Abstract—This study presents a novel hybrid intelligent system, which focuses on the optimisation of machine parameters for dental milling purposes. The basis of this approach is hybridizing two bio-inspired algorithms, as Neural Networks with Genetic Algorithms for choosing and modelling the feature subset that best describes the operation conditions. These operating conditions are given as parameters for a dental drill machine. The aim of this approach is twofold: a feature selection process is carried out while the modelling of the operating conditions is achieved. The reliability of the proposed novel hybrid system is validated with a real industrial use case, based on the optimisation of a high-precision machining centre with five axes for dental milling purposes.

I. INTRODUCTION

Over recent years there have been a high increase in the use of artificial intelligence and Soft Computing methods to solve real world problems [1, 2]. Many of those applications have been reported: the use of Exploratory Projection Pursuit (EPP) [3, 4] and ARMAX for modelling the manufacture of steel components [5], EPP and Neural Networks (NN) for determining the operation conditions in face milling operations [6] and in pneumatic drilling process [7], genetic algorithms and programming for trading rule extraction [8] and low quality data in lighting control systems [9], feature selection and association rule discovery in high dimensional spaces [10] or NN and principal component analysis and EPP in building energy efficiency [1, 11].

It is known that the complexity inherited in most of new real world problems increases with the computing capabilities. Higher performance requirements with a lower amount of data examples is needed due to the costs of generating new instances, specially in those processes where new technology arises.

The optimisation process of machine parameters could significantly help to increase companies’ efficiencies and substantially contributes to costs reductions in preparation and setting machines processes and it also helps in the production process using new materials.

Nevertheless, the variables and parameters setting processes are a well-known problem that has not been fully resolved yet. Several different techniques are proposed in the literature. In [12], is used a Taguchi orthogonal array to optimise effect of injection parameters.

Moreover, in [13] the influence of operating parameters of ultrasonic machining is studied using Taguchi and F-test method. In [14] is researched as to improve the quality of the KrF excimer laser micromachining of metal using the orthogonal array-based experimental design method.
Conventional methods can be greatly improved through the application of soft computing techniques [15]. This approach proposes a wrapper style feature selection algorithm that makes use of NN for modelling the different outputs from the data set. The proposed algorithm includes a Genetic Algorithm (GA) as the metaheuristic for searching the best feature subset. Finally, the NN model obtained in the last step is used as fitness function to be optimised in the genetic algorithm.

NN have been used to find relationships between the mechanical properties of different real world problems [16]. NN have been also applied for identification of the parameters for operating conditions [17, 18].

This research is focused in determining the main parameters for dental drill machinery. The main objective is to find the most relevant feature subset; the second objective is to obtain a decision support system for designing the operation conditions of the dental drill.

The parameter to analyse is the **manufacturing time error**, which perhaps is the most relevant one. The rest of the parameters are left as future work.

### III. BIO-INSPIRED AND HYBRID SYSTEMS FOR IMPROVING MACHINERY PERFORMANCE

In order to achieve the objectives of this study the following proposal has been considered. A wrapper feature selection method is applied to reduce the input space dimensionality and to select the most relevant features for estimating the time error for manufacturing. It is known that for this kind of problems the wrapper approach for feature selection performs better than filter solutions [19, 20, 21, 22]. These studies proposed wrapper feature selection methods using genetic algorithms (GA) for dealing with the

### TABLE I. DIFFERENT FEATURES FROM THE PROCESS, THEIR UNITS AND RANGES

<table>
<thead>
<tr>
<th>Variable (Units)</th>
<th>Range of values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of work</td>
<td>One Locator attachment of cobalt-chromium, single-implant crown of cobalt-chromium, four unit implant bridge of cobalt-chromium, single crown of cobalt-chromium, and others</td>
</tr>
<tr>
<td>Thickness (mm)</td>
<td>8 to 15</td>
</tr>
<tr>
<td>Size of tool (mm)</td>
<td>T2 to T23</td>
</tr>
<tr>
<td>Radius (mm)</td>
<td>0.25 to 1.5</td>
</tr>
<tr>
<td>Tool</td>
<td>toric, spherical, plain, drill</td>
</tr>
<tr>
<td>Number of pieces</td>
<td>1 to 4</td>
</tr>
<tr>
<td>Revolutions per minute (RPM)</td>
<td>9,600 to 38,000</td>
</tr>
<tr>
<td>Feed rate X (mm. per minute)</td>
<td>0 to 3,000</td>
</tr>
<tr>
<td>Feed rate Y (mm. per minute)</td>
<td>0 to 3,000</td>
</tr>
<tr>
<td>Feed rate Z (mm. per minute)</td>
<td>75 to 2,000</td>
</tr>
<tr>
<td>Advance in angle (mm. per minute)</td>
<td>0 to 550</td>
</tr>
<tr>
<td>Advance in rotation (mm. per minute)</td>
<td>0 to 550</td>
</tr>
<tr>
<td>Initial Diameter tool (mm)</td>
<td>91.5035 to 110.4407</td>
</tr>
<tr>
<td>Initial Temperature (°C)</td>
<td>24.8 to 30.4</td>
</tr>
<tr>
<td>Estimate work time (s)</td>
<td>12 to 2,034</td>
</tr>
<tr>
<td>Time error for manufacturing (s)</td>
<td>-28 to -449</td>
</tr>
<tr>
<td>Difference of temperature in the machine (°C)</td>
<td>0.9 to 6.7</td>
</tr>
<tr>
<td>Difference of diameter of the tool (mm)</td>
<td>0.00080 to 0.11950</td>
</tr>
</tbody>
</table>

There are three main parameters to estimate: the time error for manufacturing, the difference of temperature in the machine and the difference of diameter in tool. All the variables and their ranges are presented in Table I. The time error for manufacturing is the difference between the estimated time by the machine itself and real work time - negative values indicates that real exceeds estimated time -.

#### A. The problem definition

The main aim of this study is to select the most relevant feature subset for estimating the operating conditions, that is, the machinery parameters. A second aim is to model the different parameters to achieve a bio-inspired decision support system for designing the operation conditions of the dental drill.

This study analyses one of the different parameters, that is, applies feature selection and NN modelling to estimate one of the machine parameters so it could be decided if the proposed approach allows achieving the objectives.

The parameter to analyse is the **manufacturing time error**, which perhaps is the most relevant one. The rest of the parameters are left as future work.

### II. HIGH PRECISION DENTAL DRILL MACHINERY

This study deals with manufacturing dental pieces (Fig. 1) and the selection of the main features involved in the operating conditions.

A dynamic high-precision machining centre with five axes was applied in this research. This real industrial use case is described by an initial data set of 115 samples obtained by a dental scanner in the manufacturing of dental pieces characterized by 15 input variables.

The input variables are the type of work, the thickness, the size of the tool, the number of pieces, the radius of the tool, the revolutions of the drill, the feed rate in each of the dimensions (X, Y and Z), the advance in the angle, the advance in the rotation, the initial tool diameter, the initial temperature and the estimated duration of the work.

Figure 1. Metal pieces manufactured by a dynamic high-precision machining centre with five axes
feature subset selection, that is, an individual is a feature subset. To evaluate each individual a modelling technique has been applied: the former proposed a lazy learning model as the K-Nearest Neighbour, the latter made use of a neural network (NN) method that iteratively fixes the number of hidden neurons. Different approaches in the literature should be analyzed, the cross-validation method among them [23, 24].

Different approaches in the way the NN is learnt have been studied. In [25] a GA approach to fingerprint feature selection is proposed and selected features are input to NN for fingerprint recognition, while in [26] a similar approach was applied to automatic digital modulation recognition. Moreover, this type of approach was reported of better performance instead of using statistical models [27], despite, Support Vector Machines (SVM) have been also used in conjunction with evolutionary feature selection to reduce the input space dimensionality [28, 29].

In this study we adopt a hybridized method of GA evolving the feature subsets and a NN model is learnt for modelling the desired output. The GA is a steady state approach with the percentage of elite individuals to be defined as a GA parameter. The algorithm has been implemented in Matlab [30], using the neural network toolbox.

In the following subsections each part of the approach is described. Next Subsection outlines the algorithm; Subsection III.B introduces the GA algorithm to evolve the feature selection, while in subsection III.C the NN models used are outlined. Finally, the scheme for evaluating each individual is presented.

A. The wrapper feature selection algorithm

The feature selection algorithm is totally detailed in Fig. 2 and Fig 3. The former outlines how each individual is evaluated, while the latter describes the used method for choosing the most suitable features.

To evaluate each individual it is proposed to apply a cross validation (cv) scheme so that the feature subset is used in training and validating different models. The goal is to obtain the mean of the mean square error (mse) values for all the testing sequences.

This mean of mse would resemble better the accuracy of the models that could be obtained with the selected feature subset. Moreover, introducing these cv scheme in the evaluation of each individual would allow to use the standard deviation also as a new second objective to be minimized as well as the mse. The idea of using multi objective algorithms is left as future work.

B. The feature selection method

GA is proposed for feature selection purposes. The typical steady state GA parameters, that is, an elite population is kept between generations. The number of individuals in the elite population is fixed as a GA parameter, also.

The individual representation is the string of Booleans, each Boolean marks if a feature is or not included in the feature subset that the individual proposes. The number of features N to be chosen is given as a GA parameter to the method. Future work includes improving the algorithm to also choose the most promising feature subset dimension.

Algorithm IND EVALUATION

Requires: I the input variables data set
Requires: O the output variable data set
Requires: ind the individual to evaluate, with its feature subset

Returns: model {the best model learned for ind }
Returns: mse ≈ 0 {the associated mean of Mean Square Error for ind }

Algorithm

indMSE = 0 {best MSE found in the cross validation}
for k = 1 to 10 do
{run the k fold in the cross validation scheme}
generate the train and test reduced feature data set
initialize the model indModel
train indModel with the train data set
indKMSE ← calculate the MSE for indModel el with the test data set
mse + = indKMSE
if k == 1 or indMSE > indKMSE then
indMSE = indKMSE
model = indModel
end if
end for
mse = mse / 10
return [model, mse]

Figure 2. Algorithm for evaluating each individual

Algorithm GA+ Feature Selection

Requires: I the input variables data set
Require: O the output variable data set
Require: N the feature subset size

Returns: FS ← {} {the best feature subset}
Returns: model {the model learned for FS}
Returns: mse{the associated mean of Mean Square Error for FS}

mse = 0
Generate the initial population, Pop
for all individual ind in Pop do
{run the k fold in the cross validation scheme}
generate the train and test reduced feature data set
initialize the model indModel
train indModel with the train data set
indKMSE ← calculate the MSE for indModel el with the test data set
mse += indKMSE
if k == 1 or indMSE > indKMSE then
indMSE = indKMSE
model = indModel
end if
end for

Figure 3. Algorithm for feature selection and modelling

The crossover operator is the classical one point crossover, that is, at the crossover probability (given as a GA parameter) a random crossing point is generated for each
crossing operation and the string of Booleans are interchanged at that point. In this study, the mean square error is used as fitness function; the use of different measures, the R2 among them, is left as future work.

The mutation is run for each position in the string of Booleans, and with the mutation probability the value is changed randomly provided that the final individual is valid. An individual is valid if the number of selected features is N after the mutation. The tournament selection is implemented.

C. Modelling with Neural Networks

The Matlab NN toolbox has been used in modelling the time error of manufacturing. In this study supervised fed forward NN’s with the Levenberg-Marquardt algorithm for learning the NN’s were used. Provided the number of features in the data set is reduced to size N, the number of hidden neurons was fixed for all the models. Previously, a test was carried out to determine this NN parameter.

D. A discussion on the different validation schemes

The fitness of each individual is calculated, as outlined before, using cv scheme. The main aim of this evaluation is to estimate the goodness of a feature subset in operating conditions, when the sample given has not been presented for training. It has been found relevant in those problems for which the data set includes quite few samples [31].

Nevertheless, many different schemes can be used, the leave one out, the k-fold cv or the 5x2 cv among them. The selection of the cv relies on the data set dimensions.

The mean error among the whole set of models learned is proposed as the fitness function; for all cases in this study, the validation has been calculated with the test data set.

However, it is worth noting that it is possible to use a validation data set for evaluating the mean error and so the fitness of the individuals.

Consequently, the cv should extract three data sets from the original one: the train, the test and the validation data sets.

Clearly, choosing the kind of cv will depend on the data set dimensionality; when not enough samples are available the solution will tend to leave one out cv without a specific validation data set. Automatic selection and developing of the cv scheme is also left as future work.

Interesting enough to mention is which of the models for the same individual is kept. In this study we kept the best model found, that is, the one with lower mse, but if multi objective were used then a different criteria should be selected.

IV. EXPERIMENTATION AND RESULTS

Analysing the complexity of the data set is the first task to complete as it would help to infer the parameters needed for the whole method, e.g., the number of hidden neurons in case of NNs. Once the data complexity is pre-processed, the complexity analysis is carried out.

We sorted these tasks in this order because the NN models will face the pre-processed data. It is known that some of the measures to determine the complexity of a data set are highly dependent on the data.

As far as it use this information to estimate the parameters of the method it thought that it could be more interesting to analyse the pre-processed data set.

Finally, the method detailed in Section III is applied and the feature subset and the corresponding model are obtained. In the following Subsections these three tasks are detailed.

A. Data set pre-processing

Two different pre-processing methods were applied. Firstly, the included missing values were filtered, that is, they were taken out from the data set. The second pre-processing method applied was normalizing the data set to zero mean and 1 standard deviation in order to train NN models.

B. The data set complexity

There are plenty of measures to evaluate the complexity of a data set. In this study, the measures and criteria detailed in [32] were applied. In Table II the complexity measure values are shown. As can be seen, the problem is highly uncorrelated and nonlinear, which forces to use relative high number of hidden neurons.

C. Modelling the time error of manufacturing

In this study, the experimentation carried out just tries to identify how the parameters of the whole method should be fixed. It is though that the relevance of their values could need further analysis. Nevertheless, the main problem is the time consumed in computing, that is, the learning of the NNs.

As a compromise of the precision and the time needed for training the NN was chosen with 12 nodes in the hidden layer, the rest of NN parameters were left as the default values from the toolbox. Similarly, the number of individuals was kept small and fixed to 10 individuals in the population.

The elite subpopulation was fixed to 2. The cv scheme is the 10-fold cross validation. The number of generations was also set to 100, a relative small value. Finally, the crossover and mutation GA operators probabilities were fixed to 0.75 and 0.25, respectively. The number of features to be chosen in each feature subset was fixed to 5.

The best M individuals found, with M fixed to 5 in this study, are analysed in depth. The objective is to evaluate the evolution of the learning process for all the folds in the cv.

Results from the experiment are shown in Table III and in Fig. 4. In Table III, it is shown the mean of the mse for the M best individuals found, that is, the best feature subsets.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Value</th>
<th>Measure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>0.0000</td>
<td>N2</td>
<td>12.7799</td>
</tr>
<tr>
<td>F2</td>
<td>∞</td>
<td>N4</td>
<td>-1.0000</td>
</tr>
<tr>
<td>F3</td>
<td>1.0000</td>
<td>T1</td>
<td>0.9649</td>
</tr>
<tr>
<td>N1</td>
<td>0.9561</td>
<td>T2</td>
<td>7.6000</td>
</tr>
</tbody>
</table>
TABLE III. BEST INDIVIDUAL MEAN MSE AND THE CORRESPONDING CHOSEN FEATURES

<table>
<thead>
<tr>
<th>Mean mse</th>
<th>Feature Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1996</td>
<td>1 6 7 11 13</td>
</tr>
<tr>
<td>0.1846</td>
<td>2 7 10 11 14</td>
</tr>
<tr>
<td>0.1979</td>
<td>1 6 7 10 11</td>
</tr>
<tr>
<td>0.1685</td>
<td>2 6 9 11 13</td>
</tr>
<tr>
<td>0.2086</td>
<td>1 7 10 11 14</td>
</tr>
</tbody>
</table>

In Fig. 4, the boxplot of the mse of each of the 10 folds from the cv is depicted for the M best individuals. Nevertheless, there were not big differences between the best and the worst individuals in the final population.

Several ideas arise from the results. The first of all, though a relative good mean mse for a so complex problem was obtained, it is though that the results could be improved. Then, an in-depth study of the different NN learning methods and parameters should be done, i.e., the learning rate, etc.

Moreover, the boxplot gives an idea of what was expected: the dispersion of the different folds in the cv may have a high impact in the final decision of the best individual. Recall that the problem is to find the best feature subset for estimating the operation conditions, so if the wrong features are chosen the optimization of the problem would be hardly reached.

Consequently, there are several lines of future work that should be accomplished. Firstly, it is necessary to include an intermediate step for tuning the parameters of the NN learning algorithms. Where to introduce it as a pre-process step or just before the learning of the NN for each feature subset needs also to be studied.

Secondly, considering the problem as a multi objective one, involving the minimizing of both the mean mse and the mse dispersion seems that should obtain better results. Also, introducing better fitness functions instead of the mean of the mse as well as upper dispersion bounds and weighted sums according to the dispersion are issues worthy of further consideration.

Finally, it is needed to compare results using different cv schemes, that is, one-leave-out against k-fold cv. On the one hand, the behavior of the feature selection process could be highly sensible to the cv scheme, specially if multi objective techniques are used. On the other hand, a mixing scheme could be found for relative low number of samples problems.

V. CONCLUSIONS AND FUTURE WORK

The problem of finding the best operation conditions in real world problems represents a big challenge for the AI and Soft Computing methods. In this study, a real case of a dental drill machinery is analysed.

The main objective of this study is choosing the feature subset that best describe the operation conditions while modelling the operation conditions at the same time. For these purposes, a hybrid intelligent system is designed. A wrapper feature selection method using NNs is the responsible of searching for the best feature subset and for modelling the time error for manufacturing, one of the operation conditions that needs to be estimated.

Nevertheless, there is plenty of future work to be done, including the analysis of different cv schemes, the study of the learning parameters tuning for the NNs, automatic determination of the feature subset dimension, different fitness functions and multi objective algorithms comparisons, among others.

ACKNOWLEDGMENT

This research has been funded by the Spanish Ministry of Science and Innovation, under project TIN2008-06681-C06-04, the Spanish Ministry of Science and Innovation PID 560300-2009-11, the Spanish Ministry of Science and Innovation TIN2010-21272-C02-01 (funded by the European Regional Development Fund.), the Junta de Castilla y León [CCTT/10/BU/0002] and by the ITCL project CONSOCO.

REFERENCES


