A novel methodology for the characterization of cutting conditions in turning processes using Machine Learning models and Acoustic Emission Signals

Telmo Fernández de Barrena[1], Juan Luis Ferrando[1], Ander García[1], Pedro Jose Arrazola[2], Jose Manuel Abete[2], Diego Herrero[3]

[1] Vicomtech Foundation, Basque Research and Technology Alliance (BRTA), Mikeletegi 57, 20009 Donostia-San Sebastian (Spain)
[2] Faculty of Engineering, Mondragon University, Mondragon 20500, Spain
[3]Sidenor I+D, Barrio Ugarte s/n, Basauri 48970, Spain

Abstract. In the last few years, the industry requires to know in real-time the condition of their assets. Acoustic Emission (AE) technique has been widely used to understand the real-time condition of manufacturing processes such as the degree of tool wear or tool breakage. Traditionally, to fulfil that goal, the information extracted from the signal sensors of the machines has been processed with mathematical models. This methodology is changing, and instead of developing complex physical models (where an in-depth knowledge of the system being modelled is required), the current trend is to use Machine Learning (ML) models which are based on previous data. Signal pre-processing and feature extraction is a complex task that usually generates a high amount of predicting variables. Therefore, this paper proposes a methodology to identify the best pre-processing tools, AE features and ML models to characterize cutting condition processes. This methodology is validated identifying cutting conditions in a turning process based on AE signals. To classify the cutting condition with the highest accuracy, several techniques are applied, (including wavelet transform for multiresolution analysis, Recursive Feature Elimination (RFE) technique, different classifiers (Decision Tree (DT), Random Forests (RF), Support Vector Machine (SVM), Gaussian Process (GP), K-Nearest Neighbor (KNN) and Multilayer Perceptron (MLP) classifiers) and different signal segmentation lengths. These techniques were evaluated using the data captured in a turning process when cutting a 19NiMoCr6 steel under pre-established cutting conditions. The best accuracy of predicting the cutting conditions based on AE signals was 99.7%, and it was achieved combining the wavelet packet transform (WPT) with RFE, with a segmentation time of 0.05 s and RF as classifier.

Keywords: Machine learning, acoustic emission, cutting characterization, wavelet transform, recursive feature elimination.

1 Introduction

Nowadays, the manufacturing industry, with the help of emerging technologies, is working to achieve different goals, such as the zero defects manufacturing, reduction of costs, reduction of emissions etc. One of the keys is to be able to predict the tool wear condition of the machines, characterize the cutting conditions for detecting abnormal cutting conditions, etc. 20% of the time that a machine is inactive is due to tool failure, leading to a reduction productivity and economic losses [1]. Therefore, there are many researchers working on the study of different signals for the identification of cutting conditions and for the development of different tool wear monitoring techniques. In the industrial machining environment, usually force, AE and electric current signals are used as source of data [2]. Diei and Dornfeld [3] studied both the cutting forces and AE signals are closely correlated with the tool wear condition of a milling machine.

On the one hand, in the literature, several researchers have studied the relation of AE signals with different parameters of cutting conditions. The characterization of the cutting conditions can help to detect immediately abnormal machining conditions, thus allowing to react to these situations rapidly and reducing cost: reduce defective parts, avoid excessive tool wear, reduce excessive energy consumption, etc. Teti and Dornfeld [4] studied the correlation between AE signal energy and cutting conditions of orthogonal cutting process. They observed that the root mean square (RMS) value of the AE signal is sensitive to the cutting speed, whereas the influence of depth of cut and feed rate on the RMS voltage of AE signals was observed to be negligible. These results are also consistent with results reported by Rangwala [5] and Lan [6]. Thus, the characterization of the cutting process seems to be ideal for different cutting speed values, but harder for the conditions in which the cutting speed is constant and the feed rate and depth of cut change.

On the other hand, several research works have used ML models for monitoring machining process in real-time. Wu, D. [7] predicted the tool wear in milling operations using three popular ML algorithms: Artificial Neural Networks (ANNs), Support Vector Regressor (SVR) and RFs. First, some statistical features were extracted from three signals: cutting forces, vibrations, and AEs. The results showed that RFs generate more accurate predictions than the other algorithms.

As the whole frequency range of the AE signals could not be interesting for the study, the WPT can be applied. This is a well-known technique when useful information from the signal can be extracted at different frequency scales. This technique combined with AE signals has been successful in different scenarios [8]. Leng et al. [9] showed a correlation between the RMS value of the AE signals and the energy of the wavelet packet with the tool wear in drilling process.

However, applying the WPT to the signals and extracting the features could result on an excessive number of features, some of them irrelevant (leading to an increase in the error of the ML models). For that reason, to reduce the error and computation demand of ML classification models, feature selection techniques are an efficient tool to select the meaningful information from predicting variables. They can be classified in three categories: filter, wrapper and embedded methods, depending on how it combines the feature selection procedure with the construction of the learning model [10]. RF-RFE is a popular embedded method, and it is much more robust to data over-fitting than other feature selection techniques. Deshpande [11] successfully applied RF-RFE as a feature selection tool with AE sensor and ML frameworks to classify different wear categories simulated with a customized pin-on-disc tribometer.

The main objective of this work is to present methodology that uses different processing techniques and ML models to improve the accuracy of classifying the cutting conditions of machines based on AE signals. This study is based on WPT and is combined with RFE to remove redundant and highly-correlated variables. Besides, different classifiers are used to predict with high accuracy the cutting condition. A segmentation study is presented in this paper to identify the optimum segment length. In addition, a brief analysis on the difference between different cutting conditions is presented.

2 Methodology

This section presents the methodology followed in this work for tool condition prediction using the previously captured AE signals with WPT, RF-RFE, 5 different features (RMS, max, Kurtosis, Skewness and crest-factor) and different ML classifiers (DT, RF, SVM, GP, KNN and MP classifiers):

Step 1- Pre-processing: After ensuring a good signal-to-noise ratio and detecting the noise in the frequency domain, three pre-processing techniques were applied to the signals:

1- Application of a Butterworth filter with cutting frequency, $f_c = 150kHz$ was used with order 5. Subsequently, different features were extracted (RMS, Maximum, Kurtosis, Skewness and the Crest-factor). For extracting those features, the signals were segmented with different times (0.05, 0.1, 0.15, 0.2 and 0.3 seconds)

2- Application of the WPT with 3 levels to be able to compute a multiresolution analysis considering all frequencies (8 nodes of 62 kHz of band width were generated). As well as in the previous point, different segmentation times were considered for extracting the features of each node (in this case, 40 features are generated, 5 in each node).

3- Application of the WPT with 5 levels to be able to compute a multiresolution analysis considering frequencies lower than 150 kHz (32 nodes of 15.62 kHz of band width were generated, and just the first 10 were taken). Besides, different segmentation times were considered for extracting the features of each node (in this case, 50 features are generated, 5 in each node).

Step 2- ML models training and testing: After obtaining the features of the three different pre-processing techniques, 6 different ML classifiers were trained and tested for each pre-processing technique and segmentation time. After this, all the results were compared, and the best scenario was chosen.

Step 3- Dimensionality reduction using RF-RFE: To improve even more the accuracy of the best scenario, RFE was applied. From x_i predictors, by using algorithm with the RF classifier (its performance is the best in "Step 3"), correlated and redundant features were eliminated, reducing the number of predictors to x_j where $j \leq i$. This algorithm decreases the prediction time of ML algorithms and the prediction error.

Step 4- Best scenario training and testing with selected features: After the best features were selected, the best scenario was trained and tested again with these reduced predicting variables. After that, the results were compared with the best scenario that considers all features.

The description of the processing techniques for each step are presented in the sections below.

2.1 Wavelet transform and Wavelet packet

For improving the results of the models, the WPT was applied to the AE signals. The WT is a mathematical tool which transforms sequential data in time axis to the spectral data in both time and frequency [12]. In contrast with sinusoids, Wavelets are localized in both the time and frequency domains, so Wavelet signal processing is suitable for non-stationary signals, whose spectral content changes over time.

The Continuous Wavelet Transform (CWT) is a decomposition of an input function using scaled and translated versions of a Wavelet function known as a mother Wavelet. Mathematically the Wavelet coefficients are extracted using the function below:

$$WT_{\psi}\{x\}(a,b) = \langle x, \psi_{a,b} \rangle = \int x(t)\psi_{a,b}(t)d(t) \tag{1}$$

where $\psi_{a,b}$ is the mother Wavelet.

The Discrete Wavelet Transform (DWT) is the analogous mathematical tool of CWT for discrete functions. It is used for digital signal analysis. The DWT consists of identifying the parameters c_{ke} $d_{j,k}$, $k \in \mathbb{N}, j \in \mathbb{N}$ of the equation:

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \phi(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} d_{j,k} \psi(2^j t - k) \tag{2}$$

where ϕ (t) and ψ (t) are the functions known respectively as father Wavelet and mother Wavelet. The father Wavelet is in fact a scaling function that depends on the mother Wavelet. The ϕ (t) and ψ (t) can be calculated as sequences $h = \{h_n\}$ $n \in \mathbb{Z}$ and $g = \{g_n\}$ $n \in \mathbb{Z}$:

$$h_n = \langle \psi_{1,n}, \psi_{0,0} \rangle_e \psi(t) = \sqrt{2} \sum_{n\in\mathbb{Z}} h_n \phi(2t-n) \tag{3}$$

and

$$g_n = \langle \phi_{1,n}, \phi_{0,0} \rangle_e \phi(t) = \sqrt{2} \sum_{n\in\mathbb{Z}} h_n \phi(2t-n) \tag{4}$$

These two sequences are the base of the DWT.

The common procedure of applying the DWT is through filter bank where the filter determined by the coefficients $h = \{h_n\}$ $n \in \mathbb{Z}$ corresponds to a high-pass filter and $g = \{g_n\}$ $n \in \mathbb{Z}$ corresponds to a low-pass filter. The filters h and g are linear operators that can be applied to a digital input signal x as a convolution:

$$c(n) = \sum_k g(k)x(n-k) = g * x \tag{5}$$

and

$$d(n) = \sum_k h(k)x(n-k) = h * x \tag{6}$$

The signal $c(n)$ is known as approximation and $d(n)$ as detail.

It is possible to repeat the filters shown in Eqs. (5) and (6) generating a cascade of high-pass and low-pass filters. It generates a tree known as a filter bank. However, this decomposition filter may not be precise enough to obtain necessary information from the signal. A more detailed frequency resolution can be obtained by implementing WPT transform to the signal. In a similar way to the DWT, the WPT decomposition tree is obtained by:

$$c_{j+1}^{2p}[m] = \sqrt{2} \sum_{n=-\infty}^{\infty} g[n-2m] c_j^p[n] \tag{7}$$

$$c_{j+1}^{2p+1}[m] = \sqrt{2} \sum_{n=-\infty}^{\infty} h[n-2m] c_j^p[n] \tag{8}$$

where j is the depth of the node and p indexes the nodes in the same depth, every c_j^p with p even is associated to approximations and every c_j^p with p odd is associated to details. The WPT is a generalization of the Wavelet decomposition that offers further decomposition.

2.2 RFClassifier

This classifier, like its name implies, is a combination of tree classifiers that operate as an ensemble. A DT is a classifier in the form of a tree structure (represent rules) that can be understood by humans and used in knowledge system such as database. It is composed by: interior nodes (each one tests a variable), branches (each one corresponds to a variable value and leaf nodes, each one is labelled with a class (class node)).

The tree follows a top-down construction by recursively selecting the "best attribute" to use at the current node in the tree, based on the examples belonging to this node. The "best attribute" is the variable that maximizes the expected reduction of impurity. That impurity can be measured with the information gain and gini index.

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{9}$$

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2 \tag{10}$$

When coming to predictions, each tree predicts a class, and the class with the most votes is the one selected by the model.

In a normal DT, all training data and all features are used to train the model. However, to minimize correlations between different trees in RF, two main techniques are used:

- Bagging: DTs are sensitive to the data they are trained with. Hence, many subsamples with replacement (the same point can be selected multiple times) are created from the training dataset, and each of them is used to train a DT.
- Use of different features: In a normal DT, when a feature must be selected in a node, all features are considered. However, in RF, just one among a random subset of features is selected.

So, with these two techniques, each tree is trained until it reaches the maximum depth on new training data. One of the major advantages of the RF classifier over other DT methods is that the fully grown trees are not pruned.

2.3 RFE

The RFE algorithm is used as a dimensionality reduction measure. The RFE technique, implements a backward elimination of the AE features by ranking their importance to an initial model using all the predictors. The RFE selection method is a recursive process that ranks features according to their importance. It is a greedy optimization procedure used to find the optimum performing subset of features. The RFE needs a classifier (to predict) and an estimator (to select features). In this case RF was used because compared with other methods such as SVM-RFE, RF-RFE has been proven to be more effective, which can use fewer features to get a higher classification accuracy [13]. Detailed information about the RF-RFE algorithm can be found in [14].

3 Experimental procedure

The experimental procedure was based in the orthogonal cutting operation. In this case, the cutting was performed on slotted bars (see Fig. 1) and was carried out on a Danumerik CNC lathe. This approach was selected to reduce the amount of force directions acting against the tool, constraining the process only to the cutting and feed forces. Basically, the slotted bar is clamped to the jaw chuck of the lathe, and the material left between the slots is machined with a linear movement of the tool in the radial direction of the bar. The width of the material to be machined was 2±0.02 mm.

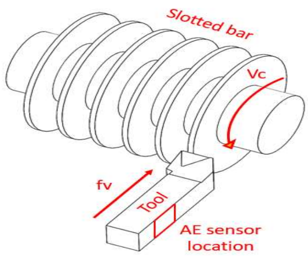

Figure 1: Schematic representation of the orthogonal cutting procedure and location of the AE sensor

During the process, the AEs were recorded with a Kistler 8152B sensor coupled with a Type 5125B conditioning system. This was clamped to the tool holder, as shown in Fig. 2. The acquisition frequency was set to 1MHz, using a National Instrument cDAQ-9171 with an analog input module NI-9223.

The material employed for the trials was a 19NiMoCr6 steel. Tests were done at different cutting conditions shown in table 1:

Table 1. Description of cutting conditions.

Condition	Cutting speed (V_c, m/min)	Feed rate (F_v, mm/rev)
1	100	0.1
2	100	0.2
3	200	0.1
4	200	0.2

4 Results

To ensure that the signals were correctly captured, Figure 2 displays the time domain (measured values with the sensor) and frequency domain spectrums (obtained applying the Fast Fourier Transform (FFT) to the time domain) of the experiment with higher cutting speed and feed rate. It clearly shows higher amplitude when the contact between the tool and piece occurs, between second 7 and 13 approximately. The frequency domain plot shows both frequencies attributed to real AE (0-150 kHz) shown in green and energy associated with electronic noise (150-500 kHz) shown in red.

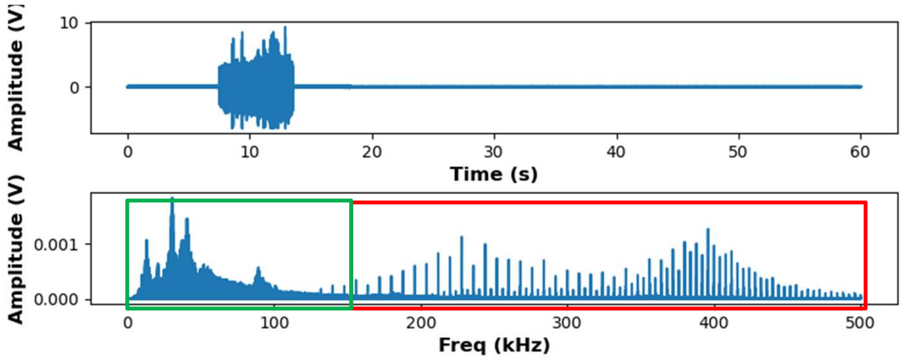

Figure 2: *Time/frequency spectrum of the AE signal with higher cutting speed and feed rate*

Figure 3 displays a comparison among the 4 cutting conditions. First, the cutting region of the signals was selected (initial-final transients values were not taken) and a Butterworth low-pass filter (cutting frequency = 150 kHz, order = 5) was applied. Then, each condition was segmented in 0.05 seconds and the RMS of each segment was calculated (black lines represent the standard deviation of the RMS of each condition). It shows that the cutting speed affects to the amplitude of the signals, while the feed rate is not such significant, which is consistent with the results obtained in [6, 7, 8].

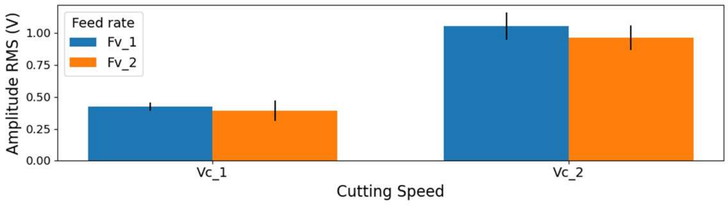

Figure 3: *Comparison of the RMS of the amplitude of the different cutting conditions*

Figure 4 displays the accuracies of the different ML models for the best pre-processing technique along with different segmentation times (the one with WPT 5 levels on frequencies below 150Hz). The models were trained 15 times and the black lines represent the standard deviation of the results. The best results come from the mentioned pre-processing technique, with the RF (99.2% accuracy) and with 0.05 seconds of segmentation time. It is worth mentioning that MLP using 0.1 seconds of segmentation time shows a similar accuracy but its computational demand for training is higher.

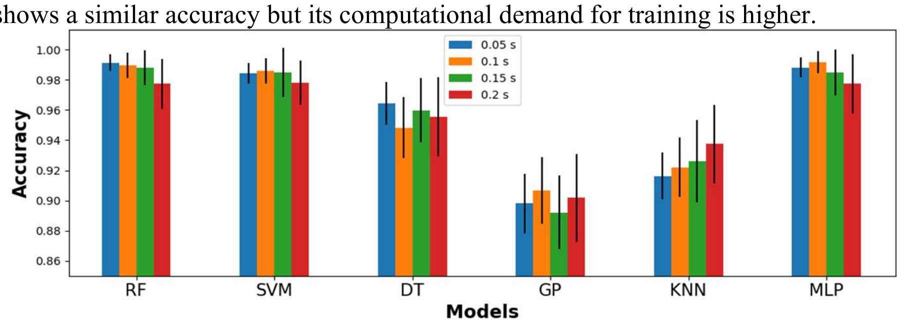

Figure 4: *Accuracies of different models, partition times, WTP and frequencies below 150 kHz*

Figure 5 displays the different accuracies obtained with different number of features by implementing RFE due to it is a stochastic method, it has been executed 20 times. Triangles represent the mean and the lines the standard deviation. Here, a classification model, to classify the data, and an estimator model, to evaluate the best features, are needed. With RF-RFE the optimum number of features was evaluated. It shows that the optimum number is 8 features.

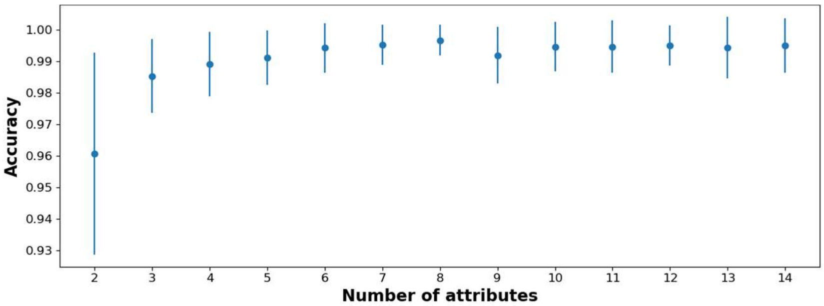

Figure 5*: Evaluation of accuracies with different number of features*

To improve the accuracies obtained when using the RF classifier along with RFE, 4 different estimator models were tried with the 8 best features (Logistic Regression, Perceptron, DT Classifier and RF Classifier). The best estimators were the DT and the RF. As the DT is much faster to compute, the first one was selected.

After applying this feature selection technique with 8 features and the DT as estimator, the following features were selected: 'RMS0', 'RMS1', 'RMS2', 'RMS3', 'RMS4', 'RMS5', 'Kurtosis5', 'Kurtosis7'. The numbers after the feature correspond to the frequency band that each feature belongs to in the WPT decomposition. Since the decomposition level is 5, 32 nodes in total are generated, being 15.62 kHz the bandwidth of each node. For example, RMS0 is the RMS value of the frequencies from 0 to 15.62 kHz, RMS1 is the RMS value of the frequencies from 15.62 to 31.24 kHz and so on.

Finally, the RF Classifier was trained and tested with the previously mentioned conditions and features, and a mean accuracy of 99,7% was achieved (data was split in a balanced way ('stratify'), to ensure that the 4 cutting conditions are present in the training and testing data). Figure 6 displays the corresponding confusion matrix of the model test which is the mean of 50 tests to obtain consistent results. It clearly shows that the model performs with very high accuracy and it is able to predict nearly all the cutting conditions. As it was expected, the model predicts better the situations in which the cutting speed changes and the feed rate remains constant rather than those in which what changes is the feed rate. However, the model also shows a great performance when the cutting speed is constant and when what changes is the feed rate (condition 1 VS 2, condition 3 VS 4).

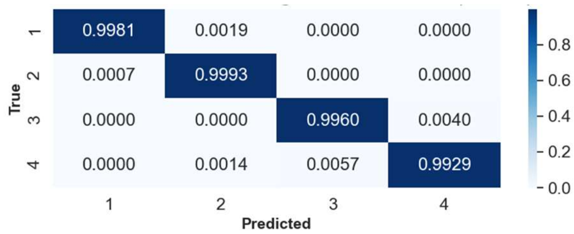

Figure 6: *Confusion matrix using RF Classifier and 8 best features*

The time required to execute the complete method, from filtering the AE signal until the RF forest, identified as the optimum classifier, for 0.05 of signal captured at 1MHz, takes 259± 9 ms. The computer used for data pre-processing and classification is an Intel i5-1035G with 8 GBytes of RAM DDR4 and 256 GB SSD. The data was processed using Python 3.8. Therefore, the algorithm can be used in real-time applications without the need high computational requirements.

5 Conclusions

In this study, a methodology to identify the best ML models and preprocessing tools to characterize cutting condition processes is presented. This methodology is validated constructing a model based on AE for predicting the cutting conditions in a turning process. The effectiveness of different models with different pre-processing techniques was investigated. The results showed that the best performance was achieved applying the WPT (5 levels) and taking the frequencies from 0 to 150 kHz of the AE signals (which is the frequency band that contains the real information about the cutting process), with a data segmentation of 0.05 seconds, selecting 8 features of the 50 possible ones ('RMS0', 'RMS1', 'RMS2', 'RMS3', 'RMS4', 'RMS5', 'kurt5' and 'kurt7') and using the RF Classifier. The achieved results showed a very high accuracy and a model able to classify the 4 conditions with very high accuracy.

Comparing the best models of the pre-processing with and without the WPT, it improves by 0.3% the accuracy of the results (from 98.9% to 99.2%). Besides, by applying the RFE method for feature selection, the accuracy is improved from 99.2% to 99.7%.

The methodology proposed was able define the tools to classify with high accuracy the conditions in which the cutting speed changes. Besides, although the RMS values of the AE signals do not change significantly between conditions of constant cutting speed and different feed rate, the model was able to classify them correctly. This study shows a very high accuracy detecting pre-stablished typical cutting conditions.

The proposed method could be exploited in industrial environments bringing benefits in terms of detecting abnormal or unprogrammed cutting conditions, reducing costs associated with faulty pieces and reducing the production of scrap. Since AE signals are inherently of very low amplitude (µv to mV) the AE signal must be amplified as close as possible to the AE sensor and shielded cable must be used. If high levels of noise are coupled to the AE signal, could have a major impact on the method. The complete process of extracting features from Wavelet nodes and applying RF-RFE feature classification must be applied for each specific case since frequency components of AE signal can vary according to several parameters such as sensor distance to the cutting tool, sensor frequency response, couplant used, etc. Thus, the optimum AE features identified by RF-RFE algorithm could differ completely from one case to the next.

In future investigations, the ML and signal processing techniques selection methodology will be applied and further validated to detect abnormal cutting conditions, such as incorrect lubrication or tool breakage.

Acknowledgement

This research was partially funded by the department of industry of the Basque government within Elkartek programme (KK-2020/00099).

References

[1] C. Bradley and S. Kurada, “A review of machine vision sensors for tool condition monitoring,” Computers in Industry, vol. 34, no. 1, pp. 55-72, 1997.

[2] G. Byrne, D. Dornfeld, I. Inasaki, G. Ketteler, W. König and R. Teti, “The Status of Research and Industrial Application,” CIRP Annals, vol. 44, no. 2, pp. 541-567, 1995.

[3] E. N. Diei and D. A. Dornfeld, “Acoustic Emission from the Face Milling Process,” Journal of Engineering for Industry, vol. 109, pp. 92-99, 1987.

[4] R. Teti and D. Dornfeld, “Modeling and Experimental Analysis of Acoustic Emission from Metal Cutting,” Journal of Engineering for Industry, vol. 111, no. 3, 1989.

[5] S. Rangwala and D. Dornfeld, “A study of acoustic emission generated during orthogonal metal cutting—1: Energy analysis,” International Journal of Mechanical Sciences, vol. 33, no. 6, p. 471–487, 1991.

[6] M. Lan, “Investigation of tool wear, fracture and chip formation in metal cutting using acoustic emission,” PhD Dissertation, University of California at Berkeley, 1983.

[7] D. Wu, C. Jennings and J. Terpenny , “A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests,” Journal of Manufacturing Science and Engineering, vol. 139, no. 7, p. 071018, 2017.

[8] J. F. Chacon, V. Kappatos, W. Balachandran and T. Gan, “A novel approach for incipient defect detection in rolling bearings using acoustic emission,” Applied Acoustics, vol. 89, pp. 88-100, 2015.

[9] S. Leng, Z. Wang, T. Min, Z. Dai and G. Chen, “"Detection of Tool Wear in Drilling CFRP/TC4 Stacks by Acoustic Emission,” Journal of Vibration Engineering & Technologies, vol. 8, pp. 463-470, 2019.

[10] Y. Saeys, I. Inza and P. Larranaga, “A review of feature selection techniques in bioinformatics,” Bioinformatics, vol. 23, pp. 2507-2517, 2007.

[11] P. Deshpande, V. Pandiyan, B. Meylan, K. Wasmer, «Acoustic emission and machine learning based classification of wear generated using a pin-on-disc tribometer equipped with a digital holographic microscope,» Wear, 2001.

[12] P. S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance, 1st edition : Taylor & Francis, 2002.

[13] N. Voyle, «A pathway based classification method for analyzing gene expression for Alzheimer’s disease diagnosis,» J. Alzheimer's Dis., vol. 49, p. 659–669, 2016.

[14] Q. Chen, Z. Meng, Xinyi Liu, Q. Jin and R. Su, «Decision Variants for the Automatic Determination of Optimal Feature Subset in RF-RFE,» Genes, vol. 301, nº 9, 2018.