TOWARDS AN ACTIVE LEARNING APPROACH TO TOOL CONDITION MONITORING WITH BAYESIAN DEEP LEARNING

Giovanna Martínez-Arellano and Svetan Ratchev
Institute for Advanced Manufacturing
University of Nottingham
UK
E-mail: giovanna.martinezarellano@nottingham.ac.uk
svetan.ratchev@nottingham.ac.uk

KEYWORDS

Smart Manufacturing, Tool Condition Monitoring, Bayesian Convolutional Neural Networks, Time Series Imaging, Active Learning.

ABSTRACT

With the current advances in the Internet of Things (IoT), smart sensors and Artificial Intelligence (AI), a new generation of condition monitoring solutions for smart manufacturing is starting to emerge. Computer Numerical Control (CNC) machines can now be sensorised and the vast amount of data generated can be processed using Machine Learning (ML) techniques. These can provide insights about the condition of the machine or tool in real-time, which can then be used by decision makers. This is fundamental in order to reach a new level of manufacturing capabilities in the context of Industry 4.0 (Lasi et al, 2014). Most current monitoring solutions rely on the off-line generation of models before they can be used online. This is not ideal when the data holds complex evolving features. There is a lack of approaches that are capable of determining what to learn and when to learn. This paper presents preliminary results on a new deep learning approach based on Bayesian Convolutional Neural Networks (BCNN) for online tool condition classification. Based on the uncertainty of the model, the proposed approach can determine using an entropy acquisition function if the incoming data cannot be classified, and therefore needs to be labelled and used for re-training. This constitutes the first step towards an online active learning tool condition monitoring approach. We demonstrate using a machine tool data set that the active learning approach can achieve similar accuracy of a deterministic Convolutional Neutral Network (CNN) with a smaller training data set.

INTRODUCTION

Industry 4.0 is a new paradigm which proposes the integration of Information and Communication Technology (ICT) into a "decentralized" production (Lasi et al, 2014). With sensorised manufacturing machines connected to a wireless network, controlled by advanced computational intelligence techniques, and a range of smart solutions for the monitoring, adaptation, simulation and optimisation of factories, Industry 4.0 is looking to attain new levels of manufacturing capability and adaptability.

Intelligent machine condition monitoring is an important task in smart manufacturing. For machining operations with cutting tools such as drilling, milling or turning, the early detection of tool degradation is extremely important because worn tools have a negative effect on the surface quality of the workpiece and may even damage the machining system (Bonifacio and Diniz, 1994; Ambhore et al, 2015). Having a Tool Condition Monitoring (TCM) system to detect this degradation on time, can avoid either removing the tool too soon or too late, maintaining the quality of the workpiece.

Real-time tool wear measurement is difficult to put in practice, as the tool is in continuous contact with the workpiece. The tool or workpiece would be analysed at the end of the machining cycle, through optical measurement, surface finishing measurement, chip size measurement, etc. If these procedures were to be done during the machining cycle, it would require frequent production stops to be able to detect degradation on time. Many of these stops would be unnecessary when the tool is in good condition, incurring in additional costs. To tackle this problem, *prognostic* approaches that work on indirect measurements have been proposed. These are normally based on sensor signals such as forces (Gosh et al., 2007), vibrations (Gouarir et al. 2018), acoustic emissions (Kima et al., 1999) and power consumption (Ambhore et al, 2015).

Prognostic approaches can be classified into two types; model-based and data-driven. The major limitation of model-based solutions is that they rely on the prior understanding of the underlying dynamics of the system to be modeled (Hou et al., 2014). For this reason, there has been an increasing interest on the use of data-driven approaches, where models are discovered using ML techniques. Support Vector Machines (SVMs), for example, have been successfully applied for tool condition monitoring in (Sun et al, 2004). The authors use Automatic Relevance Determination (ARD) on

acoustic emission data, to select nine features as inputs for classification. In (Salgado and Alonso, 2007), a least squares version of an SVM (LS-SVM) is used to estimate tool condition. The approach relies on the extraction of features of current and sound signals using techniques such as Singular Spectrum Analysis (SSA). Artificial Neural Networks (ANNs) have also been extensively applied for tool wear prediction. These commonly use a combination of cutting parameters such as cutting speed, feed rate, axial cutting length and statistical features of forces, vibrations and acoustic emission (Chungchoo and Saini, 2002; Özel et al, 2005; Sanjay et al, 2005; Palanisamy et al, 2008). In applications such as drilling and milling, it has been shown how ANNs can outperform regression models. In (Wu et al, 2016), a tool wear prediction method based on Random Forests (RFs) is proposed. Although this approach has outperformed ANN and SVM based methods, it relies on the manual selection of features in order to build the internal classification structures. A thorough review on the use of sensor signals for indirect tool-wear monitoring can be found in (Dimla, 2000; Abellan-Nebot, 2010).

Although classical ML techniques have been successful in TCM applications to some extent, there are several aspects that still need to be addressed. The large amount of data in smart manufacturing imposes challenges such as the proliferation of multi-variate data, high dimensionality of feature space and multicollinearity among data measurements (Wuest et al., 2016). In addition, most data-driven methods derive models from historical data, in a batch learning approach, which is not compatible with real-time processing. These methods assume that the distribution of data will not change through time, and require a complete dataset covering all possible situations in the monitoring process and a complete retraining from scratch when new patterns are observed. Data-driven methods lack in general of a way to determine when to learn and what to learn. In the Industry 4.0 context, industries are integrating Internet of Things (IoT) technologies and techniques, which demand more advanced solutions that can cope with online dynamic characteristics of the machining process (Pratama, 2017).

This paper explores the application of BCNNs for the online classification of tool condition using sensor data streams. While the deep learning aspect of the approach allows automatic feature selection, the introduction of a probability distribution on the weights using Bayes offers better robustness to over-fitting on small data sets compared to traditional deep learning approaches. In addition, Bayesian deep learning allows the representation of the model uncertainty. This uncertainty, coupled with an acquisition function, can allow the implementation of an online active learning approach to tool condition monitoring.

The rest of this paper is organised as follows: further details on related work on deep learning and active learning is provided. The methodology is then presented, providing the details of the classification method used, and how the data was generated and pre-processed. Then preliminary experiments and results of the proposed approach are presented followed by a discussion and future work.

RELATED WORK

Deep learning has offered better solutions than classical ML techniques when dealing with high dimensional evolving features. Its success has led to an emerging study of deep learning methods for condition monitoring. CNNs (LeCun, 2015), for example, have been used for the detection of faulty bearings (Li et al., 2017) by feeding raw vibration data, reducing the computational complexity of feature extraction. In construction, CNNs have also been applied for the real-time detection of structural damage in joints. In (Terrazas et al., 2018), a time series image encoding method together with a CNN is used to perform the classification of a machine tool wear in real-time. The approach achieves good accuracy, however, the CNN is always expected to provide a classification, with no information regarding the confidence level of such classification. Recurrent Neural Networks (RNN) have been successful for the long-term prognosis of rolling bearing health status (Malhi et al., 2011) and for the prediction of tool wear, gear fault and bearing fault diagnosis (Zhao et al., 2018). Another architecture that has been used is the bi-directional long short-term memory (LSTM) CNN. In (Zhao et al, 2017), this approach extracts local features from raw sensor data for the prediction of tool wear during milling. The technique achieves good accuracy when compared to other methods like RNN, although it performs a substantial reduction of the data, making unclear to what extent it affects the temporal correlations in the data. Although current deep learning solutions claim to have achieved high accuracy in different condition classification applications, the vast majority of these still carry out some sort of feature extraction, such as wavelets and Fourier Transforms (FFT) as a data pre-processing step, that could be argued is against the philosophy of deep learning (Khan and Yairi, 2018).

In addition to the use of manual feature extraction, deep learning techniques that have been applied to TCM lack an active learning feature. Models need to be able to adapt as the characteristics of the sensor data streams change. A big challenge in many deep learning applications is obtaining large quantities of labelled data. A framework where a system could learn from small amounts of data and choose by itself what data needs labelling would make deep learning more flexible and more widely applicable. The idea behind active learning is to develop a model on an initial small training set and then use an *acquisition function*, often based on the model's uncertainty, to decide which data points to ask
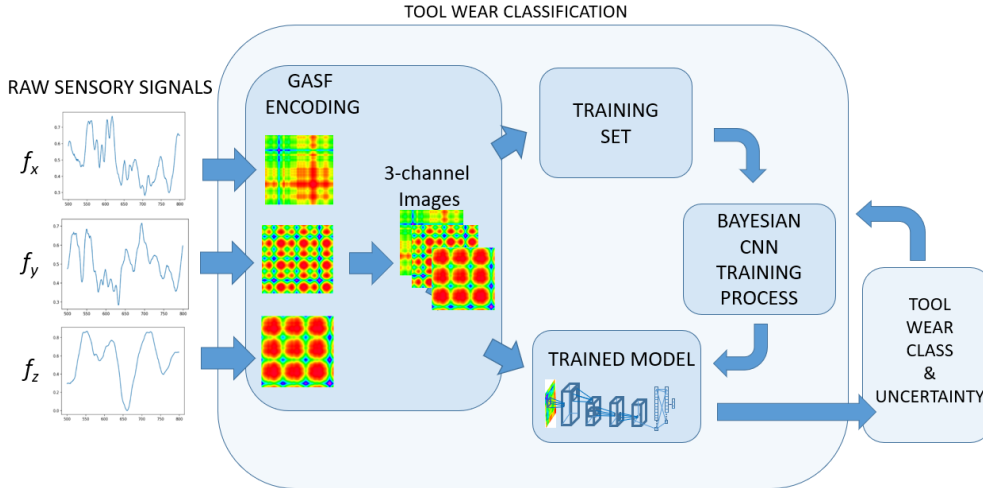
Figure 1. Proposed methodology combining time series encoding and BCNN for active learning

an external *oracle* for a label. The acquisition function selects one or several data points from a pool of unlabeled data which is not currently on the training set. These selected data points are labelled by the oracle (usually the human expert) and then added to the training set for a new model to be trained. The advantage of these types of approaches, is that, in general, they tend to need less data to train than a conventional ML technique (Ghahramani et al., 2017).

There have been some efforts on the application of active learning approaches for condition monitoring. Moshour et al. apply One Class Classifiers based on Self Organising Maps (OCSOM) and Support Vector Machines (OCSVM) to detect changes in vibration patterns for the diagnosis of broken bearings (Moshou et al, 2014). The proposed approach is able to progressively learn different stages of faults by generating new training sets as the one-class classifiers detect outliers. The construction of such one-class classifiers, however, depends on the manual extraction of features on the vibration data. Nguyen et al. present an active learning approach for the detection of the partial discharge of electrical assets in power grids (Nguyen et al., 2015). This approach relies on the feature extraction of power signals prior to training. To decide what new data points to select for training, a calculation of the posterior probability from the trained model is used. It is concluded that using an active learning approach, good accuracy can be achieved without having to train with a complete labelled data set. Pratama et al. develop an online tool condition monitoring based on an ensemble of fuzzy base classifiers and cutting forces (Pratama et al., 2018). The approach uses FFT to pre-process the dynamic occurring forces measured, and together with the mean value form the input data samples. Despite extensive literature review, there is no published implementation of a deep learning approach with automatic feature extraction from sensor data for condition monitoring that can provide uncertainty estimation of the tool state. The advantage of providing an uncertainty measure is that it can allow the

implementation of an active learning method that can deal with changes in the distribution of sensor data streams.

METHODOLOGY

This section presents the two main steps of the methodology: the imaging of sensor signals using Gramian Angular Summation Fields (GASF) (Wang and Oates, 2015) and the classification and uncertainty characterisation of tool condition using a BCNN and an entropy-based acquisition function. This two-step methodology is presented in Figure 1.

Time Series Imaging

With the success of deep learning approaches for image classification, there has been a recent interest in reformulating time series data as images in order to improve their classification. In this work, we use the GASF method proposed by Wang and Oates to encode sensor data as images that will be later used for training the model. This image encoding method consists of two steps. First, the time series is represented in a polar coordinate system instead of the typical Cartesian coordinates. Thus, given a time series $X = x_1, x_2, …, x_n$ of n real-valued observations, X is rescaled so that all values fall in the interval [-1,1]. The time series can then be represented in polar coordinates by encoding the value as the angular cosine and the time stamp as the radius. In a second step, the angular perspective is exploited by considering the trigonometric difference between each point to identify the temporal correlation within different time intervals. Given a time series of size n, the resulting GASF image will be a matrix of $n \times n$. Figure 2 shows an example of the steps explained.

To reduce the size of the image, Piecewise Aggregation Approximation (PAA) is applied to smooth the time series while keeping the trends (Keogh and Pazzani, 2000). As explained in the Experiments and Results

Polar coordinates representation of the raw signal

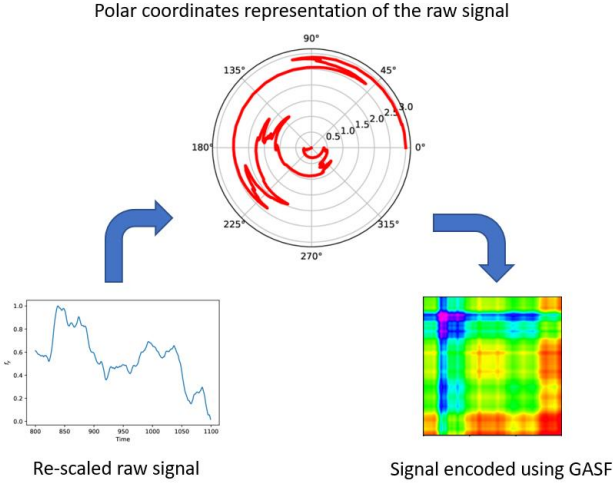Re-scaled raw signal

Signal encoded using GASF

Figure 2. Imaging of time series using GASF

section, the amount of time series data that is acquired from the sensors is large, so PAA is fundamental to keep the images at a reasonable size without losing time coherence.

Bayesian Convolutional Neural Networks

CNNs have been very successful for image processing by extracting complex features automatically (Krizhevsky et al, 2012). However, they heavily rely on the availability of large amounts of data to avoid overfitting. Bayesian Neural Networks (BNNs), on the other hand, are robust to over-fitting when learning from small training sets. In addition, BNNs offer a probabilistic interpretation of deep learning models by inferring distributions over the models' weights. This is done by adding dropout layers after all convolutional layers as well as inner-product ones (Gal and Ghahramani, 2016).

Bayesian CNNs are CNNs with prior probability distributions placed over a set of model parameters $w = \{W_1,…,W_L\}$:

$$w \sim p(w) \qquad (1)$$

with for example a standard Gaussian prior $p(w)$. For the case of classification, the likelihood model is defined as:

$$p(y=c|x,w) = \text{softmax}(f^w(x)) \qquad (2)$$

where $f^w(x)$ is the model output with parameters (w). To perform approximate inference in the Bayesian CNN model, stochastic regularisation techniques such as dropout are applied. Inference is done by training a model with dropout before every weight layer, and by performing dropout at test time as well to sample from the approximate posterior. Dropout can be interpreted as a variational Bayesian approximation, where the approximating distribution is a mixture of two Gaussians

with small variances and the mean of one of the Gaussians is fixed to zero.

The uncertainty in weights induces prediction uncertainty by marginalising over the approximate posterior using Monte Carlo integration, which can be written as:

$$p(y=c|x,D_{train}) \approx (1/T)(\Sigma_{t=1} p(y=c|x,\hat{w}_t)) \qquad (3)$$

This is equivalent to performing T stochastic forward passes through the network and averaging the results. The uncertainty information provided by the model can then be used with an acquisition function appropriate for image classification to determine what new incoming images should be labeled. There are several acquisition functions such as Variation Ratios, Mean Standard Deviation, Information Gain (BALD), Entropy, among others (Gal et al, 2017). Here the Entropy maximisation was used (Eq. 4).

$$H[y|x,D_{train}] = -\Sigma_c p(y= c|x,D_{train})\log p(y= c|x,D_{train}) \qquad (4)$$

EXPERIMENT AND RESULTS

The proposed approach is studied for the classification of tool condition in an active learning scenario and compared to a deterministic CNN. In order to test the approach in the described scenarios, it was important to capture sensor signals accurately and to harvest and characterise the right data beforehand. To achieve this, a machining experiment was designed.

Measurement of Forces During Dry Milling

Cutting force is an important feature in milling, closely related to tool design geometry. For this reason, it was decided to monitor the cutting force as a preliminary experiment. This was done using a dynamometer (Kistler 9255B) placed on the table of a Hermle C20U CNC milling machine. Forces (on the three axes) produced while face milling a mild steel workpiece were amplified and captured with a National Instruments data acquisition system. The workpiece (180×125×25mm) surface was machined line-by-line along the x axis with a 6mm two-flute cutter. After finishing one pass along the x axis, the tool was retracted to start a new pass. This was done until half of the surface (half layer) was removed. Then the tool was removed from the tool holder and taken to a digital microscope, where high resolution images of the tool flutes were taken. These were then manually processed to measure the flank wear (VB_b). The tool was returned to the tool holder to machine the remaining half layer of the workpiece and the flank wear measurement was taken again. This process was repeated until 9 layers of material were removed, point at which the tool was completely worn out. The cutting parameters used in the machining experiment were fixed to $S = 4775$ RPM, $f = 287$mm/min, $a_e = 2.7$mm and $a_p = 0.3$mm. The experimental setup is shown in Figure 3.
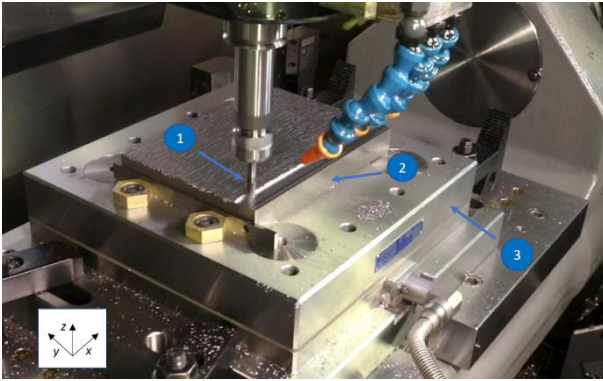
Figure 3. Machining experimental setup. Interior of the Hermle CNC machine with (1) end milling tool, (2) mild steel workpiece and (3) Kistler 9255B dynamometer

Data Pre-Processing

Once the sensory data was acquired, it was cleaned to remove those force measurements that were taken while the tool was not in contact with the workpiece (end and beginning of each tool pass). From all the data collected while machining, it was decided to only use the data that corresponded to the middle part of the workpiece in order to disregard any potential noise that could be affecting the force signals when entering and exiting the workpiece. From each layer and middle section, 4800 samples of 2000 measurements each were taken. The cutting forces of each sample, F_x, F_y and F_z, were encoded as separate images using GASF, and then reduced with PAA from an image of 2000×2000 pixels to an image of 256×256 pixels. The three images corresponding to one sample were then combined as one 3-channel image.

To label the images, four classes were defined, according to how the tool flank wear progressed through the machining experiment. Classes are defined as Break-in, Steady, Severe and Failure as shown in Figure 4. Once the classes were defined, the images were labeled according to the removed layer they belong to.
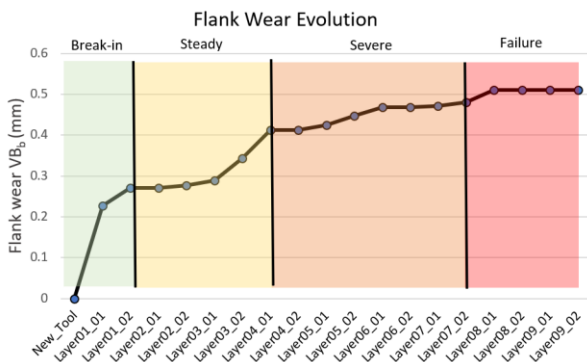

Figure 4. Evolution of the tool flank wear across the Break-in, Steady, Severe and Failure classes

Model Training and Testing

The Bayesian CNN and the active learning algorithm were implemented in Python using Keras libraries (following the implementation from Gal et al. (Gal et al., 2016)). The structure used was as follows: convolution-relu, maxpooling, convolution-relu, maxpooling, dropout, dense (384 units), relu, dropout, dense (192 units), relu, dropout, dense softmax, with 64 convolution kernels (16x16) in the first convolution, 64 convolution kernels (8x8) in the second convolution and dropout probabilities 0.25 and 0.5.

As it can be seen in Figure 4, the sensor data is unbalanced, with less samples of the Break-in class compared to other classes. However, since the data set is large (4800 images per layer), a subset of these was taken as an initial training and test set, making sure classes were represented equally (balanced). The balanced dataset set comprised of 1428 images, of which 70% was used for training (250 images per class) and 30% for testing. The active learning approach was implemented as follows: the model described above was first trained for 100 epochs with a small data set of 100 images, 25 of each class, using a validation set of 200 images. The rest of the training samples (600 images) were used as the pool from which the new images will be selected for labelling and training. Once trained, the approach determines which images from the pool maximise the predictive entropy and selects the top 100 to be labelled and incorporated in the existing training set. The model is re-trained (100 epochs) and this is repeated until good accuracy is achieved. A result previously reported in (Terrazas et al, 2018) was used as reference accuracy value. The reported results were obtained on a deterministic CNN with the same architecture and same training/test set.

Results

The CNN experiment reported in (Terrazas et al., 2018) achieved an accuracy of 78% using the complete training set (1000 images) and executed for 100 epochs. This was used as a reference value in the active learning experiment. Two experiments of the active learning approach were executed, one using the entropy acquisition function, and another one using a random acquisition function, which samples from a uniform distribution. This was done to make sure the entropy acquisition function could provide an improvement over a purely random approach. Figures 5 and 6 show the training and test set accuracy results obtained at each acquisition iteration, starting from acquisition 0 (initial random training set) up to acquisition 3 for both experiments.
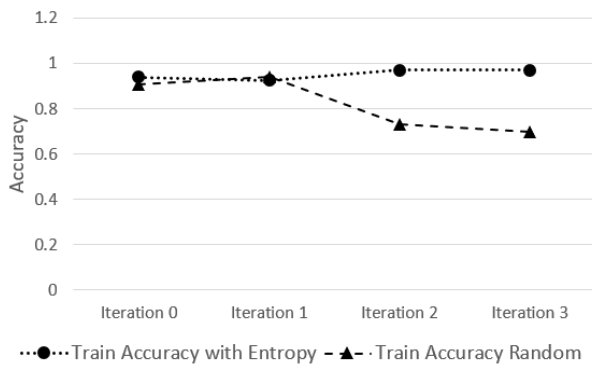
Figure 5. Accuracy on the training set at different acquisition iterations using entropy maximization and a random approach

As it can be observed, the experiment using the entropy acquisition function can achieve a similar accuracy as the deterministic CNN with far less images. Also, it can be seen that the entropy compared to the random acquisition, allows the approach to make use of the uncertainty information to determine which new images need to be labelled and used for re-training, improving, as a result, the test accuracy over time. It can also be observed in Figure 6 that at iteration step 3 the accuracy on the test set decreases for the first experiment. A possible explanation for this is that the algorithm selected in this iteration noisy images that increased the uncertainty of the model, decreasing the overall accuracy. An inspection of the images acquired at each iteration could provide more information to explain the decrease. The active learning approach nevertheless provides an advantage over the use of a complete, labelled data set.
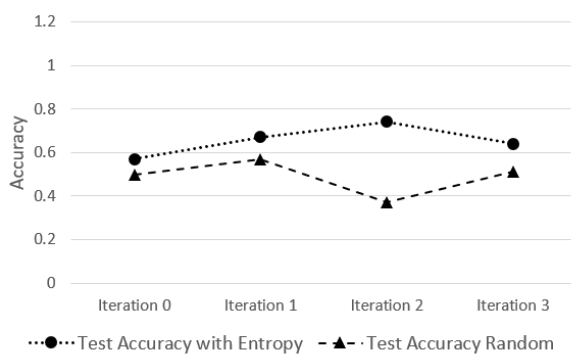

Figure 6. Accuracy on the test set at different acquisition iterations using entropy maximization and a random approach

CONCLUSIONS

This paper presents an active learning approach to tool condition monitoring using Bayesian Deep Learning. Preliminary results show the ability of the approach to select those images that need to be labelled and used for re-training when using an acquisition function compared to a random selection. The active learning approach can

achieve comparable results with less images than the deterministic CNN. Future work will include the experimentation with different architecture and training hyper parameters to improve accuracy as well as experimentation with other acquisition functions such as information gain maximisation (BALD), variation ratios, among others. Analysis of the images acquired at each iteration step will be performed to have a better understanding of how the acquisition function is working.

REFERENCES

Abellan-Nebot, J.V. and F. Romero Subirón. 2010. "A review of machining monitoring systems based on artificial intelligence process models". *International Journal of Advanced Manuf. Tech*. 47, No 1-4, pp. 237-257.

Ambhore, N.; D. Kamble; S. Chinchanikar; V. Wayal. 2015."Tool Condition Monitoring System: A Review". *Materials Today: Proceedings*. 2, No.4-5, pp. 3419-3428.

Bonifacio, M. and A. Diniz. 1994. "Correlating tool wear, tool life, surface roughness and tool vibration in finish turning with coated carbide tools". *Wear*. 174, No. 1, 137-144.

Chungchoo, C. and D. Saini. 2002. "On-line Tool Wear estimation in CNC turning operations using Fuzzy Neural Network Model". *International Journal of Machine Tools and Manufacture*. 4, No. 1, pp. 29-40.

Dimla, D. E. 2000. "Sensor signals for tool-wear monitoring in metal cutting operations-a review of methods". *International Journal of Machine Tools and Manufacture*. 40, No. 8, pp. 1073-1098.

Gal, Y. and Z. Ghahramani. 2016. "Bayesian convolutional neural networks with Bernoulli approximate variational inference". *ICLR workshop track*, arXiv:1506.0215.

Ghahramani, Z.; Y. Gal; R. Islam. 2017. "Deep Bayesian Active Learning with Image Data". *Proceedings of Machine Learning Research*, 70, pp, 1183-1192.

Ghosh, N; Y.B. Ravi; A. Patra; S. Mukhopadhyay; S. Paul; A.R. Mohanty; A.B. Chattopadhyay. 2007. "Estimation of tool wear during CNC milling using neural network-based sensor fusion". *Mechanical Systems and Signal Processing*. 21, No. 1, pp. 466-479.

Gouarir, A; G. Martínez-Arellano; G. Terrazas; P. Benardos; S. Ratchev. 2018. "In-process tool wear prediction system based on machine learning techniques and force analysis". *Procedia CIRP*. 77, pp. 501-504.

Hou, Y.; D. Zhang; M. Luo; B. Wu. 2014. "Tool flank wear recognition based on the variation of milling force vector in end milling". 2014. *IEEE/ASME Int. Conference on Advanced Intelligent Mechatronic,* pp. 1622-1627.

Khan, S.; T. Yairi. 2018. "A review on the application of deep learning in system health management". *Mech. Syst. Signal Process*. 107, pp.241-265.

Kima, J.S.; M.C. Kangb; B.J. Ryuc; Y.K. Jic. 1999. "Development of an on-line tool-life monitoring system using acoustic emission signals in gear shaping". *Int J Mach Tools Manuf*. 39, pp. 1762-1777.

Krizhevsky, A.; I. Sutskever; G. E. Hinton. 2012. "ImageNet classification with deep convolutional neural networks". *In Proceedings of the 25th International conference on Neural Information Processing Systems*. 1, pp. 1097-1105.

Lasi, H.; P. Fettke; H-G. Kemper; M. Hoffmann. 2014. "Industry 4.0". *Business and Information Systems Engineering*. 6, No. 4, pp. 239-242.

LeCun, Y.; Y. Bengio; G. Hinton; 2015. "Deep learning". *Nature*. 521, pp. 436-444.

Li, S.; G. Liu; X. Tang; J. Lu; J. Ju. 2017. "An Ensemble Deep Convolutional Neural Network Model with Improved D-S Evidence Fusion for Bearing Fault Diagnosis". *Sensors*. 17, pp. 1729.

Li, X.; B.S. Lim; J.H. Zhou; S. Huang; S.J. Phua; K.C. Shaw; M.J. Er; 2009. "Fuzzy Neural Network Modelling for Tool Wear Estimation in Dry Milling Operation". *In Proceedings of the annual conference of the Prognostics and Health Management Society, 2009.*

Malhi, A; R. Yan; R.X. Gao. 2011. "Prognosis of defect propagation based on recurrent neural networks". *IEEE Transactions on Instrumentation and Measurement*. 60, No. 3, pp. 703-711

Moshou, D; X. Eirini Pantazi; D. Kateris; E. Tsikos; N. Sawalhi. 2014. "A machine learning approach for the condition monitoring of rotating machinery". *DOI:10.1007/s12206-013-1102-y.*

Özel, T; Y. Karpat. 2005. "Predictive modeling of surface roughness and tool wear in hard turning using regression and neural networks". *International Journal of Machine Tools and Manufacture*. 45, No. 4, pp. 467-479.

Palanisamy, P.; I. Rajendran; S. Shanmugasundaram; 2008. "Prediction of Tool Wear using Regression and ANN models in End-milling Operation". *International Journal in Advanced Manufacturing Technology*. 37, No.1-2, pp. 29-41.

Pratama, M.; E. Dimla; E. Lughofer; W. Pedrycz. 2018. "Online tool condition monitoring based on parsimonious ensemble+". *IEEE Transactions on Cybernetics*, pp. 1-14, DOI: 10.1109/TCYB.2018.2871120.

Salgado, D.R. and F.J. Alonso; 2007. "An Approach based on Current and Sound Signals for In-process Tool Wear Monitoring". *International Journal of Machine Tools and Manufacture*. 47, No. 14, pp. 2140-2152.

Sanjay, C.; M. Neema; C. Chin. 2005. "Modeling of Tool Wear in drilling by Statistical Analysis and Artificial Neural Network". 2005. *Journal of Materials Processing Technology*. 170, No. 3, pp. 494-500.

Sun, J.; M. Rahman; Y.S. Wong; G.S. Hong; 2004. "Multiclassification of Tool Wear with Support Vector Machine by manufacturing loss consideration". *International Journal of Machine Tools and Manufacture*. 44. No. 11, pp. 1179-1187.

Terrazas, G.; G. Martínez-Arellano; P. Benardos; S. Ratchev. 2018. "Online tool wear classification during dry machining using real time cutting force measurements and a CNN approach". *Journal of Manufacturing and Material Processing*. 2. No. 4, pp. 72.

Wang, Z; T. Oates. 2015. "Imaging time-series to improve classification and imputation". *Proceedings of the 24th International Conference on Artificial Intelligence IJCAI'15,* pp. 3939-3945.

Wu, D.; C. Jennings; J. Terpenny; S. Kumara. 2016. "Cloud-based machine learning for predictive analytics: Tool wear prediction in milling". *IEEE International Conference on Big Data (Big Data)*, pp. 2062-2069, DOI: 10.1109/BigData.2016.7840831.

Wuest, T.; D. Weimer; C. Irgens; K.D. Thoben. 2016. "Machine learning in manufacturing: advantages, challenges, and applications". *Production and Manufacturing Research*. 4, No. 1, pp. 23-45.

Zhao, R; D. Wang; R. Yan; K. Mao; F. Shen; J. Wang. 2018. "Machine health monitoring using local feature-based gated recurrent unit networks". *IEEE Transactions on Industrial Electronics*. 65, No. 2, pp. 1539-1548.

Zhao, R; R. Yan; J. Wang; K. Mao. 2017. "Learning to monitor machine health with convolutional bi-directional LSTM networks". *Sensors*. 17, No.2, pp. 273.

AUTHOR BIOGRAPHIES

GIOVANNA M. ARELLANO has a PhD in Computer Science from Nottingham Trent University and is currently a Research Fellow in Data Science at the Institute for Advanced Manufacturing at the University of Nottingham. Email address: `giovanna.martinez-arellano@nottingham.ac.uk`.

SVETAN RATCHEV is the Cripps Chair of Production Engineering and Director of the Institute for Advanced Manufacturing at the University of Nottingham. He has established a strong international reputation for his work in informatics enabled manufacturing and automated assembly, as well as strategic and policy advice engagement. Email address: `svetan.ratchev@nottingham.ac.uk`