Sensors and Materials, Vol. 35, No. 3 (2023) 795–811
MYU Tokyo

S & M 3211

Vector Deep Fuzzy Neural Network for Breast Cancer Classification

Cheng-Jian Lin,[1,2*] Mei-Yu Wu,[3] Yi-Hsuan Chuang,[1] and Chin-Ling Lee[4]

[1]Department of Computer Science & Information Engineering, National Chin-Yi University of Technology, Taichung 411, Taiwan
[2]College of Intelligence, National Taichung University of Science and Technology, Taichung 404, Taiwan
[3]Department of Business Management, National Taichung University of Science and Technology, Taichung 404, Taiwan
[4]Department of International Business, National Taichung University of Science and Technology, Taichung 404, Taiwan

(Received September 12, 2022; accepted November 24, 2022)

Keywords: breast cancer classification, deep learning, fuzzy neural network, Taguchi method

Breast cancer is one of the most common cancers in women worldwide and the leading cause of death in women. Medical experts use histopathological images to diagnose breast cancer, but such analysis for the effective diagnosis or detection of breast cancer is challenging. Therefore, we propose a vector deep fuzzy neural network (VDFNN) to classify breast cancer effectively and automatically from histopathological images. The VDFNN model uses four sets of vector product and pooling layers to extract features and retain important feature information. Then, a feature fusion layer uses global average pooling to reduce the dimension of the extracted feature information. Finally, a fuzzy neural network performs breast cancer classification. The VDFNN model parameters are selected using the trial-and-error method. However, we also propose the Taguchi-VDFNN (T-VDFNN), which employs the Taguchi method to determine the optimal combination of model parameters. The experimental breast cancer classification accuracy of the proposed VDFNN was 92.18%. After the application of the Taguchi method to identify the optimal parameter combination, the experimental accuracy of the proposed T-VDFNN model was 94.37%, 2.19 percentage points higher than that of the basic VDFNN model.

1. Introduction

According to 2020 cancer statistics, the incidence of breast cancer is 24.5%, and the mortality rate is 15.5%, which are the highest for cancers in women.(1) Early diagnosis can reduce mortality due to breast cancer, and early breast cancer treatment can cure up to 90% of cases.(2) Effective imaging research extracts useful information from abundant raw data to aid in the detection of pathological lesions, clinical treatment of patients, and differential diagnoses. The most commonly used medical imaging modalities are mammography, ultrasonography, magnetic resonance imaging, computed tomography, and histopathological imaging.(3) Rubin and

*Corresponding author: e-mail: cjlin@ncut.edu.tw
https://doi.org/10.18494/SAM4121

ISSN 0914-4935
https://myukk.org/

Strayer[4] used histopathological imaging to obtain and examine microscopic images of tissue; such imaging is the gold standard for cancer diagnosis. However, histopathological analysis is time consuming and depends on the skill and experience of the pathologist. Diagnosis is sometimes subjective and can be influenced by multiple human factors, such as fatigue and inattention.[5] Therefore, accurate histopathological analysis methods are urgently needed.

Within the rapidly growing field of artificial intelligence, machine learning uses real data for training and learning to construct predictive models. In medical diagnosis, machine learning can be used to process raw medical data quickly and provide useful information that can help pathologists analyze large quantities of medical data efficiently. Thus, machine learning has greatly improved the early diagnosis and prediction of cancer.[6] Machine learning techniques, including simple Bayesian classifiers,[7] *K*-nearest neighbor algorithms,[7] decision trees,[8] support vector machines,[9] and artificial neural networks,[10] have been widely applied to breast cancer classification problems. However, machine learning requires feature extraction to be performed in advance. Moreover, such feature extraction is based on the experience of experts, the process is complex, and the number of extracted features affects the prediction results.

In recent years, deep learning techniques have been widely used to overcome the difficulty of manually defining features in various domains. Ortac and Ozcan[11] used 1D, 2D, and 3D convolutional models to achieve effective classification. Their 3D convolutional neural network (CNN) achieved higher classification accuracy than other state-of-the-art models. In another study,[12] a CNN architecture was used to classify materials in multispectral remote sensing images for simplified model building. Houssein *et al.*[3] proposed new applications of CNNs related to breast cancer detection and classification. When a data set is large, CNNs outperform traditional machine learning in breast cancer diagnosis.[13] Traditional clinical diagnosis is often uncertain and ambiguous, and fuzzy neural networks (FNNs) are often used to reduce this uncertainty.[14] FNNs are based on a hybrid approach that combines the semantic transparency of rule-based fuzzy systems with the learning capabilities of neural networks.[15] Because CNNs have too many parameters and require high-performance hardware, scholars have employed convolution operations with FNNs to reduce the number of network parameters. For example, Abiyev and Helwan[16] used FNNs to obtain good performance in breast cancer classification.

Because CNNs require numerous parameters, the trial-and-error method is widely used for parameter selection. To reduce the time and cost of experiments, the Taguchi method[17] can be used to statistically optimize parameter selection by using an orthogonal array of influencing factors and their levels.

In this paper, a vector deep FNN (VDFNN) that automatically and effectively classifies breast cancer from histopathological images is proposed. The VDFNN model architecture uses four sets of vector product and pooling layers to extract features and retain important feature information. A global average pooling (GAP) method is applied in a feature fusion layer to reduce the dimension of the feature information. Finally, to classify breast cancer, an FNN is used instead of the fully connected network used by traditional CNNs. To reduce the time and cost of parameter selection, the Taguchi-VDFNN (T-VDFNN) method, based on the Taguchi experimental design, is used, resulting in a slightly modified T-VDFNN model. The major contributions of this study are as follows:

1. The VDFNN is proposed for breast cancer classification problems. A GAP method is used in a feature fusion layer, and an FNN is used instead of a fully connected network to reduce the number of parameters.
2. The T-VDFNN method, based on the Taguchi experimental design, is proposed for optimal parameter selection.

The remainder of this paper is organized as follows. Section 2 describes the structure of the proposed VDFNN and presents the Taguchi method used to develop the T-VDFNN for optimal parameter selection. Section 3 presents the experimental results obtained using the proposed VDFNN and T-VDFNN models. Finally, the conclusions of this study and recommendations for future work are presented in Sect. 4.

2. Materials and Methods

In this section, the overall network architecture of the VDFNN is introduced. Because the VDFNN relies on the trial-and-error method for parameter selection, a VDFNN based on the Taguchi method, called T-VDFNN, is used for optimal parameter selection.

2.1 Architecture of VDFNN

The VDFNN architecture comprises a vector convolutional layer, a pooling layer, a feature fusion layer, and FNN (Fig. 1). The vector convolutional layer is used to extract features from input images, and dimension reduction operations are performed on extracted feature maps in the max pooling layer. The feature fusion layer replaces the traditional flatten layer; it integrates feature information to obtain less but more useful feature information. Finally, the FNN is used for classification.

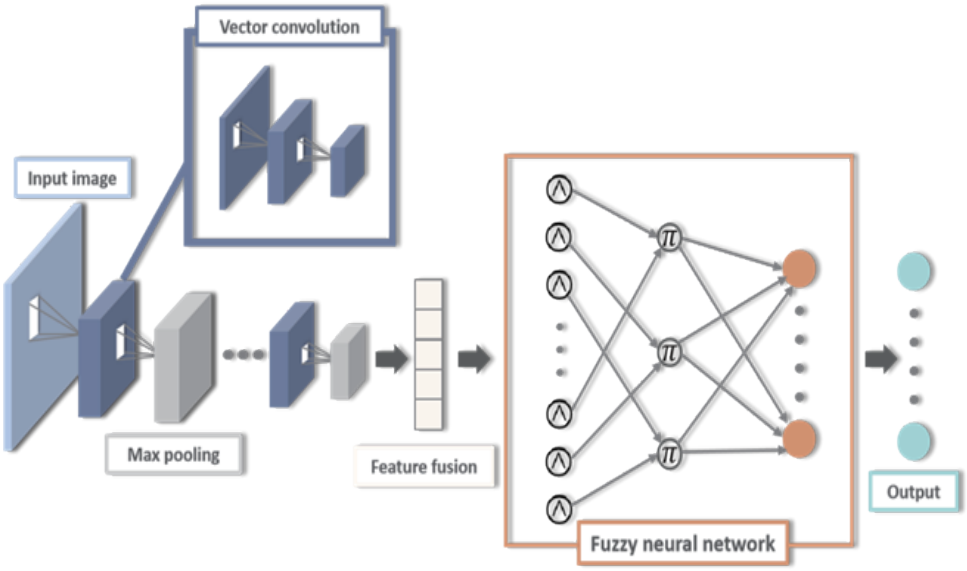


Fig. 1. (Color online) Architecture of VDFNN.

2.1.1 Vector convolutional layer

The vector convolutional layer employs vector convolution that divides the traditional convolutional layer into two layers.[18] The main purpose is to compress the convolution kernel to eliminate redundant network parameters. The principle is to use $n \times 1$ and $1 \times n$ vector kernels to obtain an $n \times n$ matrix through matrix operation. The result obtained by calculating the original 3×3 matrix from the matrices of the 3×1 and 1×3 convolution kernels is shown in Fig. 2. The formula is as follows:

$$A_{n \times n} = B_{n \times 1} * C_{1 \times n}, \tag{1}$$

where $A_{n \times n}$ is an $n \times n$ matrix, $B_{n \times 1}$ and $C_{1 \times n}$ are $n \times 1$ and $1 \times n$ matrices, respectively, and * denotes matrix multiplication.

In traditional convolution operations, if a 3×3 convolution kernel is used, nine parameters are required. If a traditional convolution kernel is divided into two vector convolution kernels and then subjected to data compression, the vector convolution kernels require only $3 + 3 = 6$ parameters. Therefore, the number of parameters required by the convolution kernel and thus the corresponding number of operations can be reduced, as depicted in Fig. 3.

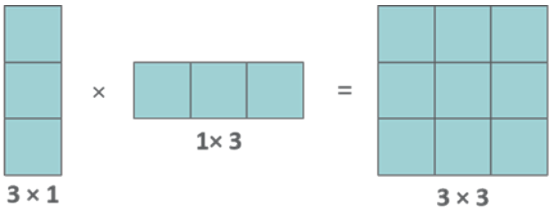


Fig. 2. (Color online) Vector convolution.

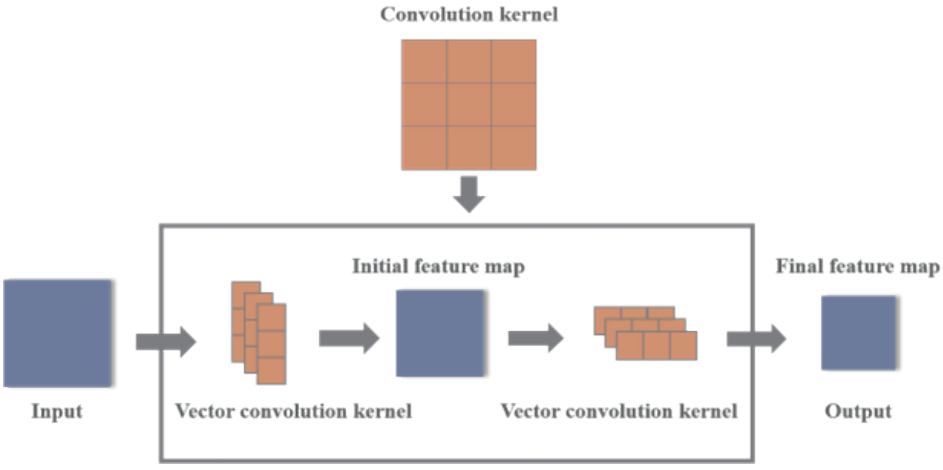


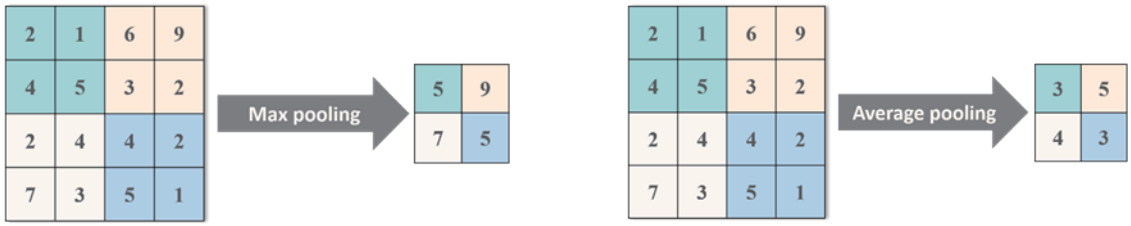
Fig. 3. (Color online) Vector convolution kernel operation.

2.1.2 Pooling layer

The pooling layer reduces the size of the image features while retaining relevant feature information. Commonly used operations in pooling layers are the max pooling method [illustrated in Fig. 4(a)] and average pooling method [illustrated in Fig. 4(b)]. The pooling operation works similarly to the convolution operation. A matrix of a specific size is used to slide the input image features, and the maximum or average value in the overlapping area is taken as the feature map output after the operation.

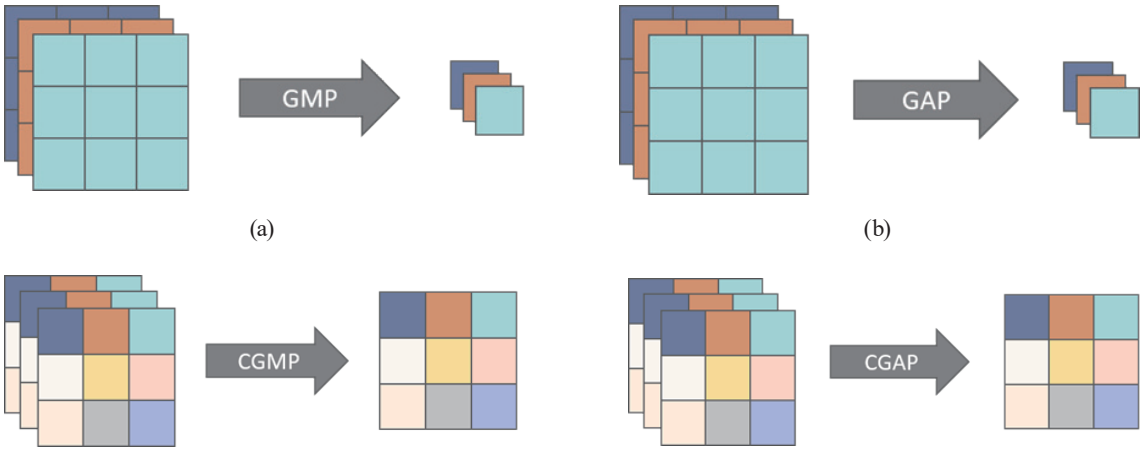
2.1.3 Feature fusion layer

The feature maps obtained in the pooling layer are integrated through various fusion methods to reduce the feature dimension and obtain more useful feature information. In this study, four fusion methods were considered. Global pooling is used to perform operations on each feature map and fuse maps separately and can be either global max pooling [GMP; Fig. 5(a)] or GAP [Fig. 5(b)], and channel global pooling is used to fuse all channels of each feature map and can be either channel GMP [CGMP; Fig. 5(c)] or channel GAP [CGAP; Fig. 5(d)] operations.



(a) (b)

Fig. 4. (Color online) (a) Max pooling method and (b) average pooling method.



(a) (b)

(c) (d)

Fig. 5. (Color online) (a) GMP, (b) GAP, (c) CGMP, and (d) CGAP.

2.1.4 FNN

An FNN operates using fuzzy logic to imitate human thinking and possesses neural-network-like learning ability. Therefore, it is capable of effectively managing data ambiguity and preventing the influence of noise. The architecture of the FNN used in the VDFNN is illustrated in Fig. 6. The network comprises a fuzzification layer, a fuzzy rule layer, and a defuzzification layer. After the feature fusion layer, the input feature map is fuzzified using a Gaussian membership function in the fuzzification layer, and the fuzzified map is processed in the if~then~ fuzzy rule layer for inference. The established fuzzy rules are as follows:

$$\text{Rule}_j : \text{If } x_1 \text{ is } A_{1j} \text{ and } x_2 \text{ is } A_{2J} \ldots \text{ and } x_n \text{ is } A_{nj}, \text{then } y_j \text{ is } w_j , \tag{2}$$

where x_i is the input, A_{ij} is the membership function, w_j is the output, and j is the jth fuzzy rule.

The corresponding membership degree is obtained in the fuzzy rule layer as follows:

$$\mu_{ij}(x) = \exp\left(-\frac{\left[x_i - m_{ij} \right]^2}{\sigma_{ij}^2} \right), \tag{3}$$

where x_i is the input, m_{ij} is the mean value, and σ_{ij} is the standard deviation.

The membership values corresponding to each input are combined to obtain the firing strength of each fuzzy rule. In this study, a product operation is used to obtain the rule firing strength as follows:

$$R_j = \prod_{i=1}^{n} \mu_{ij} . \tag{4}$$

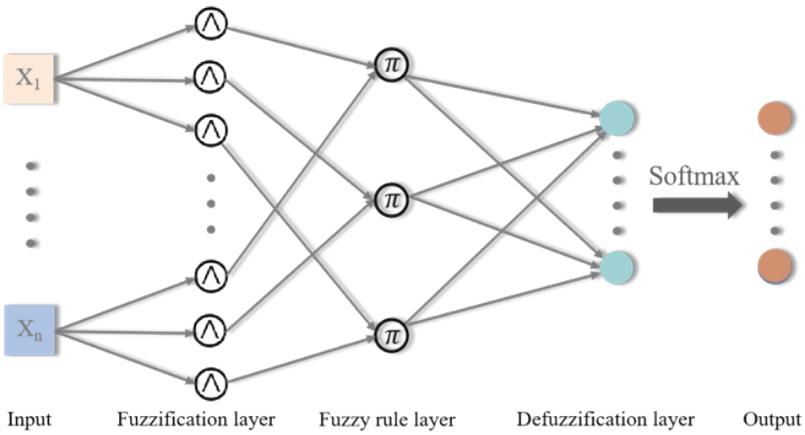


Fig. 6. (Color online) Structure of FNN.

Finally, the fuzzy output is defuzzified to a crisp output as follows:

$$y_k = \sum_{j=1}^{r} R_j w_{jk} , \tag{5}$$

where y_k is the kth output, R_j is the excitation intensity of the jth fuzzy rule, and w_{jk} is the output of the jth fuzzy rule and kth output.

2.2 Proposed T-VDFNN model

The parameters of the general VDFNN model are selected using the trial-and-error method. Therefore, the T-VDFNN was devised to determine the optimal parameter selection. The Taguchi method,[19] proposed in 1950, is an experimental design method that combines complex mathematical and statistical techniques and is used in academic and industrial research. Experimental design using an orthogonal table can effectively reduce the number of necessary experiments while achieving high quality. It is a low-cost, high-efficiency quality engineering method, and it is widely used to find the optimal parameter selection during system optimization.[20–22] A flowchart of the proposed T-VDFNN method is illustrated in Fig. 7. In brief, breast cancer images are preliminarily classified into training and test sets. The training images are input into the VDFNN model to train it. Then, the impact factors and their levels are selected, and the Taguchi method is used to determine the optimal parameter selection. When the experimental results do not meet the user's requirements, the impact factor or level is

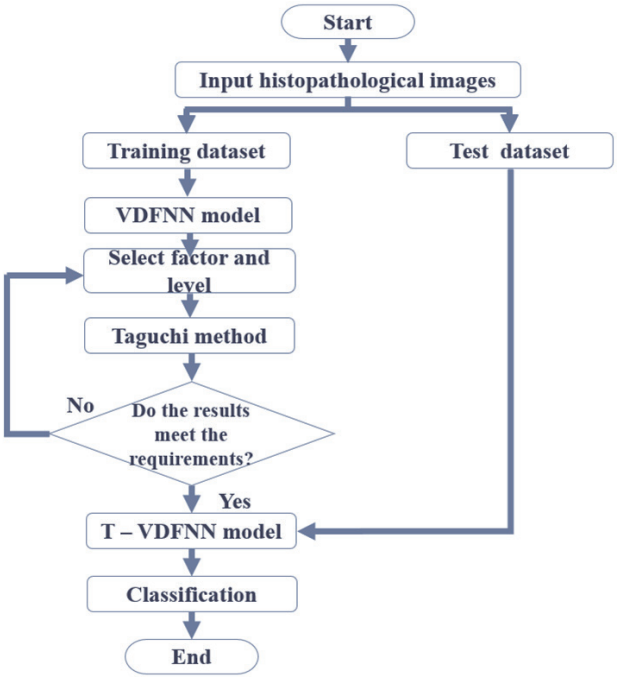


Fig. 7. (Color online) Flowchart of proposed T-VDFNN.

reselected. If the results meet the user's requirements, the trained T-VDFNN model is tested for its ability to classify breast tumors as benign or malignant.

2.2.1 Orthogonal table

Orthogonal tables are used in statistics for experimental design. By using the data obtained in experiments, a mathematical formula can be derived to predict the output of factor combination. The obtained parameter selection is used to conduct a confirmation experiment. To reduce the number of experiments required to obtain the optimal parameter selection, a Taguchi experiment with a standard orthogonal table (as presented as Table 1) is used for the T-VDNN. During selection, the numbers of factors and levels must be determined before the configuration of the experimental combinations.

2.2.2 Signal-to-noise ratio

The quality loss of a product can be calculated and quantified mathematically. The concept of quality loss has evolved to mean the signal-to-noise (*S/N*) ratio. The *S/N* is a logarithmic function of the output used to measure quality characteristics, and it is useful for performing data analysis and predicting optimal results.[23] Three quality characteristics are described as follows:

Nominal-the-best:

$$S/N = -10\log\left[\frac{\sum_{i=0}^{n}(y_i - m)^2}{n}\right]. \tag{6}$$

Table 1
Standard orthogonal table for Taguchi experiment.

Number of experiments	Maximum number of factors	Maximum level			
		2	3	4	5
L4	3	3	—	—	—
L8	7	7	—	—	—
L9	4	—	4	—	—
L12	11	11	—	—	—
L16	15	15	—	—	—
L'16	5	—	—	5	—
L18	8	1	7	—	—
L25	6	—	—	—	6
L27	13	—	13	—	—
L32	31	31	—	—	—
L'32	10	1	—	9	—
L36	23	11	12	—	—
L'36	16	3	13	—	—
L50	12	1	—	—	11
L54	26	1	25	—	—
L64	63	63	—	—	—
L'64	21	—	—	21	—
L81	40	—	40	—	—

Smaller-the-better:

$$S/N = -10\log\left[\frac{\sum_{i=0}^{n}(y_i)^2}{n}\right]. \tag{7}$$

Larger-the-better:

$$S/N = -10\log\left[\frac{\sum_{i=0}^{n}\left(\frac{1}{y_i}\right)^2}{n}\right]. \tag{8}$$

2.3 Evaluation metrics

A confusion matrix integrates model predictions in matrix form for simple evaluation. In this study, we focused on the binary classification of breast cancer as benign or malignant. An example 2 × 2 confusion matrix for breast cancer classification is presented in Table 2. A confusion matrix clearly visualizes each performance metric of the predictive classification model. Therefore, a confusion matrix is often used to evaluate the performance of machine learning and deep learning models employed for classification.[24,25]

From the true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) statistics, the accuracy, sensitivity, and specificity of the model can be calculated as follows:

$$\text{Accuracy: } \frac{TP+TN}{TP+FP+FN+TN}. \tag{9}$$

$$\text{Sensitivity: } \frac{TP}{TP+FN}. \tag{10}$$

$$\text{Specificity: } \frac{TN}{TN+FP}. \tag{11}$$

3. Experimental Results

We designed experiments to test the effectiveness of the proposed VDFNN and T-VDFNN models. Section 3.1 describes the data source. Section 3.2 describes the initial architecture of the

Table 2
Confusion matrix for binary breast cancer classification.

		Actual	
		Malignant (Positive)	Benign (Negative)
Prediction	Malignant (Positive)	True Positive (*TP*)	False Positive (*FP*)
	Benign (Negative)	False Negative (*FN*)	True Negative (*TN*)

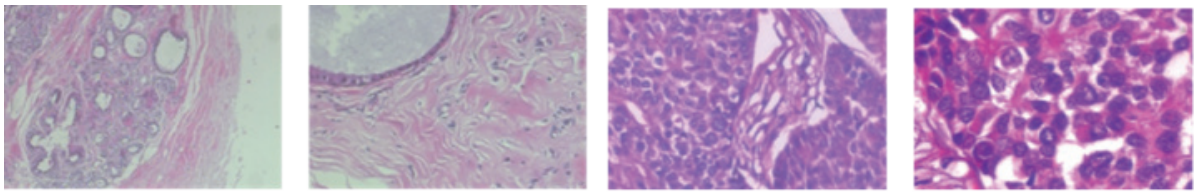
VDFNN model and the experimental results obtained using different fusion methods. The experimental results of the Taguchi method for optimizing the parameter combination for the T-VDFNN model are presented in Sect. 3.3.

3.1 Data set

In this study, we used the BreakHis data set,[26] which consists of 7909 images of 82 patients at four magnifications (40×, 100×, 200×, and 400×), as illustrated in Fig. 8 and summarized in Table 3. This data set was compiled by the Brazilian Laboratory of Pathological Anatomy and Cytopathology through pathologist staining of breast biopsy slides with hematoxylin and eosin. The size of the PNG images in the database is 700 × 460 pixels, and the images use the primary three-channel red–green–blue color mode with eight bits of depth per channel.

3.2 Classification results obtained using VDFNN

In this study, TensorFlow and Keras were used as the deep learning environment and development tool, respectively. The parameter settings of the proposed VDFNN model are summarized in Table 4. The input image size was set to 224 × 224 × 3, and four vector convolutional and pooling layers were used for feature extraction. In each vector convolutional layer, 3 × 1 and 1 × 3 convolution kernels were used for feature extraction. The feature dimension is then reduced through a 2 × 2 max pooling layer to reduce the number of computations. In the vector convolution layer, 32, 64, 128, and 64 are used as the number of four-layer vector convolution kernels to extract various feature combinations. The feature fusion layer is used to reduce the dimensionality of the features of the previous layer. Finally, the fuzzy rule layer is used for classification output.



(a) (b) (c) (d)

Fig. 8. (Color online) Breast cancer images at (a) 40 ×, (b) 100 ×, (c) 200 ×, and (d) 400 × magnification.

Table 3
Distribution of malignant and benign images in BreakHis data set at each magnification.

Image magnification	Benign	Malignant	Total
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total number of images	2480	5429	7909

Table 4
Parameters settings of the proposed VDFNN model.

Layer	Number of filters	Kernel size	Stride	Padding
Convolution_1-1	32	3 × 1	2, 1	0
Convolution_1-2	32	1 × 3	1, 2	0
Max_pooling_1	—	2 × 2	—	—
Convolution_2-1	64	3 ×1	1, 1	0
Convolution_2-2	64	1 × 3	1, 1	0
Max_pooling_2	—	2 × 2	—	—
Convolution_3-1	128	3 × 1	1, 1	0
Convolution_3-2	128	1 × 3	1, 1	0
Max_pooling_3	—	2 × 2	—	—
Convolution_4-1	64	3 × 1	1, 1	0
Convolution_4-2	64	1 × 3	1, 1	0
Max_pooling_4	—	2 × 2	—	—
Feature fusion	64	—	—	—
FuzzyRuleLayer	64	—	—	—
DeFuzzify Layer	2	—	—	—

Table 5
Experimental results obtained using four fusion methods.

	Fusion method	Accuracy (%)
VDFNN	GAP	92.98
	GMP	92.10
	CGAP	91.58
	CMP	91.78

For the VDFNN, four fusion methods were considered: GMP, GAP, CGMP, and CGAP. We conducted experiments using these four fusion methods and obtained classification results. As summarized in Table 5, the highest classification accuracy, 92.98%, was obtained using the GAP method.

3.3 Classification results obtained using T-VDFNN

Because the basic VDFNN architecture relies on the trial-and-error method for parameter setting, the proposed T-VDFNN uses the Taguchi method to obtain the optimal parameter selection for the VDFNN architecture. In T-VDFNN, five factors (i.e., number of filters and number of fuzzy rules) are selected as three levels, and eight factors (i.e., kernel size and padding) are selected as two levels. Therefore, 13 factors are selected, as summarized in Table 6.

Considering the factors and levels presented in Table 6, the influential factors were determined using an orthogonal table. Thirty-six experiments were conducted, and the *S/N* ratios were calculated, as summarized in Table 7.

Given that accuracy is a crucial classification metric for classifiers, the results of the 36 experiments in the orthogonal table suggest that the higher the *S/N* ratio, the higher the quality of classification. The *S/N* ratio of the 30th parameter combination was −0.48126 and the corresponding accuracy was 94.61%, better than the other 35 parameter combinations.

Table 6
Thirteen factors and their levels in T-VDFNN.

	Factor	Level 1	Level 2	Level 3
A	Conv1_Filter(C1_F)	8	16	32
B	Conv1_Kernel size(C1_K)	3	5	
C	Conv1_Padding(C1_P)	0	1	
D	Conv2_Filter(C2_F)	16	32	64
E	Conv2_Kernel size(C2_K)	3	5	
F	Conv2_Padding(C2_P)	0	1	
G	Conv3_Filter(C3_F)	32	64	128
H	Conv3_Kernel size(C3_K)	3	5	
I	Conv3_Padding(C3_P)	0	1	
J	Conv4_Filter(C4_F)	16	32	64
K	Conv4_Kernel size(C4_K)	3	5	
L	Conv4_Padding(C4_P)	0	1	
M	Number of fuzzy rules	32	64	128

Table 7
Accuracy and *S/N* ratio of each factor and level combination.

No	A	B	C	D	E	F	G	H	I	J	K	L	M	Accuracy	*S/N* ratio
1	8	3	0	16	3	0	32	3	0	16	3	0	32	0.9052	−0.8651
2	16	3	0	32	3	0	64	3	0	32	3	0	64	0.9279	−0.6499
3	32	3	0	64	3	0	128	3	0	64	3	0	128	0.9425	−0.5143
4	8	3	0	16	3	1	32	3	1	16	3	1	64	0.909	−0.8287
5	16	3	0	32	3	1	64	3	1	32	3	1	128	0.9191	−0.7327
6	32	3	0	64	3	1	128	3	1	64	3	1	32	0.9235	−0.6912
7	8	3	1	16	3	0	64	5	0	64	5	0	32	0.902	−0.8958
8	16	3	1	32	3	0	128	5	0	16	5	0	64	0.9406	−0.5319
9	32	3	1	64	3	0	32	5	0	32	5	0	128	0.9425	−0.5143
10	8	3	1	16	5	0	128	3	1	32	5	1	32	0.8837	−1.0739
11	16	3	1	32	5	0	32	3	1	64	5	1	64	0.9229	−0.6969
12	32	3	1	64	5	0	64	3	1	16	5	1	128	0.945	−0.4913
13	8	3	1	32	5	1	128	5	0	16	3	1	128	0.6865	−3.2671
14	16	3	1	64	5	1	32	5	0	32	3	1	32	0.9374	−0.5615
15	32	3	1	16	5	1	64	5	0	64	3	1	64	0.9178	−0.7450
16	8	3	0	32	5	1	128	5	1	32	5	1	32	0.9349	−0.5847
17	16	3	0	64	5	1	32	5	1	64	5	0	64	0.9343	−0.5902
18	32	3	0	16	5	1	64	5	1	16	5	0	128	0.9115	−0.8048
19	8	5	0	32	3	0	32	5	1	64	5	1	128	0.9204	−0.7204
20	16	5	0	64	3	0	64	5	1	16	5	1	32	0.9273	−0.6555
21	32	5	0	16	3	0	128	5	1	32	5	1	64	0.9317	−0.6144
22	8	5	1	32	3	1	64	5	1	64	3	0	128	0.9279	−0.6499
23	16	5	1	64	3	1	128	5	1	16	3	0	32	0.9444	−0.4968
24	32	5	1	16	3	1	32	5	1	32	3	0	64	0.9179	−0.7440
25	8	5	1	64	3	1	64	3	0	16	5	1	64	0.9128	−0.7924
26	16	5	1	16	3	1	128	3	0	32	5	1	128	0.9317	−0.6144
27	32	5	1	32	3	1	32	3	0	64	5	1	32	0.8963	−0.9509
28	8	5	0	64	5	0	64	5	0	32	3	1	64	0.9223	−0.7025
29	16	5	0	16	5	0	128	5	0	64	3	1	128	0.9267	−0.6612
30	32	5	0	32	5	0	32	5	0	16	3	1	32	0.9461	−0.4812
31	8	5	0	64	5	1	128	3	0	64	5	0	64	0.9125	−0.7953
32	16	5	0	16	5	1	32	3	0	16	5	0	128	0.6865	−3.2671
33	32	5	0	32	5	1	64	3	0	32	5	0	32	0.9381	−0.5550
34	8	5	1	64	5	0	32	3	1	32	3	0	128	0.9128	−0.7924
35	16	5	1	16	5	0	64	3	1	64	3	0	32	0.9134	−0.7867
36	32	5	1	32	5	0	128	3	1	16	3	0	64	0.9008	−0.9074

According to the experimental results in Table 8, the greater the factor difference, the stronger the influence of the *S/N* ratio. Therefore, C4_F had the strongest influence on accuracy, whereas C4_P had the weakest influence on accuracy.

Figure 9 depicts the optimal level of each of the factors: C1_F = 32, C1_K = 3, C1_P = 0, C2_F = 64, C2_K = 3, C2_P= 0, C3_F = 64, C3_K = 5, C3_P = 1, C4_F = 32, C4_K = 3, C4_P = 0, and number of fuzzy rules = 32 or 64.

Table 8
Optimal parameter selections from T-VDFNN method.

		Factor						
		C1_F	C1_K	C1_P	C2_F	C2_K	C2_P	C3_F
Level	1	−0.9974	−0.8356	−0.8175	−0.9918	−0.6924	−0.6976	−0.9178
	2	−0.8538	−0.8438	−0.8619	−0.894	−0.9869	−0.9818	−0.7052
	3	−0.6679			−0.6332			−0.8961
	Difference	0.3295	0.0083	0.0444	0.3586	0.2945	0.2843	0.2126
	Ranking	4	12	10	3	5	6	8
	Best level	3	1	1	3	1	1	2
	Optimal parameter combination	32	3	0	64	3	0	64

		Factor						
		C3_K	C3_P	C4_F	C4_K	C4_P	Number of fuzzy rules	C3_K
Level	1	−0.8893	−0.9648	−1.1158	−0.8377	−0.8304	−0.7166	−0.8893
	2	−0.7901	−0.7164	−0.6784	−0.8417	−0.849	−0.7166	−0.7901
	3			−0.7249			−1.0859	
	Difference	0.0991	0.2502	0.4375	0.004	0.0186	0.3693	0.0991
	Ranking	9	7	1	13	11	2	9
	Best level	2	2	2	1	1	1, 2	2
	Optimal parameter combination	5	1	32	3	0	32/64	5

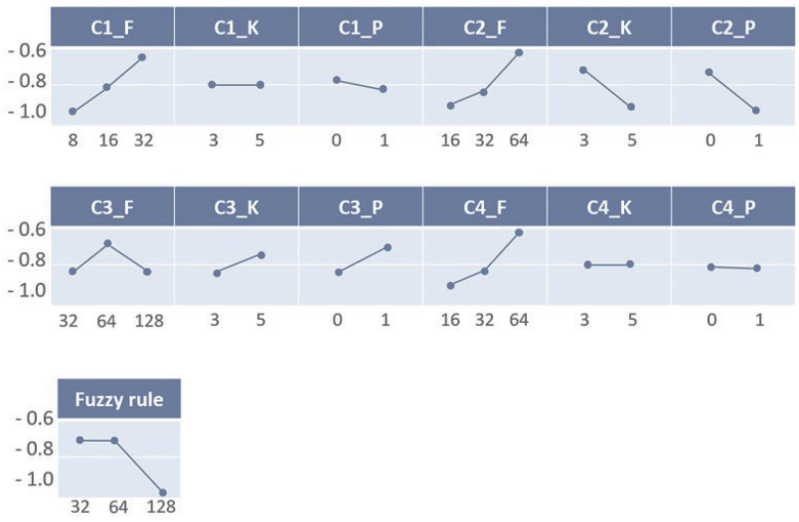


Fig. 9. (Color online) *S/N* ratio of each factor and level.

Table 9 presents the results of an analysis of variance and indicates the degree of influence of each factor. The number of degrees of freedom equals the number of levels minus 1, SS is the sum of squared errors, and F denotes the significance of a factor in the overall experiment. According to Table 9, the contribution of the Con4_Filter (C4_F) factor was the highest at 22%.

The confusion matrix was used to analyze the classification results of the VDFNN and T-VDFNN models. The accuracy, sensitivity, and specificity of the two models were calculated. Three experiments were conducted for each model, and the average was used to assess the model performance. The average accuracy, sensitivity, and specificity of the proposed VDFNN model were 92.18, 93.89, and 88.44%, respectively (Table 10). The average accuracy, sensitivity, and specificity of the proposed T-VDFNN model were 94.37, 96.53, and 89.65%, respectively (Table 11). According to the experimental results, the average accuracy, sensitivity, and specificity of the proposed T-VDFNN model were 2.19, 2.64, and 1.21% higher than those of the VDFNN model, respectively.

Table 9
Analysis of variance results.

Factor	Degrees of freedom	Seq SS	F	Contribution (%)
Con1_Filter	2	0.6551	0.83	10
Con1_Kernel size	1	0.0006	0.00	0
Conv1_Padding	1	0.0177	0.04	0
Conv2_Filter	2	0.8248	1.04	13
Conv2_Kernel size	1	0.7807	1.98	12
Conv2_Padding	1	0.7272	1.84	11
Con3_Filter	2	0.3284	0.42	5
Con3_Kernel size	1	0.0884	0.22	1
Con3_Padding	1	0.5632	1.43	9
Con4_Filter	2	1.3856	1.76	22
Con4_Kernel size	1	0.0001	0.00	0
Con4_Padding	1	0.0031	0.01	0
Number of fuzzy rules	2	1.0911	1.38	17

Table 10
Average accuracy, sensitivity, and specificity of proposed VDFNN model.

Number of experiments	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	93.17	94.11	91.13
2	92.41	94.84	87.10
3	90.96	92.73	87.10
Average	92.18	93.89	88.44

Table 11
Average accuracy, sensitivity, and specificity of proposed T-VDFNN model.

Number of experiments	Accuracy (%)	Sensitivity (%)	Specificity (%)
1	95.13	96.59	91.94
2	93.87	97.05	86.90
3	94.12	95.95	90.12
Average	94.37	96.53	89.65

Table 12
Accuracy of various models.

Model		Fusion method	Accuracy (%)
RF classifier + PFTAS[27]		—	81.82
LeNet−5(Sgdm)[28]		—	80.69
ResHist[29]		—	92.52
Single-layer CNN[30]		—	77.50
Our method	VDFNN	GAP	92.18
	T-VDFNN	GAP	94.37

To verify the effectiveness of the T-VDFNN model, we compared it with other neural network models, such as the textural feature descriptor,[27] LeNet-5,[28] ResHist,[29] and single-layer CNN.[30] The experimental results indicated the superiority of the T-VDFNN model to the other models, as summarized in Table 12.

4. Conclusions

We proposed a VDFNN model to classify breast cancer effectively and automatically from histopathological images. The architecture of the proposed VDFNN model comprises four sets of vector product and pooling layers, a feature fusion layer, and an FNN. In the feature fusion layer, GAP is used to reduce the dimension of the feature information. To simplify and optimize parameter selection for the VDFNN model, we also proposed the T-VDFNN model based on the Taguchi method. The T-VDFNN model achieved breast cancer classification accuracy, sensitivity, and specificity of 94.37, 96.53, and 89.65%, respectively, 2.19, 2.64, and 1.21 percentage points higher than those of the VDFNN model.

Because we used a small volume of medical data in this study, we intend to use a generative adversarial network or Google's AutoAugment method in the future to increase the number of training samples and improve the classification accuracy and stability of our model. To compare the classification and identification performance of deep learning networks, most scholars have used indicators such as accuracy, sensitivity, and specificity. However, these differ with each run of network training, and using a single value or the average value to judge the image classification ability of a network may not be suitable. Therefore, in the future, we will use statistical methods to define a performance index to overcome this problem.

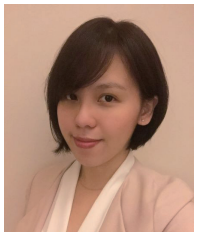
References

1 H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray: CA Cancer J. Clin. **71** (2020) 209. https://doi.org/10.3322/caac.21660
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 E. H. Houssein, M. M. Emam, A. A. Ali, and P. N. Suganthan: Expert Syst. Appl. **167** (2021) 114. https://doi.org/10.1016/j.eswa.2020.114161
4 R. Rubin and D. S. Strayer: Rubin's Pathology: Clinicopathologic Foundations of Medicine (Wolters Kluwer/Lippincott Williams & Wilkins, Philadelphia 2008).
5 H. Yang, J. Y. Kim, H. Kim, and S. P. Adhikari: IEEE Trans. Med. Imaging **39** (2020) 1306. https://doi.org/10.1109/TMI.2019.2948026

6 E. A. Bayrak, P. Kırcı, and T. Ensari: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (2019) 1–3.
7 M. Amrane, S. Oukid, I. Gagaoua, and T. Ensarİ: 2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT, Istanbul, Turkey, 2018). https://doi.org/10.1109/EBBT.2018.8391453
8 R. Hazra, M. Banerjee, and L. Badia: 2020 11th IEEE Annu. Information Technology, Electronics and Mobile Communication Conf. (IEMCON) (2020). https://doi.org/10.1109/IEMCON51383.2020.9284936
9 A. Alqudah and A. M. Alqudah: IETE J. Res. **68** (2019) 59. https://doi.org/10.1080/03772063.2019.1583610
10 Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner: Proc. IEEE **86** (1998) 2278. https://doi.org/10.1109/5.726791
11 G. Ortac and G. Ozcan: Expert Syst. Appl. **182** (2021) 115280. https://doi.org/10.1016/j.eswa.2021.115280
12 C. H. Lin and T. Y. Wang: Signal Process. Image Commun. **97** (2021) 116329. https://doi.org/10.1016/j.image.2021.116329
13 G. Chugh, S. Kumar, and N. Singh: Cogn. Comput. **13** (2021) 1451. https://doi.org/10.1007/s12559-020-09813-6
14 K. P. Korshunova: 2018 3rd Russian-Pacific Conf. Computer Technology and Applications (RPC), Vladivostok, Russia (2018). https://doi.org/10.1109/RPC.2018.8482211
15 M. M. Ebadzadeh and A. Salimi-Badr: Neurocomputing **148** (2015) 430. https://doi.org/10.1016/j.neucom.2014.07.021
16 R. H. Abiyev and A. Helwan: J. Med. Imaging Health Inf. **8** (2018) 817. https://doi.org/10.1166/jmihi.2018.2308
17 A. Rezania, S. A. Atouei, and L. Rosendahl: Mater. Today Energy **16** (2020) 100376. https://doi.org/10.1016/j.mtener.2019.100376
18 J. Ou and Y. Li: Neurocomputing **330** (2019) 253. https://doi.org/10.1016/j.neucom.2018.11.028
19 M. Lagzian, S. E. Razavi, and M. Goharimanesh: Biomed. Signal Process. Control **77** (2022) 103734. https://doi.org/10.1016/j.bspc.2022.103734
20 H. K. Wang, Z. H. Wang, and M. C. Wang: Comput. Ind. Eng. **148** (2020) 106635. https://doi.org/10.1016/j.cie.2020.106635
21 C. J. Lin, S. Y. Jeng, and C. L. Lee: Sens. Mater. **33** (2021) 315. https://doi.org/10.18494/SAM.2021.3015
22 B. Oemar and W. C. Chang: Int. J. Adv. Manuf. Technol. **107** (2020) 4609. https://doi.org/10.1007/s00170-020-05344-4
23 C. J. Lin and Y. C. Li: Electronics **9** (2020) 1066. https://doi.org/10.3390/electronics9071066
24 B. Akbugday:2019 Medical Technologies Congr. (TIPTEKNO) (2019) 1–4. https://doi.org/10.1109/TIPTEKNO.2019.8895222
25 J. Y. Kim and S. B. Cho: Neurocomputing **452** (2021) 395. https://doi.org/10.1016/j.neucom.2019.10.123
26 F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte: IEEE Trans. Biomed. Eng. **63** (2016) 1455. https://ieeexplore.ieee.org/document/7312934
27 M. Sharma, R. Singh, and M. Bhattacharya: 2017 IEEE Int. Conf. Bioinformatics and Biomedicine (BIBM) (IEEE, 2017). https://doi.org/10.1109/BIBM.2017.8217811
28 C. Darken, J. Chang, and J. Moody: Neural Networks for Signal Processing II Proc. 1992 IEEE Workshop (IEEE, 1992). https://doi.org/10.1109/NNSP.1992.253713
29 M. Gour, S. Jain, and T. S. Kumar: Int. J. Imaging Syst. Technol. **30** (2020) 621. https://doi.org/10.1002/ima.22403
30 M. E. Nejad, L. S. Affendey, R. B. Latip, and I. B. Ishak: Proc. Int. Conf. Imaging, Signal Processing and Communication (2017) 50–53. https://doi.org/10.1145/3132300.3132331

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)

Mei-Yu Wu received her M.S. and Ph.D. degrees from the Institute of Information Management, National Chiao Tung University, Taiwan, in 1999 and 2005, respectively. Currently, she is an associate professor of the Department of Business Management, National Taichung University of Science and Technology, Taichung, Taiwan, R.O.C. Her research interests include information security management, role-based access control, intelligent agriculture systems, and artificial intelligence applications and management. (mywu@nutc.edu.tw)



Yi-Hsuan Chuang received her M.S. degree from Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung, Taiwan, in 2022. Her current research interests include deep learning, neural fuzzy systems, and computer vision and applications. (lofo1987@gmail.com)

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.
(merrylee@nutc.edu.tw)