

Tool Condition Monitoring for the Chipboard Drilling Process Using Automatic, Signal-based Tool State Evaluation

Bartosz Świdorski,^a Izabella Antoniuk,^a Jarosław Kurek,^{a,*} Michał Bukowski,^a Jarosław Górski,^b and Albina Jegorowa^b

An automatic approach to tool condition monitoring is presented, with the best solution achieving overall accuracy of 94.33% and 9 misclassification errors. In the wood industry, cutting tools need to be evaluated periodically. This is especially the case when drills are concerned; since when dulled, the resulting poor-quality product may generate loss for the manufacturing company, due to the need to discard it during quality control. Each tool can be classified either as useful or useless, and the second type should be exchanged as fast as possible. Manual evaluation of tools is time consuming, which results in production downtime. This problem requires a faster, automated, and precise solution for the work environment. In response to this issue, an ensemble algorithm was developed. Different signals were collected for the input data, including feed force, cutting torque, noise, vibrations, and acoustic emission. Based on those signals, a set of 152 initial features was generated, while after feature selection 19 of them were used by the classifiers. Different algorithms were tested and evaluated in terms of overall accuracy and number of errors. The best classifiers were used to prepare ensemble solution, which was able to classify the tools accurately, with very few errors between recognized classes.

DOI: 10.15376/biores.17.3.5349-5371

Keywords: Deep learning; Classifier ensemble; Drill wear classification; Laminated chipboard; Tool condition monitoring

Contact information: a: Institute of Information Technology, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland; b: Institute of Wood Sciences and Furniture, Warsaw University of Life Sciences, Nowoursynowska 159, 02-776 Warsaw, Poland;

* Corresponding author: jaroslaw_kurek@sggw.edu.pl

INTRODUCTION

Achieving high production quality and ongoing monitoring of both tools and created product is an important aspect in the wood industry. For the best possible results, there are two main approaches: tool condition monitoring (TCM) and process condition monitoring (PCM). Regarding TCM, a wide selection of problems can be considered, such as detecting the beginning and end of the cutting process, noticing catastrophic tool failure, quality control of the cutting process, *etc.* All these problems should be realized as precisely as possible and with relatively low costs (Scheffer *et al.* 2003; Botsaris and Tsanakas 2008).

The interest in automatic monitoring of cutting tools is steadily growing, due to the increased use of flexible automation during wood production process (Dimla and Lister

2000; Abu-Zahra and Yu 2003; Jemielniak *et al.* 2012). At the same time, the general complexity of this process, when combined with different noise-influencing registered signals and specific properties of the production environment, tends to increase requirements for the existing, sensor-based solutions in terms of both accuracy and automatization. Taking into account recent, high accuracy solutions (such as Ranjan *et al.* 2020 describing the tool monitoring in micro-drilling process focuses on hole quality prediction, Yu *et al.* 2021 specifically on chisel edge wear of drills, or solutions described in an overview of current methods provided by Górski 2022), it is even more important to create more automated, diagnostic methods for cutting tools used in the wood industry, especially elements such as drills.

Due to mechanical, thermal, and chemical influence of the work piece, the drill is steadily blunting during material processing. The drill wear and the resulting decrease in cutting edge properties are important issues in machining science (Dimla and Lister 2000; Lemaster *et al.* 2000) that are influenced by numerous factors such as hard element contamination in used materials, glue, friction of wood, *etc.* (Silva *et al.* 2000; Jemielniak *et al.* 2012). The most effective cutting requires a tool that is sharp and in a good state to obtain the best quality of the machined surface. According to Porankiewicz (2003), the condition of the blade (or its bluntness) is one of the most important factors determining the quality of cutting. During machining, the tool blade wears off, and the increasing surface of mutual contact causes the acceleration of physical phenomena, which deteriorates the cutting conditions and the quality of the machined surface. The loss of cutting properties occurs due to changes in the geometry of the blade, which can be caused by mechanical, thermal, electrical, and chemical factors. The tool wear process has not been fully investigated, as it is a very complex system of blade material interaction and mechanical, thermal, chemical, and other influences (Porankiewicz and Wieloch 2008). All the above factors make it difficult for the expert to estimate the time or number of drillings after which the tool will be worn out, hence the need for automatic monitoring of tool state in that aspect.

When an automatic approach is considered in similar applications, the process depends on the appropriate selection of sensors to collect various signals. Usually elements such as force, electric power, acoustic emission, vibration, or acoustic pressure are measured (Abu-Zahra and Yu 2003; Wilkowski and Górski 2011). In the case of TCM, research on metal working has shown that force sensors achieve the best results. However, these sensors (force or torque transducers) are relatively expensive and tricky to mount on the work piece or cutting tool. Thus, in the wood industry it is easier to mount vibration sensors, even though they are less accurate because of collected background noise (Jemielniak *et al.* 2012; Kuo 2000).

Some methods measure the change in physical values from the cutting zone using appropriately selected sensors (Szwajka and Zielinska-Szwajka 2008a,b,c,d; Górski *et al.* 2009). Although the sensors are adaptable to different environments, the accuracy of the measurements is not reliable. Many different factors may affect physical signals, such as selection of signals and appropriate measures of these signals related to the wear of the blade; measurement accuracy; external factors such as disturbances accompanying the signal recording (*e.g.*, noise); and method used for mounting used sensors.

After initial signals are collected, feature extraction methodology, is used to select the best possible parameters that clearly distinguish between chosen classes. In this case two sets are considered: drills that are classified as “useful”, which are still in a good shape and can be further used in the production process, and tools classified as “useless”, which

should be immediately replaced. In the second case if the drill is not exchanged quickly, it can result in poor product quality and hence lead to financial loss for the manufacturing company. While manual evaluation of the state of the drill is possible, it is also time consuming and interrupts production, possibly resulting in downtime. To avoid that problem and speed up the entire process, an automatic approach is necessary.

This work presents an automatic approach to drill condition monitoring for the furniture manufacturing process using laminated chipboard. Signals such as feed force, cutting torque, noise, vibrations, and acoustic emission are collected and used to generate features for final classification. The resulting method can distinguish between two drill classes of useful or useless. To obtain best possible classification results, different algorithms were evaluated. The original approach presented in this paper was initially based on the TreeBagger approach, and later expanded with classifier ensemble methodology. Achieved results were compared with solutions such as SVM (or a support vector machine), Naive Bayes classification, Discriminant Analysis, K Nearest Neighbours algorithm, a single classification tree as well as a deep learning approach. Since different approaches show that using groups of classifiers instead of single methods can increase the classification rate (for more details see Jegorowa *et al.* 2021), the fusion of classifiers was also tested in three different combinations: containing all initial classifiers and combining three best ones. The final result was achieved by majority voting of all included classifiers.

EXPERIMENTAL

Materials and Tools

Drilling was performed on a domestic three-layer laminated chipboard (Fig. 1). This material was chosen because it is widely used by the domestic furniture industry, and it is difficult to process due to its physical and mechanical properties. This is directly related to the multilayer structure of the material, specific to the laminated chipboard, which often results in significant differences in its density, which are impossible to predict and are likely to influence the overall drill wear rate. Standard melamine-faced particleboard (Swiss Crono Poland, U 511 SM) with thickness of 18 mm was used. To determine the density of the board a GreCon DAX device was used (Fagus-GreCon Greten GmbH & Co. KG, Alfeld, Germany). An example profile of used chipboard is presented in Fig. 2. The parameters of the material were as follows: bending modulus of rupture, 15.4 MPa; bending modulus of elasticity, 2950 MPa; and surface Brinell hardness (HB), -2.1.

The material chosen for experiments is mainly used in the manufacture of furniture fronts, countertops, and visible elements that require high quality processing. Breakage of the laminate due to cutting is not desirable. Breakages are related to the degree of wear of the cutting tool blades.

All data were acquired using a standard Buselatto JET 100 vertical machining centre (Busellato, Thiene, Italy). It is used for processing wood and wood products. For the drilling process, laminated chipboard (Kronopol U 511 SM; Swiss Krono Sp. z o. o., Żary, Poland) was used with standard FABA WP -01 (Faba SA, Baboszewo, Poland) and drills with tungsten carbide tips. The drill overall length was 70 mm, with shank length of 25 mm, flute length of 40 mm, 10 mm shank diameter, and 12 mm drill diameter. The clearance angle, rake angle, and helix angle were 15.45°, 0°, and 15°, respectively.

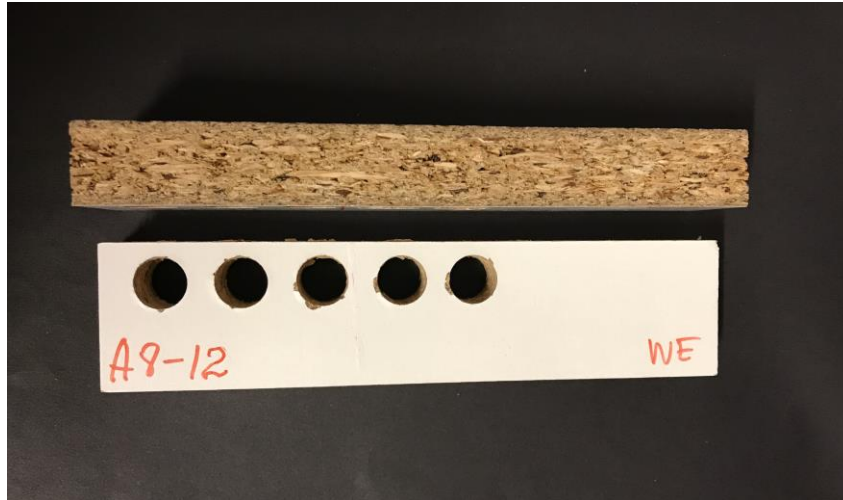


Fig. 1. Three-layer laminated chipboard (Kronopol U 511 SM), $b = 18$ mm, side section (top), and top view (bottom)

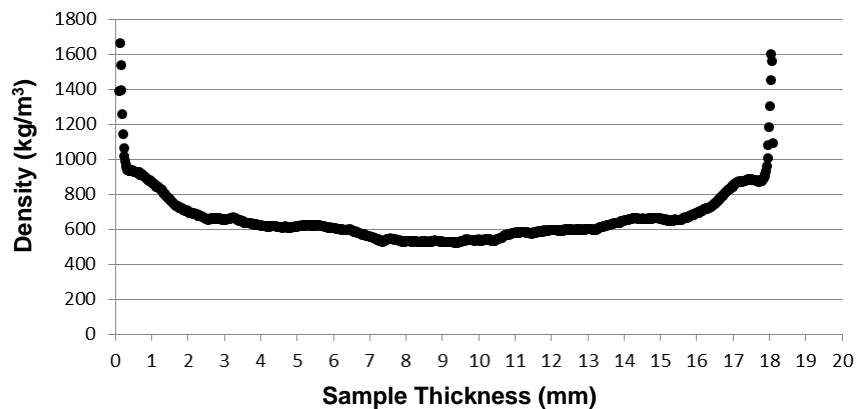


Fig. 2. Example profile of the laminated chipboard used during experiments

During the drilling process, the condition of drill bits was monitored and assessed directly, using standard workshop microscope (TM – 505; Mitutoyo, Kawasaki, Japan). Based on the measurement of the cutting edge, according to the manufacturer recommendations, when the wear exceeded 0.2 mm, the drill was considered as dull. The drill type is depicted in Fig. 3.



Fig. 3. Double edge drill FABA WP-01 with HW blades for through drilling, $\varnothing = 12$ mm

The machine centre is shown in Fig. 4. To ensure that measured signals had the best possible quality, specialized sensors were used during measurement, as follows: AE-acoustic emission measuring system (contact sensor, Kistler 8152B, amplifier, Kistler 5125B, (Winterthur, Switzerland)); V-mechanical vibration measuring system (accelerometer, Kistler 8141A, amplifier, Kistler 5127B); C-noise (sound pressure) measuring system (microphone and preamplifier, B&K 4189, amplifier, B&K NEXUS 2690, (Nærum, Denmark)); and F and M dynamometer with Kistler 9345A sensor and ICAM5073A amplifier.



Fig. 4. Busellato JET 100 standard CNC machining center

The arrangement and mounting of sensors are depicted in Fig. 5. Due to different sampling requirements, two different acquisition cards were used (NI PCI-6111 and NI PCI-6034E), where AE was measured with the faster one. The full schema of the research stand, with all key elements, are presented in Fig. 6.

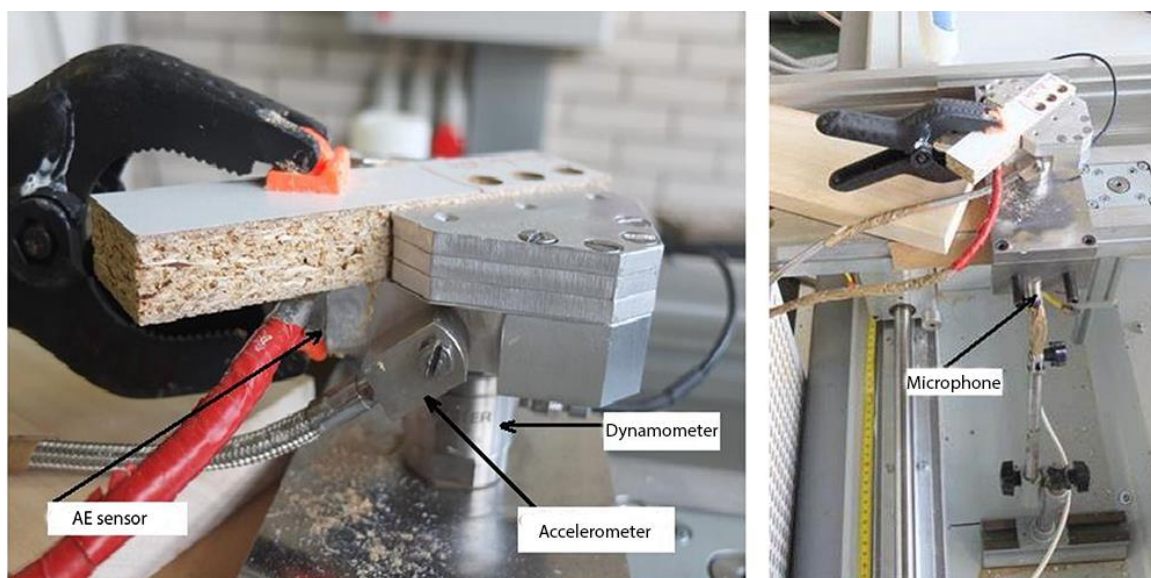


Fig. 5. Arrangement and mounting of sensors

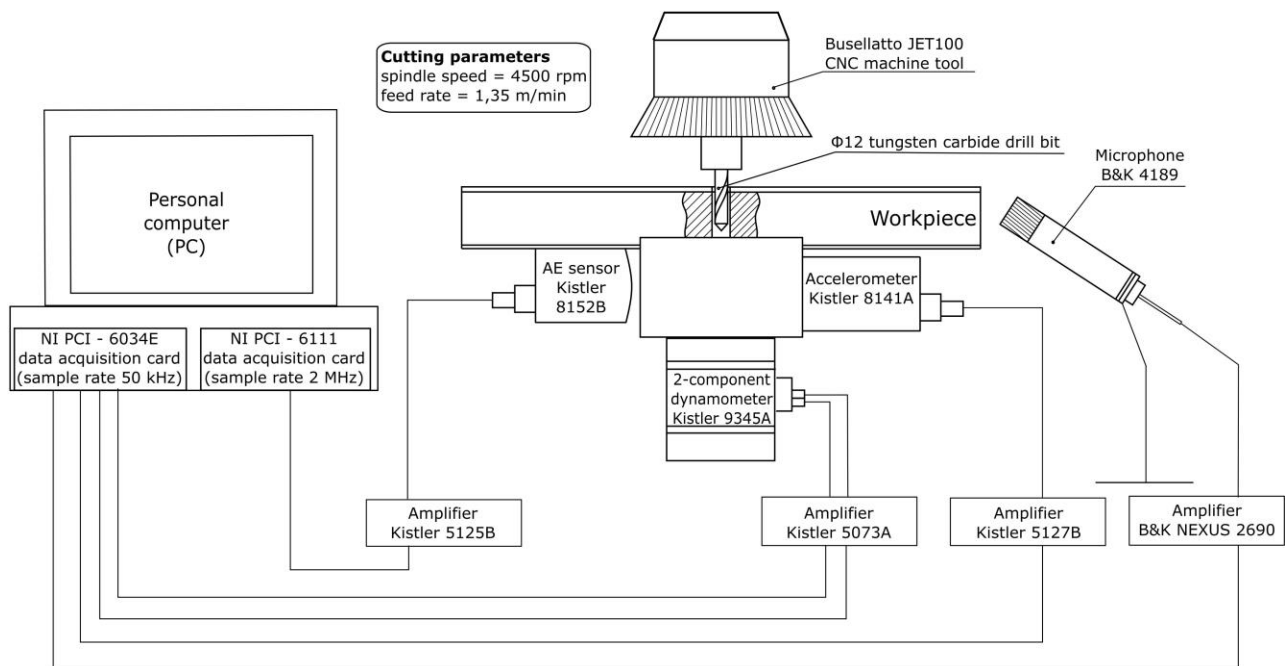


Fig. 6. The schema of full research stand with all key equipment included

Cutting Conditions

While generating features, different approaches can be used, each of them requiring specific types of data. One of the key elements in the chosen research subject is appropriate selection of collected tool parameters. In this case a set of signals was measured, to ensure possibly wide coverage of this element. The chosen elements were feed force (F), cutting torque (M), noise (C), vibration (V), and acoustic emission (AE).

The above signal set was chosen because many different features can be generated based on its components. Single measurements for F , M , C , and V contained 60,000 samples, while for AE this number was higher and amounted to 2,400,000 samples. The research process consisted of making a series of holes by each drill to acquire data. After that, each drill was subjected to blunting cycles, followed by another cycle of signal recording while drilling another series of holes. Blunting consisted of making holes in the laminated chipboard with each drill until the wear increment was at least 0.05 mm after each cycle. The wear was monitored with a microscope with simultaneous acquisition of images of the drill blades wear. The cutting conditions equalled n equal to 4500 turns per minute and u equal to 1.35 meters per minute.

The material tests were performed in accordance with CEN EN 1534 (2020) and CEN EN 310 (1994), using an Instron 3382 testing machine (Norwood, MA, USA) as well as a Brinell CV 3000LDB tester (CV Instruments, Surrey, UK) respectively.

The samples used for the tests had the dimensions of $150 \times 35 \times 18$ mm. The size of tested elements was determined based on the requirements for mounting them on the dynamometer in the platform holder.

The recording of selected signals was carried out on a PC computer using software from National Instruments, *i.e.* Lab ViewTM (National Instruments Corporation, ver. 2015 SP1, Austin, Texas, USA) environment using the NI PCI - 6034E and NI PCI - 6111 (Austin, Texas, USA) data acquisition cards. The use of two cards was justified by the

presence of signals of different frequency. AE recording required the use of a relatively high sampling frequency - 2 MHz, with the measuring window of 0.3 s. The other signals were recorded at a frequency of 50 kHz, with measuring window equal to 1.1s. The signals reached the cards through the BNC-2110 connection boxes, separately for each frequency range.

All sensors used to record the measured signals were kept the same position in relation to the workpiece and the cutting zone.

Data Set

The final database prepared for the numerical experiments was obtained using 6 drills, which were used repeatedly to generate appropriate data. During the first phase new drills were used on the laminated chipboard in the drilling process to register proper signals for all chosen sensors. This process was repeated five times, and all samples belonged to first class (tool that can be used in the production process without any risks).

After this stage single drill (no. 6) was treated as a reference and left without any blunting, while the remaining five were subjected to this process gradually, by drilling successive holes, in multiple phases. After each phase the drilling process was repeated 5 times, while measuring signals for the chosen sensor set. At this point human expert decided what the classification of the drill was (either “useful” or “useless”), and obtained samples were denoted accordingly. The reference drill was used only for 27 hole drillings, to avoid any significant blunting. It remained the only drill that was classified as “useful” through entire process. Specific drillings were organized as follows:

- For drills no. 1, 2 and 5: total of 5 trials at 3 stages when they were classified as “useful” or sharp enough with total of 15 signal registration classified as belonging to class 1 (3 stages x 5 measured signals), and 5 trials at 6 stages of “useless” or extensively worn drill state, with total of 30 signals registered (6 stages x 5 trials). Total measurement count for those drills equalled 135 (3 drills x 45 measurements).
- For drill no. 3: 4 stages during “useful” drill state, 5 trials each, with 20 signals registered (4x5) belonging to class 1, following with 4 stages when the drill was classified as “useless”, also with 5 trials each, resulting in 20 signals assigned to class 2. Total measurements done for this drill equalled to 40.
- For drill no. 4: 2 stages at “useful” state, with 5 trials each (10 signals belonging to class 1), and 7 stages with 5 trials each for “useless” state (35 signals assigned to class 2), with total of 45 measurements for this drill.
- For drill no. 6 (reference drill): total of 27 signals registered for the “useful” drill state.

In sum, a total of 247 signals was registered, with 102 of them assigned to the class 1 (“useful”) and 145 for class 2 (“useless”). During experiments, drill no. 3 deteriorated faster than the other ones, resulting in less trials for this tool. The data set prepared in such a way could be easily used during further experiments. Typical signals representing 5 physical quantities for class 1 (sharp drill) and class 2 (worn state drill) are depicted in Fig. 7. All signals are expressed in millivolts.

Feature Generation

The main problem for generating features is that not all of them will represent significant, diagnostic information, concerning tool state. To ensure that the chosen set can differentiate the drill states as accurately as possible, the first set of used methods ensured that the best possible features were chosen. The set of signals consists of 5 different elements, which can be represented as a following signal matrix, with assigned 0 or 1 flag, representing affiliation with either positive or negative class, as shown in Eq. 1.

$$Y = [y_1, y_2, y_3, y_4, y_5] = \begin{bmatrix} y_{11} & \cdots & y_{15} \\ \vdots & \ddots & \vdots \\ y_{51} & \cdots & y_{55} \end{bmatrix} \quad (1)$$

Table 1. Distribution of Samples in the Data Set

Drill No.	No. of Trials Useful (class 1)	No. of Trials Useless (class 2)
Drill 0	27	0
Drill 1	15	30
Drill 2	15	30
Drill 3	20	20
Drill 4	10	35
Drill 5	15	30
Summarize	102	145

There are two problems with initial form of used signals. Because the AE signal had higher sampling speed, this signal needs to be interpolated from an initial 2,400,000 data items to the 60,000 items that were collected for the other signals. In this case linear interpolation was used for each value in the table row, containing set of signals for single measurement.

After the interpolation, all 5 used signals can make up five-dimensional signal in time in the form shown in Eq. 2. Each signal was normalized using Eq. 3:

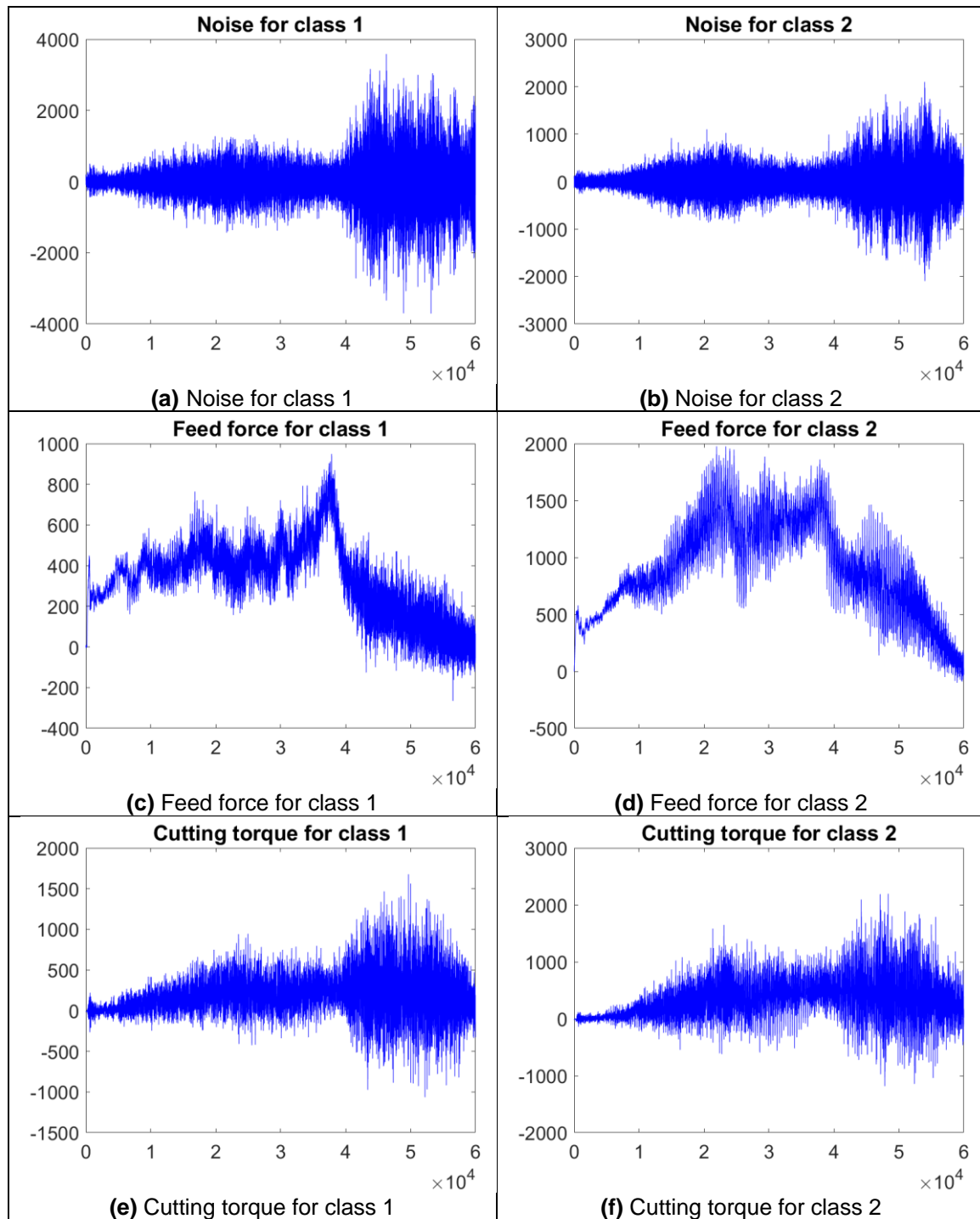
$$X = [V, C, F, M, AE] \quad (2)$$

$$X[:, i] = \frac{(X[:, i] - \text{mean}(X[:, i]))}{\text{std}(X[:, i])} \text{ for } i = 1:5 \quad (3)$$

After initial preparation of a given dataset, the actual feature extraction process was started. The given signal set can be used to generate different features, but only those that can clearly distinguish recognized classes should be used during classification process. Those elements were calculated in three different blocks, with separate requirements. The first set contains features from the vector autoregressive model, with a delay magnitude equal to 1. It can be assumed that X_t is a sample from the collected data series (5-dimensional, in time t) with size $[1, 5]$. Using those relations, the dependency presented in Eq. 4 can be modelled.

$$\begin{aligned} X_t &= X_{t-1} \cdot A + \varepsilon_t \quad \text{where:} \\ &- X_t, X_{t-1}, \varepsilon_t \text{ have size } [1 \times 5] \\ &- A \text{ has size } [5 \times 5] \end{aligned} \quad (4)$$

Using the A matrix coefficients, the first 25 features were generated. At this point, the first version of the model (also called the vector autoregression model) was prepared.



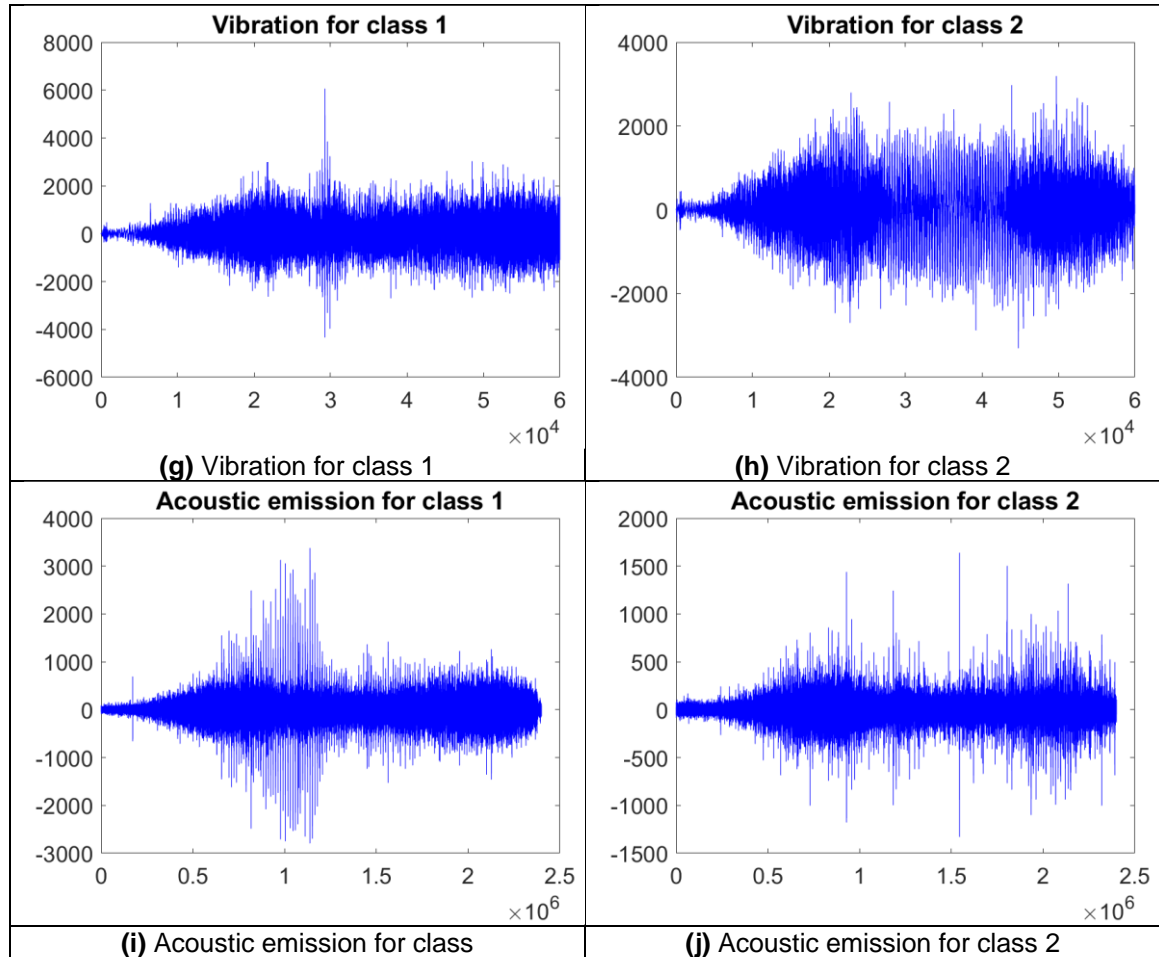


Fig. 7. The signals of (a) and (b) noise, (c) and (d) feed force, (e) and (f) cutting torque, (g) and (h) vibration and (i) and (j) acoustic emission representing the sharp state of drill (left column) and worn out state (right column); horizontal axis represents time

While it can be used to model analysed process, the system was also very defective at this point, making many mistakes. At this point errors made by initial model can be analysed and used to create second set of features. To check the overall quality of prepared model the absolute error and the forecast MAPE error are measured. For this purpose, the vectors X_t are compared to corresponding predictions $\hat{X}_t = X_{t-1} \cdot A$. For errors on individual channels: $X[:, i] - \hat{X}[:, i]$, as well as for all the channels collectively ($X(:) - \hat{X}(:)$), the following elements can be taken into consideration:

- 5 - features: total of 5 mean errors: one for each channel representing collected signal type,

$$\text{mean}(\text{abs}(X - \hat{X})) \quad (5)$$

- 1 - feature: one mean error over all 5 channels collectively:

$$\text{mean}(\text{abs}(X(:) - \hat{x}(:))) \quad (6)$$

- 6 - features: analogically, 6 median values can be denoted, one collective, and one for each of the channels,

- 6 - features: 6 standard deviation (std) errors,
- 6 - features: 6 interquartile range:

$$(\text{quartile3} - \text{quartile1}) \quad (7)$$

- 6 - features: 6 positional coefficients of variation:

$$\frac{(\text{quartile3} - \text{quartile1})}{\text{median}} \quad (8)$$

- 6 - features: 6 classical coefficients of variation:

$$\left(\frac{\text{std}}{\text{mean}} \right) \quad (9)$$

- 6 - features: 6 skew coefficient,
- 6 - features: 6, 3rd order cumulates
- 6 - features: 6, 4th order cumulates
- 6 - features: 6, 5th order cumulates

There was a total of 60 features at the level of absolute error, which were used to evaluate and improve initial model. The same set of parameters can then be calculated for the MAPE error (Eq. 10), which results in additional 60 elements that can be used for analysis and 120 features total.

$$MAPE = \text{mean} \left(\text{abs} \left(\frac{(X - \hat{X})}{X} \right) \right) \quad (10)$$

Finally, third block of features contains matrix conditioning index: $X^T X$, which can be used to indicate variables collinearity and indicators based on the data reconstruction error by the principal component analysis method (or PCA). The main idea behind this reasoning is that if data is “healthy”, after compressing and decompressing it, the reconstruction error should be lower than in case of “unhealthy” data. Using above index in overall evaluation, following features can be distinguished:

- 1 - feature: Condition Index (CI),
- 5 - features: PCA – at this point compression with 2 components is used, and the error of reconstructing each of the 5 initial columns of matrix X is calculated as:

$$\text{mean} \left(\frac{\text{abs}(X[:,i] - \hat{X}[:,i])}{X[:,i]} \right) \quad (11)$$

- 1 - feature: similarly defined PCA, for the hole matrix:

$$\text{mean} \left(\frac{\text{abs}(X - \hat{X})}{X} \right) \quad (12)$$

After calculating all of the above errors, a total of 152 features was generated (25 + 120 + 7 in the following sets), where each of them focuses on different aspects of the prepared model. The first set describes the general model parameters for the AR (25 initial factors). The second feature block defines in various ways places, where the initial model is not fitted enough. The final 7 features checks the collinearity for the parameters and the reconstruction error after compression.

After obtaining the final features set, the next step is to select those, which will offer

the best distinction between the recognized classes. The feature selection is done per set (one of initial three blocks of features defined earlier) and within each set each feature is evaluated using AUC statistic (Area Under roc Curve). Initially all features, for which the value of this statistic exceeds 0.7 are selected. If the number of parameters selected in such a way is greater than ten, then features are narrowed down to this number, using best representatives. In the result of the feature selection process, the following elements were chosen:

- features 7 and 16 from block containing 25 elements,
- all elements from the block containing 7 features,
- 10 elements from the block containing 120 features.

In this way 19 independent variables were selected, and these were later used during the classification process.

Discriminant Analysis

Discriminant analysis was used to classify the chosen drill classes. This method is used in statistics and other fields to find a combination of different features that allows for such classification of single example, which in can be clearly assigned to one of the classes. The combination of features obtained in such a way is often treated as a step before the actual classification, as a method to reduce number of dimensions, but it can also be used for the main classification. It was decided to incorporate this method to achieve a good base value for the classification algorithm presented in this paper.

Classification Tree

Because the solution presented in this paper is based on a set of classification trees, it was decided, for the comparison purposes, that a single instance of this classifier would also be used.

Classification trees, also referred to as decision trees, are predictive models used to represent the road from the observations of chosen problem (including various features and variables associated with it), to the final classification of given example to one of the existing classes. The name of this classifier is not accidental, as the structure of the classifier is often represented as a tree, with branches representing the decision process, nodes showing the test that needs to be performed at each decision point and leaves, showing the final classification.

The actual tree is prepared by choosing appropriate split points for the original data set, starting with the root node. For each of such nodes, the split condition is assigned (based on one of classification features). The goal is to achieve the best possible division for each split condition, while the process is repeated, until either all examples in leaves of current node can be clearly assigned one specific class, or predefined complexity is exceeded. For additional reference see (Breiman *et al.* 1984; Loh and Shih 1997; Coppersmith *et al.* 1999; Loh 2002; Tiitta *et al.* 2020).

Naive Bayes Classifier

One of the most commonly used classifiers is the Naive Bayes Classifier. It is a probabilistic model for machine learning problems, based on the Bayes Theorem, as shown in Eq. 13.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (13)$$

The general assumption in this classifier is that two elements are considered: hypothesis A and evidence B. For this approach to work, the features used for the classifier need to be independent (meaning that the presence of one feature does not affect the other, existing features, which is also the reason why this classifier is denoted as naive), and in this case the probability of A happening with the B which already occurred, can be found. For the classification purposes, the element with highest possibility is then chosen.

K-Nearest Neighbours Classifier

Another algorithm chosen for comparison focuses on local function approximation. An approach called the K- Nearest Neighbours algorithm (Fix and Hodges 1951; Altman 1992; Tiitta *et al.* 2020) can be used both in classification and regression problems. This approach is based on the surroundings of each given examples.

The algorithm consists of few phases. First, using training examples (where each of them contains set of features in multidimensional space and class label associated with this example), consists of storing those elements. In the next stage, the classification phase happens. In this case the k (which is a user specified value) is used to evaluate exact number of closest neighbours to the current example. The distances can be calculated using various metrics, but they are based on the values stored in each example feature vector. The tested example is then assigned the class most commonly occurring in the neighbourhood specified by the k value.

With the kNN algorithm, the main problem lies with the specification of appropriate value for the k parameter. This value is heavily dependent on the used data. In general, the higher values will reduce classification noise, but at the same time this will result in a less clear class division. Since this can be heavily affected by the presence of features that are either noisy or irrelevant to the current problem, the usage of kNN algorithm is usually preceded by extensive data preparation. In the case of two class classification, this value should be an odd number to avoid problems with tied votes.

Support Vector Machine

Another algorithm, that can be used both for regression and classification problems is Support Vector Machine (or SVM). The objective in this case is to find the hyperplane in a n-dimensional space (where n is the number of features taken into account during classification process), which offers best classification for given data points. Since using given data set different hyperplanes can be found, the objective here is to maximize the margin between separate classes (where the margin can be defined as minimal distance between the data points belonging to different classes). Support vectors in this context are the data points that are closest to considered hyperplanes. By checking the distance each support vector in relation to different hyperplanes the most optimal margin can be found (for additional information see Tiitta *et al.* (2020).

Tree Bagger

While decision trees can be quite good classifiers, they also often have a problem with overfitting. Due to changing conditions in the production environment, the accuracy and further adjustment of final solution can pose significant problems. The approach used in this paper is based on the tree bagging methodology.

In general Bagging (or Bootstrap Aggregation) is used to reduce the variance for the decision tree approach. The idea is to create randomly chosen subsets of data from the training samples. Each of such subsets will be used to train different decision tree, which results in creation of set of different models that can be used for classification. Obtaining final classification by using averaged predictions from set of trees in general will be more accurate, than using single decision tree.

Deep Learning

Another solution was based on deep learning. Deep neural networks can be trained to solve regression or classification tasks using non-image or nonsequence data. In this case the Matlab implementation of "trainNetwork" method was used. For Matlab R2020b and later versions, the "featureInputLayer" method can be used when data set consists of set of numeric scalars, representing different features (meaning that the data used does not include any spatial or time dimensions). The graphical representation of the network structure is represented in Fig. 8, while the detailed outline of all network layers is presented in Fig. 9.

The network contained following layers:

- featureInputLayer(numFeatures,'Normalization', 'zscore')
- fullyConnectedLayer(150)
- batchNormalizationLayer
- reluLayer
- fullyConnectedLayer(150)
- batchNormalizationLayer
- reluLayer
- fullyConnectedLayer(numClasses)
- softmaxLayer
- classificationLayer

For the network settings, the mini Batch size was set to 60, while training options included using 'adam' optimizer and maximal epochs were set at 130.

Ensemble of Classifiers

In previous work, fusions of classifiers (ensembles is another term that is often used to describe such solutions) were used to increase overall classification accuracy. A few examples of this methodology are shown in Porankiewicz *et al.* (2008). In case of ensemble methods, instead of a single model, a set of different methods is used and combined in order to improve results of the entire classifier. Such an approach has been used in several machine learning competitions. Because this approach is heavily dependent on the individual quality of each classifier, two different combinations were proposed and checked during the research presented in this paper: (1) ensemble of all tested classifiers; and (2) fusion containing three best classifiers from the initially checked set (TreeBager, SVM and K-NN). For the ensemble approach, the final classification results were obtained by majority voting of all involved classifiers. For this method each of the classifiers predicts the final class of given example, and the final result is obtained by choosing a class with the highest vote count.

RESULTS AND DISCUSSION

To validate each classifier, separate tests were formed on each of the initial data sets (six in total). The summary information is presented in Table 2. According to the expert suggestions in terms of machine learning algorithms and their requirements, the final tests were performed on 5 data sets (including the first, reference drill), while a sixth set was used as a test set, to evaluate each solution.

Table 2. Summary of Information about the Sets Participating in Numerical Experiments

Test No.	Training Set	Test Set
1	220	27
2	202	45
3	202	45
4	207	40
5	202	45
6	202	45

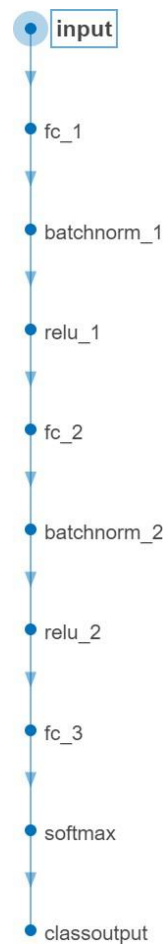


Fig. 8. Structure of Deep Learning network

The first set of experiments evaluated each solution individually. For each of the original methods (including discriminant analysis, classification tree, naive Bayes classifier, K-nearest neighbours, SVM, TreeBagger, and deep learning approaches), both

their accuracy, as well as error rates between different classes were evaluated. Some of the used methods required prior determination of various hyperparameters. In that aspect, the following operations were made:

- The number of neighbours for the K-NN algorithm was searched from the range of (3,300), with step equal to 1, with best value chosen.
- For the TreeBagger approach, the number of trees was searched in the range between 10 and 500, with step equal to 1.
- To find the C and Gamma hyperparameters for the used SVM algorithm, the following ranges of parameters were searched: C from 100 to 10000 with step 100 and Gamma from 0.001 to 0.5 with step 0.001.

Accuracy scores for each individual algorithm are presented in Tables 3, 4, 5, and 6. Table 3 shows accuracy results for the discriminant analysis, classification tree, naive Bayes classification, and deep learning algorithms; those solutions did not require calculation of any additional hyperparameters. The remaining tables show results for the algorithms that had those additional values, and apart from the general accuracy score for each data set, they also show the number of nearest neighbours used in the case of k-NN algorithm (Table 4), C and Gamma values for the SVM approach (Table 5) and number of generated trees for the TreeBagger solution (Table 6).

While the general accuracy results are important, one additional parameter that needs to be considered is the amount of errors, when a useless tool is mistaken for a useful one. This situation generates losses due to poor product quality. To better evaluate each algorithm in that aspect, the confusion matrices containing specific numbers and percentage values for each error were generated. Corresponding results for the initial set of algorithms are presented in Figs. 10, 11, 12, 13, 14, 15, and 16.

	Name	Type	Activations	Learnables
1	input 19 features with 'zscore' normalization	Feature Input	19	-
2	fc_1 150 fully connected layer	Fully Connected	150	Weights 150×19 Bias 150×1
3	batchnorm_1 Batch normalization with 150 channels	Batch Normalization	150	Offset 150×1 Scale 150×1
4	relu_1 ReLU	ReLU	150	-
5	fc_2 150 fully connected layer	Fully Connected	150	Weights 150×150 Bias 150×1
6	batchnorm_2 Batch normalization with 150 channels	Batch Normalization	150	Offset 150×1 Scale 150×1
7	relu_2 ReLU	ReLU	150	-
8	fc_3 2 fully connected layer	Fully Connected	2	Weights 2×150 Bias 2×1
9	softmax softmax	Softmax	2	-
10	classoutput crossentropyex with classes '1' and '2'	Classification Output	2	-

Fig. 9. Detailed outline of Deep Learning network structure

Table 3. Accuracy Results for Algorithms from the Initial Set that Didn't Require Obtaining Additional Hyperparameters

Test No.	DA	CT	NBC	DL
1	100.00%	100.00%	100.00%	100.00%
2	95.56%	93.33%	93.33%	97.78%
3	93.33%	95.56%	91.11%	91.11%
4	80.00%	87.50%	82.50%	95.00%
5	93.33%	93.33%	97.78%	97.78%
6	86.67%	75.56%	84.44%	95.56%

Discriminant analysis (DA), classification tree (CT), Naive Bayes classifier (NBC) as well as deep learning approach (DL)

For the TreeBagger approach, the graph showing the overall classification error frequency depending on number of used trees was prepared, and it is presented at Fig. 17.

After analyzing the initial results, especially when error rates are considered, the algorithms with best results were solutions using SVM (5 misclassifications of useless class as useful), TreeBagger (4 errors), and deep learning (also 4 errors). All of those solutions also achieved relatively high overall accuracy results, with lowest equal to 82.50% for the SVM solution in data set 4. All other accuracy results were higher or equal to 90%, which is a relatively good result (Jegorowa *et al.* 2021).

Table 4. Accuracy Results for All Training Data Sets for k- Nearest Neighbor

Test No.	No. of nearest neighbors	Accuracy
1	3	100.00%
2	8	91.11%
3	24	93.33%
4	8	80.00%
5	93	100.00%
6	170	88.89%

Table 5. Accuracy Results for All Training Data Sets for SVM

Test No.	C	Gamma	Accuracy
1	100	0.001	100.00%
2	1100	0.473	100.00%
3	600	0.232	95.56%
4	900	0.039	82.50%
5	300	0.029	97.78%
6	9600	0.038	91.11%

Table 6. Accuracy Results for All Training Data Sets for TreeBagger

Test No.	No. of Trees	Accuracy
1	10	100.00%
2	10	97.78%
3	10	97.78%
4	20	90.00%
5	50	100.0 %
6	50	91.11%

In the next step, the classifier ensembles were tested, with fusion of all used classifiers as a base value for comparison. In this type of approach, the quality of each individual classifier can directly affect the final result (since all classifiers are voting equally, a greater amount of high-quality models can result in higher accuracy, with exactly the opposite situation in case of using inferior solutions). Accuracy results for both classifiers are presented in Table 7. Similarly, as with individual classifiers, also in the case of ensembles, the confusion matrices presenting the error rates for each misclassification types were generated. Results are presented in Figs. 18 and 19. For further reference, average accuracy values for each of the classifiers and classifier ensembles are presented in Table 8.

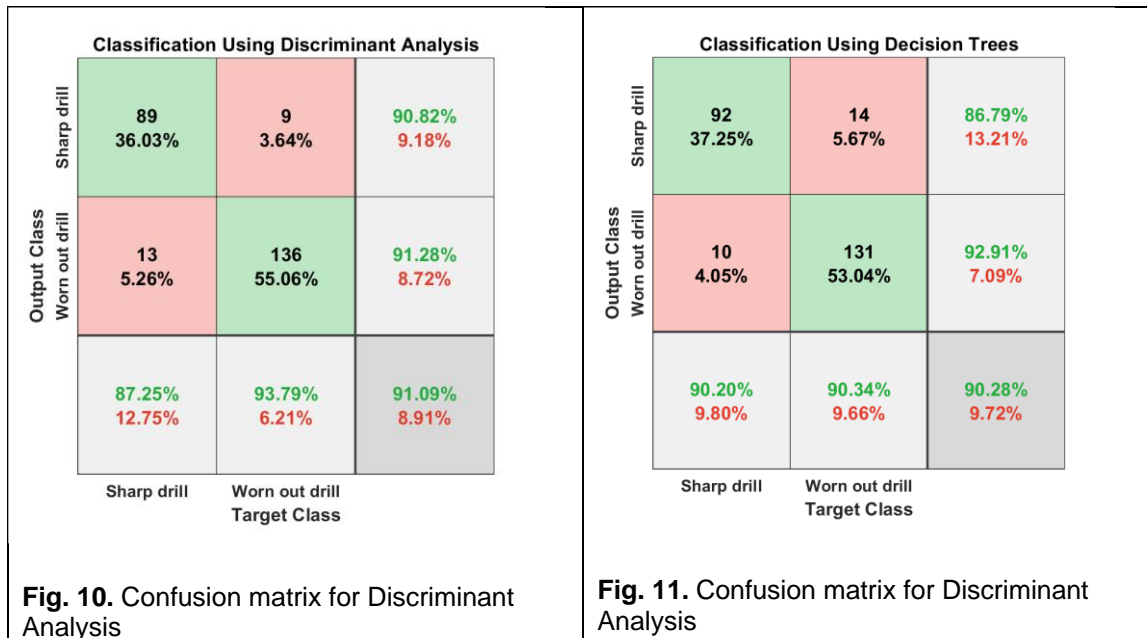


Fig. 10. Confusion matrix for Discriminant Analysis

Fig. 11. Confusion matrix for Decision Trees

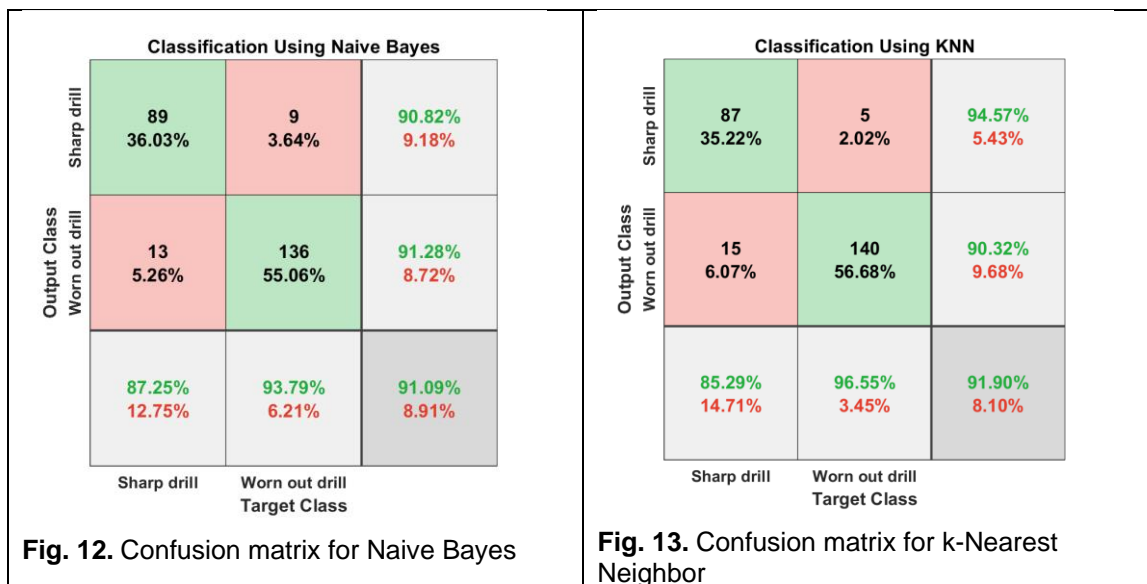


Fig. 12. Confusion matrix for Naive Bayes

Fig. 13. Confusion matrix for k-Nearest Neighbor

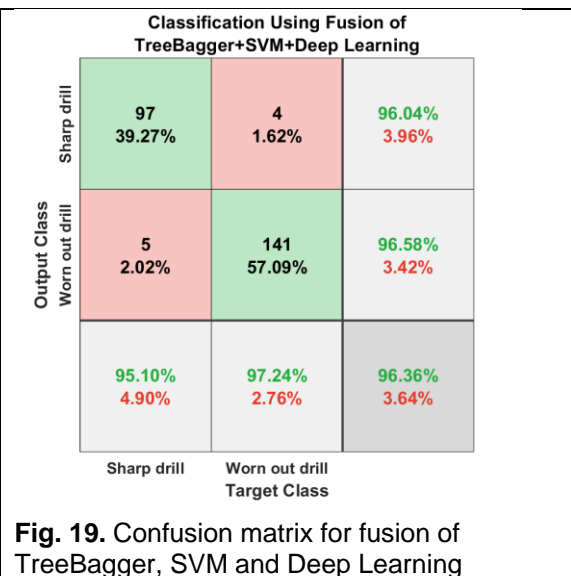
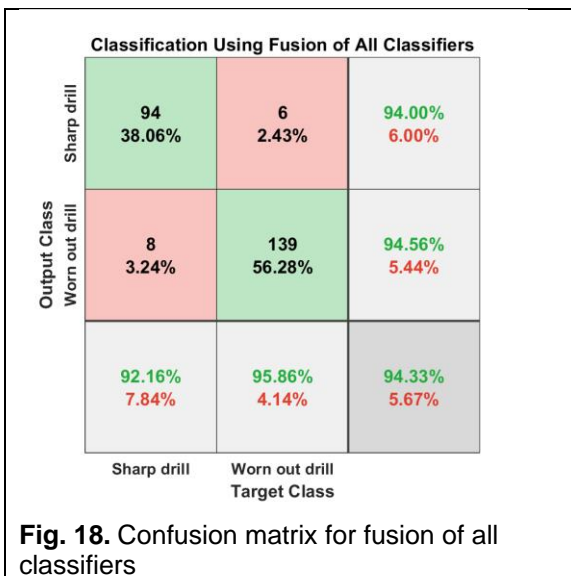
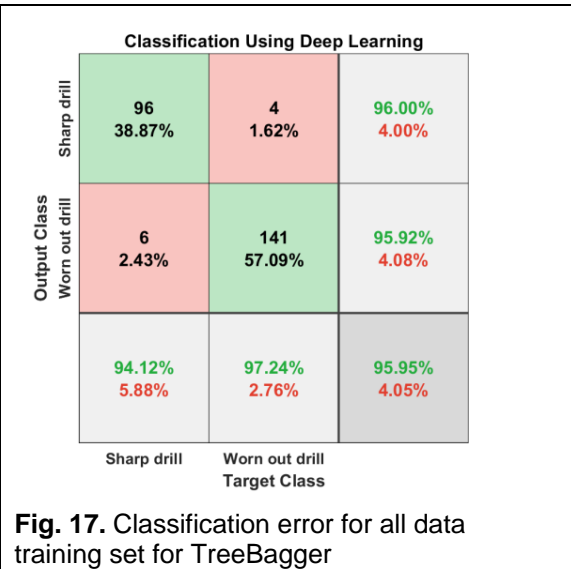
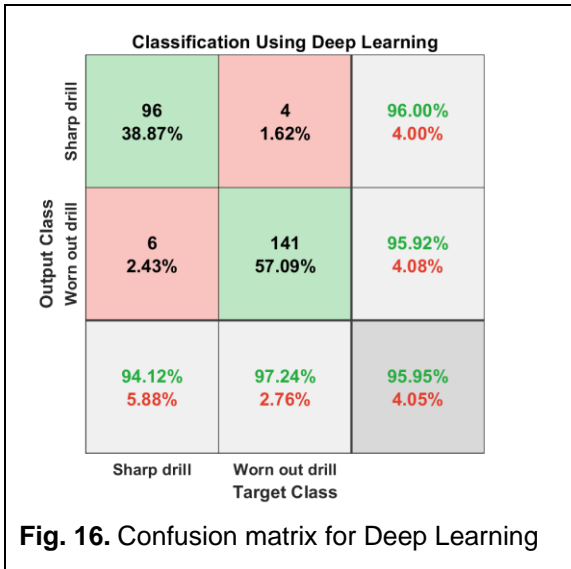
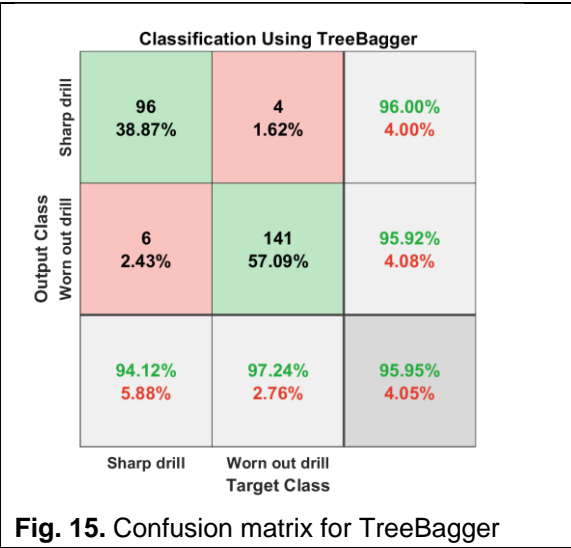
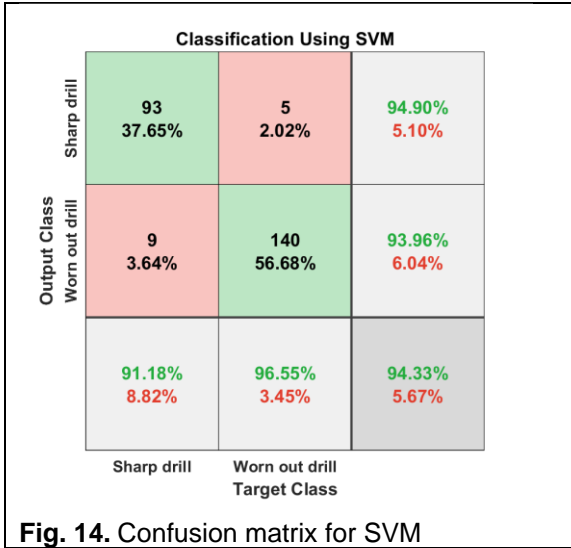


Table 7. Accuracy Results for the Classifier Ensembles, with Final Classifiers Containing All Algorithms Used in Previous Step (All) and Fusion of Three Best Classifiers (SVM+TB+DL)

Test No.	All	SVM+TB+DL
1	100.00 %	100.00%
2	100.00 %	97.78%
3	95.56 %	97.78%
4	85.00 %	92.50%
5	100.00 %	100.00%
6	86.67 %	91.11%

Both final classifier ensembles, containing all and three best classifiers for the initial set, were able to achieve more than satisfactory results. With average accuracy equal to 94.33% for all classifiers combined and 96.36% for the three-model classifier, and minimal accuracy for the used data sets remaining over 90% for the second solution (91.11%), the prepared algorithm is precise, and allows for accurate classification. In case of errors rate, the three-model classifier made only a few mistakes, wrongly classifying 4 tools as useful, and 5 as useless. The misclassification rate for the solution using an ensemble of all classifiers was higher in that aspect, with 6 errors with useless tool classified as useful, and 8 mistakes for the opposite classification. Taking into account the complexity of solution containing all classifiers, the approach with three best algorithms was judged to be better in terms of accuracy, error rate, and computational cost.

Overall, the prepared solution obtained more than satisfactory results, with precise classification and a low number of errors. Thus, the solution could be further used in an actual production environment. For future work, the data needs to be expanded, and the developed methodologies should be evaluated in different work conditions to determine the method adaptability.

Table 8. Comparison of All Tested Classifiers, Using Average Accuracy for All Training Data Sets

Classifier	Accuracy
TreeBagger	95.95 %
SVM	94.33 %
Naive Bayes	91.09 %
Discriminant Analysis	91.09 %
Classification Tree	90.28 %
Nearest Neighbor	91.90 %
Deep Learning	95.95 %
Fusion of all classifiers	94.33 %
Fusion of SVM+TreeBagger+Deep Learning	96.36 %

CONCLUSIONS

1. The work presented in this paper proposes an automatic solution for the drill classification task. The solution is based on a set of collected signals (feed force, cutting torque, noise, vibrations and acoustic emission) and uses ensemble of best classifiers from the evaluated set.

2. As the final classifiers, two ensembles were created: one containing all of the evaluated models, and one using the three best performing models (SVM, TreeBagger and Deep Learning approach).
3. For the given task, the solution containing three best classifiers was better, both in case of overall accuracy (96.36%, when compared to 94.33%) and total number of classification errors (9 when compared to 14).
4. Both final classifier ensembles allow for accurate classification of the drill state, and the chosen signal set allows for easy usage during production process, without any downtime and meet initial requirements chosen for such solution.

REFERENCES CITED

- Abu-Zahra, N., and Yu, G. (2003). "Gradual wear monitoring of turning inserts using wavelet analysis of ultrasound waves," *International Journal of Machine Tools and Manufacture* 43(4), 337-343. DOI: 10.1016/S0890-6955(02)00274-2
- Altman, N. S. (1992). "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician* 46(3), 175-185.
- Botsaris, P. N., and Tsanakas, J. A. (2008). "State-of-the-art in methods applied to tool condition monitoring (tcm) in unmanned machining operations: A review," in: *Proceedings of the International Conference of COMADEM*, Prague, Czech Republic, pp. 73-87. DOI: 10.13140/2.1.4086.7840
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*, CRC Press, Boca Raton, FL, USA.
- Coppersmith, D., Hong, S. J., and Hosking, J. R. (1999). "Partitioning nominal attributes in decision trees," *Data Mining and Knowledge Discovery* 3(2), 197-217. DOI: 10.1023/A:1009869804967.
- CEN EN 310 (1994). "Wood-based panels: Determination of modulus of elasticity in bending and of bending strength," European Committee for Standardization, Brussels, Belgium.
- CEN EN 1534 (2020). "Wood flooring and parquet - Determination of resistance to indentation - Test method," European Committee for Standardization, Brussels, Belgium.
- Dimla, Sr, D. E., and Lister, P. M. (2000). "On-line metal cutting tool condition monitoring. I: Force and vibration analyses," *International Journal of Machine Tools and Manufacture* 40(5), 739-768. DOI: 10.1016/S0890-6955(99)00084-X
- Fix, E., and Hodges, J. L. (1951). "Nonparametric discrimination: Consistency properties," Randolph Field, Texas, Project, 21-49.
- Górski, J., Wilkowski J., Czarniak P. (2009). "Introduction to automatic supervision of wood machining system, Wood machining and processing – product quality and waste characteristic," *WULS-SGGW Press*, 5-26.
- Górski, J. (2022). "The review of new scientific developments in drilling in wood-based panels with particular emphasis on the latest research trends in drill condition monitoring," *Forests* 13(2), 242. DOI: 10.3390/f13020242

- Jegorowa, A., Kurek, J., Antoniuk, I., Dołowa, W., Bukowski, M., and Czarniak, P. (2021). "Deep learning methods for drill wear classification based on images of holes drilled in melamine faced chipboard," *Wood Science and Technology* 55(1), 271-293. DOI: 10.1007/s00226-020-01245-7
- Jemielniak, K., Urbański, T., Kossakowska, J., and Bombiński, S. (2012). "Tool condition monitoring based on numerous signal features," *International J. Advanced Manufacturing Technology* 59(1-4), 73-81. DOI: 10.1007/s00170-011-3504-2
- Kuo, R. J. (2000). "Multi-sensor integration for on-line tool wear estimation through artificial neural networks and fuzzy neural network," *Engineering Applications of Artificial Intelligence* 13(3), 249-261. DOI: 10.1016/S0952-1976(00)00008-7
- Lemaster, R. L., Lu, L., and Jackson, S. (2000). "The use of process monitoring techniques on a CNC wood router. Part 1. Sensor selection," *Forest Products Journal* 50(7/8), 31-38.
- Loh, W. Y., and Shih, Y. S. (1997). "Split selection methods for classification trees," *Statistica Sinica* 815-840.
- Loh, W. Y. (2002). "Regression trees with unbiased variable selection and interaction detection," *Statistica Sinica* 361-386.
- Porankiewicz, B. (2003). "Tępienie się ostrzy i jakość przedmiotu obrabianego w skrawaniu płyt wiórowych," *Roczniki Akademii Rolniczej w Poznaniu. Rozprawy Naukowe* (241), 1-169.
- Porankiewicz, B., and Wieloch, G. (2008). "Drill wear during the boring of particle board: A multi-factor analysis including effects of mineral contaminants," *BioResources* 3(2), 425-436. DOI: 10.15376/biores.3.2.425-436
- Ranjan, J., Patra, K., Szalay, T., Mia, M., Gupta, M. K., Song, Q., Krolczyk, G., Chudy, R., Pashnyov, V. A., and Pimenov, D. Y. (2020). "Artificial intelligence-based hole quality prediction in micro-drilling using multiple sensors," *Sensors* 20(3), article no. 885. DOI: 10.3390/s20030885
- Scheffer, C., Kratz, H., Heyns, P., and Klocke, F. (2003). "Development of a tool wear-monitoring system for hard turning," *International Journal of Machine Tools and Manufacture* 43(10), 973-985. DOI: 10.1016/S0890-6955(03)00110-X
- Silva, R. G., Baker, K. J., Wilcox, S. J., and Reuben, R. L. (2000). "The adaptability of a tool wear monitoring system under changing cutting conditions," *Mechanical Systems and Signal Processing* 14(2), 287-298. DOI: 10.1006/mssp.1999.1286
- Szwajka, K., and Zielinska-Szwajka J. (2008a). "Sensor signals for tool-wear monitoring in wood cutting operations-A review of methods, Part 1," *Annals of Warsaw University of Life Sciences-SGGW. Forestry and Wood Technology*, 66.
- Szwajka K., and Zielinska-Szwajka J. (2008b). "Sensor signals for tool-wear monitoring in wood cutting operations - A review of methods, Part 2," *Annals of Warsaw University of Life Sciences-SGGW. Forestry and Wood Technology*, 66.
- Szwajka K., and Zielinska-Szwajka J. (2008c). "Sensor signals for tool-wear monitoring in wood cutting operations-a review of methods, Part 3," *Annals of Warsaw University of Life Sciences-SGGW. Forestry and Wood Technology*, 66.
- Szwajka K., and Zielinska-Szwajka J. (2008d). "Sensor signals for tool-wear monitoring in wood cutting operations-a review of methods, Part 4," *Annals of Warsaw University of Life Sciences-SGGW. Forestry and Wood Technology*, 66.
- Tiitta, M., Tiitta, V., Heikkinen, J., Lappalainen, R., and Tomppo, L. (2020). "Classification of wood chips using electrical impedance spectroscopy and machine learning," *Sensors* 20(4), article no. 1076. DOI: 10.3390/s20041076.

- Wilkowski, J., and Górski, J. (2011). “Vibro-acoustic signals as a source of information about tool wear during laminated chipboard milling,” *Wood Research* 56(1), 57-66.
- Yu, J., Cheng, X., Lu, L., and Wu, B. (2021). “A machine vision method for measurement of machining tool wear,” *Measurement* 182, article no. 109683. DOI: 10.1016/j.measurement.2021.109683.

Article submitted: April 25, 2021; Peer review completed: September 11, 2021; Revised version received: June 10, 2022; Accepted: July 24, 2022; Published: July 29, 2022.
DOI: 10.15376/biores.17.3.5349-5371