Hyperparameter Optimization of Deep Learning Networks for Classification of Breast Histopathology Images

Cheng-Jian Lin,[1,2*] Shiou-Yun Jeng,[1] and Chin-Ling Lee[3]

[1]Department of Computer Science & Information Engineering, National Chin-Yi University of Technology,
No.57, Sec.2, Zhongshan Rd., Taiping Dist., Taichung 411, Taiwan
[2]College of Intelligence, National Taichung University of Science and Technology,
No.129, Sec.3, Sanmin Rd, North Dist., Taichung 404, Taiwan
[3]Department of International Business, National Taichung University of Science and Technology,
No.129, Sec.3, Sanmin Rd, North Dist., Taichung 404, Taiwan

After tumor detection in their breasts, women typically fear mastectomy; this affects curative care outcomes. Most tumors are benign. After resection and pathological examination, because of advances in medicine and treatment, the success rate of early breast cancer treatment can reach 60 to 90%. An accurate assessment of tumor extent is essential. In this study, a novel method of hyperparameter optimization of deep learning networks was proposed to classify tumors as malignant or benign. When setting hyperparameters in deep learning networks, most users use trial and error to determine them. In our experiments, the Taguchi method was used to select the impact factors. The orthogonal table design was used to conduct experiments. Then, the best combination of parameters was determined and significant impact factors were analyzed. The Breast Cancer Histopathological Database was used for analysis. This database was built in collaboration with the P&D Laboratory and contains 2480 benign and 5429 malignant samples. The experimental results showed that the proposed method obtained a high accuracy of 83.19%.

1. Introduction

Breast cancer is the principal cause of death from cancers in women worldwide.[1] Early diagnosis can reduce breast cancer death rates and consequently improve the success rate of early breast cancer treatment to 90%.[2] Although histological image classification is critical in breast cancer diagnoses, the detection and categorization of abnormalities are difficult even for experienced pathologists.[3,4] Recently, machine learning techniques have been successfully used for handling classification in the medical field. Therefore, automatic detection and diagnosis of tumors through histological image processing and machine learning algorithms can increase the accuracy of breast cancer diagnosis.[5] Automated machine-learning-based tools will assist physicians in image analysis.[6] The diagnosis of medical images of breast cancer obtained by ultrasound by machine learning showed significantly better performance than that

of models trained on clinical features.[7] The complex processes of traditional machine learning algorithms, such as tedious preprocessing, segmentation, and manual feature extraction steps reduce their efficiency and accuracy.[8] To overcome these drawbacks of traditional machine learning techniques, deep learning networks have been used for efficient classification.[9] The present study focused on the application of deep learning networks for the classification of breast histopathology images.

Deep learning networks have multiple levels of abstraction, representation, and information, and have been successfully applied in various domains.[10–12] For example, convolutional neural networks (CNNs) based on the deep learning network architecture have been reported to be powerful tools for the automated classification of human cancer histopathology images.[13] Khosravi *et al.* designed several computational methods based on CNNs to effectively classify various histopathology images across different types.[14] Sabeena *et al.* used modified CNN-based deep transfer learning for automatic mitosis detection in breast histopathology images.[15] In addition, a successful application of the CNN is the LeNet-5 system, which consists of two convolutional layers, two pooling layers, and a fully connected layer.[16] The following methods are used for the defect classification of benign tumor and malignant tumor features: single-layer CNN,[17] rotation forest classifier and a parameter-free version of threshold adjacency statistics,[18] LeNet-5 stochastic gradient descent with momentum,[19] LeNet-5 adaptive moment estimation,[20] and LeNet-5 root mean square propagation.[21] Although traditional methods are capable of segmenting and detecting breast tumors using the Breast Cancer Histopathological Database (BreakHis), classification accuracy is limited.[22] To improve the accuracy of mitosis detection in breast histopathology images, we propose a hyperparameter-optimized CNN approach based on deep learning networks.

However, the training of a CNN based on deep learning networks typically requires a large amount of data to optimize the large number of network parameters, such as the number of filters, convolution kernel size, and number of output feature maps.[23] Therefore, hyperparameter optimization in deep learning models is effective for prototype selection and classifier selection, and it has been widely used for the optimization of general machine learning models.[24] Although several methods have been proposed for hyperparameter optimization, the Taguchi method is a unique experimental statistical design, which provides a reliable design solution, improves quality, and reduces costs because it allows optimization with the minimum number of experiments.[25] The Taguchi method is a statistical approach for optimizing process parameters by using an orthogonal array consisting of factors and levels to classify the results.[26] Debnath *et al.* applied the Taguchi method for network optimization in deep learning networks. They achieved superior classification using hyperparameter optimization to speed up training.[27] In this study, to develop deep learning networks, we presented a simple and efficient framework for the classification of breast histopathology images by using hyperparameter optimization with the Taguchi method.

The objective of this study was to use hyperparameter optimization of deep learning networks for the classification of breast histopathology images. The major contributions of the study are as follows: (1) obtaining the best combination, such as the convolution kernel size, number of filters, stride, and padding of hyperparameters by the Taguchi method; (2) adjusting

the hyperparameters in the network architecture to reduce the number of experiments compared with the traditional manual adjustment of hyperparameters; and (3) by analyzing the importance of each impact factor, the hyperparameters in CNN can be effectively optimized.

This study is organized as follows. Section 2 presents hyperparameter optimization methods for deep learning networks. Section 3 discusses the experimental results and data analysis. The final section presents conclusions and future studies.

2.   Materials and Methods

The purpose of this study was to optimize the hyperparameters of a CNN. On the basis of deep learning networks, we obtained the best hyperparameter combinations by performing experiments with the Taguchi method. In these experiments, BreakHis was used to improve the network performance. First, the experimental factors were set for many hyperparameters and the variation level of each factor was determined. The orthogonal arrays required for the experiment were designed. The results can be used to calculate the signal-to-noise (S/N) ratio, and the contribution of each factor can be analyzed through the variance to obtain the best hyperparameter combination for the network.

2.1   Deep learning networks

This section introduces the architecture and hyperparameters of the deep learning networks, which contain selectable factors. We performed experiments based on the LeNet network structure. Figure 1 depicts the proposed deep learning network architecture, including an input, three-layer convolutional layer, two-layer pooling layer, fully connected layer, and final classification.

(a) Input image size: The input image size in this study was 50×50 instead of 32×32 (original LeNet). To select hyperparameters of convolutional layers as experimental factors, more levels can be added so that the final output feature image is not less than 0.

(b) Convolutional layers: The proposed network architecture contained three convolutional layers. The selected hyperparameters included convolution kernel size, number of filters, stride, and padding. The convolution kernel size affects the fineness of feature extraction, and the number of filters is related to the number of feature maps. Too many filters may
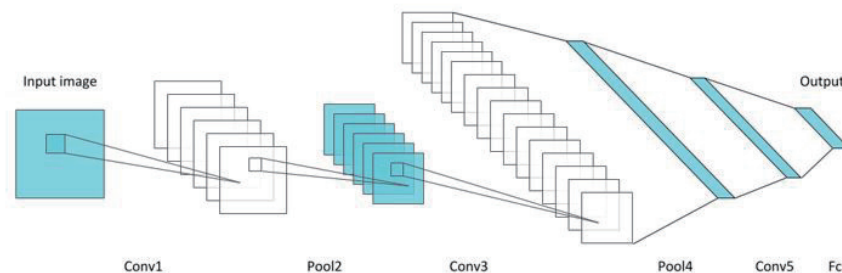

Fig. 1.   (Color online) Proposed deep learning network architecture.

affect the training time. The stride is the step of the convolution kernel, which increases each time the feature is captured. The increase makes the extracted feature map smaller, and the calculation complexity also decreases. The padding was set to 0 for each input image periphery, which is used to increase the size of the feature map. The remaining layer changes the convolution kernel to 1 × 1 to reduce the dimension. After testing, the kernel does not significantly affect the network performance, and it is connected to the last fully connected layer.

(c) Pooling layers: By referring to the LeNet network architecture, two pooling layers were inserted between the three convolutional layers. The hyperparameters of the pooling layer were not adjusted, and its size of 2 × 2 was maintained.

(d) Activation function: After each convolutional layer, the ReLU activation function was added, which can ensure efficient gradient descent.

2.2   Hyperparameter optimization using the Taguchi method

The Taguchi method is an engineering method that was introduced by Dr. Kazuhiro Taguchi. In this method, statistics are used to perform experiments and data analysis. The Taguchi method is used to design orthogonal arrays to reduce the number of experiments and observe the degree of variation among experiments. Variance analysis is used to determine the importance of factors, optimize quality, and reduce costs. Figure 2 depicts the experimental procedure of the Taguchi method.
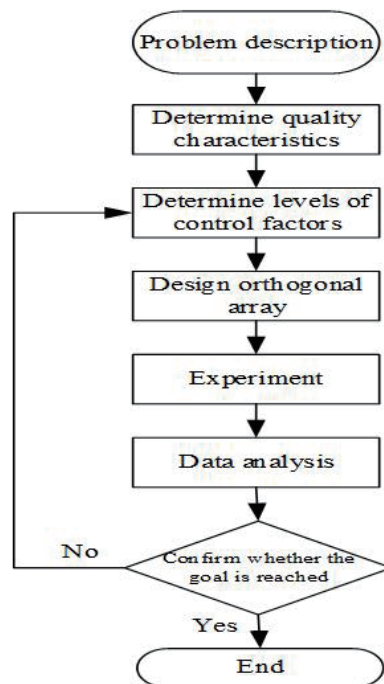

Fig. 2.   Flowchart of the Taguchi method experiment.

(a) Quality characteristics and factor selection

The orthogonal arrays in the standard Taguchi method are predefined according to the number of control factors and levels. Therefore, the quality characteristics are evaluated through experiments performed according to the defined orthogonal arrays. In this experiment, we defined the quality characteristics as the performance of the hyperparameters of the CNN based on deep learning networks and we calculated the S/N ratio for each experiment from multiple observations. The S/N ratio of the experimental results indicates the robustness of the network performance. In this experiment, the S/N ratio for the case of "larger is better" was evaluated as follows:

$$S/N = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right)db,\tag{1}$$

where n is the number of observations, y_i is the efficiency of the ith observation, and the unit of the S/N ratio is the *db* value.

As presented in Table 1, Levels 2 and 3 were the levels of the experiment, A–D were the hyperparameters of the first convolution layer, and E–H were the hyperparameters of the second convolution layer. The original hyperparameters in LeNet were the 5 × 5 convolution kernel, which was amplified to 7 × 7 or reduced to 3 × 3 in the table because we wanted to determine if the recognition rates increase for large- or small-feature extraction. The stride and padding were increased to 2 and 1, respectively, to avoid an output image smaller than zero. After the first convolution layer, the output feature image was reduced, and some features were lost. Therefore, we increased the number of filters in the second convolutional layer and retained more features to increase the classification performance.

(b) Design of the orthogonal array

After the control factors and levels were determined, the orthogonal array was designed according to the results of Table 1. First, we calculated the minimum number of experiments and selected the orthogonal array applicable to the experiment by comparing the total number of columns in the orthogonal array (see Table 2) and determined the number of experiments. The minimum number of experiments is the total degrees of freedom (DoF) plus one, as described in Eq. (2) and shown in Table 3. After the orthogonal array required for the experiment is selected, the orthogonal array is set according to the selected factors and levels.

Table 1
Levels of control factors.

No.	Factors	1	2	3
A	Conv1_Kernel size	3	5	7
B	Conv1_Filter	4	6	12
C	Conv1_Stride	1	2	
D	Conv1_padding	0	1	
E	Conv2_Kernel size	3	5	7
F	Conv2_Filter	8	16	32
G	Conv2_Stride	1	2	
H	Conv2_padding	0	1	

Table 2
Calculation of DoF of each factor.

Orthogonal array	Experiments	Columns	Maximum number of rows at these levels			
			2	3	4	5
L_4	4	3	3			
L_8	8	7	7			
L_9	9	4		4		
L_{12}	12	11	11			
L_{16}	16	15	15			
L'_{16}	16	5				5
L_{18}	18	8	1	7		
L_{25}	25	6				3
L_{27}	27	13		13		
L_{32}	32	31	31			
L'_{32}	32	10	1		9	
L_{36}	36	23	11	12		
L'_{36}	36	16	3	13		
L_{50}	50	12	1			11
L_{54}	54	26	1	25		
L_{64}	64	63	63			
L'_{64}	64	21			21	
L_{81}	81	40			40	

Table 3
Calculation of DoF of each factor.

No.	Number of Levels	DoF
A	3	2
B	3	2
C	2	1
D	2	1
E	3	2
F	3	2
G	2	1
H	2	1
Total		12

$$\text{Minimum number of experiments} = \text{Total DoF} + 1 \qquad (2)$$

3. Experimental Results

The first subsection introduces the BreakHis data set, which is used in the CNN based on deep learning networks to classify benign and malignant tumors and determine the optimal hyperparameter combination using the Taguchi method. In the second section, the experimental results obtained using the orthogonal array and the factor significance analysis are presented.

3.1 Data set

BreakHis was established by the Federal University of Parana (UFPR) and Pathological Anatomy and Cytopathology (P&D) Laboratory.[28] The data set is composed of 7909 histopathological images with different magnifications of breast tumors taken from 82 patients, as presented in Table 4 and Fig. 3. The samples were collected through surgical open biopsy, namely, partial mastectomy or excisional biopsy, performed at a hospital with the patient under general anesthesia. Compared with needle aspiration cytology methods, this type of procedure can remove large tissue samples. The data set consisted of 2480 and 5429 images of benign and malignant tumors, respectively, with 700×460 pixels. The images were three-channel RGB images with 8-bit depth in each channel and were in PNG format. Images were divided into two main groups, those of benign tumors and malignant tumors. Histologically benign is a term

Table 4
Number distribution of BreakHis data set at each magnification level.

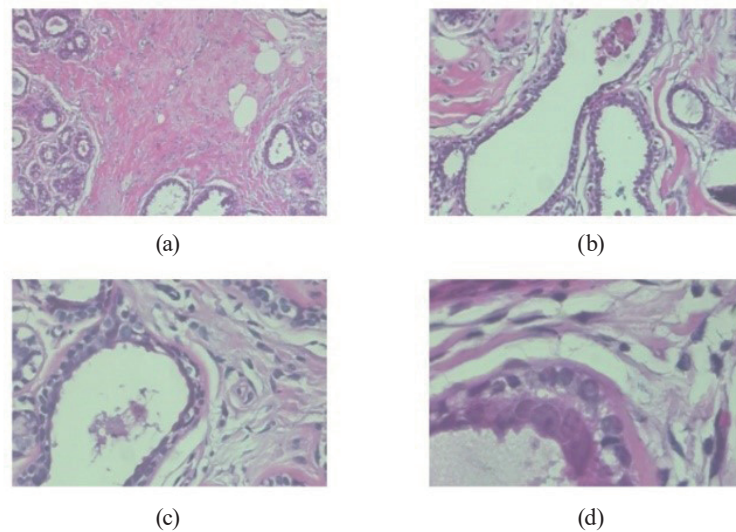| Magnification	Benign	Malignant	Total
40×	652	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total no. of images	2480	5429	7909

(a)

(b)

(c)

(d)

Fig. 3.   (Color online) Examples of images with magnification levels of (a) 40×, (b) 100×, (c) 200×, and (d) 400×.

referring to a lesion that does not satisfy any criterion of malignancy, such as mitosis, marked cellular atypia, basement membrane disruption, and metastasis. Malignant tumors represent the occurrence of cancer, and the lesion can invade and spread to distant sites and cause death.

3.2   Experimental results and analysis

The minimum number of experiments was 13. With reference to Table 3, we selected the orthogonal array more than 16 times. This experiment adopted a mixture of levels of 2 and 3. To satisfy the level number, $L_{36}(2^{11}, 3^{12})$ was selected as the experimental orthogonal array. A total of 36 experiments were conducted with the use of up to 11 factors with 2 levels and 12 factors with 3 levels. In this experiment, levels 2 and 3 had four factors, and eight factors were classified as A–H. Table 5 depicts the experimental results. Each experiment contained three observations used to perform three tests on the same hyperparameter combination. In each test, the network training parameters should be cleared, and each observation is independently trained. The S/N ratio was calculated using Eq. (1). In this experiment, we used the S/N ratio with larger-is-better characteristics. Therefore, the larger the S/N ratio, the higher the robustness of the network composed of this factor.

Table 5
Orthogonal array and the S/N ratios in the experiment.

No.	A	B	C	D	E	F	G	H	y_1	y_2	y_3	Avg	S/N
1	3	4	1	0	3	8	1	0	78.00	80.70	82.43	80.38	−1.904
2	5	6	1	0	5	16	1	0	80.15	71.85	82.39	78.13	−2.190
3	7	12	1	0	7	32	1	0	82.30	81.96	83.02	82.43	−1.679
4	3	4	1	0	3	16	1	0	82.89	82.85	80.28	82.01	−1.726
5	5	6	1	0	5	32	1	0	83.10	82.09	83.10	82.76	−1.644
6	7	12	1	0	7	8	1	0	68.31	68.65	68.31	68.42	−3.296
7	3	6	1	1	7	8	2	1	82.81	82.76	76.57	80.71	−1.879
8	5	12	1	1	3	16	2	1	82.43	80.78	82.26	81.82	−1.744
9	7	4	1	1	5	32	2	1	82.09	81.80	82.81	82.23	−1.699
10	3	12	2	0	5	8	2	1	81.54	83.19	83.48	82.74	−1.647
11	5	4	2	0	7	16	2	1	84.62	82.81	77.58	81.67	−1.777
12	7	6	2	0	3	32	2	1	82.30	82.93	82.09	82.44	−1.677
13	3	12	2	1	3	32	1	1	83.61	82.09	84.37	83.36	−1.583
14	5	4	2	1	5	8	1	1	82.39	82.17	82.76	82.44	−1.677
15	7	6	2	1	7	16	1	1	81.88	81.96	82.26	82.03	−1.720
16	3	12	2	1	5	8	2	0	83.94	83.06	82.13	83.04	−1.615
17	5	4	2	1	7	16	2	0	83.48	83.69	83.52	83.56	−1.560
18	7	6	2	1	3	32	2	0	83.19	82.81	83.10	83.03	−1.615
19	3	4	1	1	7	32	2	0	82.17	81.29	83.48	82.31	−1.692
20	5	6	1	1	3	8	2	0	82.55	81.88	83.94	82.79	−1.642
21	7	12	1	1	5	16	2	0	82.81	80.40	83.19	82.13	−1.713
22	3	6	1	1	7	32	1	1	83.90	83.78	82.68	83.45	−1.572
23	5	12	1	1	3	8	1	1	82.60	82.72	83.16	82.83	−1.637
24	7	4	1	1	5	16	1	1	82.76	83.40	83.02	83.06	−1.612
25	3	6	1	0	3	16	2	1	83.90	83.73	83.10	83.58	−1.559
26	5	12	1	0	5	32	2	1	82.85	80.53	81.96	81.78	−1.749
27	7	4	1	0	7	8	2	1	83.02	83.31	75.22	80.52	−1.912
28	3	6	2	1	5	16	1	0	82.81	83.14	82.68	82.88	−1.631
29	5	12	2	1	7	32	1	0	82.30	81.33	82.30	81.98	−1.727
30	7	4	2	1	3	8	1	0	82.55	84.24	81.92	82.90	−1.630
31	3	12	2	0	7	16	2	0	84.53	81.21	83.57	83.10	−1.611
32	5	4	2	0	3	32	2	0	84.32	82.09	82.68	83.03	−1.617
33	7	6	2	0	5	8	2	0	82.05	83.19	82.75	82.66	−1.654
34	3	4	2	0	5	32	1	1	82.22	82.89	82.01	82.37	−1.685
35	5	6	2	0	7	8	1	1	83.40	82.26	81.71	82.46	−1.676
36	7	12	2	0	3	16	1	1	83.35	82.68	80.40	82.14	−1.712

The response table for the S/N ratio, significance, and percentage contribution of each factor is presented in Table 6. When the level in the factor is changed independently, the robustness to the network performance changes. If the level of variation changes, then the greater the degree of variation in the network performance, the greater the effect this factor has on the network. We listed a factor as a significant one if the significance was higher than a threshold. When the optimization does not reach the goal, the level can be adjusted for the significance factor. Finally, the best parameter combination we selected was A1, B2, C2, D2, E1, F3, G2, and H2. The significance parameters were ranked in the order F > E > C > A > D > G > B > H. The percentage contribution order of the factors with the hyperparameter optimization to obtain a high accuracy rate is as follows: Conv1_Stride > Conv2_Filter > Conv2_Kernel size > Conv1_Kernel size > Conv1_padding > Conv2_Stride > Conv1_Filter > Conv2_padding.

Table 6
Response table for S/N ratio, significance, and percentage contribution of the factor.

Factor	A	B	C	D	E	F	G	H
Level 1	−1.675	−1.708	−1.825	−1.802	−1.670	−1.848	−1.794	−1.786
Level 2	−1.720	−1.705	−1.656	−1.679	−1.710	−1.713	−1.687	−1.695
Level 3	−1.828	−1.809	—	—	−1.842	−1.662	—	—
Delta	0.153	0.104	0.169	0.123	0.172	0.186	0.107	0.091
Rank	4	7	3	5	2	1	6	8
Contribution (%)	12.11	6.99	21.01	11.10	15.90	18.22	8.6	6.07
Best level	1	2	2	2	1	3	2	2

Table 7
Comparison of network performance between the original network and our method.

ABCDEFGH	y_1	y_2	y_3	Avg
Original CNN	83.48	77.92	80.66	80.69
Our method	83.10	83.23	83.23	83.19

Table 8
Comparison of performances using various methods.

Method	Performance (%)
Single−layer CNN[17]	77.50
RF classifier+PFTAS[18]	81.28
LeNet−5(Sgdm)[19]	80.69
LeNet−5(Adam)[20]	82.22
LeNet−5(RMSprop)[21]	82.58
Our method	83.19

Table 7 shows a comparison of the performance of the novel network with the original LeNet network. We trained the same data set (BreakHis) and tested the original network and the optimized network three times without retaining the training parameters. As shown in Table 7, compared with the original network, the proposed method has more average observations in the three tests, which proves that the proposed optimization method improved the robustness of the network.

Table 8 presents a comparison of the performances using various methods. The single-layer CNN network architecture includes a convolutional layer and a pooling layer. The number of channels in the convolutional layer is 50, which extracts more feature maps.[17] For comparison with traditional manual feature extraction, the extracted features are added to the CNN training method.[18] In LeNet-5, the accuracy rates are obtained using different gradient descent methods.[19–21] Therefore, the experimental results prove that the performance of our method reaches 83.19%, which is significantly (5.69%) higher than that of the single-layer CNN method. The proposed optimized network architecture exhibited the best performance.

4.   Conclusions

We proposed the hyperparameter optimization of deep learning networks. The Taguchi method was used to perform experiments on the BreakHis data set to classify tumors as malignant or benign. On the basis of the experimental results, the best hyperparameter

combination of a CNN based on deep learning networks can be obtained. The optimum parameters are as follows: Conv1_Kernel size = 3, Conv1_Filter = 6, Conv1_Stride = 2, Conv1_padding = 1, Conv2_Kernel size = 3, Conv2_Filter = 32, Conv2_Stride = 2, and Conv2_padding = 1. The influence of each control factor on the network performance provides the basis for the adjustment of each hyperparameter and improves its performance. The results show that compared with the unoptimized CNN based on deep learning networks, the performance of the optimized networks was improved, and the proposed method is superior to other methods. However, the proposed hyperparameter method was limited to two convolution layers and only applied to the BreakHis data set in this study. Therefore, the proposed hyperparameter optimization of deep learning networks should be performed on a different data set to classify other medical histopathological images in the future.

References

1 T. Wan, J. Cao, J. Chen, and Z. Qin: Neurocomputing **229** (2017) 34. https://doi.org/10.1016/j.neucom.2016.05.084
2 S. Khan, N. Islam, Z. Jan, I. U. Din, and P. C. Rodrigues: Pattern Recognit. Lett. **125** (2019) 1. https://doi.org/10.1016/j.patrec.2019.03.022
3 T. Kausau, M. Wang, M. Idrees, and Y. Lu: Biocybern. Biomed. Eng. **39** (2019) 967. https://doi.org/10.1016/j.bbe.2019.09.003
4 B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore: Pattern Recognit. **84** (2018) 345. https://doi.org/10.1016/j.patcog.2018.07.022
5 P. Alirezazadeh, B. Hejrati, A.M. Esfahani, and A. Fathi: BiocybernBiomed Eng. **38** (2018) 671. https://doi.org/10.1016/j.bbe.2018.04.008
6 S. Moazemi, Z. Khurshid, A. Erle, S. Lütje, M. Essler, T. Schultz, and R. A. Bundschuh: Diagnos. **10** (2020) 622. https://doi.org/10.3390/diagnostics10090622
7 A. F. Moustafa, T. W. Cary, L. R. Sultan, S. M. Schultz, E. F. Conant, S. S. Venkatesh, and C. M. Sehgal: Diagnos. **10** (2020) 631. https://doi.org/10.3390/diagnostics10090631
8 A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwiggelaar: Med. Image Anal. **47** (2018) 45. https://doi.org/10.1016/j.media.2018.03.006
9 H. Feng, J. Cao, H. Wang, Y. Xie, D. Yang, J. Feng, and B. Chen: Magn. Reson. Imaging. **69** (2020) 40. https://doi.org/10.1016/j.mri.2020.03.001
10 A. Kumar, S. K. Singh, S. Saxena, K. Lakshmanan, A. K. Sangaiah, H. Chauhan, S. Shrivastava, and R. K. Singh: Inf. Sci. **508** (2020) 405. https://doi.org/10.1016/j.ins.2019.08.072
11 R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao: Mech. Syst. Sig. Process. **115** (2019) 213. https://doi.org/10.1016/j.ymssp.2018.05.050
12 A. Kamilaris and F. Prenafeta-Boldú: Comput. Electron. Agric. **147** (2018) 70. https://doi.org/10.1016/j.compag.2018.02.016
13 R. Rasti, M. Teshnehlab, and S. L. Phung: Pattern Recognit. **72** (2017) 381. https://doi.org/10.1016/j.patcog.2017.08.004
14 P. Khosravi, E. Kazemi, M. Imielinski, O. Elemento, and I. Hajirasouliha: EBioMedicine **27** (2018) 317. https://doi.org/10.1016/j.ebiom.2017.12.026
15 B. K. Sabeena, S. N. Madhu, and G. R. Bindu: Biocybern. Biomed. Eng. **39** (2019) 214. https://doi.org/10.1016/j.bbe.2018.10.007
16 L. Ningbo, X. Yanan, T. Yonghua, M. Hongwei, and W. Shuliag: Future Gener. Comput. Syst. **94** (2019) 524. https://doi.org/10.1016/j.future.2018.11.036
17 M. E. Nejad, L. S. Affendey, R. B. Latip, and I. B. Ishak: 2017 Int. Conf. Imaging, Signal Processing and Communication (2017) 50–53. https://doi.org/10.1145/3132300.3132331
18 M. Sharma, R. Singh, and M. Bhattacharya: 2017 IEEE Int. Conf. Bioinformatics and Biomedicine (IEEE, 2017) 1110–1113. https://doi.org/10.1109/BIBM.2017.8217811
19 C. Darken, J. Chang, and J. Moody: Neural Networks for Signal Processing II Proc. 1992 IEEE Workshop (IEEE, 1992) 3–12. https://doi.org/10.1109/NNSP.1992.253713
20 D. Kingma and J. Ba: Proc. 3rd Int. Conf. Learning Representations (2015) 1–15.
21 M. C. Mukkamala and M. Hein: Proc. 34th Int. Conf. Machine Learning **70** (2017) 2545–2553.

22 Y. Zhao, K. Hao, H. He, X. Tang, and B. Wei: Neurocomputing **380** (2020) 259. https://doi.org/10.1016/j.neucom.2019.10.067

23 C. Li, X. Wang, W. Liu, L.J. Latecki, B. Wang, and J. Huang: Med. Image Anal. **53** (2019) 165. https://doi.org/10.1016/j.media.2019.01.013

24 Z. Cai, Y. Long, and L. Shao: Pattern Recognit. Lett. **125** (2019) 396. https://doi.org/10.1016/j.patrec.2019.05.021

25 F. N. Idris, M. M. Nadzir, and S. R. A. Shukor: J. Environ. Chem. Eng. **8** (2020) 103766. https://doi.org/10.1016/j.jece.2020.103766

26 A. Rezania, S. A. Atouei, and L. Rosendahl: Mater. Today Energy **16** (2020). https://doi.org/10.1016/j.mtener.2019.100376

27 S. Debnath, M. Reddy, and Q. Yi: Measurement. **78** (2016) 111. https://doi.org/10.1515/amm-2017-0276

28 F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte: IEEE Trans. Biomed. Eng. **63** (2016) 1455. https://doi.org/10.1109/TBME.2015.2496264

About the Authors

Cheng-Jian Lin received his B.S. degree in electrical engineering from Ta Tung Institute of Technology, Taipei, Taiwan, R.O.C., in 1986, and his M.S. and Ph.D. degrees in electrical and control engineering from National Chiao-Tung University, Taiwan, R.O.C., in 1991 and 1996, respectively. Currently, he is a chair professor of the Computer Science and Information Engineering Department, National Chin-Yi University of Technology, Taichung, Taiwan, R.O.C., and Dean of Intelligence College, National Taichung University of Science and Technology, Taichung, Taiwan, R.O.C. His current research interests are machine learning, pattern recognition, intelligent control, image processing, intelligent manufacturing, and evolutionary robots. (cjlin@ncut.edu.tw)

Shiou-Yun Jeng received her Ph.D. degree in industrial engineering and management from National Yunlin University of Science & Technology, Taiwan, R.O.C., in 2019. Currently, she is a postdoctoral research fellow of the Computer Science and Information Engineering Department, National Chin-Yi University of Technology, Taichung City, Taiwan, R.O.C. Her current research interests are artificial intelligence, fuzzy theory, sustainable supply chain management, resource efficiency, green production management, and big data analysis.

Chin-Ling Lee received her B.S. degree in English literature from Tamkang University, Taiwan, R.O.C., in 1986, her M.S. degree in English literature from Central Missouri State University, U.S.A., in 1990, and her Ph.D. degree in industrial education from National Taiwan Normal University, Taiwan, R.O.C., in 2005. From August 2003 to July 2005, she was an assistant professor in the Department of Applied Foreign Languages, Nan-Kai Institute of Technology, Nantou, Taiwan, R.O.C. Currently, she is an associate professor in the International Business Department, National Taichung University of Science and Technology. Her current research interests are English teaching, time series prediction, machine learning, and intelligent systems.