Maderas-Cienc Tecnol 22(2):2020
Ahead of Print: Accepted Authors Version
DOI:10.4067/S0718-221X2020005XXXXXX

USE OF NEAREST NEIGHBORS (k–NN) ALGORITHM IN TOOL CONDITION IDENTIFICATION IN THE CASE OF DRILLING IN MELAMINE FACED PARTICLEBOARD

Albina Jegorowa[1]*, Jarosław Górski[1], Jarosław Kurek[2] and Michał Kruk[2]
[1]Faculty of Wood Technology, Warsaw University of Life Sciences, Warsaw, Poland.
[2]Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, Warsaw, Poland.
***Corresponding author:** albina_jegorowa@sggw.pl
Received: January 22, 2019
Accepted: December 25, 2019
Posted online: December 26, 2019

ABSTRACT

The purpose of this study was to develop an automatic indirect (non-invasive) system to identify the condition of drill bits on the basis of the measurement of feed force, cutting torque, jig vibrations, acoustic emission and noise which were all generated during machining. The k-nearest neighbors algorithm classifier (k-NN) was used. All data analyses were carried out in MATLAB (MathWorks – USA) environment. It was assumed that the most simple (but sufficiently effective in practice) tool condition identification system should be able to recognize (in an automatic way) 3 different states of the tool, which were conventionally defined as "Green" (tool can still be used), "Red" (tool change is necessary) and "Yellow" (intermediate, warning state). The overall accuracy of classification was 76 % what can be considered a satisfactory result at this stage of studies.

Keywords: Drilling, melamine faced particleboard, k-NN classifier, tool condition identification.

INTRODUCTION

The term “Industry 4.0” (in short: “I 4.0”), used for several years, means the latest direction of technical development involving intelligent automation and data exchange in manufacturing technologies. There is more and more talk about the 4th industrial revolution (4IR) taking place in front of our eyes (e.g. Schwab 2017; Wagner *et al.* 2017). As part of this revolution, we are heading to intelligent manufacturing in which the idea of cyber-physical Production Assistance System (cPAS) is to play an important role. This idea includes development of “autonomous devices, sensors and machines that monitor themselves” which are able to “perform condition-based, decentralized small tasks for continuous monitoring and self-diagnosis” (Bergweiler 2016). It should be honestly admitted that both the wood industry and furniture production are not the leading industrial branches in terms of intelligent manufacturing. Despite some research successes in this field (e.g. Iskra and Hernandez 2012) it is obvious that there is a lot to catch up to compared to, for example, the machine industry (especially the car industry). One of these arrears is the lack of any commercial or even prototype offer for automatic tool condition monitoring (TCM) systems when processing wood and wood-based materials. Serious research on this subject has been conducted for many years (e.g. Lemaster 2000a,b; Szwajka and Górski 2006; Wilkowski and Górski 2011; Świderski *et al.*; 2017 Górski *et al*. 2019; Jegorowa *et al.* 2019), yet there is still a lot to do before the problem is solved. Therefore, the purpose of this study was to develop an automatic indirect (non-invasive) system to identify the condition of drill bits on the basis of the selected signals generated in the machining zone, such as feed force, cutting torque, acoustic emission, noise and vibrations. All data analyses were carried out in MATLAB (MathWorks – USA) environment. Such a system could be a support for the machine tool operator, telling him when it is necessary to replace the worn tool with a new one. In the long-term perspective, as part of the 4th industrial revolution, such a system would become one of the elements of some more advanced Production Assistance System and further reduce the human role in controlling the machining process.

K–nearest neighbours (k–NN) algorithm

K–nearest neighbors is one of the simplest classification algorithms. The idea of this method is not new (Fix and Hodges 1951) and belongs to lazy algorithms. It is characterized by the fact that it does not create an internal representation (model) of the training data set and looks for a solution only when the object requiring classification appears. The classification of the new x object consists in its comparison with the nearest neighbors and classifying it to the class that is represented by the majority of its k nearest neighbors. In order to determine the affiliation of the new object to the given class, the distance between it and all other objects belonging to the training data set is calculated. This distance is calculated in the multidimensional feature space using, e.g. Euclidean distance, which is defined in Equation 1: follows:

$$D(x,y)^2 = \sum_{i=1}^{n} (x_i - y_i)^2 \qquad (1)$$

where:

x – the object currently classified, belonging to the test data set (with unknown class label);

y – the object belonging to the training data set (with known class label).

n – number of features of the objects which are taking into account (feature space dimension).

The use of the algorithm requires the selection of the parameter k. General, useful in practice methods for optimal choice of k have apparently not been known (Hall *et al.* 2008). It is simply necessary to select k that will give satisfactory classification effects in a specific case.

Formally the parameter k can take values from 1 to the total number of objects in the set. For example, if k = 1, the new object will be simply assigned to the class that matches the class of its nearest neighbor. If k is too small in relation to the total number of objects belonging to the training data set, the algorithm will not be resistant to noise, and thus the quality of the classification will be poor. With too high k, the complexity of calculations is not only excessively increased, but what's worse, the object will be incorrectly identified too often as belonging to the most represented class in the training data set. The advantages of the k–NN classifier are, first of all, an unlimited number of classes, a simple method of operation and the ease of implementation in a wide range of practical applications.

MATERIALS AND METHODS

The standard CNC (Computerized Numerical Control) machining center (Busellato Jet 100), Φ12 tungsten carbide drill bits (WP-01, FABA – Poland) and melamine faced particleboard (U511SM – Swiss Krono Group) were used in experimental part of study. Cutting parameters (spindle speed 4500 rpm, feed rate 1,35 m/min) were adopted on as recommended by the drills manufacturer. The experimental set up enabled the measurement and digital recording of 5 signals generated in the machining zone: feed force, cutting torque, acceleration of jig vibration, audible noise and ultrasonic acoustic emission. The recording of these signals was performed in the NI LabView (National Instruments - USA) environment using 2 data acquisition cards. The outline of the test stand along with details on the measuring system is shown in Figure 1. General view of the jig holding the workpiece is shown in Figure 2.

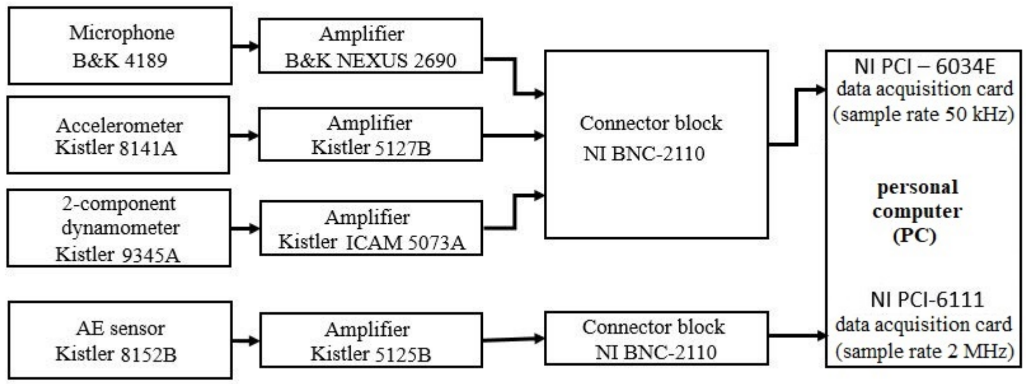

Figure 1: Structure of measuring system used in experiment.

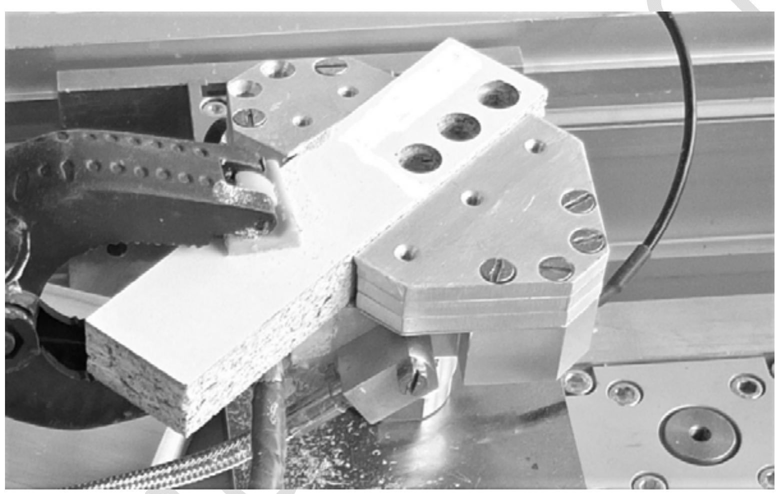

Figure 2: General view of the jig holding the workpiece (in front of the jig base the accelerometer is visible).

The tool condition was monitored in a traditional way - the size of wear of the external corner of the drill was measured using a microscope with a digital camera (Mitutoyo – 505 – Mitutoyo Corporation, Japan). The size of this wear was determined separately for each of the two drill bit blades, and then averaged. The final measurement result (W) was given in millimeters.

Six drill bits (5 experimental and 1 control) were used during the study. Each experimental drill bit was subjected to 8 operating cycles. More specifically, one of them was subjected to 7 cycles, because in this case the tool life turned out to be shorter than in other cases, which surprised and disappointed the experimenter. This way, all experimental tools gradually changed their state from perfectly sharp (W=0 mm) to definitely worn out (W > 0,5 mm). These cycles resulted in the execution (outside the measuring stand) of a series of holes as long as an increase in the wear of the drill bit corner of at least 0,05 mm (the real state of the drill bit was periodically monitored using the microscope) was achieved. After each such cycle, the experimental drill bit was sent to the measuring stand and there a series of 5 holes was made with its use, recording signals: feed force, cutting torque, acceleration of jig vibration, audible noise and ultrasonic acoustic emission. Then, a series of 3 holes was made by means of the control drill bit (all 5 diagnostic signals were still recorded). The sixth drill bit played the role of the control drill bit all the time and was used only on the measuring stand. Throughout the experiment, the control drill bit remained sharp (W<0,2 mm).

It was assumed that the most simple (but sufficiently effective in practice) tool condition identification system should be able to recognize (in an automatic way) 3 different states of the tool, which were conventionally named as “Green”, “Yellow” and “Red” condition. The explanation of these code names is as follows. “Green” state means that the wear of the external corner (W) of the drill is less than 0,2 mm and the tool is ready to work. “Red” state means that the tool wear (W) is greater than 0,35 mm, so the drill is absolutely worn out and cannot work. “Yellow” means intermediate state (0,2 mm < W < 0,35 mm), which should be interpreted as a warning.

After the experimental research was completed, the standard machine learning (ML) procedure was performed using MATLAB environment. The training and test data sets were established and next the k–nearest neighbours (k–NN) algorithm has been used. By creating and testing various options of the identification system, the test data set was always established, which included features of all signals recorded with 1 experimental drill bit (it played the role of the test drill bit). At the same time, the training data set was established, i.e.

the database concerning 5 remaining drill bits (there was always the control drill bit and 4 remaining experimental drill bits among the test drill bits), which covered both knowledge about their real current states ("Green", "Yellow" or "Red"), determined on the basis of the microscopic measurement, as well as features of all signals recorded with their participation on the measuring stand. Then, the k-NN algorithm was activated, which was to recognize the changing over time state of the test drill bit on the basis of features of signals included in the test data set. All features used in the study (91 standard features such as: root mean square, arithmetic mean, standard deviation, kurtosis, parametres based on fast Fourier transform and wavelet transform etc.) were extracted by means of functions available in MATLAB Signal Processing Toolbox and Wavelet Toolbox (MathWorks – USA). Feature space dimension (total number of features) was 455 (91 features for each out of 5 measuring channels). Therefore, the fundamental (from the suggested identification system's point of view) mathematical formula (which was based, of course, on formula no 1, presented earlier) had the following form:

$$D(x,y)^2 = \sum_{i=1}^{455}(x_i - y_i)^2 \quad (2)$$

where:

x - the currently classified case (belonging to the test data set with unknown color labels);

x_i – value of i-th feature (one of 455 features) which characterized the case "x";

y - the case "y", belonging to the training data set (with known color label);

y_i – value of i-th feature (one of 455 features) which characterized the case "y";

D (x,y) – the distance (defined in the multidimensional feature space) between the case "x" and the case "y".

This way, the distances (defined in the multidimensional feature space) between the currently classified case (belonging to the test data set with unknown color label) and all cases belonging to the training data set (with known color label) were calculated. Next, the case with unknown color label was automatically assigned to the class that was represented by the majority of its k nearest neighbors belonging to the training data set.

The aforementioned procedure was performed many times, using different values of the parameter k (from 1 to 15). In each case (for each k value), 5 different options of the test data set (including, one by one, each of the 5 experimental drill bits) were developed. In this way, 5 separate tests were performed, but only one integrated (collective) matrix confusion was developed on their basis. This matrix always included 215 cases constituting an integrated (collective) test data set. On the basis of such a matrix the classification quality indicators shown below were calculated. The overall accuracy (Acc_o), i.e. the ratio of the number of cases of correct classification of the test drill bit state to the total number of cases that were attempted to classify. In addition, for each of the three classes, the following parameters were defined separately:

TP - number of true positive identifications,

TN - number of true negative identifications,

FP - number of false positive identifications,

FN – number of false negative identifications.

On this basis, 5 standard, detailed indicators of the effectiveness of its identification by the k-NN classifier were calculated for each class separately (Eq. 3-7):

Sn – sensitivity (recall),

$$\mathrm{Sn} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3}$$

Sp – specificity,

$$Sp = \frac{TN}{TN + FP} \quad (4)$$

Pr – precision,

$$Pr = \frac{TP}{TP + FP} \quad (5)$$

Acc – accuracy,

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \quad (6)$$

Fscore,

$$F_{score} = 2\frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} = \frac{2TP}{(2TP + FP + FN)} \quad (7)$$

Each of the above 5 indicators can take values in the range from 0 (which would mean the lack of any correct identification) to 1 (in case of 100 % accuracy of identification).

RESULTS AND DISCUSSION

Figure 3 shows the impact of the parameter k on the overall accuracy of classification. When analyzing this impact, it was arbitrarily found that it was not worth increasing the value of the parameter k above 12 (at this point the clear upward trend ended). Overall accuracy of classification for k=12 was 0,76.

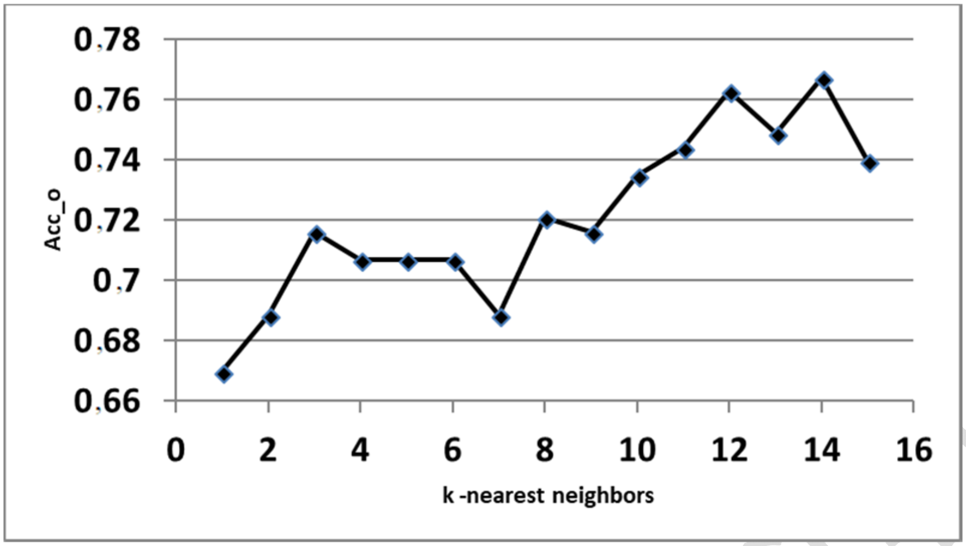

Figure 3: The effect of constant k (user-defined, basic parameter of k-NN algorithm) on the overall accuracy of classification (Acc_o) i.e. the ratio of the number of cases of correct classification of the test drill bit state to the total number of cases that were attempted to be classified.

Detailed information about the effectiveness of the k-NN algorithm for k=12 is included in the confusion matrix (Table 1).

Table 1: Confusion matrix for k=12 (total number of identification cases: 215), described in the text below.

Output tool condition	"Green"	**29,8 %**	6,5 %	0 %
	"Yellow"	5,1 %	**11,2 %**	1,9 %
	"Red"	0 %	10,2 %	**35,3 %**
		"Green"	"Yellow"	"Red"
		Target tool condition		

This is a percentage (not numerical) form of a standard confusion matrix - a specific table used for visualization of the performance of the classification algorithm. The rows represent the classes resulting from (either correct or not) decisions made by the classifier. Columns, on the other hand, reflect real classes. If it was a numerical form of a confusion matrix, then the number at the intersection of j-th column and i-th row would be equal to the number of cases from the j-th class that have been classified as belonging to the i-th class. However, the percentage form of a matrix confusion (such as in table 1) contains the percentages of cases classified to particular classes calculated as the ratio of the count of cases classified to a class to the total cases count. This means that in cells located on the diagonal (bold font), there are percentages that reflect the frequency of correct identifications of the tool state. The correct identification means the overlap between the target (real) and the output (predicted) tool condition. For example: value 29,8 % (at the intersection of first column and first row of table 1) means that 29,8 percent of the total number of cases (that were attempted to be classified) belonged to the "Green" class (W<0,2 mm) and were correctly classified as such. On the other hand, all other cells showed (also in percentages) how often different types of identification errors occurred. For example: value 5,1 % (at the intersection of first column and second row of table 1) means that 5,1 percent of the total number of cases that belonged to the "Green" class (W<0,2 mm) and have been incorrectly classified as "Yellow".

A particularly positive fact is that the two cells, furthest from the diagonal mentioned above, contain zero values. This means a complete lack of the most compromising (inexcusable from the practical point of view) mistakes of the k-NN algorithm. The zero value in the right top cell indicates that not once was the really "Red" (W>0,35 mm) drill bit incorrectly classified as the "Green" (W<0,2 mm) drill bit. The zero value in the bottom left cell indicates that none of the cases belonging to the "Green" class were ever classified as "Red". Unfortunately, sporadically (1,9 %) the real "Red" drill bit was considered to be the "Yellow" drill bit.

The table 2 contains detailed indicators of the effectiveness of identification of particular tool states by the k-NN algorithm for k=12. On the basis of the data shown in the table 2, it can be concluded that Acc and Fscore for the "Green" and "Red" classes were at a very similar, relatively high level (0,84÷0,88). The "Yellow" class was much less recognizable (Acc=0,76 and Fscore=0,48). This is quite understandable because this class is adjacent to both the "Green" and "Red" classes. In addition, the wear range, characteristic for this class, is relatively narrow (W=0,2÷0,35 mm).

Table 2: The effectiveness of tool condition identification - quantitative quality indicators calculated separately for each class.

Tool condition	Acc	Sn	Sp	Pr	Fscore
„Green”	0,88	0,85	0,90	0,82	0,84
„Yellow”	0,76	0,40	0,90	0,62	0,48
„Red”	0,88	0,95	0,84	0,78	0,85

CONCLUSIONS

The use of the k-NN algorithm in engineering practice requires the arbitrary establishment of the parameter k. In the analyzed case, it was considered that it was not worth increasing the k above 12 (for k>12 there were no unambiguously better effects of the algorithm).

The overall accuracy of classification observed for k=12 was 0,76 what can be considered a satisfactory result at this stage of studies and especially that the real “Red” drill bit was never considered to be the “Green” drill bit by the classifier. Similarly, the real “Green” drill bit was never considered to be the “Red” drill bit by the classifier. This means a complete lack of the most compromising (inexcusable from the practical point of view) mistakes of the k-NN algorithm. Unfortunately, sporadically (1,9 %) the real “Red” drill bit was considered to be the “Yellow” drill bit.

The worst recognizable drill bit class was the “Yellow” class. This is evidenced by the relatively low (less than 0,5) Sn and Fscore: 0,4 and 0,48 values. For comparison - for the “Red” class, these parameters were 0,95 and 0,85 respectively.

REFERENCES

Bergweiler, S. 2016. Smart Factory Systems - Fostering Cloud-based Manufacturing based on Self-Monitoring Cyber-Physical Systems. *International Journal on Advances in*

Systems and Measurements 9(1-2): 91-101. URL: http://www.iariajournals.org/systems_and_measurements/tocv9n12.html

Fix, E.; Hodges JR, J.L. 1951. *Discriminatory analysis, nonparametric discrimination, consistency properties*. Project 21-49-004. Report No. 4. USAF School of Aviation Medicine Randolph Field. Texas, USA. URL: https://apps.dtic.mil/dtic/tr/fulltext/u2/a800276.pdf

Górski, J.; Szymanowski, K.; Podziewski, P.; Śmietańska, K.; Czarniak, P.; Cyrankowski, M. 2019. Use of cutting force and vibro-acoustic signals in tool wear monitoring based on multiple regression technique for compreg milling. *Bioresources* 14(2): 3379–3388. URL: https://ojs.cnr.ncsu.edu/index.php/BioRes/article/view/BioRes_14_2_3379_Gorski_Cutting_Force_Vibro_Acoustic_Signal

Hall, P.; Park, B.U.; Samworth, R.J. 2008. Choice of neighbor order in nearest - neighbor classification. *Ann Stat* 36(5): 2135–2152. URL: https://projecteuclid.org/euclid.aos/1223908087

Iskra, P.; Hernandez, R. E. 2012. Toward a process monitoring of CNC wood router. Sensor selection and surface roughness prediction. *Wood Sci Technol* 46(1-3): 115–128. https://doi.org/10.1007/s00226-010-0378-7

Jegorowa, A.; Górski, J.; Kurek, J.; Kruk, M. 2019. Initial study on the use of support vector machine (SVM) in tool condition monitoring in chipboard drilling. *Eur J Wood Wood Prod* 77: 957-959. https://doi.org/10.1007/s00107-019-01428-5

Lemaster, R. L.; Lu, L.; Jackson, S. 2000a. The use of process monitoring techniques on a CNC wood router. Part 1. Sensor selection. *Forest Prod J* 50(7/8): 31–38. URL: https://search.proquest.com/docview/214640104/fulltextPDF/8BFC61230D8E4583PQ/1?accountid=17248

Lemaster, R. L.; Lu, L.; Jackson, S. 2000b. The use of process monitoring techniques on a CNC wood router. Part 2. Use of vibration accelerometer to monitor tool wear and workpiece quality. *Forest Prod J* 50(9): 59–64. URL: https://search.proquest.com/docview/214622388/fulltextPDF/A8F5CD37FAC14C6FPQ/1?accountid=17248

Schwab, K. 2017. *The Fourth Industrial Revolution*. Portfolio Penguin. London, United Kingdom. 192 p. URL: https://www.penguin.co.uk/books/304/304971/the-fourth-industrial-revolution/9780241300756.html

Szwajka, K.; Górski, J. 2006. Evaluation tool condition of milling wood on the basis of vibration signal. *J Phys: Conf Ser* 48: 1205–1209. https://doi.org/10.1088/1742-6596/48/1/225

Swiderski, B.; Kurek, J.; Osowski, S.; Kruk, M.; Jegorowa, A. 2017. Diagnostic system of drill condition in laminated chipboard drilling process. In *The 21st International Conference on Circuits, Systems, Communications and Computers*. *MATEC Web of Conferences* 125: 04002. https://doi.org/10.1051/matecconf/201712504002

Wagner, T.; Herrmann, C.; Thiede, S. 2017. Industry 4.0 impacts on lean production system. *Procedia CIRP* 63: 125–131. https://doi.org/10.1016/j.procir.2017.02.041

Wilkowski, J.; Górski, J. 2011. Vibro-acoustic signals as a source of information about tool wear during laminated chipboard milling. *Wood Res-Slovakia* 56(1): 57–66. URL: http://www.woodresearch.sk/wr/201101/06.pdf