Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 72 (2018) 1451–1456

www.elsevier.com/locate/procedia

51st CIRP Conference on Manufacturing Systems

Benchmarking of Pattern Recognition Techniques for Online Tool Wear Detection

M. Hassan[a], A. Damir[b]*, H. Attia[a,b], V. Thomson[a]

[a]Department of Mechanical Engineering, McGill University, Montreal, Quebec, Canada
[b]Structures, Materials and Manufacturing Lab (SMM), National Research Council of Canada, 5145 Avenue Decelles, Montreal, Quebec, H3T2B2, Canada

* Corresponding author. Tel.: +1-514-283-9304; fax: +1-514-283-9662. *E-mail address:* ahmed.damir@nrc.ca

Abstract

Pattern recognition techniques have been implemented in real-time tool condition monitoring (TCM) systems to improve their robustness and reliability. The performance and accuracy of these techniques vary depending on their algorithm and the dataset properties. This research benchmarks six pattern recognition techniques to optimize the learning effort, classification accuracy and calculation time for TCM in milling of Al-Alloys using spindle-drive feedback. The techniques were tested using a generalized dataset where the tool condition has a dominant effect over the cutting conditions. The analysis demonstrated the high capability of the linear discriminant analysis technique compared to other techniques.

Peer-review under responsibility of the scientific committee of the 51st CIRP Conference on Manufacturing Systems.

Keywords: Tool condition monitoring; Pattern recognition; Milling

1. Introduction

The evolution of tool wear in milling operations is affected by many parameters due to the process complexity. Currently, milling processes can be accurately evaluated by only inspecting the final workpiece [1], by which time, any damage cannot be avoided. Therefore, advanced investigations of the tool condition monitoring systems are required to achieve the industrial demands for automated machining systems, such as reducing the process cost and standardizing the produced parts quality. These systems overcome the uncertainty of tool life prediction by estimating the tool condition based on process-born features [2]. Several signals such as forces, acoustic emissions, vibrations and spindle motor current were reported as good indicators of tool failure detection [3, 4]. However, the later has high potential for industrial applications due to its low cost, high flexibility and unobtrusive nature.

Signal processing methods used in TCM systems cover the majority of conventional processing techniques, including time, frequency and time-frequency domain analysis. In addition, pattern recognition techniques have been used for tool condition classification and decision making [2]. A comparative study has been carried out on machine learning algorithms for tool wear prediction in high speed milling [5]. The study was carried out at single cutting condition/tool geometry with considerable learning effort, which is necessary for prediction purpose. Although pattern recognition techniques have revealed high potential for TCM applications, they have limitations [6]. Most of these techniques are difficult to be used to estimate the state of the tool condition at different cutting conditions throughout a single process, due to the extracted features sensitivity to the cutting conditions [7]. Additionally, TCM systems need extensive experimental work for system learning in order to build a reliable database of features corresponding to different cutting conditions, tool geometries and tool paths methods. Such database needs a considerable training data that is difficult to be calibrated for tool wear monitoring. This is because actual tool wear must be measured

2212-8271 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 51st CIRP Conference on Manufacturing Systems.
10.1016/j.procir.2018.03.201

after the cutting operation is interrupted, which provides few training data for the correlation stage [8]. Furthermore, the pattern recognition techniques are depended on probabilistic and optimization results of the training dataset and not physical meaning models [6].

The main drawback of previous work related to tool detection in TCM is applying the classification techniques on datasets that were gathered from determined-problem experimental results. Therefore, the performance, accuracy and efficiency of these techniques cannot be generalized. In addition, most of this work has ignored the practicality of their classification techniques selection in the TCM systems with respect to the learning effort. Recently, new approaches have increased the pattern recognition TCM systems certainty, key feature extraction, standardization and generalization [9]. This is mainly by pre-processing the acquired signals to mask the effect of the cutting feed and depth of cut, and emphasize the tool condition effect prior to feature extraction. However, the high dynamics of the milling process still limits the extraction of a stand-alone describing feature. This necessitates fusing several features in a pattern recognition classification technique. Hence, it is essential to analyse the performance of different classification techniques using the acquired signals in the machining process to highlight their efficiency.

This research benchmarks six common pattern recognition classification techniques that have been used in literature, namely Binary Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), K-Nearest Neighbour (KNN), Neural Network (NN), Naïve Bayes (NB), and Decision Trees (DT). These techniques have been considered and compared with respect to their conservation characteristics and computational efficiency to optimize the learning effort, classification accuracy and calculation time for TCM using spindle-drive feedback signals in milling processes of Al-Alloys. The comparison is based on a dataset of high-speed roughing operations that is independent on the cutting conditions. The following sections describe the signal processing technique and the techniques used for pattern recognition, illustrate the test matrix and experimental setup, and discuss the benchmarking approach and results. The comparison outcome demonstrates that the linear discriminant analysis technique is superior compared to other techniques.

2. Signal processing and classification methods

2.1. Signal processing and analysis:

In this work, the spindle-drive feedback current signals have been used to classify the tool condition. A novel processing technique has been adapted to generate a generalized dataset of features to be used for pattern recognition techniques comparison [9, 10]. This technique depresses the effect of the cutting conditions on the extracted features and emphasizes the tool condition effect. It first filters the acquired signals using the second passing frequency as a low pass filter to reduce the signal noises. Filtered signals are then segmented per tool rotation using overlapping windows and each segment is normalized with respect to its maxima. The segmentation process provides comparable patterns owing to the repetitive nature of the milling process. While the normalization process minimizes the cutting forces effect represented in the depth of cut and feed. After processing, a feature vector is extracted in the time and frequency domain for each segment. This vector can be used in a pattern recognition technique to classify the tool condition. In [9], the extracted features have been ranked according to their sensitivity to the tool condition using the results of an N-way analysis of variance test. Table 1 shows the ranking value for the signal mean (M), maximum peak of periodogram (P_p), root mean square (rms), peak to root mean square ratio (P2rms), mean frequency (F_{mean}), band power (BP), median frequency (F_{med}), maximum peak of welch power spectral energy (P_w), kurtosis (K), minimum (min) and variance (Var). The feature ranking score, R_i, varies from -∞ to 1. A value of $R_i = 1$ indicates ultimate sensitivity to tool condition, while $R_i = 0$ shows very low sensitivity to the tool condition. Oppositely, a value of $R_i < 0$ denotes that this feature is sensitive to the cutting condition. In this work, these 11 features were used in the benchmarking process.

Table 1 Ranking score of extracted features according to [9]

Feature	M	P_p	rms	P2rms	F_{mean}	BP
R_i	0.9	0.88	0.86	0.86	0.84	0.84
Feature	F_{med}	P_w	K	min	Var	
R_i	0.8	0.71	0.41	0.25	0.17	

2.2. Pattern recognition techniques:

The most frequent pattern recognition techniques, which were used in literature, have been considered in this study. To apply these techniques, the machine learning toolbox of MATLAB© software has been used and empirical prior probabilities depending on the trained data relative frequencies were used whenever applied. Low number of training samples may produce undefined classifier, while high number may lead to over-defined classifier. In both cases, the error increases. Therefore, different number of training samples have been tested to reach near-optimum classifiers.

2.2.1. Support Vector Machine (SVM)

The support vector machine solves the problem of separation of two classes by finding a linear function *f*, called *hyperplane* that separates the classes and finds the widest margin between them by minimizing *w* as follows [11]:

$$f(x) = (x \cdot w) + b; \begin{Bmatrix} f(x) > 0 & for\ x \in class\ 1 \\ f(x) < 0 & for\ x \in class\ 2 \end{Bmatrix} \quad (1)$$

This is done usually using the sequential minimal optimization technique. For nonlinearly separable data, a slack variable is allowed of samples from boundaries of the separation margin with a penalization parameter. Support vector machine has been proved less vulnerable for overfitting problem and higher generalization ability since it is designed to minimize structural risk [12]. In addition, the technique does not require a large number of training samples and can solve the learning problem even when only a small amount of training samples are available [11]. A linear SVM with a regularization parameter equal to 1 has be used in this work.

2.2.2. Linear Discriminant Analysis (LDA)

Linear discriminant analysis is a mathematical model to classify multivariate data based on statistical analysis. It is based on assumptive Gaussian distributions of the learned data classes. The score function of the LDA model can be expressed as follows [13]:

$$Sf(\beta) = \frac{Var_{between}}{Var_{within}} = \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \quad (2)$$

The function goal is to maximize the variance between classes over the variance of the data within the same class. The assumed distribution parameters are used to search for a linear combination of variables that best separate the learned data classes. This linear combination is used to determine the class of the tested data.

2.2.3. K-Nearest Neighbor (KNN)

K-Nearest Neighbor classification is a fundamental classification method that is recommended when it is difficult to determine reliable parametric estimates of probability densities [14]. It is based on learning by analogy. It measures the distance between the tested observation and the closest K-nearest neighbors in the learned dataset in an n-dimensional space. In this work, Euclidean distance metric is utilized to determine the neighbors' closeness. The tested observation class is determined using the assumption that objects near each other are similar. Hence, it categorizes tested data based on the classes of their nearest neighbors in the learned dataset.

2.2.4. Neural Network (NN)

Neural Networks are a computational model inspired by the neural structure of the human brain, which find data structures and algorithms for learning and classification of data. They consist of interconnected group of multi-layers of neurons that relate the inputs to the desired outputs. The network is trained by iteratively modifying the strength of the structure connections based on the information flow through the network to map the inputs to the correct response [15]. Usually NN is used to model complex relationships and find patterns in nonlinear data. A one-layer NN of size 10 with one output has been constructed using the Levenberg-Marquardt training algorithm and the extracted features in this work.

2.2.5. Naïve Bayes (NB)

Naïve Bayes is a statistical technique to construct classifiers that predict the probabilities of each class feature. They are based on Bayes theorem with strong (naive) independence assumption between the features. This assumption states that the conditional probability of a feature for a class is independent from the conditional probabilities of other features given that same class. They use the joint probabilities of a new data to predict its class depending on the dataset features probability distribution [16]. They classify new data based on the highest probability of its belonging to a particular class. Despite the incorrectness in this assumption, as regularly features are dependent, NB classifiers are simple to be applied and usually provide high accuracy [16]. Normal distribution was assumed to calculate the predictor distribution parameters within each class.

2.2.6. Decision Trees (DT)

A decision tree is a hierarchical model composed of decision rules that recursively splits independent variables into homogeneous zones [17]. It is a flowchart-like tree structure of decision rules to predict a new observation class from a set of input features. At each tree branching condition (node), the new observation features value is compared to a weight obtained from the training dataset. The number of branches and the values of weights are determined in the training process. During the training process, the DT technique identifies and removes branches that may reflect noises in the training dataset to improve the classification accuracy. Decision trees has been applied successfully in many real-world situations for classification and prediction. The Gini's diversity index was used in this work as a split criterion to classify the tool condition.

3. Experimental setup

High speed rouging milling tests of high strength aluminum alloy workpieces have been performed on a 5-axis DMU 100P duoBlock machining center with 28 kW spindle and maximum speed of 18,000 rpm. The experimental setup is shown in Fig. 1(a). Three AC-DC and pulsed current signals transducers have been used to measure the spindle current signals. The transducers have a measuring frequency band and reaction time 100 kHz and 0.5 μs, respectively. The three transducers where mounted on the three phases between the spindle motor and its pulse width modulation PWM driving module. A National Instrument data acquisition card type NI 4472 Series has been used to digitalize and store the acquired signals.

Milling operations of straight slots have been performed using 2, 3 and 4 flutes carbide end mills at speed of 14,000 rpm. Table 2 shows the full factorial matrix of the cutting conditions (i.e. feed and depth of cut a_e), the tool diameters (Dia), corner radius (Cr) and flank wear levels (VB) used in the slotting tests. A total of 80 slot tests, including one replicate for each set, have been performed to induce the effect of the tool and cutting parameters on the extracted features.

According to the ISO standards [18], tool flank wear (VB) is the phenomenon of tool life deterioration. It has maximum acceptable uniform value of 0.3 mm. In this work, uniform tool wear has been used as tool life criterion. The tested tool conditions have been categorized into two ranges; namely, fresh ($0 \leq VB < 0.07$) and worn tool ($0.25 \leq VB < 0.3$), as shown in Fig. 1(b). To generate tool wear prior to the milling test, the targeted tools have been subjected to severe cutting conditions under controlled machining conditions to accelerate the induced uniform tool wear.

Table 2 Cutting conditions and test matrix

Tool	Dia. (mm)	Cr (mm)	VB (mm)	Feed (mm/tooth/rev)	a_d (mm)
T1	16	0.4			
T2		3.3	0 - 0.07	0.1	3
T3	20	0.4			
T4		3.3	0.25 - 0.3	0.14	5
T5	25	0.4			

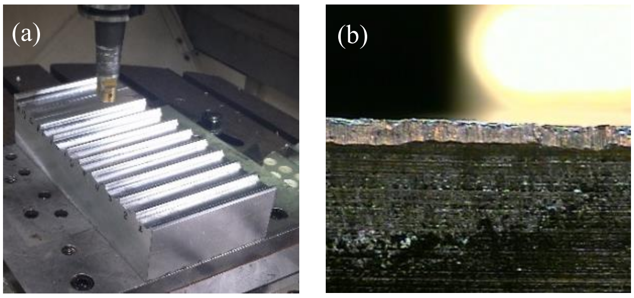

Fig. 1. (a) Experimental setup and (b) worn tool, VB=0.28mm

4. Comparison approach, results and discussion

In this section, the benchmarking approach of the six pattern recognition techniques and the outcome results are demonstrated and discussed. The techniques performance has been compared according to their practicality in tool condition monitoring applications.

4.1. General approach of comparison

The six pattern recognition techniques have been applied using the following approach, as shown in Fig. 2:

1. The resultant current signals for each slot cut have been filtered, segmented per tool revolution and normalized with respect to its maxima.
2. For each processed signal segment, the top-ranked features, reported in Table 1, have been extracted and sorted in according to R_i. Figs. 3(a) and (b) show the normalized mean and standard deviation values of the top ranked features extracted from the current signals acquired from the fresh and worn conditions of the five tested tools before and after processing. The processing technique has successfully separate the features extracted from the two tool conditions into two distinct distinguishable clusters depending on the tool condition by separating the features mean values and limiting their deviation, as shown in Fig. 3(b) compared to Fig. 3(a).
3. The extracted features dataset has been divided to two distinct datasets for training and testing. Due to the signal processing technique capability to mask the cutting conditions effect on the extracted features, the training dataset has been generated from the features extracted from only one cutting feedrate and depth of cut combination per tool condition. Cuts with feed and axial depth of cut of 0.1 mm/rev/tooth and 5 mm respectively have been selected for training. In total, 10 cuts have been used as a training database for the five tested tools. The rest of the slot cuts have been used for classification models testing.
4. For classification model training, five training subsets *Ti* have been created using 20%, 40%, 60%, 80% and 100% of the training dataset. Each subset has been used for generating 11 different models of the same classifier technique using an m number of features features Ti_m. Each subset Ti_m represents m number of features, where m varies from 1 to 11 according to their ranking score *Ri* in sequence. In total, 55 models have been generated and analysed for each pattern recognition technique. A 5-fold Cross validation method has been implemented in the training process in order to train the models by 10 stages of the training data.
5. The testing dataset has been used to test and analyze the accuracy and computational time for the generated models.

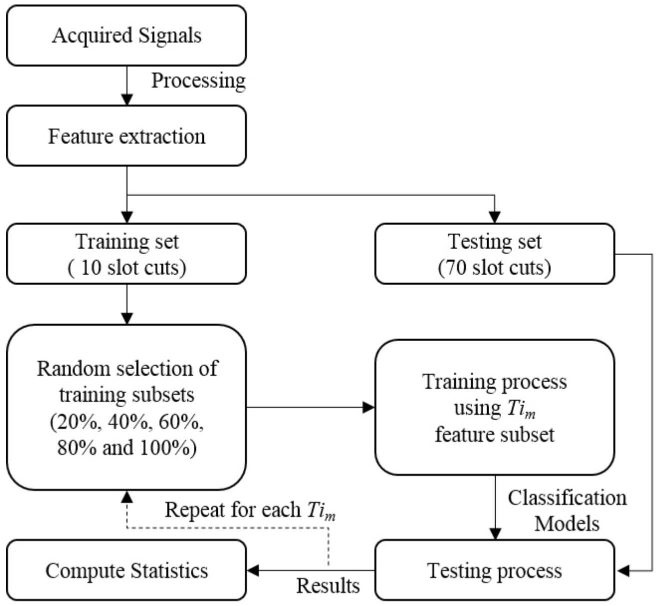

Fig. 2 Benchmarking general approach

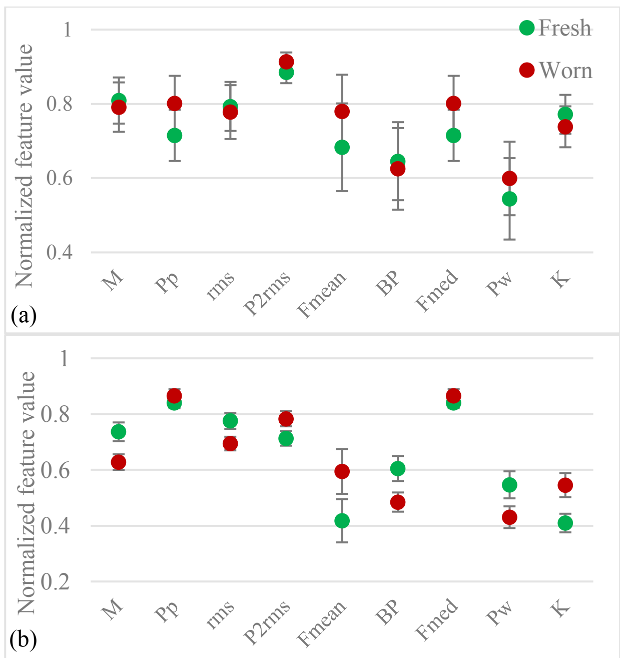

Fig. 3. The mean of the top-ranked features extracted from the current signals of the five tested tools (a) before processing (b) after processing

The techniques have been compared according to the classification accuracy, computational time and learning effort, which consists of the number of training segments and the number of extracted features needed in the learning process. For a more representation of the classification techniques accuracy from a practical point of view, two types of false classification errors of the tool condition are introduced, namely, Safe False Alarm (SFA) and Unsafe False Alarm (USFA) rates. USFA occurs when the tool is worn but the technique classified it as fresh. The SFA is the opposite where the tool is classified as worn while it is in a good state. While SFAs may reduce productivity, they will not affect the part quality. In contrast, the USFA condition could lead to more damage as the part surface integrity could be affected before tool replacement. Both errors rates have been computed for all the classification techniques.

4.2. Learning effort and accuracy

Figs. 4(a) and (b) show the maximum accuracy achieved by the six classification techniques with respect to (a) the training dataset size and (b) the number of features used for training respectively. In general, the LDA and SVM showed higher accuracy than the other techniques. In addition, they were able to achieve an almost 90% accuracy using only 20% of the training data. This accuracy was almost the same when five or more features were used regardless of the training dataset size. Hence, 20% of a training set consisting of the five top ranked features were enough to achieve high accuracy for both techniques. The high performance of the LDA technique can be referred to its assumption of training data normality and not adapting the data distribution. Therefore, the LDA prediction errors are due to the errors in estimating a representative mean and variance out of the training dataset. Whereas the SVM ability to minimize structural risk has limited its classification errors to the difficulty of calculating global boundaries for the separation margin using the training data. However, the LDA has showed higher accuracy compared to the SVM even when only one feature was used for training. Hence, the LDA classification technique is recommended for applications with a limited training data as in the TCM system learning process for machining operations.

On the other hand, DT and KNN techniques, which adopt the training set distribution, showed the lowest accuracy, with a maximum of 84.2% and 84.6% respectively, through all the training dataset sizes. However, the KNN technique presented an increasing trend by increasing the number of training samples. The NB and NN showed a steady accuracy around 85% regardless of the number of features used. The NN had a decreasing trend by increasing the number of training samples. It also provided the highest accuracy when five features were used, which has decreased afterwards. This performance illustrates the technique sensitivity to the training dataset size, which is in agreement with the results found in [5]. Such performance may not provide a generalized approach for TCM systems.

Fig. 5 shows the safe and unsafe false alarms rates achieved by the classification techniques at their highest accuracy levels. The LDA showed the lowest SFA and USFA rates with values of 6.2% and 3.6% respectively, followed by the SVM technique. On the other hand, the DT, KNN and NN techniques showed an USFA rate as high as 10%, while the NB technique reached the highest SFA rate. From these results, a conclusion can be drawn that implementing the LDA and SVM techniques in TCM systems should increase the systems accuracy and the machining process productivity.

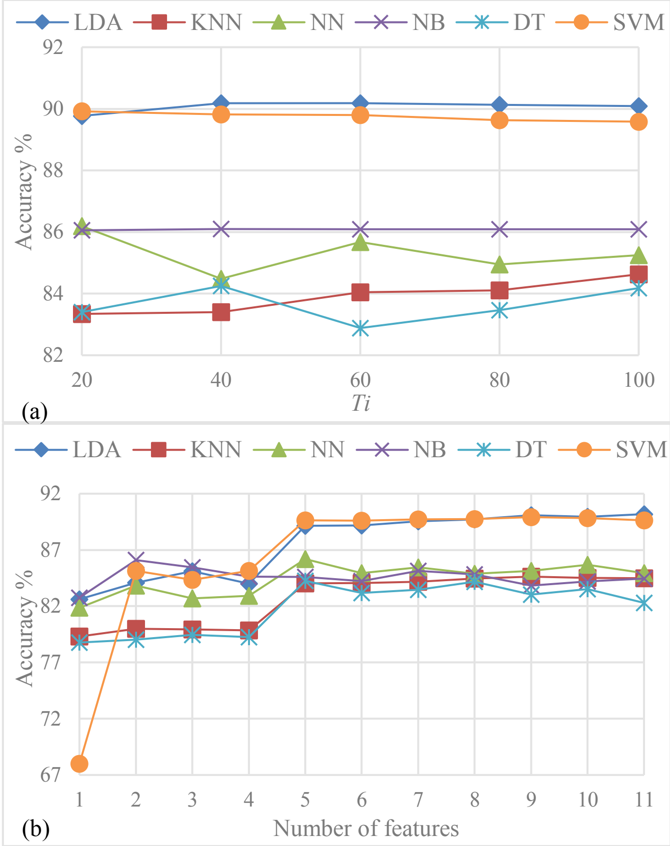

Fig. 4 Classification technique accuracy

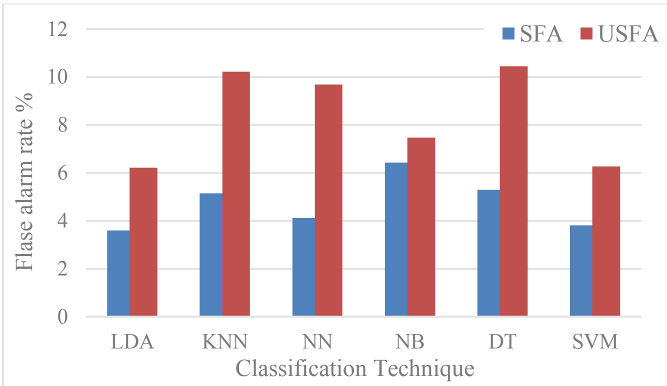

Fig. 5 Safe and Unsafe false alarms at the highest accuracy values

It should be noted that the tool diameter and corner radius have not been included in the study performed in [9]. In this work, each model has included five different tools. The results showed the capability of fusing the adopted processing technique with pattern recognition techniques to mask the effect of different tool diameters and corner radius while preserving the same accuracy level reported in [9].

4.3. Computational time

Early detection of tool wear will minimize the worn tool impact on the workpiece surface integrity. Hence, the time needed per one revolution will be used as a reference for comparing the pattern recognition techniques classification time. In this work, a speed of 14000 rpm was applied. At this speed, 4.28 ms is required for full tool rotation. Fig. 6 shows the average time needed for each classification technique to classify one segment. The time ranges have been calculated for all the tested sample sizes Ti_m. The results show that the DT has the lowest classification time of 6.8µs and the lowest deviation as well, followed by the NB and LDA respectively. While the KNN showed the highest classification time of 114.7 µs. Although these time ranges are too low compared to the time needed for 1 revolution at high speed machining applications, it represents the relative computational effort and memory needed to classify the tool condition.

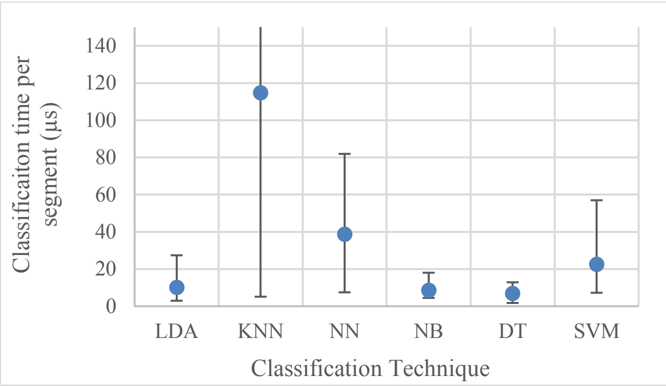

Fig. 6 Classification time per segment

5. Conclusion

The practicality of applying different pattern recognition techniques in TCM systems has been benchmarked in this work. The following conclusions can be drawn from the conducted tests, analysis and comparison:

- The linear discriminant analysis, followed by the support vector machine, are the most recommended classification techniques for TCM applications.
- The LDA has shown the highest classification accuracy and lowest USFA rate using a limited learning effort with an applicable classification computational time. This shows the high performance and applicability of such technique when the provided training data is limited as in TCM systems learning process.
- Decision trees and k-nearest neighbor classification techniques have provided the lowest accuracy and highest USFA rate. Hence, the application of these techniques in TCM systems should be minimized.
- The neural network and naïve Bayes classification techniques have provided a steady accuracy regardless of the training size. However, the neural network technique has provided higher USFA rates.
- Fusing the adopted processing technique with pattern recognition techniques can mask the effect of the tool diameter and corner radius on the learning and classification process in TCM systems.

Acknowledgements

This work was conducted under the partial financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC), which the authors greatly appreciate. The authors would like also to acknowledge the support of the Structures, Materials and Manufacturing Laboratory of the National Research Council of Canada, where all the tests were conducted.

References

[1] Downey J, O'Leary P, Raghavendra R. Comparison and analysis of audible sound energy emissions during single point machining of HSTS with PVD TiCN cutter insert across full tool life. Wear Wear. 2014;313:53-62.
[2] Lauro CH, Brandão LC, Baldo D, Reis RA, Davim JP. Monitoring and processing signal applied in machining processes – A review. Measurement. 2014;58:73-86.
[3] Abellan-Nebot J, Romero Subirón F. A review of machining monitoring systems based on artificial intelligence process models. Int J Adv Manuf Technol. 2010;47:237-57.
[4] Teti R, Jemielniak K, O'Donnell G, Dornfeld D. Advanced monitoring of machining operations. CIRP Annals-Manufacturing Technology. 2010;59:717-39.
[5] Wu D, Jennings C, Terpenny J, Gao RX, Kumara S. A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. Journal of Manufacturing Science and Engineering. 2017;139:071018--9.
[6] Benardos P, Vosniakos G-C. Predicting surface roughness in machining: a review. International Journal of Machine Tools and Manufacture. 2003;43:833-44.
[7] Nouri M, Fussell BK, Ziniti BL, Linder E. Real-time tool wear monitoring in milling using a cutting condition independent method. International Journal of Machine Tools and Manufacture. 2015;89:1-13.
[8] Zhang D. An adaptive procedure for tool life prediction in face milling. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology. 2011:1350650111414332.
[9] Hassan M, Sadek A, Attia MH, Thomson V. A Novel Generalized Approach for Real-Time Tool Condition Monitoring. Journal of Manufacturing Science and Engineering. 2017;140:021010--8.
[10] Hassan M, Sadek A, Attia H, Thomson V. A Novel Generalized Approach for Real-Time Tool Condition Monitoring. Journal of Manufacturing Science and Engineering. 2017.
[11] Widodo A, Yang B-S. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical Systems and Signal Processing. 2007;21:2560-74.
[12] Shi D, Gindy NN. Tool wear predictive model based on least squares support vector machines. Mechanical Systems and Signal Processing. 2007;21:1799-814.
[13] McLachlan G. Discriminant analysis and statistical pattern recognition: John Wiley & Sons; 2004.
[14] Peterson LE. K-nearest neighbor. Scholarpedia. 2009;4:1883.
[15] Bose NK, Liang P. Neural Network Fundamentals with Graphs, Algorithms and Applications, McGraw-Hill Series in Electrical and Computer Engineering. 1996.
[16] Murphy KP. Naive bayes classifiers. University of British Columbia. 2006.
[17] Myles AJ, Feudale RN, Liu Y, Woody NA, Brown SD. An introduction to decision tree modeling. Journal of Chemometrics. 2004;18:275-85.
[18] ISO8688-2. Tool Life Testing in Milling – Part 2: End Milling. International Standard, first edition.: International Organization for Standardization, Geneva, Switzerland; 1989.