

# Modeling for a Small-hole Drilling Process of Engineering Plastic PEEK by Taguchi-based Neural Network Method

**Dar-Yuan Chang**

Chinese Culture University

**Chien-Hung Lin** (✉ [ljh13@ulive.pccu.edu.tw](mailto:ljh13@ulive.pccu.edu.tw))

Chinese Culture University

**Xing-Yao Wu**

Chinese Culture University

---

## Research Article

**Keywords:** Neural networks, Stepped-learning procedure, Taguchi's method, Small-Hole drilling, Thrust force, Diameter shrinkage

**Posted Date:** July 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-621267/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at The International Journal of Advanced Manufacturing Technology on January 16th, 2022. See the published version at <https://doi.org/10.1007/s00170-021-08431-2>.

# **Modeling for a small-hole drilling process of engineering plastic PEEK by Taguchi-based neural network method**

Dar-Yuan Chang, Chien-Hung Lin\*, Xing-Yao Wu

Department of Mechanical Engineering,  
Chinese Culture University,  
Taipei, Taiwan (R.O.C.)

Corresponding Author: Chien-Hung Lin

55, Hwa-Kang Road,  
Yang-Ming-Shan, Taipei,  
Taiwan 11114, Republic of China  
Phone: 886-2-28610511 # 33311  
Email: [ljh13@ulive.pccu.edu.tw](mailto:ljh13@ulive.pccu.edu.tw)

# **Modeling for a small-hole drilling process of engineering plastic PEEK by Taguchi-based neural network method**

## **Abstract**

Engineering plastics have specific properties in the strength, hardness, impact resistance, and aging persistence, often used for structural plates and electronic components. However, the holes made by the drilling process always shrink after the cutting heat dispersion due to their high thermal expansion coefficient. Especially for small-hole fabrication, drilling parameters must be discussed thoughtfully to acquire a stable hole quality. This study developed parameter models by the Taguchi-based neural network method to save the experimental resources on drilling of engineering plastic, polyetheretherketone (PEEK). A three-level full-factorial orthogonal array experiment,  $L_{27}$ , was first conducted for minimizing the thrust force, hole shrinkage in diameter, and roundness error. The experiments were operated by a peck-drilling process with cyclic lubricant, and the diameter was 1 mm. In terms of the network modeling, four variables were designated to the input layer neurons including the spindle speed, depth of peck-drilling, feed rate, and thrust force detected; and that of the output layer were the diameter shrinkage and roundness of the hole drilled. The models were trained by a stepped-learning procedure to expand the network's field information stage by stage. After three stages of training, the models developed can provide precise simulations for the network's training sets and accurately predict the hole's characteristics for the non-trained cases.

Keywords: Neural networks, Stepped-learning procedure, Taguchi's method, Small-Hole drilling, Thrust force, Diameter shrinkage

## 1. Introduction

For a mechanical part made by cutting processes, usually, the first operation is the main body machining by means of turning or milling process, and uses the drilling process to finish the last hole features. [Esim and Yildirim \[1\]](#) indicated that many undesired effects occur during the drilling process due to the complexity of the drill and cutting geometry. [Roth et al. \[2\]](#) also believed the use of multiple cutting edges in the drilling process results in variable cutting speeds along these edges, which increase interactions between the chips and the cutting tool during chip evacuation and altered the heat transfer. Drilling is one of the most frequent machining operations, accounting for over 30% of all cutting operations in the industry [\[2\]\[3\]](#). As the product trends in modernization, lightweight, space reduction, and sustainability, fabrication of miniature parts and components is crucial in industry. Correlations between cutting parameters and characteristics of the hole are worth investigating especially for a small-hole process.

Artificial neural network (ANN) has an excellent ability to learn complex non-linear and multivariable relationships between process parameters and makes them very useful in many engineering applications. For instance, [Zerti et al. \[4\]](#) conducted turning experiments of the stainless steel of AISI 420 for minimizing the surface roughness, cutting force, and power. Two ANN models for surface roughness and thrust force were built based on the experimental parameter data of cutting speed, feed rate, and depth of cut. [Daniel et al. \[5\]](#) developed an ANN model for milling aluminum metal matrix composites using the back-propagation neural network (BPNN) algorithm. The input vectors were the mass fraction of SiC, particle size of SiC, speed, feed, and depth of cut, and the output vectors were the surface roughness, temperature, material removal rate (MRR), and thrust force components of  $F_x$ ,  $F_y$ , and  $F_z$ . [Yadav et al. \[6\]](#) presented an ANN model for electrical discharge diamond face grinding to predict the output responses of MRR and surface roughness for a given set of input parameters including the wheel speed, pulse current, pulse on-time, and duty factor. [Biswas et al. \[7\]](#) presented an ANN model developed by the BPNN for modeling the laser microdrilling process

of TiN-Al<sub>2</sub>O<sub>3</sub> composite. Five drilling parameters were designated to the input layer neurons, and the neurons for the output layer were the hole's taper and circularities at entry and exit. [Sarikaya and Yilmaz \[8\]](#) investigated the effects of process parameters on the profile of drilled holes for micro-electrical discharge drilling of the stainless steel of AISI 304. Neural network (NN) models were established for predicting the process efficiency and the hole's characteristics. A two-hidden-layer architecture was adopted in training, which can derive satisfactory models for output estimations.

In kinds of literature, drilling parameters of cutting velocity, spindle speed, feed rate, depth of drilling, and point angle are mostly analyzed in mechanical drilling processes; cutting characteristics of thrust force, torque, vibration, cutting temperature, driving power, and tool wear are discussed; and hole's quality characteristics of displacement of hole center, surface roughness of drilled wall, roundness, hole oversize, and burr are measured [\[9-10\]](#). Thrust force is an important cutting characteristic in drilling. Several ANN models have been presented as follows. [Patra et al. \[11\]](#) developed an ANN model based on thrust force signals and drilling parameters to predict the drilled hole number in micro peck-drilling process of the tool steel of AISI P20. The network was trained by BPNN with Levenberg-Marquardt (LM) algorithm. The predicted accuracy by the ANN model was better than that by the regression model done. [Corne et al. \[12\]](#) built an ANN model based on the data of spindle power to predict tool wear and to detect drill breakage during the drilling of Inconel steel. Their results showed that the power data integrated with an ANN system can be used for real-time tool monitoring and process control. [Murthy et al. \[13\]](#) analyzed the drilling process of glass fiber reinforced plastic (CFRP) composite laminates using Taguchi's method. The influences of drilling parameters on the thrust force were discussed. They proposed a dynamics modeling approach through a causal loop diagram, and the results showed a good agreement of more than 93% acceptance level with the experimental data. [Efkolidis et al. \[14\]](#) performed drilling experiments of the aluminum material of Al 7075 to investigate the cutting characteristics by the full-factorial method. Models for predicting the thrust force and torque were established by the ANN and response surface methodology

(RSM) methods.

Vibration induced during the machining process has adverse effects on the life span of the cutting tool, measurement accuracy of the workpiece, and quality of the surface machined. [Ulas et al. \[15\]](#) conducted drilling experiments of the tool steel of AISI D3 to determine the optimum cutting parameters by measuring the vibrations during the process, and saved the resources by creating a high-reliability ANN model. [Esim et al. \[1\]](#) presented two NN predictors modeling by BPNN and radial basis neural network to predict the vibration variation for both cases of steel and aluminum during the drilling system. [Caggiano et al. \[16\]](#) presented an ANN model for tool wear in the CFRP drilling based on the multiple sensor signals of thrust force, torque, acoustic emission, and vibration. For each training example, two principal component analysis features and the specific hole number were designated as the input layer neurons and the tool flank wear was the output layer neuron. The diagnosis of tool flank wear can be accurately simulated by the BPNN model developed.

ANN has excellent ability in the function mapping for nonlinear behaviors and derives the predictions of output variables of the input training patterns quickly. Therefore, it is very suitable to be used for estimating the quality characteristics in machining processes. [Vrabel et al. \[17\]](#) presented an ANN adaptive control system for the drilling process of a nickel-based superalloy of Udimet 720. The surface roughness was assessed based on the process monitoring indices such as torque, thrust force, power, feed rate, cutting spend, and tool flank wear. [Mondal et al. \[18\]](#) conducted an experimental study for analyzing burrs in aluminum alloy drilling using an  $L_{27}$  orthogonal array (OA). An ANN model was built and determined the optimal process parameters by flower pollination algorithm for diminishing the burr height. [Ahn and Lee \[19\]](#) conducted microdrilling experiments of ductile metals of copper and brass to observe the burr size and type, and established a burr predicted network by BPPN. The variables of the input layer neurons were the cutting speed, feed, drill diameter, hardness ratio, and tensile strength, and that of the output layer was the type of burr. The prediction accuracy was more than 80%. [Akıncioğlu et al. \[20\]](#) presented an ANN model for predicting the surface

roughness and roundness in the drilling of the cold work tool steel of AISI D2 by means of BPNN. The parameters analyzed were cutting tool, cutting speed, and feed rate. Results of the network's experiment showed that the best learning algorithm was the scale conjugate gradient algorithm. Cruz et al. [21] installed a multiple sensors system to monitor cutting conditions in the drilling of a metal sandwich plate. The signals detected included the acoustic emission, electric power, tri-axial cutting forces, and acceleration, and used to establish the ANN model for monitoring the hole diameter and roundness in operations.

Several scholars used the BPNN to develop their network models for the input patterns simulation or prediction in machining processes [1, 3-8, 11-12, 14-21]. Data of the experimental trials are divided randomly, approximately 60-70% for network training and the remaining for validating and testing. When the network training has met one of the converging conditions, the validation and testing procedures can then be started. The performance of the NN model can be evaluated by the mean square error (MSE) and the correlation coefficient ( $R$ ). For instance, Cruz et al. [21] proposed 162 trials, and the experimental data obtained were divided, with 60% for network training, 20% for validation, and 20% for testing. Ahn and Lee [19] conducted 60 datasets for NN modeling and sliced them into a training set of forty and a testing set of twenty. Yadav et al. [6] presented 31 sets of experimental data which were used as the inputs in network training, and a separate 8 experimental trials were arranged for network testing. Mondal et al. [18] performed the training ratio of 80%, testing ratio of 10%, and validation ratio of 10% used in NN modeling. At the same time, the prediction was validated with another  $L_9$  OA experimental data.

In the literatures cited, for the input patterns (training sets) used in network's training, the models built can derive precise simulations after some NN's experiments by different hidden layer architectures or training parameters. However, for the non-trained cases (beyond the training sets), the models might predict the output characteristics inaccurately and even occur overfitting. It is owing to the lack of filed knowledge. This study presents a stepped-learning procedure to advance the network's

training process. Through the differential analyses of the network’s training set, to decide some additional drilling trails and training set for the next stage training. Accordingly, the field information of the network model can be expanded in a planned way, and the prediction reliability would be improved as well. Figure 1 shows the flowchart of the stepped-learning procedure.

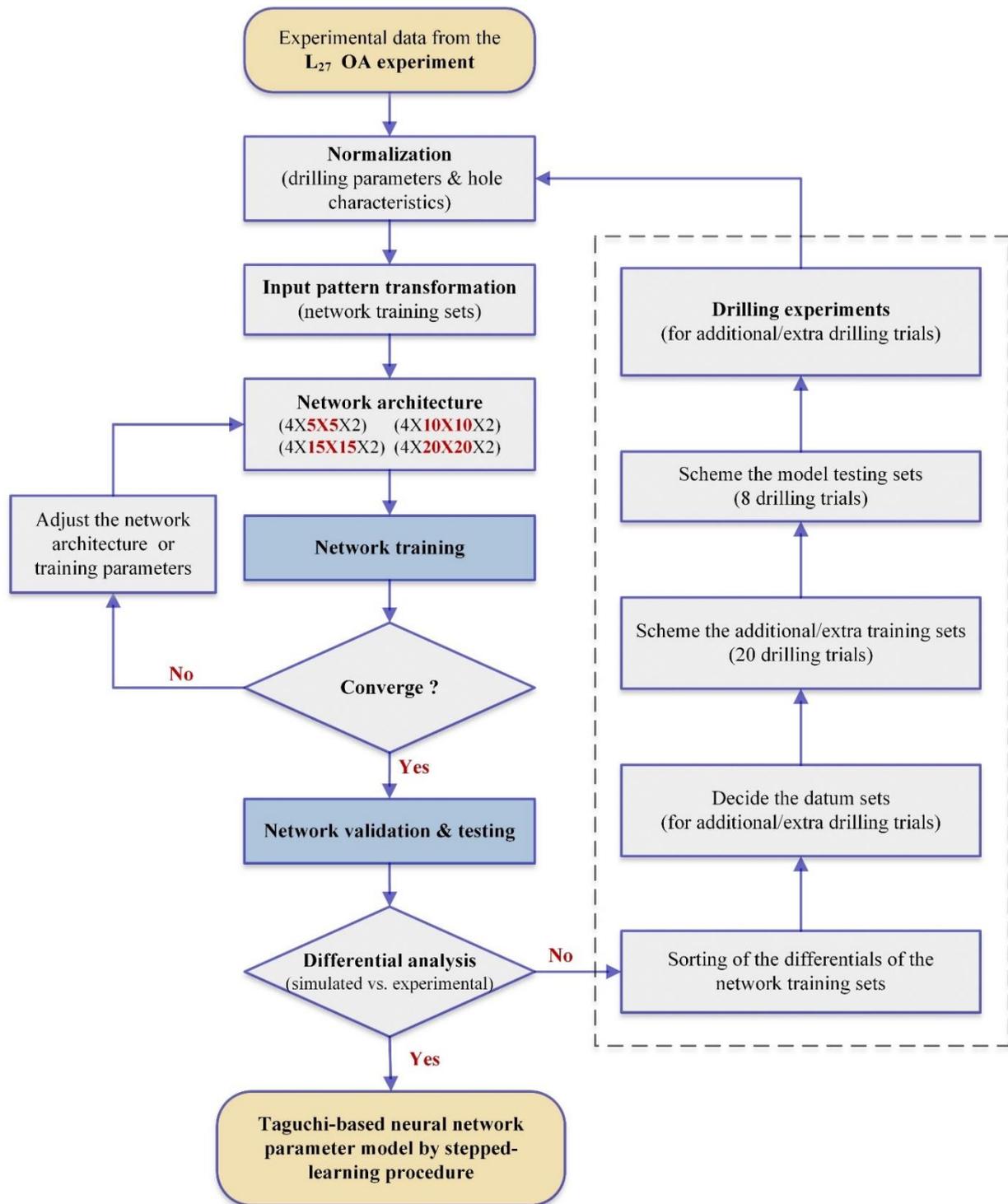


Fig. 1 Taguchi-based neural network model by a stepped-learning method

The remainder of this paper is presented as follows. Section 2 focuses on the machining setup and experimental plan, followed by the method of the Taguchi-based neural network in Section 3. Section 4 reports the results of  $L_{27}$  OA drilling experiment. Section 5 describes the procedure of NN modeling with data, followed by conclusions in Section 6.

## 2 Machining setup and experimentation

### 2.1 Material and Tools

Thermoplastic plastics have a specific property of high thermal expansion coefficient. It makes the cutting reaction different to the metal in the drilling process. The holes drilled will be shrunk while the workpiece is cooled. Therefore, the final hole diameter would be smaller than the drill diameter. In order to acquire a stable quality of the hole drilled, the selection of drilling parameters is very important. This study conducted small-hole drilling experiments of the polyetheretherketone (PEEK), MDS 100, which produced by Quadrant Systems Ltd, UK. [Table 1](#) presents its important properties.

**Table 1** Material properties of the PEEK—MDS 100

Property	Typical average value
Specific Gravity @73 °F (22.8 °C)	1.51 g/cm <sup>3</sup>
Tensile Strength	14,700 psi (101.4 MPa)
Elongation, at break	1,500,000 psi (10.3 GPa)
Flexural Strength	1.5 %
Flexural Modulus of Elasticity	20,500 psi (141.3 MPa)
Hardness, Rockwell	R121
Coefficient of Linear Thermal Expansion (-40 °F to 300 °F; -40 °C to 148.9 °C)	1.1×10 <sup>-5</sup> in/in/°F (1.98×10 <sup>-5</sup> mm/mm/°C)

Source: [www.mcam.com](http://www.mcam.com)

To avoid the tool “walking” on the workpiece surface at the beginning of the operation, a piloting hole was first made by a short-rigid drill of 0.5-mm diameter with a drill body of 0.8-mm length and a point angle of 90°. The drilling depth was 0.2 mm. After that, the peck-drilling cycle of the target hole of 1.0-mm diameter was started as shown in [Fig. 2](#). The retreating point was set at 0.2 mm above the workpiece surface. A tungsten carbide twist drill with a point angle of 118° was used made by Sphinx Tools Ltd, Swiss.

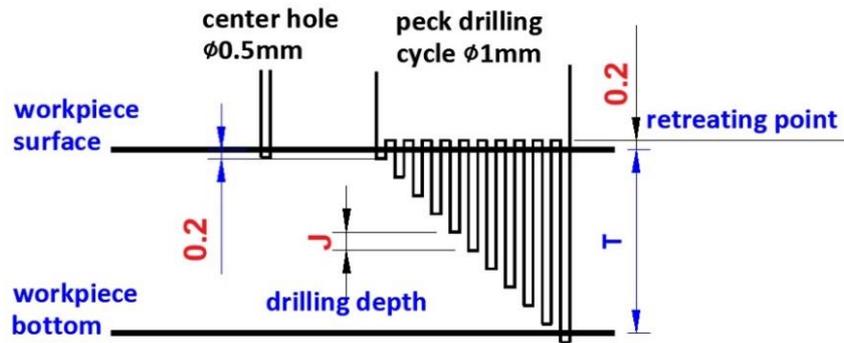


Fig. 2 Tool lifting in the peck-drilling cycle

## 2.2 Apparatus and machining setup

A vertical machine center (NXV-560A made by YCM Ltd, Taiwan) was used as shown in Fig. 3. The force system consists of a tri-axial force sensor (261A01, made by PCB Ltd, USA) and a data acquisition card (NI 9234, made by National Instrument, USA). Sensitivity of the force sensor on the Z axis is 0.56 mV/N, broadband resolution is 0.027 N-rms, and its low frequency response is 0.01 Hz. After the analogy signals transformed into the digital, import them to the computer and execute the signal analysis by a dynamic acquisition software (m+p analyzer, v. 5.2.1, made by m+p international, Germany).

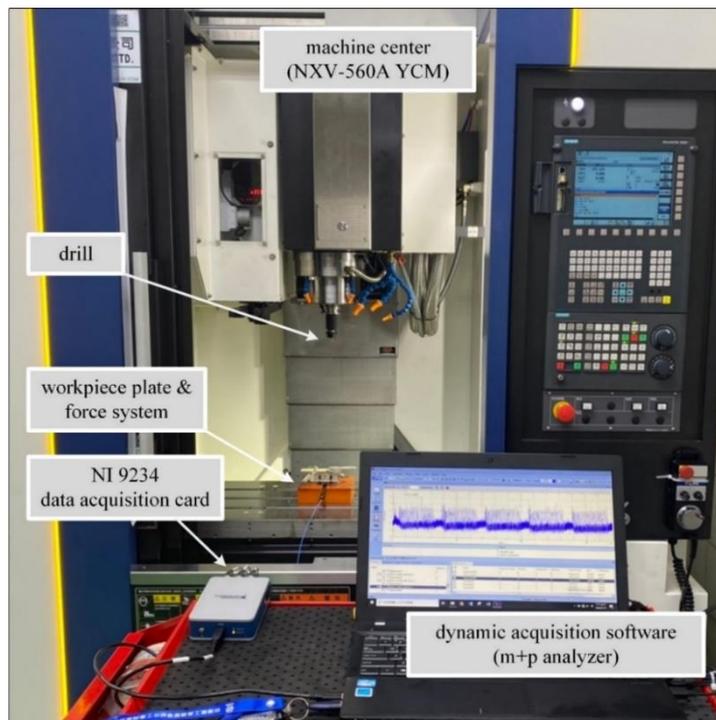


Fig. 3 Machine tool and force measurement system

This study designed a bakelite working base on the machine's platform for mounting the force sensor as shown in Fig. 4. An acrylic workpiece fixture was screwed on the top of the force sensor, and the workpiece can then be fasten by the four clamps. Figure 5 shows a PEEK workpiece plate of  $80\text{ mm} \times 30\text{ mm} \times 4.5\text{ mm}$ . Twenty-one position holes of 2-mm diameter are distributed over the workpiece plate. Every plate could arrange 20 drilling zones at the most. Because that burrs are easy to be formed during the plastic machining, a shallow square of  $7\text{ mm} \times 7\text{ mm} \times 0.25\text{ mm}$  was first milled for each drilling zone on both surfaces to prevent the measuring errors. Thus, the total drilling length of the peck-drilling cycle was actually 4 mm. To make sure that the drilling zone of each experimental trial can be positioned correctly, two pins with 2-mm diameter were inserted in the acrylic fixture for coupling with the position holes of the drilling zone. Each zone had five experimental holes, but only the third was the hole drilled directly above the force sensor (refer the drilling sequence marked in the right-enlarged drawing). Therefore, the force signals of the third hole were collected and used in the following thrust force analysis.

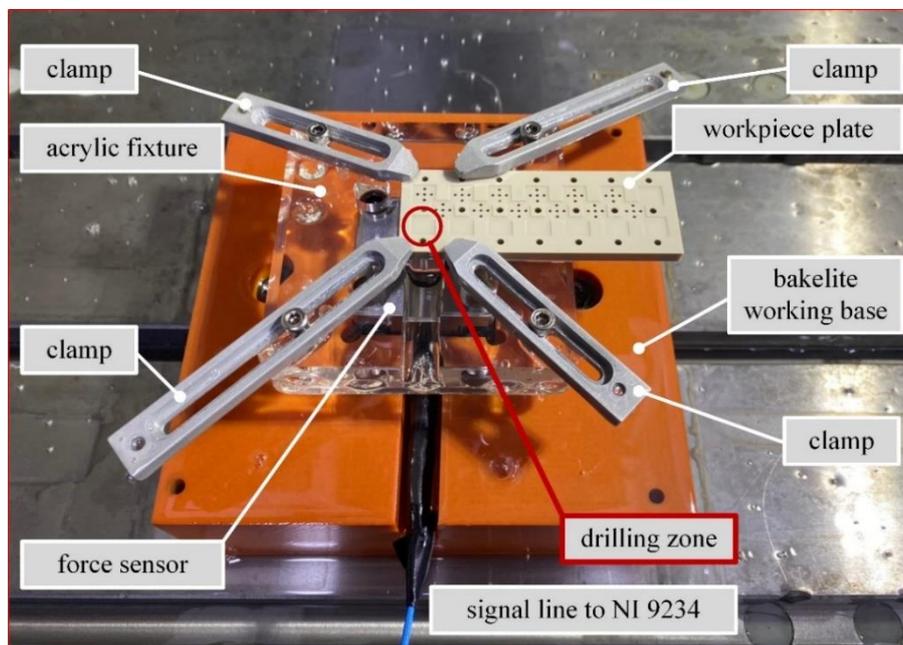


Fig. 4 Installation of the workpiece and force sensor.

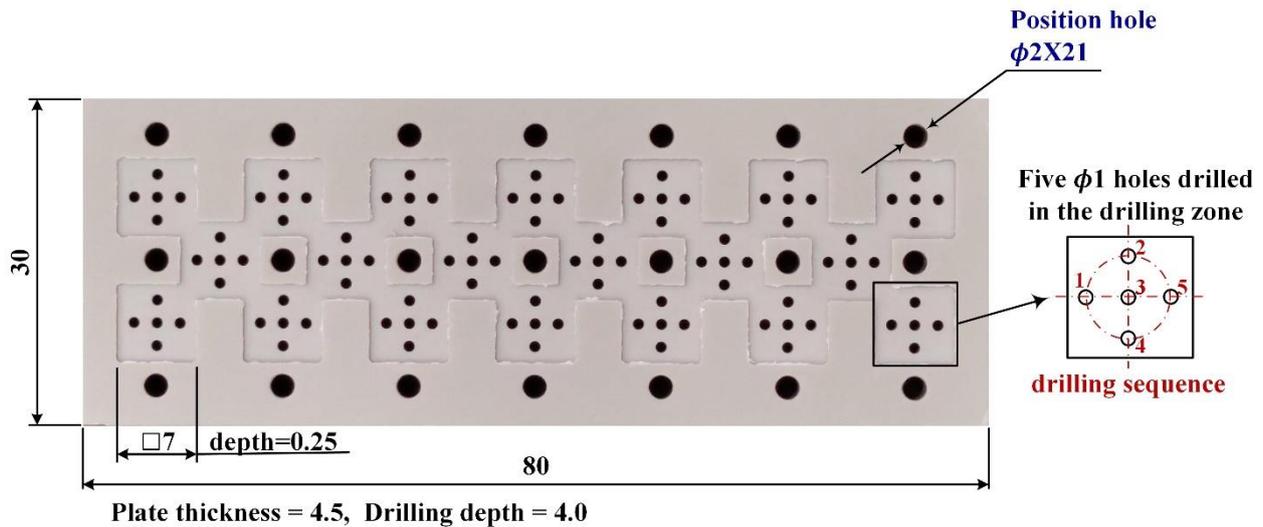


Fig. 5 A workpiece plate finished 20 experimental trials

In terms of the hole characteristics, an image measurement instrument (Baty-6490 Venture plus, UK) was used here. Its resolution is  $0.5 \mu\text{m}$  and measurement accuracy is  $2.5 \mu\text{m} + L/150$ .  $L$  is the measuring length in millimeter. To carry out the automatic inspection, a measurement original was first designated, and the five holes drilled in the identical trial were specified as a measuring group. Then an automatic measurement can be executed according to the sequence of the groups. Both characteristics of the hole diameter and the roundness were measured.

### 2.3 Experimental plan

Most research believes that cutting speed and feed rate are two important parameters affecting the performance in drilling processes, and they are usually selected as control factors in experiments. In order to avoid the adverse effects of chip piling in the small-hole fabrication, the holes drilled in this study were machined by the peck-drilling process and cooling cyclically with an oil lubricant (C-1170A-1, made by Peisun Chemical Co., LTD., Taiwan).

Accordingly, a full-factorial experiment with three three-level factors of spindle speed, depth of peck-drilling, and feed rate was considered in this work based on the Taguchi  $L_{27}$  OA. On the basis of the machining information provided by the material vender (www.mcam.com) and some preliminary trials, the levels of the control factors are presented in Table 2. Each parameter set in the  $L_{27}$  OA had

three replicate trials, and then data of the 81 drilling trials were obtained, included the thrust force and the two hole characteristics measured.

**Table 2** Control factors and corresponding levels for  $L_{27}$  OA experiment

Control factors	<i>A</i> Spindle speed (rpm)	<i>B</i> Depth of peck- drilling (mm)	<i>C</i> Feed rate (mm/min)
Level 1	3000	0.3	200
Level 2	4000	0.4	300
Level 3	5000	0.5	400

A smaller thrust force indicates the drilling resistance of the tool is low and the machining is in stable. A less shrinkage in diameter or a lower roundness indicates that the hole's quality is good. Therefore, the three characteristics mentioned in this experiment all belong to the smaller-the-better (STB) problem. According to the Taguchi method [22], the definition of single-to-noise (S/N) ratio is

$$S/N_{STB} = -10 \cdot \log(\bar{y}^2 + S^2), \quad (1)$$

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad (2)$$

$$S = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}, \quad (3)$$

where  $\bar{y}$  is the average value,  $S$  is the standard deviation, and  $n$  is the number of replication of each parameter set. As to the hole shrinkage, this study quantified the diameter shrinkage,  $\Delta D$  by the difference between the hole's diameter measured ( $D_{mea}$ ) and the tool diameter ( $D_{tool}$ ) expressed as

$$\Delta D = D_{tool} - D_{mea}. \quad (4)$$

### 3 Taguchi-based neural network model by a stepped-learning procedure

This study presents a stepped-learning procedure to accomplish the work of network training. Field information of the network model can be expanded step by step to achieve the goal of network training. Details of the training procedure are described as follows.

#### 3.1 Back-propagation neural network

ANN is a mathematical or computational model that is inspired by the structure or functional aspects

of biological neural networks. If a network is able to compute some functional relationship between its input and its output, this neural network is called a mapping network [23]. The BPNN is a layered, feedforward network, and fully interconnected by layers. There are no feedback connections and no connections that bypass one layer to go directly to a later layer in this network. It can be used to solve complex pattern-matching problems. Figure 6 shows the architecture of a BPNN network with a double-hidden-layer built in this study. Four variables were designated to the input layer neurons including the spindle speed, depth of peck-drilling, feed rate, and thrust force, and the variables of the output layer were two hole characteristics of diameter shrinkage and roundness.

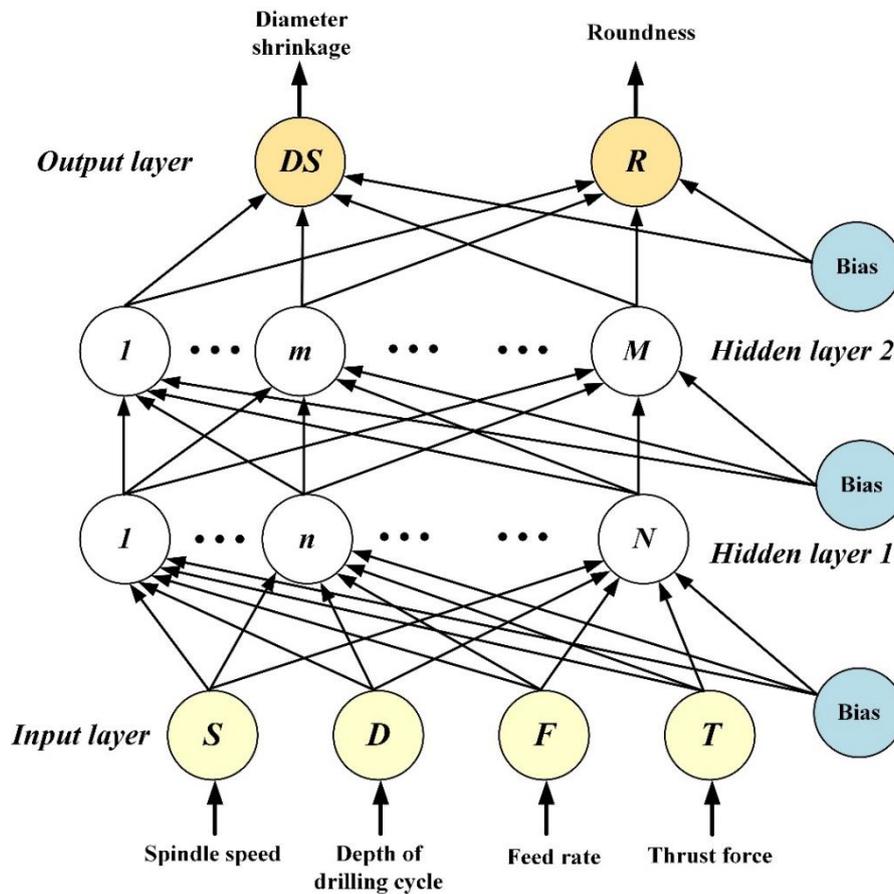


Fig. 6 Architecture of the BPNN model with double hidden layers

Several NN training algorithms have been proposed such as LM, bayesian regularization, and scaled conjugate gradient. LM algorithm is a combination of two training processes: steepest descent algorithm and Gauss-Newton algorithm, the network's convergence speed can be improved significantly [11]. Sarikaya & Yilmaz [8] also indicated that LM algorithm operates quicker once it

trains a medium-sized feed forward neural network. This study built the drilling parameters model by BPNN network trained with LM algorithm. For a neural network, learning means setting a proper architecture with sufficient neurons to find appropriate connected weights for modeling. In this study, four types of double-hidden-layer neurons, 5×5, 10×10, 15×15, and 20×20, were experimented to determine the network architecture (see Fig. 6). The network was developed by the MATLAB neural network tool box (v. 2019a).

### 3.2 Input pattern transformation

Before network training, experimental data obtained from the  $L_{27}$  OA experiment must be transformed into the input patterns by means of a normalization process. Values of the input and output variables have to be normalized between 0 and 1. The range in the normalization of each variable is schemed by the parameter values experimented with or the hole characteristics measured. However, to expand the field information of the network model and to predict the output characteristics beyond the training sets the range in the normalization of each variable is slightly greater than that of experimental parameter values and hole characteristic measurement values. This means that the minimum and maximum values in the normalization are slightly lower and higher than the minimum and maximum values of the experimental data, respectively. The normalized value ( $x_{nor}$ ) is calculated by

$$x_{nor} = \frac{x - x_{min}}{x_{max} - x_{min}} \in [0, 1], \quad (5)$$

where  $x$  is the value of the variable,  $x_{min}$  and  $x_{max}$  are the minimum and maximum values in normalization, respectively.

### 3.3 Training by a stepped-learning procedure

This study trained the NN parameter model by a stepped-learning procedure as shown in Fig. 1. In the first-stage training, the sampling distribution is selected as the defaulted ratio of MATLAB; i.e., 70% for training (57 out of 81), 15% for validation (12 out of 81), and the remaining 15% for testing, randomly sampled from the 81 input patterns. The transfer function of the hidden layer is tan-sigmoid

(Tansig), and that of the output layer is pure linear (Purelin). Performance of sample regression can be evaluated in terms of the MSE and  $R$  values. The MSE is given as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - y_{i\_exp})^2, \quad (6)$$

where  $n$  is the number of the training sets,  $y_i$  is the simulated outputs of the training set examined, and  $y_{i\_exp}$  is its expected value in training. The  $R$  of the model is calculated by

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (7)$$

where  $x$  and  $y$  are the experimental values and the model's simulated values, respectively, and  $\bar{x}$  and  $\bar{y}$  are their averages. A lower MSE and a larger  $R$  indicate a better performance.

Three converging conditions are set here to terminate the NN training as follows:

- (1) The maximum number of epochs in training is reached in 1000.
- (2) The minimum gradient descent in training is below  $1 \times 10^{-7}$ .
- (3) The MSE of the validation sets cannot be descended during 6 continuous epochs.

When the NN training is stopped, differential analyses are conducted for checking the deviation of every training set in network simulations. Afterward, to select the training sets that underperformed in network training as the “*datum sets*” for planning the additional training sets and the model's testing sets. The former is used to expand the field information that the model lacked, and the latter is prepared for estimating the prediction accuracy only, not involved in the training process.

The second-stage refining procedure aims to improve the simulation precision and prediction accuracy by adding some additional drilling trials that underperformed in the first-stage training. In this stage, all the trials obtained from the drilling experiments are designated as the network's training sets, no validation sets and testing sets again. Prediction accuracy of the non-trained cases is estimated by a specific model's testing sets, and termination of the network training is determined by the MSE.

In this study, three MSE conditions were experimented including  $MSE < 0.001$ ,  $MSE < 0.005$ , and  $MSE < 0.010$ .

The model's prediction accuracy achieved can be estimated in terms of root mean square error, RMSE, calculated by the simulated value of the output variable ( $y_j$ ) and its corresponding measured data ( $y_{j\_mea}$ ).

$$RMSE = \sqrt{\frac{1}{m} \sum_{j=1}^m (y_j - y_{j\_mea})^2} \quad (8)$$

## 4 Results of Taguchi's drilling experiment

According to the factorial levels presented in [Table 2](#), the  $L_{27}$  OA experiment was performed and obtained the data of 81 drilling trials. The results were reported below.

### 4.1 Cutting characteristic—thrust force

Three levels of peck-drilling depth (Factor  $B$ ) were analyzed here including 0.3, 0.4, and 0.5 mm, whereas the numbers of drilling cycles to finish a hole were 14, 10, and 8, respectively. [Figure 7](#) shows time-domain diagrams of the thrust force in the trials of Trial-1-1, Trial-4-1, and Trial-7-1 that they were operated by an identical spindle speed of 3000 rpm and a feed rate of 200 mm/min but with different peck-drilling depths. We found that the thrust force peak during the first drilling cycle was higher than that of the following cycles, and the force decreased abruptly when the drill pierced the bottom surface. Hence, the force signals of the first and last cycles were excluded in analyses. [Table 3](#) presents the thrust force data and S/N ratio of the  $L_{27}$  OA experiment. The averages of thrust force peaks of the three trials above mentioned were 4.50, 4.26, and 4.07 N, respectively. The high-level drilling depth ( $B3$ ) resulted in a lower thrust force (Trial-7-1).

In the drilling process of metal with a lower feed rate, as number of holes increases, tool wear increases rapidly due to ploughing and rapid heating that contribute to average thrust force increasing [\[11\]](#). However, in our drilling experiments of plastics—PEEK, the cutting heat results from a lower feed rate would soften the material and decrease the thrust force. [Figure 8\(a\)](#) shows the factorial level

effects analyzed by the S/N ratio. The sequence affecting the thrust force was feed rate (Factor  $C$ ), spindle speed (Factor  $A$ ), and depth of peck-drilling (Factor  $B$ ). A high-level spindle speed ( $A3$ ) with a low-level feed rate ( $C1$ ) obtained the best performance in the thrust force. The effect of drilling depth (Factor  $B$ ) was insignificant. From the statistics of the full-factorial experiment,  $L_{27}$  OA (Table 3), the optimal factorial level set was  $A3, B2,$  and  $C1$  (Trial-22) that had the smallest thrust force of 3.01 N. On the contrary, the worst one was  $A1, B2,$  and  $C3$  (Trial-6) that it enlarged to 8.55 N.

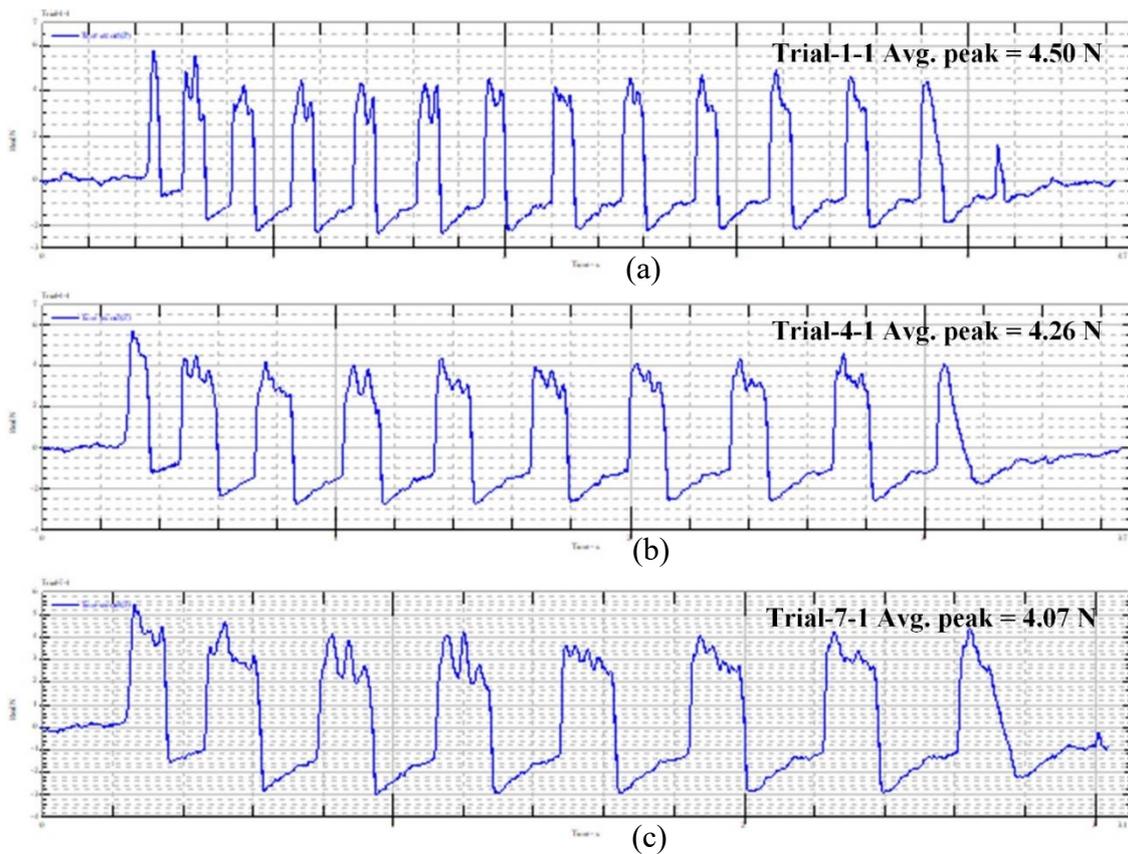


Fig. 7 Time-domain diagrams of the thrust force: (a) Trial-1-1, (b) Trial-4-1, and (c)

Trial-7-1

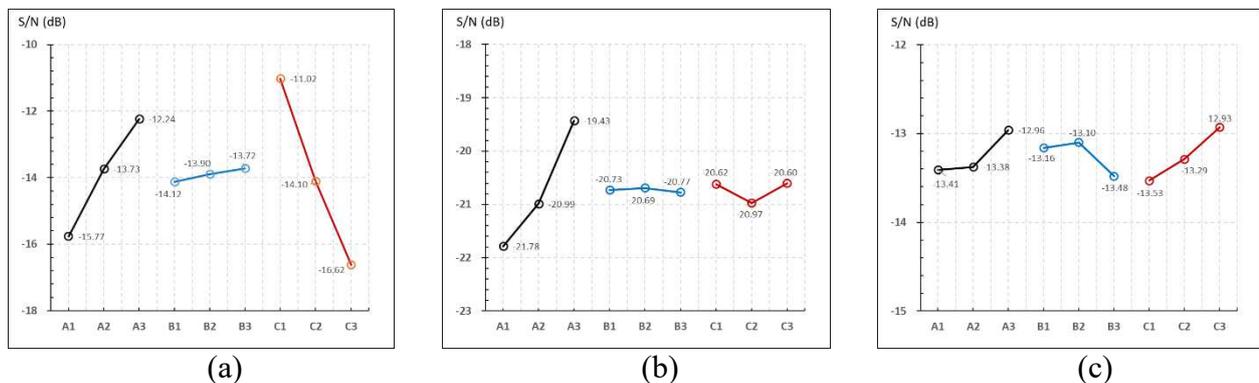


Fig. 8 Factorial level effect plots: (a) thrust force, (b) diameter shrinkage, and (c) roundness

**Table 3** Data of the thrust force and S/N ratio of the  $L_{27}$  OA experiment

$L_{27}$ OA	Control factors			Avg. of thrust force peaks (N)			Mean (N)	Std. dev. (N)	S/N (dB)
	A	B	C	N1	N2	N3			
Trial-1	1	1	1	<b>4.50</b>	4.47	4.38	4.45	0.049	-12.97
Trial-2	1	1	2	6.61	6.55	6.55	6.57	0.031	-16.35
Trial-3	1	1	3	8.21	8.45	8.52	8.39	0.133	-18.48
Trial-4	1	2	1	<b>4.26</b>	4.17	4.16	4.20	0.046	-12.46
Trial-5	1	2	2	6.50	6.42	6.50	6.47	0.037	-16.22
<b>Trial-6</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>8.56</b>	<b>8.58</b>	<b>8.50</b>	<b>8.55</b>	<b>0.035</b>	<b>-18.64</b>
Trial-7	1	3	1	<b>4.07</b>	4.18	4.03	4.09	0.062	-12.24
Trial-8	1	3	2	6.34	6.28	6.23	6.29	0.044	-15.97
Trial-9	1	3	3	8.49	8.50	8.54	8.51	0.020	-18.60
Trial-10	2	1	1	3.65	3.59	3.62	3.62	0.025	-11.18
Trial-11	2	1	2	5.07	5.11	5.00	5.06	0.044	-14.08
Trial-12	2	1	3	6.84	6.81	6.83	6.83	0.015	-16.69
Trial-13	2	2	1	3.51	3.39	3.40	3.43	0.052	-10.71
Trial-14	2	2	2	4.96	4.99	4.93	4.96	0.025	-13.91
Trial-15	2	2	3	6.78	6.82	6.78	6.79	0.016	-16.64
Trial-16	2	3	1	3.33	3.34	3.35	3.34	0.008	-10.47
Trial-17	2	3	2	4.60	4.78	4.79	4.72	0.089	-13.48
Trial-18	2	3	3	6.63	6.59	6.68	6.63	0.037	-16.43
Trial-19	3	1	1	3.18	3.13	3.17	3.16	0.021	-9.99
Trial-20	3	1	2	4.21	4.27	4.24	4.24	0.026	-12.55
Trial-21	3	1	3	5.55	5.56	5.42	5.51	0.062	-14.82
<b>Trial-22</b>	<b>3</b>	<b>2</b>	<b>1</b>	<b>2.99</b>	<b>2.99</b>	<b>3.05</b>	<b>3.01</b>	<b>0.027</b>	<b>-9.57</b>
Trial-23	3	2	2	4.15	4.07	4.02	4.08	0.052	-12.22
Trial-24	3	2	3	5.42	5.44	5.43	5.43	0.007	-14.69
Trial-25	3	3	1	2.99	3.04	3.03	3.02	0.022	-9.61
Trial-26	3	3	2	4.03	4.09	4.01	4.04	0.034	-12.14
Trial-27	3	3	3	5.34	5.36	5.35	5.35	0.005	-14.57

## 4.2 Hole characteristics—diameter shrinkage and roundness

There were 5 holes drilled in each drilling zone (see Fig. 5); therefore, the factorial level effect was evaluated by the average of the hole characteristics measured. Three replicas were done for every parameter set in the  $L_{27}$  OA experiment. Table 4 presents the statistics of the two hole characteristics, diameter shrinkage ( $\Delta D$ ) and roundness, and the S/N ratio. For the diameter shrinkage, the factorial level effects analyzed by the S/N ratio are shown in Fig. 8(b). The factor of the spindle speed (Factor A) presented the greatest significance, and the high-level spindle speed (A3) still obtained the smallest hole shrinkage in diameter. It is because that the peck-drilling process was adopted and the drill was lifted up above the workpiece surface during peck-drilling cycle (see Fig. 2), the chips can be smoothly removed from the machining hole. In addition, the liquid lubricant was sufficiently applied;

consequently, the adverse effect of the high-level spindle speed on the hole shrinkage was unobvious. From the statistics of the  $L_{27}$  OA experiment (see Table 4), the parameter sets of Trial-23 and Trial-20 had the smallest means of 8.22 and 8.24  $\mu\text{m}$ , respectively. The optimal factorial level set was  $A_3, B_3$ , and  $C_2$  (Trial-23).

**Table 4** Data of hole characteristics and S/N ratio of the  $L_{27}$  OA experiment

$L_{27}$ OA	Control factors			Avg. of dia. shrinkage ( $\Delta D$ )			Avg. of roundness		
	A	B	C	Mean ( $\mu\text{m}$ )	Std. dev. ( $\mu\text{m}$ )	S/N (dB)	Mean ( $\mu\text{m}$ )	Std. dev. ( $\mu\text{m}$ )	S/N (dB)
Trial-1	1	1	1	11.33	0.961	-21.11	4.39	0.560	-12.91
Trial-2	1	1	2	13.41	2.047	-22.65	4.86	0.867	-13.87
Trial-3	1	1	3	11.41	0.877	-21.17	3.89	0.462	-11.85
Trial-4	1	2	1	12.13	2.435	-21.85	4.92	0.082	-13.84
Trial-5	1	2	2	12.42	1.119	-21.92	5.00	0.118	-13.98
Trial-6	1	2	3	10.95	0.710	-20.81	3.99	0.367	-12.05
Trial-7	1	3	1	12.01	0.318	-21.60	5.33	0.590	-14.58
Trial-8	1	3	2	13.85	2.250	-22.94	4.62	0.850	-13.44
Trial-9	1	3	3	12.47	1.787	-22.00	5.08	0.445	-14.15
Trial-10	2	1	1	12.84	0.753	-22.19	5.15	0.292	-14.26
Trial-11	2	1	2	9.49	1.131	-19.61	4.53	0.537	-13.19
Trial-12	2	1	3	12.03	0.084	-21.61	4.32	0.230	-12.72
Trial-13	2	2	1	9.80	0.564	-19.84	4.10	0.706	-12.38
Trial-14	2	2	2	15.61	1.060	-23.89	5.48	0.297	-14.79
Trial-15	2	2	3	10.71	0.090	-20.59	4.79	0.324	-13.62
Trial-16	2	3	1	9.93	1.126	-20.00	4.94	0.671	-13.95
Trial-17	2	3	2	10.95	2.509	-21.01	4.58	0.227	-13.23
Trial-18	2	3	3	10.09	1.275	-20.14	4.09	0.223	-12.24
Trial-19	3	1	1	8.84	0.798	-18.96	4.23	0.598	-12.62
Trial-20	3	1	2	8.24	0.300	-18.32	4.27	0.544	-12.67
Trial-21	3	1	3	11.03	1.440	-20.93	5.18	0.654	-14.36
Trial-22	3	2	1	10.13	2.143	-20.30	4.77	1.335	-13.90
Trial-23	3	2	2	<b>8.22</b>	0.255	-18.30	<b>3.48</b>	0.212	-10.85
Trial-24	3	2	3	8.62	0.730	-18.74	4.08	1.059	-12.50
Trial-25	3	3	1	9.67	0.745	-19.73	4.57	0.713	-13.31
Trial-26	3	3	2	10.12	0.757	-20.13	4.75	0.514	-13.58
Trial-27	3	3	3	9.35	0.467	-19.42	4.37	0.431	-12.86

In terms of the roundness, Fig. 8 (c) shows the factorial level effects plotted by the S/N ratio. The level effects of the three factors were approximate. In principle, the use of a high-level spindle speed ( $A_3$ ) with a high-level feed rate ( $C_3$ ) could obtain a better performance in the hole's roundness. According to the statistics, Trial-23 was also the optimal parameter set with the smallest mean of 3.48  $\mu\text{m}$ .

## 5 Parameter models trained by stepped-learning procedure

## 5.1 The first-stage training

### 5.1.1 Network training experiments

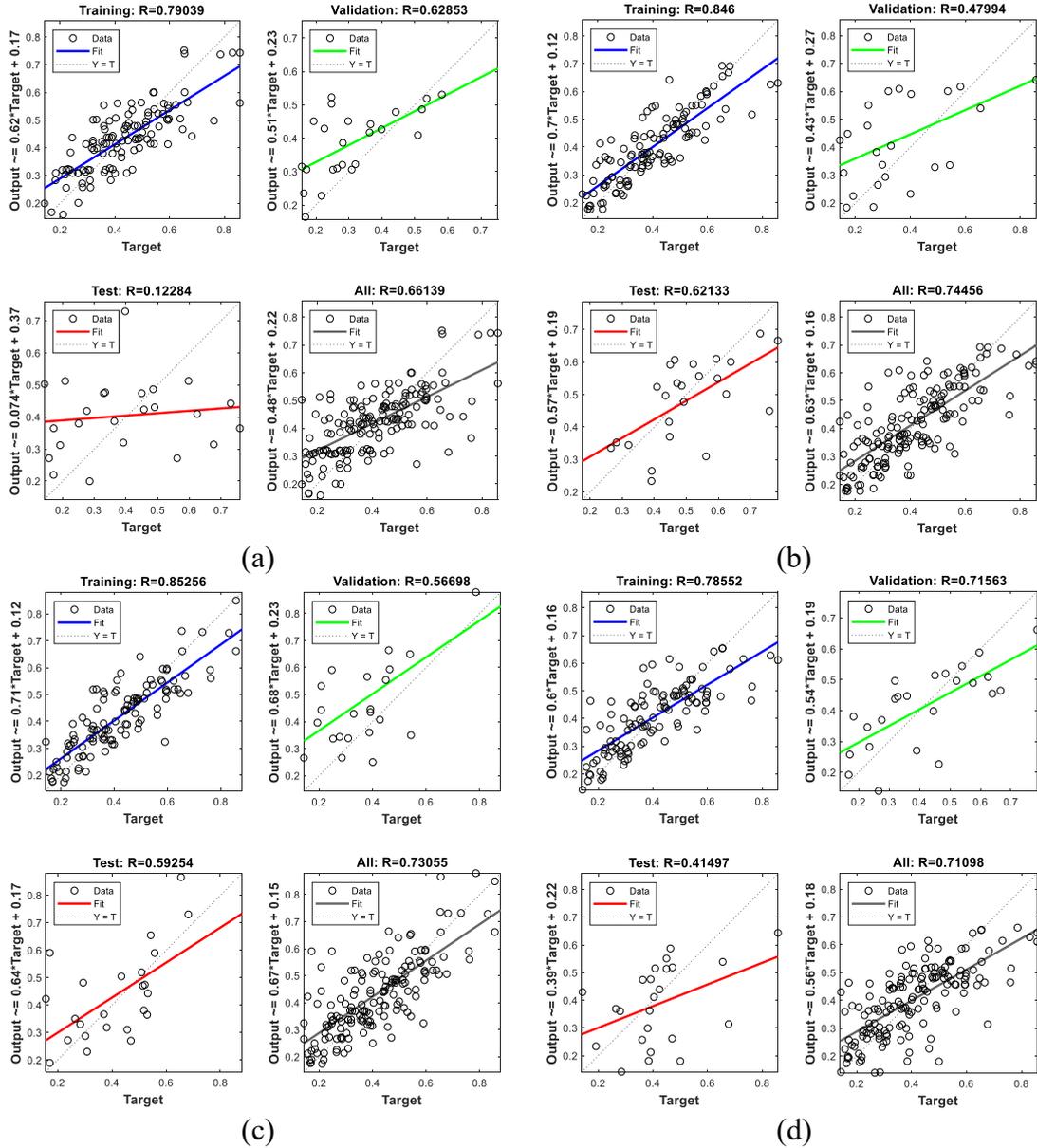
First, the experimental data were transformed into the input patterns required in the network training with normalization. Then the 81 input patterns were randomly divided into three groups of training, validation, and testing by the ratio of 57: 12: 12 as the MATLAB neural network tool box defaulted. In terms of the hidden-layer architecture, four types of  $5 \times 5$ ,  $10 \times 10$ ,  $15 \times 15$ , and  $20 \times 20$ , were arranged in the experiments.

Table 5 presents the training data in the first-stage experiment. The terminal moment was determined by the performance of the validation sets. The training of the four networks stopped at the 18<sup>th</sup>, 33<sup>rd</sup>, 11<sup>th</sup>, and 16<sup>th</sup> epochs, respectively, when their MSEs did not improve during 6 continuous epochs. From the MSE of the training sets, both models of  $10 \times 10$  and  $15 \times 15$  had better performance whose MSEs were below 0.0065.

Table 5 Training data in the first-stage training

Hidden layer architecture	Number of training cycle (epochs)	MSE of training sets	Correlation coefficient ( $R$ )			
			Training sets	Validation sets	Testing sets	All patterns
$5 \times 5$	18	0.00824	0.79039	0.62853	0.12284	0.66139
$10 \times 10$	33	<b>0.00629</b>	0.84600	0.47994	0.62133	0.74456
$15 \times 15$	11	<b>0.00644</b>	0.85256	0.56698	0.59254	0.73055
$20 \times 20$	16	0.00887	0.78552	0.71563	0.41497	0.71098

Model's performance in the output simulations was evaluated by the  $R$ , discussed by training sets, validation sets, testing sets, and all input patterns, respectively. The  $R$  value of the training sets evaluated by the network with hidden-layer neurons of  $5 \times 5$  achieved 0.79039, but that of the all patterns was 0.66139 only. The simulation precision was insufficient. While the hidden-layer neurons increased to  $10 \times 10$ ,  $15 \times 15$ , and  $20 \times 20$ , the  $R$  values of the all patterns were all above 0.7. Regression plots of the four NN models are shown in [Appendix A](#).



Appendix A. Models' performance presented by correlation coefficient in the first-stage training:

(a)  $4 \times 5 \times 5 \times 2$ , (b)  $4 \times 10 \times 10 \times 2$ , (c)  $4 \times 15 \times 15 \times 2$ , (d)  $4 \times 20 \times 20 \times 2$ .

### 5.1.2 Differential analyses for the additional training sets

When the network training is terminated, the simulated values of the output-layer variables can be derived through a simple NN's calculation. Because the NN model is trained by the training sets sampled from the drilling trials of the  $L_{27}$  OA experiment only, learning information of the model is limited. In order to enhance the simulation precision, a differential analysis of each training set is executed for calculating the deviation between the simulated output and its expected value. Moreover,

these data are used to compare the learning outcomes of the models trained by different hidden-layer architectures.

According to the factorial level effects of the thrust force and hole characteristics, refer to Fig. 8, the spindle speed (Factor *A*) and feed rate (Factor *C*) were significant factors. Therefore, the additional drilling trials were designed based on the parameters of spindle speed and feed rate only. Each factor has three levels, so there are 9 possible combinations. Based on the differential analysis results of the network training in the  $L_{27}$  OA experiment, the worse five parameter sets of Trial-22, Trial-4, Trial-2, Trial -24, and Trial-14 were selected as the datum sets for planning the additional drilling trials. Four trials were determined for each datum set, which designated with a spindle speed interval of  $\pm 500$  rpm and a feed rate interval of  $\pm 50$  mm/min. Accordingly, twenty drilling trials were planned as presented in Table 6. After actual drilling work, the thrust force and hole characteristics were measured as the variable data of the additional training sets for the second-stage training.

**Table 6** Variable data of the additional training sets for the second-stage training

No	Datum sets	Input variables				Output variables	
		Spindle speed (rpm)	Depth of peck drilling (mm)	Feed rate (mm/min)	Avg. thrust force peaks (N)	$\Delta D$ ( $\mu\text{m}$ )	Roundness ( $\mu\text{m}$ )
Add-1	Trial-22	4500	0.4	150	2.50	8.1	6.3
Add-2	Trial-22	4500	0.4	250	3.51	8.2	5.1
Add-3	Trial-22	5500	0.4	150	2.23	7.7	5.8
Add-4	Trial-22	5500	0.4	250	3.06	7.2	4.8
Add-5	Trial-2	2500	0.3	250	6.10	9.0	5.8
Add-6	Trial-2	2500	0.3	350	8.25	11.2	6.3
Add-7	Trial-2	3500	0.3	250	4.47	9.9	6.9
Add-8	Trial-2	3500	0.3	350	6.27	10.2	5.1
Add-9	Trial-24	4500	0.4	350	4.80	8.8	6.8
Add-10	Trial-24	4500	0.4	450	6.21	8.8	6.9
Add-11	Trial-24	5500	0.4	350	3.92	9.4	7.5
Add-12	Trial-24	5500	0.4	450	5.19	8.0	5.6
Add-13	Trial-4	2500	0.4	150	3.54	11.2	6.9
Add-14	Trial-4	2500	0.4	250	5.76	9.8	6.7
Add-15	Trial-4	3500	0.4	150	2.96	8.7	5.2
Add-16	Trial-4	3500	0.4	250	4.28	8.9	8.3
Add-17	Trial-14	3500	0.4	250	4.21	7.8	6.7
Add-18	Trial-14	3500	0.4	350	6.15	8.9	6.2
Add-19	Trial-14	4500	0.4	250	3.50	8.2	5.5
Add-20	Trial-14	4500	0.4	350	4.76	11.0	5.7

### 5.1.3 Trials for the model's testing sets

For a NN model, a higher  $R$  value indicates that the model's simulation of the training sets is good, but it cannot ensure the prediction reliability of the non-trained cases. To improve the accuracy in the model's prediction, in addition to the five datum sets decided for the additional drilling trials, two datum sets that learning badly in the first-stage training are more selected for planning the model's testing sets. In this case, Trial-20 and Trial-13 were chosen as the datum sets, and each set designated four drilling trials as well, as Table 7 presents. After actual drilling work, the characteristics measured are also listed in the table, that is, the variable data of the 8 model's testing sets. They only used for the model prediction test, not involved in the network training.

**Table 7** Variable data of the model's testing sets

No	Datum sets	Input variables				Output variables	
		Spindle speed (rpm)	Depth of peck-drilling (mm)	Feed rate (mm/min)	Avg. thrust force peaks (N)	$\Delta D$ ( $\mu\text{m}$ )	Roundness ( $\mu\text{m}$ )
Test-1	Trial-20	4500	0.3	250	3.60	8.9	4.8
Test-2	Trial-20	4500	0.3	350	4.92	8.0	5.2
Test-3	Trial-20	5500	0.3	250	3.19	10.0	5.7
Test-4	Trial-20	5500	0.3	350	4.24	9.4	5.1
Test-5	Trial-13	3500	0.4	150	2.81	9.3	5.5
Test-6	Trial-13	3500	0.4	250	4.26	9.3	4.9
Test-7	Trial-13	4500	0.4	150	2.53	8.4	3.6
Test-8	Trial-13	4500	0.4	250	3.59	8.9	4.6

Figure 9 shows the results of RMSE for the model's testing sets by the four models derived in the first-stage training. The predicted errors of the diameter shrinkage were larger than the roundness. The model trained by the hidden-layer of  $5 \times 5$  had a better performance, but only 57 sets of training data used here, the information for the network was insufficient. It still needed a network refining procedure to improve the model's prediction accuracy.

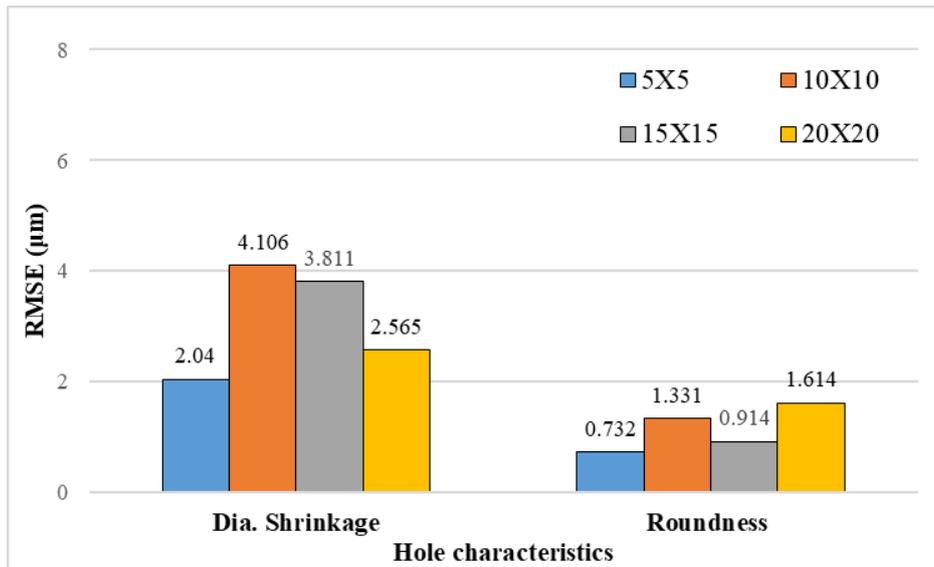


Fig. 9 Results of RMSE for the model’s testing sets evaluated by the first-stage models

## 5.2 The second-stage training

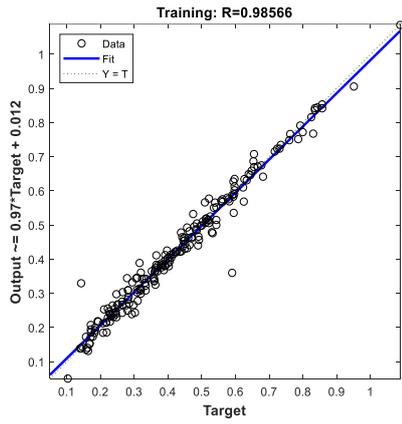
### 5.2.1 Network refining experiments

In the second-stage refining experiment, in addition to the 81 trials of the  $L_{27}$  OA experiment, 20 additional drilling trials were added. After the input pattern transformation, 101 training sets were prepared for NN refining experiments. In terms of the network architecture, because the R value of the model of hidden-layer  $5 \times 5$  evaluated by all patterns was only 0.66139 in the first-stage training (see Table 5), the simulation precision was unsatisfied due to the insufficiently connected weights among the neurons. Hence, only three hidden-layer architectures of  $10 \times 10$ ,  $15 \times 15$ , and  $20 \times 20$  were experimented here. The outcome of network learning can be estimated by the MSE of the network’s training sets. A larger MSE would result in poor precision, but a smaller MSE may incur the model to be overfitting. In this stage, three converging conditions for training termination were designated, included  $MSE < 0.001$ ,  $MSE < 0.005$ , and  $MSE < 0.010$ . As a result, 9 models were built after the network experiments. The training data and models’ performance are presented in Table 8.

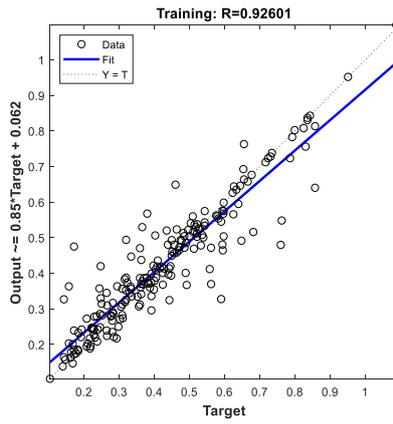
**Table 8** Training data in the second-stage training

Neurons of hidden layer	Converging condition (MSE)	Number of training cycles (epochs)	MSE of network's training sets	Correlation coefficient ( $R$ )
10×10	< 0.001	1066	0.000998	0.98566
	< 0.005	115	0.005	0.92601
	< 0.010	24	0.00968	0.85114
15×15	< 0.001	302	0.000988	0.98582
	< 0.005	55	0.00499	0.92612
	< 0.010	9	0.00975	0.84994
20×20	< 0.001	380	0.000991	0.98581
	< 0.005	74	0.00496	0.92650
	< 0.010	5	0.00955	0.85385

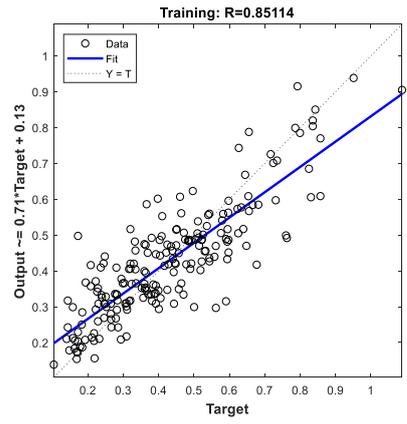
The  $R$  values of the three models trained by the condition of  $MSE < 0.001$  were all above 0.98. This indicated that the simulated outputs were very close to the expected value having a good performance in the training sets simulation. When the converging condition enlarged to  $MSE < 0.005$ , the  $R$  values were slightly decreased near 0.92 (0.92601, 0.92612, 0.92650). However, when the MSE more increased to  $MSE < 0.010$ , the terminal epochs of the three models were only 24, 9, and 5, respectively, but their  $R$  values were dropped below 0.85 (0.85114, 0.84994, 0.85385). The learning performance was fair. As to the training efficiency, in the experiments of the hidden-layer neurons of 15×15 and 20×20, the maximum training epochs to satisfy the converging conditions were about 300 which were lower than the network of 10×10 obviously (see [Table 8](#)). The more neurons used in the hidden layer can acquire higher training efficiency. Regression plots of the 101 training sets by the 9 NN models are shown in [Appendix B](#).



MSE < 0.001

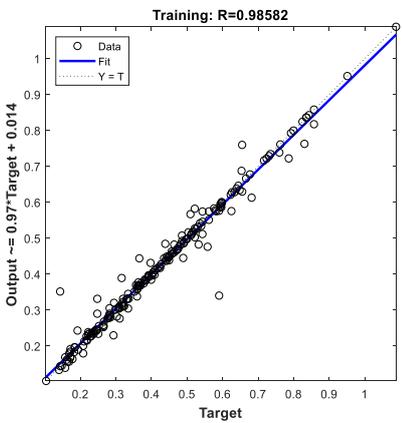


MSE < 0.005

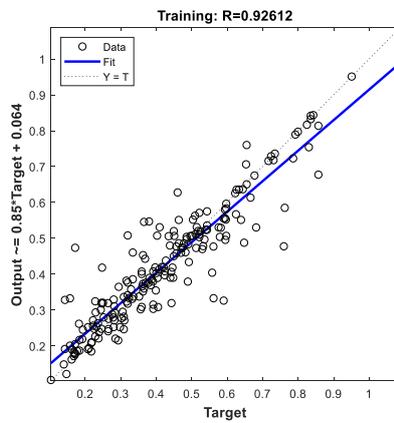


MSE < 0.010

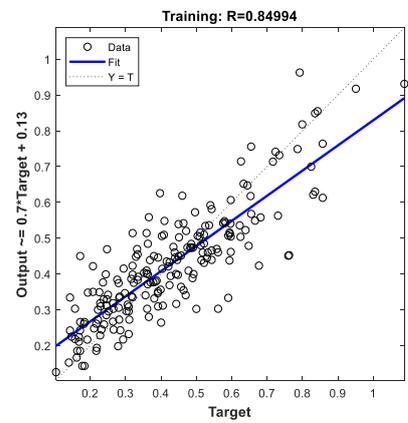
(a)



MSE < 0.001

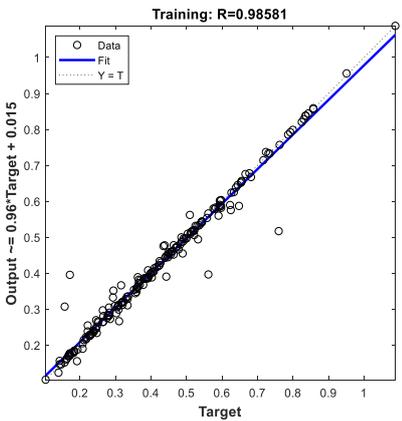


MSE < 0.005

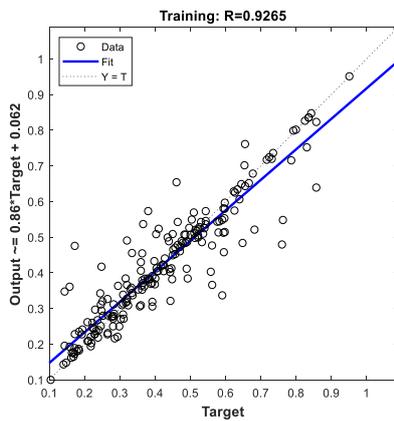


MSE < 0.010

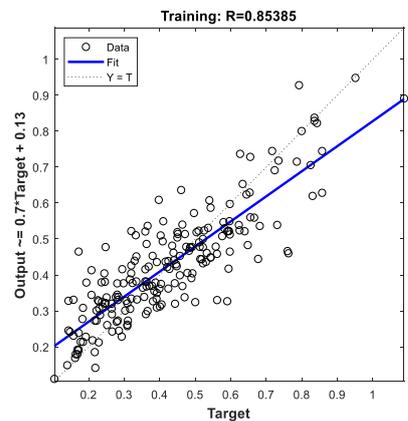
(b)



MSE < 0.001



MSE < 0.005



MSE < 0.010

(c)

Appendix B. Models' performance presented by correlation coefficient in the second-stage

training: (a)  $4 \times 10 \times 10 \times 2$ , (b)  $4 \times 15 \times 15 \times 2$ , (c)  $4 \times 20 \times 20 \times 2$ .

### 5.2.2 Analyses of the model's testing sets

After training, import the 8 model's testing sets into the 9 models obtained from the NN experiments to derive their predicted values of the output variables. The prediction accuracy of each model was evaluated by the RMSE as shown in Fig. 10. The models trained by the converging condition of  $MSE < 0.001$  had very high RMSEs. This indicated that these models couldn't derive an accurate prediction for the non-trained cases (the model's testing sets), and they were in an overfitting state. However, for the three models whose converging condition loosened to  $MSE < 0.010$ , their RMSEs were descended obviously. These models had a better performance in predictions, but their accuracy was still high, all above  $1\mu m$ . More training sets were required to provide field information for network learning.

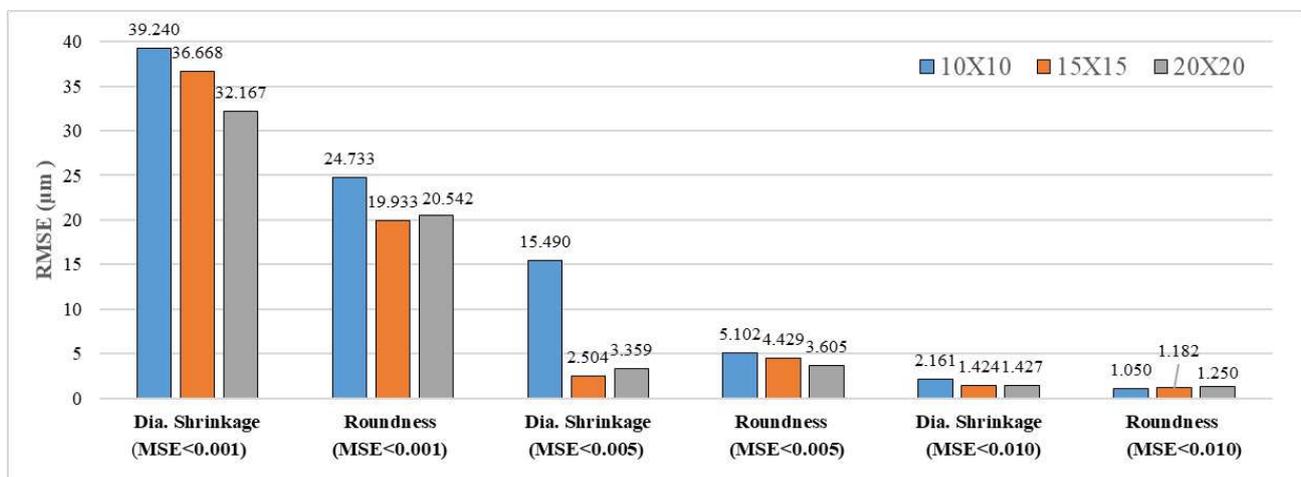


Fig. 10 RMSEs of the model's testing sets evaluated by the second-stage models

### 5.2.3 Differential analyses for the extra training sets

From the previous subsection, when the converging condition was specified at  $MSE < 0.010$ , the RMSEs were the least, and no overfitting occurred in the predictions of the non-trained cases. Therefore, the differential analyses were executed based on the simulated outputs derived by the models of  $MSE < 0.010$ . Table 9 presents the training sets which underperformed in the second-stage training. From the 101 training set, five training sets were decided as the datum sets for planning the extra drilling sets required for the next stage. They were Trial-18, Add-20, Add-17, Add-19, and Add-18. Since some datum sets had been experimented in the second-stage (coded in Add-), but their learning outcomes were still underperformed. More information required in these cases. Hence, here the intervals setting for the extra drilling trials were only half of that in the previous stage, that is, 250

rpm for the spindle speed and 25 mm/min for the feed rate. Similarly, four trials were designated for each datum set, and the extra drilling trials were presented in Table 10. After actual experiments, characteristic values of the thrust force, diameter shrinkage, and roundness were measured and prepared for the extra training sets used in the third-stage training.

**Table 9** Results of the differential analyses in the second-stage training

Neurons of hidden layer	Diameter shrinkage ( $\Delta D$ )					Roundness				
	Worst-1	Worst-2	Worst-3	Worst-4	Worst-5	Worst-1	Worst-2	Worst-3	Worst-4	Worst-5
10×10	18-3	Add-9	18-2	Add-18	Add-20	Add-16	Add-9	Add-17	Add-20	Add-19
15×15	18-3	Add-20	18-2	Add-9	Add-11	Add-7	Add-17	Add-16	Add-9	Add-20
20×20	Add-5	18-3	Add-9	18-2	Add-20	Add-16	Add-17	Add-9	Add-20	15-1

**Table 10** Variable data of the extra training sets for the third-stage training

No	Datum sets	Input variables				Output variables	
		Spindle speed (rpm)	Depth of peck drilling (mm)	Feed rate (mm/min)	Avg. thrust force peaks (N)	$\Delta D$ ( $\mu\text{m}$ )	Roundness ( $\mu\text{m}$ )
Extra-1	Trial-18	3500	0.5	350	6.06	13.5	4.6
Extra-2	Trial-18	3500	0.5	450	7.81	8.2	5.1
Extra-3	Trial-18	4500	0.5	350	4.79	6.5	4.6
Extra-4	Trial-18	4500	0.5	450	6.35	7.7	4.4
Extra-5	Add-20	4250	0.4	325	4.72	7.2	4.1
Extra-6	Add-20	4250	0.4	375	5.54	8.9	3.8
Extra-7	Add-20	4750	0.4	325	4.30	8.4	4.4
Extra-8	Add-20	4750	0.4	375	5.01	8.7	4.1
Extra-9	Add-17	3250	0.4	225	4.11	8.4	4.6
Extra-10	Add-17	3250	0.4	275	4.93	6.4	3.0
Extra-11	Add-17	3750	0.4	225	3.64	6.7	3.4
Extra-12	Add-17	3750	0.4	275	4.23	7.4	4.2
Extra-13	Add-19	4250	0.4	225	3.33	7.2	5.0
Extra-14	Add-19	4250	0.4	275	3.82	13.7	5.2
Extra-15	Add-19	4750	0.4	225	3.21	6.9	4.4
Extra-16	Add-19	4750	0.4	275	3.60	6.9	4.1
Extra-17	Add-18	3250	0.4	325	5.98	8.7	3.9
Extra-18	Add-18	3250	0.4	375	6.92	7.4	3.8
Extra-19	Add-18	3750	0.4	325	5.24	9.4	4.4
Extra-20	Add-18	3750	0.4	375	5.80	14.1	5.8

## 5.3 The third-stage training

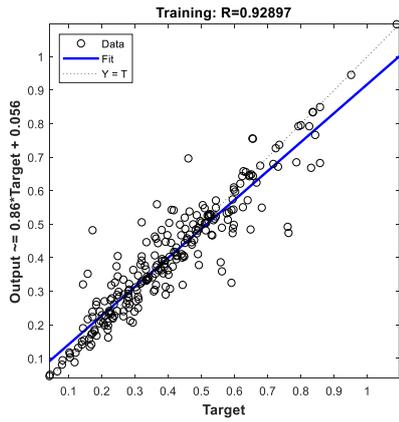
### 5.3.1 Network refining experiments

In the third-stage training, there were 121 training sets including 81 sets obtained from Taguchi's  $L_{27}$  OA experiment, 20 sets from the additional drilling trials, and 20 sets from the extra drilling trials. Because the models trained by the converging condition of the  $MSE < 0.001$  had obvious overfitting in the second-stage training, only two conditions of  $MSE < 0.005$  and  $MSE < 0.010$  were remained here. In addition, a condition of  $MSE < 0.0125$  was added for observing the effect of enlarging the MSE in training. In terms of network architecture, the setting was the same as the second-stage.

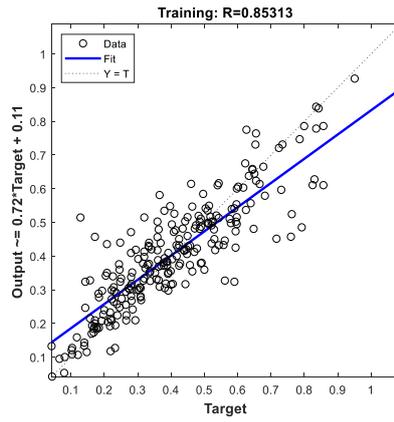
Table 11 presents the training data of nine models in the third-stage refining experiments. From the  $R$ , the converging condition of  $MSE < 0.005$  presented the best learning results for the training sets as well. The  $R$  values were all above 0.92 for the three architectures. However, when the converging condition specified to  $MSE < 0.0125$ , only 20, 11, and 8 epochs required to terminated the training procedure, respectively. But the  $R$  values were dropped near 0.81 (0.81293、0.81700、0.81269). Compared with the second-stage training, the number of training sets was increased from 101 to 121, the field information of the network model was more integrated. But, the training cycles needed for the models of  $15 \times 15$  and  $20 \times 20$  were around 40 epochs to meet the converging condition. The training efficiency was improved. Regression plots of the 121 training sets by the 9 NN models are shown in Appendix C.

Table 11 Training data in the third-stage training

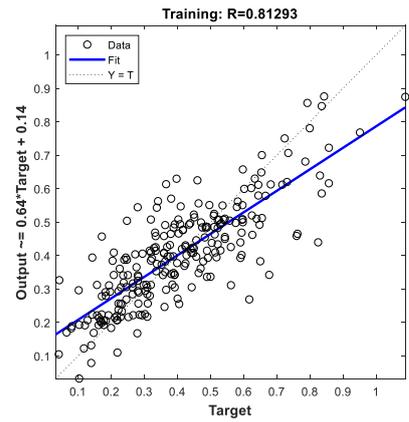
Neurons of hidden layer	Converging condition (MSE)	Number of training cycles (epochs)	MSE of network's training sets	Correlation coefficient ( $R$ )
10×10	< 0.005	141	0.005	0.92897
	< 0.010	28	0.00993	0.85313
	< 0.0125	20	0.0124	0.81293
15×15	< 0.005	41	0.005	0.92906
	< 0.010	26	0.00991	0.85372
	< 0.0125	11	0.0122	0.81700
20×20	< 0.005	39	0.005	0.92906
	< 0.010	9	0.00999	0.85299
	< 0.0125	8	0.0125	0.81700



MSE < 0.005

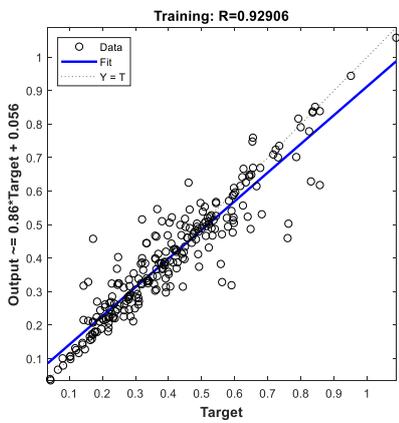


MSE < 0.010

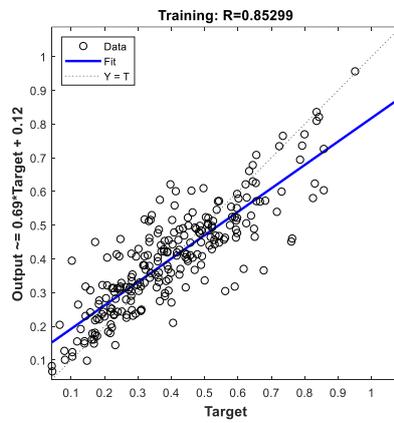


MSE < 0.0125

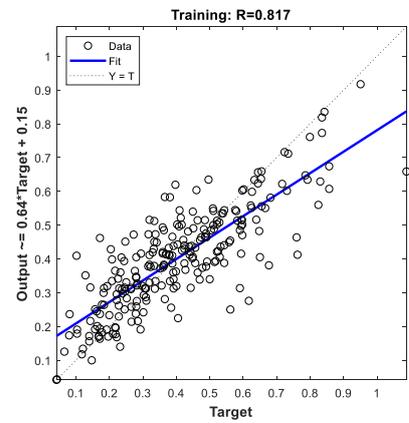
(a)



MSE < 0.005

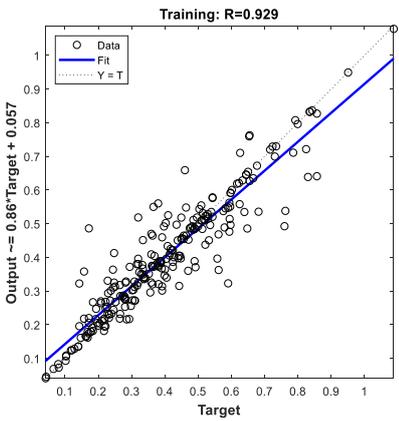


MSE < 0.010

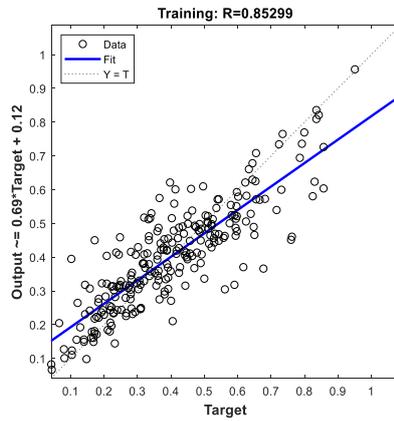


MSE < 0.0125

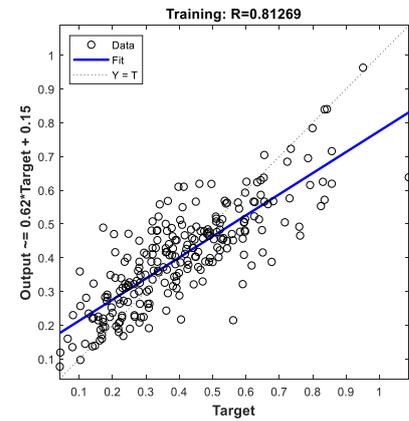
(b)



MSE < 0.005



MSE < 0.010



MSE < 0.0125

(c)

Appendix C. Models' performance presented by correlation coefficient in the third-stage training:

(a)  $4 \times 10 \times 10 \times 2$ , (b)  $4 \times 15 \times 15 \times 2$ , (c)  $4 \times 20 \times 20 \times 2$ .

### 5.3.2 Analyses of the model's testing sets

After the network refining experiment, the predicted outputs of the model's testing sets can be derived by the 9 NN models. Figure 11 shows the RMSEs of the model's testing sets. Although the prediction accuracy of the models trained by the  $MSE < 0.005$  was still large, the RMSE of the others, the six models derived by  $MSE < 0.010$  and  $MSE < 0.0125$ , were all below  $2 \mu m$  and no overfitting occurred. The optimal model for both characteristics was trained by the hidden-layer architecture of  $20 \times 20$  with the converging condition of  $MSE < 0.0125$ . The RMSEs for the diameter shrinkage and roundness were  $0.821$  and  $0.635 \mu m$ , respectively.

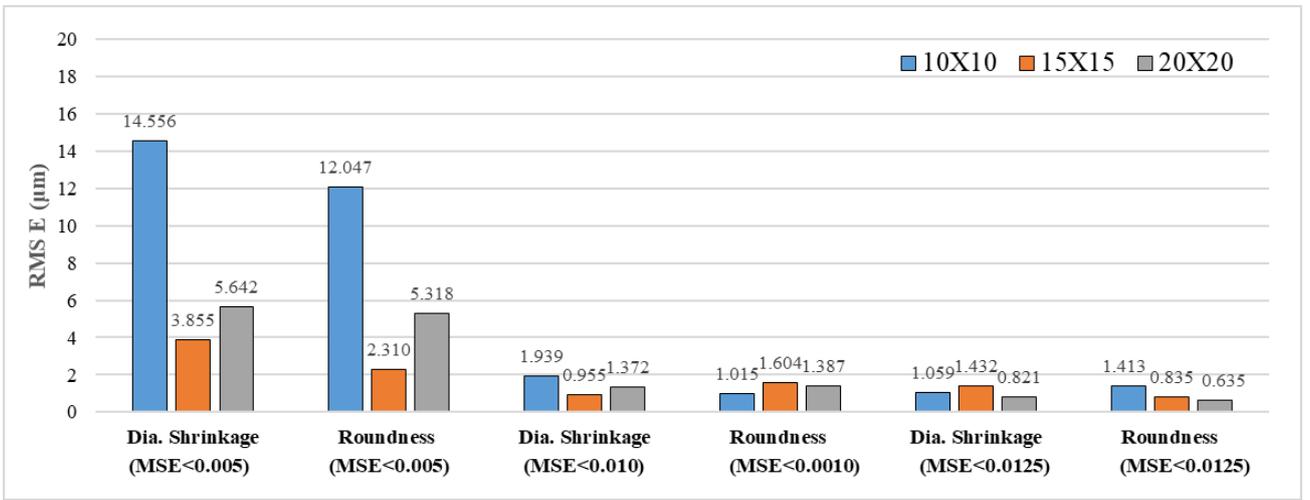


Fig. 11 RMSEs of the model testing sets evaluated by the third-stage models

## 5.4 Discussion

In the network training, the model's precision can be estimated by the  $R$  based on the simulated outputs of the training sets. However, the prediction accuracy for the non-trained sets (model's testing sets) may be estimated by the average RMSE of the output variables examined. Two hole's characteristics discussed in this study, thus we have

$$RMSE_{avg} = \frac{RMSE_{DS} + RMSE_R}{2}, \quad (9)$$

where  $RMSE_{DS}$  is the RMSE of the hole shrinkage in diameter and  $RMSE_R$  is that of the roundness.

Table 12 summarizes the models' performance during the three training stages. Several points are found:

- (1) In the first-stage training, only 57 training sets were used. The  $R$  values of the models trained by different hidden-layer neurons were all below 0.75. As to the second-stage training, the number of training sets increased to 101. The  $R$  values of the three models trained by the converging condition of  $MSE < 0.001$  were high to 0.98. This indicates that the model's simulation precision of the training sets was excellent; nevertheless, their predicted errors ( $RMSE_{avg}$ ) were so large for the non-trained cases (model's testing sets). These models were in an overfitting state.
- (2) In the second- and third-stage training, for the models terminated at  $MSE < 0.005$ , their  $R$  values were near 0.92. Although their simulation precisions were slightly decreased to the models trained by  $MSE < 0.001$ , their predicted errors of the model's testing sets ( $RMSE_{avg}$ ) were improved. When the converging condition enlarged to  $MSE < 0.010$ , the  $R$  values were dropped again, near 0.85; however, the average predicted errors were more improved to 1.2794 and 1.3797  $\mu m$ , respectively. As to the three models trained by the condition of  $MSE < 0.0125$ , their  $R$  values were near 0.81 in the third-stage training.
- (3) In the case analyzed, the optimal model for output variables prediction was trained by the hidden-layer neurons of  $20 \times 20$  with the converging condition of  $MSE < 0.0125$ . The average RMSE of the model's testing sets (non-trained cases) was 0.7281  $\mu m$ .

**Table 12** Summary of the models' performance during the three training stages

Training stage		Training (MSE)	$4 \times 10 \times 10 \times 2$		$4 \times 15 \times 15 \times 2$		$4 \times 20 \times 20 \times 2$	
			$R$	$RMSE_{avg}$	$R$	$RMSE_{avg}$	$R$	$RMSE_{avg}$
<b>1<sup>st</sup></b>	<b>57</b> Training sets	System defaulted	<b>0.74456</b>	2.7186	<b>0.73055</b>	2.3627	<b>0.71098</b>	2.0892
	<b>101</b> Training sets	0.001	<b>0.98566</b>	31.9867	<b>0.98582</b>	28.3004	<b>0.98581</b>	26.3544
<b>2<sup>nd</sup></b>	Training sets	0.005	<b>0.92601</b>	10.2962	<b>0.92612</b>	3.4662	<b>0.92650</b>	3.4821
		0.010	0.85114	1.6055	0.84994	1.3031	0.85385	1.3384
<b>3<sup>rd</sup></b>	Training sets	0.005	<b>0.92897</b>	13.3013	<b>0.92906</b>	<b>3.0822</b>	<b>0.92900</b>	<b>5.4802</b>
		0.010	0.85313	1.4767	0.85372	<b>1.2794</b>	0.85299	<b>1.3797</b>
		0.0125	<b>0.81293</b>	1.2362	<b>0.81700</b>	1.1337	<b>0.81269</b>	<b>0.7281</b>

## 6. Conclusions

Hole shrinkage always produces in the drilling of engineering plastics, as a result, the dimensions and profile of the hole drilled are hard to control. This study conducted the Taguchi  $L_{27}$  OA experiment of the PEEK—MDS 100 to derive the optimal parameter conditions for lowering the thrust force and minimizing the hole's shrinkage and roundness in drilling a 1-mm-diameter hole. In addition, a stepped-learning procedure is presented for developing a Taguchi-based NN model that can save the experimental data, simulate the hole's characteristics of the network's training sets, and predict the non-trained cases. For the case analyzed, the field information of the network model was expanded by three stages of training, and the model's prediction accuracy was improved gradually. The ANN models for a small-hole drilling process of the PEEK—MDS 100 were successfully built, which achieved a satisfying prediction accuracy, below 1  $\mu\text{m}$ .

**Authors' contributions** DYC and CHL conceived this research and designed experiments; XYW performed experiments and analysis; DYC and CHL wrote the paper and participated in the revisions of it. All authors discussed the results and contributed to the final manuscript.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

**Availability of data and material** Data can be made available upon request subject to the approval of all parties involved in the research.

## **Declarations**

**Ethical approval** No ethical approval was required for this research.

**Consent to participate** Not applicable.

**Consent for publication** All authors have read and agreed to the published version of the manuscript.

**Code availability** Due to ongoing studies, the code associated with the publication is currently not available.

**Conflict of interest** The authors declare that they have no potential conflict of interest in relation to the study in this paper.

## References

- [1] Esim E, Yildirim Ş (2017) Drilling performance analysis of drill column machine using proposed neural networks. *Neural Comput & Applic* 28:79-90. <https://doi.org/10.1007/s00521-016-2322-8>
- [2] Roth JT, Djurdjanovic D, Yang X, Mears L, Kurfess T (2010) Quality and Inspection of Machining Operations: Tool Condition Monitoring. *ASME J Manuf Sci Eng* 132(4):041015-410516. <https://doi.org/10.1115/1.4002022>
- [3] Choi YJ, Park MS, Chu CN (2008) Prediction of drill failure using features extraction in time and frequency domains of feed motor current. *Int J Mach Tools Manuf* 48(1):29-39. <https://doi.org/10.1016/j.ijmachtools.2007.08.009>
- [4] Zerti A, Yallese MA, Meddour I, Belhadi S, Haddad A, Mabrouki T (2019) Modeling and multi-objective optimization for minimizing surface roughness, cutting force, and power, and maximizing productivity for tempered stainless steel AISI 420 in turning operations. *Int J Adv Manuf Technol* 102:135–157. <https://doi.org/10.1007/s00170-018-2984-8>
- [5] Daniel SAA, Pugazhenthir R, Kumar R, Vijayananth S (2019) Multi objective prediction and optimization of control parameters in the milling of aluminium hybrid metal matrix composites using ANN and Taguchi -grey relational analysis. *Def Technol* 15(4):545-556. <https://doi.org/10.1016/j.dt.2019.01.001>
- [6] Yadav RN, Yadava V, Singh GK (2013) Multi-objective optimization of process parameters in Electro-Discharge Diamond Face Grinding based on ANN-NSGA-II hybrid technique. *Front Mech Eng* 8:319-332. <https://doi.org/10.1007/s11465-013-0269-3>

- [7] Biswas R, Kuar AS, Biswas SK, Mitra S (2009) Artificial neural network modeling of Nd:YAG laser microdrilling on titanium nitride alumina composite. *Proc Inst Mech Eng B J Eng Manuf* 224(3):473- 482. <https://doi.org/10.1243/09544054JEM1576>
- [8] Sarıkaya M, Yılmaz V (2018) Optimization and predictive modeling using S/N, RSM, RA and ANNs for micro-electrical discharge drilling of AISI 304 stainless steel. *Neural Comput & Applic* 30:1503-1517. <https://doi.org/10.1007/s00521-016-2775-9>
- [9] Kumar SPL, Jerald J, Kumanan S, Prabakaran R A (2014) Review on Current Research Aspects in Tool-Based Micromachining Processes. *Mater Manuf Process* 29(11-12):1291-1337. <https://doi.org/10.1080/10426914.2014.952037>
- [10] Hyacinth SX, Natarajan U, Ramasubbu N (2015) A review of accuracy enhancement in microdrilling operations. *Int J Adv Manuf Technol* 81:199-217. <https://doi.org/10.1007/s00170-015-6900-1>
- [11] Patra K, Jha AK, Szalay T, Ranjan J, Monostori L (2017) Artificial neural network based tool condition monitoring in micro mechanical peck drilling using thrust force signals. *Precis Eng* 48:279-291. <https://doi.org/10.1016/j.precisioneng.2016.12.011>
- [12] Corne R, Nath C, Mansori MEI, Kurfess T (2017) Study of spindle power data with neural network for predicting real-time tool wear/breakage during inconel drilling. *J Manuf Syst* 43(2):287-295. <https://doi.org/10.1016/j.jmsy.2017.01.004>
- [13] Murthy BRN, Vijay GS, Narayan S, Naik N, Sooriyaperakasam N, Karthik A, Borkhade R (2019) Mechanical modelling and simulation of thrust force in drilling process in GFRP composite laminates: A novel system dynamics approach. *Cogent Eng* 6(1):1706981. <https://doi.org/10.1080/23311916.2019.1706981>
- [14] Efkolidis N, Hernández CG, Talón, JH, Kyratsis P (2018) Modelling and Prediction of Thrust Force and Torque in Drilling Operations of Al7075 Using ANN and RSM Methodologies. *J Mech Eng Sci* 64(6):351-361. <https://doi.org/10.5545/sv-jme.2017.5188>

- [15] Ulas HB, Ozkan MT, Malkoc Y (2019) Vibration prediction in drilling processes with HSS and carbide drill bit by means of artificial neural networks. *Neural Comput & Applic* 31:5547-5562. <https://doi.org/10.1007/s00521-018-3379-3>
- [16] Caggiano A, Angelone R, Napolitano F, Nele L, Teti R (2018) Dimensionality Reduction of Sensorial Features by Principal Component Analysis for ANN Machine Learning in Tool Condition Monitoring of CFRP Drilling. *Procedia CIRP* 78:307-312. <https://doi.org/10.1016/j.procir.2018.09.072>
- [17] Vrabel M, Maňková I, Beňo J (2016) Monitoring and Control of Manufacturing Process to Assist the Surface Workpiece Quality When Drilling. *Procedia CIRP* 41:735-739. <https://doi.org/10.1016/j.procir.2015.12.092>
- [18] Mondal N, Mandal S, Mandal MC (2020) FPA based optimization of drilling burr using regression analysis and ANN model. *Measurement* 152:107327. <https://doi.org/10.1016/j.measurement.2019.107327>
- [19] Ahn Y, Lee SH (2017) Classification and prediction of burr formation in micro drilling of ductile metals. *Int J Prod Res* 55(17):4833-4846. <https://doi.org/10.1080/00207543.2016.1254355>
- [20] Akıncioğlu S, Mendi F, Çiçek A, Akıncioğlu G (2013) ANN-based prediction of surface and hole quality in drilling of AISI D2 cold work tool steel. *Int J Adv Manuf Technol* 68:197-207. <https://doi.org/10.1007/s00170-012-4719-6>
- [21] Cruz CED, Aguiar PR, Machado ÁR, Bianchi EC, Contrucci JG, Neto FC (2013) Monitoring in precision metal drilling process using multi-sensors and neural network. *Int J Adv Manuf Technol* 66:151–158. <https://doi.org/10.1007/s00170-012-4314-x>
- [22] Fowlkes WY, Creveling CM (1984) *Engineering Methods for Robust Product Design: Using Taguchi Methods in Technology and Product Development*. Addison-Wesley, Massachusetts, chapter 4 and chapter 5
- [23] Freeman JA, Skapura DM (1991) *Neural Networks: Algorithms, Application, and Programming*

Techniques. Addison-Wesley, Massachusetts, chapter 3