Int J Adv Manuf Technol (2009) 42:1082–1088
DOI 10.1007/s00170-008-1678-z

ORIGINAL ARTICLE

Regression analysis, support vector machines, and Bayesian neural network approaches to modeling surface roughness in face milling

B. Lela · D. Bajić · S. Jozić

Received: 26 March 2008 / Accepted: 18 July 2008 / Published online: 12 August 2008

Abstract This study examines the influence of cutting speed, feed, and depth of cut on surface roughness in face milling process. Three different modeling methodologies, namely regression analysis (RA), support vector machines (SVM), and Bayesian neural network (BNN), have been applied to data experimentally determined by means of the design of experiment. The results obtained by the models have been compared. All three models have the relative prediction error below 8%. The best prediction of surface roughness shows BNN model with the average relative prediction error of 6.1%. The research has shown that, when the training dataset is small, both BNN and SVR modeling methodologies are comparable with RA methodology and, furthermore, they can even offer better results. Regarding the influence of the examined cutting parameters on the surface roughness, it has been shown that the feed has the largest affect on it and the depth of cut the least.

Keywords Face milling · Surface roughness · Regression · Support Vector Machines · Bayesian neural network

1 Introduction

The quality of milled components is expressed by required product specifications such as: dimensions, surface finish, reflective properties, etc. In all machining processes, surface quality is one of the most specified customer requirements, and the major indication of surface quality on machined parts is surface roughness [1]. Demanded surface roughness cannot be attained as easily as physical dimensions because many factors influence on it. The most important of them are machining parameters, built-up edge, tool geometry, machining time, tool and workpiece material, tool wear, dynamic behavior of machining system, application of cooling and lubrication agent, etc. [2].

Some of these factors can be controlled and some cannot. Factors, such as tool, workpiece and machine vibration, tool wear and degradation, chip formations and workpiece, and tool material variability cannot be controlled as easily [3]. The tool geometry, feed, cutting speed, and depth of cut can be controlled and set up in advance. It is always desirable to have an accurate and reliable mathematical model that is able to predict the surface roughness and thereby the quality of a product too. That is the reason why many researchers have made a great effort in obtaining a model relating cutting parameters and surface roughness.

In science and engineering practice, a system is typically modeled by fitting a model to the gathered data. Models usually involve some free parameters, and fitting a model to the data implies inferring the values of those parameters.

A survey of the literature indicates that many different approaches have been applied to surface roughness prediction. A detailed review of methodologies and practice on the prediction of surface profile and roughness in machining processes can be found in Lu [4].

Regarding the prediction of surface roughness in the process of face milling, design of experiment (DOE) and regression analysis (RA) as well as neural networks have been already applied to the problem of its modeling. For example, Bajić and Belajić [2] and Oktem et al. [5] used response surface methodology, while Benardos and Vosniakos [6] and

B. Lela (✉) · D. Bajić · S. Jozić
Faculty of Electrical Engineering, Mechanical Engineering and Naval Architecture, University of Split,
Ruđera Boškovića bb,
21000 Split, Croatia
e-mail: blela@fesb.hr

El-Sonbaty et al. [7] used back-propagation neural network approach to the problem.

Bayesian neural network (BNN) and support vector machines (SVM) have been also used in modeling face milling process not for the surface roughness prediction but for the tool wear estimation and tool breakage detection. For instance, Dong et al. [8] used Bayesian multilayer perceptrons and Bayesian SVM for tool wear estimation, while Choa et al. [9] and Hsueh and Yang [10] used SVM methodology for tool breakage detection.

The goal of this study is to obtain a mathematical model that relates the surface roughness with the cutting parameters in face milling. Precisely, the study examines the influence of:

the cutting speed, v_c,
the feed, f, and
the depth of cut, a_p,

on the surface roughness.

To get a mathematical model, three different approaches have been used in this work. The first approach is a DOE together with the analysis of variance (ANOVA) and RA. The second approach is modeling by means of SVM, and the third modeling method is BNN. This study compares the three modeling methodologies that differently infer the parameters of models.

2 Surface roughness

The surface parameter used to evaluate surface roughness in this study is the roughness average (Ra). This parameter is also known as the arithmetic mean roughness value, arithmetic average (AA), or centerline average (CLA) [3]. Parameter Ra is the most commonly used surface roughness assessment parameter. The average roughness is the area between the roughness profile and its centerline or the integral of the absolute value of the roughness profile height over the evaluation length [3]. Therefore, Ra is specified by the following equation:

$$\mathrm{Ra} = \frac{1}{l}\int_0^l |y(x)|\,dx, \tag{1}$$

or when evaluated from digital data, the integral is approximated by a trapezoidal rule:

$$\mathrm{Ra} \approx \frac{1}{n}\sum_{i=1}^{n} |y_i|, \tag{2}$$

where Ra is the arithmetic average deviation from the mean line (micrometers), l is the sampling length, and y is the ordinate of the profile curve.

3 Experimental settings

Test samples made of steel St 52-3 (Deutsches Institut für Normung [DIN] designation), with dimensions 230×100×100 mm, were prepared to remove rust, grooves, and all damages from the top surface that is to be machined. Chemical composition and mechanical properties of test samples are given in Table 1.

The face milling experiments were performed by a face mill Helido 45° S845 F45SX D063-05-22-R16 using inserts with eight helical right-hand cutting edges, produced by Iscar. The width of cut was 50% of tool diameter, and the milling process has been performed by the down-milling method. The type of machine used for the milling test was machining center VC 560 manufactured by Spinner. All experiments were carried out with cooling and lubrication agent utilizing system integrated within the milling machine.

The surface roughness values of finish-milled workpieces were measured by Surtronic 3+ instrument, produced by Rank Taylor Hobson. The measurements were repeated ten times. To measure the roughness of a surface formed by processing the workpiece, the cutoff length is taken as 0.8 mm and the sampling length as 5.6 mm. The temperature of environment was 20±1°C.

4 DOE and RA

The milling is characterized with many factors, which directly or interconnected act on the course and outcome of an experiment. It is necessary to manage the experiment with the statistical multifactor method due to the statistical character of a machining process. In this work, the DOE was achieved using the rotatable central composite design (RCCD) [11]. In the experimental research, modeling, and adaptive control of multifactor processes, the RCCD of experiment is very often used because it offers optimization possibility.

The RCCD models the response using the empirical second-order polynomial:

$$y = b_0 + \sum_{i=0}^{k} b_i \cdot X_i + \sum_{1\le i<j}^{k} b_{ij} \cdot X_i \cdot X_j + \sum_{i=1}^{k} b_{ii} \cdot X_i^2, \tag{3}$$

where

–b_0, b_i, b_{ij}, and b_{ii} are the regression coefficients and
–X_i and X_j are the coded values of input parameters.

To determine the required number of experimental points for RCCD, the following expression is used:

$$N = 2^k + 2k + n_0 = n_k + n_\alpha + n_0, \tag{4}$$

where

–k is the number of parameters,

Table 1 The chemical composition and mechanical properties of specimens

C (%)	Mn (%)	Si (%)	P (%)	S (%)	N (%)	Al (%)	Hardness (HB)	Tensile strength (MPa)	Yield strength (MPa)
0.20	1.70	0.55	0.045	0.045	0.009	0.02	180	580	340

–n_0 is the repeated design number of the average level, and
–n_α is the design number on the central axes.

Rotatability provides equal precision of estimation in all directions, and the central composite design is made rotatable by adding the points $\alpha=\pm1.682$ to the central axes.

The three-factorial RCCD of experiment demands eight experiments (three factors on two levels, 2^3), six experiments on the central axes, and six experiments on the average level, which makes a total of 20 experiments.

In this study, the region of interest for the cutting speed, depth of cut, and feed is 150 to 180 m/min, 1.3 to 2.7 mm, and 0.24 to 0.36 mm/tooth, respectively. Therefore, the points on the central axes have not been used in the testing procedure of the models.

The original values of the cutting parameters used in the experiment as well as their coded values are presented in Table 2. To collect data for ANOVA and RA, software package Design-Expert was used to generate (using the data points presented in Table 2) 20 experimental points.

Applying RA on the experimentally determined data, the regression coefficients were obtained and thereby the regression equation for the surface roughness too. After omitting the insignificant factors, the equation is defined as:

$$\begin{aligned}\mathrm{Ra} = 4.927 - 0.0574 \times v_c + 1.181 \times f - 0.0273 \\ \times a_p + 0.000178 \times v_c^2 + 11.618 \times f^2\end{aligned} \tag{5}$$

To allow performing both SVR and BNN modeling and the verification of all models, additional 15 experiments were conducted. For those experiments, the values of the cutting parameters were randomly chosen within the range of interest mentioned above. Altogether, 35 experiments were carried out, and 30 data pairs have been chosen for the procedure of training and testing of both SVR and BNN models. Five experiments were discarded because RCCD demands the six repetitions at the center point. Furthermore, ten data pairs have been randomly selected so as to form the testing data set. These ten data have been chosen in a way to obtain the cutting parameters having the values within the range of interest. Those ten chosen data can be seen in Table 3. The remaining 20 data pairs have been used in the training process.

Before the training and testing process, all input and output data were scaled within the interval −0.9 and 0.9.

Table 2 Physical and coded values of input factors

Coded parameters	Physical parameters		
	a_p (mm)	v_c (m/min)	f (mm/tooth)
−1.682	0.82	139.8	0.2
−1.0	1.30	150	0.24
0	2.00	165	0.30
1	2.70	180	0.36
1.682	3.18	190.2	0.4

5 SVM modeling

SVMs are a specific class of algorithms that are referred to as kernel methods that are well-known tools for classification and regression tasks. They have properties such as good generalization ability, few free adjusting parameters, and no requisite for experimentation for the purpose of finding the learning machine architecture [12]. SVMs were introduced by Vapnik [13, 14]. Although SVMs were originally developed for the classification tasks, they can be applied to regression problems (called support vector regression [SVR]) by the introduction of a loss function that includes a distance measure [14]. SVR constructs a linear model after the input has been mapped into a higher dimensional feature space using some nonlinear mapping (usually by reproducing kernels). A linear model in the feature space is:

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \tag{6}$$

The quality of estimation is measured by the ε-insensitive loss function proposed by Vapnik:

$$L_\varepsilon(y) = \begin{cases} 0 & \text{for} \quad |f(\mathbf{x}) - y| < \varepsilon \\ |f(\mathbf{x}) - y| - \varepsilon & \text{otherwise} \end{cases}, \tag{7}$$

where ε allows for some deviations between the targets y and the function $f(\mathbf{x})$. This can be visualized as a band or tube around the function $f(\mathbf{x})$, and points outside the tube can be viewed as training errors [12]. Along with the linear regression, SVM algorithm controls the model complexity by minimizing $\|w\|^2$. Measuring the deviation of training data points outside the ε-insensitive zone is performed by introducing nonnegative slack variables ξ_i, ξ_i^*. Therefore, the task of SVR is to minimize the following functional [12]:

$$\min\left[\|\mathbf{w}\|^2 + C\sum_{i=1}^{m}\left(\xi_i + \xi_i^*\right)\right], \tag{8}$$

Table 3 Testing the capability of all models in the predictions of surface roughness

v_c (m/min)	f (mm/tooth)	a_p (mm)	Experiment Ra (μm)	RA	BNN	SVR	Relative error obtained by RA (%)	Relative error obtained by BNN (%)	Relative error obtained by SVR (%)
173	0.28	2	0.45	0.4345	0.4627	0.4722	3.43	2.82	4.94
180	0.36	2.7	0.88	0.8927	0.9467	0.8164	1.45	7.58	7.23
175	0.25	2.5	0.37	0.3732	0.3783	0.3613	0.87	2.24	2.35
150	0.36	2.7	1.1	1.1601	1.1452	1.1533	5.47	4.11	4.85
174	0.31	1.8	0.59	0.5393	0.5853	0.6247	8.60	0.79	5.88
168	0.29	2	0.55	0.4996	0.5584	0.5683	9.16	1.54	3.33
165	0.35	1.9	1.01	0.8117	0.9318	0.9609	19.63	7.75	4.86
169	0.26	2.2	0.4	0.3960	0.4245	0.4133	1.00	6.13	3.32
150	0.24	2.7	0.47	0.5355	0.5476	0.5709	13.94	16.50	21.46
163	0.26	1.3	0.42	0.3573	0.3715	0.3363	14.93	11.56	19.93
Average relative error (%)							7.85	6.10	7.82

subject to

$$y_i - \mathbf{w}\cdot\mathbf{x}_i - b \le \varepsilon + \xi_i, (\mathbf{w}\cdot\mathbf{x}_i + b) - y_i \le \varepsilon + \xi_i^*, \xi_i, \xi_i^* \ge 0, i = 1, \ldots, m. \tag{9}$$

To be able to estimate a nonlinear function, a kernel function is substituted, and the dual Lagrangian objective function is obtained as:

$$W(\alpha,\alpha^*) = \sum_{i=1}^{m} y_i\left(\alpha_i + \alpha_i^*\right) - \varepsilon\sum_{i=1}^{m}\left(\alpha_i + \alpha_i^*\right) - -\frac{1}{2}\sum_{i,j=1}^{m}\left(\alpha_i + \alpha_i^*\right)\left(\alpha_j + \alpha_j^*\right)K\left(x_i, x_j\right) \tag{10}$$

where α and α* are Lagrangian multipliers. The objective function is maximized subjected to constraints:

$$\sum_{i=1}^{m}\left(\alpha_i - \alpha_i^*\right) = 0, 0 \le \alpha_i, \alpha_i^* \le C, i = 1, \ldots, m. \tag{11}$$

The regression function is given as:

$$f(\mathbf{z}) = \sum_{i=1}^{SV}\left(\alpha_i - \alpha_i^*\right)K(\mathbf{x}_i, \mathbf{z}) + b \tag{12}$$

SV is the number of support vectors, and bias b is computed by considering Karush–Kuhn–Tucker conditions for regression. The optimization problem has been solved by linear programming using interior point method [15].

It is obvious that estimation accuracy of SVM depends on kernel parameters, capacity C and insensitivity region ε. The value of ε influences the number of support vectors used to form the regression function. If ε increases, fewer support vectors are chosen, and the smoothness of the regression function increases too. The capacity C represents a trade-off cost between the empirical error and the model complexity (flatness).

In this research, radial basis function kernel was used. The kernel parameter σ and the parameter ε were determined by a leave-one-out (LOO) cross-validation procedure [16, 17]. All 30 data points were used in the LOO procedure. First, the capacity C was chosen to be constant and set to 10, and ε was set to 0.21. Then, the LOO procedure was preformed to find σ, which minimizes the mean squared error (MSE). The minimal MSE was achieved at σ=1.7. Afterward, the same procedure was repeated but now to find ε, which minimizes the MSE, and in this case, σ was set to 1.7. It was found that ε=0.13 gives the minimal value of the MSE. In both cases, the minimal values of the MSE were found using simplex optimization algorithm [18]. The results of the LOO procedure in a graphical form are shown on Fig. 1, precisely for kernel parameter (Fig. 1a) and parameter ε (Fig. 1b).

After both parameters have been found, the SVM model is ready for the learning process. After the learning process, the final number of support vectors was 9. The effectiveness of training of the SVM model is depicted on Fig. 2a and its ability to predict correct values for unseen data on Fig. 2b. The values of R very close to 1 for both training and testing indicate that the model has been very well learned and that also has the excellent generalization ability.

6 BNN modeling

Feed forward neural networks are well-known modeling tools for complex nonlinear problems. The aim is always to infer the function from the given data set. The data set usually consists of input vectors $\boldsymbol{x}$ and corresponding target vectors $\boldsymbol{t}$. Neural networks parameterize the unknown function by means of parameter vector $\boldsymbol{w}$, which leads to a nonlinear function $y(\boldsymbol{x};\ \boldsymbol{w})$ [19]. By inferring the parameter $\boldsymbol{w}$, the function $y(\boldsymbol{x};\ \boldsymbol{w})$ is inferred as well. This is achieved by adjusting $\boldsymbol{w}$ so as to minimize an error function [20], which

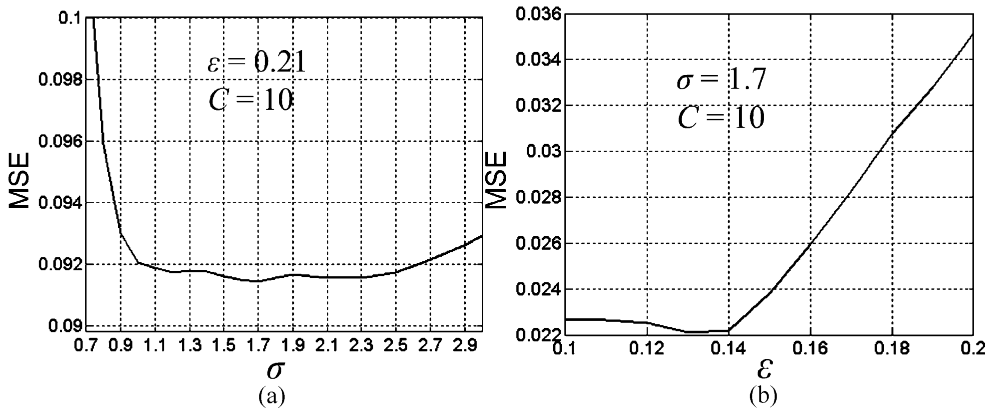

Fig. 1 Results of LOO procedure in determining kernel parameter σ (**a**) and parameter ε (**b**)

is usually the sum-of-squares error (SSE). The process of adjusting $\boldsymbol{w}$ is also called learning or training. This research uses the BNN modeling approach. BNN gives a probabilistic interpretation to the network learning process. The SSE function is expressed in terms of the likelihood function representing probability of the observed data when the parameters are known [20]. This function is usually taken to be a separable Gaussian assuming zero-mean Gaussian noise on the target variables and the hyperparameter β controls the variance of the noise [19]. BNN includes regularization, and the regularizer is interpreted as prior probability distribution over the parameters. For a weight-decay regularizer, prior distribution is a Gaussian where variance is governed by the hyperparameter α [9].

Once the likelihood function and a prior distribution have been determined, the Bayes' theorem can be used to find the posteriori distribution of the network weights. Having found the posteriori distribution, predictions can be made by marginalization over the parameters [20]. BNN automatically controls the network complexity, so there is no need for validation data set and cross-validation procedure.

BNN model used in this study consists of three layers and a hidden layer of five neurons. The input layer is composed of three neurons. The output layer has one neuron and a linear activation function. To optimize BNN weights $\boldsymbol{w}$, the scaled conjugate gradient algorithm [20] has been used. The final values of the hyperparameters have been obtained as α=2.88, β=20.39.

The R values of 0.987 and 0.984 have been obtained for training and testing, respectively.

7 Analysis of results and discussion

The validation of all models was performed with the testing data set. Relative errors obtained using RA, BNN, and SVR methodologies have been compared, and the results of testing are shown in Table 3.

The results from Table 3 indicate that BNN model offers the best prediction capability. The average relative errors for RA and SVR models are nearly identical.

After the validation of the models, a simulation of the prediction, using the best model, was performed. The

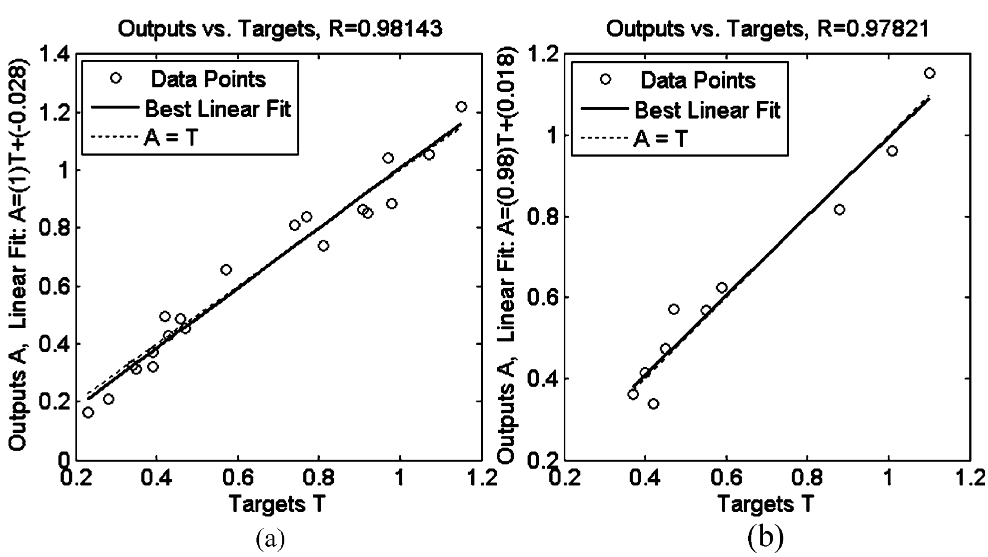

Fig. 2 Results of SVR for training (**a**) and testing (**b**)

simulation was carried out by varying two cutting parameters, while the third is kept constant. The results obtained that way, using BNN model, are given in a graphical form on Figs. 3, 4, and 5.

Figure 3 shows how the surface roughness depends on the feed and depth of cut in the case when the cutting speed of 165 (m/min) is kept constant. It can be seen that both factors influence on *Ra* but the feed is by far more dominant factor.

Figure 4 shows the influence of the cutting speed and feed on the surface roughness for the invariable dept of cut of 2 (mm). It is obvious that Ra decreases when the cutting speed increases and the feed simultaneously decreases. The feed is again, by far, a more influential factor.

The influence of the cutting speed and depth of cut on the surface roughness for the constant feed of 0.3 (mm/tooth) is shown on Fig. 5. Both factors show the similar intensity of influence on Ra. The surface roughness decreases if the cutting speed increases. The depth of cut has the opposite effect; that is, Ra decreases when a_p decreases.

Generally, the surface roughness decreases with the increase of cutting speed. This is usually explained relating to the type of chip formation and built-up edge generated from the machining process. At very low cutting speed, discontinuous chip formation occurs, which gives a poor surface finish. As the cutting speed increases, the chip formation becomes less discontinuous, and the surface finish improves. Further increase of cutting speed reduces the size of the built-up edge until a continuous chip is formed, and then, surface roughness approaches a steady low value.

Considering the fact that the research has been made above the cutting speed range where built-up edge appears, the increase of cutting speed leads to the decrease of surface roughness. Further increase of the cutting speed causes tool

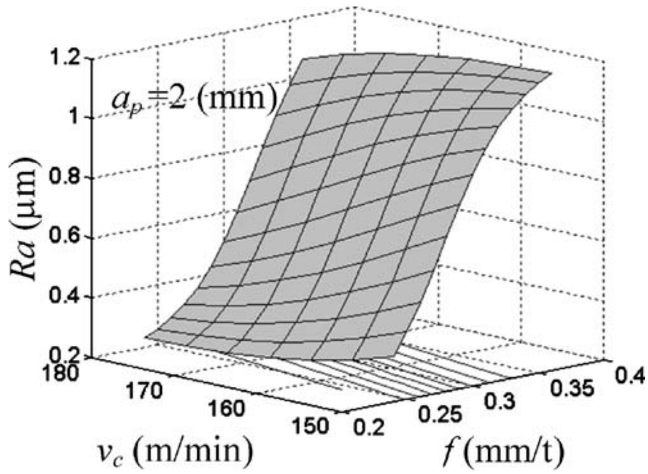

Fig. 4 Influence of cutting speed and feed on surface roughness for constant depth of cut of 2 (mm)

wear, and that maintains approximate constant value of surface roughness.

The feed is the most significant factor associated with surface roughness. When feed increases, surface roughness increases too. This is expected, as it is well known that, for a given tool, nose radius, the theoretical surface roughness ($\mathrm{Rt} \cong f^2/(8 \times r_\varepsilon)$) is mainly a function of the feed. The selection of the feed must be performed carefully, because apart from the strong influence on Ra, excessive feed increases cutting forces, tool deflections, tool wears, chipping, etc.

In comparison with cutting speed and feed, the depth of cut has a minor influence on surface roughness. From a geometrical point of view, depth of cut has not influence surface roughness because the height and form of the roughness profile are independent of the depth of cut. The depth of cut has indirect influence on surface roughness through the formation of the built-up edge, chip deformation, cutting force, cutting temperature, vibration, etc.

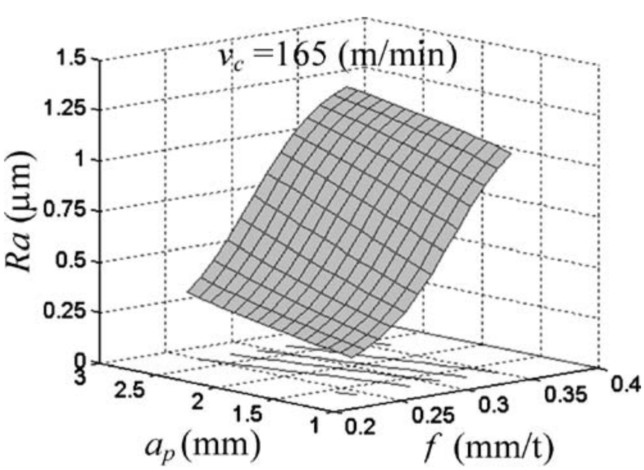

Fig. 3 Influence of feed and depth of cut on surface roughness for constant cutting speed of 165 (m/min)

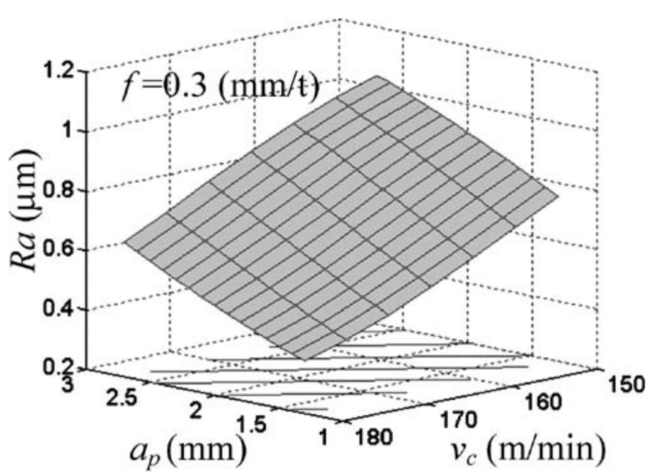

Fig. 5 Influence of cutting speed and depth of cut on surface roughness for constant feed of 0.3 (mm/tooth)

It is possible to select a combination of cutting speed, depth of cut, and feed for achieving the desired surface finish. The application of the present approach to obtain machining conditions may be quite useful in computer-aided process planning. Thus, with the known boundaries of the desired surface roughness parameter and machining parameters, machining can be performed with a relatively high rate of success.

8 Conclusion

In this study, the influences of cutting speed, feed, and depth of cut on surface roughness in face milling process have been examined. Experiments have been carried out on the steel St 52-3 (DIN designation), and the results have been analyzed by means of RA, BNN, and SVR methodologies. All three models have the relative prediction error below 8%. The best prediction of surface roughness shows BNN model with the average relative prediction error of 6.1%. All models have been trained and tested with 20 and 10 experimental points, respectively.

The research has shown that, when the training dataset is relatively small (as in the study), both BNN and SVR modeling methodologies are comparable with RA methodology, and furthermore, they can even offer better results.

This research has also shown that the feed has the largest affect on surface roughness and the depth of cut the least.

Acknowledgments This work was supported by the Ministry of Science, Education and Sport of the Republic of Croatia.

References

1. Colak O, Kurbanoglu C, Kayacan MC (2005) Milling surface roughness prediction using evolutionary programming methods. Mater Des 28:657–666
2. Bajić D, Belaić A (2006) Mathematical modeling of surface roughness in milling process. In: Proceedings of the 1st International Scientific Conference on Production Engineering (ISC), Lumbarda, Croatia, June/July 2006, pp 109–115
3. Lou MS, Chen JC, Li CM (1998) Surface roughness prediction technique for CNC end-milling. J Inf Technol 15(1):2–6
4. Lu C (2008) Study on prediction of surface quality in machining process. J Mater Process Technol 205:439–450 doi:10.1016/j.jmatprotec.2007.11.270
5. Oktem H, Erzurumlu T, Kurtaran H (2005) Application of response surface methodology in the optimization of cutting conditions for surface roughness. J Mater Process Technol 170:11–16 doi:10.1016/j.jmatprotec.2005.04.096
6. Benardos PG, Vosniakos GC (2002) Prediction of surface roughness in CNC face milling using neural networks and Taguchi's design of experiments. Robot Comput Integr Manuf 18:343–354 doi:10.1016/S0736-5845(02)00005-4
7. El-Sonbaty IA, Khashaba UA, Selmy AI, Ali AI (2008) Prediction of surface roughness profiles for milled surfaces using an artificial neural network and fractal geometry approach. J Mater Process Technol 200:271–278 doi:10.1016/j.jmatprotec.2007.09.006
8. Dong J, Subrahmanyam KVR, Wong YS, Hong GS, Mohanty AR (2006) Bayesian-inference-based neural networks for tool wear estimation. Int J Adv Manuf Technol 30:797–807 doi:10.1007/s00170-005-0124-8
9. Choa S, Asfoura S, Onarb A, Kaundinya N (2005) Tool breakage detection using support vector machine learning in a milling process. Int J Mach Tools Manuf 45:241–249 doi:10.1016/j.ijmachtools.2004.08.016
10. Hsueh, YW, Yang CY (2008) Tool breakage diagnosis in face milling by support vector machine. J. Mater. Process. Tech. doi:10.1016/j.jmatprotec.2008.01.033
11. Box GEP, Draper NR (1987) Empirical model building and response surfaces. Wiley, New York
12. Campbell C (2002) Kernel methods: a survey of current techniques. Neurocomputing 48:63–84 doi:10.1016/S0925-2312(01)00643-9
13. Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
14. Vapnik VN (1998) Statistical learning theory. Wiley, New York
15. Nash S, Sofer A (1996) Linear and non-linear programming. McGraw-Hill, New York
16. Twomey JM, Smith AE (1998) Bias and variance of validation methods for function approximation neural networks under conditions of sparse data. IEEE Trans Syst Man Cybern 28(3):417–430
17. Rivals I, Personnaz L (1999) On cross-validation for model selection. Neural Comput 11(4):863–870 doi:10.1162/089976699300016476
18. Lagarias JC, Reeds JA, Wright MH, Wright PE (1998) Convergence properties of the Nelder–Mead simplex method in low dimensions. SIAM J Optim 9(1):112–147 doi:10.1137/S1052623496303470
19. Mackey DJC (1992) Bayesian methods for adaptive models. PhD Dissertation, California Institute of Technology
20. Bishop C (1995) Neural networks for pattern recognition. Oxford Univ. Press, UK