This is the Pre-Published Version.

The following publication Q. Lou, Z. Deng, G. Wang and K. S. Choi, "Reliability Learning for Interval Type-2 TSK Fuzzy Logic System with its Application to Medical Diagnosis," 2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineering (ISKE), Dalian, China, 2019, pp. 43-50 is available at https://doi.org/10.1109/ISKE47853.2019.9170445.

Reliability Learning for Interval Type-2 TSK Fuzzy Logic System Classifier

Zaohong Deng, *IEEE Senior Member*, Kup-Sze Choi, Shitong Wang

***Abstract*—To apply intelligent model in serious practical applications like medical diagnosis, the reliability and interpretability of the model are very important to users. Among the existing intelligent models, type-2 fuzzy systems are distinctive in interpretability and modeling uncertainty. However, like most existing models, the determination of the reliability of fuzzy systems trained for recognition task is still an unsolved problem at present. In this study, the minimax probability interval type-2 TSK fuzzy logic system classifier (MP-IT2TSK-FLSC) is proposed to construct a type-2 fuzzy system classifier based on reliability learning. The proposed classifier can provide the lower bound of correct classification of the model, which is an important indicator to quantifying the model reliability. Experimental results on medical datasets have demonstrated the advantages of this method, exhibiting remarkable interpretability and reliability of the proposed fuzzy classifier.**

***Index Terms*—Type-2 fuzzy logic system, minimax probability decision, model reliability, classification.**

I. INTRODUCTION

As a typical type of intelligent models, fuzzy logic systems (FLSs) have been extensively applied for system modeling and pattern recognition. FLS has experienced two important stages of development. The first stage concerned type-1 (T1) FLSs based on the classical fuzzy set and the second stage concentrated on the advanced type-2 (T2) FLSs using the T2 fuzzy set. T1 FLSs have been comprehensively studied from both theoretical and application aspects since 1970 [1-5], while T2 FLSs have been attracting more attention only in the last decade [6-8].

Compared with T1 FLSs, T2 FLSs possess superior performance in modeling uncertainty which is essential for real world applications. In addition, the interpretation and learning abilities of T2 FLSs are as strong as that of T1 FLSs. The Interval T2 FLSs (IT2 FLSs) are the most extensively used T2 FLSs because of their efficiency and simplicity. In this study, we will focus on this type of FLSs to investigate the model reliability of FLS based classifier.

This work was supported in part by the National Natural Science Foundation of China under Grants 61170122, in part by the Ministry of Education Program for New Century Excellent Talents under Grant NCET-120882, in part by the Hong Kong Research Grants Council under Grant PolyU 5134/12E.

Z.H. Deng is with the School of Digital Media, Jiangnan University, Wuxi 214122, China and the Department of Biomedicine Engineering, University of California, Davis, Davis 95616, USA (e-mail: dzh666828@yahoo.com.cn and zhdeng@ucdavis.edu).

K.S. Choi is with the Centre Smart Health of the School of Nursing and the Interdisciplinary Division of Biomedical Engineering, Hong Kong Polytechnic University(e-mail: kschoi@ieee.org).

S.T. Wang is with the School of Digital Media, Jiangnan University, Wuxi 214122, China (e-mail: wxwangst@yahoo.com.cn).

The learning of the model parameters is a critical task for IT2 FLS construction. At present, data-driven learning is the most important and popular method. Various effective approaches have been proposed for IT2 FLSs, including methods based on gradient learning [9], support vector learning [12] and self-organizing evolution [13,14], as well as fast type-2 fuzzy extreme learning algorithm based on extreme learning mechanism [47], singular value decomposition-QR decomposition (SVD-QR) method [10], dynamical optimal training method [11], hybrid learning algorithms [15-21] and bio-inspired methods [22-27]. These algorithms have been used for training different IT2 FLS models, such as IT2 Mamdani-Larsen FLSs (IT2 ML FLSs) and IT2 TSK FLSs. Although the methods have demonstrated promising performance for some specific applications, a critical challenge in model training remains, i.e., all these algorithms do provide a reliability index for the model obtained, which severely prohibits extensive applications of IT2 FLSs. In some real-world applications, such as medical diagnosis, the decisions resulting from models with unknown reliability, which could be unconvincing or even unacceptable to users. Currently, research on model reliability is still scarce for IT2 FLSs. The study of new IT2 FLS learning methods that is able to indicate the reliability of the model is therefore very significant for extensive applications of IT2 FLSs.

Some researchers have begun to pay attention to the model reliability of FLSs recently. For example, model reliability was studied for evolving fuzzy systems in the context of data stream processing [58] [59] by introducing the concepts of “conflict” and “ignorance”. However, this method is only suitable for a specific application. A more general model reliability learning strategy is thus desired. In the past decade, the minimax probability decision technique has been considered as an important approach to determine model reliability. Considerable research has been conducted with this technique in different fields [28-34]. For example, a minimax probability machine (MPM) for novelty detection and classification has been proposed. The distinctive characteristic with this intelligent decision making technique is that the obtained model provides the lower bound for the accuracy of the decision made.

In this study, we introduce the minimax probability decision technique to realize a reliability-learning-based fuzzy classifier. Accordingly, a minimax probability IT2 TSK FLS classifier, named as MP-IT2TSK-FLSC, is proposed to train the classifier and determine the model reliability simultaneously. For the proposed MP-IT2TSK-FLSC, the lower bound of correct classification is available for users, making it a practical approach for real-world applications. The proposed method has been evaluated on medical datasets for diagnostic purposes and the effectiveness has been confirmed. As such, the

MP-IT2TSK-FLSC possesses superior performance in modeling uncertainty, high level of interpretation, and better transparency in model reliability.

The rest of this paper is organized as follows. The related concepts about IT2 TSK FLS and the minimax probability machine are reviewed in Section II. In Section III, the MP-IT2TSK-FLSC is proposed based on the minimax probability decision technique. The experimental results on several medical datasets for diagnostic purposes are reported in Section IV. Conclusions and future work are given in the last section.

II. Related Work

In this section, we first introduce the typical IT2 TSK FLS. The classical minmax probability decision technique for classification is then described in brief, which will be used to develop the proposed MP-IT2TSK-FLSC.

A. IT2 TSK FLS

Among T2 FLSs, IT2 TSK FLS is widely used for the ease of design and implementation. The commonly used fuzzy inference rules for IT2 TSK FLSs are given as follows [12],

Rules of IT2 TSK FLS R^k :

If x_1 is $\tilde{A}_1^k \wedge x_2$ is $\tilde{A}_2^k \wedge \cdots \wedge x_d$ is $\tilde{A}_d^k$

Then $w^k = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d$, $k = 1, \cdots, K$,

where $\tilde{A}_i^k$ is an IT2 fuzzy subset associated with the input variable x_i for the k-th rule; k is the number of fuzzy rules and $\wedge$ is a fuzzy conjunction operator; $\mathbf{p}^k = [p_0^k, p_1^k, \cdots,, p_d^k]^T$ denotes the consequent parameters of the k-th fuzzy rule. Each rule is premised on the input vector $\mathbf{x} = [x_1, x_2, \cdots, x_d]^T$, mapping the fuzzy sets in the input space $\tilde{A}^k \subset R^d$ to a varying singleton, i.e., w^k .

Different types of primary membership function (MF) can be defined for the IT2 fuzzy set $\tilde{A}_i^k$. For example, Gaussian primary membership function with a fixed standard deviation σ_i^k and an uncertain mean within $[m_i^{k1}, m_i^{k2}]$ can be expressed as

$$\mu_{\tilde{A}_i^k}(x_i) = \exp\left[-\frac{1}{2}\left(\frac{x_i - m_i^k}{\sigma_i^k} \right)^2 \right] \equiv N(x_i, m_i^k, \sigma_i^k), \quad m_i^k \in [m_i^{k1}, m_i^{k2}], \tag{1.a}$$

with m_i^k and σ_i^k as the mean and width of the Gaussian membership function respectively; the lower primary MF $\underline{\mu}_{\tilde{A}_i^k}(x_i)$ and upper primary MF $\bar{\mu}_{\tilde{A}_i^k}(x_i)$ of Eq. (1.a) can be expressed below.

$$\underline{\mu}_{\tilde{A}_i^k}(x_i) = \begin{cases} N(x_i;\ m_i^{k2}, \sigma_i^k) & x_i \le \dfrac{m_i^{k1} + m_i^{k2}}{2} \\ N(x_i;\ m_i^{k1}, \sigma_i^k) & x_i > \dfrac{m_i^{k1} + m_i^{k2}}{2} \end{cases}, \tag{1.b}$$

$$\bar{\mu}_{\tilde{A}_i^k}(x_i) = \begin{cases} N(x_i;\ m_i^{k1}, \sigma_i^k) & x_i < m_i^{k1} \\ 1 & m_i^{k1} \le x_i \le m_i^{k2} \\ N(x_i;\ m_i^{k2}, \sigma_i^k) & x_i > m_i^{k2} \end{cases}. \tag{1.c}$$

Each rule of the IT2 TSK FLS performs a fuzzy meet operation using the algebraic product. The output of the *if-part* of a rule is the firing strength given by the following interval T1 (IT1) fuzzy set,

$$F_k = [\underline{f}_k, \bar{f}_k], \tag{2.a}$$

$$\underline{f}_k = \prod_{i=1}^{d} \underline{\mu}_{\tilde{A}_i^k} \tag{2.b}$$

$$\bar{f}_k = \prod_{i=1}^{d} \bar{\mu}_{\tilde{A}_i^k}. \tag{2.c}$$

The commonly used output of the *then-part* is the linear combination of the input values, i.e.,

$$w^k = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d = \sum_{i=0}^{d} p_i^k x_i. \tag{3}$$

By using a T2 reducer, the following IT1 fuzzy set $[y_l, y_r]$ can be computed for the IT2 TSK FLS:

$$[y_l, y_r] = \int_{w^1} \cdots \int_{w^K} \int_{f_1 \in [\underline{f}_1, \bar{f}_1]} \cdots \int_{f_K \in [\underline{f}_K, \bar{f}_K]} 1 \Big/ \frac{\sum_{k=1}^{K} f_k w^k}{f_k}. \tag{4}$$

Since there is no direct theoretical solution for Eq. (4), iterative methods have been proposed to compute the reduced set. A commonly used method is the Karnik-Mendel (KM) iterative procedure [7, 35], in which the consequent values should be re-arranged in ascending order. Denoting the original rule-ordered consequent values as $\mathbf{w} = [w^1, \cdots, w^K]^T$ and the re-ordered sequence as $\tilde{\mathbf{w}} = [\tilde{w}^1, \cdots, \tilde{w}^K]^T$ ($\tilde{w}^1 \le \tilde{w}^2 \le \cdots \le \tilde{w}^K$), the relationship between $\mathbf{w}$ and $\tilde{\mathbf{w}}$ can be obtained by

$$\tilde{\mathbf{w}} = \mathbf{Q}\mathbf{w}, \tag{5}$$

where $\mathbf{Q}$ is a $K \times K$ permutation matrix [9, 12]. The elementary vectors in this permutation matrix are taken as the columns, and they are arranged so that the elements in $\mathbf{w}$ are moved to new locations in the transformed vector $\tilde{\mathbf{w}}$ in ascending order. The elementary vector contains unity element in a specified position while the remaining elements are all zeros. A detailed description about the construction of $\mathbf{Q}$ can be found in [9]. Accordingly, the rule orders $\underline{f}_k$ and $\bar{f}_k$ are rearranged to yield $\underline{\tilde{f}}_k$ and $\bar{\tilde{f}}_k$, respectively. The outputs y_l and y_r in Eq. (4) can then be computed as follows,

$$y_l = \frac{\sum_{k=1}^{L} \bar{\tilde{f}}_k \tilde{w}^k + \sum_{k=L+1}^{K} \underline{\tilde{f}}_k \tilde{w}^k}{\sum_{k=1}^{L} \bar{\tilde{f}}_k + \sum_{k=L+1}^{K} \underline{\tilde{f}}_k} \tag{6.a}$$

and

$$y_r = \frac{\sum_{k=1}^{R} \underline{\tilde{f}}_k \tilde{w}^k + \sum_{k=R+1}^{K} \bar{\tilde{f}}_k \tilde{w}^k}{\sum_{k=1}^{R} \underline{\tilde{f}}_k + \sum_{k=R+1}^{K} \bar{\tilde{f}}_k}, \tag{6.b}$$

where L and R are the switch points obtained by the KM algorithm or its modified versions [7, 35]. With the IT1 fuzzy set $[y_l, y_r]$, the final output can be computed by averaging y_l and y_r , i.e.,

$$y = (y_l + y_r)/2. \tag{7}$$

Besides KM algorithm, its variants have also been proposed to obtain the switch points with reduced computational time [36, 37]. For example, the enhanced KM algorithm (EKM) [36] is an efficient one that has been used in our experiments.

B. Minimax Probability Decision Technique

The minimax probability decision technique has been extensively investigated to cater for different applications [28-34]. The aim of the methods based on the minimax probability principle is to obtain the lower bound of correct decision for the modeling tasks. Here, we will briefly review the principle of a classical minimax probability decision based classifier, i.e., MPM [29], which is closely related to the proposed MP-IT2TSK-FLSC in this study.

Given a dataset containing two classes sampled from two random variables $\mathbf{x} \sim (\mathbf{u}_+, \Sigma_+)$ and $\mathbf{x} \sim (\mathbf{u}_-, \Sigma_-)$, where $\mathbf{u}_+, \Sigma_+$ and $\mathbf{u}_-, \Sigma_-$ denote the means and standard deviations of the two classes respectively. MPM defines the following optimization objective to obtain the classification hyperplane $\mathbf{w}^T\mathbf{x} - b = 0$,

$$\begin{aligned} &\max_{\alpha,\mathbf{w},b} \ \alpha \\ &s.t. \inf_{\mathbf{x}\sim(\mathbf{u}_+,\Sigma_+)} pr(\mathbf{w}^T\mathbf{x} \ge b) \quad \ge \quad \alpha \\ &\quad\ \ \inf_{\mathbf{x}\sim(\mathbf{u}_-,\Sigma_-)} pr(\mathbf{w}^T\mathbf{x} \le b) \quad \ge \quad \alpha \end{aligned} \tag{8}$$

where $\inf_{\mathbf{x}\sim(\mathbf{u}_+,\Sigma_+)} pr(\mathbf{w}^T\mathbf{x} \ge b)$ denotes the infimum of the probability for the condition: $\mathbf{w}^T\mathbf{x} \ge b$ with $\mathbf{x} \sim (\mathbf{u}_+, \Sigma_+)$; α is the lower bound of correct classification. The optimization objective in Eq.(8) implies that for a two-class data sampled from random variables $\mathbf{x} \sim (\mathbf{u}_+, \Sigma_+)$ and $\mathbf{x} \sim (\mathbf{u}_-, \Sigma_-)$, there exists an optimal hyperplane $\left(\mathbf{w}^*\right)^T \mathbf{x} - b^* = 0$ which specify the lower bound of correct classification of a future datum point as maximum. The following lemmas and theorem [29] are used to obtain the solution of Eq.(8).

Lemma 1. *With* $\mathbf{u}_-, \mathbf{\Sigma}_-$ *being positive definite,* $\mathbf{w} \ne \mathbf{0}$, *a given* b *such that* $\mathbf{w}^T\mathbf{u}_- - b \le 0$, *and* $\alpha \in [0,1)$, *the condition* $\inf_{\mathbf{x}\sim(\mathbf{u}_-,\mathbf{\Sigma}_-)} pr(\mathbf{w}^T\mathbf{x} - b \le 0) \ \ge \ \alpha$ *holds if and only if* $b - \mathbf{w}^T\mathbf{u}_- \ge \kappa(\alpha)\sqrt{\mathbf{w}^T\mathbf{\Sigma}_-\mathbf{w}}$ *,where* $\kappa(\alpha)=\sqrt{\frac{\alpha}{1\text{-}\alpha}}$.

Lemma 2. *With* $\mathbf{u}_+, \mathbf{\Sigma}_+$ *being positive definite,* $\mathbf{w} \ne \mathbf{0}$, *a given* b *such that* $\mathbf{w}^T\mathbf{u}_+ - b \ge 0$, *and* $\alpha \in [0,1)$, *the condition* $\inf_{\mathbf{x}\sim(\mathbf{u}_+,\mathbf{\Sigma}_+)} pr(\mathbf{w}^T\mathbf{x} - b \ge 0) \ \ge \ \alpha$ *holds if and only if* $\mathbf{w}^T\mathbf{u}_+ - b \ge \kappa(\alpha)\sqrt{\mathbf{w}^T\mathbf{\Sigma}_+\mathbf{w}_g}$ *,where* $\kappa(\alpha)=\sqrt{\frac{\alpha}{1\text{-}\alpha}}$.

Based on Lemmas 1 and 2, Eq.(8) can be transformed to the following optimization problem.

$$\begin{aligned} &\max_{\mathbf{w},b,\alpha} \ \alpha \\ &s.t. \ \mathbf{w}^T\mathbf{u}_+ - b \ge \kappa(\alpha)\sqrt{\mathbf{w}^T\mathbf{\Sigma}_+\mathbf{w}} \\ &\qquad b - \mathbf{w}^T\mathbf{u}_- \ge \kappa(\alpha)\sqrt{\mathbf{w}^T\mathbf{\Sigma}_-\mathbf{w}} \end{aligned} \tag{9}$$

Based on Eq.(9), the Theorem 1 presented below can be used to obtain the solution variables in Eq.(8) [29].

Theorem 1. *If* $\mathbf{u}_+ = \mathbf{u}_-$, *then the minimax probability decision problem in Eq.(8) does not have a meaningful solution: the optimal worst-case misclassification probability obtained is* $1-\alpha^*$ *=1. Otherwise, an optimal hyperplane* $\mathrm{H}(\mathbf{w}^*, b^*)$ *exists and can be determined by solving the convex optimization problem*

$$\begin{aligned} &\kappa(\alpha)^* = \min_{\mathbf{w}} \ \sqrt{\mathbf{w}^T\mathbf{\Sigma}_+\mathbf{w}} + \sqrt{\mathbf{w}^T\mathbf{\Sigma}_-\mathbf{w}} \\ &s.t. \ \mathbf{w}^T(\mathbf{u}_+ - \mathbf{u}_-) = 1 \end{aligned} , \tag{10.a}$$

and setting b^* *as follows,*

$$b^* = (\mathbf{w}^*)^T\mathbf{u}_+ - \kappa^*\sqrt{(\mathbf{w}^*)^T\mathbf{\Sigma}_+\mathbf{w}^*} , \tag{10.b}$$

where $\mathbf{w}^*$ *is the optimal solution of* $\mathbf{w}$ *. The optimal worst-case misclassification probability is obtained via*

$$\begin{aligned} 1-\alpha^* &= \frac{1}{1+(\kappa(\alpha)^*)^2} \\ &= \frac{\left(\sqrt{(\mathbf{w}^*)^T\mathbf{\Sigma}_+\mathbf{w}^*} + \sqrt{(\mathbf{w}^*)^T\mathbf{\Sigma}_-\mathbf{w}^*}\right)^2}{1+\left(\sqrt{(\mathbf{w}^*)^T\mathbf{\Sigma}_+\mathbf{w}^*} + \sqrt{(\mathbf{w}^*)^T\mathbf{\Sigma}_-\mathbf{w}^*}\right)^2} \ . \end{aligned} \tag{10.c}$$

If either $\mathbf{\Sigma}_+$ *or* $\mathbf{\Sigma}_-$ *is positive definite, the optimal hyperplane* $\mathrm{H}(\mathbf{w}^*, b^*)$ *is unique.*

Eqs. (10.a) is a second order core program (SOCP) optimization problem [38]. It can be effectively solved by tools such as SeDuMi [39]. A specific algorithm was also proposed in [29] for this problem.

By introducing the kernel trick, the objective of the kernelized version of MPM is proposed as follows,

$$\begin{aligned} &\max_{\alpha,\mathbf{w},b} \ \alpha \\ &s.t. \inf_{\mathbf{x}\sim(\mathbf{u}_+,\Sigma_+)} pr(\mathbf{w}^T\varphi(\mathbf{x}) \ge b) \quad \ge \quad \alpha \ , \\ &\quad\ \ \inf_{\mathbf{x}\sim(\mathbf{u}_-,\Sigma_-)} pr(\mathbf{w}^T\varphi(\mathbf{x}) \le b) \quad \ge \quad \alpha \end{aligned} \tag{11}$$

where $\varphi(\mathbf{x})$ is a mapping function that maps the data $\mathbf{x}$ in the original space to $\varphi(\mathbf{x})$ in the kernel feature space. The kernelized MPM can also be solved by transforming it to a second order core program (SOCP) optimization problem [29].

III. MP-IT2TSK-FLSC

A. The Proposed IT2TSK-FLSC Model and Minimax Probability Objective Criterion

1) The IT2TSK-FSLC Model

Classification can be regarded as a special case of regression with discrete labels as the outputs of the input-output pairs in the regression dataset. Thus, IT2 TSK FLS, as a regression model, can be used for classification [40-43]. However, a better way of classification by IT2 TSK FLS is to train the FLS using a specific learning mechanism for classification. This is commonly achieved by developing a binary classification decision function for FLS, i.e.,

$$y = sign\left(f_{FLS}(\mathbf{x})\right) = \begin{cases} 1 & \text{if } f_{FLS}(\mathbf{x}) \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (12)$$

Based on the above decision function, fuzzy-system-based classifiers for binary classification can be developed [44-46], for example, the self-organizing TS-type fuzzy network with support vector learning (SOTFN-SV) fuzzy classification method [44]. By using a similar approach, the IT2 TSK FLS based classification model is proposed below.

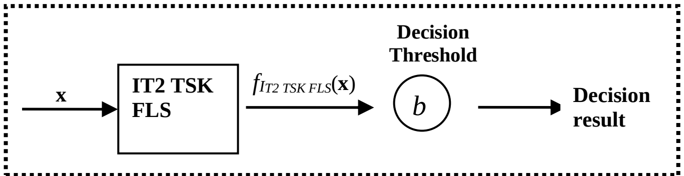


Fig. 1 The proposed IT2TSK-FLSC Model

The proposed classification model consists of two parts, i.e., the classical IT2 TSK FLS and a decision threshold, as shown in Fig. 1. With this model, the final decision function for binary classification is expressed as follows.

$$y = sign\left(f_{IT2TSK-FS}(\mathbf{x}) - b\right) = \begin{cases} 1 & \text{if } f_{TSK-FS}(\mathbf{x}) > b \\ -1 & \text{otherwise} \end{cases} \quad (13)$$

2) Minimax Probability Objective Criterion

Based on the minimax probability decision theory, the following objective is proposed to train the proposed IT2TSK-FLSC model,

$$\begin{aligned} &\max_{\mathbf{\Theta},\alpha} \ \alpha \\ &s.t. \inf_{\mathbf{x}\sim(\mathbf{u}_+,\mathbf{\Sigma}_+)} pr(f_{IT2TSK-FS}(\mathbf{x}) - b \geq 0) \quad \geq \alpha , \quad (14) \\ &\quad\ \inf_{\mathbf{x}\sim(\mathbf{u}_-,\mathbf{\Sigma}_-)} pr(f_{IT2TSK-FS}(\mathbf{x}) - b \leq 0) \quad \geq \alpha \end{aligned}$$

where $\mathbf{\Theta}$ is the parameter set of the proposed IT2TSK-FLSC, including the parameters of IT2 TSK FLS and the decision threshold b . With this optimization criterion, we intend to train the IT2 TSK-FLSC model and obtain the corresponding lower bound of correct classification. However, it is a non-trivial task to solve the objective function in Eq. (14) directly. To overcome this issue, Eq.(14) is replaced with a modified objective function to readily solve the optimization problem readily.

B. Learning Framework of MP-IT2TSK-FSLC

Generally, for a FLS, the antecedents and consequents can be determined independently. A popular way for determining antecedents is by using clustering techniques [47-49] or other partition techniques, such as the self-evolving learning method [12, 44], to partition the input space for a modeling task. Once the antecedents are determined with the above strategy, the consequents and the decision threshold of the proposed MP-IT2TSK-FSLC model can be easily learned with the minimax probability decision technique. The framework and details to construct the MP-IT2TSK-FLSC and its algorithm will be described in the following section. An overview of the major procedure of the MP-IT2TSK-FLSC algorithm is first presented here.

The proposed training method for the MP-IT2TSK-FLSC consists of three stages, as illustrated in Fig. 2. In Stage 1, according to the adopted clustering technique or other partition techniques, the parameters of the antecedents are assigned and then fixed in the subsequent stages. In Stage 2, the parameters of the consequents and the decision threshold are initialized by solving for a decision hyperpalne with the minimax decision technique, where the KM algorithm is *not* used to compute the switch points for type-2 type reducer since the initial parameters in the consequents are not available yet. In Stage 3, the consequents and the decision threshold are further refined based on the result obtained in Stage 2 by using the KM algorithm to obtain the switch points for the type reducer. The corresponding decision hyperpalne is solved again with the minimax probability decision technique.

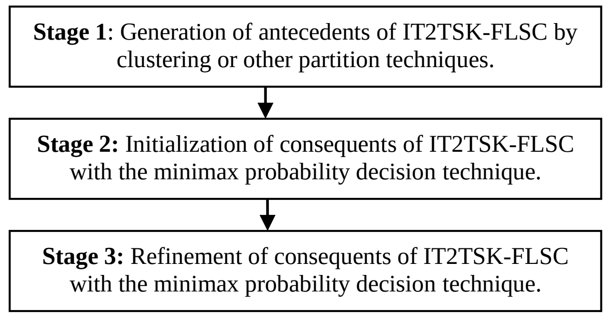


Fig.2 The framework of the proposed leaning method for MP-IT2TSK-FLSC

C. Three Stages of Parameter Learning

1) Generation of Antecedents

In this stage, Eqs. (1.a)-(1.c) are used to model the fuzzy set of the antecedents. The parameters to be determined are the fixed standard deviations σ_i^k and the bounds of the uncertain mean, i.e., $[m_i^{k1}, m_i^{k2}]$. First, by using partition techniques, e.g. FCM clustering technique [47-49] or self-evolutive organization technique [12, 44], the parameters of the antecedents, i.e., $\bar{m}_i^k$ and σ_i^k , are assigned and then fixed in the subsequent stages. Note that according to the extreme learning mechanism, the parameters in the antecedents can also be randomly assigned. Although the approximation abilities remain strong with the randomization strategy, the interpretability of the FLS obtained may be weakened. Further, we use $\bar{m}_i^k$ and the fixed minor deviation $\Delta\bar{m}_i^k$ to obtain $m_i^{k1} = \bar{m}_i^k - \Delta\bar{m}_i^k$ and $m_i^{k2} = \bar{m}_i^k + \Delta\bar{m}_i^k$.

2) Initialization of Consequents and Decision Threshold

In this stage, since the initial parameters of the consequents are not available, the outputs y_l and y_r in Eqs. (6.a) and (6.b) are approximated by the following two equations without using the KM iterative procedure [47].

$$y_l = \frac{\sum_{k=1}^{K} \underline{f}_k w^k}{\sum_{k=1}^{K} \underline{f}_k} = \sum_{k=1}^{K} \underline{f}_k' w^k \ , \ \underline{f}_k' = \frac{\underline{f}_k}{\sum_{k'=1}^{K} \underline{f}_{k'}} \quad (13.a)$$

and

$$y_r = \frac{\sum_{k=1}^{K} \bar{f}_k w^k}{\sum_{k=1}^{K} \bar{f}_k} = \sum_{k=1}^{K} \bar{f}_k' w^k \ , \ \bar{f}_k' = \frac{\bar{f}_k}{\sum_{k'=1}^{K} f_{k'}} . \quad (13.b)$$

Based on Eqs. (13.a) and (13.b), Eq. (7) becomes

$$y = \frac{y_l + y_r}{2} = \frac{1}{2}\left(\sum\nolimits_{k=1}^{K} \underline{f}_k' w^k + \sum\nolimits_{k=1}^{K} \bar{f}_k' w^k\right). \tag{14}$$

Since $w^k = p_0^k + p_1^k x_1 + \cdots + p_d^k x_d$, we have

$$y_l = \sum\nolimits_{k=1}^{K} \underline{f}_k' \left(p_0^k + p_1^k x_1 + \cdots + p_d^k x_d \right) \tag{15.a}$$

and

$$y_r = \sum\nolimits_{k=1}^{K} \bar{f}_k' \left(p_0^k + p_1^k x_1 + \cdots + p_d^k x_d \right). \tag{15.b}$$

Let

$$\phi_l(\mathbf{x}) = \left[\underline{f}_1', \underline{f}_1' x_1, \cdots, \underline{f}_1' x_d, \cdots, \underline{f}_K', \underline{f}_K' x_1, \cdots, \underline{f}_K' x_d \right]^T, \tag{16.a}$$

$$\phi_r(\mathbf{x}) = \left[\bar{f}_1', \bar{f}_1' x_1, \cdots, \bar{f}_1' x_d, \cdots, \bar{f}_K', \bar{f}_K' x_1, \cdots, \bar{f}_K' x_d \right]^T, \tag{16.b}$$

$$\begin{aligned} \phi(\mathbf{x}) &= \frac{1}{2}\left(\phi_l(\mathbf{x}) + \phi_r(\mathbf{x})\right) \\ &= \frac{1}{2}\Big[\left(\underline{f}_1' + \bar{f}_1'\right), \left(\underline{f}_1' + \bar{f}_1'\right)_1 x_1, \cdots, \left(\underline{f}_1' + \bar{f}_1'\right) x_d, \cdots, \\ &\qquad \left(\underline{f}_K' + \bar{f}_K'\right), \left(\underline{f}_K' + \bar{f}'\right) x_1, \cdots, \left(\underline{f}_K' + \bar{f}'\right) x_d \Big] \in R^{(d+1)K} \end{aligned} \tag{16.c}$$

and

$$\mathbf{p} = \left[p_0^1, \cdots, p_d^1, \cdots, p_0^K, \cdots, p_d^K \right]^T \in R^{(d+1)K}, \tag{17}$$

Eq. (14) can then be expressed as the linear system

$$f_{IT2TSK-FS}(\mathbf{x}) = \mathbf{p}^T \phi(\mathbf{x}), \tag{18}$$

and the output of the proposed model is given by

$$\begin{aligned} y &= sign(f_{IT2TSK-FS}(\mathbf{x}) - b) \\ &= sign(\mathbf{p}^T \phi(\mathbf{x}) - b) \end{aligned}. \tag{19}$$

For a given training dataset $D = \{\mathbf{x}_i, y_i\}$ of a binary classification task, when the antecedents of the IT2 TSK FLS are determined (e.g. by fuzzy clustering), we can construct the mapped dataset $\tilde{D} = \{\phi(\mathbf{x}_i), y_i\}$ by fuzzy inference, where $\phi(\mathbf{x}_i)$ is generated using Eq. (16.c). Thus, the training of the proposed MP-IT2TSKFLSC model can be transformed into the parameter learning of the corresponding linear system. According to Eq.(8), the following objective can be adopted for parameter learning based on the minimax probability decision technique,

$$\begin{aligned} &\max_{\mathbf{p},b,\alpha} \ \alpha \\ &s.t. \inf_{\phi(\mathbf{x}) \sim (\tilde{\mathbf{u}}_+^{\phi(\mathbf{x})}, \tilde{\boldsymbol{\Sigma}}_+^{\phi(\mathbf{x})})} pr(\mathbf{p}^T \phi(\mathbf{x}) - b \geq 0) \geq \alpha, \\ &\quad\ \inf_{\phi(\mathbf{x}) \sim (\tilde{\mathbf{u}}_-^{\phi(\mathbf{x})}, \tilde{\boldsymbol{\Sigma}}_-^{\phi(\mathbf{x})})} pr(\mathbf{p}^T \phi(\mathbf{x}) - b \leq 0) \geq \alpha \end{aligned} \tag{20}$$

where $\phi(\mathbf{x}) \sim (\mathbf{u}_+^{\phi(\mathbf{x})}, \boldsymbol{\Sigma}_+^{\phi(\mathbf{x})})$ denotes the distribution of mapped data. The solution to Eq. (20) can be obtained with the following theorem.

Theorem 2 The parameter learning of Eq.(20) can be regarded as a special case of the classical MPM, where the training data $\mathbf{x}$ are mapped to $\phi(\mathbf{x})$ in the new feature space constructed by the fuzzy inference rules using the strategy described with Eqs. (16.a)-(16.c).

Proof. It is evident by comparing Eq.(20) with Eqs.(8) and (11), that the equations are of the same form. Thus Eq. (20) can be regarded as a special case of the minimax probability machine in [29]. The distinctive characteristic of Eq.(20) is that the training data is the mapped data in a feature space constructed by using Eqs. (16.a)-(16.c) with the corresponding fuzzy inference mechanism.

3) Refinement of Consequents and Decision Threshold

The final consequent parameters p_i^k are determined in this stage with the initial values obtained in Stage 2. Since the values of p_i^k are now available by computing the consequent values w^k for all the rules, the K-M iterative procedure can now be used to obtain the outputs of the IT2 TSK FLS accurately. Let $\mathbf{w} = [w^1, \cdots, w^K]^T$, $\underline{\mathbf{f}} = [\underline{f}_1, \cdots, \underline{f}_K]^T$, $\bar{\mathbf{f}} = [\bar{f}_1, \cdots, \bar{f}_K]^T$, where the firing strengths are expressed according to the original rule order. According to Eq.(5), the relationship between the original rule-ordered consequent values $\mathbf{w}$ and the re-ordered sequence $\tilde{\mathbf{w}}$ ($\tilde{w}^1 \leq \tilde{w}^2 \leq \cdots \leq \tilde{w}^K$) can be expressed as $\tilde{\mathbf{w}} = \mathbf{Qw}$, where $\mathbf{Q}$ is a $K \times K$ permutation matrix as described in Eq.(5). Then, $\underline{\mathbf{f}}$ and $\bar{\mathbf{f}}$ are rearranged to yield the re-ordered sequences $\tilde{\underline{\mathbf{f}}} = [\tilde{\underline{f}}_1, \cdots, \tilde{\underline{f}}_K]^T$ and $\tilde{\bar{\mathbf{f}}} = [\tilde{\bar{f}}_1, \cdots, \tilde{\bar{f}}_K]^T$ respectively, with $\tilde{\underline{\mathbf{f}}} = \mathbf{Q}\underline{\mathbf{f}}$ and $\tilde{\bar{\mathbf{f}}} = \mathbf{Q}\bar{\mathbf{f}}$. Therefore, in Eq. 6(a), $\sum\nolimits_{k=1}^{L} \tilde{\bar{f}}_k \tilde{w}^k$, $\sum\nolimits_{k=L+1}^{K} \tilde{\underline{f}}_k \tilde{w}^k$, $\sum\nolimits_{k=1}^{L} \tilde{\bar{f}}$ and $\sum\nolimits_{k=L+1}^{K} \tilde{\underline{f}}_k$ can be expressed in a more compact matrix form respectively as $\bar{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_1^T \mathbf{E}_1^T \mathbf{Qw}$, $\underline{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_2^T \mathbf{E}_2^T \mathbf{Qw}$, $\sum\nolimits_{k=1}^{L} (\mathbf{Q}\bar{\mathbf{f}})_k$ and $\sum\nolimits_{k=L+1}^{K} (\mathbf{Q}\underline{\mathbf{f}})_k$. Finally, Eq. (6.a) can be re-written in the rule-ordered form

$$y_l = \frac{\bar{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_1^T \mathbf{E}_1^T \mathbf{Qw} + \underline{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_2^T \mathbf{E}_2^T \mathbf{Qw}}{\sum\nolimits_{k=1}^{L} (\mathbf{Q}\bar{\mathbf{f}})_k + \sum\nolimits_{k=L+1}^{K} (\mathbf{Q}\underline{\mathbf{f}})_k} = \boldsymbol{\psi}_l^T \mathbf{w}, \tag{21.a}$$

where

$$\boldsymbol{\psi}_l^T = [\psi_{l,1}, \cdots, \psi_{l,K}] = \frac{\bar{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_1^T \mathbf{E}_1^T \mathbf{Q} + \underline{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_2^T \mathbf{E}_2^T \mathbf{Q}}{\sum\nolimits_{k=1}^{L} (\mathbf{Q}\bar{\mathbf{f}})_k + \sum\nolimits_{k=L+1}^{K} (\mathbf{Q}\underline{\mathbf{f}})_k} \in R^K, \tag{21.b}$$

$$\mathbf{E}_1 = \left[\mathbf{e}_1, \cdots, \mathbf{e}_L, \mathbf{0}, \cdots, \mathbf{0}\right] \in R^{L \times K}, \tag{21.c}$$

$$\mathbf{E}_2 = \left[\mathbf{0}, \cdots, \mathbf{0}, \boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_{K-L}\right] \in R^{(K-L) \times K}, \tag{21.d}$$

with $\mathbf{e}_i \in R^L$ $(i = 1, \cdots, L)$ and $\boldsymbol{\varepsilon}_i \in R^{(K-L)}$ $(i = 1, \cdots, K-L)$ as the elementary vectors, i.e., all the elements equal to 0 except for the *i*-th element which is 1.

Similarly, Eq. (6.b) can be re-written in the rule-ordered form

$$y_r = \frac{\underline{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_3^T \mathbf{E}_3^T \mathbf{Qw} + \bar{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_4^T \mathbf{E}_4^T \mathbf{Qw}}{\sum\nolimits_{k=1}^{L} (\mathbf{Q}\underline{\mathbf{f}})_k + \sum\nolimits_{k=L+1}^{K} (\mathbf{Q}\bar{\mathbf{f}})_k} = \boldsymbol{\psi}_r^T \mathbf{w}, \tag{22.a}$$

where

$$\boldsymbol{\psi}_r^T = [\psi_{r,1}, \cdots, \psi_{r,K}] = \frac{\underline{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_3^T \mathbf{E}_3^T \mathbf{Q} + \bar{\mathbf{f}}^T \mathbf{Q}^T \mathbf{E}_4^T \mathbf{E}_4^T \mathbf{Q}}{\sum\nolimits_{k=1}^{L} (\mathbf{Q}\underline{\mathbf{f}})_k + \sum\nolimits_{k=L+1}^{K} (\mathbf{Q}\bar{\mathbf{f}})_k} \in R^K, \tag{22.b}$$

$$\mathbf{E}_3 = \left[\mathbf{e}_1, \cdots, \mathbf{e}_R, \mathbf{0}, \cdots, \mathbf{0}\right] \in R^{R \times K}, \tag{22.c}$$

$$\mathbf{E}_4 = \left[\mathbf{0}, \cdots, \mathbf{0}, \boldsymbol{\varepsilon}_1, \cdots, \boldsymbol{\varepsilon}_{K-R}\right] \in R^{(K-R) \times K}, \tag{22.d}$$

with $\mathbf{e}_i \in R^R$ $(i = 1, \cdots, R)$ and $\boldsymbol{\varepsilon}_i \in R^{(K-R)}$ $(i = 1, \cdots, K-R)$ as the elementary vectors. Based on Eqs. (21.a) and (22.a), Eq. (7) becomes

$$y = \frac{y_l + y_r}{2} = \frac{1}{2}(\boldsymbol{\psi}_l^T + \boldsymbol{\psi}_r^T)\mathbf{w} = \frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\psi}_l + \boldsymbol{\psi}_r)_k w^k$$
$$= \frac{1}{2}\sum_{k=1}^{K}(\boldsymbol{\psi}_l + \boldsymbol{\psi}_r)_k \left(\sum_{j=0}^{d} x_j p_j^k\right),\ x_0 \equiv 1. \quad (23)$$

Furthermore, let

$$\tilde{\phi}(\mathbf{x}) = \frac{1}{2}\Big[(\psi_{l,1} + \psi_{r,1})x_0, \cdots, (\psi_{l,1} + \psi_{r,1})x_d, \cdots, (\psi_{l,K} + \psi_{r,K})x_0, \cdots, (\psi_{l,K} + \psi_{r,K})x_d\Big]^T \in R^{K(d+1)} \quad (24)$$

and with $\mathbf{p}$ defined in Eq. (17), Eq. (24) can be expressed as the linear system

$$f_{IT2TSK-FS}(\mathbf{x}) = \mathbf{p}^T \tilde{\phi}(\mathbf{x}), \quad (25)$$

and the output of the proposed classification model is given by

$$y = sign(f_{IT2TSK-FS}(\mathbf{x}) - b) = sign(\mathbf{p}^T \tilde{\phi}(\mathbf{x}) - b). \quad (26)$$

As in Stage 2, for a given training dataset $D = \{\mathbf{x}_i, y_i\}$ of a binary classification task, when the antecedents of the IT2 TSK FLS are fixed, we can construct the dataset $\tilde{D} = \{\tilde{\phi}(\mathbf{x}_i), y_i\}$ where $\tilde{\phi}(\mathbf{x}_i)$ is generated using Eq. (24). Thus, the training of the proposed classification model can be transformed into the parameter learning of the corresponding linear system. To solve for the linear system in Eq. (26), according to Eq.(8), the objective in Eq. (27) can be adopted for parameter learning by using the minimax probability decision technique

$$\max_{\mathbf{p},b,\alpha}\ \alpha$$
$$s.t.\ \inf_{\tilde{\phi}(\mathbf{x})\sim(\mathbf{u}_+^{\tilde{\phi}(\mathbf{x})}, \boldsymbol{\Sigma}_+^{\tilde{\phi}(\mathbf{x})})} pr(\mathbf{p}^T\tilde{\phi}(\mathbf{x}) - b \geq 0) \geq \alpha, \quad (27)$$
$$\inf_{\tilde{\phi}(\mathbf{x})\sim(\mathbf{u}_-^{\tilde{\phi}(\mathbf{x})}, \boldsymbol{\Sigma}_-^{\tilde{\phi}(\mathbf{x})})} pr(\mathbf{p}^T\tilde{\phi}(\mathbf{x}) - b \leq 0) \geq \alpha$$

where $\tilde{\phi}(\mathbf{x}) \sim (\mathbf{u}_+^{\tilde{\phi}(\mathbf{x})}, \boldsymbol{\Sigma}_+^{\tilde{\phi}(\mathbf{x})})$ denotes the data mapped from $\mathbf{x} \sim (\mathbf{u}_+, \boldsymbol{\Sigma}_+)$ by the fuzzy inference mechanism described with Eqs. (21-24). The solution to Eq. (27) can be obtained with the theorem below.

Theorem 3 The parameter learning of the consequents in Eq. (27) can be regarded as a special case of the classical MPM, where the training data $\mathbf{x}$ are mapped as $\tilde{\phi}(\mathbf{x}_g)$ in a new feature space constructed by the fuzzy inference rules using the strategy discussed with Eqs. (21-24).

Proof. By comparing Eq.(27) with Eqs.(8) and (11), it is clear that they are of the same form and Eq. (27) can be regarded as a special case of the minimax probability machine in [29]. The distinctive characteristic of Eq.(27) is that the training data is the mapping data in a feature space constructed by using Eqs. (21)-(24) with the corresponding type-2 fuzzy inference mechanism.

D. The MP-IT2TSK-FSC Learning Algorithm

Based on the three-stage training method described above, the corresponding MP-IT2TSK-FLSC learning algorithm is given as follows.

Algorithm: MP-IT2TSK-FSC

Stage 1:	Assign the parameters of the antecedents according to clustering techniques or other partition techniques.
Stage 2:	Initialize the parameters of the consequents using Eqs. (13)-(20).
Stage 3:	Refine the parameters of the consequents using Eqs. (21)- (27).

E. Remarks

A similar three-stage learning procedure was also used in existing interval type-2 fuzzy neural network with support vector regression algorithm (IT2FNN-SVR) [12] and type-2 fuzzy extreme learning algorithm (T2FELA) [47]. In [12], the IT2FNN-SVR algorithm first generates the antecedents by a self-evolving strategy and then learns the consequents using a two-stage SVR with the antecedents fixed. In [44], T2FELA generates the antecedents by assigning the parameters in the antecedents randomly based on the extreme learning theory, followed by the learning of the consequents using a two-stage fast learning approach for solving the corresponding matrix equations based on the extreme learning mechanism with the antecedents fixed.

It should be emphasized that the aim of the proposed method is to provide a T2 fuzzy classifier with transparent model reliability rather than improving the classification accuracy. Thus, the classification accuracy may only be at a level comparable to that of the existing methods. The advantage of the model obtained by the proposed method is the availability of a reliability index that makes it more transparent and acceptable to users in many applications.

IV. EVALUATION

The performance of the proposed MP-IT2TSK-FLSC has been evaluated on several benchmarking medical datasets and compared with the related methods. The experimental studies are organized as follows. In section IV-A and IV-B, the experiment setting and the datasets are described respectively. In section IV-C, the classification model obtained by the MP-IT2TSK-FLSC algorithm is analyzed. The results of the comparison with the related methods are reported in section IV-D.

A. Experiment Settings

1) Methods for Comparison: The MP-IT2TSK-FLSC algorithm is implemented with fuzzy c-means clustering (FCM) and self-evolving learning strategy (SE), and compared with 11 methods to evaluate the performance. Refer to the descriptions and abbreviations given in Table I, the methods include two mimimax probability decision based methods (MPM (linear) and MPM (kernel))[29], three IT2 FLS methods (T2FELA) [47], GL-IT2FLS[9], SEIT2FNN [13]), three T1 TSK-FLS based methods (SOTFN-SV [44], ε-TSK-FS [50] and L2- TSK-FS [48]), and three classical non-fuzzy systems based classification methods (KNN [51], SVC [52] and Naïve Bayes classifier [53]).

Among the algorithms, MP-IT2TSK-FLSC, MPM, SOTFN-SV, KNN, SVC and the Naïve Bayes classifier are developed for classification. The rest are originally developed for regression but can still be used for classification tasks, which is achieved in our experiments by employing a simple strategy: the class labels are directly used as the outputs of the regression datasets for the model training. When a future sample is tested, the output of the regression model is compared with different class labels and the nearest one is taken as the class label of the testing sample.

For the methods based on kernel technique, i.e., MPM (kernel) and SVC, the radius basis function (RBF) is adopted as the kernel function for its effectiveness. For all the methods related to fuzzy systems, Gaussian function is adopted as the fuzzy membership function in the antecedents.

TABLE I
METHODS ADOPTED FOR PERFORMANCE COMPARISON

Method	Description
MP-IT2TSK-FLSC (FCM) and MP-IT2TSK-FLSC (SE)	The proposed IT2 TSK FLS classifier by using minimax probability decision for classification task: MP-IT2TSK-FLSC (FCM) means the antecedents are determined by fuzzy c-means (FCM) clustering technique, and MP-IT2TSK-FLSC (SE) means the antecedents are determined by self-evolving (SE) learning strategy [13, 44].
MPM (linear) and MPM (kernel) [29]	Linear and kernel minimax probability machine by using minimax probability decision
T2FELA [47]	Type-2 fuzzy extreme learning algorithm for fast training of T2 TSK FLSs based on the extreme learning mechanism
GL-IT2FLS [9]	Gradient learning based training algorithm for IT2 FLS
SEIT2FNN [13]	Self-evolving interval type-2 fuzzy neural network
SOTFN-SV [44]	Self-organizing TS-type fuzzy network with support vector learning
ε-TSK-FS(IQP) [50]	ε-insensitive criterion based TSK-FS training method with IQP optimization technique
L2-TSK-FS [48]	L2 norm penalty and ε-insensitive criterion based TSK-FS training method
KNN [51]	K near neighbors classifier
SVC [52]	Support vector classification
Naïve Bayes classifier [53]	Naïve Bayes classifier

2) Datasets: Three medical datasets are adopted for performance evaluation. They are the *epileptic electroencephalograph (EEG)* dataset, *heart disease* dataset and *breast cancer* dataset. In our experiments, each attribute of the inputs of the data is normalized into the range [-1, 1] for the whole dataset. The details of these datasets are described in section IV-B.

3) Parameter Settings: For all the algorithms, the five-fold cross-validation (CV) strategy is used to determine the optimal setting within the given grids for the related hyper parameters. The corresponding hyper parameters in different methods and the search grids for CV are listed in Table II.

TABLE II
THE HYPER PARAMETERS IN DIFFERENT METHODS AND THE SEARCH GRIDS USED FOR CV

Method	**Description of the hyper parameters and the search grid used for CV**
MP-IT2TSK-FLSC (FCM)	Scale parameter of width in Gaussian membership function: $h \in \{10^{-5},\cdots,10^{0},\cdots,10^{5}\}$, the number of fuzzy rules: $K \in \{4,9,16,25,36,49,64,81,100,121\}$.
MP-IT2TSK-FLSC (SE)	Self-organization learning threshold parameter: $\sigma_{th} = \{0.2,0.3,0.4,0.5,0.6,0.7,0.8\}$, the number of fuzzy rules: $K \in \{4,9,16,25,36,49,64,81,100,121\}$.
MPM(linear)	No hyper parameters
MPM(kernel)	RBF kernel width parameter: $\sigma \in \{10^{-12},\cdots,10^{0},\cdots,10^{12}\}$
T2FELA	Scale parameter of width in Gaussian membership function: $h \in \{10^{-5},\cdots,10^{0},\cdots,10^{5}\}$, the number of fuzzy rules: $K \in \{4,9,16,25,36,49,64,81,100,121\}$.
GL-IT2FLS	The number of fuzzy rules: $K \in \{4,9,16,25,36,49,64,81,100,121\}$.
SEIT2FNN [12]	Self-organization learning threshold parameter: $\sigma_{th} = \{0.2,0.3,0.4,0.5,0.6,0.7,0.8\}$.
SOTFN-SV	Self-organization learning threshold parameter: $\sigma_{th} = \{0.2,0.3,0.4,0.5,0.6,0.7,0.8\}$, regularization parameter for support learning: $C \in \{10^{-12},\cdots,10^{0},\cdots,10^{12}\}$.
ε-TSK-FS(IQP) L2-TSK-FS	Scale parameter of width in Gaussian membership function: $h \in \{10^{-5},\cdots,10^{0},\cdots,10^{5}\}$, the number of fuzzy rules: $K \in \{4,9,16,25,36,49,64,81,100,121\}$, regularization parameter: $\tau \in \{10^{-12},\cdots,10^{0},\cdots,10^{12}\}$.
KNN	The number of near neighbors: $K \in \{1,\cdots,12\}$
SVC (RBF)	Regularization parameter: $C \in \{10^{-12},\cdots,10^{0},\cdots,10^{12}\}$; RBF kernel width parameter: $\sigma \in \{10^{-12},\cdots,10^{0},\cdots,10^{12}\}$.
Naïve Bayes classifier	No hyper parameters

4) Evaluation Index: The classification performance is evaluated with the index defined in Eq. (28), i.e., classification accuracy.

$$J_{clas} = \frac{\text{Number of the test samples with correct classification}}{\text{Number of the test samples}} \quad (28)$$

For performance comparison, the means and standard deviations of classification accuracies of the different methods, obtained with the optimal parameters determined by the CV strategy in the given search grids, are reported and compared.

5) Computing Environment: All the algorithms are implemented with MATLAB codes on a computer with 2GB RAM and 1.66 GHz CPU.

B. *Medical Datasets*

The three medical datasets used for performance evaluation are described as follows.

(1) *Epileptic EEG*. The epileptic EEG data used is publicly available on the web from the University of Bonn, Germany [54]. The complete data archive contains five groups of data (denoted by groups A to E), each containing 100 single channel EEG segments of 23.6 seconds duration. The sampling rate of all datasets was 173.6Hz. Group A and B consist of segments acquired from surface EEG recordings performed on five healthy volunteer subjects, while groups C, D and E are data obtained from volunteer subjects with epilepsy. In our experiment, groups A and B are referred to as the healthy class and groups C-E are the patient class. For the epileptic EEG data, feature extraction has been conducted by using short time Fourier transform (STFT). Finally, data with five features associated with the energy of different frequency sub-bands are obtained [55].

(2) *Breast cancer*. The breast cancer dataset is obtained from the UCI machine learning repository [56]. It contains 458 instances of the benign class and 241 instances of the malignant class. Each instance is described by 9 attributes.

(3) *Heart disease*. The heart disease dataset is also obtained from the UCI machine learning repository [56], which includes 120 instances with heart disease and 150 instances without heart disease. Each instance is described by 13 attributes.

C. Model Analysis of the MP-IT2TSK-FLSC

In this subsection, the trained MP-IT2TSK-FLSC model is analyzed to study the characteristics. In Table III, an MP-IT2TSK-FLSC with nine rules trained at a certain time point on the epileptic EEG dataset is presented for illustration. The constructed MP-IT2TSK-FLSC contains three parts and they are explained individually as follows.

(1) The first part is the fuzzy rules base as shown in Part A of Table III, which is used for fuzzy inference and presents a final real value as the IT2 TSK FLS output. With the fuzzy rule base, the fuzzy inference rules can be linguistically interpretable with expert knowledge. In Fig. 3, the corresponding IT2 fuzzy membership functions of each fuzzy subset in the antecedents of the 1st, 4th and 8th fuzzy rules are shown. Each membership function corresponds to an IT2 fuzzy subset which can be explained by experts using medical terms or concepts.

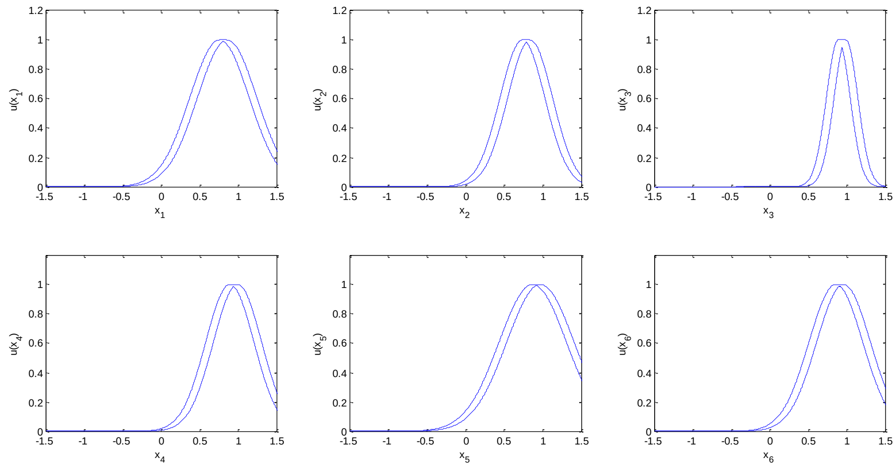
(2) The second part presents a decision threshold which is introduced for the classification task in the MP-IT2TSK-FLSC. The decision threshold and the consequents of the IT2 TSK FLS are learned based on the minimax probability decision principle. With the real output of the trained TSK-FLS and the decision threshold, the final decision can be obtained for the classification task.

(3) The last part provides the reliability of the trained model for the classification task, which characterizes the lower bound of the correct classification α.

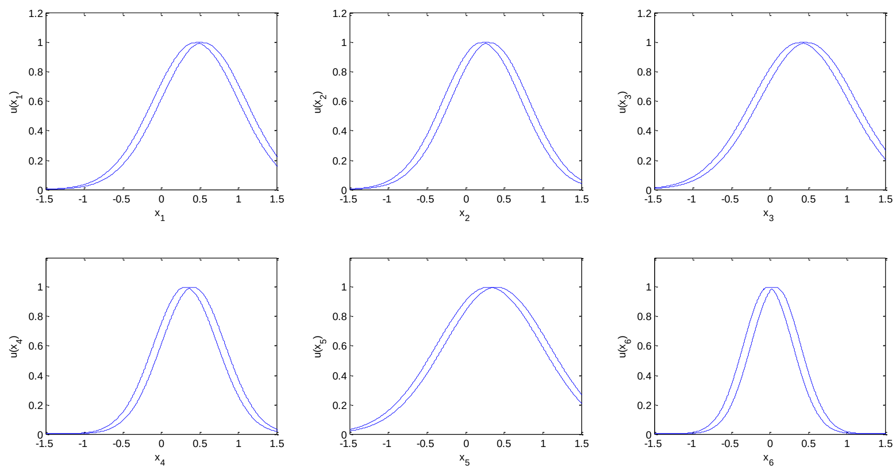
From Table III, we can see that MP-IT2TSK-FLSC is a practical expert system where *linguistically interpretable* fuzzy rules are employed and model reliability index is available, making it highly human-interpretable and very suitable for many reliability-critical applications like for medical diagnosis.

TABLE III
THE MP-IT2TSK-FLSC WITH NINE RULES TRAINED AT A CERTAIN TIME ON EPILEPTIC EEG DATASET

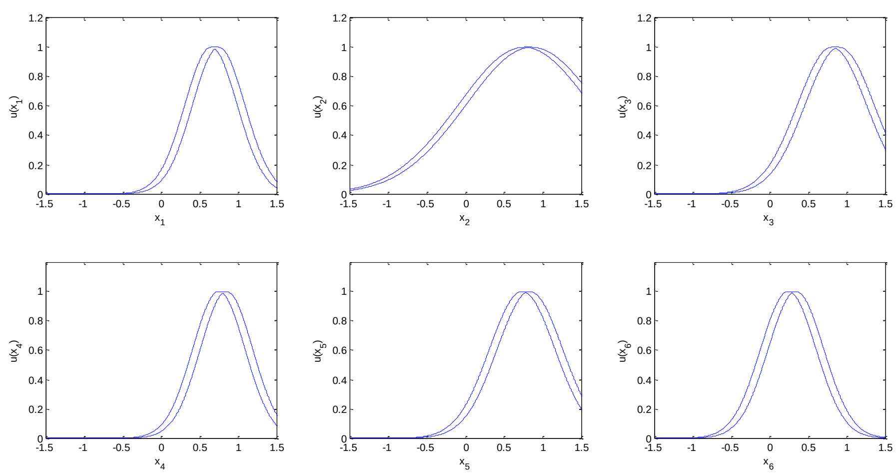
Part A: Fuzzy rules base		
TSK Fuzzy Rule R^k : IF x_1 is $A_1^k(c_1^k,\delta_1^k)$ $\wedge x_2$ is $A_2^k(c_2^k,\delta_2^k)$ $\wedge\cdots\wedge x_d$ is $A_d^k(c_d^k,\delta_d^k)$, Then $f_k(\mathbf{x}) = p_{k0} + p_{k1}x_1 + \cdots + p_{kd}x_d$.		
No. of rules	**Antecedent parameters** (Gaussian membership function parameters)	**Consequent parameters** (linear function parameters)
k	$\mathbf{c}^k = (c_1^k,\cdots,c_d^k)^T$, $\boldsymbol{\delta}^k = (\delta_1^k,\cdots,\delta_d^k)^T$	$\mathbf{p}_k = (p_{k0}, p_{k1},\cdots, p_{kd})^T$
1	$\mathbf{c}^1 =$ [0.6274, 0.6622, 0.7091, 0.5682, 0.5147, -0.4565] $\boldsymbol{\delta}^1 =$ [1.99e-05, 1.44e-05, 1.23e-05, 2.22e-05, 2.606e-05, 1.10e-05]	$\mathbf{p}_1 =$ [0.4213, -1.0270, 0.5407, -0.0451, 1.4879, -1.4751, 0.2835]
2	$\mathbf{c}^2 =$ [-0.0586, 0.7156, 0.6102, 0.3266, 0.3973, -0.2281] $\boldsymbol{\delta}^2 =$ [2.11e-05, 1.87e-05, 2.11e-05, 1.85e-05, 1.48e-05, 6.54e-06]	$\mathbf{p}_2 =$ [6.6533, -1.4751, -0.0907, 0.4213, -1.0270, 0.1655, -0.0451]
3	$\mathbf{c}^3 =$ [0.5304, -0.1475,0.4119, 0.5431, 0.9242, -0.9219] $\boldsymbol{\delta}^3 =$ [3.57e-05, 3.70e-05, 3.80e-05, 3.45e-05, 3.65e-05, 3.90e-05]	$\mathbf{p}_3 =$ [-1.0270, 0.1655, 0.3608, 1.4879, -1.4751, 0.2835, 0.4213]
4	$\mathbf{c}^4 =$ [0.4461, 0.0854, 0.2350, 0.1030, 0.0985, -0.8656] $\boldsymbol{\delta}^4 =$ [2.43e-05, 1.92e-05, 1.45e-05, 1.38e-05, 1.55e-05, 1.09e-05]	$\mathbf{p}_4 =$ [-1.4751, 0.2835, 0.4213, -1.0270, 0.1655, -0.0451, 1.4879]
5	$\mathbf{c}^5 =$ [-0.6666, -0.5807, -0.6221, -0.7431, -0.8156, -0.3132] $\boldsymbol{\delta}^5 =$ [1.62e-05, 1.201e-05, 1.44e-05, 1.40e-05, 1.11e-05, 1.88e-05]	$\mathbf{p}_5 =$ [0.1655, -0.0451, 1.4879, -1.4751, 0.2835, 0.4213, -1.0270]
6	$\mathbf{c}^6 =$ [-0.2397, -0.5931, -0.3827, -0.4711, -0.5062, -0.9506] $\boldsymbol{\delta}^6 =$ [2.74e-05, 1.78e-05, 2.47e-05, 2.23e-05, 2.50e-05, 7.49e-06]	$\mathbf{p}_6 =$ [0.2835, 0.6366, -1.0270, 0.1655, -0.0451, 1.4879, -1.4751]
7	$\mathbf{c}^7 =$ [-0.2844, 0.3527, 0.2058, -0.0151, -0.2410, -0.2229] $\boldsymbol{\delta}^7 =$ [1.62e-05, 1.602e-05, 1.31e-05,9.73e-06, 1.30e-05, 3.71e-06]	$\mathbf{p}_7 =$ [-0.0451, 1.4281, -1.4751, 0.2835, 0.4213, -1.0270, 0.1655]
8	$\mathbf{c}^8 =$ [-0.5508, -0.5821, 0.0715, -0.2276, -0.3817, 0.3099] $\boldsymbol{\delta}^8 =$ [1.60e-05, 1.39e-05, 1.94e-05, 1.08e-05, 1.40e-05, 1.30e-05]	$\mathbf{p}_8 =$ [0.4213, -0.8599, 0.1655, -0.0451, 1.4879, -1.4751, 0.2835]
9	$\mathbf{c}^9 =$ [0.5521, 0.6120, 0.9045, 0.8486, 0.8170, 0.7474] $\boldsymbol{\delta}^9 =$ [1.61e-05, 1.01e-05, 2.83e-06, 1.23e-05, 1.07e-05, 2.00e-05]	$\mathbf{p}_9 =$ [1.4879, -1.2547, 0.2835, 0.4213, -1.0270, 0.1655, -0.0451]
Part B: Decision threshold for classification		
Decision threshold of MP-IT2TSK-FLSC: b = -0.2693		
Part C: Reliability of the classification model		
Correct classification lower bound: α =0.8816		



(a)



(b)



(c)

Fig. 3 The corresponding membership functions of each fuzzy subset in the antecedents of three fuzzy rules: (a) the 1st, (b) the 4th and (c) the 8th.

D. Comparison with Related Methods

The classification performance of the proposed method is compared with the 11 methods described in section IV-A-1. Tables IV-VI present the means and standard deviations (SD) of the classification accuracies of different methods, which are obtained with the three medical datasets under the optimal parameter setting determined by the CV strategy. From these results, the following observations are revealed:

(1) The proposed MP-IT2TSK-FLSC demonstrates highly

competitive generalization abilities when compared with the exiting state-of-the-art methods.

(2) For all the fuzzy-system-based methods, i.e., four IT2 FLS based methods (MP-IT2TSK-FLSC, T2FELA, GL-IT2FLS and SEIT2FNN) and three T1 FLS based method (SOTFN-SV, ε-TSK-FS(IQP), and L2- TSK-FS), a high level of interpretation ability is demonstrated. However, only the proposed MP-IT2TSK-FLSC can specify the reliability of the trained model.

(3) Of the three minimax probability-based methods, i.e., MP-IT2TSK-FLSC, MPM (linear) and MPM (RBF), the generalization abilities of MP-IT2TSK-FLSC are better than that of MPM (linear) while comparable to that of MPM (RBF). However, MP-IT2TSK-FLSC is obviously advantageous over MPM (RBF) in that it is more transparent to the user in terms of model reliability and has a high level of interpretation ability, whereas MPM (RBF) is more like a black box that is based on a hyper plane in an unknown kernel feature space.

(4) When compared with the classical classification methods, e.g. SVC and KNN, the advantages of proposed MP-IT2TSK-FLSC is clearly demonstrated with (i) superior generalization abilities, (ii) high interpretability, and (iii) better transparency/reliability of the trained model with the lower bound of classification accuracy provided.

TABLE IV
PERFORMANCE COMPARISON WITH RELATED METHODS ON EPILEPTIC EEG DATASET

Method	MP-IT2TSK-FLSC⁺		MPM		T2FELA	GL-IT2FLS	SEIT2FNN
	(FCM)	(SE)	(linear)	(RBF)			
Mean *SD*	0.9620 0.0409	0.9640 0.0329	0.9480 0.0277	0.9600 0.0346	0.9620 0.0342	0.9560 0.0297	0.9500 0.0400
α *	0.8402 0.0192	0.8637 0.0227	0.7724 0.0105	0.8024 0.0200			
Method	SOTFN-SV		ε-TSK-FS (IQP)	L2- TSK-FS	SVC (RBF)	KNN	Naïve Bayes
Mean *SD*	0.9660 0.0296		0.9620 0.0370	0.9200 0.0430	0.9560 0.0364	0.9580 0.0311	0.9480 0.0432

* Lower bound of correct classification.

TABLE V
PERFORMANCE COMPARISON WITH RELATED METHODS ON BREAST CANCER DATASET

Method	MP-IT2TSK-FLSC		MPM		T2FELA	GL-IT2FLS	SEIT2FNN
	(FCM)	(SE)	(linear)	(RBF)			
Mean *SD*	0.9700 0.0145	0.9643 0.0131	0.9685 0.0128	0.9715 0.01665	0.9657 0.0091	0.9657 0.0094	0.9615 0.0159
α *	0.8924 0.0048	0.9089 0.0080	0.8362 0.0046	0.8576 0.0063			
Method	SOTFN-SV		ε-TSK-FS (IQP)	L2- TSK-FS	SVC (RBF)	KNN	Naïve Bayes
Mean *SD*	0.9728 0.0116		0.9629 0.0144	0.9200 0.0430	0.9560 0.0364	0.9580 0.0311	0.9480 0.0432

* Lower bound of correct classification.

TABLE VI
PERFORMANCE COMPARISON WITH RELATED METHODS ON HEART DISEASE DATASET

Method	MP-IT2TSK-FLSC		MPM		T2FELA	GL-IT2FLS	SEIT2FNN
	(FCM)	(SE)	(linear)	(RBF)			
Mean *SD*	0.8074 0.0280	0.8000 0.0275	0.8296 0.0479	0.8222 0.0465	0.8259 0.0675	0.8000 0.0479	0.7926 0.0647
α*	0.6279* 0.0245	0.6411* 0.0521	0.5515* 0.0180	0.5561* 0.0182			
Method	SOTFN-SV		ε-TSK-FS (IQP)	L2- TSK-FS	SVC (RBF)	KNN	Naïve Bayes
Mean *SD*	0.8037 0.0712		0.7852 0.0483	0.8481 0.0576	0.8222 0.0426	0.8148 0.0571	0.8222 0.0384

V. CONCLUSIONS

In this study, a minimax probability IT2 TSK FLS classifier MP-IT2TSK-FLSC is proposed to simultaneously train an IT2 fuzzy system-based classifier and to provide an index of the reliability of the trained model, i.e., the lower bound of correct classification. Experimental results show that the classifier not only exhibits high level of interpretability as other fuzzy systems, but also demonstrates the better transparency in model reliability.

Despite of the distinctive characteristics and promising performance, the proposed minimax probability classifier requires further investigation in many aspects. For example, other minimax probability decision based fuzzy system models can be studied, e.g. the ML-type fuzzy model [57]. In addition, minimax probability-based fuzzy systems can also be investigated for other modeling tasks, such as outlier detection and regression. Future work will be conducted to address these issues.

REFERENCES

[1] L.X. Wang, *Adaptive fuzzy systems and control: design and stability analysis.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1994

[2] J.S.R. Jang, C.T. Sun, and E. Mizutani, *Neuro-fuzzy and soft-computing*. Upper Saddle River, NJ: Prentice-Hall, 1997.

[3] J. Leski, "TSK-fuzzy modeling based on ε-insensitive learning," *IEEE Trans. Fuzzy Systems*, vol. 13, no.2, pp: 181-193, 2005

[4] L. A. Zadeh, "Fuzzy sets", *Inf. Control*, vol. 8, pp. 338-353, 1965.

[5] L. A. Zadeh, "Fuzzy logic, neural networks, and soft computing," *Commun. ACM*, vol. 37, pp. 77–84, 1994.

[6] Q. Liang, and J. M. Mendel, "Interval type-2 fuzzy logic systems: Theory and design," *IEEE Trans. on Fuzzy Systems*, vol. 8, pp: 535-550, 2000.

[7] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice-Hall, Upper-Saddle River, NJ, 2001

[8] J. M. Mendel, "Type-2 fuzzy sets and systems: An overview," *IEEE Computational Intelligence Magazine*, vol. 2, pp: 20-29, 2007.

[9] J. M. Mendel, "Computing derivatives in interval type-2 fuzzy logic systems," *IEEE Trans. on Fuzzy Systems*, vol.12, no. 1, pp: 84-98, 2004.

[10] Q. L. Liang and J. M. Mendel, "Designing interval type-2 fuzzy logic systems using an SVD-QR method: Rule reduction," *Int. J. Intell. Syst.* vol.15, no. 10, pp: 939-957, 2000.

[11] C. H. Wang, C. S. Cheng, and T. T. Lee, "Dynamical optimal training for interval type-2 fuzzy neural network (T2FNN)," *IEEE Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 3, pp: 1462-1477, 2004.

[12] C. H. Juang, R. B. Huang, and W. Y. Cheng, "An interval type-2 fuzzy-neural network with support-vector regression for noisy regression problems," *IEEE Trans. on Fuzzy Systems*, vol. 18, no. 4, pp:686-699, 2010.

[13] C. F. Juang and Y. W. Tsao, "A self-evolving interval type-2 fuzzy neural network with online structure and parameter learning," *IEEE Trans. on Fuzzy Systems*, vol. 6, no. 6, pp: 1411-1424, 2008.

[14] C. F. Juang, Y. W. Tsao, "A Type-2 self-organizing neural fuzzy system and its FPGA implementation," *IEEE Trans. on Syst., Man, and Cybern., Part B*, vol. 38, no. 6, pp: 1537-1548, 2008.

[15] G. M. Mendez, "Orthogonal-back propagation hybrid learning algorithm for type-2 fuzzy logic systems," *IEEE Proceedings of the NAFIPS 04 International Conference on Fuzzy Sets*, 2: 899-902, 2004.

[16] J. R. Castro, O. Castillo, P. Melin, A. R. Díaz, "A hybrid learning algorithm for a class of interval type-2 fuzzy neural networks," *Inf. Sci.*, vol. 179(13), pp. 2175-2193, 2009.

[17] W. H. R. Jeng, C. Y. Yeh, S. J. Lee, "General type-2 fuzzy neural network with hybrid learning for function approximation," Proc. the IEEE Conference on Fuzzy Systems, Jeju, Korea, 2009, pp. 1534–1539.

[18] G.. M. Méndez and M. A. Hernandez, "Hybrid learning mechanism for interval A2-C1 type-2 non-singleton type-2 Takagi-Sugeno-Kang fuzzy logic systems," *Information Sciences*, vol. 220, pp. 149-169, 2013

[19] G. M. Méndez, A. Hernandez, A. Cavazos, M. T. Mata, "Type-1 non-singleton type-2 Takagi-Sugeno-Kang fuzzy logic systems using the hybrid mechanism composed by a Kalman type filter and back propagation methods," *Lecture Notes in Artificial Intelligence*, HAIS 2010, pp. 429-437, Volume 6076, 23 June 2010.

[20] G. M. Méndez, M. De los Angeles Hernandez, "Hybrid learning for interval type-2 fuzzy logic systems based on orthogonal least-squares and back-propagation methods," *Information Sciences*, vol. 179, no. 13, pp. 2146-2157, 2009.

[21] G. M. Méndez, M. De los Angeles Hernandez, "Interval type-2 non-singleton type-2 Takagi-Sugeno-Kang fuzzy logic systems using the hybrid learning mechanism recursive-least-square and back-propagation methods," In the 11th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 710-714, 2010.

[22] O. Castillo, P. Melin, "Optimization of type-2 fuzzy systems based on bio-inspired methods: A concise review," *Inf. Sci.*, vol. 205, pp. 1-19, 2012.

[23] Y. Maldonado, O. Castillo, P. Melin, "Particle swarm optimization of interval type-2 fuzzy systems for FPGA applications," *Appl. Soft Comput.*, vol. 13, no. 1, pp. 496-508, 2013.

[24] D. Hidalgo, P. Melin, O. Castillo, "An optimization method for designing type-2 fuzzy inference systems based on the footprint of uncertainty using genetic algorithms," *Expert Syst. Appl.*, vol. 39, no. 4, pp. 4590-4598, 2012.

[25] O. Castillo, P. Melin, A. A. Garza, O. Montiel, R. Sepúlveda, "Optimization of interval type-2 fuzzy logic controllers using evolutionary algorithms," *Soft Comput.*, vol. 15, no. 6, pp. 1145-1160, 2011.

[26] R. A. Aliev, W. Pedrycz, B.G. Guirimov, R.R. Aliev, U. Ilhan, M. Babagil, S. Mammadli, "Type-2 fuzzy neural networks with fuzzy clustering and differential evolution optimization," *Inf. Sci.*, vol. 181, no. 9, pp. 1591–1608, 2011.

[27] R. Martinez, A. Rodriguez, O. Castillo, L.T. Aguilar, "Type-2 fuzzy logic controllers optimization using genetic algorithms and particle swarm optimization," *Proc. the IEEE International Conference on Granular Computing*, GrC 2010, 2010, pp. 724–727.

[28] G.R.G. Lanckriet, L.E. Ghaoui, and M.I. Jordan, "Robust novelty detection with single-class MPM," *in: Advances in Neural Information Processing Systems* 15, Cambridge, MA:MIT Press, 2002.

[29] G.R.G. Lanckriet, L.E. Ghaoui, C. Bhattacharyya and M.I. Jordan, A robust minimax approach to classification, *Journal of Machine Learning Research*, 2003, vol. 3, pp. 555–582.

[30] K.Z. Huang, H.Q. Yang, I. King, M. R. Lyu, and L. Chan, "Minimum error minimax probability machine," *Journal of Machine Learning Research*, vol.5, pp.1253-1286, 2004.

[31] K.Z. Huang, H.Q. Yang, I. King, and M.R. Lyu, "Imbalanced learning with biased minimax probability machine," *IEEE Trans. on System, Man, and Cybernetics–Part B*, vol.36, no.4, pp.913-923, 2006.

[32] T.R. Strohmann, A. Belitski, G.Z. Grudic, and D.M. DeCoste, "Sparse greedy minimax probability machine classification," Proc. of Neural Information Processing (NIPS'2003), 2003.

[33] Z.H Deng, S.T Wang, F.L Chung, "A minimax probabilistic approach to feature transformation for multi-class data," *Appl. Soft Comput.* vol.13, no. 1, pp. 116-127, 2013.

[34] T. Strohmann and G.Z. Grudic, "A formulation for minimax probability machine regression," *Neural Information Processing Systems: Natural and Synthetic*. 2003.

[35] N. N. Karnik and J. M. Mendel, "Centroid of a type-2 fuzzy set," *Inf. Sci.*, vol. 132, pp. 195–220, 2001.

[36] D. Wu and J. M. Mendel, "Enhanced Karnik-Mendel Algorithms," *IEEE Trans. Fuzzy Systems*, vol. 17, pp. 923-934, 2009.

[37] D. Wu and M. Nie, "Comparison and Practical Implementation of Type-Reduction Algorithms for Type-2 Fuzzy Sets and Systems," *IEEE International Conference on Fuzzy Systems*, Taipei, Taiwan, June 2011.

[38] F. Alizadeh and D. Goldfarb, "Second-order cone programming," *Mathematical Programming*, vol.95, no.1, pp.3-51, 2004.

[39] J. Sturm, "Using SeDuMi 1.02, A MATLAB toolbox for optimization over symmetric cones," Special issue on Interior Point Methods, *Optimization Methods and Software*, vol.11-12, pp.625-653, 1999.

[40] H. Ishibuchi, T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Systems*, 2001, vol. 9, no. 4, pp. 506-515, 2001.

[41] H. Ishibuchi and Y. Takashi, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Systems*, vol. 13, no. 4, pp. 428-435, 2005.

[42] K. Nozaki, H. Ishibuchi, and H. Tanaka, "Adaptive fuzzy rule-based classification systems," *IEEE Trans. Fuzzy Systems*, vol. 4, no. 3, pp. 238-250, 1996.

[43] H. Ishibuchi, and T. Nakaskima, "Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes," *IEEE Trans. Industrial Electronics*, vol. 46, no. 6, 1057-1068, 1999.

[44] C.F. Juang, S.H. Chiu, and S.W. Chang, "A self-organizing TS-type fuzzy network with support vector learning and its application to classification problems," *IEEE Trans. Fuzzy Systems*, vol. 15, no. 5, pp. 998-1008, 2007.

[45] C.F. Juang, S.J. Shiu, "Using self-organizing fuzzy network with support vector learning for face detection in color images," *Neurocomputing*, vol, 71, no. 16, pp. 3409-3420, 2008.

[46] G.D. Wu and P. H. Huang, "A maximizing-discriminability-based self-organizing fuzzy network for classification problems," *IEEE Trans. Fuzzy Systems*, vol. 18, no. 2, pp. 362-373, 2010.

[47] Z.H. Deng, K.S. Choi, L.B. Cao, S.T. Wang, "T2FELA: Type-2 Fuzzy Extreme Learning Algorithm for Fast Training of Interval Type-2 TSK Fuzzy Logic System," *IEEE Trans. Neural Networks and Learning Systems*, DOI:10.1109/TNNLS.2013.2280171

[48] Z.H. Deng, K.S. Choi, F.L. Chung, S.T. Wang, "Scalable TSK fuzzy modeling for very large datasets using minimal-enclosing-ball approximation," *IEEE Trans. Fuzzy Systems*, vol. 19, no.2, pp.210-226, 2011.

[49] Z.H. Deng, Y.Z. Jiang, K.S. Choi, F.L. Chung, S.T. Wang,

"Knowledge-Leverage-Based TSK Fuzzy System Modeling," IEEE Trans. Neural Networks and Learning Systems, vol. 24, no. 8, pp. 1200-1212, 2013.

[50] J. Leski, "TSK-fuzzy modeling based on ε-insensitive learning," IEEE Trans. Fuzzy Systems, vol. 13, no.2, pp. 181-193, 2005

[51] R. Duda, P.Hart, and D. Stork, *Pattern Classification*. NewYork:Wiley, 2000.

[52] C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[53] D. Grossman, and P. Domingos, "Learning Bayesian network classifiers by maximizing conditional likelihood," *In Proceedings of the twenty-first international conference on Machine learning*, pp. 46. ACM, 2004.

[54] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6: 061907, 2001.

[55] K.S. Choi, Y. Zeng, J. Qin, "Using sequential floating forward selection algorithm to detect epileptic seizure in EEG signals," *The 11th International Conference on Signal Processing*, Beijing, China, 21-25 October, 2012.

[56] K. Bache, and M. Lichman, UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.

[57] E.H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," IEEE Trans. Computer, vol. C-26 (12), pp. 1182- 1191, 1977.

[58] E. Lughofer and O. Buchtala, "Reliable all-pairs evolving fuzzy classifiers," IEEE Transactions on Fuzzy Systems, vol. 21, no. 4, pp. 625-641, 2013.

[59] E. Lughofer, "Single-pass active learning with conflict and ignorance," Evolving Systems, vol. 3, no. 4, pp. 251-271, 2012.