A new predictive model based on the PSO-optimized support vector machine approach for predicting the milling tool wear from milling runs experimental data

Article *in* The International Journal of Advanced Manufacturing Technology · September 2016

DOI: 10.1007/s00170-015-8148-1

CITATIONS
19

READS
236

4 authors, including:

Paulino Jose Garcia Nieto
University of Oviedo
202 PUBLICATIONS **3,147** CITATIONS
SEE PROFILE

J. A. Vilan
University of Vigo
59 PUBLICATIONS **635** CITATIONS
SEE PROFILE

Abraham SEGADE Robleda
University of Vigo
30 PUBLICATIONS **104** CITATIONS
SEE PROFILE

Int J Adv Manuf Technol
DOI 10.1007/s00170-015-8148-1

ORIGINAL ARTICLE

A new predictive model based on the PSO-optimized support vector machine approach for predicting the milling tool wear from milling runs experimental data

P. J. García-Nieto[1] · **E. García-Gonzalo**[1] · **J. A. Vilán Vilán**[2] · **A. Segade Robleda**[2]

Received: 20 April 2015 / Accepted: 23 November 2015

Abstract The main aim of this research work is to build a new practical hybrid regression model to predict the milling tool wear in a regular cut as well as entry cut and exit cut of a milling tool. The model was based on particle swarm optimization (PSO) in combination with support vector machines (SVMs). This optimization mechanism involved kernel parameter setting in the SVM training procedure, which significantly influences the regression accuracy. Bearing this in mind, a PSO-optimized SVM (PSO–SVM)-based model was successfully used here to predict the milling tool flank wear (output variable) as a function of the following input variables: the duration of experiment, depth of cut, feed, type of material, etc. The second aim is to determine the factors with the greatest bearing on the milling tool flank wear with a view to proposing milling machine's improvements. Firstly, regression with optimal hyperparameters was performed and a determination coefficient of 0.95 was obtained. Secondly, the main advantages of this PSO–SVM-based model are its capacity to produce a simple, easy-to-interpret model; its ability to estimate the contributions of the input variables; and its computational efficiency. Finally, the main conclusions of this study are exposed.

Keywords Statistical learning techniques · Support vector machines (SVMs) · Particle swarm optimization (PSO) · Milling tool wear · Hyperparameter selection

✉ P. J. García-Nieto
lato@orion.ciencias.uniovi.es

[1] Department of Mathematics, Faculty of Sciences, University of Oviedo, 33007 Oviedo, Spain

[2] Department of Mechanical Engineering, University of Vigo, 36200 Vigo, Spain

1 Introduction

Many physical systems require the description of the mechanical interaction across interfaces if they are to be successfully analyzed. Examples in the engineered world range from the milling machining process to characterization of the response and durability of the head–disk interfaces in computer magnetic storage devices, etc. Specifically, milling is the machining process of using rotary cutters to remove material from a workpiece advancing (or feeding) in a direction at an angle with the axis of the tool. It covers a wide variety of different operations and machines, on scales from small individual parts to large, heavy-duty gang milling operations [1–3]. It is one of the most commonly used processes in industry and machine shops today for machining parts to precise sizes and shapes. Milling can be done with a wide range of machine tools. The original class of machine tools for milling was the milling machine (often called a mill). After the advent of computer numerical control (CNC), milling machines evolved into machining centers (milling machines with automatic tool changers, tool magazines or carousels, CNC control, coolant systems, and enclosures), generally classified as vertical machining centers (VMCs) and horizontal machining centers (HMCs). The integration of milling into turning environments and that of turning into milling environments begun with live tooling for lathes, and the occasional use of mills for turning operations led to a new class of machine tools, multitasking machines (MTMs), which are purpose built to provide for a default machining strategy of using any combination of milling and turning within the same work envelope.

Accordingly, milling is a cutting process that uses a milling cutter to remove the material from the surface of a workpiece. The milling cutter is a rotary cutting tool, often with multiple cutting points. As opposed to drilling, where the tool is advanced along its rotation axis, the cutter in milling is usually

Published online: 23 December 2015

moved perpendicular to its axis so that cutting occurs on the circumference of the cutter. As the milling cutter enters the workpiece, the cutting edges (flutes or teeth) of the tool repeatedly cut into and exit from the material, shaving off chips (swarf) from the workpiece with each pass. The cutting action is shear deformation. The metal is pushed off the workpiece in tiny clumps that hang together to more or less extent (depending on the metal type) to form chips. This makes metal cutting a bit different (in its mechanics) from slicing softer materials with a blade [1–3].

As might be suspected, the milling process removes the material by performing many separate, small cuts. This is accomplished by using a cutter with many teeth, spinning the cutter at high speed, or advancing the material through the cutter slowly; most often, it is some combination of these three approaches. The speeds and feeds used are varied to suit a combination of variables. The speed at which the piece advances through the cutter is called feed rate, or just feed; it is most often measured in length of material per full revolution of the cutter. To fix ideas, there are two major classes of milling process [1–3]:

- In face milling, the cutting action occurs primarily at the end corners of the milling cutter. Face milling is used to cut flat surfaces (faces) into the workpiece or to cut flat-bottomed cavities.
- In peripheral milling, the cutting action occurs primarily along the circumference of the cutter, so that the cross section of the milled surface ends up receiving the shape of the cutter. In this case, the blades of the cutter can be seen as scooping out material from the work piece. Peripheral milling is well suited to the cutting of deep slots, threads, and gear teeth.

Note that a high-quality product often implies high-quality surface finish and dimensional accuracy. Ideally, a sharp tool should be maintained at all times. A dull one deforms the surface to a greater depth and may tear the surface which, in turn, may lower the fatigue resistance. A worn tool also results in more friction which, in turn, results in higher cutting temperatures. Unwanted effects may arise from these temperatures; e.g., it may produce untempered martensite in heat-treatable steel [4–6]. Therefore, the milling tool wear has to be controlled. In order to estimate the best conditions, i.e., the values of the most influential parameters in milling machines, as well as to predict the milling tool wear, the flank wear was predicted from the other parameters [4–6] by applying an innovative and new regression technique described here.

Indeed, the main goal of this research work was to obtain the dependence relationship of the milling tool wear (output variable) as a function of the 10 milling operation input variables [4–6] indicated later. To fix ideas, the objective of this study is to evaluate the application of support vector machines (SVMs) in combination with the particle swarm optimization (PSO) technique to identify the milling tool flank wear (see Fig. 1). SVM models are based on the statistical learning theory and are a new class of models that can be used for predicting values from very different fields [7–12]. SVMs are a set of related supervised learning methods used for classification and regression and possess the well-known ability of being universal approximators of any multivariate function to any desired degree of accuracy. The statistical learning theory and structural risk minimization are the theoretical foundations for the learning algorithms of SVMs [11, 13]. In order to carry out the optimization mechanism corresponding to the kernel optimal hyperparameters setting in the SVM training, the PSO technique was used here with success. The PSO technique is a population-based search algorithm based on the simulation of the bird flocking. Similar to other evolutionary computation algorithms, PSO exploits the model of social sharing of information [14–16]. Similar to other evolutionary computation algorithms such as the ant colony optimization [17, 18], artificial bee colony (ABC) technique [19–21], and so on, PSO optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. For the abovementioned purpose, hybrid PSO-optimized SVM (PSO–SVM)-based models [22, 23] were used as automated learning tools, training them in order to predict the milling tool wear values from other input operation parameters in an industrial milling process. According to previous research, the SVM technique has been proven to be an effective tool to predict natural parameters, being successfully used in a wide range of environmental fields: forest modeling

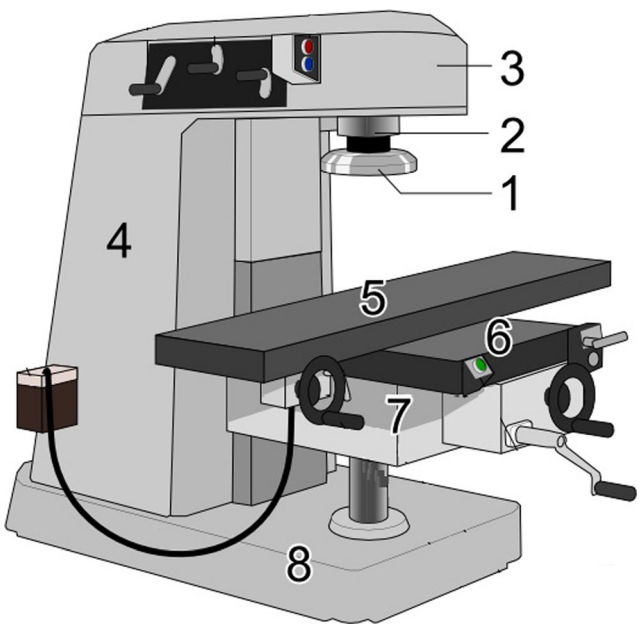

Fig. 1 Vertical milling machine. *1* Milling cutter. *2* Spindle. *3* Top slide or overarm. *4* Column. *5* Table. *6* Y-axis slide. *7* Knee. *8* Base

[24], solar radiation estimation [25, 26], prediction of the air quality [27], study of water properties [28, 29], and so on.

In summary, this paper is organized as follows: firstly, Section 2 describes the materials, methods, and dataset; Section 3 presents and discusses the results of the new hybrid PSO–SVM-based model; and finally, Section 4 draws the main conclusions of this research work.

2 Materials and methods

2.1 Experimental dataset

The data in this set represents experiments from runs on a milling machine under various operating conditions. In particular, tool wear was investigated [5, 6] in a regular cut as well as entry cut and exit cut. The data sampled by three different types of sensors (acoustic emission sensor, vibration sensor, and current sensor) were acquired at several positions. The total number of data processed was 167 samples, and the variables used in this study are shown in Table 1.

There are 16 cases with varying number of runs. The number of runs was dependent on the degree of flank wear that was measured between runs at irregular intervals up to a wear limit (and sometimes beyond). Flank wear was not always measured, and at times when no measurements were taken, no entry was made. The 16 cases are enumerated in Table 2.

The setup of the experiment is depicted in Fig. 2. The basic setup encompasses the spindle and the table of the Matsuura machining center (MC) at 510 V. An acoustic emission sensor and a vibration sensor are each mounted to the table and the spindle of the machining center. The signals from all sensors are amplified and filtered and then fed through two root mean square (RMS) converters before they enter the computer for data acquisition. The signal from a spindle motor current sensor is fed into the computer without further processing.

Table 1 Set of milling operation input variables used in this study and their description

Variable name	Description
Case	Case number (1–16)
Run	Counter for experimental runs in each case
VB (mm)	Flank wear, measured after runs; measurements for VB were not taken after each run
Time (mm)	Duration of experiment (restarts for each case)
DOC (mm)	Depth of cut (does not vary for each case)
Feed (mm/rev)	Feed (does not vary for each case)
Material	Steel or cast iron
smcAC	AC spindle motor current
smcDC	DC spindle motor current
vib_table	Table vibration
vib_spindle	Spindle vibration
AE_table	Acoustic emission at table
AE_spindle	Acoustic emission at spindle

Table 2 Experimental conditions

Case	Depth of cut (mm)	Feed (mm/rev)	Material
1	1.5	0.5	Cast iron
2	0.75	0.5	Cast iron
3	0.75	0.25	Cast iron
4	1.5	0.25	Cast iron
5	1.5	0.5	Steel
6	1.5	0.25	Steel
7	0.75	0.25	Steel
8	0.75	0.5	Steel
9	1.5	0.5	Cast iron
10	1.5	0.25	Cast iron
11	0.75	0.25	Cast iron
12	0.75	0.5	Cast iron
13	0.75	0.25	Steel
14	0.75	0.5	Steel
15	1.5	0.25	Steel
16	1.5	0.5	Steel

The matrix for the parameters chosen for the experiments was guided by industrial applicability and recommended manufacturer's settings. Therefore, the cutting speed was set to 200 m/min which is equivalent to 826 rev/min. Two different depths of cut (1.5 and 0.75 mm, respectively) were chosen.

In this study, two feeds were taken, 0.5 and 0.25 mm/rev, which translate into 413 and 206.5 mm/min, respectively. Furthermore, two types of material (cast iron and stainless steel J45) were used. All experiments were done a second time with the same parameters with a second set of inserts. The size of the workpieces was 483×178×51 mm.

2.1.1 Data acquisition and processing

As described in the previous section, the data were sent through a high-speed data acquisition board with a maximal sampling rate of 100 kHz. The sampled output of the data was used for the signal processing software. LabVIEW® [30] was used for this task. This software is a general purpose programming development system which uses a graphical language (G). With G, programs are created in block diagram form. The chosen layout allowed for data acquisition, storage, presentation, and processing. Data were stored to allow for real-time simulation and also later analysis.

Several sensor signals underwent preprocessing. In most cases, the signal was amplified to be able to meet threshold requirements of equipment. In particular, the signals from the

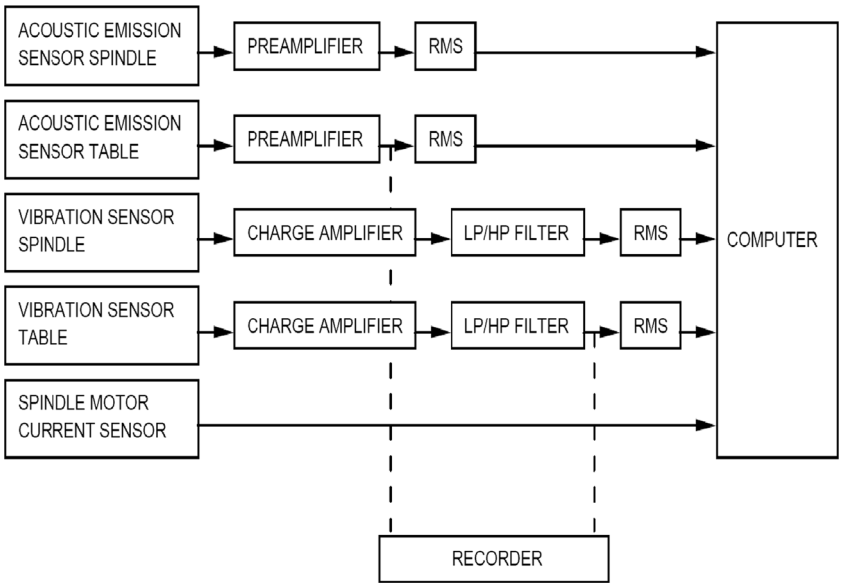

Fig. 2 The setup of the experiment

acoustic emission sensors (acoustic emission sensor model WD 925) and from the vibration sensors (an accelerometer model 7201-50, Endevco) were amplified to be in the range of ±5 V for maximum load, considering the maximum allowable range of the equipment. The signals were filtered using a high-pass filter, and the vibration sensor signals were additionally filtered with a low-pass filter. Corner frequencies were chosen according to the noise that could be observed on an oscilloscope. The periodical noise of 180 Hz was observed on the oscilloscope for the vibration signal corresponding to the third harmonic of the main power supply. Therefore, the chosen corner frequency for the low-pass filter was 400 Hz. For the high-pass filter, 1 kHz was chosen. Above 8 kHz, the range of the acoustic emission sensor ends. That is, readings above that frequency cannot be attributed to any occurrence in the machining process. Since it clutters the signal unnecessarily, it was filtered out. Acoustic emission and vibration signals were fed through an RMS device. Its use smoothes the signal and makes it more accessible to signal processing. The RMS is a statistical measure of the magnitude of a varying quantity, and it is proportional to the energy contents of the signal. The RMS of a function f for a period of time ΔT is defined by [5, 6]

$$\mathrm{RMS} = \sqrt{\frac{1}{\Delta T}\int_0^{\Delta T} f^2(t)dt} \tag{1}$$

where ΔT is the time constant and $f(t)$ is the signal function. In this case, as the data is discrete, the formula used was [5, 6]

$$\mathrm{RMS} = \sqrt{\frac{1}{n}\sum_{k=1}^{n} (f(t_k))^2} \tag{2}$$

where n is the number of samples. There is only a value of the variables *case*, *run*, *VB*, and *time* per each structure array (in the dataset). Furthermore, the variables depth of cut (DOC), *feed*, and *material* do not vary within each case. Additionally, the variables AC spindle motor current, DC spindle motor current, table vibration, spindle vibration, acoustic emission at table, and acoustic emission at spindle are curves with 9000 points each one. The RMS has been obtained for each of these curves.

2.1.2 Tool wear

Tool wear comes in different forms. Apart from the intuitive rounding of the cutting edge, the crater wear on the rake face due to the abrasion of the sliding of the chip on the rake face and the flank wear due to friction of the tool on the workpiece occur. The speed of cutting, more than other parameters, influences the rate of wear; the depth of cut and the feed rate also affect the tool life. In our experiments, we measured the *flank wear* VB as a generally accepted parameter for evaluating tool wear (see Fig. 3) [1–6].

The flank wear VB is measured as the distance from the cutting edge to the end of the abrasive wear on the flank face of the tool. The flank wear was observed during the experiments. The insert was taken out of the tool, and the wear was measured with the help of a microscope.

2.2 Support vector machine method

Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression [7, 31–35]. The SVMs were originally developed for classification and were later generalized to solve regression problems [31, 35–37]. This last method is called *support vector*

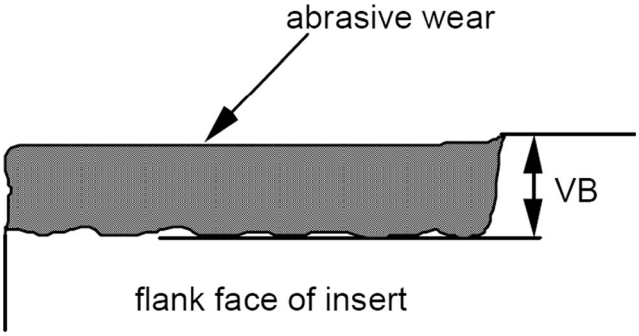

Fig. 3 The flank wear VB as the accepted parameter to evaluate the milling tool wear

regression (SVR). The model produced by SVR only depends on a subset of the training data, because the cost function for building the model tries to ignore any training data that are close (within a threshold $\in$) to the model prediction. When the regression SVM is applied to nonlinear separable data, it is necessary to use the *kernel trick*. The reason that this kernel trick is useful is that there are many regression problems that are not linearly regressable in the space of the inputs $\mathbf{x}$, which might be in a higher-dimensionality feature space given a suitable mapping $x \mapsto \psi(x)$ [31, 35, 37].

The basic idea of SVR is briefly described here. Instead of attempting to classify new unseen variables $\tilde{\mathbf{x}}$ into one of two categories $\tilde{y}=\pm 1$, we now wish to predict a real-valued output y for the observed value t so that our training data is a set of L points of the form $\{\mathbf{x}_i, t_i\}$, where $i=1,2,\ldots,L$, $y \in \Re$, and $\mathbf{x} \in \Re^D$ [24, 35, 38, 39] so that

$$y_i = f(\mathbf{x}_i) = \mathbf{w} \cdot \mathbf{x}_i + b \tag{3}$$

where "$\cdot$" denotes the dot product, x_i is the D-dimensional real input vector, $\mathbf{w}$ is the normal vector to the maximum-margin hyperplane, and y_i is the predicted output value. The parameter $\frac{b}{\|\mathbf{w}\|}$ determines the offset of this hyperplane from the origin along the normal vector $\mathbf{w}$. The SVR uses a more sophisticated penalty function: a penalty is not imposed if the predicted value y_i is less than the distance ε away from the actual value t_i, i.e., if $|t_i - y_i| < \varepsilon$. Referring to Fig. 4, the region bound by $y_i \pm \varepsilon \, \forall \, i$ is called the ε-insensitive tube. Another modification to the penalty function is that output variables outside the tube are allocated one of two slack variable penalties, depending on whether they lie above (ξ^+) or below (ξ^-) the tube, where $\xi^+ > 0$ and $\xi^- > 0 \, \forall \, i$.

$$t_i \leq y_i + \varepsilon + \xi^+ \tag{4}$$

$$t_i \geq y_i - \varepsilon - \xi^- \tag{5}$$

The task is then to find a functional form f that can correctly predict new cases that the SVM has not been presented with before. This can be achieved by training the SVM model on a sample set called the training set, a process that involves the sequential optimization of an error function. Depending on the definition of this error function, two types of SVM models can be recognized and the resulting SVM problem can be formulated as follows [24, 31, 40]:

(a) Regression SVM type 1 (also known as ε–SVM regression): for this type of SVM, we have to solve an optimization problem, minimizing the following general risk function [24, 31, 35]:

$$R[\mathbf{w}, b, \xi] = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{L}\left(\xi_i^+ + \xi_i^-\right) \tag{6}$$

subject to

$$\left\{ \begin{array}{c} \langle \mathbf{w}, \psi(\mathbf{x}_i) \rangle + b - y_i \leq \varepsilon + \xi_i^+ \\ y_i - \langle \mathbf{w}, \psi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_i^- \\ \xi_i^+, \xi_i^- \geq 0 \end{array} \right\} \; i = 1, \ldots, L \tag{7}$$

(b) Regression SVM type 2 (also known as ν–SVM regression): for this SVM model, it is necessary to solve the following optimization problem, minimizing the following general risk function [24, 31, 35, 38]:

$$R[\mathbf{w}, b, \xi] = \frac{1}{2}\|\mathbf{w}\|^2 - C\left[\nu\varepsilon + \frac{1}{L}\sum_{i=1}^{L}\left(\xi_i^+ + \xi_i^-\right)\right] \tag{8}$$

subject to

$$\left\{ \begin{array}{c} \langle \mathbf{w}, \psi(\mathbf{x}_i) \rangle + b - y_i \leq \varepsilon + \xi_i^+ \\ y_i - \langle \mathbf{w}, \psi(\mathbf{x}_i) \rangle - b \leq \varepsilon + \xi_j^- \\ \xi_j^+, \xi_j^- \geq 0 \end{array} \right\} i = 1, \ldots, L \tag{9}$$

where $\psi : \mathbf{X} \rightarrow \mathbf{Z}$ is the transformation of the input space into a new space $\mathbf{Z}$, usually a larger dimension space, where we define an inner product by means of a positive definite function k (kernel trick) [24, 31, 35, 38]

$$\langle \psi(\mathbf{x}), \psi(\mathbf{x}') \rangle = \sum_i \psi_i(\mathbf{x})\psi_i(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}') \tag{10}$$

The above problem is quadratic with linear constraints, and so, the Karush–Kuhn–Tucker (KKT) optimality conditions are necessary and sufficient. The solution, which can be obtained from the dual problem, is a linear combination of a subset of sample point-denominated support vectors (s.v.) as follows [39–42]:

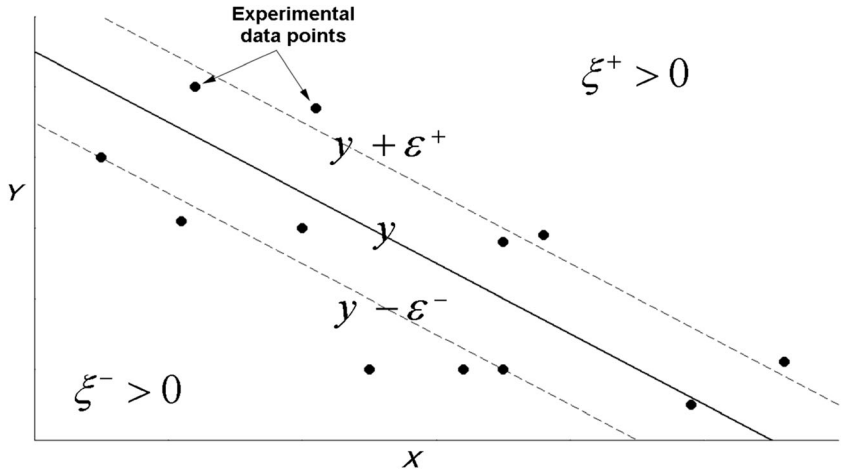

Fig. 4 Regression with ε-insensitive tube

$$\mathbf{w} = \sum_{\text{s.v.}} \beta_i \psi(\mathbf{x}_i) \Rightarrow$$
$$f_{\mathbf{w},b}(\mathbf{x}) = \sum_{\text{s.v.}} \beta_i \langle \psi(\mathbf{x}_i), \psi(\mathbf{x}) \rangle + b = \sum_{\text{s.v.}} \beta_i k(\mathbf{x}_i, \mathbf{x}) + b \tag{11}$$

The reason that this kernel trick is useful is that there are many regression problems that cannot be linearly regressed in the space of the inputs $\mathbf{x}$, which might be in a higher-dimensional feature space given a suitable mapping. Different kernel functions are described in the bibliography as for example [24, 35, 38–42]

- Radial basis function

$$k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\sigma \|\mathbf{x}_i - \mathbf{x}_j\|^2} \tag{12}$$

- Polynomial kernel

$$k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + a)^b \tag{13}$$

where σ, a and b are the parameters defining the kernel's behavior.

To sum up, to use an SVM to solve a regression problem for data that is not linearly separable, we need to first choose a kernel and relevant parameters that can be expected to map the nonlinearly separable data into a feature space where it is linearly separable.

2.3 The particle swarm optimization algorithm

The particle swarm optimization (PSO) is an evolutionary optimization algorithm, where a population of particles or proposed solutions evolves with each iteration, moving towards the optimal solution of the problem [14–16]. The ant colony system [17], genetic algorithms [22], and differential evolution [18, 23] also belong to this group where the evolution of the population has a stochastic component. Indeed, a new population in the PSO algorithm is obtained by shifting the positions of the previous one in each iteration. In its movement, each individual is influenced by its neighbor's trajectory and its own trajectory.

The parameters, or possible set of solutions, are contained in a vector $\mathbf{x}_i$ that is called the *particle* of the swarm and represents its position in the search space of possible solutions. The particle dimension is the number of parameters. The particle position $\mathbf{x}_i^0$ and its velocity $\mathbf{v}_i^0$ are chosen randomly. The value of the fitness function is then calculated for each particle, and the velocities and positions are updated, taking into account these values. The algorithm updates the positions and the velocities of the particles following the equations

$$\mathbf{v}_i^{k+1} = \omega \mathbf{v}_i^k + \phi_1 (\mathbf{g}^k - \mathbf{x}_i^k) + \phi_2 (\mathbf{l}_i^k - \mathbf{x}_i^k) \tag{14}$$

$$\mathbf{x}_i^{k+1} = \mathbf{x}_i^k + \mathbf{v}_i^{k+1} \tag{15}$$

The velocity of each particle, i, at iteration k, depends on three components:

- The previous step velocity term, $\mathbf{v}_i^k$, affected by the constant inertia weight, ω.
- The cognitive learning term, which is the difference between the particle's best position so far found (called $\mathbf{l}_i^k$, local best) and the particle current position $\mathbf{x}_i^k$.
- The social learning term, which is the difference between the global best position found thus far in the entire swarm (called $\mathbf{g}^k$, global best) and the particle's current position $\mathbf{x}_i^k$.

These two last components are affected by $\phi_1 = c_1 r_1$ and $\phi_2 = c_2 r_2$ where r_1 and r_2 are the random numbers distributed

uniformly in the interval [0,1] and c_1 and c_2 are the constants. The particles of the swarm make up a cloud that covers the whole search space in the initial iteration and gradually contracts its size as iterations advance, performing the exploration. So, in the initial stages, the algorithm performs an exploration searching for plausible zones and, in the last iterations, the best solution is improved.

The PSO implementation of the algorithm has been refined over the years, and many variants were created. In this paper, the standard 2011 PSO has been used. It contemplates some improvements in the implementation [14–16], and the PSO parameters are set to the values

$$\omega = \frac{1}{2\ln 2} \quad \text{and} \quad c_1 = c_2 = 0.5 + \ln 2 \tag{16}$$

The swarm topology defines how particles are connected between them to interchange information with the global best. In the actual standard PSO, each particle informs only K particles, usually three, randomly chosen.

2.4 The goodness of fit of this approach

In this study, an SVR technique in combination with the PSO approach has been implemented in order to predict the milling tool flank wear values using radial basis function (RBF) kernels [38]. The goodness of fit of a statistical model describes how well it fits a set of observations. Indeed, it is important to select the model that best fits the experimental data. The following criterion was considered here [24, 43, 44]:

- The coefficient of determination (R^2): as it is well known, in statistics, the coefficient of determination (R^2) is used in the context of statistical models whose main purpose is the prediction of future outcomes on the basis of other related information [8, 9]. If an intercept is included, then R^2 is simply the square of the sample correlation coefficient between the outcomes and their predicted values. To fix ideas, this ratio indicates the proportion of total variation in the dependent variable explained by the model (milling tool wear in our case). A dataset takes the value t_i, each of which has an associated modeled value y_i. The former is called the observed value, and the latter is often referred to as the predicted value. The variability in the dataset is measured through different sums of squares [24, 43, 44]:

- $\mathrm{SS_{tot}} = \sum_{i=1}^{n} (t_i - \bar{t})^2$: the total sum of squares, proportional to the sample variance.
- $\mathrm{SS_{reg}} = \sum_{i=1}^{n} (y_i - \bar{t})^2$: the regression sum of squares, also called the explained sum of squares.
- $\mathrm{SS_{err}} = \sum_{i=1}^{n} (t_i - y_i)^2$: the residual sum of squares.

In the previous sums, $\bar{t}$ is the mean of the n observed data

$$\bar{t} = \frac{1}{n} \sum_{i=1}^{n} t_i \tag{17}$$

Bearing in mind the above sums, the general definition of the coefficient of determination is

$$R^2 \equiv 1 - \frac{\mathrm{SS_{err}}}{\mathrm{SS_{tot}}} \tag{18}$$

The coefficient of determination value of 1.0 indicates that the regression curve fits the data perfectly.

3 Analysis of results and discussion

The operation input variables considered in this research work are shown in Table 3 [1–6]. The total number of predicting variables used to build the hybrid PSO–SVM-based model was 10 (see Table 1). The predicted output variable is the flank wear (VB) measured in micrometers. In this study, we do not use the input variable DC spindle motor current (smcDC) because it is highly correlated with AC spindle motor current (smcAC). Therefore, the final number of input variables used to predict the flank wear was 9 (see Table 3). The VB missing values have been removed, and 145 samples remain. Furthermore, the input variable material is a categorical variable. The variables smcDC and smcAC are highly correlated, with a correlation coefficient of 0.9515. Thus, only one of them has been used as an input variable (smcAC).

Additionally, it is well known that the SVM techniques are strongly dependent on the SVM hyperparameters: the regularization factor C (see Eqs. (6) and (8)), the hyperparameter ε that defines the ε-insensitive tube (allowable error), and σ that represents the kernel parameter if an RBF is chosen. There exists a vast body of literature regarding the choice of hyperparameters for SVMs [31, 38]. Some methods often

Table 3 Operation input variables used in this study and their names

Input variables	Name of the variable
Time (mm)	Time
DOC (mm)	Depth of cut
Feed (mm/rev)	Feed
Material	Material
smcAC	AC spindle motor current
vib_table	Table vibration
vib_spindle	Spindle vibration
AE_table	Acoustic emission at table
AE_spindle	Acoustic emission at spindle

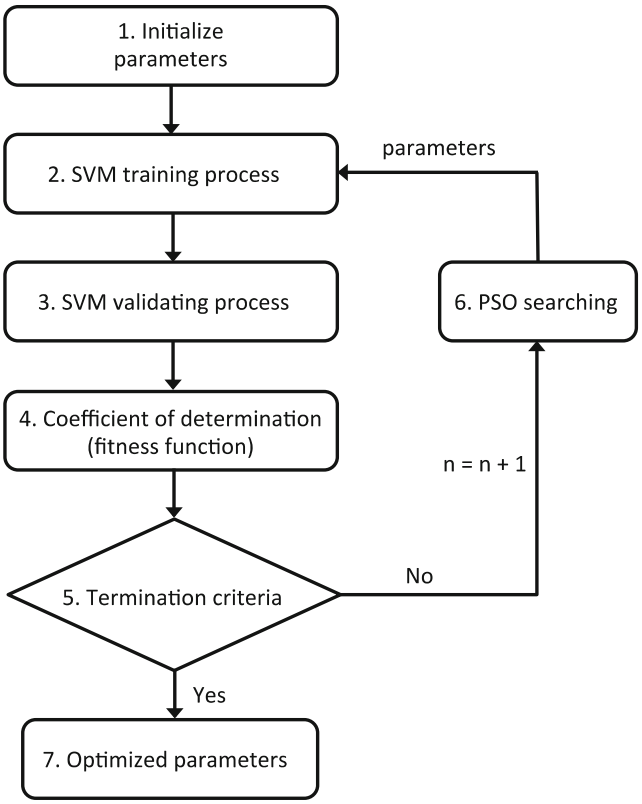

Fig. 5 Flowchart of the new hybrid PSO–SVM-based model

used to determine suitable hyperparameters are [24, 31, 38] cross-validation, grid search, random search, Nelder–Mead search, heuristic search, genetic algorithms, pattern search, etc.

In other words, a novel hybrid PSO–SVM-based model was applied to predict the milling tool wear (output variable) from the other nine remaining variables (input variables) in a milling process [24, 38], studying their influence in order to optimize its calculation through the analysis of the coefficient of determination (R^2) with success. Figure 5 shows the flowchart of this new hybrid PSO–SVM-based model developed in this study.

The determination coefficient is the statistical measure of how well a regression curve approximates real data points. Furthermore, it is a descriptive measure ranging from zero to one, indicating how good one term is predicted by another one. Thus, $R^2=1$ indicates the best approximation and $R^2=0$ the worst one.

Table 4 Initial ranges of the three hyperparameters of the PSO–SVM-based model with an RBF kernel

SVR hyperparameters	Lower limit	Upper limit
C	10^{-6}	10^{-4}
ε	10^{-10}	10^{-4}
σ	10^{-6}	10^{-4}

Table 5 Optimal hyperparameters of the fitted RBF–SVM-based model found with a PSO technique

Kernel	Values of optimal hyperparameters
RBF	Regularization factor C=57.8773, ε=0.014021, σ=0.5778

Cross-validation was the standard technique used here for finding the real coefficient of determination (R^2) [45]. The data set is randomly divided into l disjoint subsets of equal size, and each subset is used once as a validation set, whereas the other $l-1$ subsets are put together to form a training set. In the simplest case, the average accuracy of the l validation sets is used as an estimator for the accuracy of the method. The combination of the hyperparameters with the best performance is chosen [31, 38]. In this way, 10-fold cross-validation was used here.

As it has been previously pointed out, in order to guarantee the prediction ability of the PSO–SVM-based model, an exhaustive 10-fold cross-validation algorithm was used [45]. The referred algorithm consists in splitting the sample into 10 parts and using 9 of them for training and the remaining 1 for testing. This process was performed 10 times using each of the parties of the 10 divisions for testing and calculating the average error. Therefore, all the possible variabilities of PSO–SVM-based model parameters have been evaluated in order to get the optimum point, looking for those parameters that minimize the average error. With these optimal hyperparameters, the error criterion was calculated from the built model using 90 % of the sample and tested with the remaining 10 %. In this way, we are able to simulate as much as possible the real conditions under which the model would be built in order to later fit it to new observation data unrelated to the construction of the model.

The regression modeling has been performed with SVR–ε using the LIBSVM library [46]. The searching in the parameter space has been made, taking into account that the SVM algorithm changes its results significantly when its parameters increase or decrease by a power of 10. We have worked with powers of 10, and the searched parameters have been the exponents being the 3-dimensional search space $[-6, 4]\times[-10,4]\times[-6,4]$. The binds (initial ranges) of the space of solutions used in particle swarm optimization (PSO) technique are shown in Table 4. The number of particles used has

Table 6 Coefficient of determination (R^2) and correlation coefficient for the hybrid PSO–SVM-based model with an RBF kernel fitted in this study

Kernel	Coefficients of determination (R^2)/correlation coefficients (r)
RBF	0.95/0.98

Table 7 Evaluation of the importance of the variables according to their weights in the fitted PSO–SVM-based model with an RBF kernel

Input variable	Weight
Time	4.1405
smcAC	1.5900
vib_spindle	−1.4923
Material	1.3610
AE_spindle	1.2298
AE_table	0.9176
vib_table	−0.8449
DOC	−0.8228
Feed	0.0551

been 20. The stopping criterion is fulfilled if there is no improvement after the 10 iterations, along with the maximum number of iterations equal to 1000.

To optimize the SVM parameters, the PSO module is used. The PSO searches for the best C, σ, and ε parameters by comparing the forecasting error in every iteration. Search space is organized in three dimensions, one for each parameter. Main fitness factor is the coefficient of determination (R^2). Table 5 shows the optimal hyperparameters of the fitted RBF–SVM-based model found with the PSO technique.

Table 6 shows the coefficient of determination and correlation coefficient for the fitted PSO–RBF–SVM-based model from 2006 to 2010. According to this statistic, the SVM with the RBF kernel is the best model for estimating the flank wear in the milling process, since the fitted SVM with the RBF kernel has a coefficient of determination (R^2) equal to 0.95 and a correlation coefficient equal to 0.98. These results indicate an important goodness of fit, that is to say, a good agreement is obtained between our model and the observed data.

The importance ranking of the nine input operation variables in order to predict the milling tool wear (output variable) in this high nonlinear complex problem is shown in Table 7 and Fig. 6.

Finally, this research work was able to predict the milling tool flank wear in agreement to the actual milling tool wear values observed experimentally using this hybrid PSO–SVM-based model with great accurateness and success. Indeed, Fig. 7 shows the comparison between the measured flank wear (VB) in micrometers and the flank wear predicted using the PSO–SVM-based model with an RBF kernel in the milling process using the optimal hyperparameters calculated previously using the PSO technique and Fig. 8 shows the absolute error of the predicted model. Therefore, the use of an SVM model with an RBF kernel is necessary in order to achieve an effective approach to nonlinearities present in the regression problem. Obviously, these results coincide again with the outcome criterion of *goodness of fit* (R^2) so that the SVM model with an RBF kernel has been the best fitting.

In summary, the PSO–SVM-based model [7–16] is a suitable tool in modeling and assessment of singular problems, such as the study of the milling tool wear (flank wear in this case) in an industrial milling process. The most important operation input variables were used to fully characterize the problem.

4 Conclusions

Based on the experimental and numerical results, the main findings of this research work can be summarized as follows:

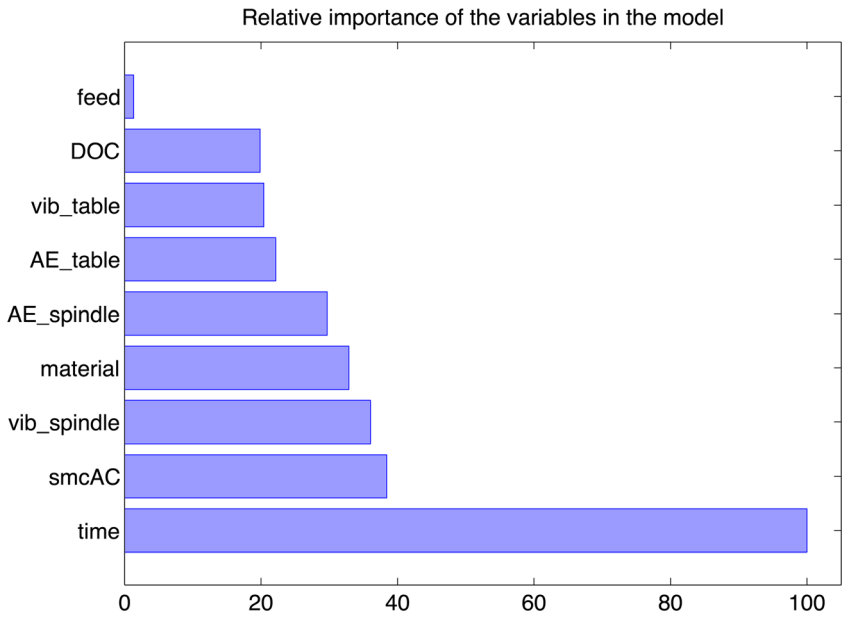

Fig. 6 Relative importance of the input variables to predict the milling tool wear in the fitted PSO–SVM-based model with an RBF kernel

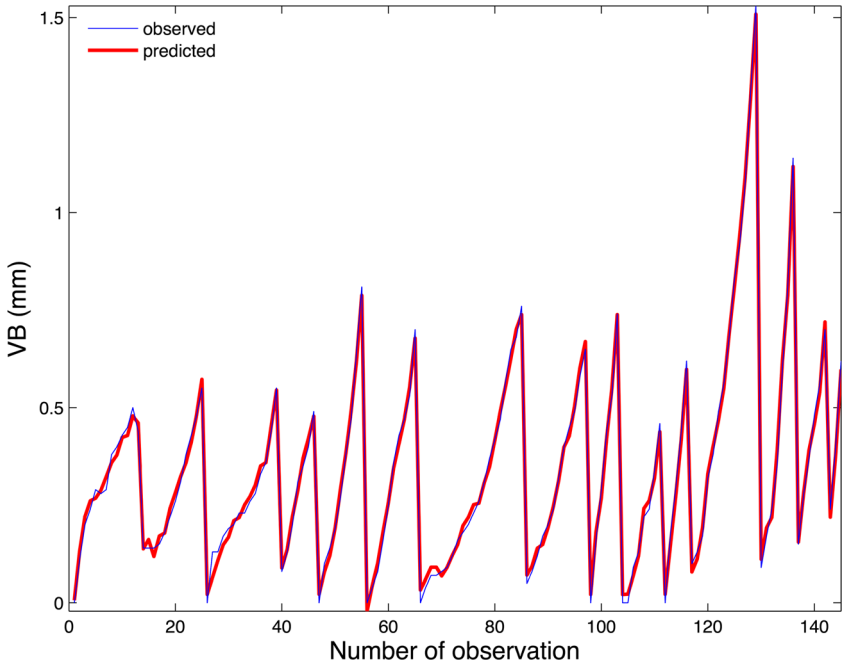

Fig. 7 The comparison between the measured flank wear (VB) in millimeters and the flank wear predicted using the PSO–SVM-based model with an RBF kernel in the milling process using the optimal hyperparameters (R^2=0.95)

- Firstly, the milling tool flank wear can be accurately modeled using a hybrid PSO–SVM-based model with an RBF kernel.
- Secondly, this hybrid PSO–RBF–SVM-based model to predict the flank wear allows to lower costs in the quality's assessment of the milling process.
- Thirdly, a high coefficient of determination equal to 0.95 was obtained when this hybrid PSO–RBF–SVM-based model was applied to the experimental dataset (see Figs. 7 and 8).
- Fourthly, the significant order of the input variables involved in the prediction of the milling tool flank wear values was set. Specifically, the input operation variables *duration of the experiment* (time) and *AC spindle motor current* (smcAC) could be considered as the most influential parameters, respectively.

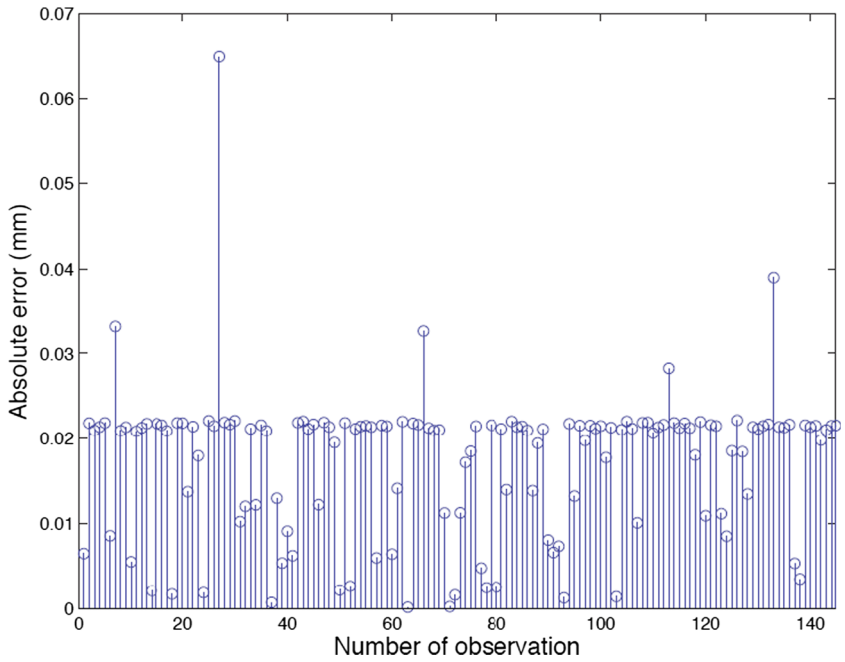

Fig. 8 The absolute error in millimeters for the flank wear predicted using the PSO–SVM-based model with an RBF kernel in the milling process using the optimal hyperparameters (R^2=0.95)

- Fifthly, the influence of the kernel parameter setting of the SVMs on the flank wear regression performance was established.
- Finally, the results verify that this hybrid PSO–SVM-based regression method significantly improves the generalization capability achievable with only the SVM-based regressor.

In summary, this innovative methodology can be applied to other similar industrial processes with success, but it is always necessary to take into account the specificities of each process. Additionally, authors of this research work have confidence that the results obtained in this study will be useful to promote new future research works in this direction.

Acknowledgments The authors wish to acknowledge the computational support provided by the Department of Mathematics at the University of Oviedo. The English grammar and spelling of the manuscript have been revised by Anthony Ashworth, a teacher and an international lecturer.

References

1. Rolt LTC (1965) A short history of machine tools. MIT, Cambridge
2. Hall H (2004) Milling: a complete course. Trans-Atlantic, New York
3. Bray S (2011) Milling. Crowood, Ramsbury
4. Schey JA (1977) Introduction to manufacturing processes. McGraw-Hill, New York
5. Goebel K (1966) Management of uncertainty in sensor validation, sensor fusion, and diagnosis of mechanical systems using soft computing techniques. Ph.D. Dissertation, Department of Mechanical Engineering, University of California at Berkeley
6. Agogino A, Goebel K (2007) BEST lab. In: UC Berkeley, Milling data set. NASA Ames Prognostics Data Repository. http://ti.arc.nasa.gov/project/prognostic-data-repository, NASA Ames Research Center, Moffett Field, CA
7. Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20:273–297
8. Vapnik V (1995) The nature of statistical learning theory. Springer, New York
9. Vapnik V (1998) Statistical learning theory. Wiley-Interscience, New York
10. Hansen T, Wang CJ (2005) Support vector based battery state of charge estimator. J Power Sources 141:351–358
11. Li X, Lord D, Zhang Y, Xie Y (2008) Predicting motor vehicle crashes using support vector machine models. Accident Anal Prev 40:1611–1618
12. Álvarez Antón JC, García Nieto PJ, Blanco Viejo C, Vilán Vilán JA (2013) Support vector machines used to estimate the battery state of charge. IEEE T Power Electr 28(12):5919–5926
13. Kecman V (2005) Support vector machines: an introduction. In: Wang L (ed) Support vector machines: theory and applications. Springer, Heidelberg, pp 1–48
14. Eberhart RC, Shi Y, Kennedy J (2001) Swarm intelligence. Morgan Kaufmann, San Francisco
15. Clerc M (2006) Particle swarm optimization. Wiley-ISTE, London
16. Olsson AE (2011) Particle swarm optimization: theory, techniques and applications. Nova Science, New York
17. Dorigo M, Stützle T (2004) Ant colony optimization. MIT, Cambridge
18. Panigrahi BK, Shi Y, Lim M-H (2011) Handbook of swarm intelligence: concepts, principles and applications. Springer, Berlin
19. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J Global Optim 39(3):459–471
20. Karaboga D, Akay B (2009) A survey: algorithms simulating bee swarm intelligence. Artif Intell Rev 31(1):68–85
21. Karaboga D, Gorkemli B (2014) A quick artificial bee colony (qABC) algorithm and its performance on optimization problems. Appl Soft Comput 23:227–238
22. Simon D (2013) Evolutionary optimization algorithms. Wiley, New York
23. Yang X-S, Cui Z, Xiao R, Gandomi AH, Karamanoglu M (2013) Swarm intelligence and bio-inspired computation: theory and applications. Elsevier, London
24. García Nieto PJ, Martínez Torres J, Araújo Fernández M, Ordóñez Galán C (2012) Support vector machines and neural networks used to evaluate paper manufactured using *Eucalyptus globulus*. Appl Math Model 36:6137–45
25. Chen J-L, Li G-S, Wu S-J (2013) Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energ Convers Manage 75:311–318
26. Zeng J, Qiao W (2013) Short-term solar power prediction using a support vector machine. Renew Energ 52:118–127
27. García Nieto PJ, Combarro EF, del Coz Díaz JJ, Montañés E (2013) A SVM-based regression model to study the air quality at local scale in Oviedo urban area (northern Spain): a case study. Applied Mathematics and Computation 219(17):8923–8937
28. García Nieto PJ, Alonso Fernández JR, de Cos Juez FJ, Sánchez Lasheras F, Díaz Muñiz C (2013) Hybrid modelling based on support vector regression with genetic algorithms in forecasting the cyanotoxins presence in the Trasona reservoir (northern Spain). Environ Res 122:1–10
29. Vilán Vilán JA, Alonso Fernández JR, García Nieto PJ, Sánchez Lasheras F, de Cos Juez FJ, Díaz Muñiz C (2013) Support vector machines and multilayer perceptron networks used to evaluate the cyanotoxins presence from experimental Cyanobacteria concentrations in the Trasona reservoir (northern Spain). Water Resour Manag 27(9):3457–3476
30. Essick J (2012) Hands-on introduction to LabVIEW for scientists and engineers. Oxford University Press, New York
31. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, New York
32. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D (2000) Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16:906–914
33. Guo G, Li SZ, Chan KL (2001) Support vector machines for face recognition. Image Vision Comput 19:631–638
34. Taboada J, Matías JM, Ordóñez C, García Nieto PJ (2007) Creating a quality map of a slate deposit using support vector machines. J Comput Appl Math 204(1):84–94
35. Fletcher T (2009) Support vector machines explained: Introductory course. University College London (UCL), London, pp 10–15, Technical internal report
36. Suárez Sánchez A, García Nieto PJ, Riesgo Fernández P, del Coz Díaz JJ, Iglesias-Rodríguez FJ (2011) Application of a SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). Math Comput Model 54:1453–1466
37. Safavi HR, Esmikhani M (2013) Conjunctive use of surface water and groundwater: application of support vector machines (SVMs) and genetic algorithms. Water Resour Manag 27(7):2623–2644

38. Steinwart I, Christmann A (2008) Support vector machines. Springer, New York
39. de Cos Juez FJ, García Nieto PJ, Martínez Torres J, Taboada Castro J (2010) Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model. Math Comput Model 52:1177–1184
40. Matías JM, Taboada J, Ordóñez C, García Nieto PJ (2007) Machine learning techniques applied to the determination of road suitability for the transportation of dangerous substances. J Hazard Mater 147: 60–66
41. Schölkopf B, Smola AJ (2002) Learning with kernels: support vector machines, regularization, optimization and beyond. MIT, Cambridge
42. Shawe-Taylor J, Cristianini N (2004) Kernel methods for pattern analysis. Cambridge University Press, New York
43. Wasserman L (2003) All of statistics: a concise course in statistical inference. Springer, New York
44. Freedman D, Pisani R, Purves R (2007) Statistics. W.W. Norton & Company, New York
45. Picard R, Cook D (1984) Cross-validation of regression models. J Am Stat Assoc 79(387):575–583
46. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM T Int Syst Technol 2:1–27