Integrated Computer-Aided Engineering 23 (2016) 349–367
DOI 10.3233/ICA-160513
IOS Press

Interpreting tree-based prediction models and their data in machining processes

Andres Bustillo[a], Maciej Grzenda[b,*] and Bohdan Macukow[b]
[a]*Department of Civil Engineering, University of Burgos, Burgos, Spain*
[b]*Faculty of Mathematics and Information Science, Warsaw University of Technology, Warszawa, Poland*

Abstract. Machine-learning techniques frequently predict the results of machining processes, based on pre-determined cutting tool settings. By doing so, key parameters of a machined product can be predicted before production begins. Nevertheless, a prediction model cannot capture all the features of interest under real-life industrial conditions. Moreover, careful assessment of prediction credibility is necessary for accurate calibration; aspects that should be addressed through appropriate modeling and visualization techniques. A machine process test problem is proposed to analyze data-visualization techniques, in which a real data set is analyzed that describes deep-drilling under different cutting and cooling conditions. The main objective is the efficient fusion of visualization techniques with the knowledge of industrial engineers. Common modeling and visualization techniques were first surveyed, to contrast standard practice with our novel approach. A hybrid technique combining conditional inference trees with dimensionality reduction was then examined. The results show that a process engineer will be able to estimate overall model accuracy and to verify the extent to which accuracy depends on industrial process settings and the statistical significance of model predictions. Moreover, evaluation of the data set in terms of its sufficiency for modeling purposes will help assess the credibility of these decisions.

Keywords: Visualization, deep drilling, machining processes, prediction, dimensionality reduction

1. Introduction

1.1. The modeling of machining processes

Industrial experts managing modern machining processes have to maximize process productivity and they simultaneously have to ensure appropriate final product quality. Machine learning [20] models such as neural networks can be used to model quality parameters [45]. The accuracy of such models is usually assessed with performance indicators such as mean prediction error rates. Dimensionality reduction techniques such as Principal Component Analysis (PCA) [40] are also frequently applied to reduce error rates [23,56] and eliminate noisy data [1].

*Corresponding author: Maciej Grzenda, Faculty of Mathematics and Information Science, Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warszawa, Poland. Tel.: +48 22 621 93 12; Fax: +48 22 625 74 60; E-mail: M.Grzenda@mini.pw.edu.pl.

Studies primarily emphasize the development of novel data-processing and modeling techniques, which results in a variety of approaches even for a single machining process. So, there are a large group of possible models, depending on the most critical key features of the machined workpiece [5]. Key features of the product may be predicted on the basis of such models: surface roughness and tool wear among others, as well as geometrical errors, the most critical of which is surface roughness, due to its influence on product performance in many different industries, among which the production of moulds and dies [5].

Machine learning techniques involved in process modeling include neural networks, Bayesian networks, fuzzy-logic, and ensembles of different kind of classifiers, etc. [11,13,61]. In all these examples, certain principal conclusions may in all cases be applied: the first is the inevitable development of a separate prediction model for each cutting tool, as the list of measured parameters varies between different industrial environ-

ISSN 1069-2509/16/$35.00

ments and possible cutting processes. Moreover, even similar tools from which data records are collected under the same set of cutting conditions may require their own prediction models, because the diameter and general geometry of tool shapes and tool materials could substantially vary their behavior [5]. As a consequence, there is a need to use a significant group of different models for the modeling of individual tools.

An important aspect for the development of a model is the availability of experimental data. The development of a process model with machine-learning techniques requires data collection in real time when the industrial tool is in use. Data acquisition is done through a series of experiments, such as deep drilling experiments at different tool settings, which go beyond typical tool settings. In addition, the experiments should be repeated to capture the variability of the experimental results under the same tool settings. Furthermore, the quality of the result such as borehole roughness has to be measured. Not surprisingly, this results in substantial experimental costs. Hence, a tradeoff is observed between the collection of large data sets and the minimization of the experimental costs. The number of experimental records has to be largely constrained, so as not to diminish the potential for real use of the process model. As stated before, this is largely because of the fact that combining the data from different tools is frequently impossible, due to differences in measurements and tool characteristics. Hence, the data sets composed of a limited number of records (i.e. up to 250 records) are frequently used. In particular, A. Çiçek et al. used only 36 records to model thrust forces in the drilling of AISI 316 stainless steel [12], Y.-H. Tsai et al. used a data set composed of 48 records to develop a surface recognition system with neural networks [62] and Hashmi et al. [29] used 84 records to optimize feed rate using fuzzy rules in the drilling of 4 different materials. One of the largest reported data sets is composed of 250 records and was used in [54] to monitor surface roughness in ball-end milling operations. A detailed study of recent reviews in the application of machine-learning modeling to machining processes [11,13,61] would suggest that it is very rare to find data sets with more than 250 instances Hence, for any solutions to be applicable in real industrial settings, they have to be developed and validated with data sets of limited size. Importantly, such data sets require particular attention when interpreting the data models. Ideally, a process engineer using a model should not only be informed of the predicted results of a process under planned tool settings, but also of the confidence level related to such predictions, which may vary over different tool settings.

However, even very accurate models that predict process results can not capture all features of interest, such as tool wear and tool reconfiguration costs [42]. Moreover, measurement of every variable that contributes to product quality is no easy task. Direct measurement of certain variables-for instance, material inhomogeneity and machine performance-is extremely difficult. Relevant data may however be provided by the measurement of other independent variables-such as cutting vibrations-that indirectly provide us with input data for the prediction model. What is even more challenging is the fact that these very same features also function as output features. For example, some process settings outside the optimal range of production parameters may increase tool wear and machine vibrations. From this perspective, tool wear and machining vibrations can be treated as both input and output features of the process, which contributes to the complexity of prediction models and their development.

Other real-life process-configuration-related variables that the process engineer has to consider include tool cost and energy consumption per machined workpiece. This illustrates the fact that the prediction models that only address some aspects of the process can not be treated as black boxes under industrial conditions, as they are not sufficient in themselves to control the process. More precisely, such models have to be merged with the expert knowledge of the process engineer who will take the final decisions. One of the ways to merge such knowledge is to use visualization techniques, which show the mutual relation between the measurable variables and the dependencies revealed by the model. There is therefore a need for visualization of the process data. For instance, as outlined in Section 2, many authors plot 3D figures for different cutting processes that show the impact of selected process settings on the surface roughness of the machined workpiece. This approach reflects the unprecedented opportunities in this area owing to modern stationary and mobile devices, the display capabilities of which are undergoing constant improvement. So much so that on-site data visualization and investigation of various figures is now an advantageous option.

However, a majority of works that deal with the application of machine-learning models under industrial conditions provide only limited visualization of the models and the data used to develop them, as outlined in Section 2. One possible reason for this lack of research into visualization techniques is the fact that the data arising from machining processes is multidimensional, which makes such visualization techniques all the more challenging.

1.2. The transparency of prediction models

The visualization of interpretable prediction models can satisfy the need of the process engineer for clearly stated prediction rules. A process engineer should be clearly informed of the dependencies observed in the process data for a particular tool and the way in which the predicted results at specific tool settings were developed. The latter expectation is not unique for the modeling of machining processes and is directly related to model transparency.

Model transparency [50], also referred to as comprehensibility [50] or interpretability [21] can be defined as the ability of a human user to *understand what the model consists of, leading ideally to the ability to apply it to new observations* [50]. Another aspect, is the transparency of model development process [17,55] i.e. the way the model was built, including parameter settings and underlying assumptions. This is vital for decision making in such domains as health care [17,55], in relation to applications for epilepsy detection [22], among others. In this study, we focus on the ability of a human user to understand and to apply prediction rules contained in the model. Hence, transparency will be used in this context interchangeably with interpretability.

Not all machine learning models exhibit transparency. Many categories of models such as Support Vector Machines [20] and neural networks [3] are capable of high accuracy modeling of the data. However, their interpretability is extremely limited, especially when even more complex models such as probabilistic neural networks are applied [4].

Whether interpretability is a major concern depends largely on the application domain. In domains such as image processing, model transparency is difficult to attain. Image preprocessing, feature extraction, and image-data-based reasoning are inevitably very complex. Hence, models of low interpretability frequently involving the use of neural networks and related techniques are applied in this domain. These are used for tasks such as image classification [63]. Other application-oriented tasks, such as the analysis of video images [33], emotion recognition [39] or sound processing [16] share the same assumptions i.e. they try to automate the tasks that can not be formalized by human users, especially while using a limited set of reasoning rules.

In some areas, such as financial applications, including bankruptcy prediction [50], transparency is largely desired. This is also a desirable feature when a physical interpretation and inspection for validation purposes is needed and possible, as observed in [66] in the case of fault diagnosis. In these and similar cases, a domain expert can compare the inference rules embedded in interpretable process model with domain knowledge. The objective is to validate the model, but also to discover new possible insights on the process that is of interest.

The visualization of both process data and interpretable process models is of particular interest when machining processes are considered. Such visualization promotes real-life adoption of models built with machine-learning techniques. Hence, our primary objective is the investigation of the techniques that could be used both to predict process results and to provide experts from industry with sufficient insight into the inter-dependencies observed in the data. Moreover, the degree of credibility attached to these predictions should be made clear and the significance of individual predictions should be depicted for easy viewing by industrial experts. In fact, the resulting product parameter for some tool settings may be predicted with high accuracy, while for some other cases, the prediction will be of low accuracy and low confidence: facts that should also be clearly reported to the industrial engineer. Finally, a frequent practice is to apply dimensionality reduction techniques such as PCA to the data in the preprocessing stage. One of the techniques is the removal of some of the resulting components based on the Proportion of Variance Explained (PVE), as suggested in a review of machining monitoring systems based on artificial intelligence process models [1]. At the same time, it follows from our previous works that *a priori* transformation of the data with PCA combined with standard dimensionality selection techniques such as PVE may in some cases be suboptimal [24,25]. Thus, a method for making a decision on whether to use PCA for data pre-processing is proposed in this work.

Importantly, no error rates are investigated in this study, unlike most works on modeling industrial processes. Instead, the emphasis is placed on proposing methods that can help assess other attributes of the prediction model: its interpretability and the credibility and the sufficiency of the data set (i.e. whether further experiments are needed to increase its prediction credibility).

The remainder of this work is organized as follows:

- A survey is first presented in Section 2 of the state-of-the-art techniques applied to model machining processes and for visualization of input features that impact on dependent features.
- Then, a reference industrial process, a deep drilling process, is described in Section 3, together with a data set collected expressly for that purpose.
- Section 4 contains a discussion of the techniques for easy investigation of the process data.
- The proposals in Section 5 are intended to show the impact of tool settings on the output feature of interest, while simultaneously revealing the credibility of individual predictions.
- The role of dimensionality reduction, its impact on model interpretability and the search for a balance between model accuracy and interpretability is discussed in Section 6.
- Finally, the conclusions are presented and future lines of research are proposed.

2. Survey of visualization techniques applied to machining processes

Most studies on machine-learning models running under industrial conditions provide only limited visualization attributes. One reason for this lack of research into visualization techniques is perhaps that the machining processes generate multidimensional data, which makes the visualization task much more challenging. Most research on modeling machining processes, whether using analytical models or artificial intelligence techniques, only involves two dimensional graphics, to illustrate the relationship between variables: an input variable of the machining process (X-axis) and an output variable (Y-axis). The authors are therefore seeking to demonstrate a clear relation between one pair of variables. However, although these representations shed some light on the dynamics of the cutting process, they provide a very limited panorama of multivariable processes, creating specific sets of rules (knowledge) that will be difficult to generalize and to apply under alternative industrial conditions. So as to include more information, the X-axis of the graph in some examples is split into more than one region, covering up to 4 inputs [35,57]. This inclusion of further information makes it possible to discuss the relation between 2–4 inputs and one output with only one graph. However, this type of graph requires a less-

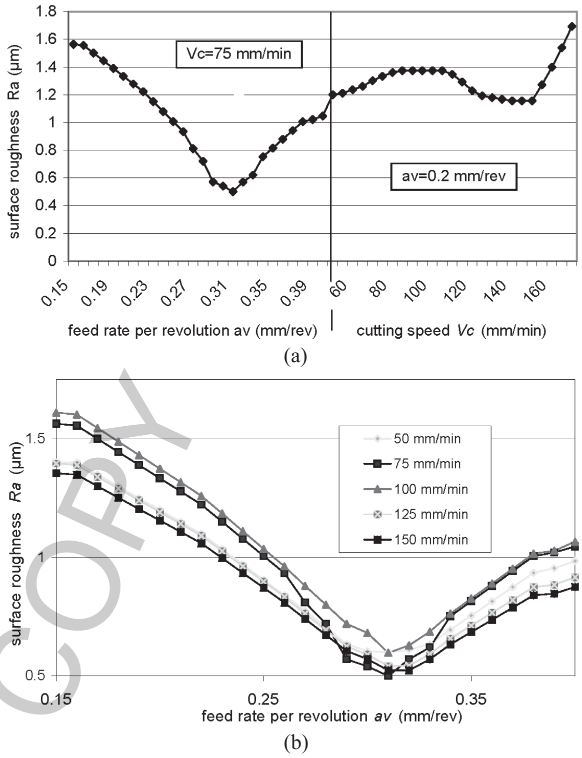

Fig. 1. Examples of 2D plots showing the influence on surface roughness of feed rate per revolution and cutting speed in a drilling process.

complex relation between variables and not very extensive data sets, otherwise the graph is packed with too much information [57]. A virtual experiment is therefore proposed to illustrate the different type of graphs presented in the bibliography, where the surface roughness of the machined workpiece (Ra) is measured after being machined under different cutting conditions that include the variation of 2 of the main cutting settings: feed rate per revolution and cutting speed. A more detailed explanation of the physical meaning of these settings will be done in Section 3. Figure 1(a) shows an example of such graphs [35,57]: the feed rate per revolution (*av*) and the cutting speed (Vc) are represented on the X axis and the surface roughness (Ra) for a given cutting process on the Y axis. The resulting graph allows us to discuss the influence of these 2 process parameters on surface roughness.

Another common way of including further variables in a 2D graph is to draw more than one line in each graph [37,38,43]; each line representing the relation between one input and one output for a certain values of a second input. Different lines mean different values of the second input. Figure 1(b) shows an example:

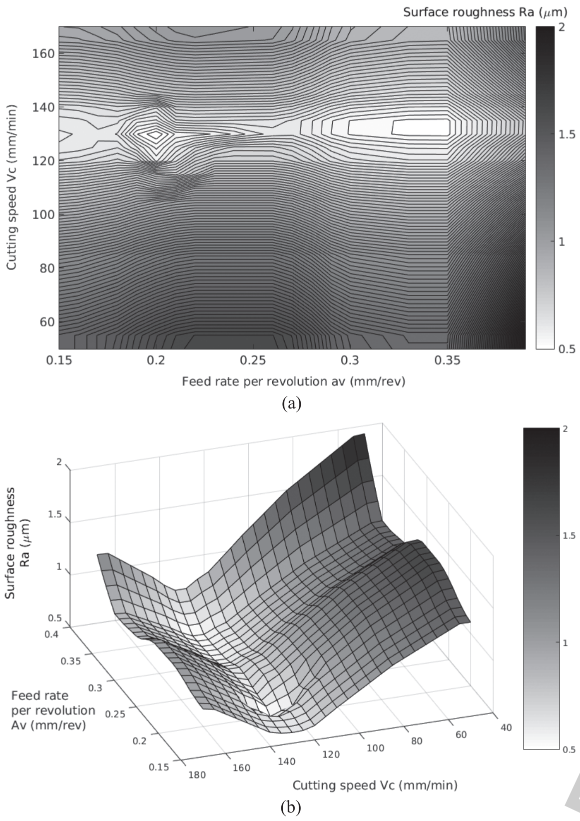

Fig. 2. Examples of surface plots, 2D (a) and 3D (b), showing the influence of feed rate per revolution and cutting speed on surface roughness in a drilling process.

the X-axis represents the feed rate per revolution, the Y-axis represents the surface roughness and each line represents a certain cutting speed. The use of color or grey codes in Fig. 1(b) differentiates the data to facilitate an in-depth discussion of the influence of process inputs on roughness quality. Finally, the most complex 2D graphs use both the X and the Y-axis for two inputs and draw the predicted value of a given output on a colored or grey scale [9,53]. This representation can only be used to show the usability of a prediction model, to provide information on the best cutting conditions for the process engineer in an industrial setting. Figure 2(a) shows an example of these graphs: the X-axis represents the feed rate per revolution, the Y-axis the cutting speed, and the grey scale shows surface roughness.

Although these last two examples of 2D graphs, Figs 1(b) and 2(a), might provide a rough idea of the relation between three variables, these relations are only clearly shown in 3D graphs: two process inputs on both the X and the Y-axes and one output on the Z-axis. 3D graphs are mainly used in two ways in the current literature. The first is to analyze the influence of two cutting parameters on a process output, usually surface roughness [14,30,37,54,60]. This kind of representation needs an extensive experimental test of different cutting conditions to generate a homogenous 3D surface, which is a very expensive solution. The second option is to show how a process engineer can find the best combination of two process inputs for a desired value of an output, because the 3D surface is generated by a prediction model and not by real experiments [7,10,26,51,59,60]. An example is shown in Fig. 2.b: feed rate per revolution and cutting speed are represented on the X and the Y-axes, respectively, and surface roughness on the Z-axis. This graph is illustrative in a discussion on the influence of these 2 process parameters on roughness.

It can be seen that there is no difference between the information shown in this graph and the information shown in the last example of the 2D graphs: Figs 2.a and 2.b respectively.

However, any cutting process is a complex multivariable process that depends on more than 4 inputs with complex relationships between the variables [5]. Different authors have proposed different approaches with no clear solution to the analysis of a cutting process in which over 4 process parameters are simultaneously considered. Even more importantly, such plots do not actually test the significance of the dependency between input variable(s) and output variables, irrespective of the number of input variables under consideration. 2D graphs have in some cases been drawn to address both high dimensionality and the search for the influence of input variables on output variable, where the X axis represents all the process inputs and the Y axis, the weight of each input in the variation of an output [46]. The same conclusions can be obtained with a table that shows the results of an ANOVA analysis [59,65]. A similar solution is based on PCA and the construction of a scree plot to analyze the number of features that influence the Artificial Intelligence (AI) model, although the authors of the study emphasized that the analysis of such a plot is partly subjective [25]. Although these solutions might help to decide which inputs are less significant and of negligible influence in a study case, they do not illustrate the type of relationship between inputs and output.

Although they have very seldom been used in machining processes so far, dimensionality reduction techniques [56] involve more complex visualization techniques than correlation plots and they are also promising avenues with which to develop and to analyze prediction or diagnosis models for industrial tasks in the manufacturing industry. For example, a dimensionality reduction technique, called t-Stochastic Neighbor Embedding, has been proposed to detect a chatter fault for the rolling process in the manufacturing industry [52]; it projects all feature vectors on a visual 2D map that describes the vibrational states of the rolling process. Decision trees have been used in fault diagnostics for feature extraction or dimensionality reduction in the vibration analysis of roller bearings in rotary machines [58]; in this case the reduction of dimensions is necessary, because the analysis of the vibrational spectrum of the machine provides too many inputs. In the same way, a discrete wavelet decomposition procedure has been used to reduce dimensionality in acoustic emission signals for grinding-wheel-surface-condition diagnosis [44]. Finally, the use of scatter plots has also been proposed for the design of cellular manufacturing systems [28]. This proposal involves the construction of a correlation matrix. It then uses PCA to find the eigenvalues on the correlation matrix and, finally, it creates a scatter plot for cluster analysis creating machine groups and part families, while maximizing the correlation between elements. The only example of scatter plots used in machining operations refers to the analysis of cutting forces and torques that depend on cutting parameters, such as feed rate and cutting speed for face-milling operations [37]. The scatter plot is very limited because the machining tests use only 3 inputs and analyze up to 5 outputs; an example that will be discussed later in Section 5 in comparison with this work.

The conclusion that came after our survey was that most of the existing research on prediction models for machining processes used different types of correlation plots to show the relationship between inputs and outputs in such processes [7,9,10,14,26,30,35,37,38,43,51,53,54,57,59,60]. Only very few works used other techniques rather than correlation plots, such as scatter plots [37] and scree plots [25]. Nevertheless, recent research projects on other tasks in the manufacturing industry have used complex visualization or dimensionality reduction techniques [28,37,44,52,58]. These papers are usually on vibrational or acoustic analysis, in which many statistical variables can be defined, and dimensionality reduction is a critical and mandatory task. On the other hand, due to the high cost of experimentation and the complex relation between inputs, most of the existing works related to machining processes take no more than 7 inputs into account [7,10,14,26,30,35,37,43,51,53,54,57,59,60,65] and very few extend their research to a broader group of process inputs [9,25,53]. It should be remarked, that, in most cases where the research into machining processes involves vibrational analysis, very few variables extracted from vibration signals are considered [10,26,30]. Therefore, most of the proposed prediction models for machining processes use very few process inputs and visualization techniques and little data analysis, concentrating their efforts on the accuracy of prediction models. Moreover, it remains unclear whether to use dimensionality reduction where there are several input variables. Dimensionality reduction, if done, is used to replace the original data with transformed data [25]. However, whether there is actually any need for such transformation, which inevitably reduces interpretability of the prediction models, is not addressed.

Finally, the figures relating one or many variables to another variable do not respond to the need for an evaluation of the significance [9,30,35,37,38,43,54,55,60]. The question remains whether the trends observed in the figures and, in all likelihood, based on a limited number of experimental records are significant and if so, whether they are equally significant. A further aspect of this issue is whether the collection of extra data could increase the confidence of inference based on the input variables and considering the expected result of the process. This could provide a basis for iterative data collection i.e. performing additional experiments aiming to collect extra data records for only some of the tool settings i.e. tool settings for which the confidence of process result prediction is reported by a model to be limited.

3. A survey of state-of-the-art deep drilling techniques

Our data set in this research describes the deep drilling of steel. Deep drilling of steel components is an especially interesting industrial process, due to the broad range of steel products with deep holes, such as coolant circuits for the manufacture of moulds and dies and boreholes for knockout pins. Drilling is a cutting process, in which the cutting tool rotates at fixed speeds while moving along the axis of the tool. Milling is usually the main machining process to generate a complex geometry, but drilling is necessary to create channels or any cylindrical geometry in the component. Drilling is in most cases more critical than milling, because the chip removal process is aided by the vertical helices of the drill, a complex process that requires fluids circulating under high pressure that displace the chips from the open side of the cutting edge, unlike when milling, where the waste moves away from the trailing edge of the cutting tool. Deep drilling is a critical drilling operation, because drill length further complicates the chip removal process and the cooling of the tool tip; it is usually defined by considering the length-to-diameter ratio of the drilled hole: drill lengths longer than 3 times the drill diameter are considered deep drills with conventional cutting fluid, because almost no cutting fluid reaches the drill tip under such conditions [36]. For instance, the limit for Minimum Quantity Lubrification (MQL), a cooling technique developed over the last 15 years, is at a depth no greater than 4 times the diameter [26]. In this cutting process, the cutting speed is the speed of the cutting tool in the drilling direction (*Vc*), the feed rate per revolution is the linear movement of the cutting tool in relation to the number of tool revolutions in the workpiece. The hole length (L) is the distance from the workpiece surface to the end of the hole, the hole diameter is the same as the tool diameter (D) and the cutting fluid is the fluid that is used to facilitate the chip and the heat removal during the drilling process. Figure 3 illustrates most of these parameters, which define the cutting conditions in a drilling process that the process engineer can set before the cutting process begins. In this case, process variables such as vibrations cannot be easily measured, due to the greater rigidity of the milling machine head on the Z-Axis along which drilling usually takes place [8]. Therefore, real-time process monitoring measures the axial force of the milling machine. This variable might give information on all the cutting parameters that the process engineer cannot establish, when designing the cutting process, such as tool wear, material heterogeneities, etc.

Different approaches have been considered to predict surface roughness in drilling operations [5], but only AI approaches are considered in this research. The most widespread AI technique is the Artificial Neural Network (ANN), although other techniques include fuzzy logic, Bayesian networks, genetic algorithms and support vector machines as outlined in some recent reviews [11,13,61]. Most of these works refer to cutting conditions that cannot be considered deep drilling or do not use MQL techniques. The use of MQL in deep drilling operations is fairly recent, which means that most of the research on this topic is focused on understanding the physical behavior of this process, such as tool wear [30], tool life [19,31], chatter phenomena [47] and the effect of MQL and cutting conditions on surface quality [15,64], rather than trying to use AI techniques to model the process. Datasets related to deep drilling are therefore very seldom found. Only a few study cases appear in the bibliography: fuzzy logic has been used to predict surface quality on MQL deep drilling of aluminum [49], evolutionary algorithms and Bayesian networks for deep drilling of steel [10,26] and better cutting conditions in deep drilling of steel with conventional flood cooling [29]. In the case of deep drilling of steel under MQL conditions, accurate models for surface prediction have been built with ANNs [26] and Bayesian networks [10], although the tuning of the ANN parameters is a very sensitive process [2,26,27]. The size of these data sets is also small, in line with datasets describing other manufacturing processes like milling or turning as explained before, and varies from 36 [12] to 170 [10], with some examples in between [29,49].

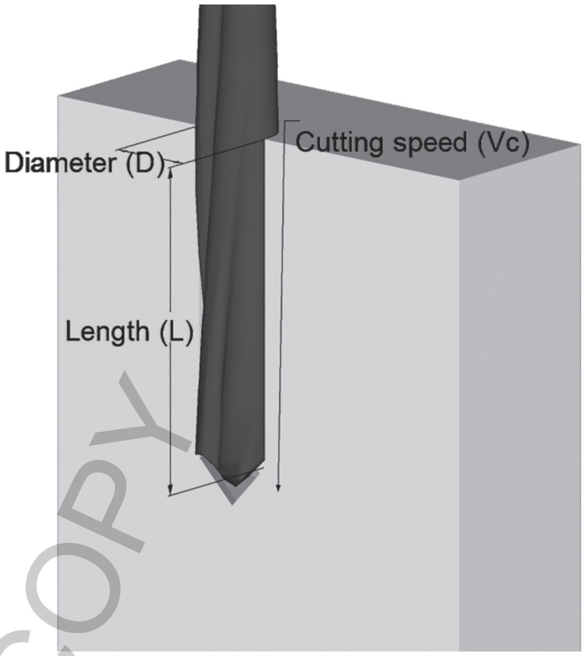

Fig. 3. Scheme of a drilling process.

The dataset used in this work has previously been presented elsewhere [10]. The drilling tests were performed on two different milling centers: one for MQL tests and the other for traditional coolant tests. In both cases, the blank material was F114 steel. In view of the industrial application of knockout pins, two hole diameters were chosen for testing: 5 and 10 mm. Two hole lengths were tested for each diameter: 5 and 8 times the diameter (5xD and 8xD). Tools from two

Table 1
Description of the variables used in the drilling data set

Variable (units)	Range	Relationship and nature
Tool (none)	1,2,3,4	Fixed by the process engineer
Diameter D (mm)	5,10	Fixed by the process engineer
Hole length L (mm)	25,40,50,80	Fixed by the process engineer
av ($\times$ 10−2 mm/rev)	10,12,14,15,20,25,30	Fixed by the process engineer
Vc (m/min)	70,80,90,100,125,130,156	Fixed by the process engineer
Coolant type (none)	1 (traditional), 2 (MQL)	Fixed by the process engineer
Axial force AF (N)	502–3012	Measured during cutting process
Productivity (cm^3/min)	8.7–93.7	Fixed by the process engineer
Roughness Ra (μm)	**0.27–4.96**	**Output, measured after the cutting process**

tool providers were tested to assure different geometries within them: HAM and Mitsubishi. Three cutting parameters were considered: cutting speed (*Vc*), feed rate per revolution (*av*) and coolant type. A second group of variables takes into account the difference in the cutting tools and the hole geometry: the type of tool (*Tool)*, tool and hole diameter (*diameter*) and hole length (*length*). These variables define the productivity as the material removal rate or the quantity of metal cut by time unit. Finally, the axial force measured in the milling head was included in the dataset. As the tests were performed along the Z-Axis of the milling centers, the Z-Axis feed motor consumption provided the axial cutting force (*AF*) during the drilling operation, because both variables were proportional [10]. Once the drilling tests had been performed, roughness was measured on the inner side of the holes in accordance with ISO standard 4288:1996 [34] and the Ra parameter was calculated. Surface properties play an important role in the performance of a finished part. They have an enormous influence on several relevant characteristics of the final product such as dimensional accuracy, friction coefficient, wear, thermal resistance, electric resistance, fatigue limit and behavior, corrosion, post-processing requirements, appearance and cost [53]. The *Ra* parameter is the arithmetic average of the vertical deviations (y) from the nominal surface converted to an absolute value for a specified distance over which the surface deviations are measured (*Lm*), as shown in Eq. (1). The Ra parameter, used in academic testing and industrial production, is the most common way of evaluating the roughness of machined workpieces [5].

$$Ra = \int_{0}^{Lm} \frac{|y|}{Lm} dx \tag{1}$$

All the cutting conditions used in the drilling tests belong to the cutting range proposed by the tool manufacturer. The drilling tests included 90 different cutting conditions and the tests were repeated to increase the amount of data. Thus, a data set of 165 records was obtained, less than foreseen because the acquisition procedure failed in 15 cases, delivering incomplete records. Each record is composed of the 8 variables described above. Table 1 shows the main information on the variables: units, variation range and origin of the data (measured or fixed by the process engineer before the cutting process). The output variable, surface roughness, is shown in bold in Table 1. Although this data set includes 8 inputs, one of them (productivity) is calculated on the basis of other variables. Hence, it can be considered redundant and will only be used in the data analysis presented in Section 4 but not in further research summarized in Sections 5 and 6, as will be explained later on.

4. Visualization of independent variables

Visualization of multidimensional data is a challenging task. The investigation of individual independent variables, frequently takes place followed by the selection of some of them, before generating the visualization.

Hence, the first stage is the use of visualization to augment data analysis and preprocessing. Even though many variables are observed in the data, some techniques can be applied at this stage to analyze multidimensional data.

First of all, a star plot can be developed, which helps to identify typical relations in the set and even more importantly to identify possible outliers. An example of a star plot for a selection of 50 records present in the drilling data set and sorted by ascending roughness values is shown in Fig. 4. This form of plot shows individual records in the data set. Each separate plot in the figure shows the values of the individual independent parameters in the form of separate sections of a

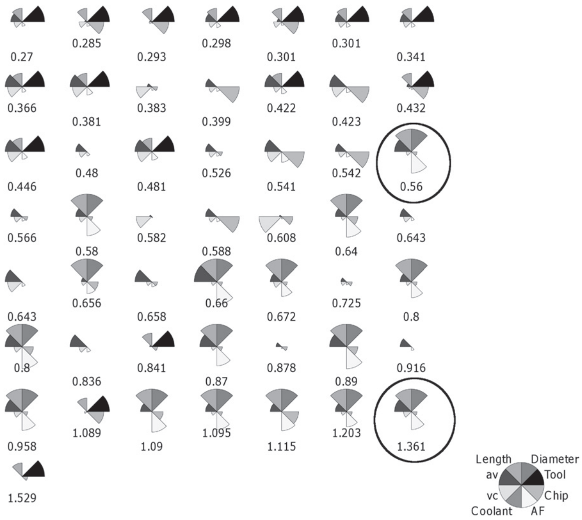

Fig. 4. A star plot showing the experimental settings and their impact on the roughness value. The interpretation of each section of individual pie charts is shown at the bottom of the plot. *Vc* denotes cutting speed, *av* – feed rate per revolution and *AF* – axial cutting force.

pie chart. The size of each section reflects the relative value of the parameter that is visualized. A larger section means a relatively larger value of the parameter. The meaning of each section is shown in the corner. In addition, the corresponding roughness value is shown below each plot.

Hence, this plot is particularly useful when the data are scarce, which is frequently the case as experimental data are difficult or very expensive to obtain. This form of figure is particularly useful to analyze the data at the level of individual records. In the case under analysis, it clearly follows from the figure that larger diameter values and larger axial forces are associated with higher roughness. Moreover, they are usually accompanied by a deeper hole length.

Moreover, the combined impact of the values of all independent variables on the feature of interest is seen. In particular, this provides the basis for additional verification of the data and identification of possible outliers. For instance, we see that large diameter values in general yield high roughness values.

However, the tool settings that produced the roughness of 0.56 μm (upper plot marked with a circle) are in fact very similar to the settings, typically resulting in a roughness of 1 μm or even more. In particular, the tool settings for the latter record are virtually identical to those present in the penultimate record (lower plot marked with a circle), having a roughness value of 1.361 μm. The record with an unlikely and small roughness value of 0.56 μm could undergo additional verification. The analysis of star plots is especially promising when a limited number of records constitute the data set. However, the question regarding possible approaches to larger data sets arises. Obviously, these datasets will involve an aggregated visualization of the data and their tendencies. Both correlation plots and matrices of these plots can be developed for both large and limited data sets. The informational value of such matrices may be increased by combining them with histograms, and correlation coefficients. An example of a matrix in the latter form is shown in Fig. 5. First of all, the values of individual independent variables are shown on the main diagonal histograms, which improve their readability, especially when they depict hundreds of experiments. More precisely, the distribution of the values of each parameter becomes clear. Moreover, linear regression is applied to individual variables to develop the best linear dependency between the paired variables under analysis. Finally, the regression coefficients are shown above the main diagonal, clearly depicting the extent to which the value of one variable explains the value of another variable.

Independent variables with the largest linear impact on the feature of interest, which in this case is roughness Ra, can be identified in Fig. 5. In addition, the relation between pairs of independent variables can be easily investigated. In this way, the relation between the values of individual variables can be observed. This includes typical clusters of values. Moreover, discrete and continuous variables can be easily determined. Some clear conclusions arise directly from this kind of representation: there is strong dependency between av and productivity (marked by a solid ellipse and a (1) in the figure). Moreover, larger tool diameters mean larger cutting forces (marked by ellipse (2)).

These results are expected by a process engineer and their visualization would help to guarantee that the experimentation is correct. Other results clearly show that the drilling process is a complex task: in many cases, smaller diameters imply higher productivities than larger tools, or the relation between *Vc* and roughness *Ra* is not clear, apart from the fact that some outliners can, in this way, be easily identified (ellipse marked with (3)). In the case under analysis, it can be observed that *Tool* is the variable with the largest impact on *Ra* from among all those under analysis, which is shown by the coefficient equal to 0.34. At the same time, the correlation between diameter and axial force is the largest of the other variable pairs and is equal to 0.92. In other words, a very strong dependency between diameter and axial force is shown. However, the

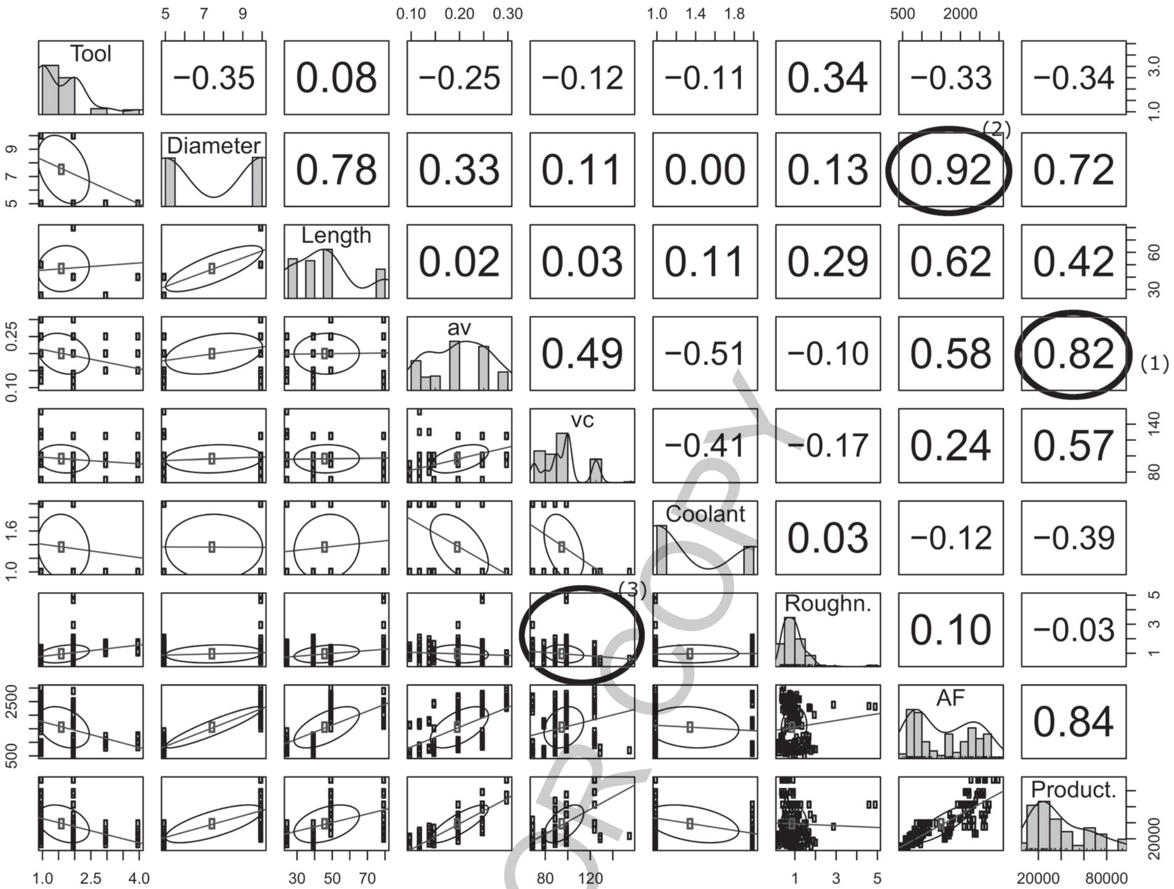

Fig. 5. A scatter plot matrix showing combinations of individual variables. Variable names and histograms of these variables are shown on the main diagonal of the matrix. The correlation coefficients for each variable pair are shown above the main diagonal, while the correlation plots for the same pairs are provided below. Vc denotes cutting speed, av - feed rate per revolution and *AF* – axial cutting force.

figure shows that only two diameter values were used in the experiments. An observation that confirms the merits of the visualization applied at this stage to the data set.

It follows directly from the analysis depicted in Fig. 5 that there is no single independent variable that shows a strong linear impact on the feature of interest; in this case, roughness. Should there be such an independent feature, it would have a correlation coefficient close to one, as shown in the Roughness column in the matrix shown in Fig. 5. It is worth noting here that productivity is calculated on the basis of other variables. Hence, it can be considered redundant and will not be included in the remainder of the analysis in this study. Finally, it may be noted that other authors have found similar, but more limited results with scatter plots [37], because they consider smaller numbers of inputs in the machining process.

5. Modeling and visualization of the roughness prediction process performed with raw data

5.1. Development of prediction models

Questions arise over the extent of the combined impact of the independent features on roughness and which features are the most important from this perspective. Moreover, a further point is whether the data that has been collected is sufficient for modeling purposes and what the credibility of the predictions are i.e. whether the scale of uncertainty regarding process results is the same for all process settings.

Various machine learning techniques [20] can be used to investigate this impact. These include neural networks [3] and most notably multilayer perceptrons [24,25], decision trees [59] and random forests. While neural networks enable nonlinear dependency between the input signals and output signal(s), the vi-

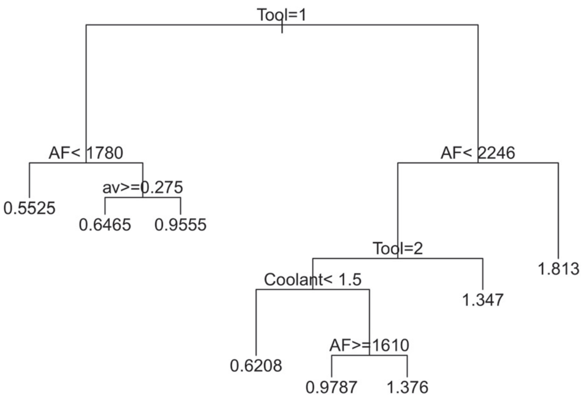

Fig. 6. A regression tree showing the prediction of roughness value based on the values of independent variables. *Vc* denotes cutting speed, *av* – feed rate per revolution and *AF* – axial cutting force. If a condition in the node is satisfied, the left branch is selected. For instance, if Tool is 1, then the condition $AF < 1780$ is verified. Every leaf of the tree contains the actual roughness value predicted for all process settings represented by the sequence of conditions followed from the tree root. As an example, if Tool $=$ 1, $AF >= 1780$ and $av >= 0.275$, *Ra* is predicted to be 0.6465 μm.

sualization of the mapping function that they represent is a difficult task in a multidimensional space. Similarly random forests – a technique using a group of regression trees that can arrive at a decision concerning roughness prediction, suffers from a similar inherent complexity.

More precisely, visualization of a random forest would be impractical as it would mean visualization of 100 individual regression trees. This complexity hinders the visualization and understanding of the process setting mappings on roughness value that random forest performs.

When the investigation of the impact of individual independent features on the feature of interest is an issue, regression trees can be considered [7,59]. Frequently, they are constructed with the CART algorithm proposed by Breiman et al. [6] or C4.5 proposed by R. Quinlan and used, among others, by Elangovan et al. for condition monitoring of a single point cutting tool [18]. A sample regression tree constructed with the CART algorithm is shown in Fig. 6.

It is important to note here that the tree yields the roughness prediction for different *Tool*, *AF*, *Coolant* and *av* values. Other independent variables were not relevant enough to be included in the tree, even though they were considered in the tree development process. Each value in the terminating nodes (i.e. leaves of the tree) is the roughness that is predicted under different conditions. However, an inherent feature of the CART algorithm is that it relies on the parameters controlling the tree construction process. One example of such a parameter is the minimum number of training records needed to create a condition node i.e. an additional split in a multidimensional space of independent variables. Moreover, the algorithm does not use the concept of statistical significance [32]. The CART algorithm is based on excessive tree construction followed by tree pruning [48]. The ultimate structure of the tree is usually selected in the cross-validation process. Similarly, the selection of parameters for C4.5 requires the experimental verification of various settings and their impact on model accuracy [18].

This process makes the selection of the optimal tree-the tree minimizing roughness prediction error-possible, but provides no answer to the need for a more thorough investigation of the data. First of all, the statistical significance of individual splits is not investigated by the CART algorithm. In particular, it remains unclear whether more accurate predictions could be made, but with lower confidence settings. A positive response to this doubt would suggest further data acquisition to support more accurate and reliable roughness prediction.

Hence, to answer these needs another approach is proposed in this study: the use of conditional inference trees [32]. Unlike CART trees, Conditional Inference Trees (CIT) are created in one stage i.e. recursive partitioning of the input space is continued, as long as it is significant in terms of the criteria it establishes. It is not followed by a pruning process. In the case of drilling data, a CIT was first developed using a univariate test type and a minimum criterion set to 0.95. This tree is shown in Fig. 7. A major difference, in comparison with CART or C4.5 trees is that the significance of individual splits made in each of the CIT nodes is shown by p values. Lower p values clearly indicate high confidence splits, which would provide an industry expert with a better understanding of the prediction process performed by the tree. In the case under analysis, it also follows from the analysis of this tree that it is only capable of distinguishing between different roughness values for Tool 1, as all the data for other tools is grouped in one node. More precisely, the roughness prediction for other tools is based on returning the same value, irrespective of the tool settings, with a relatively large dispersion of values and some outliers reaching a value of 5 μm. It was therefore impossible to make more accurate credible predictions and the same average roughness value was predicted under all conditions for Tools 2, 3, and 4. Besides, detailed prediction was successfully made for Tool 1.

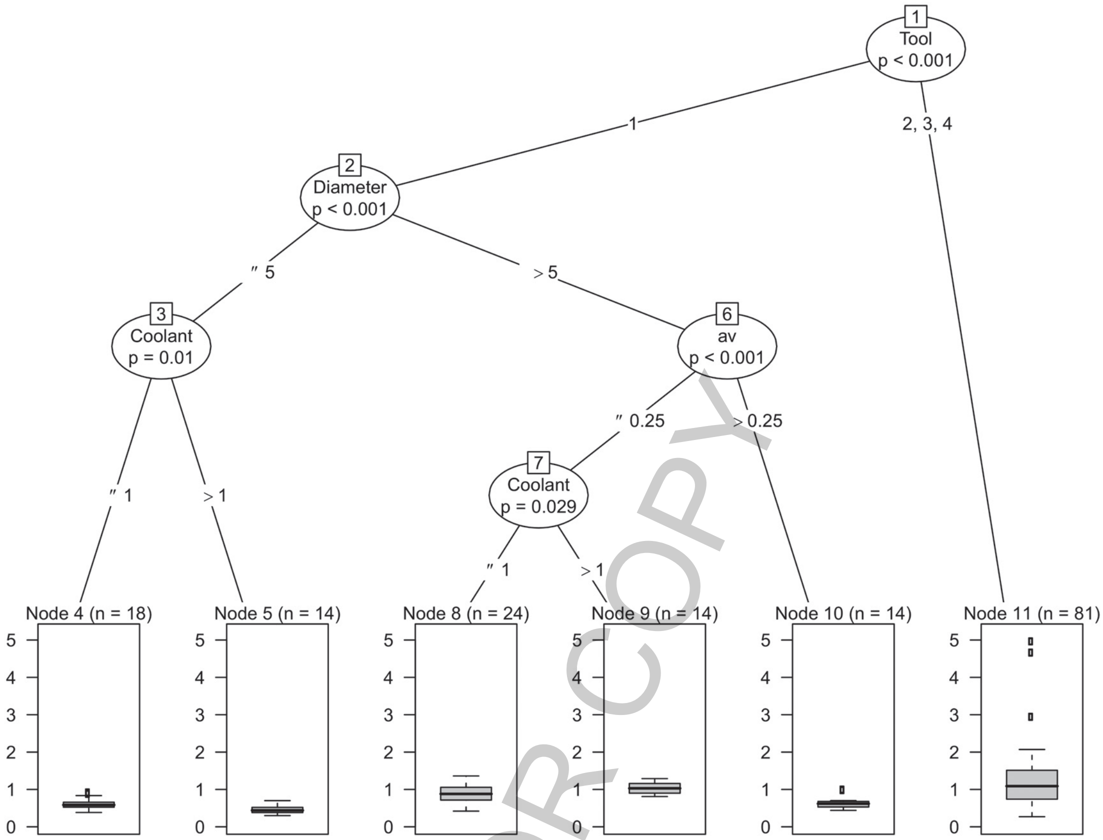

Fig. 7. A conditional inference tree with the minimum criterion set to 0.95. At each node, an average roughness value is shown with a horizontal line in the box. Moreover, the distribution of real records present in the training data and matched with each terminal node is also shown in the node. For instance, it can be seen that there were $n = 81$ records at node 11 i.e. the node was used for all records of Tools 2, 3 and 4. The average roughness value in this group was in excess of 1 μm, although the dispersion of values, which varied between 0.7 μm and 1.5 μm with some outliers reaching the value of 5 μm, was quite significant.

What is even more important, is the fact that the real roughness cases, behind each of the individual leaves (4,5,8,9,10) of the tree shown in Fig. 7, show relatively limited variability. Hence, the open issue is whether more accurate predictions could also be attained for other tools. In particular, the question arises as to whether more accurate prediction could be attained once the minimum criterion is reduced to 0.9.

Figure 8 shows the result of CIT construction with the aforementioned minimum criterion setting. First of all, it can be observed that the prediction for Tool 1, represented by the left subtree, is identical, as shown in Fig. 7. However, hitherto new dependencies, not known before, were revealed for Tools 2–4. In particular, for AF values below 2233.9 N, a division was made in node 12. As a consequence, roughness values were predicted for Tools 3 and 4, treated together, that differed from the roughness value predicted for Tool 2. While, all 81 records were grouped into one node for Tools 2–4, in the tree shown in Fig. 7, with one average prediction assigned, 5 different nodes may be observed with a high precision of roughness prediction for the same experimental cases, in Fig. 8. Moreover, in Fig. 8, significant dispersion of values is only observed for AF values exceeding 2233.9 N. There are only $n = 12$ cases in this category. All of which are represented by node 19.

It should be emphasized that a number of objectives may be reached by following this process for drilling and other machining data sets. First of all the prediction is made. Moreover, the credibility of individual predictions is revealed. Not only is the distribution of individual records mapped to each node shown, but also the statistical significance of space partitioning can be controlled. In the case under analysis, it is clear that the predictions made for Tool 1 are of greater credibility, than predictions for Tools 2–4, with the tree shown in Fig. 8. Simply put, these latter predictions

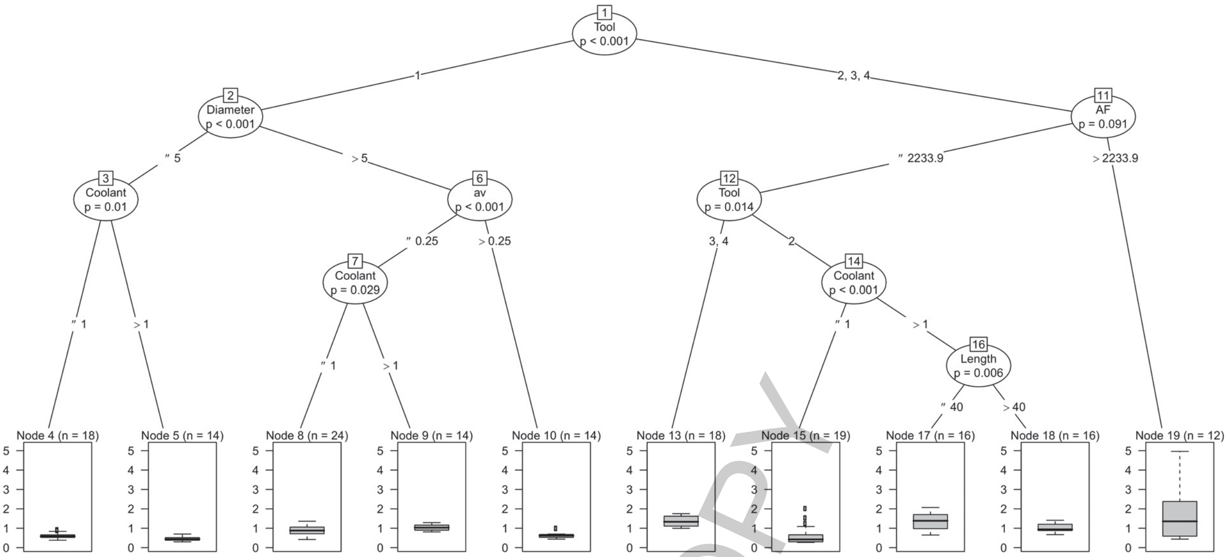

Fig. 8. A conditional inference tree with the minimum criterion set to 0.9. Lower significance results in additional splits compared to the tree shown in Fig. 7. These splits, represented by nodes 11 ($p = 0.091$), 12, 14 and 16, result in varied predictions of various roughness values for Tools 2,3 and 4, depending on the values of AF, Tool, Coolant and Diameter. Hence, prediction for Tools 2, 3, and 4 is more precise than in Fig. 7. However, the confidence level is lower for this part of the tree.

are only revealed when the statistical significance is set to a lower value. Moreover, it may be seen that the dispersion of actual roughness values was quite large for large AF values. Finally, it can also be seen that more accurate predictions of greater significance could potentially be made for some tool settings. However, more data would be needed for Tools 2–4, and especially for these tools and large AF values. A larger data set might confirm these dependencies also at the significance level of 0.95. Finally, such an analysis clearly identifies the impact of individual process parameters on roughness values, unlike the 2D and the 3D figures described in Section 2. These figures include no concept of significance and make no selection of actually relevant features from among those measured for the process of interest; a task that is usually entrusted to the expertise of the process engineer.

An important outcome of the analysis performed here is not only the selection of the features with a direct impact on the roughness value, but also the analysis of the sufficiency of the experimental data. In the case under study, it can be seen that some of the splits in the tree are based on a Diameter $\leqslant 5$. However, it follows directly from Table 1 that only two diameter values were present in the experimental data providing the basis for tree development: 5 mm and 10 mm. Hence, the split made at Diameter $= 5$ mm is quite arbitrary. In particular, new experimental data are needed for reliable use of the model with other diameter values e.g. 7 mm. It is important that the tree construction results do not suggest that other variables are relevant in terms of roughness prediction. For instance, hole length was not used in prediction process. This suggests that further experiments using new hole-length settings would not be likely to contribute to improvements in prediction accuracy and credibility. In particular, the figures that analyze the value of this variable against roughness, such as these discussed in Section 2, could be prepared, although they are not necessary, as roughness variability is explained by other input variables. This simplifies the visualization of the machining process in the multidimensional space, by performing feature selection.

5.2. *The automation of the planning of further experiments*

Importantly, further industrial experiments may be guided by the investigation of conditional inference trees. Let us propose Algorithm 1 to guide such a process. It constructs a conditional inference tree, C, based on available experimental data including tool settings, D, and process output data, P. It then identifies problematic regions, S, of an input space partitioned by the tree. Problematic regions refer to regions linked to the decision nodes of low confidence i.e. confidence lower than α. S also contains regions related to leaf nodes that are identified on the basis of high confidence splits, but which have a significantly large standard deviation of the target feature, which in this case is roughness.

Input: $D \subset \mathbb{R}^n$ - matrix of input attributes, $P \subset \mathbb{R}$ - vector of corresponding output features, $card(D)$=$card(P)$, T - the maximum number of additional experiments to perform, $(\alpha, c_{\min}, d_{\min})$ - parameters controlling data acquisition process, where α - minimum confidence level, $c_{\min}$ - minimum cardinality, $d_{\min}$ - minimum standard deviation
Result: $E \subset \mathbb{R}^n$ - tool settings to develop new experiments with, $card(E) \leq T$
begin
  C = DevelopCIF(D,P);
  S = FindProblematicRegions($C, \alpha, c_{\min}, d_{\min}$);
  A = AttributesUsedIn(C);
  $\Omega = \phi$;
  for $r \in S$ **do**
    for $i \in 1 : T$ **do**
      for $a \in 1 : card(A)$ **do**
        x_a=drawRandomAttributeValue(r, a);
      end
      $\Omega = \Omega \cup \{\mathbf{x}\}$;
    end
  end
  E = sample(Ω,T);
end

Algorithm 1 The selection of tool settings to perform additional experiments with. First, selection of regions of input space i.e. tool settings is made. Next, T random samples are drawn from candidate tool settings.

The identification of problematic regions is performed in a recursive manner by the *FindProblematicRegions()* procedure, defined in Algorithm 2. Importantly, leaf node and the region of input space to which it is linked can be included in a list of problematic regions, S, only when its cardinality is sufficient to exceed c_{min}. In other words, it is acceptable to have low cardinality nodes that have a significant value of standard deviation, exceeding the threshold value of d_{min} as long as the number of corresponding tool settings is too limited to justify further experiments. Details of problematic region selection are summarized in Algorithm 2. Once all regions are identified, a number of tool setting combination vectors is randomly generated for each region. Next, T random samples are drawn from candidate tool settings. In this way, a process engineer can balance the cost of additional experiments and the need for additional data by setting a T value. Moreover, the process of generating tool settings with which to perform additional experiments is automated. Importantly this process can be repeated in an iterative manner to minimize the cost of experiments.

6. The use of dimensionality reduction

Dimensionality reduction [41] is a frequent practice when processing multivariate data, so as to use it to

Input: C - conditional inference tree, $(\alpha, c_{\min}, d_{\min})$ - parameters controlling data acquisition process, where α - minimum confidence level, $c_{\min}$ - minimum cardinality, $d_{\min}$ - minimum standard deviation
Result: S - the regions of input space that require the collection of extra experimental data
begin
  if $\neg$*isLeafNode(C)* **then**
    if *confidence(C)* $< \alpha$ **then**
      return *region(C)* ;
    end
    else
      N = GetChildTrees(C);
      $S = \phi$;
      for $n \in N$ **do**
        $S = S \cup$ FindProblematicRegions($n, \alpha, c_{\min}, d_{\min}$);
      end
      return S;
    end
  end
  if *isLeafNode(C)* **then**
    if *recordCount(C)* $\geq c_{\min} \wedge$ *stdev(P(C))* $\geq d_{\min}$ **then**
      return *region(C)* ;
    end
    else
      return ϕ;
    end
  end
end

Algorithm 2 FindProblematicRegions() procedure i.e. the procedure performing the selection of regions of input space that require additional data based on a conditional inference tree or its subtree.

reduce data redundancy and by doing so, to improve interpretability. The most frequently used technique is PCA [23,40]. The use of PCA yields transformed data composed of the set of variables that are uncorrelated with each other, yet based on linear combinations of raw variables. Hence, another representation of the same data set is attained. The data in transformed form can be used as an input for various data processing techniques, including the regression tree construction algorithm. Typically, the transformed data serve to replace the original data. The use of PCA in this way as a feature extraction technique in the context of machining processes is described *inter alia* in [1]. Hence, the analysis will start from this option. However, the selection of features arising from dimensionality reduction including the selection of reduced dimensionality generally requires in-depth analysis. Appropriate selection of reduced dimensionality, going beyond the use of the scree plot [40] or a proportion of explained variance [41] may yield substantial performance gains [25,26].

The tree constructed with transformed data is shown in Fig. 9, in which a more complex mapping of process settings to their corresponding roughness values may be observed. However, the interpretability of the tree is largely reduced, when compared to the tree built with raw data. In this case, conditions in individual tree nodes are placed on components which are linear

Table 2
Definition of components

	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
Tool	0.000	0.000	0.000	0.705	0.527	−0.474	0.000
Diameter	0.000	0.000	0.000	−0.695	0.646	−0.312	0.000
Length	0.000	0.380	−0.923	0.000	0.000	0.000	0.000
Av	0.000	0.000	0.000	0.000	0.000	0.000	−0.999
Vc	0.000	−0.924	−0.381	0.000	0.000	0.000	0.000
Coolant	0.000	0.000	0.000	−0.142	−0.551	−0.822	0.000
AF	−1.000	0.000	0.000	0.000	0.000	0.000	0.000

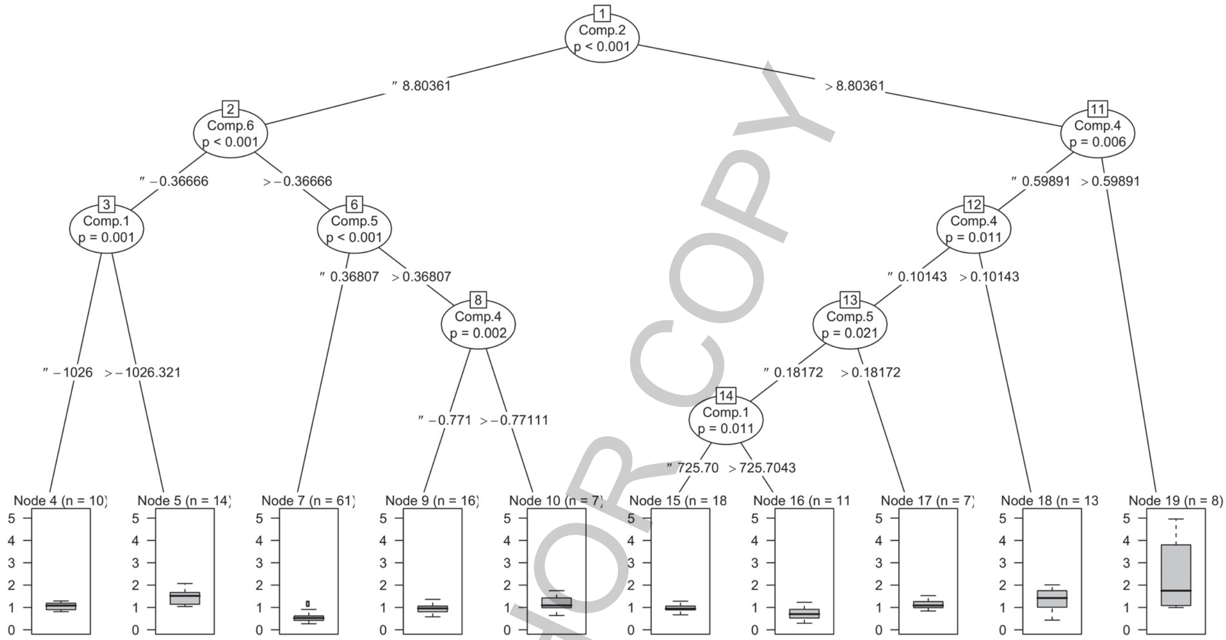

Fig. 9. A conditional inference tree built with the data arising from PCA transformation (minimum criterion set to 0.95). As the data were transformed, original independent variables such as diameter or coolant were replaced with PCA-based components. Moreover, the precision of the prediction in each leaf was improved. Unlike in Fig. 7, there were no nodes representing over 80 patterns i.e. predicting the same average roughness for a large group of experiments.

transformations of raw data. Moreover, different components are involved, namely Components 1, 2, 4, 5 and 6. Even though the definition of each component can be investigated (see Table 2 for component definitions), the interpretability of such a tree is obviously largely reduced.

Rather than only transformed data, our proposal is to use both raw and transformed data as an input for tree development. In the general case, this means that the raw data are replaced with transformed data that are generated through the use of a vector of reduction functions $c_i : R^n \rightarrow R^{k_i}$ and a vector of feature selection functions $f_i : R^{k_i} \rightarrow R^{j_i}$. This combined vector is of the form $[c_1, f_{1,}, \ldots, c_n, f_n]$. Our proposal is that this vector should include both $f_1(c_1(x)) = x$ i.e. the inclusion of a copy of raw attributes and the use of $c_2 : R^n \rightarrow R^n$, which is Principal Component Analysis. In this case, the data are transformed without reducing dimensionality, as *a priori* dimensionality reduction may diminish model performance. In particular, $f_2(x) = x$ (i.e. no selection of components before construction of the model). Moreover, $[c_1, f_{1,}, \ldots, c_n, f_n]$ can include other dimensionality reduction techniques such as kernel Principal Component Analysis [41]. This means that the tree construction process supplied with the data transformed with $[c_1, f_{1,}, \ldots, c_n, f_n]$, will select some of the components (i.e. transformed variables) and some of the raw independent variables depending on which of them yields the higher significance of splits made in the multidimensional space.

The result of this approach is shown in Fig. 10. It can be seen that successful predictions are made for Tool 1. This is based on the values of the variables Coolant, *av*

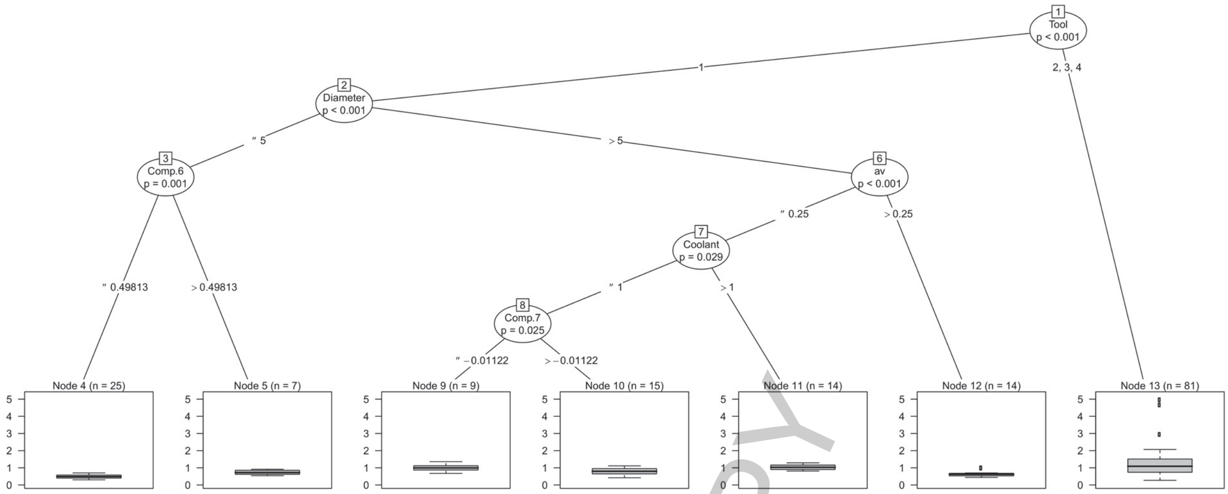

Fig. 10. A conditional inference tree constructed with both raw variables and principal components (minimum criterion set to 0.95). The tree construction algorithm selected raw variables, namely Tool (node 1), Diameter (node 2), av (node 6), and Coolant (node 7), but also some principal components for space splits. Unfortunately, better interpretability of the tree is combined with lower prediction accuracy, as all patterns belonging to Tools 2, 3 and 4 were grouped again into one leaf i.e. node 13.

and Diameter combined with the value of Components 6 and 7.

By using the approach proposed in this study, a process engineer may let the tree construction algorithm decide whether raw variables or components should be used for the individual splits in the tree. Then, a decision may be taken as to whether potential accuracy benefits arising from the inclusion of dimensionality reduction are sufficient compensation for the reduction in model interpretability. This decision can be taken on a case-by-case basis and may be different for different tools.

The composition of individual components is shown in Table 2. It can be observed that Component 6 is a linear combination of a number of variables with the most significant impact of Coolant, Diameter and Tool. Since Tool remains constant in this subtree and is equal to 1, the features having a significant impact on roughness for Tool 1 are Diameter (node 2), av (node 6), and Coolant. It is worth noting here that Component 7 is in fact the *av* variable. This example shows the way to attempt further improvements of roughness prediction model accuracy. By including more complicated transformation of an input, such as linear combinations of independent variables or even nonlinear transformation performed by multilayer perceptrons, there is a chance to improve roughness accuracy even further. However, the interpretability of the model is compromised.

7. Conclusions

Several issues have to be addressed before machine learning models can be applied in real industrial environments, such as machining processes in real workshops. Among them, investigation of the credibility of the models and their interpretability are of particular importance. Furthermore, the question of whether additional data collection could improve model quality and significance should be answered.

Many studies on machine learning model development have concentrated on the calculation of error rates. All too often, only limited attention is paid to the investigation of the data on which the model is based and its practical use. While such studies contribute to machine learning development, there is a need to combine them with the investigation of model deployment aspects. Hence, various plots have been developed in industrial studies to visualize data with 2D and 3D approaches. Some attempts have also approached higher dimensionality data. However, as our survey reveals, the most frequently used visualization techniques do not reveal the statistical significance of the trends they observe. Moreover, the impact of process settings on process results may be not clear. For instance, this impact may only be significant for some diameters or tool configurations. Finally, the visualization techniques that are usually applied do not evaluate the benefits of dimensionality reduction and its impact on the interpretability of the models.

Hence, the primary contribution of this work consists of the techniques that investigate multidimensional experimental data, upon which the prediction models are based. First of all, the techniques used to visualize both individual records and the relation between variables were selected. The use of conditional inference trees was then proposed to develop roughness prediction models for deep drilling processes. This approach has made it possible to assess the significance of predictions made for different tools and even for individual tool settings. As a consequence, a process engineer can differentiate between predictions made with lower credibility and those attained with higher levels of significance. In particular, this knowledge can be used to decide whether further experiments aimed at additional data collection are needed and if so, what tool and process settings should be used in these experiments to actually benefit from data acquisition. In addition, a formal procedure used to generate tool settings for a requested number of experiments has been proposed. Importantly, this approach addresses the fact that experimental data sets of limited size are the only feasible alternative, due to the cost of these experiments.

Finally, proposals have been advanced on the way a decision over the use of dimensionality reduction can be taken. Both model accuracy and interpretability can be taken into account in the proposed procedure. Hence, the decisions on the use of dimensionality reduction can be taken in accordance with the needs of the process engineer.

All these conclusions have been obtained from experimentation with a real data set that describes a deep drilling process. Deep drilling is of significant industrial importance in the production of moulds and dies for the automotive industry. The data set that was used in this study is representative of the number of records that can be reasonably acquired in industrial settings. Future research may include the investigation of various dimensionality reduction techniques, in the context of industrial data processing. In particular the problem might be related to the selection of the dimensionality reduction technique that best preserves the information needed to predict process output and model interpretability. Also, the implementation of the proposed visualization techniques in micro-machining processes of high industrial interest might also be explored, where current knowledge of the main cutting parameters and the experience of process engineers in such processes are still very limited.

Acknowledgments

This work was made possible thanks to the support received from Nicolás Correa S.A. and Fundación Tecnalia, which provided the drilling data and performed all the experimental tests. The authors would especially like to thank Mrs. Iraitz Etxeberria and Mr. German Rodriguez for their kind-spirited and useful advice. The authors would also like to thank Mr Michał Karwowski for releasing under open license his R scripts for processing the data and generating the visualizations of process data and prediction models, which were partly used in this study.

References

[1] J.V. Abellan-Nebot and F.R. Subirón, A review of machining monitoring systems based on artificial intelligence process models, *International Journal of Advanced Manufacturing Technology* **47** (2010), 237–257.

[2] H. Adeli and S.L. Hung, An Adaptive Conjugate Gradient Learning Algorithm for Effective Training of Multilayer Neural Networks, *Applied Mathematics and Computation* **62**(1) (1994), 81–102.

[3] H. Adeli and S.L. Hung, Machine Learning – Neural Networks, Genetic Algorithms, and Fuzzy Systems, John Wiley and Sons, (1995), New York.

[4] M. Ahmadlou and H. Adeli, Enhanced Probabilistic Neural Network with Local Decision Circles: A Robust Classifier, *Integrated Computer-Aided Engineering* **17**(3) (2010), 197–210.

[5] P.G. Benardos and G.C. Vosniakos, Predicting surface roughness in machining: a review, *International Journal of Machine Tools and Manufacture* **43**(8) (2003), 833–844.

[6] L. Breiman et al., Classification and Regression Trees, Chapman and Hall, 1984.

[7] A. Bustillo et al., Modelling of process parameters in laser polishing of steel components using ensembles of regression trees, *International Journal of Computer Integrated Manufacturing* **24**(8) (2011), 735–747.

[8] A. Bustillo, L.M. Plaza and M. Rodriguez, New Strategy for the Optimal Design and Manufacture of High Performance Milling Heads, *Revista De Metalurgia* **47**(6) (2011), 462–476.

[9] A. Bustillo et al., Avoiding neural network fine tuning by using ensemble learning: application to ball-end milling operations, *The International Journal of Advanced Manufacturing Technology* **57**(5) (2011), 521–532.

[10] A. Bustillo and M. Correa, Using artificial intelligence to predict surface roughness in deep drilling of Steel Components, *Journal of Intelligent Manufacturing* **23**(5) (2012), 1893–1902.

[11] M. Chandrasekaran et al., Application of soft computing techniques in machining performance prediction and optimization: a literature review, *The International Journal of Advanced Manufacturing Technology* **46**(5-8) (2010), 445–464.

[12] A. Çiçek, T. Kivak, G. Samtaş and Y. Çay, Modelling of thrust forces in drilling of AISI 316 stainless steel using artificial neural network and multiple regression analysis, *Journal of Mechanical Engineering* **58**(7-8) (2012), 492–498.

[13] A.K. Choudhary, J.A. Harding and M.K. Tiwari, Data mining in manufacturing: A review based on the kind of knowledge, *Journal of Intelligent Manufacturing* **20**(5) (2009), 501–521.

[14] J. Ciurana, G. Arias and T. Ozel, Neural Network Modeling and Particle Swarm Optimization (PSO) of Process Parameters in Pulsed Laser Micromachining of Hardened AISI H13 Steel, *Materials and Manufacturing Processes* **24**(3) (2009), 358–368.

[15] J. Davim et al., Experimental studies on drilling of aluminium (AA1050) under dry, minimum quantity of lubricant, and flood-lubricated conditions, *Proceedings of the Institution of Mechanical Engineers Part B-Journal of Engineering Manufacture* **220**(10) (2006), 1605–1611.

[16] A. Deleforge, F. Forbes and R. Horaud, Acoustic Space Learning for Sound-source Separation and Localization on Binatural Manifolds, *International Journal of Neural Systems* **25**(1) (2015), 1440003 (21 pages).

[17] D.M. Eddy et al., Model Transparency and Validation: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7, *Value in Health* **15** (2012), 843–850.

[18] M. Elangovan et al., Visual analysis of a cold rolling process using a dimensionality reduction approach, *Expert Systems with Applications* **38** (2011), 4450–4459.

[19] A. Filipovic and D. Stephenson, Minimum quantity lubrication (MQL) applications in automotive power-train machining, *Mechanical Science and Technology* **10** (2006), 3–22.

[20] P. Flach, Machine Learning. The Art and Science of Algorithms That Make Sense of Data, Cambridge University Press, 2013.

[21] A. Ghandar and Z. Michalewicz, An Experimental Study of Multi-Objective Evolutionary Algorithms for Balancing Interpretability and Accuracy in Fuzzy Rulebase Classifiers for Financial Prediction, *2011 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)*, 2011, 1–6.

[22] S. Ghosh-Dastidar and H. Adeli, Improved Spiking Neural Networks for EEG Classification and Epilepsy and Seizure Detection, *Integrated Computer-Aided Engineering* **14**(3) (2007), 187–212.

[23] S. Ghosh-Dastidar, H. Adeli and N. Dadmehr, Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection, *IEEE Transactions on Biomedical Engineering* **55**(2) (2008), 512–518.

[24] M. Grzenda, Prediction-Oriented Dimensionality Reduction of Industrial Data Sets, *Lecture Notes in Artificial Intelligence* **6703** (2011), 233–241.

[25] M. Grzenda et al., Improvement of surface roughness models for face milling operations through dimensionality reduction, *Integrated Computer-Aided Engineering* **19**(2) (2012), 179–197.

[26] M. Grzenda, A. Bustillo and P. Zawistowski, A Soft Computing System Using Intelligent Imputation Strategies for Roughness Prediction in Deep Drilling, *Journal of Intelligent Manufacturing* **23**(5) (2012), 1733–1743.

[27] K.J. Gurubel, A.Y. Alanis, E.N. Sanchez and S. Carlos-Hernandez, A Neural observer with time-varying learning rate: Analysis and Applications, *International Journal of Neural Systems* **24**(1), 1450011 (15 pages).

[28] W. Hachicha, F. Masmoudi and M. Haddar, Formation of machine groups and part families in cellular manufacturing systems using a correlation analysis approach, *The International Journal of Advanced Manufacturing Technology* **36**(11-12) (2008), 1157–1169.

[29] K. Hashmi, I. Graham and B. Mills, Fuzzy logic based data selection for the drilling process, *Journal of Materials Processing Technology* **108**(1) (2000), 55–61.

[30] R. Heinemann, S. Hinduja and G. Barrow, Use of process signals for tool wear progression sensing in drilling small deep holes, *International Journal of Advanced Manufacturing Technology* **33**(3-4) (2007), 243–250.

[31] R. Heinemann et al., Effect of MQL on the tool life of small twist drills in deep-hole drilling, *International Journal of Machine Tools and Manufacture* **46**(1) (2006), 1–6.

[32] T. Hothorn, K. Hornik and A. Zeileis, Unbiased Recursive Partitioning: A Conditional Inference Framework, *Journal of Computational and Graphical Statistics* **15**(3) (2006), 651–674.

[33] J. Huo, Y. Gao, W. Yang and H. Yin, Multi-instance Dictionary Learning For Detecting Abnormal Event detection in Surveillance Videos, *International Journal of Neural Systems* **24**(3) (2014), 1430010 (15 pages).

[34] International organization for standardization. ISO:4288:1996 (1996). Geometrical product specifications (GPS) – surface texture: Profile method – rules and procedures for the assessment of surface texture.

[35] S. Kasman and I.E. Saklakoglu, Determination of process parameters in the laser micromilling application using Taguchi method: A case study for AISI H13 tool steel, *The International Journal of Advanced Manufacturing Technology* **58**(1-4) (2012), 201–209.

[36] H. Kubota and H. Tabei, Drilling of a small and deep hole using a twist drill, *Transactions of the Japan Society of Mechanical Engineers, Part C* **62**(601) (1999), 3691–3697.

[37] E. Kuljanic and M. Sortino, TWEM, a method based on cutting forces – monitoring tool wear in face milling, *International Journal of Machine Tools and Manufacture* **45**(1) (2005), 29–34.

[38] A. Kumar Dubey and V. Yadava, Laser beam machining – A review, International, *Journal of Machine Tools and Manufacture* **48**(6) (2008), 609–628.

[39] M. Kwon, S. Kavuri and M. Lee, Action-Perception Cycle Learning for Incremental Emotion Recognition In a Movie Clip Using 3D fuzzy GIST Based on Visual and EEG signals, *Integrated Computer-Aided Engineering* **21**(3) (2014), 295–310.

[40] D.T. Larose, Data Mining Methods and Models, John Wiley & Sons, 2006.

[41] J. Lee and M. Verleysen, Nonlinear Dimensionality Reduction, Springer, 2010.

[42] Y. Li, C. Liu, J.X. Gao and W. Shen, An integrated feature-based dynamic control system for on-line machining, inspection and monitoring, *Integrated Computer-Aided Engineering* **22**(2) (2015), 187–200.

[43] C. Lia, M. Tsaia and C. Yang, Study of optimal laser parameters for cutting QFN packages by Taguchi's matrix method, *Optics & Laser Technology* **39**(4) (2007), 786–795.

[44] T.W. Liao et al., A wavelet-based methodology for grinding wheel condition monitoring, *International Journal of Machine Tools and Manufacture* **47**(3-4) (2007), 580–592.

[45] R. Lostado, R.F. Martinez, B.J. Mac Donald and P.M. Villanueva, Combining soft computing techniques and the finite element method to design and optimize complex welded products, *Integrated Computer-Aided Engineering* **22**(2) (2015), 153–170.

[46] S.S. Mahapatra and A.K. Sood, Bayesian regularization-based Levenberg – Marquardt neural model combined with BFOA for improving surface finish of FDM processed part, *The In-*

ternational Journal of Advanced Manufacturing Technology **60**(9-12) (2012), 1223–1235.

[47] I. Mehrabadi, M. Nouri and R. Madoliat, Investigating chatter vibration in deep drilling, including process damping and the gyroscopic effect, *International Journal of Machine Tools & Manufacture* **49**(12-13) (2009), 939–946.

[48] O. Maimon and L. Rokach, eds., Data Mining and Knowledge Discovery Handbook, A Complete Guide for Practitioners and Researchers, Springer, 2005.

[49] A. Nandi and J. Davim, A study of drilling performances with minimum quantity of lubricant using fuzzy logic rules, *Mechatronics* **19**(2) (2009), 218–232.

[50] D.L. Olson, D. Delen and Y. Meng, Comparative analysis of data mining methods for bankruptcy prediction, *Decision Support Systems* **52** (2012), 464–473.

[51] J.A. Perez, M. Gonzalez and D. Dopico, Adaptive neurofuzzy ANFIS modeling of laser surface treatments, *Neural Computing and Applications* **19**(1) (2010), 85–90.

[52] D. Perez et al., Visual analysis of a cold rolling process using a dimensionality reduction approach, *Engineering Applications of Artificial Intelligence* **26**(8) (2013), 1865–1871.

[53] G. Quintana, A. Bustillo and J. Ciurana, Prediction, monitoring and control of surface roughness in high-torque milling machine operations, *International Journal of Computer Integrated Manufacturing* **25**(12) (2012), 1129–1138.

[54] G. Quintana, M.L. Garcia-Romeu and J. Ciurana, Surface roughness monitoring application based on artificial neural networks for ball-end milling operations, *Journal of Intelligent Manufacturing* **22**(4) (2011), 607–617.

[55] M.C.P. Ramos et al., A Systematic Review of Research Guidelines in Decision-Analytic Modeling, *Value in Health* **18** (2015), 512–529.

[56] O. Reyes, C. Morell, S. Ventura, Evolutionary feature weighting to improve the performance of multi-label lazy algorithms, *Integrated Computer-Aided Engineering* **21**(4) (2014), 339–354.

[57] B. Samanta, W. Erevelles and Y. Omurtag, Prediction of workpiece surface roughness using soft computing, Proceedings of the Institution of Mechanical Engineers, *Part B: Journal of Engineering Manufacture* **222**(10) (2008), 1221–1232.

[58] V. Sugumaran, V. Muralidharan and K.I. Ramachandran, Feature selection using Decision Tree and classification through Proximal Support Vector Machine for fault diagnostics of roller bearing, *Mechanical Systems and Signal Processing* **21**(2) (2007), 930–942.

[59] D. Teixidor et al., Modeling pulsed laser micromachining of micro geometries using machine-learning Techniques, *Journal of Intelligent Manufacturing* **26**(4) (2015), 801–814.

[60] D. Teixidor et al., Optimization of process parameters for pulsed laser milling of micro-channels on AISI H13 tool steel, *Robotics and Computer-Integrated Manufacturing* **29**(1) (2013), 209–218.

[61] J. Torabi et al., A survey on artificial intelligence technologies in modeling of high speed end-milling processes, in: Advanced Intelligent Mechatronics, 2009. AIM 2009, IEEE/ASME International *Conference on. IEEE*, 2009, pp. 320–325.

[62] Y.-H. Tsai, J.C. Chen and S.-J. Lou, An in-process surface recognition system based on neural networks in end milling cutting operations, *International Journal of Machine Tools and Manufacture* **39**(4) (1999), 583–605.

[63] Y.B. Yang, Y.N. Li, Y. Gao, H.J. Yin and Y. Tang, Structurally Enhanced Incremental Neural Learning for Image Classification with Subgraph Extraction, *International Journal of Neural Systems* **24**(7) (2014), 1450024 (13 pages).

[64] J. Zang et al., Evolutionary optimization of machining processes, *Journal of Intelligent Manufacturing* **17**(2) (2006), 203–215.

[65] J.Z. Zhang and J.C. Chen, Surface Roughness Optimization in a Drilling Operation Using the Taguchi Design Method, *Materials and Manufacturing Processes* **24**(4) (2009), 459–467.

[66] E. Zio and G. Gola, A neuro-fuzzy technique for fault diagnosis and its application to rotating machinery, *Reliability Engineering and System Safety* **94** (2009), 78–88.