Improving the accuracy of machine-learning models with data from machine test repetitions

Andres Bustillo[1,a], Roberto Reis [2,b], Alisson Rocha Machado[2,3,c], Danil Yurievich Pimenov[4,d*]

[1] *Department of Civil Engineering, Universidad de Burgos, Avda Cantabria s/n, 09006, Burgos, Spain*
[2] *Federal University of Uberlandia, School of Mechanical Engineering, Av. João Naves de Ávila, 2121, Bloco 1M, 38400-902, Uberlândia, MG, Brazil*
[3] *Graduate Program in Mechanical Engineering, Pontifícia Universidade Católica do Paraná – PUC-PR, R. Imaculada Conceição, 1155, Bairro Prado Velho, CEP 80215-901, Curitiba, PR, Brazil*
[4] *Department of Automated Mechanical Engineering, South Ural State University, Lenin Prosp. 76, 454080, Chelyabinsk, Russia*

[*] *Corresponding author; mail: a – abustillo@ubu.es; https://orcid.org/0000-0003-2855-7532; b – rr3d_@hotmail.com; c – alissonr.rocha@pucpr.br / alissonr.machado@gmail.com; https://orcid.org/0000-0002-5388-2954; d – danil_u@rambler.ru; https://orcid.org/0000-0002-5568-8928;*

Abstract: The modelling of machining processes by means of machine-learning algorithms is still based on principles that are especially adapted to mechanical approaches, in which very few inputs are varied with little repetition of experimental conditions. These principles might not be ideal to achieve accurate machine-learning models and they are certainly not aligned with the practicalities of industrial machining in factories. In this research the effect of a new strategy to improve machine-learning model accuracy is studied: experimental repetition. Tool-life prediction in the face-turning operations of AISI 1045 steel discs, depending on different cooling systems and tool geometries, is selected as a case study. Both the side rake and the relief angles of HSS tools are optimized using the Brandsma facing test under dry, MQL, and flooding conditions. Different machine-learning algorithms, such as regression trees, kNNs, artificial neural networks, and ensembles (bagging and Random Forest) are tested. On the one hand, the results of the study showed that artificial neural networks of Radial Basis Functions presented the highest model accuracy (11.4 mm RMSE), but required a very sensitive and complex tuning process. On the other hand, they demonstrated that ensembles, especially Random Forest, provided models with accuracy in the same range, but with no tuning procedure (12.8 mm RMSE)**.** Secondly, the effect of an increased dataset size, by means of experimental repetition, is evaluated and compared with traditional experimental modelling that used average values. The results showed that some machine-learning techniques, including both ensemble types, significantly improved their accuracy with this strategy**,** by up to 23%. The results therefore suggested that the use of raw experimental data, rather than their averaged values, can achieve machine-learning models of higher accuracy for tool-wear processes.

Keywords: machine learning; artificial intelligence; industry 4.0; ensembles; machining; wear; dataset size; Brandsma facing tests

1 Introduction

Metal-alloy machining processes are of well-known industrial interest. These processes are basic in the manufacture of many industrial tools, such as machine-tools, dies, molds, aerospace components, etc. High levels of industrial interest have not, however, made the optimization of those processes an easy task, due to the number of variables that can influence them, mainly in non-linear ways (Davim 2008). Each quality parameter of the workpiece, such as surface quality, surface geometry, and wear resistance will depend on dozens of machining-process variables (Benardos and Vosniakos 2002). Some of the variables are defined by the process engineer (e.g., cutting speed, feed rate, depth of cut, etc.), but others are not so easily controlled and evaluated (e.g., tool wear, workpiece unevenness, and quality, etc.). With no precise knowledge of all these process inputs, the desired quality of the workpiece cannot be assured before the cutting process takes place. Therefore, under industrial conditions, the use of trial and error methods and the expertise of the process engineer and the machine operator have become the success factors for meeting all the quality requirements of the workpiece (Sanchez-Egea and Lopez de Lacalle 2018).

Although these machining-process limitations may be controlled, immense research effort has gone into the development of reliable prediction models for the main cutting processes. The current approach towards building reliable prediction models is based on the mechanical description of the cutting process. For example, Martellotti presented the first mathematical model for peripheral milling in 1941 (Martellotti 1941). The mechanical approach limits the process inputs that are at the same time varied. The ideal condition might be the one where only one input is varied, and the output is then measured (e.g., either cutting speed or surface roughness). However, this methodology would require thousands of experiments, as machining processes depend on many inputs. The real number of experiments that may be required will be even higher, because the influence of the inputs on any one output (e.g., surface roughness) may not be

independent (Nalbant *et al*. 2009). This requirement for extensive experimental tests should be balanced with the budgetary limitations of the laboratory, due to the high costs of any machining experiment. Therefore, the use of optimized experimental design for this approach, such as Taguchi design, is a basic requirement. Taguchi experimental design is based on a limited number of input parameters and a limited number of levels for those parameters (Antony and Kaye 2000). Usually the cutting experiments are performed once; only in very limited cases are the experiments repeated, at all times with the objective of assuring the stability of the output values and low dispersion levels. If some repetitions of the same cutting conditions are performed, the averaged value of those experiments are fed into the mechanical model.

However, although the mechanical models have successfully provided useful knowledge on cutting processes, their industrial implementation has always presented difficulties, due to three main industrial constraints (Zheng *et al*. 2018). First, most of the input parameters of the cutting processes under industrial conditions will vary without their quantitative evaluation (e.g., tool wear). Second, if a certain combination of factory cutting conditions provides the required workpiece quality and production rate, this combination will not be changed; therefore, many repeated experimental conditions in industrial datasets may be expected, instead of many different experimental conditions. Third, the process engineer and the machine operator can vary many parameters (e.g., override of the spindle speed) in a continuous way, but within a narrow range around the conditions that fulfill workpiece requirements; these variations are very limited compared with the Taguchi experimental levels.

Over the last decade, the development of new communication and processing technologies has led to the growth of machine-learning solutions for industrial tasks under the umbrella of the Industry 4.0 paradigm (Oztemel *et al*. 2020). Machine-learning approaches to these industrial tasks differ from the approaches of mechanical models where the research focus has been on machining processes. In this case, only the efficiency of the machine-learning model is considered, in order to predict the workpiece quality indicator under study. Machine-learning techniques are especially suitable for extracting information from noisy (Chandrasekaran *et al*. 2010), strongly multivariable (Amini and Chang 2018), and partially incomplete data (Grzenda and Bustillo 2019). They therefore have a huge potential for use in industrial applications that concern machining processes, although they also present some limitations that should be overcome. One of these limitations refers to the experimental design that will provide the datasets for training and validating the prediction models.

The experimental *modus operandi* to develop mechanical models, based on a reduced number of experiments with very few inputs, has been extended to machine-learning modelling of machining processes (Zuperl and Cus 2003). Over the past twenty years, the suitability of machine-learning algorithms has been tested on small datasets, in terms of both their instances and inputs (Bustillo *et al*. 2018). For example, milling datasets for roughness prediction will usually have between 50 to 250 instances (Grzenda *et al*. 2012). However, these conditions are not often available in industrial environments. Besides, small-size experimental datasets might reduce the efficiency of machine-learning algorithms or produce some prediction-model malfunction such as overfitting. It has been proven that, if the dataset size increases, most of the machine-learning algorithms will provide better results (Grzenda and Bustillo 2013), although noise effects may be included. In some cases, as with ensembles, the existence of dataset dispersion or noise is a requirement for high efficiency (Kuncheva 2014). Some recent works have sought to apply small datasets to manufacturing problems (Pimenov *et al*. 2018). Strong imbalance resulting from fault diagnosis is one of the most common situations that these small databases address. Kovacs *et al*. (Kovacs 2020) proposed the use of undersampling, while the over-represented classes were combined with Ward's hierarchical clustering and Kohonen self-organizing map methods for pneumatic actuator state classification in the manufacture of electrical components. Wang *et al*. (Wang 2020) proposed a fault diagnosis model for turbo-generator sets based on the Weighted Extension Neural Network. The performance of their model was clearly better than other well-established machine-learning techniques such as Support Vector Machines, in terms of higher identification accuracy with fewer samples and strong noise-tolerant ability.

Considering the above requirement of machine-learning algorithms, this research evaluates the effect of experimental repetition in small datasets on algorithm accuracy. Therefore, the accuracy of different machine-learning algorithms will be evaluated in terms of dataset size. Dataset size will be varied by adding new repetitions of the same cutting conditions and the algorithm efficiency will be compared with the case of using the average of the different experimental repetitions. As previously mentioned, this last condition is the preferred option in mechanical modelling.

The industrial task selected for this analysis is the determination of the best combination for the tool of its side-rake and relief angles, a frequent challenge in the workshop (Kountanya 2009). In machining, the success of an operation is highly dependent on the choice of the tool material, the tool geometry, the cutting conditions, and the lubri-cooling atmosphere (Dogra *et al*. 2011). The side-rake and the relief angles are the most important in relation to tool geometry, because they determine the resistance of the cutting tool wedge; these angles are tied to the wedge angle, as the values of all three (side-rake, wedge and side-relief) angles must sum 90°. All three angles must be maximized: the rake angle to reduce chip curling; the relief angle, in order to reduce friction between the flank face of the tool and the workpiece; and,

the wedge angle to maximize tool resistance. The choice of these angles will affect the performance of the whole machining process (cutting force, heat generation, energy consumption, tool life, vibration, and workpiece surface quality) (Singh and Rao 2006). The Brandsma rapid-facing test is used to investigate the effect of tool geometry on tool life. In these tests, the tool travels from a disc center towards its periphery until the tool collapses, because of the continuous increase in cutting speed which eventually exceeds the strength of the tool. The output parameter of the test is the workpiece diameter at tool failure. The larger the diameter at tool failure, the more resistant the tool wedge. This method is considered to be a rapid tool-life test and has been successfully applied to compare the performance of cryogenically treated and non-cryogenically treated HSS tools (Da Silva *et al*. 2006).

The effect of tool geometry on machining performance has been evaluated in many studies. Saglam *et al*. (2006) used experimental tests and models to investigate the cutting force and tool tip temperature when varying the rake and approach angles of uncoated cemented carbide tools for the machining of hardened AISI 1040 steel (40 HRC). Günay *et al*. (2005) showed that the main cutting force followed a decreasing trend as the rake angle increased from negative to positive in the machining of AISI 1040 steel. Abainia and Ouelaa (2015) found the best experimental combination of tool geometries that reduced cutting force and tool vibrations more than any other and that improved machined surface roughness in the turning of carbon steel. Mir *et al*. (2017) used a smoothed particle hydrodynamics approach in simulation and experimental tests, to study the effect of the rake angle of single point diamond tools turning silicon, considering the cutting forces, chip formation, surface roughness, and tool wear. The results showed the enormous influence of tool geometry on the output parameters. Tunç and Budak (2012) simulated the end-milling process, in their study of the effect of the cutting condition and tool geometry (including the relief angle) on tool damping capacity and consequently, chatter-free cutting conditions. Li *et al*. (2020) modelled and simulated the chip formation process when machining Ti-6Al-4V using FEM, observing a reduction in chip segmentation as the rake angle increased. Zhang *et al*. (2019) developed a theoretical model and experimental tests to study the effect of the rake angle on crack propagation, surface generation, and chip formation processes of potassium dihydrogen phosphate crystals using single-point turning tools. The proposed model satisfactorily predicted the quantitative relationship between the rake angle and the crack propagation process. Other angles that affect the machining process have been studied. One example, in the work of Kattan and Currie (1996), was their experimental variation of the end cutting-edge angle, the nose angle, and the side cutting-edge angle. They also developed a mathematical model based on their results, to determine the best side-cutting edge angle that minimizes the machining forces. These studies demonstrated the importance of tool geometry in the machining process.

Another important variable is the lubri-cooling atmosphere used in the machining process. Application of a cutting fluid for example, with the turning of common and low alloy steels and superalloys with cemented carbide tools, is sometimes vital. At other times, lubri-cooling is not recommended, such as in hard machining, because the heat generated in the process is a factor in the reduction of the material strength of the workpiece, and a cutting fluid would hinder that process; and, in machining with ceramic tools, because they do not have thermal shock resistance, and therefore collapse with the application of a cutting fluid (Sales *et al*. 2001). The main functions of cutting fluids are lubrication, generally at low cutting speeds, and cooling, generally at high cutting speeds (Trent and Wright 2000). Several types and methods of application are available, and they may differ in their lubri-cooling qualities (Debnath *et al*. 2014). Flood cooling is the most common and widely investigated for machining, although the sustainable approach of the MQL technique and dry machining means that they are, whenever possible, preferable (Krolczyk *et al*. 2019). Finally, high-pressure cooling has been also tested during rough face-turning of nickel-based alloys, showing a significant reduction of flank wear and cutting forces (Suárez *et al*. 2017). Dry cutting, flood cooling, and MQL will be tested in the present investigation.

In summary, a response to the following question is sought: can the machine-learning models of a machining process take advantage of experimental tests performed under the same cutting conditions? This question will be answered for tool-life prediction in a case study of face-turning operations of AISI 1045 steel discs with varied cooling systems and tool geometries. The answer to this question opens up a new strategy to achieve machine-learning models of higher accuracy than those that exist at present. The results of this research might also, in the future, be extended to other machining processes (such as milling and drilling) and performance indicators (such as workpiece surface finish and dimensional accuracy). The block diagram in Figure 1 summarizes the work flow that was followed to answer the above question.

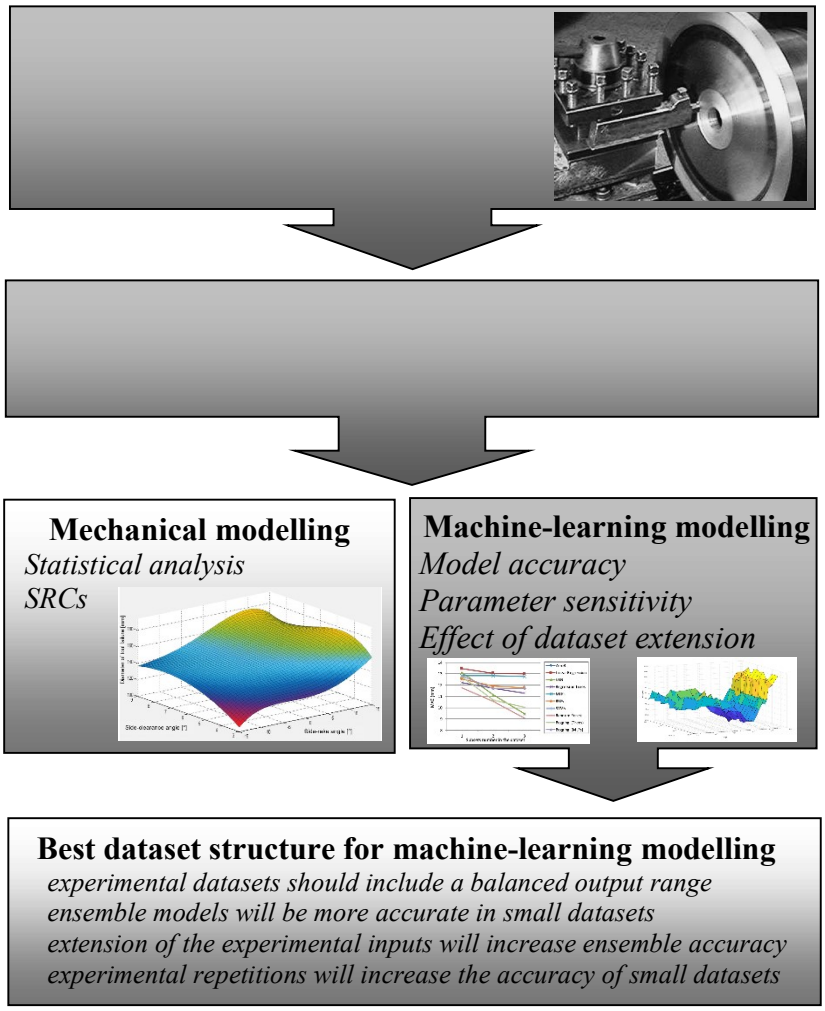


Fig. 1: Block diagram summarizing the structure of the work.

This paper will be structured as follows: in Section 2, the machine-learning techniques, used to model the case study, will be described. In Section 3, the selected study case will be defined and the best tool side-rake and relief angle settings will be determined for turning operations on carbon-steel workpieces. The results of both mechanical and machine-learning modelling will be presented in Section 4, paying special attention to the tuning sensitivity of the models and the influence of dataset size on model accuracy. Finally, in Section 5, the conclusions and future lines of work will be presented.

2 Machine-learning modelling procedure

A brief description will firstly be presented in this section of the machine-learning algorithms in this research. Then, the methodology used to train the machine-learning models and for their validation will be described.

2.1 Machine-learning techniques

Four families of machine-learning techniques were evaluated in this research: k-nearest neighbors, regression trees, artificial intelligence networks, and ensembles. Weka data-mining software was used (Witten *et al*. 2016).

K-Nearest Neighbors (kNNs) is an instance-based method (Aha and Kibler 1991). The algorithm stores the training examples; when a new instance is evaluated for output prediction, the algorithm generates the most similar instances in the training data, providing a weighted average of those instances as its output prediction. In Weka software, the kNN implementation is called IBk and uses the Euclidean distance to weigh the predicted output. KNN models have demonstrated their very accurate predictions for different machining processes, such as roughness prediction in milling

operations (Grzenda & Bustillo 2019) and dimensional accuracy in laser micromachining operations (Teixidor *et al.* 2015).

Regression trees (Quinlan 1992) are machine-learning models composed of different linear models. They are a variation of decision trees that can handle continuous outputs (Juez-Gil *et al.* 2019). Each regression tree is a tree-based structure. The tree is composed of decision nodes: depending on the input values, one branch of the tree is sequentially selected until a final leaf is reached. The leaves contain a linear model to calculate the output prediction. The branching criterion is the information gain. Each decision node includes a dilemma concerning the value of one of the attributes; if its value is higher than a given threshold, one rather than another of the branches will be selected. Decision trees are very popular, because they provide visual models, from which useful information on the machining process can be directly extracted. The Weka implementation of regression trees is called M5P (Quinlan 1992). Decision trees have demonstrated their utility at supplying accurate and visual information for different machining processes, such as surface quality in deep-drilling operations (Bustillo *et al.* 2016), and reliable prediction models for ore-crushing plate lifetimes (Juez-Gil *et al.* 2019).

Artificial Neural Networks (ANNs) are undoubtedly the most common bio-inspired machine-learning algorithm (Abiodun 2018). In fact, their capability to learn from previous experience has turned them into a standard machine-learning solution for manufacturing tasks. Small computational units, called neurons, are the basis of ANNs. Each neuron provides an output considering an activation function; this function is evaluated from a weighted transformation of the inputs. Neurons are split into, at least, three layers. The first layer is an input-feed with the input values. The middle or hidden layers evaluate the hidden patterns of the dataset. These layers are especially useful to discard background noise. Finally, the last layer calculates the output value. MultiLayer Perceptron (MLP) is the most common type of ANN (He *et al.* 2020). MLPs use a linear activation function and their neuron connections are weighted by a back–propagation algorithm. MLP models have demonstrated their high accuracy in different machining processes, such as surface roughness in milling operations (Bustillo *et al.* 2011), tool condition monitoring in milling operations (Cuka and Kim 2017), and workpiece dimensional prediction (Arnaiz-Gonzalez *et al.* 2016). However, despite such a high accuracy, MLPs present a serious drawback, in that they have many parameters to be tuned. Therefore, the tuning process must be supervised by an expert. One solution to this problem is the selection of another type of ANN with less-tunable parameters, such as Radial Basis Function (RBF) networks. The RBF is a type of ANN where the neuron activation function in the hidden layer is an RBF (Leonard and Kramer 1991). An RBF only has one layer, although this layer can directly perform non-linear calculations that require more than one layer in MLPs. RBFs have also been successfully tested for machining modelling, such as dimensional prediction in ball-end finishing operations (Arnaiz-Gonzalez *et al.* 2016) and drill-wear prediction in milling operations (Patra *et al.* 2010).

Ensembles consist of a combination of tens of simple models (e.g., regression trees) called base models (Kuncheva 2014). The base models provide different output predictions, because they are trained with different subsets of the training dataset. Finally, their predictions are combined to provide the final ensemble prediction (Kuncheva 2014). One of the most popular ensembles is Bagging (Ren *et al.* 2016). This ensemble trains each base model with a bootstrapped subsample of the original training data set with replacement. The output is a linear combination of all the base-model predictions. The best way to maximize the accuracy of Bagging-based models is the use of very unstable and fast base models. A model is considered unstable when slight differences in its training dataset will produce highly different prediction models. The speed requirement comes from the need to build hundreds of models during the training step. Taking into account those two requirements, regression trees are the most promising base models for Bagging ensembles, although other machine-learning algorithms, such as MLPs, can also be used for this task. In this research, both MLPs and regression trees are considered as base models to analyze ensemble capability at extracting information from very simple models, such as regression trees, compared with more complex models, such as MLPs. Random Forest is another popular type of ensemble (Breiman 2001). Random Forest uses random trees as base models. Random trees are a subtype of regression trees. Although they use the same training dataset, each random tree uses only a random subset of the attributes for building each decision node, thereby providing a set of very different regression trees. It has been demonstrated that ensembles have lower tuning parameter sensitivity than MLPs for some manufacturing tasks such as surface quality in milling operations (Bustillo *et al.* 2011). Besides, the suitability of Random Forest for machining datasets has been widely demonstrated in tool-wear prediction for milling operations (Wu *et al.* 2017), dimensional prediction of stent manufacturing by micro-milling (Maudes *et al.* 2017), and surface-quality prediction in milling operations (Grzenda *et al.* 2019).

Besides these four families of machine-learning algorithms, two additional models were used in this study as baseline methods: a naïve approach and a linear regression. The naïve approach will always predict the average output value of the training dataset, regardless of the input values. The linear regression will calculate the output value as a linear combination of the inputs that minimize the squared error of the training dataset. Even though very simple, linear

regression models provide reliable predictions for many manufacturing tasks when the input variation range is very limited. For instance, Suresh *et al.* (2012) showed the effectiveness of linear regression models for evaluating cutting force, power, specific cutting force, tool wear, and surface roughness for the machining of hardened AISI 4340 steel. In Weka software, these models are called ZeroR and Linear Regression, respectively. It can be expected that the prediction error of the machine-learning models will be significantly lower than those of the baseline methods, due to their simplicity, and the complexity of the machining processes.

2.2 Machine-learning training and validation procedure

The procedure to build machine learning-training models requires the definition of 1) the most suitable performance indicators to evaluate model accuracy; 2) the training–validation scheme; and, 3) the description of the tuning procedure for the machine-learning parameters. Two performance indicators were selected in this work: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Both indicators were considered, because there is no general agreement over which one might be better to measure model accuracy in regression tasks for machining processes (Juez-Gil 2019). The units of both indicators, in mm, were the same as the predicted target units. This information facilitated the evaluation of model performance. The equations for calculating MAE and RMSE are as follows:

$$MAE = \frac{\sum_{t=1}^{n} (\hat{y}_t - y_i)}{n} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (\hat{y}_t - y)^2}{n}} \tag{2}$$

where, n is the number of instances in the test subset, i identifies the instance used for the current prediction, $\hat{y}_i$ is the predicted value, and y_i is the actual value of the output variable.

Regarding the training and validation scheme, a 10x10 cross-validation scheme (Kohavi 1995) was selected due to the small dataset size. In a cross-validation scheme, the model is only evaluated with new data (not previously used in the training stage). For this purpose, the original dataset is randomly divided into 10 subsets, 9 of which are used for training and the remaining one is used for validation. The process is repeated by changing the selected subset for validation until the 10 subsets have been used for validation. The whole training and validation scheme is repeated 10 times with different random subsets, to assure the statistical validity of the whole process. In summary, 100 models will be trained and validated with different subsets of the dataset and the final accuracy will be the averaged result of all 100 models.

Finally, the tuning process of the machine-learning model parameters plays a major role in model accuracy. From among the selected algorithms, MLPs showed the highest sensitivity to the tuning process, while the ensembles were almost insensitive to it (Bustillo *et al.* 2011). Two different strategies were followed to find the best parameters for both models: MLP and RBF. Both strategies were based on a grid search of the main ANN parameters, a common tuning-procedure technique (Chandrasekaran *et al.* 2010). In the first strategy, a limited variation of the main ANN parameters, Learning Rate for the MLPs, and Number of Functions for RBFs was performed. In the second strategy, up to three ANN parameters -Learning Rate, Momentum, and number of neurons in the hidden layer for the MLPs, and Number of Functions, Ridge, and Tolerance for the RBFs- were all varied. The two tuning strategies were used in different conditions, as will be described in Section 4.

The tuning ranges of the ANN parameters were as follows:

- Learning rate: from 0.00001 to 0.11 in steps of 0.005
- Momentum: from 0.01 to 0.5 in steps of 0.05
- Number of neurons in the hidden layer: 2, 3, and 4. In this work, we used only one hidden layer, due to the reduced number of inputs.
- Number of Functions: 2, 3, and 4
- Ridge: from 0.00001 to 0.1 in steps of 0.005
- Tolerance from 1.0E-14 to 0.00001 in steps of 5.0E-12

The default options in Weka were used for the other ANN parameters. The other machine-learning algorithms were not tuned at all and the WEKA default options were also used.

3 Study case: optimization of tool geometry in face-turning operations

In this Section, a description of the experimental set-up and the data-acquisition procedure during the experimental tests will be presented, followed by the datasets that were extracted from the experimental tests and later used to train and to validate the machine-learning models.

3.1 Experimental set-up and data acquisition

The Brandsma rapid-facing test was used to determine the best side-rake and clearance angle settings that will yield the maximum diameter of tool failure. Figure 2 illustrates the rapid facing test, where vc1 is the cutting speed at the beginning of the test (in m/min), vc2 the final cutting speed when tool collapse occurs (in m/min), d1and d2 are the initial and final disk diameters, respectively (mm), and n is the spindle rotation speed (rpm).

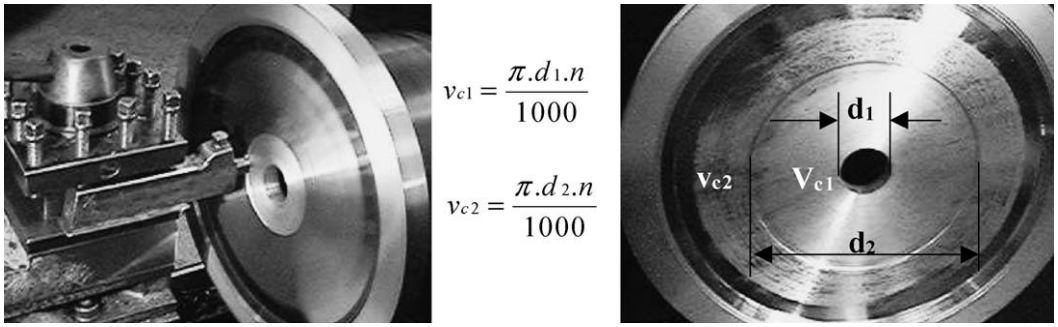


Fig. 2: Illustration of the Brandsma rapid facing test (Da Silva *et al*. 2006).

AISI 1045 steel workpieces were used with an average hardness of 211 HB, in the form of discs with diameters of 250 mm and thicknesses of 50 mm. The chemical composition of the material in use, AISI 1045 steel, is listed in Table 1. The initial diameter at the center of the disc was 31 mm, of sufficient size for introducing the HSS tool bit at the start of the tests. A Romi Imor II 520 conventional lathe, maximum spindle speed 1400 rpm, with a power consumption of 4.5 KW, was used. In this kind of test, where tool failure will occur in a single pass, High-Speed Steels (HSS) are recommended, because of their lower wear resistance compared, for instance, with cemented carbide. If cemented carbide were used, tool collapse, a prerequisite of the Brandsma test, could only be provoked with a very large disc diameter or very high-speed rotation, conditions that require a high-power lathe and that could easily compromise the rigidity of the test conditions. AISI M2 HSS tool bits in the form of square bars (10 x 10 x 100 mm^3) with an average hardness of 64 HRC were selected for the tests. The chemical composition of these tools is shown in Table 2. A Pantec dial gauge with a resolution of 0.001 mm was used, for accurate measurement of the tool failure diameter (workpiece diameter of tool collapse, due to plastic deformation of the cutting wedge).

Table 1: Chemical composition of AISI 1045 steel [% wt].

C	Mn	S	P	Fe
0.42 - 0.5	0.6 - 0.9	to 0.05	to 0.04	98.51 - 98.98

Table 2: Typical chemical composition of M2 HSS tool [% wt] (Adapted from Mesquita *et al*. 2017).

C	Mn	Si	Cr	V	W	Mo
0.85 - 1.0	0.1 - 0.4	0.1 - 0.4	4 - 4.25	1.7 - 2.1	6.0 - 6.5	4.75 - 5.25

The tool bit geometries were ground in a Universal sharpener machine manufactured by Atlasmaq Ind Com de Máquinas Ltda, model PP6025Q, using an alumina gridding wheel, mesh 60 manufactured by Race Abrasivos Ind Com Ltda. The cutting conditions were held constant, whereas the side-rake angle was varied at nine levels and the side-clearance angle at four levels, according to Table 3. This resulted in 36 tests with different combinations of tool geometries. The Brandsma tests were repeated three times for each tool geometry, to increase the statistical reliability, resulting in a total number of 108 tests. The averaged results of the three test repetitions yielded the average tool failure diameter used as the output parameter for each tool geometry.

Table 3: Tool geometry and cutting conditions used in the machining tests.

Side-rake angle [degree]	Side-clearance angle [degree]	End clearance angle [degree]	Side cutting edge angle [degree]	End cutting edge angle [degree]	Back rake angle [degree]	Depth of cut [mm]	Feed rate [mm/rev]	Rotation speed [rpm]
-15	3	3	15	15	0	0.5	0.167	180
-12	6							
-9	7.5							
-6	9							
0								
6								
9								
12								
15								

Three lubrication-cooling conditions were used in the Brandsma tests: dry, flood cooling, and MQL. An emulsion of vegetal base oil, Vasco 1000, manufactured by Blaser Swisslube, was used as the cutting fluid, at a concentration of 8% in water. Under flooding conditions, the cutting fluid was directed at the cutting zone (overhead direction) with a flow rate of 90 l/h, while in the MQL mode, the spray of fluid was directed at the cutting zone using two nozzles, forming an angle of 120° with a total flow rate of 30 ml/h.

3.2 Dataset description

Different datasets were generated from the machining tests. All the datasets included three inputs: lubrication-cooling conditions type and side-clearance and side-rake angles of the cutting tool. The output was the same in all datasets: the tool failure diameter (diameter of the workpiece at the moment of tool failure; likewise, the corresponding cutting speed at tool failure). The type of lubrication-cooling conditions was the only nominal input, while the other inputs and the output were continuous variables. The inputs and the output variables included in the dataset and their range of variation are summarized in Table 4; the output variable, the tool failure diameter, is outlined in bold.

Table 4: Dataset attributes and outputs, range of variation, and units.

Variable	**Abbreviation/Symbol**	**Range**
Lubrication type	*LUB*	d, f, m
Side-rake angle	γ_o	-15° - +15°
Side-clearance angle	α_o	+3° - +9°
Diameter of tool failure	***dtf***	**59.8 – 206.2 mm**

As the experiments were repeated three times, the first three datasets were generated using the data obtained from each experiment. These datasets included 108 instances, as 108 different cutting conditions were tested. Then, another three datasets were built, merging the results of two of the three repetitions of the experiments. These datasets included 216 instances. Besides, an additional seventh dataset was built using the results of the three experiments together and an eighth, using the average values for each cutting condition of the three repetitions of each experiment. In total, eight different datasets were generated: 4 of these datasets had 108 instances (3 datasets with data from only one experiment and 1 dataset with the average values of the 3 experiments), 3 had 216 instances (2 experiments together in each dataset), and 1 had 324 instances (the 3 experiments together). All the datasets are included as supplementary material. The standard deviation of the tool failure diameter in these datasets was within a range between 23.19 and 24.46 mm, which are very high values if compared with the variable range: between 59.8 and 206.2 mm. These values demonstrate one of the main features of the tool-wear process. Its dispersion or variability (Sick 2003) is a central factor to take into account when building prediction models for this machining indicator.

4 Results and discussion

In this Section, the results of the mechanical modelling of the tool failure diameter will be presented. Then the results of modelling the failure diameter using machine-learning algorithms will be described; the accuracy of each algorithm will be evaluated, as well as the sensitivity of the model to the tuning processes of their main parameters.

Finally, the influence of dataset size on the accuracy of each machine-learning algorithm will be compared, leading to conclusions on the best experimental design for each machine-learning model.

4.1 Mechanical approach

The results of the average values of the tool failure diameter obtained from the Brandman tests in each combination of tool geometry, according to Table 3, were used to generate models for each lubrication-cooling condition in Matlab software, 2015a. Under dry and flood-cooling conditions, fourth-order models gave reasonable coefficients of determination, while for the MQL condition, a sixth-order model had to be used. They had the following coefficients of determination, R^2: 0.82 for the dry condition, 0.84 for flood cooling, and 0.82 for MQL. They are respectively represented in Equations (3), (4) and (5).

$$f(x,y)=9.963+10.41\cdot x+33.28\cdot y+0.3587\cdot x^2-8.41\cdot x\cdot y-2.249\cdot y^2+0.01672\cdot x^3-0.05379\cdot x^2\cdot y+1.726\cdot x\cdot y^2+0.01316\cdot y^3-0.0006728\cdot x^4-0.002444\cdot x^3\cdot y+0.002878\cdot x^2\cdot y^2-0.1002\cdot x\cdot y^3; \quad (3)$$

$$f(x,y)=-0.9208-6.412\cdot x+67.03\cdot y+0.0542\cdot x^2+3.812\cdot x\cdot y-9.816\cdot y^2+0.01296\cdot x^3+0.01164\cdot x^2\cdot y-0.6455\cdot x\cdot y^2+0.4839\cdot y^3-0.0001924\cdot x^4-0.002057\cdot x^3\cdot y-0.001792\cdot x^2\cdot y^2+0.03674\cdot x\cdot y^3; \quad (4)$$

$$f(x,y)=104.9+12.97\cdot x+5.296\cdot y-6.512\cdot x\cdot y+0.02543\cdot x^2\cdot y+0.9691\cdot x\cdot y^2+0.006079\cdot x^3\cdot y+0.1378\cdot x^3\cdot y^2-0.04407\cdot x\cdot y^3-4.292\cdot 10^{-5}\cdot x^5-3.136\cdot 10^{-10}\cdot x^4\cdot y^6+0.1383\cdot x^3\cdot y^2-1.467\cdot 10^{-6}\cdot x^4*y^3, \quad (5)$$

Figures 3, 4, and 5 show the Surface Response Curves (SRC) that represent the models within the range of side-rake values and clearance angles under study. Figure 3 shows a maximum tool diameter of 164.55 mm for tool failure under dry conditions, obtained at a side-rake angle of 12° and a side-clearance angle of 7.5°, which is 0.15% higher than the experimental average value of 164.30 mm. Under flood-cooling conditions (Fig.4), the model gave a maximum tool diameter of 177.40 mm for tool failure at side-rake and clearance angles of 9°. This value was 6.99% lower than the experimental average value of 189.8 mm. The maximum tool diameter that the model generated for tool failure under the MQL condition (Fig. 5) was 182.55 mm at a side-rake angle of 12° and a side-clearance angle of 9°, which was 2.34% higher than the average experimental value of 178.37 mm for this tool geometry. These three maximums are quiet broad, showing stable expected conditions around the maximums that can minimize any uncertainty; they can therefore be considered the best conditions in terms of the lifetime of the tool for turning under the selected cutting specifications.

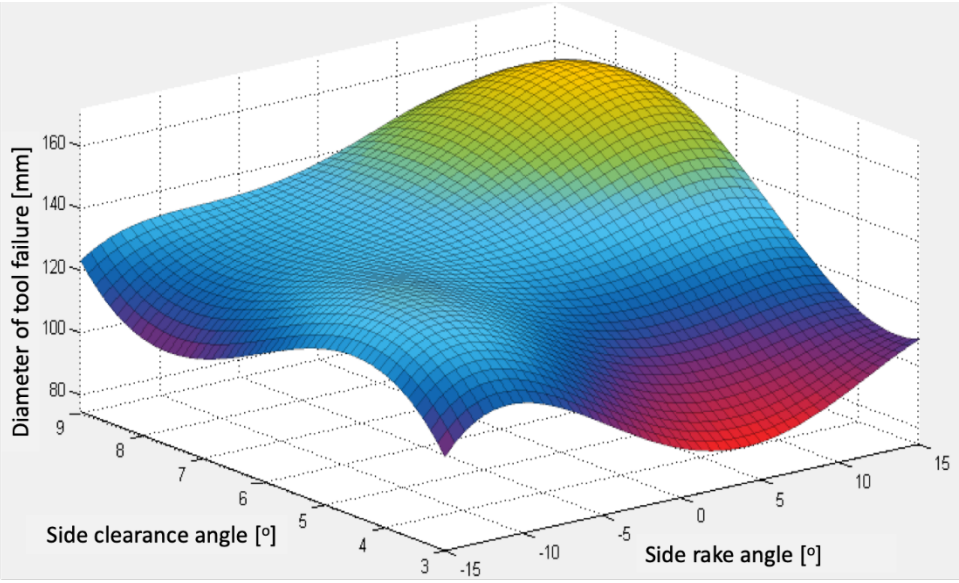


Fig.3: Surface response - dry condition.

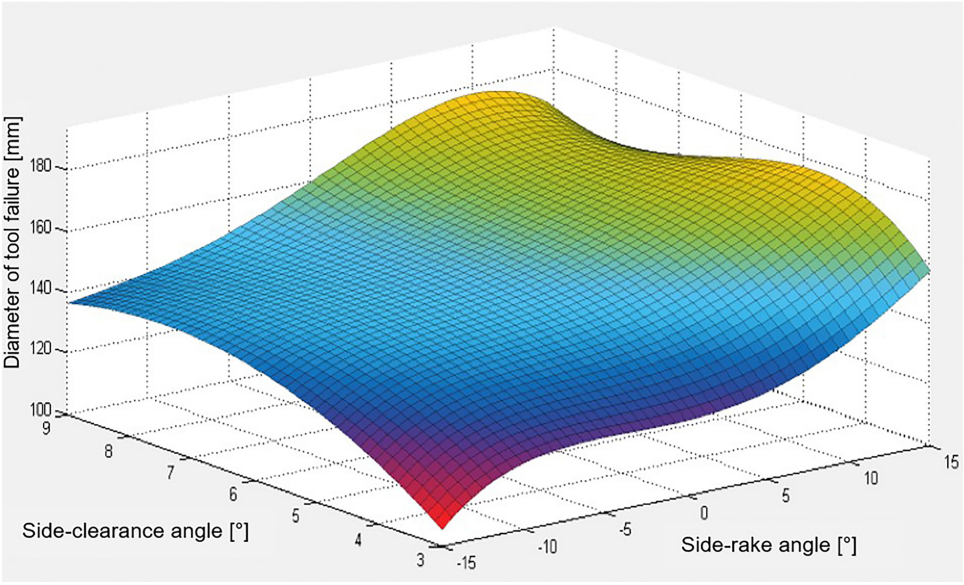


Fig. 4: Surface response - flood-cooling condition.

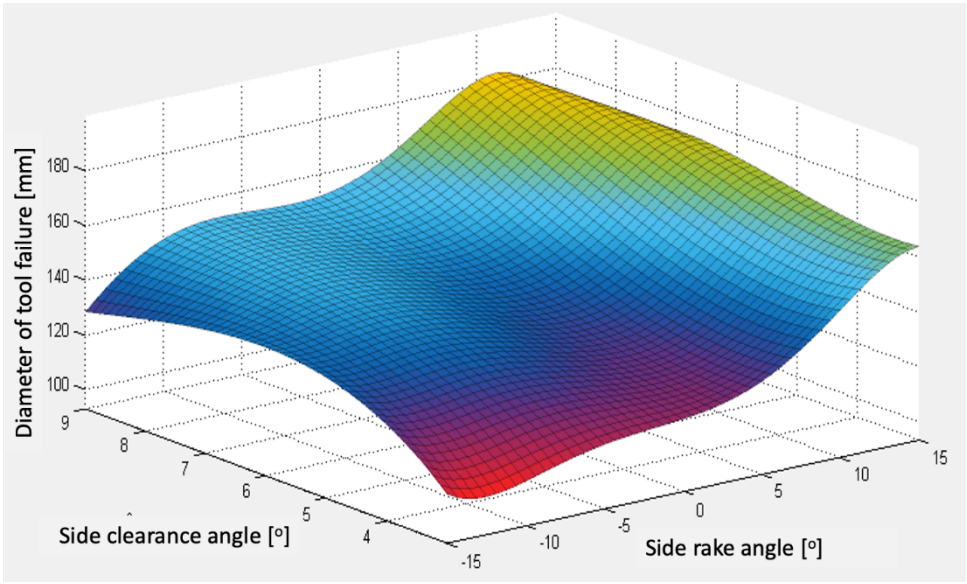


Fig. 5: Surface response - MQL condition.

Figure 6 shows the average values of the Brandsma test results for all the tool geometries tested under dry, flood cooling, and MQL conditions. The results showed that the lubrication-cooling condition is very important when turning steels. The dry condition, although the most environmentally friendly, gave the worst results. The best experimental result was achieved with the application of flood cooling that combines both lubrication and cooling, reducing friction at the chip-tool workpiece interfaces, and cutting temperatures that retard tool wear. The MQL-condition results are very close to those of flood cooling, which is an interesting alternative, as it is far less polluting, although poor system performance was reported in (Sales *et al*. 2001) under higher cutting speeds (with cemented carbide).

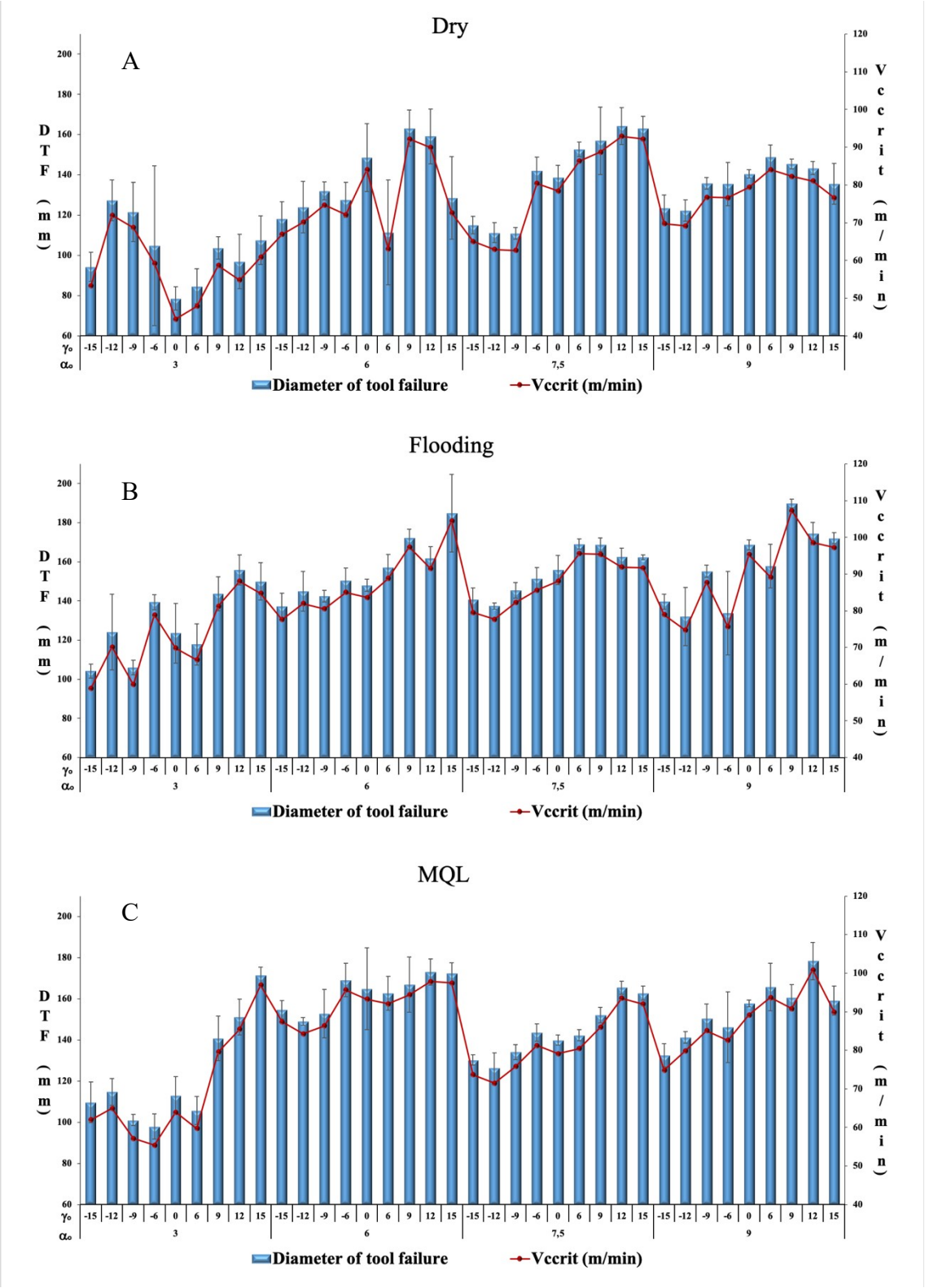


Fig. 6: Diameter of tool failure (DTF) in all tool geometry tests. A) Dry cutting; B) Flood cooling; C) MQL.

4.2 Machine-learning approach: accuracy and model-tuning sensitivity

In the first stage, the conventional strategy for machine-learning modelling of machining tasks was followed: the dataset of the averaged results for each cutting condition was used to train and to validate the models. As described in Section 2.2, most of the algorithm parameters were not tuned at all. As only the two types of ANN -MLPs and RBFs- showed that some of their parameters had a strong influence on model accuracy, only those parameters were varied. In each cases, two types of tuning were implemented; firstly, a limited variation of their main parameters, Learning Rate for the MLPs and Number of Functions for RBFs; secondly, an extensive variation of up to three parameters, Learning Rate,

Momentum, and number of neurons in the hidden layer for the MLPs, and Number of Functions, ridge and tolerance for RBFs. In this second case, the most accurate model was obtained for an MLP with 0.1 Learning Rate, 0.35 Momentum, and 3 neurons in the hidden layer and, for the RBFs, with 4 functions, 0.00002 Ridge, and 1.0E-11 Tolerance.

The accuracy of those models in terms of MAE and RMSE are summarized below in Table 5. The strongly tuned models (second tuning procedure) are marked with an asterisk. The percentage improvement in model accuracy, referring to the linear regression model, is also presented in Table 5. Bold font shows the technique with the highest accuracy.

Table 5: Accuracy of machine-learning techniques on the average dataset.

Machine-learning technique	**MAE (mm)**	**% MAE Improvement**	**RMSE (mm)**	**% RMSE Improvement**
ZeroR	18.8485	-54.9	23.2521	-54.3
Linear Regression	12.1686	0	15.0721	0
kNN	10.5487	13.3	14.4262	4.3
Regression Trees	9.9271	18.4	12.8692	14.6
MLPs	11.0229	9.4	13.9417	7.5
MLPs*	10.7392	11.7	13.6419	9.5
RBFs	10.822	11.1	13.9417	7.5
RBFs*	**8.605**	**29.3**	**11.3559**	**24.7**
Random Forest	9.7322	20	12.7974	15.1
Bagging (Trees)	10.777	11.4	13.667	9.3
Bagging (MLPs)	10.9013	10.4	14.1552	6.1

Some interesting conclusions may be extracted from Table 5. First, the results were analyzed in terms of the MAE indicator. All the machine-learning models improved the accuracy of the best baseline method –the linear regression- by at least 10%. Obviously, linear regression was already 54% better than the Naïve approach, a result that might suggest a partially-linear relationship between the three inputs and tool life. Besides, machine-learning algorithms can be split into three groups depending on their accuracy. The first group improved the accuracy of the linear regression within a range between 9 and 13%. This group comprised the kNNs, both types of ANNs (MLPs and RBFs) with a limited tuning procedure, MLPs with an extensive training procedure, and bagging ensembles (both with regression trees or MLPs as base models). A second group improved the accuracy of the linear regression by more than 18%. This second group was formed of Regression Trees and Random Forest. Finally, RBFs with an extensive training procedure achieved the best improvement (in bold in Table 5): 29.3%. However, in this case, up to three parameters of the RBFs had to be tuned across a very broad range.

When Table 5 was analyzed in terms of RMSE, the conclusions were the same as using MAE, although the differences between all the machine-learning models and the linear regression model were significantly lower.

Some characteristics of the dataset might explain these results. The difference between the MAE and the RMSE results can be explained by examining Figure 7. It shows the correlation between the experimental diameter of tool failure (X-axis) and the predicted one (Y-axis) for the Regression Tree algorithm. The cross size indicates the average error for each instance. Where Fig. 7 shows homogenous medium-size crosses, then MAE and RMSE might be very similar, and where the cross sizes vary greatly (with large and small crosses), then RMSE will significantly increase but not MAE. The reason is found in the definition of both indicators: RMSE considers the sum of the squares of the prediction errors, while MAE considers the sum of the prediction errors. A clear example of this second case is shown below in Figure 7.

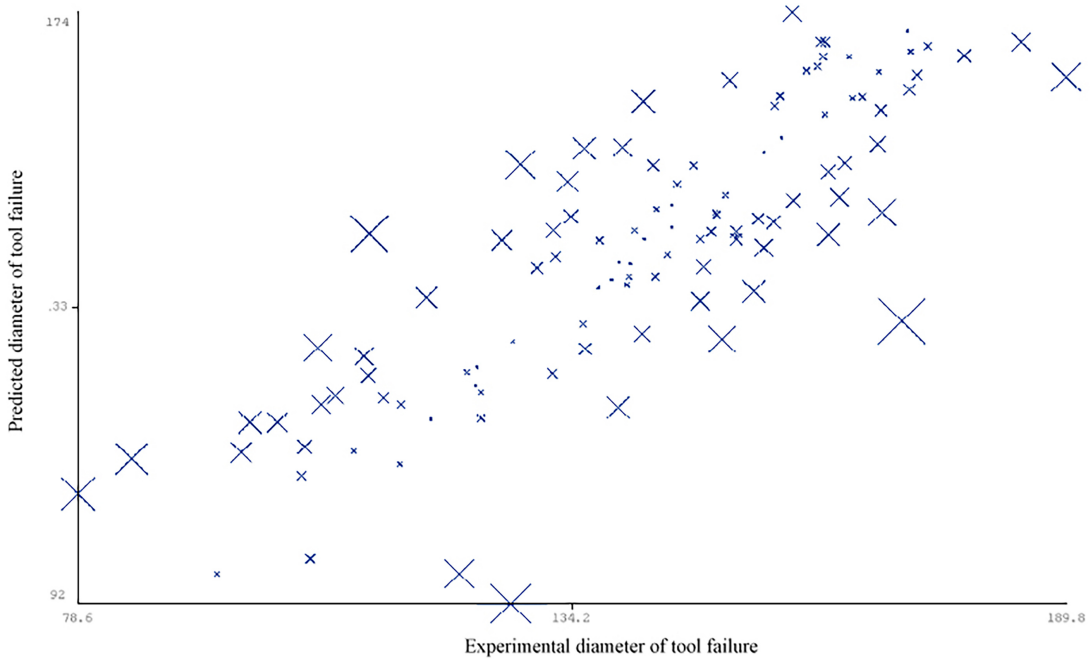


Fig. 7: Error analysis for Regression Tree models.

Figure 7 shows three different locations for the instances with large errors: the instances with 1) very low; or, 2) very high values of the tool failure diameter; and, 3) the instances within the middle range. high errors were expected for the first two groups of machine-learning algorithms; there are very few instances in those intervals and the machine-learning algorithms penalized them, because the focus of the training process was on the most populated areas with the most instances. There was no clear explanation for the second group, and instances with large crosses were commonly considered outliers in this populated area, because their behavior was far from as expected. On the one hand, outliers will decrease model accuracy for any model, baseline method or machine-learning algorithm, because they are outside the expected behavior. However, the presence of low-populated areas will especially affect machine-learning models. On the other hand, these areas will have little effect on a linear regression model, because the model will minimize the squared error on the training dataset. Therefore, with the existence of instances with large errors in low populated areas, RMSE errors for machine-learning models will be higher than MAE errors compared with linear regression.

The different performances of the various machine-learning techniques can have different reasons. First, as the dataset is very small and no close conditions are tested, kNN is not a good solution, because no close neighbors can be found by the algorithm to evaluate the new conditions included in the validation subset (Teixidor *et al*. 2015). The case of ANNs, especially MLPs, is also related to dataset size. A low number of inputs will limit their learning capability (Grzenda and Bustillo 2013), and the number of hidden layers that could be introduced in the ANN (Grzenda *et al*. 2012), as outlined in previous studies for similar tasks, such as surface prediction in roughing operations. Bagging ensemble processes have a similar disadvantage, in that they cannot provide sufficient diversity in the training subsets of the different base models, due to the limited number of inputs (only 3). In contrast, regression trees are capable of extracting much more useful information from the dataset and they manage data-set size and size variation better than other techniques. The RepTrees provide both accuracy and diversity to the Random Forest for models of even greater accuracy. This capability of Random Forest has previously been demonstrated for surface wear prediction, based on surface isotropy levels (Bustillo *et al*. 2018).

One more conclusion should be extracted from Table 5. Regardless of whether MAE or RSME are evaluated, the prediction-model error was quite high. This result was also detected in Section 4.1 for experimental models where a low coefficient of determination, R^2, was achieved for any cooling-lubrication model. If the standard deviation of the output, 23.19 mm, is considered, then the RMSE errors of most of the machine-learning models were within a range of 12-14 mm. These high errors were related to the existence of outliers and the limited number of cases with a very short or a very long tool life. In other words, wear processes, due to their nature, have very high dispersion rates (Rehorn *et al*. 2005). This fact will assure the presence of outliers in any dataset. Besides, the standard experimental design will entail only very rare examples of extreme minimum or maximum tool lives, because they will often refer to cutting conditions that are both of low industrial interest, which will penalize machine-learning model accuracy.

Finally, an analysis of RBF tuning procedures and their stability is advisable, to demonstrate that the RBF model is of most interest for the prediction of tool life under these conditions. Figure 8 shows the accuracy of the RBF models, if the value of its two most sensible parameters, Ridge and Tolerance, are varied around their optimal value. Both the X-

and the Y-axis show the tuning parameters on a logarithmic scale, while the Z-axis shows the accuracy (in terms of MAE) of the RBF model.

This figure shows a structure of peaks, due to the limited number of test values, because the extension of the test range was prioritized, to parameter fine tuning, at some distance from the optimized model (down peak at 0.00002 Ridge and 1.0E-11 Tolerance). The conclusions from Fig. 8 are: 1) Ridge parameter plays a major role rather than Tolerance; 2) the tuning process is very sensitive for models under 9.0 mm of MAE; and, 3) models with MAE at around 9.8 mm are easily obtained. Therefore, if a reduced experimental dataset is available, very accurate RBFs models (MAE<9.0 mm) will be very sensitive to the training dataset, reducing the reliability of the model; while accurate RBF models (9.0mm<MAE<10 mm) will be more reliable and of low sensitivity to the training dataset, although Random-Forest ensemble models of the same type can also be obtained with no tuning procedure.

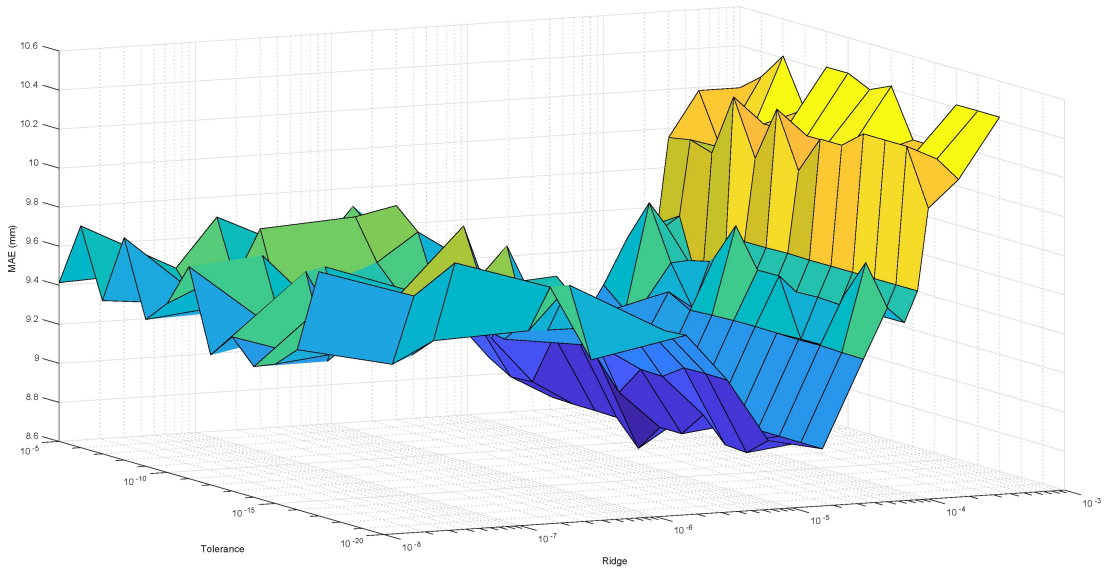


Fig. 8: Analysis of RBF model accuracy in terms of its tuning parameters.

4.3 Machine-learning approach: influence of dataset size on model accuracy

Section 4.2 has shown that machine-learning techniques provide prediction models of medium accuracy for very dispersive processes such as tool-wear evaluation. Although they fulfil industrial requirements, their accuracy can be greatly improved. These models are based on the standard strategy of mechanical modelling: the use of average results of experimental tests. However, if a different strategy were followed, then greater machine-learning technique capabilities could be exploited. Most of these techniques will foreseeably improve in accuracy, if the dataset size increases (Grzenda *et al*. 2012). These results might also be expected, although the increase is due to the additional instances with the same cutting conditions, through repeated cutting experiments.

This possibility was analyzed using the 3 repetitions of the experimental design, instead of using their average result. The new datasets were previously described in Section 3.2. There are 3 datasets that include only 1 repetition of each experiment, referred to as *1-subset*; 3 datasets that include 2 repetitions of each experiment, referred to as *2-subsets*; and, 1 dataset that includes all the 3 repetitions of each experiment, referred to as *all-subsets*. The former dataset, based on the average value of the 3 repetitions will be referred to as *Average*.

Then, the same machine-learning algorithms used in Section 4.2 were trained and validated with these datasets. If many model parameters were tuned, the tuning procedure would be significantly more complex, so only the Learning Rate of the MLPs was varied within the range of 0.01 to 0.1, to find the most accurate model. This limitation was decided upon, because the objective, at this stage of the research, was to evaluate the influence of dataset size on model accuracy and was not to identify the most accurate model. Because of this tuning limitation, the accuracy for the average dataset was lower than the one shown in Section 4.2 in Table 5 for those algorithms that required a complex tuning process such as MLPs or RBFs.

Table 6 summarizes the accuracy of those models, in terms of MAE and RMSE, for the different datasets. As there were 3 datasets with 1 repetition and with 2 repetitions, the accuracy shown in Table 6 for those cases was the average accuracy of the 3 models in each case, having been trained and validated with one of the datasets. The technique with the highest accuracy is highlighted in bold text.

Table 6: Accuracy of machine-learning techniques depending on the dataset size.

	MAE (mm)				**RMSE (mm)**			
	1-subset	**2-subsets**	**All subsets**	**Average**	**1-subset**	**2-subset**	**All subset**	**Average**
ZeroR	19.59	19.57	19.51	18.85	24.57	24.55	24.48	23.25
Linear Regression	13.45	13.04	12.99	12.17	17.31	16.72	16.74	15.07
kNN	13.10	11.13	9.48	10.55	13.88	15.98	13.51	14.43
Regression Trees	12.23	11.71	11.30	9.93	16.10	15.41	14.96	12.87
MLPs	12.83	12.83	12.75	11.02	16.79	16.58	16.54	13.94
RBFs	12.53	11.89	11.75	10.82	16.63	15.64	15.43	13.89
Random Forest	**11.76**	**10.55**	**9.06**	**9.73**	**16.43**	**14.38**	**12.84**	**12.80**
Bagging (Trees)	12.71	10.67	10.03	10.78	17.01	14.40	13.66	13.67
Bagging (MLPs)	12.81	11.68	11.71	10.90	16.99	15.19	15.46	14.16

The analysis of Table 6 can be done in terms of either eMAE or RMSE, but all the models behaved in the same way for both quality indicators. Therefore, this analysis will only be presented in terms of MAE. Only if a different behavior is observed between both quality indicators, will it be outlined. Two data analyses are shown in Table 6: 1) which methods improved their accuracy as the dataset size increased; and, 2) which methods achieved higher accuracy using all the repetitions instead of using their averaged values.

Three methods hardly improved their accuracy as the dataset size increased (less than 5%). These were the Naïve approach, Linear Regression, and MLPs. The first two cases might be expected: the Naïve approach and the Linear Regression models provided predictions that were a constant value or a linear fit, respectively; the inclusion of repeated data in the dataset will not significantly change the prediction of those models. The behavior of MLPs was more complex to understand; MLPs are not properly suited to this industrial task, as Section 4.2 has outlined, perhaps because the dataset has a very low number of inputs and instances for MLPs to develop their learning capability (Grzenda and Bustillo 2013). It has been outlined that MLP requires the use of many neurons in different hidden layers for a proper fit with non-linear problems (Grzenda *et al*. 2012), a difficult issue if the dataset is very small and with very few inputs. When the dataset size increases through repeated cutting conditions, MLPs will have no new information for a better fit with the cutting process. This fact has also been identified for other industrial problems where many repeated conditions, such as fault diagnosis, provoke a strong imbalance in the dataset (Santos *et al*. 2018).

Secondly, three methods slightly improved their accuracy as the dataset increased (between 5% and 10%): regression trees, RBFs, and Bagging of MLPs. In those cases, the number of instances within the low populated areas (very low and very high tool wear) increased along with the dataset size, though the accuracy of the algorithms was not boosted. RBFs might be extremely sensitive to the tuning parameters, as the tuning process is at that stage very limited. The behavior of MLP bagging might be explained through their mixed behavior from ensemble to MLPs (they strongly improved their accuracy as an ensemble, as it will be explained later, but as MLPs they hardly improved in accuracy at all).

Finally, three methods greatly improved their accuracy (by over 20%) as the dataset increased. These were kNNs, Random Forest, and Bagging of regression trees. In this case, the behavior of kNNs was foreseeable: if the dataset size increases with repeated cutting conditions, then the algorithm will have more instances in the training data that are similar to those presented in the validation data, and its accuracy will immediately improve. The cases of Random Forest and Bagging of regression trees showed that the diversity of the base models as the dataset size increased was highly advantageous for the ensembles (Kuncheva 2014), as other authors have also pointed out for other wear processes such as surface wear based on surface isotropy levels (Bustillo *et al*. 2018), and tool condition monitoring based on a force sensor (Wang *et al*. 2014).

The second analysis of Table 6 refers to which methods achieved higher accuracy using all the repetitions rather than only using the averaged values. Most of the methods showed equitable results whether the 3 repetitions or the averaged values were used: the Naïve Approach, linear regression, regression trees, MLPs, and Bagging of MLPs. These methods matched the algorithms that showed no significant improvement in their accuracy as the dataset size increased,

therefore yielding a foreseeable result. However, kNNs, Random Forest, and Bagging of regression trees all increased their accuracy when they used the repetition datasets. The case of the kNNs has previously been explained. In the case of the ensembles, although their improvement was minor (<10%), it clearly marked a trend, as shown in Fig. 9. It plots the accuracy of the different techniques as a function of the number of subsets included in the dataset; *1* refers to 1 subset (108 instances), *2* refers to 2 subsets (216 instances), and *3* refers to all the subsets (324 instances). While the X-Axis refers to the number of subsets, it can also be understood as the number of instances in the dataset or the dataset size. The trend towards improved accuracy as the dataset size increased was clear for Random Forest, Bagging of regression trees, and kNNs. In the case of the most accurate model, Random Forest, the use of the three repetitions instead of only one improved model accuracy by 23%.

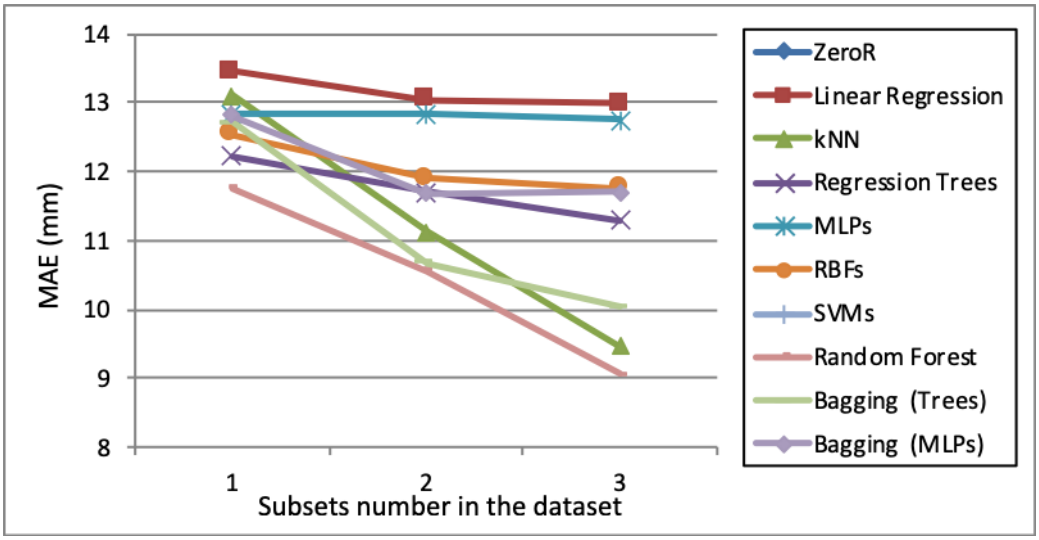


Fig. 9: Model accuracy vs number of subsets in the dataset.

In summary, increasing the dataset size by means of repeated experiments was of no advantage to some machine-learning techniques, for example, ANNs or Decision Trees, while others, such as kNN or ensembles, significantly improved their accuracy with this strategy. This result was, at least, true for small datasets of industrial tasks with a high dispersion rate, such as tool wear prediction. As the repetition of experiments is a compulsory technique, to achieve reliable results from the mechanical point of view, some machine-learning techniques should be fed with the original datasets, rather than the averaged datasets, to achieve higher model accuracy. ANNs of RBFs showed the lowest errors (8.605 mm for MAE), in relation to the most accurate machine-learning technique for tool-life modelling when considering the side-rake and the relief angles of HSS tools as Bransdma facing-test inputs under different flooding conditions. However, this accurate model needs a very sensitive and complex tuning process. Other machine-learning techniques, such as ensembles, especially Random Forest, provided models with almost the same accuracy (9.05 mm for MAE) and with no tuning procedure.

4.4 The best dataset structure for machine-learning modelling

The analysis presented in the previous sections yielded some interesting conclusions on the best experimental design, for the improvement of machine-learning modelling:

- The intrinsic nature of the wear process means that very dispersive datasets with many outliers will always reduce the accuracy of any prediction model.
- The presence of low-populated areas will strongly decrease the accuracy of machine-learning models. The experimental dataset should therefore include a balanced output range and the experimental design should target that objective. Nevertheless, some of the programmed cutting conditions during the experimental tests can be of little or no industrial interest.
- ANNs are problematic fits for non-linear relationships. If the dataset size is small with very limited numbers of inputs, repetitions of cutting conditions will not significantly improve their accuracy.
- Likewise, ensembles are not adapted to the analysis of small datasets, because they cannot achieve enough diversity in the training subsets of the different base models. However, repetitions of cutting conditions represented a good strategy for significant improvements of their accuracy. RepTrees provided both accuracy and diversity for Random Forest to achieve models that were still more accurate.

The trend of improving their accuracy as the dataset size increased was clear for Random Forest, with an accuracy improvement of 23%.

- Although some ANNs such as RBFs might provide very accurate models, their tuning process showed extreme sensitivity to the datasets, making them models of low-reliability for small datasets. Unlike RBFs, ensembles required no parameter tuning and were therefore very stable prediction models.
- Some machine-learning models provided better results using the traditional strategy (averaging the results of experiments repetitions) when the datasets were very small. Those models, regression trees, MLPs, and Bagging of MLPs, matched with the algorithms that showed no significant improvement in their accuracy as the dataset size increased.

These conclusions may be added to another previously presented by other authors:

- Both the monitoring and the variation of as many inputs as possible during the cutting processes will improve the efficiency of machine-learning models (Amini and Chang 2018), and non-linear relationships are expected, especially in the case of ANNs (Grzenda *et al*. 2012), (Grzenda and Bustillo 2013). These new inputs, although in some cases providing very little new information, will provide diversity to the ensemble modelling process to improve its accuracy (Santos *et al*. 2018).

5 Conclusions

As with any wear process, AISI 1045 steel discs under dry, MQL, and flooding conditions presented high variability in each Bransdma facing test, although the same tool geometry, cutting conditions, and workpiece material were used. The standard strategy to prevent any such variability is the repetition of the experiment (e.g., 3 times) and the calculation of the average tool life under each cutting condition. The average values of the tool failure diameter obtained from the Brandsma rapid facing test were used to generate mechanical models, taking into account tool side-rake and clearance angles. Surface response curves were used to analyze the results.

The best set of tool angles under dry conditions was $\gamma = 12^{\circ}$ and $\alpha o = 7.5^{\circ}$; for flood cooling, side-rake and clearance angles equal to 9°; and for MQL, $\gamma = 12^{\circ}$ and $\alpha o = 9^{\circ}$. Steel machine tool manufacturers usually recommend values for both side-rake and clearance angles at around 5° or 6°. Although most of the practical turning operations use cemented carbide tools for turning steels, the results for the HSS tools in the present investigation showed that these tool geometries were at a considerable distance from the optimum points under test conditions. This result may be an indication that the optimization of cemented carbide tool geometry is also necessary. HSS is used more often for the turning of small diameter workpieces and in the micro-turning of steels, so the results of this study are useful in that area.

Some new interesting conclusions have been drawn from the same dataset that was used for wear-modelling processes with machine-learning techniques. First, artificial intelligence RBF networks showed the lowest errors (8.605 mm for MAE). However, this sort of accurate model will require a very sensitive and complex tuning process. Other machine-learning techniques, such as ensembles, especially Random Forest, generated models with almost the same accuracy (9.05 mm for MAE) with no tuning procedure. Although this accuracy range fulfils industrial requirements, it could be greatly improved. These models are based on the standard strategy of mechanical modelling: the use of average values of repeated experimental tests. However, if the strategy of repeated experiments is followed, then the capabilities of machine-learning techniques can be better exploited, by adding instances to the training dataset with previously tested cutting conditions. When this strategy is tested on the experimental dataset, some machine-learning techniques, such as ANNs or Decision Trees, hardly improved their performance at all (less than 10%), while others, such as kNN or ensembles, improved significantly (more than 20%). Ensembles increased their accuracy especially when they used simple and unstable base models such as regression trees, while their accuracy hardly improved at all (less than 10%) when they used complex base models, such as MLPs. This result is at least true for small datasets of industrial tooling tasks with high dispersion rates, used for example for tool-wear prediction. If the machine-learning models are fed with the original datasets instead of the averaged datasets, then they will increase in accuracy, because the repetition of experiments is a compulsory technique to achieve reliable results from the mechanical point of view.

Future research will be focused on analyzing balancing the data and therefore increasing the dataset size and its effect on increasing model accuracy; instead of repeated experimental conditions, the new instances should be at some distance from the optimal conditions, that is: conditions with very low productivity or with very high tool wear. Usually, cemented carbide is the most highly recommended tool material for the machining of steels. Future work should therefore focus on using a rapid tool-life test (e.g., conical turning or continuous accelerated spindle-speed tests) to investigate the best set of tool geometries for this important tool material. Likewise, more robust models should be used, so that optimization processes can be applied to other important tool geometry characteristics (e.g., back-rake angle, nose angle, nose radius, cutting edge radius).

Acknowledgements

This investigation was partially supported by Project TIN2015-67534-P of the Ministerio de Economía y Competitividad of the Spanish Government and by Project BU085P17 of the Junta de Castilla y León, both co-financed through European Union FEDER funds. The work was supported through Act 211 of the Government of the Russian Federation, under contract № 02.A03.21.0011.

References

Abainia, S., & Ouelaa, N. (2015) Experimental study of the combined influence of the tool geometry parameters on the cutting forces and tool vibrations. *International Journal of Advanced Manufacturing Tech*nology, 79(5-8), 1127-1138. https://doi.org/10.1007/s00170-015-6885-9

Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., & Arshad, H. (2018) State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Aha, D., & Kibler, D. (1991) Instance-Based Learning Algorithms. *Machine Learning*, 6, 37-66. https://doi.org/10.1023/A:1022689900470

Amini, M., & Chang, S. (2018) A review of machine learning approaches for high dimensional process monitoring. *In IISE Annual Conference and Expo*, 390-395.

Antony, J., & Kaye, M. (2000) Analysis and Interpretation of Data from Taguchi Experiments. *In: Experimental Quality*. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-5293-2_9

Arnaiz-Gonzalez, A., Fernandez-Valdivielso, A., Bustillo, A., & Lopez de Lacalle, L.N. (2016) Using artificial neural networks for the prediction of dimensional error on inclined surfaces manufactured by ball-end milling. *International Journal of Advanced Manufacturing Tech*nology, 83(5), 847-859. https://doi.org/10.1007/s00170-015-7543-y

Benardos, P.G., & Vosniakos, G.C. (2002) Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments. *Robotics and Computer-Integrated Manufacturing*, 18(5-6), 343-354, https://doi.org/10.1016/S0736-5845(02)00005-4

Breiman, L. (2001) Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bustillo, A., Díez-Pastor, J.-F., Quintana, G., & García-Osorio, C. (2011) Avoiding neural network fine tuning by using ensemble learning: application to ball-end milling operations. *International Journal of Advanced Manufacturing Tech*nology, 57(5), 521-532. https://doi.org/10.1007/s00170-011-3300-z

Bustillo, A., Grzenda, M., & Macukow, B. (2016) Interpreting tree-based prediction models and their data in machining processes. *Integrated Computer-Aided Engineering*, 23(4), 349-367. https://doi.org/10.3233/ICA-160513

Bustillo, A., Pimenov, D.Y., Matuszewski, M., & Mikolajczyk, T. (2018) Using artificial intelligence models for the prediction of surface wear based on surface isotropy levels. *Robotics and Computer-Integrated Manufacturing*, 53, 215-227. https://doi.org/10.1016/j.rcim.2018.03.011

Chandrasekaran, M., Muralidhar, M., Krishna, C., & Dixit, U. (2010) Application of soft computing techniques in machining performance prediction and optimization: a literature review. *International Journal of Advanced Manufacturing Technology*, 46(5), 445–464. https://doi.org/10.1007/s00170-009-2104-x

Cuka, B., & Kim, D.-W. (2017) Fuzzy logic based tool condition monitoring for end-milling. *Robotics and Computer-Integrated Manufacturing*, 47, 22-36. https://doi.org/10.1016/j.rcim.2016.12.009

Da Silva, F.J., Franco, S.D., Machado, A.R., Ezugwu, E.O., & Souza Jr., A.M.S. (2006) Performance of cryogenically treated HSS tools. *Wear*, 261, 674–685. https://doi.org/10.1016/j.wear.2006.01.017

Davim, J.P. (2008) Machining: fundamentals and recent advances. *Machining: Fundamentals and Recent Advances*, 1-36. https://doi.org/10.1007/978-1-84800-213-5

Debnath, S., Reddy, M.M., & Yi, Q.S. (2014) Environmental friendly cutting fluids and cooling techniques in machining: a review. *Journal of Cleaner Production*, 83, 33–47. https://dx.doi.org/10.1016/j.jclepro.2014.07.071

Dogra, M., Sharma, V.S., & Dureja, J. (2011) Effect of tool geometry variation on finish turning - A review. *Journal of Engineering Science and Technology Review*, 4(1), 1-13.

Grzenda, M., & Bustillo, A. (2013) The evolutionary development of roughness prediction models. *Applied Soft Computing*, 13(5), 2913-2922. https://doi.org/10.1016/j.asoc.2012.03.070

Grzenda, M., & Bustillo, A. (2019) Semi-supervised roughness prediction with partly unlabeled vibration data streams. *Journal of Intelligent Manufacturing*, 30(2), 933-945. https://doi.org/10.1007/s10845-018-1413-z

Grzenda, M., Bustillo, A., Quintana, G., & Ciurana, J. (2012) Improvement of surface roughness models for face milling operations through dimensionality reduction. *Integrated Computer-Aided Engineering*, 19(2), 179-197. https://doi.org/10.3233/ICA-2012-0398

Günay, M., Korkut, I., Aslan, E., & Seker, U. (2005) Experimental investigation of the effect of cutting tool rake angle on main cutting force. *Journal of Materials Processing Technology*, 166, 44–49. https://doi.org/10.1016/j.jmatprotec.2004.07.092

He, C., Ma, M., & Wang, P. (2020) Extract interpretability-accuracy balanced rules from artificial neural networks: A review. Neurocomputing, 387, 346-358. https://doi.org/10.1016/j.neucom.2020.01.036

Juez-Gil, M., Erdakov, I.N., Bustillo, A., & Pimenov, D.Y. (2019) A regression-tree multilayer-perceptron hybrid strategy for the prediction of ore crushing-plate lifetimes. *Journal of Advanced Research*, 18, 173-184. https://doi.org/10.1016/j.jare.2019.03.008

Kattan, I.A., & Currie, K.R. (1996) Developing new trends of cutting tool geometry. *Journal of Materials Processing Technology*, 61, 231-237. https://doi.org/10.1016/0924-0136(96)02492-2

Kohavi, R. (1995) The power of decision tables. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 912, 174-189.

Kountanya, R., Al-Zkeri, I., & Altan, T. (2009) Effect of tool edge geometry and cutting conditions on experimental and simulated chip morphology in orthogonal hard turning of 100Cr6 steel. *Journal of Materials Processing Technology*, 209(11), 5068-5076. https://doi.org/10.1016/j.jmatprotec.2009.02.011

Kovacs, T., & Ko, A. (2009) Monitoring Pneumatic Actuators' Behavior Using Real-World Data Set. *SN Computer Science*, 1, 196. https://doi.org/10.1007/s42979-020-00202-2

Krolczyk, G.M., Maruda, R.W., Krolczyk, J.B., Wojciechowski, S., Mia, M., Nieslony, P., & Budzik, G. (2019) Ecological trends in machining as a key factor in sustainable production – a review. *Journal of Cleaner Production*, 218, 601-615. https://dx.doi.org/10.1016/j.jclepro.2019.02.017

Kuncheva, L.I. (2014) Combining Pattern Classifiers: Methods and Algorithms. 2nd edition. *Joboken (New Jersey): John Wiley and Sons. Inc*. https://doi.org/10.1002/9781118914564

Leonard, J.A., & Kramer, M.A. (1991) Radial basis function networks for classifying process faults. *IEEE Control Systems*, 11(3), 31-38. https://doi.org/10.1109/37.75576

Li, A., Zhang, J., & Zhao, J. (2020) Effect of cutting parameters and tool rake angle on the chip formation and adiabatic shear characteristics in machining Ti-6Al-4V titanium alloy. *International Journal of Advanced Manufacturing Tech*nology, 107(7-8), 3077–3091. https://doi.org/10.1007/s00170-020-05145-9

Martellotti, M.E. (1941) An analysis of the milling process. *Transactions of ASME*, 63, 667.

Maudes, J., Bustillo, A., Guerra, A.J., & Ciurana, J. (2017) Random Forest ensemble prediction of stent dimensions in microfabrication processes. *International Journal of Advanced Manufacturing Tech*nology, 91(1-4), 879-893. http://dx.doi.org/10.1007/s00170-016-9695-9

Mesquita, R.A., Barbosa, C.A., & Machado, A.R. (2017) Heat Treatment of Tool Steels, Book Chapter in Comprehensive Materials Finishing. *Elsevier, Organized by Professor Shahjahan Mridha, Ph.D., (edited by Saleem Hashmi)*, 2, 214-245. https://doi.org/10.1016/B978-0-12-803581-8.09191-8

Mir, A., Luo, X., & Cheng, K. (2017) A. Cox, Investigation of influence of tool rake angle in single point diamond turning of silicone. *International Journal of Advanced Manufacturing Tech*nology, 94(5-8), 2343-2355. https://doi.org/10.1007/s00170-017-1021-7

Nalbant, M., Gökkaya, H., Toktaş, I., & Sur, G. (2009) The experimental investigation of the effects of uncoated, PVD- and CVD-coated cemented carbide inserts and cutting parameters on surface roughness in CNC turning and its prediction using artificial neural networks. *Robotics and Computer-Integrated Manufacturing*, 25(1), 211-223. https://doi.org/10.1016/j.rcim.2007.11.004

Oztemel, E., & Gursev, S. (2020) Literature review of Industry 4.0 and related technologies. Journal of Intelligent Manufacturing, 31(1), 127-182. http://dx.doi.org/10.1007/s10845-018-1433-8

Patra, K., Pal, S.K., & Bhattacharyya, K. (2010) Fuzzy radial basis function (FRBF) network based tool condition monitoring system using vibration signals. *Machining Science and Technology*, 14(2), 280-300. https://doi.org/10.1080/10910344.2010.500954

Pimenov, D.Y., Bustillo, A., & Mikolajczyk, T. (2018) Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth. *Journal of Intelligent Manufacturing*, 29(5), 1045-1061. https://doi.org/10.1007/s10845-017-1381-8

Quinlan, J.R. (1992) Learning with continuous classes. *Proc. of the 5th Australian Joint Conference on Artificial Intelligence, 92m Singapore*, 343-348.

Rehorn, A.G., Jiang, J., & Orban, P.E. (2005) State-of-the-art methods and results in tool condition monitoring: A review. *International Journal of Advanced Manufacturing Tech*nology, 26(7-8), 693-710. https://doi.org/10.1007/s00170-004-2038-2

Ren, Y., Zhang, L., & Suganthan, P.N. (2016) Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]. *IEEE Computational Intelligence Magazine*, 11(1), 7379058, 41-53. https://doi.org/10.1109/MCI.2015.2471235

Saglam, H., Unsacar, F., & Yaldiz, S. (2006) Investigation of the effect of rake angle and approaching angle on main cutting force and tool tip temperature. *International Journal of Machine Tools and Manufacture*, 46, 132–141. https://doi.org/10.1016/j.ijmachtools.2005.05.002

Sales, W.F., Diniz, A.E., & Machado, A.R. (2001) Application of Cutting Fluids in Machining Processes. *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 23(2), 227-240. https://dx.doi.org/10.1590/S0100-73862001000200009

Sanchez-Egea, A.J., & Lopez de Lacalle, L.N. (2018) Machines, processes, people and data, the keys to the 4.0 revolution. *DYNA*, 93(6), 576-577. https://doi.org/10.6036/8807

Santos, P., Maudes, J., & Bustillo, A. (2015) Identifying maximum imbalance in datasets for fault diagnosis of Gearboxes. *Journal of Intelligent Manufacturing*, 29(2), 333-351. https://doi.org/10.1007/s10845-015-1110-0

Sick, B. (2002) On-line and indirect tool wear monitoring in turning with artificial neural networks: A review of more than a decade of research. *Mechanical Systems and Signal Processing*, 16(4), 487-546. https://dx.doi.org/10.1006/mssp.2001.1460

Singh, D., & Rao, P.V. (2006) A surface roughness prediction model for hard turning process. *International Journal of Advanced Manufacturing Tech*nology, 32(11-12), 1115–1124. https://doi.org/10.1007/s00170-006-0429-2

Suárez, A., López de Lacalle, L.N., Polvorosa, R., Veiga, F., & Wretland, A. (2017) Effects of high-pressure cooling on the wear patterns on turning inserts used on alloy IN718. *Materials and Manufacturing Processes*, 32(6), 678-686. https://doi.org/10.1080/10426914.2016.1244838

Suresh, R., Basavarajappa, S., & Samuel, G.L. (2012) Some studies on hard turning of AISI 4340 steel using multilayer coated carbide tool. *Measurement: Journal of the International Measurement Confederation*, 45(7), 1872-1884. https://doi.org/10.1016/j.measurement.2012.03.024

Teixidor, D., Grzenda, M., Bustillo, A., & Ciurana, J. (2015) Modeling pulsed laser micromachining of micro geometries using machine learning techniques. *Journal of Intelligent Manufacturing*, 26(4), 801–814. https://doi.org/10.1007/s10845-013-0835-x

Trent, E.M., & Wright, P.K. (2000) Metal Cutting, *4th Edition. Butterworths, London, UK*, 464 pgs.

Tunç, L.T., & Budak, E. (2012) Effect of cutting conditions and tool geometry on process damping in machining, *International Journal of Machine Tools and Manufacture*, 57, 10-19. https://dx.doi.org/10.1016/j.ijmachtools.2012.01.009

Wang, G., Yang, Y., & Li, Z. (2014) Force sensor based tool condition monitoring using a heterogeneous ensemble learning model. *Sensors*, 14(11), 21588-21602. https://doi.org/10.3390/s141121588

Wang, T., Wang, J., Wu, Y., Sheng, X. (2020) A fault diagnosis model based on weighted extension neural network for turbo-generator sets on small samples with noise. Chinese Journal of Aeronautics, 2020, Article in press. https://doi.org/10.1016/j.cja.2020.06.024

Witten, I.H., Frank, E., Hall, M.A., & Pal, C.J. (2016) Data Mining: Practical Machine Learning Tools and Techniques (Book). *Data Mining: Practical Machine Learning Tools and Techniques*, 1-621.

Wu, D., Jennings, C., Terpenny, J., Gao, R.X., & Kumara, S. (2017) A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 139(7), 071018. https://doi.org/10.1115/1.4036350

Zhang, S., Zhang, H., & Zonga, W. (2019) Modeling and simulation on the effect of tool rake angle in diamond turning of KDP crystal. *Journal of Materials Processing Technology*, 273, 116259. https://doi.org/10.1016/j.jmatprotec.2019.116259

Zheng, P., Wang, H., Sang, Z., Zhong, R.Y.E., Liu, Y., Yu, S., & Xu, X. (2018) Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. *Frontiers of Mechanical Engineering*, 13(2), 137-150. https://dx.doi.org/10.1007/s11465-018-0499-5

Zuperl, U., & Cus, F. (2003) Optimization of cutting conditions during cutting by using neural networks. *Robotics and Computer-Integrated Manufacturing*, 19(1-2), 189-199. https://doi.org/10.1016/S0736-5845(02)00079-0