Int. J. Fuzzy Syst. (2019) 21(7):2258–2269
https://doi.org/10.1007/s40815-019-00730-x

A *K*-means Interval Type-2 Fuzzy Neural Network for Medical Diagnosis

Tien-Loc Le[1,2] · Tuan-Tu Huynh[1,2] · Lo-Yi Lin[3] · Chih-Min Lin[1] · Fei Chao[4]

Received: 11 August 2018 / Revised: 26 November 2018 / Accepted: 21 August 2019 / Published online: 3 September 2019

Abstract This paper proposes a new medical diagnosis algorithm that uses a *K*-means interval type-2 fuzzy neural network (KIT2FNN). This KIT2FNN classifier uses a *K*-means clustering algorithm as the pre-classifier and an interval type-2 fuzzy neural network as the main classifier. Initially, the training data are classified into *k* groups using the *K*-means clustering algorithm and these data groups are then used sequentially to train the structure of the *k* classifiers for the interval type-2 fuzzy neural network (IT2FNN). The test data are also initially used to determine to which classifier they are best suited and then they are inputted into the corresponding main classifier for classification. The parameters for the proposed IT2FNN are updated using the steepest descent gradient approach. The Lyapunov theory is also used to verify the convergence and stability of the proposed method. The performance of the system is evaluated using several medical datasets from the University of California at Irvine (UCI). All of the experimental and comparison results are presented to demonstrate the effectiveness of the proposed medical diagnosis algorithm.

Keywords Classification problem · Interval type-2 fuzzy neural network · *K*-means clustering algorithm · Medical diagnosis

✉ Chih-Min Lin
cml@saturn.yzu.edu.tw

Tien-Loc Le
tienloc@lhu.edu.vn

Lo-Yi Lin
zoelin56@hotmail.com

Fei Chao
fchao@xmu.edu.cn

[1] Yuan Ze University, Taoyuan, Taiwan

[2] Department of Electrical Electronic and Mechanical Engineering, Lac Hong University, Bien Hoa, Vietnam

[3] Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan

[4] Department of Cognitive Science, Xiamen University, Xiamen, China

1 Introduction

The *K*-means clustering method was first introduced by Steinhaus in 1956 [1]. It is used to arrange data into specific groups or clusters based on their characteristics, so that data in the same group will have similar characteristics or features. Recently, because large amounts of data can be more easily processed, the *K*-means algorithm has been widely used for data mining, pattern recognition, decision support and machine learning [2–4]. In 2011, Singh et al. noted the evolving limitations in *K*-means algorithm in data mining applications [5]. In 2017, Jiang et al. proposed a method of pattern recognition for acoustic emission signals based on *K*-means clustering [6]. In 2018, Tuncer and Alkan proposed a decision support system for the detection of renal cell cancer [7]. However, *K*-means algorithms have some disadvantages such as their sensitivity to noisy data and outlier, and the number of clusters must be specified in advance [3, 8].

In the past decade, fuzzy logic has been widely used in many fields such as control problems, system identification, prediction, energy and so on. The type-1 fuzzy logic system (T1FLS) was first introduced by Zadeh in 1965 [9]. Zadeh subsequently proposed the concept of type-2 fuzzy logic system (T2FLS) in 1975 [10]. This deals with uncertainties more efficiently than the T1FLS. However,

because of its complexity of computation, the T2FLS was not widely applied in the early stages. In 2000, Liang and Mendel developed the interval type-2 fuzzy logic system (IT2FLS) [11], which only considers the upper and lower grades of the type-2 membership function (T2MF), so it is easier to implement and the computational load is reduced. Recently, the IT2FNN has been successfully applied in various fields [12–16].

The rapid development of technology leads to computer-aided diagnosis (CAD) systems allowing a faster and more accurate medical diagnosis. In 2016, Guan et al. proposed a self-validating cerebellar model neural network for computer-aided diagnosis of breast tumors [17]. In 2018, Zhou et al. used a functional link-based fuzzy brain emotional learning network to classify breast tumors [18]. Many studies also focus on the diagnosis of heart disease and fetal cardiotocography. In 2016, Dwivedi evaluated the performance of different machine learning techniques in the prediction of heart disease [19]. In 2017, Liu et al. proposed a hybrid classification system for the diagnosis of heart disease using the relief and rough set method [20]. In 2016, Yılmaz proposed the artificial neural networks for fetal state assessment from cardiotocography data [21]. In 2017, Mandal presented a clinical healthcare enterprise system using machine learning algorithms [22]. However, these diagnosis methods are complex and the diagnostic performance can be further improved.

This study proposes a new method for medical diagnosis that uses a *K*-means interval type-2 fuzzy neural network (KIT2FNN). This method combines a *K*-means clustering algorithm and an interval type-2 fuzzy neural network (IT2FNN). The *K*-means algorithm classifies the training data into k groups, depending on their characteristics. These groups are then used to train the IT2FNN for the k classifiers. Using the *K*-means algorithm, the test data are again processed to be classified to the most pertinent IT2FNN classifiers, and then they are inputted into the corresponding classifier for classification. Using this method of 2-layer classification, the performance of the proposed classifier is comparable to that for some state-of-the-art methods. To demonstrate the effectiveness of the proposed diagnostic KIT2FNN, numerical experiments are conducted using real-world medical datasets. The data using in this study include the breast cancer, the heart disease and the fetal cardiotocography, which are provided by the University of California at Irvine (UCI).

The comparisons show that the results of this study are more accurate than those of previous studies [17, 19–23], and the method is easily applied to medical diagnosis. The major contributions of this study are (i) the design of the *K*-means IT2FNN classifier for medical diagnosis; (ii) the design of the adaptive laws for updating the network parameters to guarantee stable convergence for the system; and (iii) the numerical experiments using the medical datasets illustrate the effectiveness of the proposed approach.

The remainder of this paper is organized as follows. Section 2 presents the structure of *K*-means IT2FNN classifier. The algorithm for learning the parameters is illustrated in Sect. 3. The numerical experiments using medical datasets are presented in Sect. 4. Finally, conclusions are given in Sect. 5.

2 *K*-means Interval Type-2 Fuzzy Neural Network

Figures 1 and 2 respectively show the training and testing processes for the proposed *K*-means IT2FNN classifier, which comprises a *K*-means pre-classifier and an IT2FNN as the main classifier.

2.1 The *K*-means Clustering Algorithm

A *K*-means clustering algorithm separates data into different clusters so that the data in each cluster have the similar characteristics. When the data and the expected number of clusters are inputted, the center of each cluster is determined and each data point is assigned to a specific cluster. The minimum distance from each point to all center points is

$$J = \sum_{j=1}^{n_j} \sum_{i}^{n_i} \left\| x_i^j - c_j \right\|^2 \quad i = 1, 2, \ldots, n_i \quad j = 1, 2, \ldots, n_j, \tag{1}$$

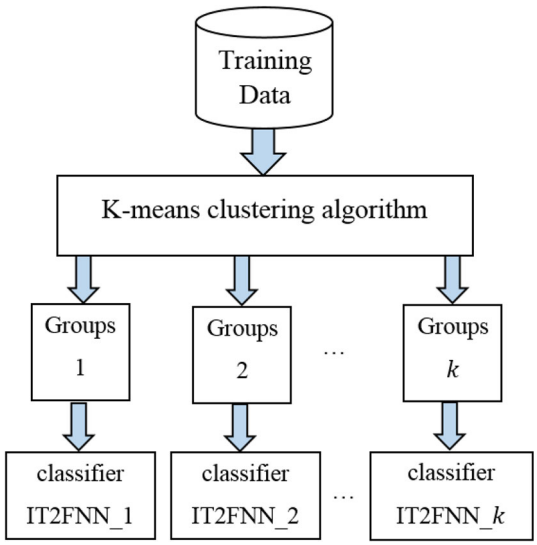

Fig. 1 The training process operation of the *K*-means IT2FNN classifier

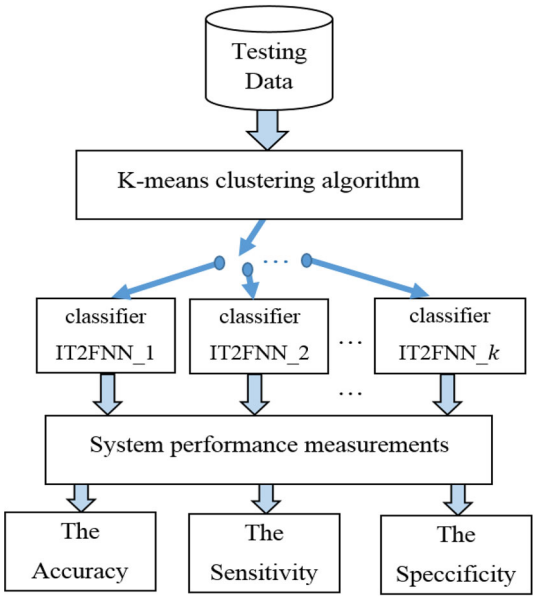

Fig. 2 The testing process operation of the *K*-means FIT2FNN classifier

where $\left\|x_i^j - c_j\right\|^2$ is the algebraic distance between a data point position x_i^j and the cluster center c_j. Figure 3 shows the flowchart for the *K*-means clustering algorithm, which has the following steps:

Step 1: Initialize the k center points.
Step 2: Calculate the distance between each point and the k center points.
Step 3: Assign each point to a cluster (based on the distance).
Step 4: Stop if the assignment is not changed.
Step 5: Update the new position for the k center points (average of each cluster).
Step 6: Go to step 2.

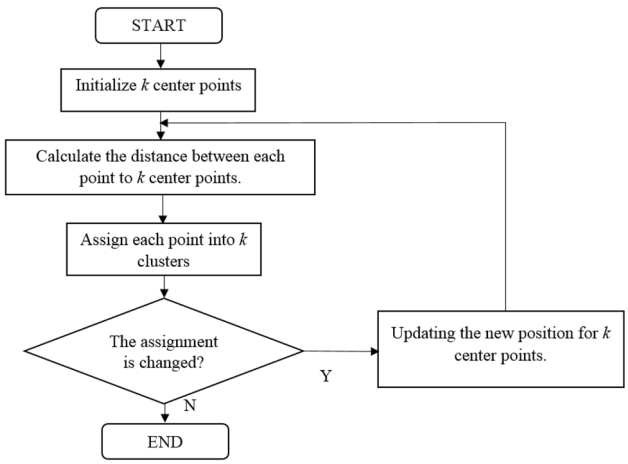

Fig. 3 The *K*-means clustering flowchart

2.2 The Interval Type-2 Fuzzy Neural Network

Figure 4 shows a block diagram of the proposed IT2FNN classification system, for which the adaptive parameter are derived using the gradient descent method and the error feedback system. Figure 5 shows the structure of the IT2FNN.

The IT2FNN realizes the fuzzy inference rules as follows:

Rule j: IF I_1 is $\tilde{\mu}_{1j}$ and... and I_i is $\tilde{\mu}_{ij}$ and... and I_n is $\tilde{\mu}_{nj}$

$$\text{THEN} \quad o_j = \tilde{w}_{jk} \tag{2}$$

where $i = 1, \ldots, n_i$; $j = 1, \ldots, n_j$ and $k = 1, \ldots, n_k$ are the indices for the ith fuzzy input, the jth rule and the kth output. $\tilde{\mu}_{ij}$ and $\tilde{w}_{jk}$ respectively denote the type-2 fuzzy membership function for the input and the output weights.

Figure 5 shows that the structure of the IT2FNN has six layers. The detailed mathematical functions for each layer are described as

Layer 1 (input layer) In this layer, a crisp input variable is fed into the membership function layer. Note that these are directly transferred into the next layer without any computation.

Layer 2 (membership function layer) In this layer, the input is fed into the type-2 Gaussian membership function $\tilde{\mu}_{ij} = \left[\underline{\mu}_{ij} \quad \bar{\mu}_{ij}\right]$, which has a fixed mean and an uncertain standard deviation $\tilde{\sigma}_{ij} = \left[\underline{\sigma}_{ij} \quad \bar{\sigma}_{ij}\right]$. The upper and lower membership functions are

$$\bar{\mu}_{ij} = \exp\left\{-\frac{1}{2}\left(\frac{I_i - m_{ij}}{\bar{\sigma}_{ij}}\right)^2\right\} \tag{3}$$

$$\underline{\mu}_{ij} = \exp\left\{-\frac{1}{2}\left(\frac{I_i - m_{ij}}{\underline{\sigma}_{ij}}\right)^2\right\}, \tag{4}$$

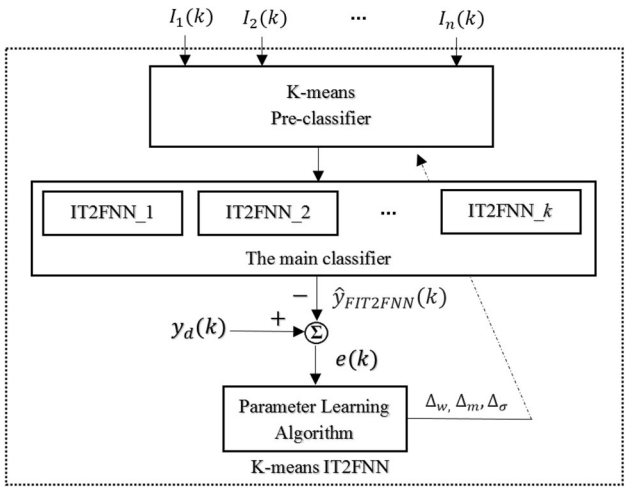

Fig. 4 Classification scheme using *K*-means IT2FNN

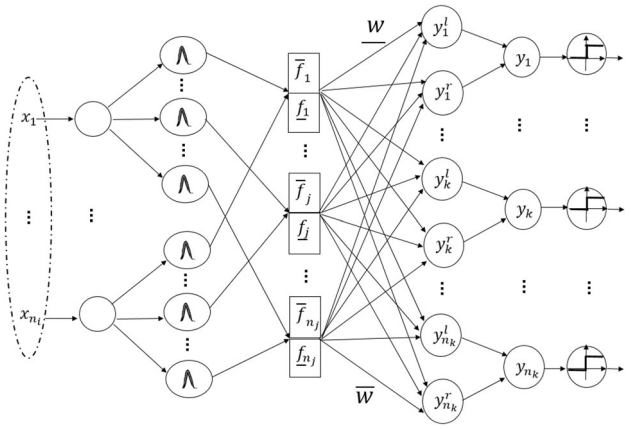

Fig. 5 Architecture of the IT2FNN classifier

where $\bar{\sigma}_{ij}$, $\underline{\sigma}_{ij}$ and m_{ij} are the upper standard deviation, the lower standard deviation and the mean of the type-2 Gaussian membership function, respectively.

Layer 3 (firing layer) This layer performs a fuzzy meet operation for all elements in the same block j. The firing of the jth rule $F_j = \left[\underline{f}_j, \bar{f}_j\right]$ can be computed by

$$\bar{f}_j = \prod_{i=1}^{n_i} \bar{\mu}_{ij} \tag{5}$$

$$\underline{f}_j = \prod_{i=1}^{n_i} \underline{\mu}_{ij}. \tag{6}$$

Layer 4 (output weight layer) This layer performs the consequent part of the IT2FNN, which is known as the weights of the network. The initial values of the weights are randomly established and are then updated by parameter learning. The interval value for the weight is $\tilde{w}_{jk} = \left[\underline{w}_{jk}\ \bar{w}_{jk}\right]$, where $\bar{w}_{jk}$ is the upper value and $\underline{w}_{jk}$ is the lower value.

Layer 5 (pre-output layer) The output of this layer is the combination of the outputs of layers 3 and 4 using a center-of-gravity algorithm. This is also known as the defuzzification operation. The Karnik–Mendel (KM) algorithm in [24] is used to adjust the contribution of the firing strength. The defuzzification output is

$$y_k^l = \frac{\sum_{j=1}^{n_j} f_j^l \underline{w}_{jk}}{\sum_{j=1}^{n_j} f_j^l} \tag{7}$$

$$y_k^r = \frac{\sum_{j=1}^{n_j} f_j^r \bar{w}_{jk}}{\sum_{j=1}^{n_j} f_j^r}. \tag{8}$$

The KM algorithm is applied to find the switch points L and R, so f_j^l and f_j^r are obtained as

$$f_j^l = \begin{cases} \bar{f}_j, j \le L \\ \underline{f}_j, j > L \end{cases} \tag{9}$$

$$f_j^r = \begin{cases} \underline{f}_j, j \le R \\ \bar{f}_j, j > R \end{cases}. \tag{10}$$

Layer 6 (output layer) The final output of the IT2FNN is obtained by applying an average operation to the interval value $\left[y_k^l,\ y_k^r\right]$ in the previous layer

$$y_{\text{IT2FNN}}^k = hardlim\left(\frac{y_k^l + y_k^r}{2}\right). \tag{11}$$

Because the prediction results have only two states, disease or not a disease, the hard-limit transfer function is used at the final output of the IT2FNN to describe the state of the samples. All of the initial parameters for the IT2FNN are randomly assigned. The adaptive laws for updating network parameters are introduced in the following section.

3 The Parameter Learning Algorithm for IT2FNN

The classification problem involves minimizing the feedback error between the desired output y_d and the final output of the K-means IT2FNN y_{IT2FNN}. As shown in Fig. 4, the structure of IT2FNN is trained using the prior dataset, which includes the sets of input that correspond to the desired output. When the training process is complete, the performance of the proposed classifier is evaluated using a test dataset.

The Lyapunov cost function is

$$E(k) = \frac{1}{2}\boldsymbol{e}^2(k), \tag{12}$$

where $e(k)$ is the output error, which is obtained by

$$\boldsymbol{e}(k) = \boldsymbol{y}_d(k) - \boldsymbol{y}_{\text{IT2FNN}}(k). \tag{13}$$

The gradient descent approach is then used to minimize the change in the Lyapunov cost function

$$\dot{E}(k) = \boldsymbol{e}(k)\dot{\boldsymbol{e}}(k). \tag{14}$$

Using the gradient descent method and the chain rule, the online tuning laws for the parameters $\hat{\underline{w}}_{jk}, \hat{\bar{w}}_{jk}, \hat{m}_{ij}, \hat{\underline{\sigma}}_{ij}$, $\hat{\bar{\sigma}}_{ij}$ are described as

$$\begin{aligned} \hat{\underline{w}}_{jk}(k+1) &= \hat{\underline{w}}_{jk}(t) - \hat{\eta}_w \frac{\partial \dot{E}(k)}{\partial \hat{\underline{w}}_{jk}} \\ &= \hat{\underline{w}}_{jk}(k) + \frac{1}{2}\hat{\eta}_w \boldsymbol{e}(k) \frac{f_j^l}{\sum_{j=1}^{n_j} f_j^l} \end{aligned} \tag{15}$$

$$\hat{\bar{w}}_{jk}(k+1) = \hat{\bar{w}}_{jk}(k) - \hat{\eta}_w \frac{\partial \dot{E}(k)}{\partial \hat{\bar{w}}_{jk}} = \hat{\bar{w}}_{jk}(k) + \frac{1}{2}\hat{\eta}_w \boldsymbol{e}(k) \frac{f_j^r}{\sum_{j=1}^{n_j} f_j^r} \tag{16}$$

$$\hat{m}_{ij}(k+1) = \hat{m}_j^i(k) - \hat{\eta}_m \frac{\partial \dot{E}(k)}{\partial \hat{m}_{ij}} = \hat{m}_j^i(k) + \frac{1}{2}\hat{\eta}_m \boldsymbol{e}(k) \left(\frac{\left(\underline{w}_{jk} - y_k^l\right)}{\sum_{j=1}^{n_j} f_j^l} \frac{\partial f_j^l}{\partial \hat{m}_{ij}} + \frac{\left(\bar{w}_{jk} - y_k^r\right)}{\sum_{j=1}^{n_j} f_j^r} \frac{\partial f_j^r}{\partial \hat{m}_{ij}} \right) \tag{17}$$

$$\hat{\bar{\sigma}}_{ij}(t+1) = \hat{\bar{\sigma}}_{ij}(t) - \hat{\eta}_\sigma \frac{\partial \dot{E}(k)}{\partial \hat{\bar{\sigma}}_{ij}} = \hat{\bar{\sigma}}_{ij}(t) + \frac{1}{2}\hat{\eta}_\sigma \boldsymbol{e}(t) \left(\frac{\left(\underline{w}_{jk} - y_k^l\right)}{\sum_{j=1}^{n_j} f_j^l} \frac{\partial f_j^l}{\partial \hat{\bar{\sigma}}_{ij}} + \frac{\left(\bar{w}_{jk} - y_k^r\right)}{\sum_{j=1}^{n_j} f_j^r} \frac{\partial f_j^r}{\partial \hat{\bar{\sigma}}_{ij}} \right) \tag{18}$$

$$\hat{\underline{\sigma}}_{ij}(t+1) = \hat{\underline{\sigma}}_{ij}(t) - \hat{\eta}_\sigma \frac{\partial \dot{E}(k)}{\partial \hat{\underline{\sigma}}_{ij}} = \hat{\underline{\sigma}}_{ij}(t) + \frac{1}{2}\hat{\eta}_\sigma \boldsymbol{e}(t) \left(\frac{\left(\underline{w}_{jk} - y_k^l\right)}{\sum_{j=1}^{n_j} f_j^l} \frac{\partial f_j^l}{\partial \hat{\underline{\sigma}}_{ij}} + \frac{\left(\bar{w}_{jk} - y_k^r\right)}{\sum_{j=1}^{n_j} f_j^r} \frac{\partial f_j^r}{\partial \hat{\underline{\sigma}}_{ij}} \right). \tag{19}$$

where $\hat{\eta}_q, \hat{\eta}_m$, *and* $\hat{\eta}_\sigma$ are the learning rates for the parameter adaptive laws. When the KM algorithm is used, the terms f_j^l and f_j^r in (15)–(19) are re-expressed as $\underline{f}_j$ or $\bar{f}_j$.

$$\frac{\partial \underline{f}_j}{\partial \hat{m}_{ij}} = \frac{\partial \underline{f}_j}{\partial \underline{\mu}_j^i} \frac{\partial \underline{\mu}_j^i}{\partial \hat{m}_{ij}} = \underline{f}_j \frac{x_j - \hat{m}_{ij}}{\left(\hat{\underline{\sigma}}_{ij}\right)^2}; \quad \frac{\partial \bar{f}_j}{\partial \hat{m}_{ij}} = \frac{\partial \bar{f}_j}{\partial \bar{\mu}_j^i} \frac{\partial \bar{\mu}_j^i}{\partial \hat{m}_{ij}} = \bar{f}_j \frac{x_j - \hat{m}_{ij}}{\left(\hat{\bar{\sigma}}_{ij}\right)^2} \tag{20}$$

$$\frac{\partial \underline{f}_j}{\partial \hat{\bar{\sigma}}_{ij}} = \frac{\partial \underline{f}_j}{\partial \underline{\mu}_j^i} \frac{\partial \underline{\mu}_j^i}{\partial \hat{\bar{\sigma}}_{ij}} = \underline{f}_j \frac{\left(x_j - \hat{m}_{ij}\right)^2}{\left(\hat{\bar{\sigma}}_{ij}\right)^3}; \quad \frac{\partial \bar{f}_j}{\partial \hat{\bar{\sigma}}_{ij}} = \frac{\partial \bar{f}_j}{\partial \bar{\mu}_j^i} \frac{\partial \bar{\mu}_j^i}{\partial \hat{\bar{\sigma}}_{ij}} = \bar{f}_j \frac{\left(x_j - \hat{m}_{ij}\right)^2}{\left(\hat{\bar{\sigma}}_{ij}\right)^3} \tag{21}$$

$$\frac{\partial \underline{f}_j}{\partial \hat{\underline{\sigma}}_{ij}} = \frac{\partial \underline{f}_j}{\partial \underline{\mu}_j^i} \frac{\partial \underline{\mu}_j^i}{\partial \hat{\underline{\sigma}}_{ij}} = \underline{f}_j \frac{\left(x_j - \hat{m}_{ij}\right)^2}{\left(\hat{\underline{\sigma}}_{ij}\right)^3}; \quad \frac{\partial \bar{f}_j}{\partial \hat{\underline{\sigma}}_{ij}} = \frac{\partial \bar{f}_j}{\partial \bar{\mu}_j^i} \frac{\partial \bar{\mu}_j^i}{\partial \hat{\underline{\sigma}}_{ij}} = \bar{f}_j \frac{\left(x_j - \hat{m}_{ij}\right)^2}{\left(\hat{\underline{\sigma}}_{ij}\right)^3} \tag{22}$$

Using these adaptive laws, the *K*-means IT2FNN gives the desired performance.

Proof of convergence

$$V(k) = \boldsymbol{E}(k) = \frac{1}{2}\boldsymbol{e}^2(k). \tag{23}$$

Therefore,

$$\Delta V(k) = V(k+1) - V(k) = \frac{1}{2}\left[\boldsymbol{e}^2(k+1) - \boldsymbol{e}^2(k)\right]. \tag{24}$$

Applying the Taylor expansion and linearization techniques gives

$$\boldsymbol{e}(k+1) = \boldsymbol{e}(k) + \Delta \boldsymbol{e}(k) \cong \boldsymbol{e}(k) + \left[\frac{\partial \boldsymbol{e}(k)}{\partial \hat{\underline{w}}_j}\right] \Delta \hat{\underline{w}}_j. \tag{25}$$

From (15),

$$\frac{\partial \boldsymbol{e}(k)}{\partial \hat{\underline{w}}_j} = -\frac{1}{2}\frac{f_j^l}{\sum_{j=1}^{n_j} f_j^l} = \boldsymbol{\psi}. \tag{26}$$

Using (26) and (15), (25) can be rewritten as

$$\boldsymbol{e}(k+1) = \boldsymbol{e}(k) - \boldsymbol{\psi}(\hat{\eta}_w \boldsymbol{e}(k)\boldsymbol{\psi}) = \boldsymbol{e}(k)\left[1 - \hat{\eta}_w \boldsymbol{\psi}^2\right]. \tag{27}$$

From (27), (24) becomes

$$\Delta V(k) = \frac{1}{2}\boldsymbol{e}^2(k)\left[\left(1 - \eta_w \boldsymbol{\psi}^2\right)^2 - 1\right] = \frac{1}{2}\boldsymbol{e}^2(k)\left[\left(\hat{\eta}_w \boldsymbol{\psi}^2\right)^2 - 2\hat{\eta}_w \boldsymbol{\psi}^2\right] = \frac{1}{2}\hat{\eta}_w \boldsymbol{e}^2(k)\boldsymbol{\psi}^2\left(\hat{\eta}_w \boldsymbol{\psi}^2 - 2\right). \tag{28}$$

This result shows that if the learning rates are chosen such that $0 < \hat{\eta}_w < \frac{2}{\boldsymbol{\psi}^2}$, then $\Delta V(k) < 0$, so the stability of the system is guaranteed. The values for $\hat{\eta}_m$ *and* $\hat{\eta}_\sigma$ can be proved using a similar method.

4 Experimental Results and Discussion

To evaluate the system performance, the accuracy (ACC), the sensitivity (SEN) and the specificity (SPE) of the medical diagnosis system are measured. To allow a fair comparison and to allow the use of average values, the experiments are conducted 100 times using random samples for the training and testing processes.

The evaluation indices are defined as

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TN} + \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \times 100\% \tag{29}$$

$$\mathrm{SEN} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \times 100\% \tag{30}$$

$$\mathrm{SPE} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \times 100\%, \tag{31}$$

where TP: (True Positives) the patient has the disease and the outcome of the prediction is correct; TN: (True Negatives) the patient does not have the disease and the outcome of the prediction is correct; FP: (False Positives) the patient

has the disease but the outcome of the prediction is not a disease; FN: (False Negatives) the patient does not have the disease but the outcome of the prediction is a disease.

4.1 Wisconsin Breast Cancer Dataset

The effectiveness of the *K*-means IT2FNN classifier is tested using the Wisconsin breast cancer dataset (WBCD), which is provided by the University of California at Irvine (UCI). This dataset includes breast cancer samples for 683 patients, each of which has 9 characteristics and 1 output target. The output target is categorized into two classes: benign cases and malignant cases. Table 1 shows the details of the characteristics. The respective ratios for the training and testing processes are 70% and 30%. Using the *K*-means algorithm, the training data are divided into k groups according to the characteristics of each data point. The k IT2FNN classifier network is then trained using the corresponding data. The test data are then used to evaluate the performance of the system. The test data are also assigned into k groups using the *K*-means pre-classifier before feed into the IT2FNN main classifiers. However, determining a suitable value for k is very important because this value significantly affects the system performance. In this study, each experiment is conducted 100 times for each $k = 1, 2, \ldots 10$.

The experimental results for different values of k are shown in Table 2, in which it is seen that the proposed method gives better results for all indices with $k = 6$. The accuracy during 2000 iterations of the training process is shown in Fig. 6 (for $k = 6$). Table 3 compares the accuracy of the proposed method with other methods [17, 18, 25–28] for the Wisconsin breast cancer dataset. It shows that the proposed algorithm gives the most accurate classification. Table 2 shows that the k value for the *K*-means clustering has a significant effect on the system performance. When k is large, the training data are divided into a greater number of groups, so the trained network does not have

Table 2 Experimental results of the *K*-means IT2FNN classifier for breast cancer

k	Avg. accuracy (%)		Avg. sensitivity (%)		Avg. specificity (%)	
	Training	Testing	Training	Testing	Training	Testing
1	100	96.21	100	96.63	100	95.44
2	100	97.61	100	98.01	100	97.05
3	100	96.72	100	97.36	100	95.79
4	100	96.91	100	97.25	100	96.34
5	100	96.97	100	96.93	100	97.12
6	100	98.07	100	98.13	100	97.98
7	100	97.14	100	97.28	100	96.96
8	100	97.52	100	97.91	100	96.92
9	100	97.36	100	98.83	100	95.14
10	100	97.73	100	98.36	100	96.75

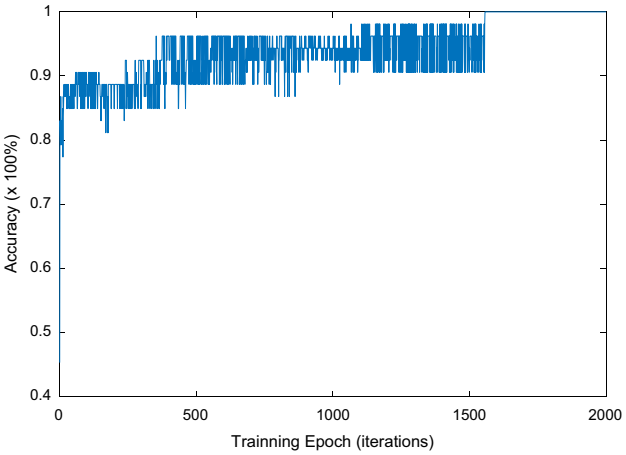

Fig. 6 The accuracy during training process of *K*-means IT2FNN for the Wisconsin breast cancer dataset (with $k = 6$)

Table 1 Wisconsin breast cancer dataset

No	Attribute	Description	Domain
1	Clump thickness	Measurement of thickness of clustered mass tissues	1–10
2	Uniformity of cell size	Degree of consistent cell size	1–10
3	Uniformity of cell shape	Having one form of shape	1–10
4	Marginal adhesion	The stable joining of parts to one another, which may occur abnormally	1–10
5	Single epithelial cell size	Number of layers present in epithelium	1–10
6	Bare nuclei	Having sufficient nucleus	1–10
7	Bland chromatin	Unperturbed genetic	1–10
8	Normal nuclei	Normal round granular body composed of protein and RNA in the nucleus of a cell	1–10
9	Mitoses	The entire process of cell division of the nucleus and the cytoplasm	1–10

Table 3 The comparison results of WBC dataset

Year	Author	Method	Accuracy (%)
2013	Stoean and Stoean [25]	SVM and evolutionary algorithm	97.23
2014	Zheng et al. [26]	*K*-means and SVM	97.38
2015	Lim and Chan [27]	BK with IVFS	95.26
2016	Guan et al. [17]	SVCMAC	96.50
2017	Khan et al. [28]	SVM	97.36
2018	Zhou et al. [18]	FL-FBELN	96.17
	Our method	*K*-means IT2FNN (k = 6)	98.07

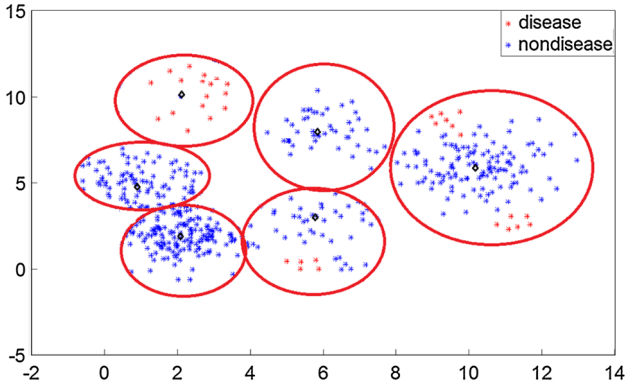

Fig. 7 A two-dimensional illustrative example, when training data are split into more groups

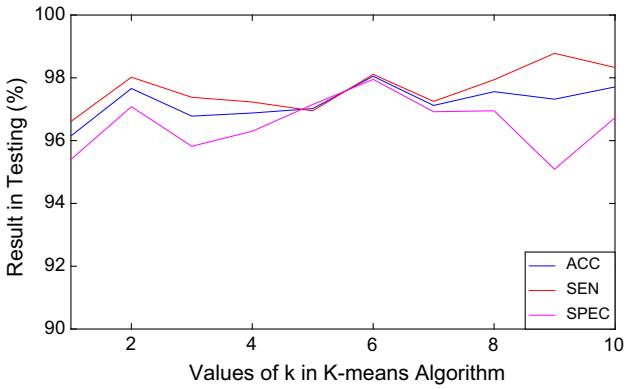

Fig. 8 The testing results of *K*-means IT2FNN for the Wisconsin breast cancer dataset

sufficient characteristics to recognize the samples in the test data. Figure 7 shows that if *K*-means are used to divide data into greater number of groups, some groups include all of the disease samples and some groups include all of the non-disease samples, so it is easy to train the network. However, when the test data for this group include both states, the trained network can give a false diagnosis. Figure 8 shows the change in the accuracy, sensitivity and specificity of the *K*-means IT2FNN for the breast cancer dataset; which corresponds to Table 2; thus, we choose $k = 6$ for this dataset.

4.2 The Heart Disease Dataset

The effectiveness of the *K*-means IT2FNN classifier was also tested using the UCI Statlog (heart) disease dataset. This dataset contains 270 sets of samples, each of which has 13 characteristics and 1 output target. The details of all characteristics are shown in Table 4. The output has two statuses, that are represented 0, which corresponds to the absence of heart disease, and 1, which corresponds to the presence of heart disease.

First, the dataset is separated into a training set and a test set in the ratio of 70% to 30%, respectively. The training data are then divided into k groups, based on the characteristics of the data point. To determine a suitable value for k, the experiment is conducted 100 times with $k = 1, 2, \ldots 10$. A comparison of the proposed method using different values of k is shown in Table 5. It is seen that the proposed method gives the best performance for $k = 4$. When the values of k are greater than 5, the training is 100% accurate. However, the accuracy of the test is decreased, because when k is larger, the amount of training data in each classifier network is decreased, so it is easy to achieve a high training accuracy. However, a small amount of training data means that the trained classifier network does not have sufficient characteristics to recognize the samples for the test data. The accuracy of the training process during 2500 iterations is shown in Fig. 9 (for $k = 4$). Figure 10 shows the change in the accuracy, sensitivity and specificity of the *K*-means IT2FNN for the heart disease dataset. Table 6 compares the accuracy of the proposed method with that for other methods using the UCI Starlog (heart) disease dataset, which shows the superiority of the proposed method.

4.3 The Fetal Cardiotocography Dataset

The effectiveness of the *K*-means IT2FNN classifier is also tested using the fetal cardiotocography dataset, which is

Table 4 UCI Starlog (Heart) disease dataset

No	Attribute	Description	Domain
1	Age		29–77
2	Sex	Male, female	0, 1
3	Chest pain type	Angina, asymptomatic, abnormal	1, 2, 3, 4
4	Resting blood pressure		94–200
5	Serum cholesterol in mg/dl		126–564
6	Fasting blood sugar > 120 mg/dl		0, 1
7	Resting electrocardiographic results	Norm, abnormal, hyper	0, 1, 2
8	Maximum heart rate achieved		71–202
9	Exercise-induced angina		0, 1
10	Old peak = ST depression induced by exercise relative to rest		0–6,2
11	Slope of the peak exercise ST segment	Up, flat, down	1, 2, 3
12	Number of major vessels (0–3) colored by fluoroscopy		0, 1, 2, 3
13	Thal	Normal, fixed defect, reversible defect	3, 6, 7

Table 5 Experimental results of the *K*-means IT2FNN classifier for heart disease

k	Avg. accuracy (%)		Avg. sensitivity (%)		Avg. specificity (%)	
	Training	Testing	Training	Testing	Training	Testing
1	85.27	83.92	83.63	89.24	86.57	81.09
2	96.78	88.86	96.45	90.61	97.08	87.71
3	98.91	91.33	98.87	91.16	99.03	91.46
4	99.97	93.81	99.96	94.08	99.93	93.58
5	97.83	92.56	97.61	93.91	98.06	91.64
6	94.16	91.34	95.13	93.74	93.45	89.76
7	100	86.39	100	89.97	100	84.28
8	100	85.24	100	82.87	100	87.04
9	100	86.47	100	87.56	100	85.76
10	100	87.62	100	85.68	100	89.07

provided by UCI. This dataset includes 2126 sets of samples, each of which has 21 characteristics and 1 output target. The output target has 3 statuses of fetal state, that are represented 1, which corresponds to the normal case; 2, represents to the suspect case; and 3, substitutes to the pathologic case. The details of all characteristics and output targets are shown in Table 7. To evaluate the system performance for 3-class dataset, the confusion matrix and the evaluation indices are defined as in Tables 8 and 9. In which, the normal case, suspect case and pathologic case are respectively represented by class A, class B, and class C.

The respective ratios for the training and testing processes are 70% and 30%. The training data are then divided into *k* groups, based on the characteristics of the data point. To determine a suitable value for *k*, the experiment is

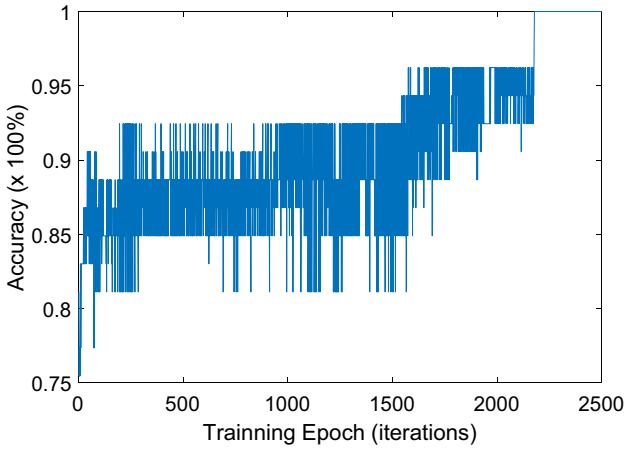

Fig. 9 The accuracy during training process of *K*-means IT2FNN for the Starlog heart disease dataset (with $k = 4$)

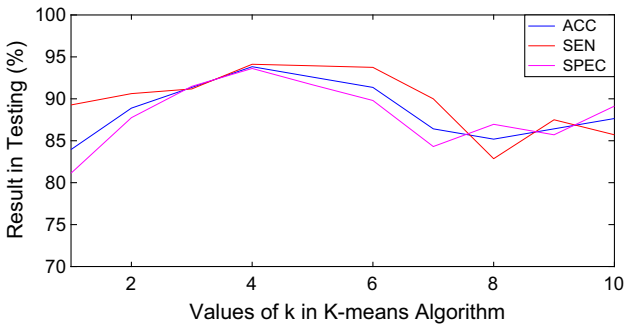

Fig. 10 The testing results of *K*-means IT2FNN for the Starlog heart disease dataset

Table 6 The comparison result of heart disease

Year	Author	Method	Accuracy (%)
2013	Buscema et al. [29]	TWIST algorithm	84.14
2014	Tomar and Agarwal [30]	Feature selection-based LSTSVM	85.59
2015	Lee [31]	The center of gravity of BSWFMs using NEWFM	87.40
2016	Dwivedi [19]	Logistic regression	85.00
2017	Liu et al. [20]	RFRS classification system	92.59
	Our method	*K*-mean IT2FNN ($k = 4$)	93.81

Table 7 The fetal cardiotocography dataset

No	Attribute	Description	Domain
1	LB	Baseline value (SisPorto)	120–160
2	AC	Accelerations (SisPorto)	0–0.02
3	FM	Fetal movement (SisPorto)	0–0.481
4	UC	Uterine contractions (SisPorto)	0–0.015
5	ASTV	Percentage of time with abnormal short-term variability (SisPorto)	0–0.015
6	mSTV	Mean value of short-term variability (SisPorto)	0, 0.001
7	ALTV	Percentage of time with abnormal long-term variability (SisPorto)	0, 0.001
8	mLTV	Mean value of long-term variability (SisPorto)	0–0.005
9	DL	Light decelerations	73–87
10	DS	Severe decelerations	0.4–7
11	DP	Prolonged decelerations	36–91
12	Width	Histogram width	42–180
13	Min	Low freq. of the histogram	62–159
14	Max	High freq. of the histogram	126–238
15	Nmax	Number of histogram peaks	2–18
16	Nzeros	Number of histogram zeros	0–10
17	Mode	Histogram mode	120–187
18	Mean	Histogram mean	137–182
19	Median	Histogram median	121–186
20	Variance	Histogram variance	1–269
21	Tendency	Histogram tendency	− 1, 0, 1
22	NSP	Normal = 1; suspect = 2; pathologic = 3	1–3

conducted 100 times with $k = 1, 2, \ldots 10$. The experimental results for different values of k are shown in Table 10, in which it is seen that the proposed method gives better results for accuracy with $k = 5$. The accuracy of the training process during 3500 iterations is shown in Fig. 11 (for $k = 5$). Figure 12 shows the change in the accuracy, sensitivity and specificity of the *K*-means IT2FNN for the fetal cardiotocography dataset; which corresponds to Table 10. The comparison results of the proposed method and the other methods using the fetal cardiotocography dataset are shown in Table 11, which also shows the superiority of the proposed method.

For all experiments involving breast cancer, heart disease and fetal cardiotocography, the proposed method gives results that are superior to other methods. In Tables 2, 5 and 10, if $k = 1$, then the *K*-means algorithm is not applied. It is seen that when the *K*-means algorithm is

Table 8 Confusion matrix for 3-class dataset

Predicted	True status		
	A	B	C
A	TP_A	E_BA	E_CA
B	E_AB	TP_B	E_CB
C	E_AC	E_BC	TP_C

where TP_A: the true samples in class A and the outcome of the prediction is correct; TP_B: the true samples in class B and the outcome of the prediction is correct; TP_C: the true samples in class C and the outcome of the prediction is correct; E_AB: the true samples in class A and the outcome of the prediction is class B; E_AC: the true samples in class A and the outcome of the prediction is class C; E_BA: the true samples in class B and the outcome of the prediction is class A; E_BC: the true samples in class B and the outcome of the prediction is class C; E_CA: the true samples in class C and the outcome of the prediction is class A; E_CB: the true samples in class C and the outcome of the prediction is class B

Table 9 Evaluation indices for 3-class dataset

Measure	Formula
Sensitivity_A	SEN_A = TP_A/(TP_A + FN_A)
Sensitivity_B	SEN_B = TP_B/(TP_B + FN_B)
Sensitivity_C	SEN_C = TP_C/(TP_C + FN_C)
SPE_A	SPE_A = TN_A/(TN_A + FP_A)
SPE_B	SPE_B = TN_B/(TN_B + FP_B)
SPE_C	SPE_C = TN_C/(TN_C + FP_C)
ACC	ACC = (TPA + TPB + TPC)/(TPA + TPB + TPC + E_AB + E_AC + E_BA + E_BC + E_CA + E_CB)
SEN	SEN = (SEN_A + SEN_B + SEN_C)/3
SPE	SPE = (SPE_A + SPE_B + SPE_C)/3

where FN_A = E_AB + E_AC; FN_B = E_BA + E_BC; FN_C = E_CA + E_CB; FP_A = E_BA + E_CA; FP_B = E_AB + E_CB; FP_C = E_AC + E_BC; TN_A = TP_B + TP_C + E_CB + E_BC; TN_B = E_AB + E_CB + TP_C + E_AC; TN_C = E_AB + E_BC + TP_B + E_AC

Table 10 Experimental results of the *K*-means IT2FNN classifier for fetal cardiotocography

k	Avg. accuracy (%)		Avg. sensitivity (%)		Avg. specificity (%)	
	Training	Testing	Training	Testing	Training	Testing
1	90.92	90.91	84.74	84.94	84.74	82.80
2	91.70	91.75	85.99	86.72	83.70	83.75
3	92.13	92.14	86.93	87.37	84.35	84.18
4	92.65	91.10	87.70	86.15	84.96	83.05
5	92.85	92.84	88.56	87.71	85.43	85.26
6	93.26	92.78	89.10	88.02	86.19	85.07
7	92.97	91.85	88.61	85.44	85.39	83.99
8	93.01	92.43	88.84	87.36	85.63	84.86
9	93.29	92.28	89.28	86.97	86.07	84.46
10	93.74	92.17	89.86	87.60	86.64	84.70

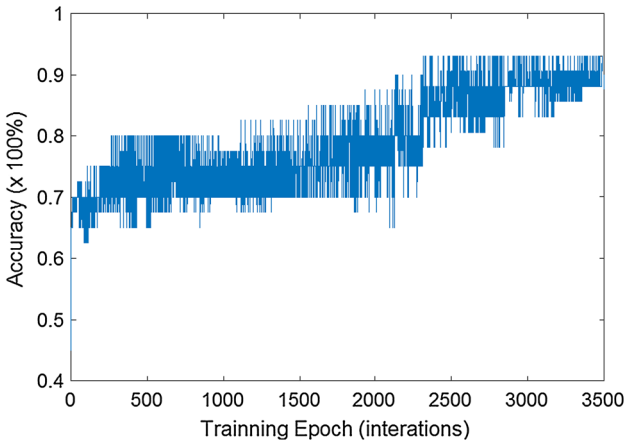

Fig. 11 The accuracy during training process of *K*-means IT2FNN for the fetal cardiotocography dataset (with $k = 5$)

used initially, the classification performance is significantly improved.

5 Conclusion

This study proposes a medical diagnosis system that uses *K*-means clustering and an IT2FNN classifier. The main contribution of this work is the design of a classifier structure that used *K*-means clustering as the pre-classifier and an IT2FNN as the main classifier. This system gives highly accurate results. The method is shown to be superior to some state-of-the-art methods. The data used in the experiment come from the UCI disease dataset. As well as

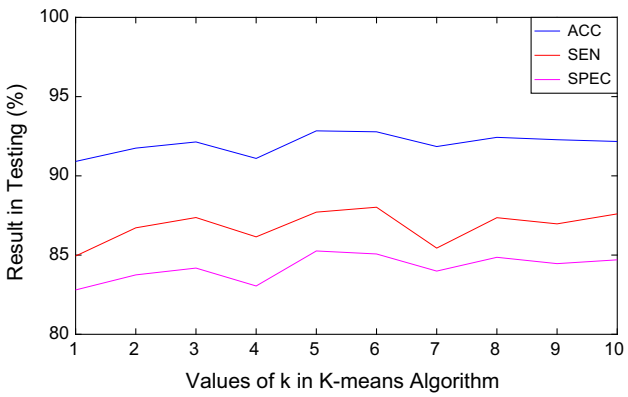

Fig. 12 The testing results of *K*-means IT2FNN for the fetal cardiotocography dataset

Table 11 The comparison result of fetal cardiotocography

Year	Author	Method	Accuracy (%)
2013	Yılmaz [32]	LS-SVM-PSO-BDT	91.62
2014	Karabulut [33]	MLPNN-based AdaBoost Ensbl	92.05
2015	Ravindran [34]	Principal Component Analysis (PCA)	92.14
2016	Yılmaz [21]	Generalized regression neural network (GRNN)	91.86
2017	Mandal [22]	Multiple classifier system (MCS)	92.74
	Our method	K-mean IT2FNN ($k = 5$)	92.84

applications for medical diagnosis, the proposed method is also suitable for many other classification systems because the computation process is easy and it is easy to implement.

Acknowledgements The authors appreciate the financial support in part from the Ministry of Science and Technology of Republic of China under Grant MOST 106-2221-E-155-MY3.

References

1. Steinhaus, H.: Sur la division des corp materiels en parties. Bull. Acad. Polonaise Sci. **1**(804), 801 (1956)
2. Xie, J., Jiang, S., Xie, W., Gao, X.: An efficient global *K*-means clustering algorithm. J. Comput. Phys. **6**(2), 271–279 (2011)
3. Vora, P., Oza, B.: A survey on *K*-mean clustering and particle swarm optimization. Int. J. Sci. Mod. Eng. **1**(3), 1–14 (2013)
4. Yunoh, M., Abdullah, S., Saad, M., Nopiah, Z., Nuawi, M.: *K*-means clustering analysis and artificial neural network classification of fatigue strain signals. J. Brazilian Soc. Mech. Sci. Eng. **39**(3), 757–764 (2017)
5. Singh, K., Malik, D., Sharma, N.: Evolving limitations in *K*-means algorithm in data mining and their removal. Int. J. Comput. Eng. Manag. **12**, 105–109 (2011)
6. Jiang, P., Zhang, L., Li, W., Wang, X.: Pattern recognition for acoustic emission signals of offshore platform T-tube damage based on *K*-means clustering. Advances in Acoustic Emission Technology, pp. 53–61. Springer, Berlin (2017)
7. Tuncer, S.A., Alkan, A.: A decision support system for detection of the renal cell cancer in the kidney. Measurement **123**, 298–303 (2018)
8. Gan, G., Ng, M.K.-P.: *K*-means clustering with outlier removal. Pattern Recogn. Lett. **90**, 8–14 (2017)
9. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**(3), 338–353 (1965)
10. Zadeh, L.A.: The concept of a linguistic variable and its application to approximate reasoning—I. Inf. Sci. **8**(3), 199–249 (1975)
11. Liang, Q., Mendel, J.M.: Interval type-2 fuzzy logic systems: theory and design. IEEE Trans. Fuzzy Syst. **8**(5), 535–550 (2000)
12. Eyoh, I., John, R., De Maere, G.: Interval type-2 intuitionistic fuzzy logic for regression problems. IEEE Trans. Fuzzy Syst. **26**(4), 2396–2408 (2017)
13. Zirkohi, M.M., Lin, T.-C.: Interval type-2 fuzzy-neural network indirect adaptive sliding mode control for an active suspension system. Nonlinear Dyn. **79**(1), 513–526 (2015)
14. Lin, C.-M., Le, T.-L., Huynh, T.-T.: Self-evolving function-link interval type-2 fuzzy neural network for nonlinear system identification and control. Neurocomputing **275**, 2239–2250 (2018)
15. Lin, C.-M., Le, T.-L.: PSO-self-organizing interval type-2 fuzzy neural network for antilock braking systems. Int. J. Fuzzy Syst. **19**(5), 1362–1374 (2017)
16. Li, H., Wang, J., Wu, L., Lam, H.-K., Gao, Y.: Optimal guaranteed cost sliding-mode control of interval type-2 fuzzy time-delay systems. IEEE Trans. Fuzzy Syst. **26**(1), 246–257 (2018)
17. Guan, J.-S., Lin, L.-Y., Ji, G.L., Lin, C.-M., Le, T.-L., Rudas, I.J.: Breast tumor computer-aided diagnosis using self-validating cerebellar model neural networks. Acta Polytech. Hung. **13**(4), 39–52 (2016)
18. Zhou, Q., Chao, F., Lin, C.-M.: A functional-link-based fuzzy brain emotional learning network for breast tumor classification and chaotic system synchronization. Int. J. Fuzzy Syst. **20**(2), 349–365 (2018)
19. Dwivedi, A.K.: Performance evaluation of different machine learning techniques for prediction of heart disease. Neural Comput. Appl. **29**(10), 1–9 (2016)
20. Liu, X., Wang, X., Su, Q., Zhang, M., Zhu, Y., Wang, Q., Wang, Q.: A hybrid classification system for heart disease diagnosis based on the RFRS method. Comput. Math. Methods Med. **2017**, 1–11 (2017)
21. Yılmaz, E.: Fetal state assessment from cardiotocogram data using artificial neural networks. J. Med. Biol. Eng. **36**(6), 820–832 (2016)
22. Mandal, I.: Machine learning algorithms for the creation of clinical healthcare enterprise systems. Enterp. Inf. Syst. **11**(9), 1374–1400 (2017)
23. Aličković, E., Subasi, A.: Breast cancer diagnosis using GA feature selection and rotation forest. Neural Comput. Appl. **28**(4), 753–763 (2017)
24. Mendel, J.M.: Uncertain rule-based fuzzy logic systems: introduction and new directions. Prentice Hall PTR, Upper Saddle River (2001)
25. Stoean, R., Stoean, C.: Modeling medical decision making by support vector machines, explaining by rules of evolutionary algorithms with feature selection. Expert Syst. Appl. **40**(7), 2677–2686 (2013)
26. Zheng, B., Yoon, S.W., Lam, S.S.: Breast cancer diagnosis based on feature extraction using a hybrid of *K*-means and support vector machine algorithms. Expert Syst. Appl. **41**(4), 1476–1482 (2014)
27. Lim, C.K., Chan, C.S.: A weighted inference engine based on interval-valued fuzzy relational theory. Expert Syst. Appl. **42**(7), 3410–3419 (2015)
28. Khan, R.A., Suleman, T., Farooq, M.S., Rafiq, M.H., Tariq, M.A.: Data mining algorithms for classification of diagnostic cancer using genetic optimization algorithms. IJCSNS **17**(12), 207 (2017)
29. Buscema, M., Breda, M., Lodwick, W.: Training with Input selection and testing (TWIST) algorithm: a significant advance in pattern recognition performance of machine learning. J. Intell. Learn. Syst. Appl. **5**(1), 29 (2013)
30. Tomar, D., Agarwal, S.: Feature selection based least square twin support vector machine for diagnosis of heart disease. Int. J. Bio-Sci. Bio-Technol. **6**(2), 69–82 (2014)

31. Lee, S.-H.: Feature selection based on the center of gravity of BSWFMs using NEWFM. Eng. Appl. Artif. Intell. **45**, 482–487 (2015)
32. Yılmaz, E., Kılıkçıer, Ç.: Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree. Comput. Math. Methods Med. **2013**, 1–8 (2013)
33. Karabulut, E.M., Ibrikci, T.: Analysis of cardiotocogram data for fetal distress determination by decision tree based adaptive boosting approach. J. Comput. Commun. **2**(09), 32–37 (2014)
34. Ravindran, S., Jambek, A.B., Muthusamy, H., Neoh, S.-C.: A novel clinical decision support system using improved adaptive genetic algorithm for the assessment of fetal well-being. Comput. Math. Methods Med. **2015**, 1–11 (2015)

Tien-Loc Le was born in Vietnam, in 1985. He received the B.S. degree in Electronics and Telecommunication Engineering from the LacHong University, Vietnam in 2009, where he is currently a lecturer; the M.S. degree in Electrical Engineering from Hochiminh University of Technical Education, Vietnam, in 2012; and the Ph.D. degree in Electrical Engineering from Yuan Ze University, Taoyuan, Taiwan, in 2018. He is currently a Postdoctoral Research Fellow with the Department of Electrical Engineering, Yuan Ze University, Taoyuan, Taiwan. His research interest includes intelligent control systems, fuzzy neural network, and cerebellar model articulation controller.

Tuan-Tu Huynh was born in Ho Chi Minh City, Viet Nam, in 1982. He received the B.S. degree from Department of Electrical & Electronics Engineering, Ho Chi Minh University of Technology and Education, Vietnam, in 2005; the M.S. degree in Automation from Ho Chi Minh City University of Transport, Viet Nam, in 2010; and the Ph.D. degree in Electrical Engineering from Yuan Ze University, Taoyuan, Taiwan, in 2018. He is currently a Research Fellow with the Department of Electrical Engineering, Yuan Ze University, Chung-Li, Taiwan. He is also a lecturer at Lac Hong University, Vietnam. His research interests include MCDM, fuzzy logic control, neural network, cerebellar model articulation controller, brain emotional learning-based intelligent controller, and intelligent control systems.

Lo-Yi Lin was born in Taipei, Taiwan in 1989. She graduated from School of Medicine, Taipei Medical University, Taipei, Taiwan in 2016. She completed the Post Graduate Year program in Far Eastern Memorial Hospital, New Taipei City, Taiwan. She is currently a resident doctor of Department of Radiology, Taipei Veterans General Hospital, Taipei, Taiwan.

Chih-Min Lin (M'87–SM'99–F'10) was born in Changhua, Taiwan in 1959. He received the B.S. and M.S. degrees from Department of Control Engineering in 1981 and 1983, respectively, and Ph.D. degree from Institute of Electronics Engineering in 1986, all in National Chiao Tung University, Hsinchu, Taiwan. He is currently a Chair Professor and the Vice President of Yuan Ze University, Taoyuan, Taiwan. He also serves as an Associate Editor of *IEEE Transactions on Cybernetics and IEEE Transactions on Fuzzy Systems.* During 1997–1998, he was the honor research fellow at the University of Auckland, New Zealand. His research interests include fuzzy neural network, cerebellar model articulation controller, intelligent control system, adaptive signal processing and classification problem. He has published more than 200 journal papers and 160 conference papers.

Fei Chao received the B.Sc. degree in Mechanical Engineering from the Fuzhou University, China, and the M.Sc. degree with distinction in Computer Science from the University of Wales, Aberystwyth, UK, in 2004 and 2005, respectively, and the Ph.D. degree in Robotics from the Aberystwyth University, Wales, UK in 2009. He is currently an Associate Professor with the Cognitive Science Department, Xiamen University, China. Dr. Chao has published more than 50 peer-reviewed journal and conference papers. His research interests include developmental robotics, reinforcement learning, and optimization algorithms. He is the vice chair of the IEEE Computer Intelligence Society Xiamen Chapter. Also, he is a member of IEEE, CCF and CAAI.