.

Finally, an ensemble of SW-ELM (SW-ELME) models is proposed with incremental learning scheme to further improve the reliability of tool wear monitoring. The proposed SW-ELME enables predicting the tool wear and estimating

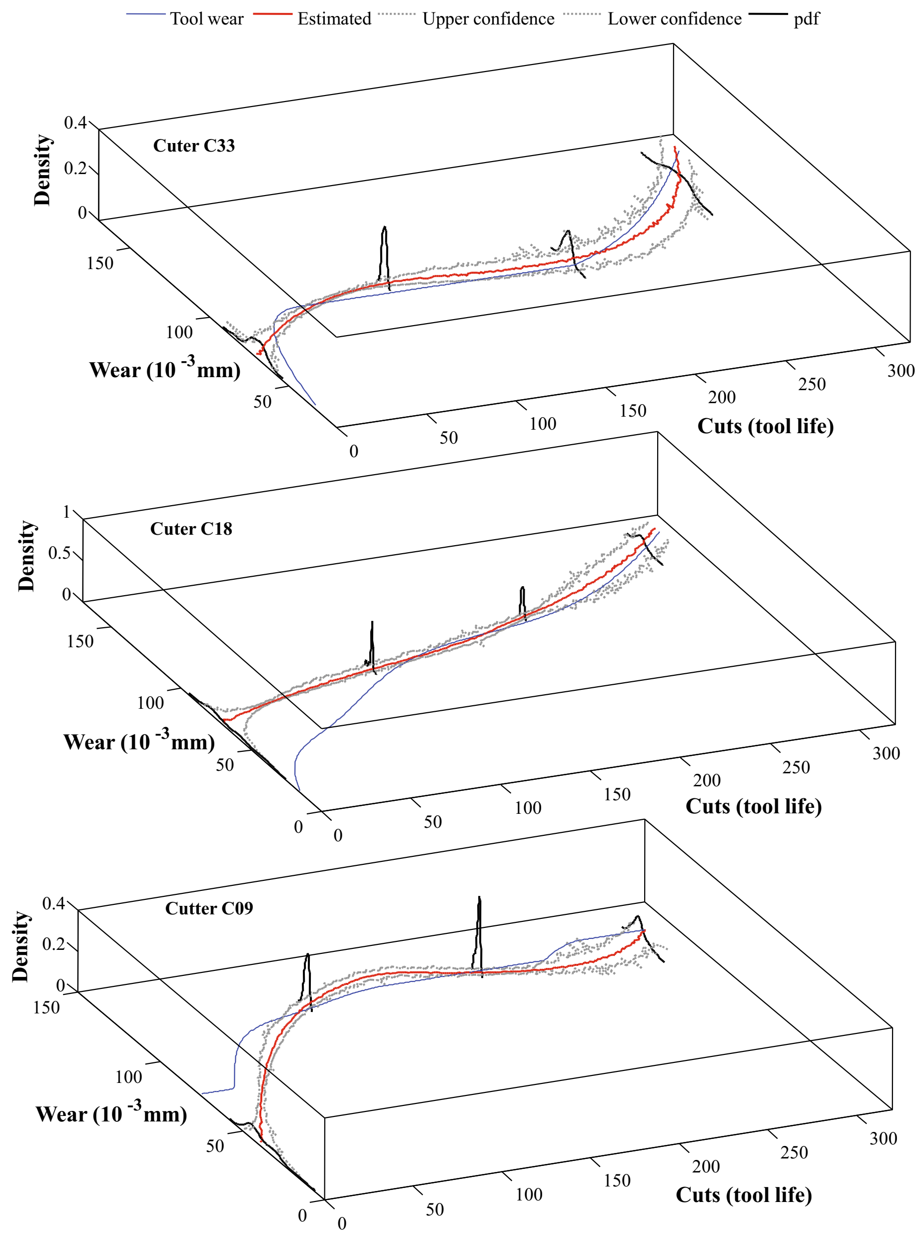

Fig. 16 Cutting tools wear estimation and uncertainty quantification

Table 6 Reliability and applicability for unknown data

Tool	Cuts	Estimated	Error	R2	Time
C33	315	313	2	0.89	119 (s)
C18	315	311	4	0.74	133 (s)
C09	315	303	12	0.52	112 (s)

life span online, with computation time around two minutes. Moreover, the SW-ELME provides confidence to predictions to facilitate decision making. Tool wear prediction results on all cutters data clearly show the significance of our proposition. However, reliability of the SW-ELME still needs to be addressed for tool condition monitoring application under variable operating conditions, which is the aim of our future works.

Acknowledgments This work was carried out within the Laboratory of Excellence ACTION funded by the French Government through the program “Investments for the future” managed by the National Agency for Research (ANR-11-LABX-01-01).

References

An, D., Kim, N. H., & Choi, J. H. (2015). Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering & System Safety*, *133*, 223–236.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2013). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing*, *26*(2), 213–223.

Bhat, A. U., Merchant, S., & Bhagwat, S. S. (2008). Prediction of melting point of organic compounds using extreme learning machines. *Industrial and Engineering Chemistry Research*, *47*(3), 920–925.

Bosnić, Z., & Kononenko, I. (2009). An overview of advances in reliability estimation of individual predictions in machine learning. *Intelligent Data Analysis*, *13*(2), 385–401.

Camci, F., & Chinnam, R. B. (2010). Health-state estimation and prognostics in machining processes. *IEEE Transactions on Automation Science and Engineering*, *7*(3), 581–597.

Cojbasic, Z., Petkovic, D., Shamshirband, S., Tong, C. W., Ch, S., Jankovic, P., et al. (2015). Surfaceroughnessprediction by extreme learning machine constructed withabrasivewater jet. *Precision Engineering*. doi:10.1016/j.precisioneng.2015.06.013.

Das, S., Hall, R., Herzog, S., Harrison, G., & Bodkin, M. (2011). Essential steps in prognostic health management. In *IEEE Conference on prognostics and health management*. Denver, CO, USA.

Ding, F., & He, Z. (2011). Cutting tool wear monitoring for reliability analysis using proportional hazards model. *The International Journal of Advanced Manufacturing Technology*, *57*(5–8), 565–574.

Echo state network. http://reservoir-computing.org/software.

NF EN 13306. (2010). Terminologie de la maintenance.

Feng, G., Huang, G. B., Lin, Q., & Gay, R. (2009). Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Transactions on Neural Networks*, *20*(8), 1352–1357.

Gao, R., Wang, L., Teti, R., Dornfeld, D., Kumara, S., Mori, M., et al. (2015). Cloud-enabled prognosis for manufacturing. *CIRP Annals-Manufacturing Technology*. doi:10.1016/j.cirp.2015.05.011.

Ghasempoor, A., Moore, T., & Jeswiet, J. (1998). On-line wear estimation using neural networks. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, *212*(2), 105–112.

Grzenda, M., & Bustillo, A. (2013). The evolutionary development of roughness prediction models. *Applied Soft Computing*, *13*(5), 2913–2922.

Haddadi, E., Shabghard, M. R., & Ettefagh, M. M. (2008). Effect of different tool edge conditions on wear detection by vibration spectrum analysis in turning operation. *Journal of Applied Sciences*, *8*(21), 3879–3886.

Hu, C., Youn, B. D., Wang, P., & Yoon, J. T. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. *Reliability Engineering & System Safety*, *103*, 120–135.

Huang, G. B., & Chen, L. (2007). Convex incremental extreme learning machine. *Neurocomputing*, *70*(16), 3056–3062.

Huang, G. B., & Chen, L. (2008). Enhanced random search based incremental extreme learning machine. *Neurocomputing*, *71*(16), 3460–3468.

Huang, G. B., Chen, L., & Siew, C. K. (2006). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, *17*(4), 879–892.

Huang, G. B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: A survey. *International Journal of Machine Learning and Cybernetics*, *2*(2), 107–122.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2004). Extreme learning machine: A new learning scheme of feedforward neural networks. In *International Joint conference on neural networks*. Budapest, Hungary.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, *70*, 489–501.

Jaeger, H. (2001). The echo state approach to analyzing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, *148*, 34.

Jaeger, H. (2002). Tutorial on training recurrent neural networks, covering BPPT, RTRL. *GMD-Forschungszentrum Informationstechnik: EKF and the echo state network approach*.

Jalab, H. A., & Ibrahim, R. W. (2011). New activation functions for complex-valued neural network. *International Journal of the Physical Sciences*, *6*(7), 1766–1772.

Javed, K. (2014). A robust & reliable data-driven prognostics approach based on extreme learning machine and fuzzy clustering. *Ph.D. thesis, Université de Franche-Comté*.

Javed, K., Gouriveau, R., & Zerhouni, N. (2014). SW-ELM: A summation wavelet extreme learning machine algorithm with a priori parameter initialization. *Neurocomputing*, *123*, 299–307.

Javed, K., Gouriveau, R., Zerhouni, N., & Nectoux, P. (2015). Enabling health monitoring approach based on vibration data for accurate prognostics. *IEEE Transactions on Industrial Electronics*, *62*(1), 647–656.

Javed, K., Gouriveau, R., Zerhouni, N., Zemouri, R., & Li, X. (2012). Robust, reliable and applicable tool wear monitoring and prognostic: approach based on an improved-extreme learning machine. In *IEEE conference on prognostics and health management*. Denver, CO, USA.

Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, *22*(9), 1341–1356.

Li, X., Lim B. S., Zhou J. H., Huang, S., Phua S. J., & Shaw, K. C. (2009). Fuzzy neural network modeling for tool wear estimation in drymilling operation. In *Annual conference of the prognostics and health management society*. San Diego, CA, USA.

Liao, L. (2010). An adaptive modeling for robust prognostics on a reconfigurable platform. *Ph.D. thesis, University of Cincinnati*.

Massol, O., Li, X., Gouriveau, R., Zhou, J. H., & Gan, O. P. (2010). An exTS based neuro-fuzzy algorithm for prognostics and tool-condition monitoring. In *11th international conference on control automation robotics & vision ICARCV'10*. Singapore, pp. 1329–1334.

Mathworks: Curve fitting toolbox. (2010). http://mathworks.com/help/toolbox/curvefit/smooth.html

Nguyen, D., & Widrow, B. (1990). Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In *International joint conference on neural networks IJCNN*. San Diego, CA, USA.

Oussar, Y., & Dreyfus, G. (2000). Initialization by selection for wavelet network training. *Neurocomputing*, *34*(1–4), 131–143.

Pal, S., Heyns, P. S., Freyer, B. H., Theron, N. J., & Pal, S. K. (2011). Tool wear monitoring and selection of optimum cutting conditions with progressive tool wear effect and input uncertainties. *Journal of Intelligent Manufacturing*, *22*(4), 491–504.

Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: A review. *International Journal Advance Manufacturing Technology*, *50*, 297–313.

Petkovi, D., Danesh, A. S., Dadkhah, M., Misaghian, N., Shamshirband, S., & Pavlovi, N. D. (2016). Adaptive control algorithm of flexible robotic gripper by extreme learning machine. *Robotics and Computer-Integrated Manufacturing*, *37*, 170–178. doi:10.1016/j.rcim.2015.09.006.

Rajesh, R., & Prakash, J. S. (2011). Extreme learning machines—A review and state-of-the-art. *International Journal of Wisdom Based Computing*, *1*, 35–49.

Rao, C. R., & Mitra, S. K. (1971). *Generalized inverse of matrices and its applications*. New York: John Wiley and Sons.

Ren, L., Lv, W., & Jiang, S. (2015). Machine prognostics based on sparse representation model. *Journal of Intelligent Manufacturing* pp. 1–9. doi:10.1007/s10845-015-1107-8.

Rizal, M., Ghani, J. A., Nuawi, M. Z., & Haron, C. H. C. (2013). Online tool wear prediction system in the turning process using an adaptive neuro-fuzzy inference system. *Applied Soft Computing*, *13*(4), 1960–1968.

Saikumar, S., & Shunmugam, M. (2012). Development of a feed rate adaption control system for high-speed rough and finish end-milling of hardened en24 steel. *International Journal Advance Manufacturing Technology*, *59*(9–12), 869–884.

Shamshirband, S., Mohammadi, K., Chen, H. L., Samy, G. N., Petkovi, D., & Ma, C. (2015). Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: A case study for Iran. *Journal of Atmospheric and Solar-Terrestrial Physics*, *134*, 109–117. doi:10.1016/j.jastp.2015.09.014.

Sikorska, J. Z., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Journal of Mechanical Systems and Signal Processing*, *26*(5), 1803–1836.

Singh, R., & Balasundaram, S. (2007). Application of extreme learning machine method for time series analysis. *International Journal of Intelligent Technology*, *2*(4), 256–262.

Wang, G., & Cui, Y. (2013). On line tool wear monitoring based on auto associative neural network. *Journal of Intelligent Manufacturing*, *24*(6), 1085–1094.

Wu, Y., Hong, G., & Wong, W. (2015). Prognosis of the probability of failure in tool condition monitoring application—A time series based approach. *The International Journal of Advanced Manufacturing Technology*, *76*(1–4), 513–521.

Zemouri, R., Gouriveau, R., & Zerhouni, N. (2010). Improving the prediction accuracy of recurrent neural network by a pid controller. *International Journal of Systems Applications, Engineering & Development*, *4*(2), 19–34.

Zhai, L. Y., Er, M. J., Li, X., Gan, O. P., Phua, S. J., Huang, S., Zhou, J. H., Linn, S., & Torabi, A. J. (2010). Intelligent monitoring of surfaceintegrity and cutter degradation in high-speed milling processes. In *Annual conference of the prognostics and health management society*. Portland, Oregon, USA.

Zhao, G., Shen, Z., Miao, C., & Man, Z. (2009). On improving the conditioning of extreme learning machine: a linear case. In *7th International conference on information, communications and signal processing. ICICS 09*. Piscataway, NJ, USA.

Zhou, J., Li, X., Gan, O. P., Han, S., & Ng, W. K. (2006). Genetic algorithms for feature subset selection in equipment fault diagnostics. *Engineering Asset Management*, *10*, 1104–1113.

Zhou, J. H., Pang, C. K., Lewis, F., & Zhong, Z. W. (2009). Intelligent diagnosis and prognosis of tool wear using dominant feature identification. *IEEE Transactions on Industrial Informatics*, *5*(4), 454–464.

Zhou, J. H., Pang, C. K., Zhong, Z. W., & Lewis, F. L. (2011). Tool wear monitoring using acoustic emissions by dominant-feature identification. *IEEE Transactions on Instrumentation and Measurement*, *60*(2), 547–559.