An Improved Fuzzy Min–Max Neural Network for Data Classification

Santhos Kumar A., Anil Kumar, *Member, IEEE*, Varun Bajaj, *Member, IEEE*, and Girish Kumar Singh

***Abstract*—Hyperbox classifier is an efficient tool for modern pattern classification problems due to its transparency and rigorous use of Euclidian geometry. Fuzzy min–max (FMM) network efficiently implements the hyperbox classifier, and has been modified several times to yield better classification accuracy. However, the obtained accuracy is not up to the mark. Therefore, in this paper, a new improved FMM (IFMM) network is proposed to increase the accuracy rate. In the proposed IFMM network, a modified constraint is employed to check the expandability of a hyperbox. It also uses semiperimeter of the hyperbox along with *k*-nearest mechanism to select the expandable hyperbox. In the proposed IFMM, the contraction rules of conventional FMM and enhanced FMM (EFMM) are also modified using semiperimeter of a hyperbox in order to balance the size of both overlapped hyperboxes. Experimental results show that the proposed IFMM network outperforms the FMM, *k*-nearest FMM, and EFMM by yielding more accuracy rate with less number of hyperboxes. The proposed methods are also applied to histopathological images to know the best magnification factor for classification.**

***Index Terms*—Enhanced fuzzy min–max (EFMM) model, fuzzy min–max (FMM) neural network, histopathological images, hyperbox classifier, pattern classification, semiperimeter.**

I. Introduction

An ARTIFICIAL neural network (ANN) is a computational model, consisting of an interconnected group of artificial neurons, similar to the biological neural network of a human brain [1]. In the past, ANNs have received considerable attention in several engineering fields like power systems [2], health care [3], robotics [4], financial economics [5], fault detection [6], human recognition [7]–[9], arrhythmias classification [10], and pattern classification [11]–[13]. Learning methods based on ANN models, like radial basis function (RBF) networks and multilayer perceptron networks, suffer from the catastrophic forgetting problem [14], [15] that results in information loss, when new information is added to the learned system. This phenomenon is also known as a stability–plasticity dilemma, which occurs due to a quick response with respect to a changing environment [16]. To understand the problem of stability–plasticity dilemma in detail, the back-propagation ANN model was examined, and authors have interpreted that when multiple pieces of information are given as input to a network, then the network finds a new solution [14], which is different from the brain characteristics. The aim of this dilemma is to address different problems associated with ANN, such as flexibility of a learning system to learn new information, and stability to contain the previously learned information, when new information is learned [16]. During the past, a number of ANN models, such as fuzzy min–max (FMM) networks [17], [18] and adaptive resonance theory (ART) [19], [20] were developed to resolve the stability–plasticity dilemma. The FMM neural network has been widely used due to its simple operation and more flexible boundary. In recent years, many researchers have applied FMM to variety of applications. An FMM can be applied to medical diagnosis [21], ECG beat analysis [22], face detection [23], fault classification [24], and color recognition [25], etc.

With the invention of fuzzy set theory [26], fuzzy sets have been extensively applied to pattern recognition and classification problems. Since then, many researchers exploited the ANN learning ability and inference properties of fuzzy systems to design the intelligent systems [27], and were further modified in [28]–[36]. Juang *et al.* [32] used a fuzzy neural network (FNN) along with a graphic processing unit to reduce the FNN training time by using multithreaded CPU processes, instead of using single-threaded CPU processes. An improved generalized ART (IGART) model has been introduced in [33], in which the Laplacian-likelihood function was used to provide a new description of vigilance function for ART. An IGART model also uses the rule extraction capabilities and network pruning to solve the real world problems effectively. A hybrid learning technique to train the fuzzy wavelet neural networks was proposed, based on adaptive back propagation methods, the recursive least squares method, and a clustering model [34]. Pratama *et al.* [35] proposed a generic evolving neuro-fuzzy inference system network, in which there is a tradeoff between the niggardly rule and high predictive accuracy, while Lin *et al.* [36] proposed an interactively recurrent self-evolving FNN (IRSFNN) to predict and identify the dynamic systems.

The FMM contains a number of hyperboxes, where each hyperbox represents some portion in the n-dimensional pattern space. Depending on the incoming data samples, an FMM creates new hyperboxes. FMM also associates a hyperbox with new classes or refine an existing hyperbox without retraining it. Several variants of FMM, such as the general fuzzy min–max

Manuscript received September 16, 2018; revised January 19, 2019 and May 7, 2019; accepted June 13, 2019. Date of publication June 21, 2019; date of current version September 1, 2020. *(Corresponding author: Santhos Kumar A.)*

S. Kumar A., A. Kumar, and V. Bajaj are with the Pandit Dwarka Prasad Mishra (PDPM) Indian Institute of Information Technology, Design, and Manufacturing Jabalpur, Jabalpur 482005, India (e-mail: hymavathisanthu@gmail.com; anilkdee@gmail.com; bajajvarun056@yahoo.co.in).

G. K. Singh is with the Indian Institute of Technology Roorkee, Roorkee 247667, India (e-mail: gksngfee@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2019.2924396

(GFMM) model, general reflex FMM (GRFMM) model, adaptive resolution classifier (ARC), and pruned ARC (PRAC) have been developed to yield better results. The GFMM model has been developed [37] to handle the unlabeled and labeled data, which are developed on the basis of principle of expansion and contraction. Supervised and unsupervised learning strategies can be combined within a single method by using a GFMM network [37]. It is also useful to avoid the uncertainty limit in processing input data and retraining. A GRFMM model was developed by Nandedkar and Biswas [38], which was based on FMM clustering and classification techniques with a human reflex mechanism concept. Furthermore, Likas [39] modified the reinforcement FMM model to yield a stochastic FMM model, which uses random hyperboxes. Thereon, ARC and PRAC were introduced, and were combined to obtain the adaptive resolution min–max network [40]. However, in these models, no restriction was applied to the size of a hyperbox during the expansion rule.

Bargiela *et al.* [41] proposed a fuzzy hyperbox classifier, which has two types of hyperboxes, i.e., inclusion and exclusion hyperboxes. Inclusion hyperboxes contain input patterns of a single class, whereas the exclusion hyperboxes contain input patterns of more than one class. The exclusion hyperbox reduces the training process from three steps to two steps, i.e., expansion, overlap test, and contraction to expansion and overlap test. Kim and Yang [42] proposed a weighted FMM network, which applies no restriction to the size of hyperbox, during expansion. In learning, this model exploits the feature distribution information to reduce the hyperbox distortion caused by removing the overlapping area of hyperbox during the contraction process.

The fuzzy min–max neural network classifier with compensatory neurons was introduced in [43], which is a supervised classification model that supports online learning. A data-core-based FMM neural network was introduced in [44], which uses different membership functions for classifying and overlapping neurons. In this methodology, membership functions were designed by considering geometric center of the hyperbox, noise, and data core. To improve the FMM classification performance, a modified FMM (MFMM) was proposed in [45]. An MFMM solves the problem associated with the smaller number of larger hyperboxes. To increase the performance of an MFMM, a fuzzy min–max neural network with a genetic-algorithm [46] was proposed, which uses the genetic-algorithm-based rule extractor for data classification. By combining FMM and classification and regression tree (CART), a hybrid model known as FMM-CART [47] was proposed that uses the offline learning technique. To perform online learning, FMM-CART was further enhanced in [48]. Mohammed and Lim [49] proposed a new efficient k-nearest expansion rule for FMM to reduce the network complexity. In this method, the k-nearest expansion rule avoids the formation of too many small hyperboxes. An enhanced FMM (EFMM) was proposed in [50], which defines a new set of hyperbox expansion rules that reduce the overlapping rate during expansion. The existing overlap test rules of an FMM are extended in an EFMM to know new overlapping cases. To resolve the abovementioned new overlapping test rules, an EFMM defines new contraction rules. The EFMM was further improved in [51], where authors have used a pruning strategy to eliminate the less efficient hyperboxes.

FMM, k-nearest FMM (k-NFMM), and EFMM have a number of inbuilt limitations. For example, the expansion process of FMM and k-NFMM increases the overlapping region between hyperboxes of different classes, whereas the expansion process of EFMM creates a large number of small-sized hyperboxes. During the expansion process, when more than one hyperbox has the same largest membership function value, FMM, k-NFMM, and EFMM randomly select a hyperbox to be checked for expansion for including the current training pattern. The contraction processes of FMM, k-NFMM, and EFMM are sometimes biased toward a particular hyperbox and sometimes the overlapping region is equally distributed among two hyperboxes of different classes, which is inappropriate. The abovementioned limitations have motivated the authors to propose an improved FMM (IFMM).

In order to accurately classify the testing patterns, two IFMM networks, namely IFMM_IV and IFMM_IX, are proposed in this paper. The main contributions of the paper are as follows:

1) k-nearest-neighbor and semiperimeter based combined strategy to select a candidate hyperbox to include the current training pattern;
2) a weighted formula to check the expandability of candidate hyperbox;
3) a new set of weighted contraction rules to remove the overlapping regions between two hyperboxes of different class.

It is worth mentioning that the proposed methods are a refinement of FMM, k-NFMM, and EFMM. While considering the inclusion of a new training pattern, the above methods randomly select probable hyperboxes for expansion when more than one hyperbox has the same largest membership function value. However, this random selection is illogical. Hence, the proposed methods use the semiperimeter of the hyperboxes to break the tie. The abovementioned state-of-the-art methods suffer from an erroneous expandability test; hence, the proposed methods use an experiment-motivated condition to check the expandability of a hyperbox. The contraction rules of the existing methods improperly resolve the overlapping regions corresponding to the respective overlap test rules. In order to accurately perform a contraction, the proposed methods incorporate the semiperimeter of hyperboxes. The proposed methods also evenly divide the overlapping regions between the hyperboxes. In order to compare the performance of proposed methods with the state-of-the-art methods, seven benchmark datasets are used. Experimental results conclude that the proposed methods provide a more accurate rate than those of the state-of-the-art methods. Simultaneously, the proposed methods generate very less number of hyperboxes.

The rest of this paper is presented as follows. In Section II, FMM, k-NFMM, and EFMM learning methods are clearly explained. Section III presents the proposed methodology. Section IV shows the experimental results. Finally, Section V concludes this paper.

II. FMM, k-NFMM, AND EFMM NETWORKS

In this section, the classical FMM [17], k-NFMM [49], and EFMM [50] networks are discussed in detail.

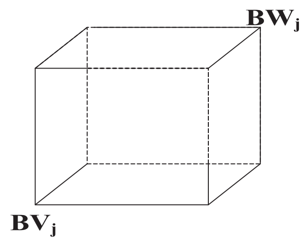

Fig. 1. Three-dimensional (3-D) hyperbox.

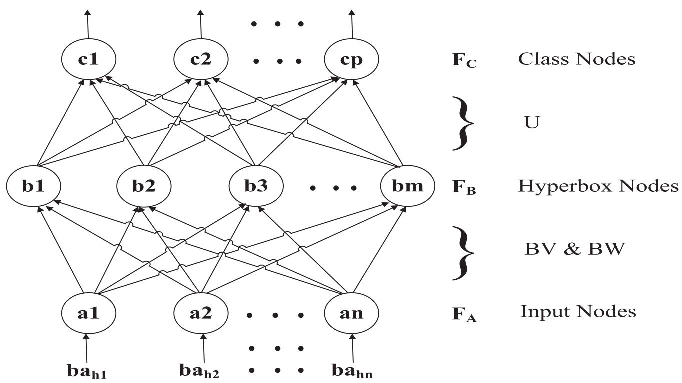

Fig. 2. Three-layer FMM network.

A. Fuzzy Min–Max

The concept of a hyperbox was first introduced by Simpson for pattern classification [17]. Every hyperbox is indicated by a set of minimum and maximum points in an *n*-dimensional space within a unit cube (I^n) along with a membership function, where *n* is the number of features per input pattern. The size of a hyperbox is maintained by its maximum and minimum points, which are regulated by a predefined expansion–overlap–contraction mechanism. The learning algorithm of FMM includes a three-step procedure, viz., expansion, overlap test, and contraction process. The FMM learning starts by using input patterns T_h, *h* = 1,…, *M*, where *M* represents the total number of data samples and a set of target classes *Ci*, *i* = *1*,…, *k* where *k* is the total number of classes. FMM generates hyperboxes on the basis of data samples. Fig. 1 shows a three-dimensional (3-D) hyperbox with its maximum point (BW_j) and minimum point (BV_j).

Fuzzy set for the *j*th hyperbox, HB_j, can be represented using

$$\mathrm{HB}_j = (\{T_h, \mathrm{BV}_j, \mathrm{BW}_j,\ f((T_h,\ \mathrm{BV}_j,\ \mathrm{BW}_j))\})\ \forall\, T_h \in I^n \tag{1}$$

where $T_h = (t_{h1}, t_{h2}, \ldots, t_{hn})$ represents the *h*th input pattern and $\mathrm{BV}_j = (\mathrm{bv}_{j1}, \mathrm{bv}_{j2}, \ldots, \mathrm{bv}_{jn}), \mathrm{BW}_j = (\mathrm{bw}_{j1}, \mathrm{bw}_{jn}, \ldots, \mathrm{bw}_{jn})$ indicates the minimum and maximum points of *j*th hyperbox, respectively.

When an input pattern is contained in a hyperbox, it is said that the input pattern has a full class membership of the corresponding hyperbox. The user-defined parameter called expansion coefficient (θ) controls the size of hyperbox. The structure of FMM consists of three layers, as shown in Fig. 2. F_A, F_B, and F_C are the input, hyperbox, and output layers, respectively. Each input node of the input layer corresponds to an input feature. Each hyperbox layer node indicates a hyperbox fuzzy set, generated during the learning process. The connections between the input layer and hyperbox layer nodes are represented using the minimum and maximum points, which are stored in two matrices BV and BW, respectively. The hyperbox membership function is defined in (3), which is the F_Bs transfer function [17]. The connections between the hyperbox layer and output layer nodes are represented using (2). The number of nodes of F_C is equal to the number of target classes. Each output of an output layer node indicates the degree to which T_h fits within the corresponding output class. The connections between the hyperbox layer and output layer nodes are binary values, and are stored in matrix *U*, as defined by

$$u_{jk} = \begin{cases} 1, & \text{if } \mathrm{b}_j \text{ is hyperbox for class} \mathrm{C}_k \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where b_j represents the *j*th hyperbox layer node and C_k indicates the *k*th output layer node. For a soft decision, the outputs can be directly employed. The winner-take-all principle [52] can be used to select the output layer node with the highest value as the expected class when a hard decision is required.

An FMM uses the membership function (MF), when a new training pattern is provided, the membership value of hyperbox HB_j for an input pattern (T_h) is computed using

$$\mathrm{HB}_j\,(T_h) = \frac{1}{2n}\sum_{i=1}^{n} \times \begin{bmatrix} \max(0, 1 - \max(0, \gamma\min(1, t_{hi} - \mathrm{BW}_{ji}))) \\ + \max(0, 1 - \max(0, \gamma\min(1, BV_{ji} - t_{hi}))) \end{bmatrix} \tag{3}$$

where T_h is the *h*th input pattern, BV_{ji} and BW_{ji} are the minimum and maximum points of HB_j, respectively, and γ is a constant that represents the sensitivity parameter, which regulates how fast the membership value decreases as the distance between T_h and HB_j increases.

The learning algorithm used by Simpson for classification using the concept of an FMM network mainly consists of expansion–overlap–contraction steps.

1) *Expansion criteria:* Simpson proposed the following equation to test whether a hyperbox can be expanded to include a new input pattern BA_h:

$$n\theta \geq \sum_{i=1}^{n} (\max(\mathrm{BW}_{ji},\ t_{hi}) - \min(\mathrm{BV}_{ji},\ t_{hi})) \tag{4}$$

where BW_{ji} and BV_{ji} are the maximum and minimum points of hyperbox, respectively. t_{hi} is the input pattern and θ is the expansion coefficient and its range is $0 \leq \theta \leq 1$. If (1) is satisfied by a hyperbox HB_j for an input pattern T_h, then HB_j can be expanded to include the input pattern BA_h.

2) *Overlap test rule:* Once a hyperbox is selected for expansion, the overlap test is performed to know the existence of overlapping between two or more hyperboxes occurred by the abovementioned expansion rule. Overlapping of hyperboxes will be there if any of the following case is satisfied:

Case 1:

$$\begin{aligned} &\mathrm{BV}_{ji} < \mathrm{BV}_{ki} < \mathrm{BW}_{ji} < \mathrm{BW}_{ki}, \\ &\quad \delta n = \min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \delta o). \end{aligned} \tag{5}$$

Case 2:

$$\mathrm{BV}_{ki} < \mathrm{BV}_{ji} < \mathrm{BW}_{ki} < \mathrm{BW}_{ji}$$
$$\delta n = \min(\mathrm{BW}_{ki} - \mathrm{BV}_{ji},\ \delta o). \quad (6)$$

Case 3:

$$\mathrm{BV}_{ji} < \mathrm{BV}_{ki} < \mathrm{BW}_{ki} < \mathrm{BW}_{ji},$$
$$\delta n = \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (7)$$

Case 4:

$$\mathrm{BV}_{ki} < \mathrm{BV}_{ji} < \mathrm{BW}_{ji} < \mathrm{BW}_{ki},$$
$$\delta n = \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (8)$$

Initially, set $\delta o = 1$. If $\delta o - \delta n > 0$, set $\Delta = i$, $\delta o = \delta n$, and proceed to check the existence of an overlapping region for the next dimension. Otherwise, set $\Delta = -1$ and the testing stops.

3) *Contraction rule:* If an overlap is detected, then an appropriate contraction rule is applied to remove the overlap between the hyperboxes. The corresponding contraction rules with respect to overlap test rules as mentioned in overlap test are as follows.

Case 1:

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta}, \mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \frac{\mathrm{BW}_{j\Delta}^{\mathrm{old}} + \mathrm{BV}_{k\Delta}^{\mathrm{old}}}{2}. \quad (9)$$

Case 2:

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}, \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \frac{\mathrm{BW}_{k\Delta}^{\mathrm{old}} + \mathrm{BV}_{j\Delta}^{\mathrm{old}}}{2}. \quad (10)$$

Case 3(a):

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}$$
$$(\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) < (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{old}}. \quad (11)$$

Case 3(b):

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}$$
$$(\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) > (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{old}}. \quad (12)$$

Case 4(a):

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta} \quad (1)$$
$$(\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) < (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{old}}. \quad (13)$$

Case 4(b):

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta} \quad (2)$$
$$(\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) > (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \mathrm{BW}_{j\Delta}^{\mathrm{old}}. \quad (14)$$

After the abovementioned three processes, the training process is completed, and hence, a list of hyperboxes is obtained to represent the FMM network. There are so many modifications available to overcome certain drawbacks of the FMM network.

B. *k-Nearest Fuzzy Min–Max*

Mohammed *et al.* [49] proposed a modified FMM model, *k*-NFMM, by applying the *k*-nearest neighbor mechanism to the classical FMM model. Unlike the FMM model, which only considers the hyperbox with largest MF value, *k*-NFMM have *k* number of hyperboxes with *k* largest MF values. In the case of FMM, if the hyperbox (only one) with the largest MF value fails the expansion test, a new hyperbox is created to include the training patterns. However, in the case of *k*-NFMM, if the hyperbox with the highest MF value fails the expansion test, then the hyperbox with the next highest MF value is checked for expansion. This process is continued up to *k*-number of hyperboxes corresponding to *k* highest MF values. If all the *k* hyperboxes fails the expandability test, then a new hyperbox is created to include the current training pattern.

C. *Enhanced Fuzzy Min–Max*

To solve the various problems in the FMM network, Mohammed and Lim [50], proposed EFMM that modifies FMM in the following three ways.

1) *Expansion criteria:* Classical FMM network uses (4) to check whether a hyperbox is able to expand or not. The abovementioned expansion creates some overlapping hyperboxes. This overlapping occurs since the sum of all dimensions is compared with $n\theta$ to determine whether or not a hyperbox can be expanded to include a particular feature vector. To overcome the abovementioned problem, EFMM uses the following to check the expandability of a hyperbox:

$$\mathrm{Max}_n(\mathrm{BW}_{ji},\ t_{hi}) - \mathrm{Min}_n(\mathrm{BV}_{ji},\ t_{hi}) \le \theta. \quad (15)$$

The advantage of (15) is that it compares each individual dimension with the expansion coefficient (θ), as a result the overlapping of hyperbox is minimized.

2) *Overlap test rule:* The overlapping test rules used by FMM are insufficient to detect each type of overlapping that exists between two hyperboxes of different classes. To resolve the above issue, EFMM uses the following set of overlap test rules:

Case 1:

$$\mathrm{BV}_{ji} < \mathrm{BV}_{ki} < \mathrm{BW}_{ji} < \mathrm{BW}_{ki}$$
$$\delta n = \min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \delta o). \quad (16)$$

Case 2:

$$\mathrm{BV}_{ki} < \mathrm{BV}_{ji} < \mathrm{BW}_{ki} < \mathrm{BW}_{ji}$$
$$\delta n = \min(\mathrm{BW}_{ki} - \mathrm{BV}_{ji},\ \delta o). \quad (17)$$

Case 3:

$$\mathrm{BV}_{ji} = \mathrm{BV}_{ki} < \mathrm{BW}_{ji} < \mathrm{BW}_{ki}$$
$$\delta n = \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (18)$$

Case 4:

$$\mathrm{BV}_{ji} < \mathrm{BV}_{ki} < \mathrm{BW}_{ji} = \mathrm{BW}_{ki}$$
$$\delta n = \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (19)$$

Case 5:

$$\begin{aligned}\mathrm{BV}_{ki} &= \mathrm{BV}_{ji} < \mathrm{BW}_{ki} < \mathrm{BW}_{ji}\\ \delta n &= \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (20)\end{aligned}$$

Case 6:

$$\begin{aligned}\mathrm{BV}_{ki} &< \mathrm{BV}_{ji} < \mathrm{BW}_{ki} = \mathrm{BW}_{ji}\\ \delta n &= \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (21)\end{aligned}$$

Case 7:

$$\begin{aligned}\mathrm{BV}_{ji} &< \mathrm{BV}_{ki} \le \mathrm{BW}_{ki} < \mathrm{BW}_{ji}\\ \delta n &= \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (22)\end{aligned}$$

Case 8:

$$\begin{aligned}&\mathrm{BV}_{ki} < \mathrm{BV}_{ji} \le \mathrm{BW}_{ji} < \mathrm{BW}_{ki}\\ &\delta n = \min(\min(\mathrm{BW}_{ji} - \mathrm{BV}_{ki},\ \mathrm{BW}_{ki} - \mathrm{BV}_{ji}),\ \delta o). \quad (23)\end{aligned}$$

Case 9:

$$\begin{aligned}\mathrm{BV}_{ki} = \mathrm{BV}_{ji} < \mathrm{BW}_{ki} = \mathrm{BW}_{ji}\\ \delta n = \min(\mathrm{BW}_{ki} - \mathrm{BV}_{ji},\ \delta o). \quad (24)\end{aligned}$$

First two overlap test rules are taken from FMM (5) and (6). Initial set $\delta o = 1$, for each dimension *i* of an *n*-dimensional feature vector, an overlapping region will be noticed if $\delta o - \delta n < 1$. Then, set $\Delta = i$, $\delta o = \delta n$ and proceed to check the existence of an overlapping region for the next dimension $i + 1$. This process stops, when $\delta o - \delta n$ becomes 1.

3) *Contraction rule*: An EFMM uses the following contraction rules that correspond to each overlap test rule, as described in Section II-C, respectively

Case 1:

$$\begin{aligned}\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta}, \mathrm{BW}_{j\Delta}^{\mathrm{new}}\\ = \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \frac{\mathrm{BW}_{j\Delta}^{\mathrm{old}} + \mathrm{BV}_{k\Delta}^{\mathrm{old}}}{2}. \quad (25)\end{aligned}$$

Case 2:

$$\begin{aligned}\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}, \mathrm{BW}_{k\Delta}^{\mathrm{new}}\\ = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \frac{\mathrm{BW}_{k\Delta}^{\mathrm{old}} + \mathrm{BV}_{j\Delta}^{\mathrm{old}}}{2}. \quad (26)\end{aligned}$$

Case 3:

$$\mathrm{BV}_{j\Delta} = \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta},\ \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \mathrm{BW}_{j\Delta}^{\mathrm{old}}. \quad (27)$$

Case 4:

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} = \mathrm{BW}_{k\Delta},\ \mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{old}}. \quad (28)$$

Case 5:

$$\mathrm{BV}_{k\Delta} = \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta},\ \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{old}}. \quad (29)$$

Case 6:

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} = \mathrm{BW}_{j\Delta},\ \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{new}}. \quad (30)$$

Case 7(a):

$$\begin{aligned}\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} \le \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}\\ (\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) < (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{old}}. \quad (31)\end{aligned}$$

Case 7(b):

$$\begin{aligned}\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} \le \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}\\ (\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) > (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{old}}. \quad (32)\end{aligned}$$

Case 8(a):

$$\begin{aligned}\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} \le \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta}\\ (\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) < (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{old}}. \quad (33)\end{aligned}$$

Case 8(b):

$$\begin{aligned}\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} \le \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta}\\ (\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) > (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}), \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \mathrm{BW}_{j\Delta}^{\mathrm{old}}. \quad (34)\end{aligned}$$

Case 9(a):

$$\begin{aligned}\mathrm{BV}_{j\Delta} = \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} = \mathrm{BW}_{k\Delta}, \mathrm{BW}_{j\Delta}^{\mathrm{new}}\\ = \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \frac{\mathrm{BW}_{j\Delta}^{\mathrm{old}} + \mathrm{BV}_{k\Delta}^{\mathrm{old}}}{2}. \quad (35)\end{aligned}$$

Case 9(b):

$$\begin{aligned}\mathrm{BV}_{k\Delta} = \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} = \mathrm{BW}_{j\Delta}, \mathrm{BW}_{k\Delta}^{\mathrm{new}}\\ = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \frac{\mathrm{BW}_{k\Delta}^{\mathrm{old}} + \mathrm{BV}_{j\Delta}^{\mathrm{old}}}{2}. \quad (36)\end{aligned}$$

The first two contraction rules were taken from FMM (9) and (10).

Equal priority syndrome: In order to eliminate the overlapping regions between two hyperboxes contraction rules used by the FMM for cases I and II, and contraction rules used by EFMM for its cases I, II, and IX, divides the overlapping regions evenly, i.e., equal priority 0.5 is given to both hyperboxes, for example, (9) can be written as $0.5 * \mathrm{BW}_{j\Delta}^{\mathrm{old}} + 0.5 * \mathrm{BV}_{k\Delta}^{\mathrm{old}}$. However, very often, different hyperboxes have different sizes; hence they are capable of including different number of training patterns, so giving equal priority to both the hyperboxes is erroneous. This problem is known as equal priority syndrome. To solve this syndrome, different priorities to different hyperboxes are applied in this paper, which are discussed in Section III.

Entire priority syndrome: In order to eliminate the overlapping regions between two hyperboxes—"contraction rules used by FMM for cases III and IV" and "contraction rules used by EFMM for its cases III, IV, V, VI, VII, and VIII"—only reduce the size of the outer (i.e., larger) hyperbox, whereas the size of the inner (i.e., smaller) hyperbox remains the same as during the overlapping test; for example, in Fig. 3(e) and (h), the size of the inner (smaller) hyperbox remains the same after a contraction. However, the abovementioned contraction rule suffers from huge information loss because here only the size of the

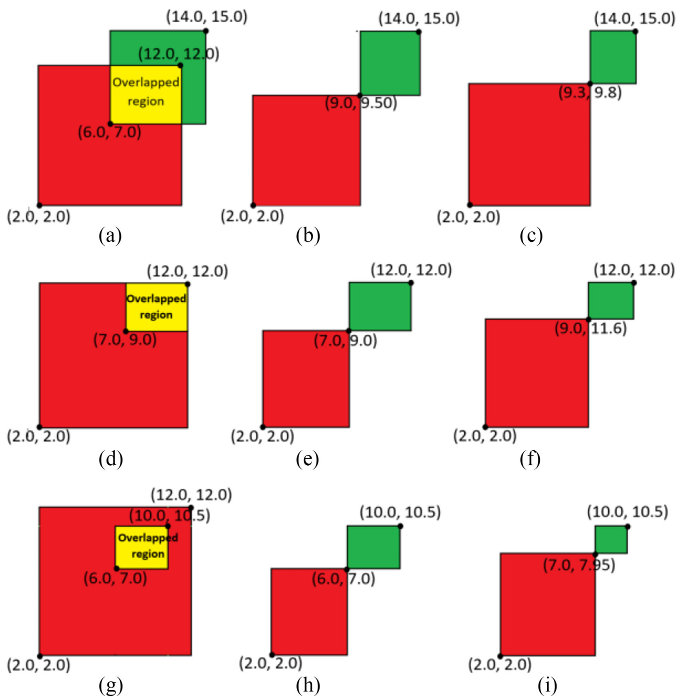

Fig. 3. Advantage of the proposed contraction rules over the rules used by FMM [17], k-NFMM [49], and EFMM [50]. (a), (d), and (g) represent three different overlapping cases. (b), (e), and (h) represent the corresponding contraction mechanism used by FMM, k-NFMM, and EFMM. (c), (f), and (i) represent the corresponding contraction mechanism used by the proposed IFMM.

larger hyperbox is affected since the number of training patterns contained by the outer hyperbox is drastically decreased. This occurs due to assigning entire priority to the inner hyperboxes (i.e., its size is totally unaffected). Hence, this drawback has been known as the entire priority syndrome. In order to resolve the above syndrome, the authors in this paper assigned a lesser priority to the inner hyperbox, as discussed in Section III.

III. Proposed Improved FMM

From Section II, it is evident that FMM, k-NFMM, and FEMM have some limitations, which can be improved. This motivated the authors to propose some modifications to FMM and EFMM for better results. In this paper, mainly three modifications have been proposed that are a combined strategy to select a hyperbox for expansion, weighted condition to check the expandability of a hyperbox, and weighted contraction rules. Various proposed modifications are described as follows.

A. Combined Strategy to Select a Hyperbox for Expansion

After selecting a list of probable hyperboxes, which could be expanded to include the current training pattern, FMM, k-NFMM, and EFMM select the hyperbox with the largest MF value as the winner hyperbox to include the current training pattern. Simultaneously, if more than one hyperbox has the largest membership value, then FMM, k-NFMM, and EFMM randomly select a hyperbox. However, this random selection of hyperbox is not appropriate since the sizes of different hyperboxes are not the same, and simultaneously different hyperboxes contain a different numbers of training patterns.

In order to resolve the abovementioned winner hyperbox problem, instead of randomly selecting a hyperbox, the proposed methods select the hyperbox with the largest semiperimeter, when more than one hyperbox has the same largest MF value. The abovementioned process is used by the proposed methods, since hyperboxes with a smaller semiperimeters are more likely to be expanded in the future. In this paper, the semiperimeter of ith hyperbox (Γ_i) can be computed using

$$\Gamma_i = \sum_{x=1}^{n} |\mathrm{BW}_{ix} - \mathrm{BV}_{ix}|. \quad (37)$$

Simultaneously, the concept of the k-nearest-neighbor mechanism is used to avoid the creation of a number of smaller hyperboxes [49]. If the hyperbox with the maximum MF value fails the expandability test, a new hyperbox is created. This unnecessarily increases the number of hyperboxes by FMM and EFMM. To resolve this, a k-nearest-neighbor [49] is used, which is described as follows: when the hyperbox with the highest membership value fails the expandability test, the hyperbox with the next higher MF value will be checked for expansion. If this hyperbox fails, check for the hyperbox with the next higher MF value. This process continues up to k hyperboxes. If the kth hyperbox also fails the expandability test, then a new hyperbox is created to represent the training pattern.

B. Weighted Condition to Check the Expandability of a Hyperbox

In the case of FMM, k-NFMM, a hyperbox, can be considered as a candidate for expansion if (4) is satisfied by that hyperbox. An EFMM includes a hyperbox into the probable candidates list for expandable hyperboxes, if (15) is satisfied by the respective hyperbox for all n-features. However, the probable expandable hyperbox selection mechanism used by EFMM is superior to FMM [50]. Moreover, checking (15) for all n-features is not required, since this can create a large number of smaller hyperboxes. For the proposed method, (15) is also used to determine whether a hyperbox can be expandable or not. However, unlike EFMM, for the proposed case, a hyperbox may be selected for expansion if at least 60% of features are satisfied by (15). In the proposed methods, 60% of features are selected, because this gives a better accuracy rate with a less number of hyperboxes, which are discussed in Section IV.

C. Weighted Contraction Rules

The contraction process used by traditional FMM, k-NFMM, and EFMM to solve the “Case 1” overlapping case divides the overlap region into four equal halves. The bottom left and top right parts of the abovementioned overlapping region were contained by two respective hyperboxes after a contraction as shown in Fig. 3(b). Using the abovementioned mechanism, after a contraction, two hyperboxes have a common point, which is located at the center of the overlap region, but the abovementioned mechanism of eliminating the overlap region has a drawback of equally distributing the proportion of the overlap region. Simultaneously, the sizes of different hyperboxes are not the same. Hence, partitioning along the center point of the overlap region to perform a contraction eliminates more regions of the bigger hyperbox, which is shown in Fig. 3. To solve the abovementioned drawback, a new contraction rule is proposed, which assigns a weight to min–max points of the hyperboxes in order to obtain the separation point, which lies in the overlapped region. Weights are assigned depending on the size of the respective hyperbox.

The proposed contraction rules are as follows.

Case 1:

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta},$$
$$\mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{new}} = (\in_1)\,\mathrm{BW}_{j\Delta}^{\mathrm{old}} + (1-\in_1)\,\mathrm{BV}_{k\Delta}^{\mathrm{old}} \quad (38)$$
$$\text{where } \in_1 = \frac{\Gamma_j}{\Gamma_j + \Gamma_k},\ \Gamma_j \text{ and } \Gamma_\mathrm{k} \text{ represent}$$
semiperimeter of jth and kth hyperbox, respectively.

Case 2:

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta},$$
$$\mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = (\in_1)\,\mathrm{BW}_{k\Delta}^{\mathrm{old}} + (1-\in_1)\,\mathrm{BV}_{j\Delta}^{\mathrm{old}}. \quad (39)$$

Case 3:

$$\mathrm{BV}_{j\Delta} = \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta}, \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BW}_{j\Delta}^{\mathrm{old}} \times \in_2. \quad (40)$$

Case 4:

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{j\Delta} = \mathrm{BW}_{k\Delta},$$
$$\mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{old}} \times (1+\in_2)\,, \text{where } \in_2 = \frac{\Gamma_k}{\Gamma_j + \Gamma_k}. \quad (41)$$

Case 5:

$$\mathrm{BV}_{k\Delta} = \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta}, \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{old}} \times \in_1. \quad (42)$$

Case 6:

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} = \mathrm{BW}_{j\Delta},\ \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{old}}\,(1+\in_1)\,. \quad (43)$$

Case 7(a):

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta} \text{ and} (\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) \le (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta}) \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BW}_{k\Delta}^{\mathrm{old}} - \left(\mathrm{BW}_{k\Delta}^{\mathrm{old}} - \mathrm{BV}_{k\Delta}^{\mathrm{old}}\right) \times \in_2. \quad (44)$$

Case 7(b):

$$\mathrm{BV}_{j\Delta} < \mathrm{BV}_{k\Delta} < \mathrm{BW}_{k\Delta} < \mathrm{BW}_{j\Delta} \text{ and } (B\mathrm{W}_{k\Delta} - \mathrm{BV}_{j\Delta}) > (\mathrm{BW}_{\mathrm{j}\Delta} - \mathrm{BV}_{\mathrm{k}\Delta}) \mathrm{BW}_{\mathrm{j}\Delta}^{\mathrm{new}} = \mathrm{BV}_{\mathrm{k}\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{old}} + \left(\mathrm{BW}_{k\Delta}^{\mathrm{old}} - \mathrm{BV}_{k\Delta}^{\mathrm{old}}\right) \times \in_2. \quad (45)$$

Case 8(a):

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta} \text{ and } (\mathrm{BW}_{k\Delta} - \mathrm{BV}_{j\Delta}) \le (\mathrm{BW}_{j\Delta} - \mathrm{BV}_{k\Delta})$$
$$\mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{old}} + \left(\mathrm{BW}_{j\Delta}^{\mathrm{old}} - \mathrm{BV}_{j\Delta}^{\mathrm{old}}\right) \times \in_1. \quad (46)$$

Case 8(b):

$$\mathrm{BV}_{k\Delta} < \mathrm{BV}_{j\Delta} < \mathrm{BW}_{j\Delta} < \mathrm{BW}_{k\Delta} \text{ and } (B\mathrm{W}_{k\Delta} - \mathrm{BV}_{j\Delta}) > (\mathrm{BW}_{\mathrm{j}\Delta} - \mathrm{BV}_{\mathrm{k}\Delta}) \mathrm{BV}_{\mathrm{k}\Delta}^{\mathrm{new}} = \mathrm{BW}_{\mathrm{j}\Delta}^{\mathrm{new}} = \mathrm{BW}_{j\Delta}^{\mathrm{old}} - \left(\mathrm{BW}_{j\Delta}^{\mathrm{old}} - \mathrm{BV}_{j\Delta}^{\mathrm{new}}\right) \times \in_1. \quad (47)$$

Case 9:

$$\mathrm{BV}_{k\Delta} = \mathrm{BV}_{j\Delta} < \mathrm{BW}_{k\Delta} = \mathrm{BW}_{j\Delta}$$
$$\text{let} \beta_k = \frac{\text{Number of points of class } k}{\text{Total number of points of class } k \text{ and class } j}$$
$$\beta_j = \frac{\text{Number of points of class } j}{\text{Total number of points of class } k \text{ and class } j}$$
$$\text{If } \beta_j > \beta_k,\ \mathrm{BW}_{j\Delta}^{\mathrm{new}} = \mathrm{BV}_{k\Delta}^{\mathrm{new}} = \left(\mathrm{BW}_{j\Delta}^{\mathrm{old}} \times \beta_j\right) + \left(\mathrm{BV}_{k\Delta}^{\mathrm{old}} \times \beta_k\right)$$
$$\text{If } \beta_k > \beta_j,\ \mathrm{BW}_{k\Delta}^{\mathrm{new}} = \mathrm{BV}_{j\Delta}^{\mathrm{new}} = \left(\mathrm{BW}_{k\Delta}^{\mathrm{old}} \times \beta_k\right) + \left(\mathrm{BV}_{j\Delta}^{\mathrm{old}} \times \beta_j\right). \quad (48)$$

For case (9), the *i*th feature is selected for which the side length of the hyperbox is maximum, because data are likely to be scattered in such a way that the center point of a hyperbox is nearly at the same distance from all the points that fall in that hyperbox. Here, except the *i*th dimension (feature), all other dimensions of both *j* and *k* hyperbox remains the same.

The advantages of the proposed contraction rules are depicted in Fig. 3. From Fig. 3, following can be concluded.

1) In Fig. 3(a), areas of larger and smaller hyperboxes are 100 and 64, respectively. If the contraction mechanism used by FMM, *k*-NFMM, and EFMM is applied to Fig. 3(a), then the total information loss by FMM, *k*-NFMM, and EFMM is [Union of hyperboxes shown in Fig. 3(a) – union of hyperboxes shown in Fig. 3(b)] = (100 + 64 − 30) − (52.5 + 27.5) = 54. However, the total information loss by the proposed IFMM mechanism is (100 + 64 − 30) − (56.94 + 24.44) = 52.5. Simultaneously, in the case of IFMM, the area of hyperboxes obtained after applying a contraction are relatively proportional to their respective sizes before the contraction. However, in the case of FMM, *k*-NFMM, and EFMM, a larger hyperboxes lose sizeable amount of information.
2) The total information loss incurred by FMM, *k*-NFMM, and EFMM for Fig. 3(d) is (100 + 15 − 15) − (35 + 15) = 50, whereas the total information loss incurred by the proposed IFMM is (100 + 15 − 15) − (67.2 + 1.2) = 31.6

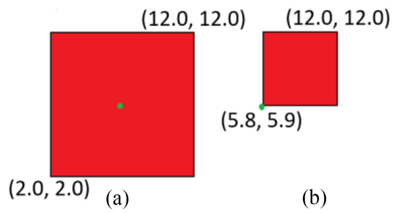

Fig. 4. Contraction rule used by EFMM for the condition $BV_{j\Delta} < BV_{k\Delta} = BW_{k\Delta} < BW_{j\Delta}$ and $BV_{k\Delta} < BV_{j\Delta} = BW_{j\Delta} < BW_{k\Delta}$ (a) overlapping of two hyperboxes (b) results of the contraction mechanism of EFMM.

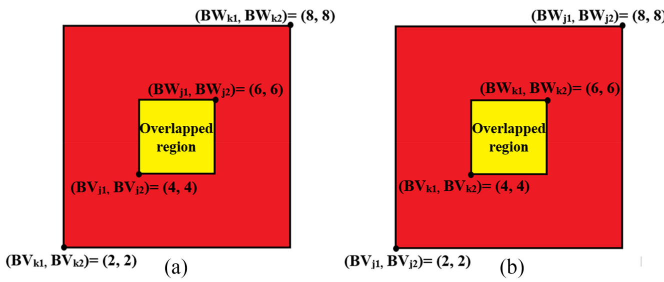

Fig. 5. Two different overlapping cases which are not considered by FMM, *k*-NFMM, and EFMM during contraction.

3) The total information loss incurred by FMM, *k*-NFMM, and EFMM for Fig. 3(g) is $(100 + 14 - 14) - (20 + 14) = 66$, whereas the total information loss incurred by the proposed IFMM is $(100 + 14 - 14) - (29.75 + 7.65) = 62.6$.

Therefore, the contraction mechanism used by FMM, *k*-NFMM, and EFMM relatively loses more area from the larger hyperbox; as a result, the classification accuracy rate decreases. Whereas, the proposed mechanism evenly modifies the size of both hyperboxes based on weight. The value of weights depends on the size of corresponding hyperbox. IFMM_IV uses (38), (39), (44)–(47), whereas IFMM_IX uses (38)–(48).

Note 1: When a single input pattern of a class is enclosed in a hyperbox of a different class $(BV_{j\Delta} < BV_{k\Delta} = BW_{k\Delta} < BW_{j\Delta}$ and $BV_{k\Delta} < BV_{j\Delta} = BW_{j\Delta} < BW_{k\Delta})$, as shown in Fig. 4(a), the contraction rule applied by EFMM is depicted in Fig. 4(b). This contraction is erroneous because of the following reasons:

1) here, a sizable amount of information is lost from the larger hyperbox, which is undesirable;
2) simultaneously, it is evident that, if another input pattern of the class represented by the single point nearer to that single point is encountered, then a hyperbox will definitely be created to represent both points; and hence, a contraction will be needed.

Note 2: FMM, *k*-NFMM, and EFMM have not considered the overlapping cases as described in Fig. 5(a) and (b) for a contraction that occurs when $(BW_{k\Delta} - BV_{j\Delta}) = (BW_{j\Delta} - BV_{k\Delta})$ and $(BW_{j\Delta} - BV_{k\Delta}) = (BW_{k\Delta} - BV_{j\Delta})$, respectively. However, the proposed contraction rules consider the above-mentioned overlapping cases.

Description of the functions used in algorithm Check_class_label(): For all the methods, it returns 0 if the system does not contain any hyperbox whose class label is equal to the class label of the current training pattern. Otherwise, it returns 1.

Computer_max_MF (): It returns the hyperbox whose class label is the same as the current training pattern and provides the maximum membership value corresponding to the current training pattern. The operation of this function with respect to FMM, *k*-NFMM, EFMM, and the proposed methods is described as follows.

FMM, EFMM: It uses (3) to compute MF of different hyperboxes corresponding to the current training pattern and returns the hyperbox corresponding to the highest MF value. However, when more than one hyperbox has the same highest MF value, it randomly selects the hyperbox.

k-NFMM: It uses (3) to compute MF for different hyperboxes corresponding to the current training pattern and returns a one-dimensional (1-D) array Y, consisting of *k*-hyperboxes corresponding to the *k*-highest MF value arranged in decreasing order of their MF values. However, when more than one hyperbox has the same highest MF value, it randomly selects the hyperbox.

IFMM: It uses (3) to compute MF for different hyperboxes for the proposed IFMM_IV and IFMM_IX. However, for both the proposed methods, it returns a 1-D array Y, consisting of *k*-hyperboxes corresponding to the *k*-highest MF value arranged in decreasing order of their MF values. However, when more than one hyperbox has the same MF value, they are ordered according to their semiperimeter. Here, a hyperbox with a larger semiperimeter has a higher priority.

Check_expandability (): It returns 0 if the input hyperbox is nonexpandable. Otherwise, it returns 1. The operation of this function with respect to FMM, *k*-NFMM, EFMM, and the proposed methods is described as follows.

FMM and k-NFMM: It uses (4) to check whether the hyperbox returned by compute_max_MF (), y can be expanded or not. It returns 0 if y cannot be expanded; otherwise, it returns 1.

EFMM: It uses (15) to check whether the hyperbox returned by compute_max_MF (), y can be expanded or not. It returns 0 if y cannot be expanded; otherwise, it returns 1.

IFMM: For the proposed methods, Check_expandability () returns 0 if all the hyperboxes present in Y fail the expandability test. Otherwise, it returns the first hyperbox (starting from first location) of Y, which satisfies the expandability test.

add_hyperbox (): For all the methods, it is used to create a new point hyperbox using the current training pattern.

modify_hyperbox (): For all the methods, it modifies the size of a hyperbox while including the current training pattern.

Check_overlap (): For all the methods, this function is used to check whether any overlapping exists between the hyperboxes of different classes. This is needed because adding a new hyperbox or modifying an existing one may sometimes create overlapping between two or more hyperboxes. For all the methods, Check_overlap returns 0 if there is no overlapping between the hyperbox of different classes. However, if any overlapping exists between the two hyperboxes j and k, Check_overlap () returns the case number corresponding to the type of overlapping.

Do_contraction (): It performs a contraction on two hyperboxes that corresponds to a particular overlapping case. The

Algorithm 1: FMM Based Classification Algorithm.

Training Phase

Input: Training set, expansion coefficient (θ), sensitivity parameter (γ)

Output: BV, BW, CL and n_b

```
Set n_b = 0
for 1=1:N
  if n_b ==0
    n_b = n_b + 1;
    add(BV, BW, CL, T(i,:), n_b);
  else
    x=Check_class_label(CL, NC(i));
    if x==0
      n_b = n_b + 1;
      add (BV, BW, CL, T(i,:), n_b);
    else
      y=Compute_max_MF(BV, BW, CL, T(i,:));
      z=Check_expandability(BV(y,:), BW(y,:), θ);
      if z==0
        n_b = n_b + 1;
        add (BV, BW, CL, T(i,:), n_b);
      else
        Modify_hyperbox(BV, BW, y, T(i,:));
      end if
    end if
    for j=1: n_b
      for k=1: n_b
        if j~=k && CL(j) ~=CL(k)
          p = Check_overlap(BV, BW, j, k);
          If p~=0
            Do_contraction(p, BV, BW, j, k);
          end if
        end if
      end for
    end for
  end if
end for
```

Testing phase:

Input: testing pattern, set of hyperboxes with minimum and maximum points, class label, t_p, BV, BW, CL, n_b and γ

Output: class label of the testing patterns

z=Compute $\left(t_p, BV, BW, CL, n_b, \gamma\right)$

output of Check_overlap () is fed into it. For all the methods, it modifies the size of both overlapped hyperboxes returned by Check_overlap ().

Compute (): For all the methods, this function takes a testing pattern, gamma, and the min–max points of all the hyperboxes along with their respective class labels as its inputs and returns an approximate class label.

TABLE I
NOMENCLATURE

Notation	Meaning
N	Number of patterns present in the learning set
f	Number of features per pattern
T	$N \times f$ matrix containing the training patterns
NC	1-D matrix, consisting of the class label of all the training patterns.
BV	Matrix consisting of the MIN points of each hyperbox
BW	Matrix consisting of the MAX points of the hyperboxes
n_b	Number of hyperboxes
CL	1-D matrix, consisting of the class label of all hyperboxes
θ	Expansion coefficient to control the size of hyperboxes
γ	Sensitivity parameter to control the membership value
t_p	Testing pattern

TABLE II
DESCRIPTION OF DATASETS

Datasets	Instances		Attributes	Classes
	Total	Used		
WBC	198	194	33	02
Lung Cancer	32	27	56	03
Hepatitis	88	80	19	02
Parkinson	197	197	22	02
Primary Tumor	339	132	17	22
Optical Recognition of Handwritten Digits	5620	500	64	10
Statlog (Landsat Satellite)	6435	500	36	07

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Dataset

A total of seven popular datasets, namely Wisconsin Breast cancer (WBC), lung cancer, hepatitis, Parkinson's [53], primary tumor, optical recognition of handwritten digits, and statlog (Landsat Satellite) databases, are used in this paper to compare different FMM-based methods. All seven databases are downloaded from [54]. Hence, for the WBC, lung cancer, hepatitis, Parkinson's, and primary tumor datasets, the instances for which there exist values for all the attributes are used in simulation. However, all instances of optical recognition of handwritten digits and statlog (Landsat Satellite) datasets contain values for all the attributes, but these datasets contain a very large number of instances. Therefore, a favorable number of instances are considered for simulation. Simulation statistics for various datasets are shown in Table II. For all datasets, 80% of the instances are randomly selected for training, and the rest are used for testing.

B. Experimental setting

FMM and k-NFMM have used (4) and EFMM has used (15) to check the expandability of hyperboxes. Due to the use of (4), FMM and k-NFMM introduce some overlapping between the

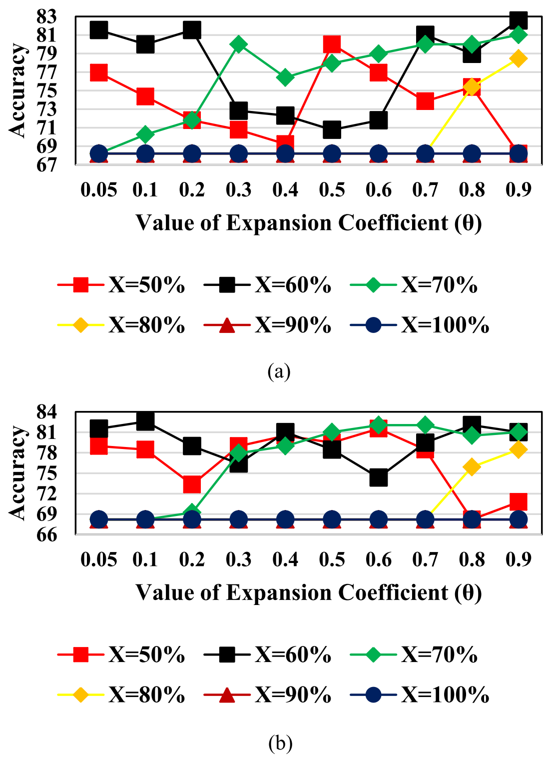

Fig. 6. Accuracy rate of the WBC (Prognostic) dataset for different X values obtained by the proposed methods. (a) IFMM_IV. (b) IFMM_IX.

hyperboxes of different classes [50]. Hence, EFMM uses (15) to check whether a hyperbox can be expandable or not. However, EFMM unnecessarily creates many small-sized hyperboxes, because it allows a hyperbox to expand if (15) is satisfied by all the dimensions of the current training pattern. Moreover, creation of such a large number of small-sized hyperboxes can be controlled to some extent by relaxing the rule of EFMM applied to (15).

In order to control the creation of a large number of small-sized hyperboxes, the proposed method uses the following rule: "First, apply (15) to current training pattern and all the feature of a hyperbox. Then, count the number of features of the corresponding hyperbox that satisfy (15). Then, a hyperbox is said to be expandable, if at least $'X\%$ of features $/X \leq 100'$ of the hyperbox satisfy (15)." In this paper, the optimal value of $'X'$ is experimentally obtained, using the WBC dataset, which is described as follows: first, apply the two proposed methods to classify the WBC dataset using different X values, where $X \in [50\%, 60\%, \ldots, 100\%]$. Fig. 6 and Table III show and list the obtained average accuracy rate of the proposed methods, respectively, for different values of X, while classifying the WBC dataset. From Table III, it is clear that, when at least 60% of features are selected ($X = 60$), a better accuracy rate is obtained. Hence, the proposed methods select a hyperbox for expansion, if (15) is satisfied by at least 60% of its features.

C. Performance Comparison

Broadly, different classification methods can be compared either qualitatively or quantitatively. In this paper, "classification accuracy rate" is used to qualitatively evaluate different methods. Simultaneously, "number of generated hyperboxes" is used for quantitative evaluation of various methods. The proposed methods are compared with FMM, k-NFMM, and EFMM using the WBC, lung cancer, hepatitis, Parkinson's, primary tumor, optical recognition of handwritten digits, and statlog (Landsat Satellite) databases. Fig. 7 shows the average accuracy rates of different methods while classifying the WBC, lung cancer, hepatitis, Parkinson's, primary tumor, optical recognition of handwritten digits, and statlog (Landsat Satellite) datasets. Simultaneously, Fig. 8 shows the average number of generated hyperboxes for different methods while classifying the above-mentioned datasets. The average accuracy rate and average number of generated hyperboxes for each method corresponding to different datasets are obtained as follows: for each dataset, first randomly select 80% samples for training and train the system for each method accordingly. Then, use the remaining 20% of samples for testing and obtain the accuracy rate and number of generated hyperboxes for each method, respectively, for different values of θ, where $\theta = 0.05,\ 0.1x/x = 1, 2, \ldots, 9$. Perform the above process eight times using different randomly selected training and testing samples. Then, the average accuracy rates of eight observations are computed, which is designated as the average accuracy rate of a particular method. Similarly, the average number of generated hyperboxes is computed. From Fig. 7, it is clear that IFMM_IV and IFMM_IX provide a more average accuracy rate than FMM, k-NFMM and EFMM. Similarly, Fig. 8 concludes that IFMM_IV and IFMM_IX generate fewer hyperboxes than FMM, k-NFMM and EFMM for all the datasets. Hence, the proposed methods provide a better performance than FMM, k-NFMM, and EFMM.

TABLE III
AVERAGE ACCURACY RATE OF WISCONSIN BREAST CANCER DATASET

X	Average Accuracy Rates	
	IFMM_IV	IFMM_IX
50%	73.74359	76.87179
60%	**77.33333**	**79.58974**
70%	76.46154	76.92308
80%	69.94872	70
90%	68.20513	68.20513
100%	68.20513	68.20513

D. Application of Proposed Method: Selecting a Proper Magnification Factor for Classification of Histopathological Images

In this section, the proposed methods are used to find a better magnification factor for classification. For this purpose, images of breast cancer histopathology (BreaKHis) database [55] are used for classification, because this is the only database that contains images at multiple magnification factors. This database contains images at four magnification factors i.e., $40\times$, $100\times$, $200\times$, and $400\times$. BreakHis Database standardized using [56] contains a total of 7909 biopsy images, out of which 1000 images per magnification rate are randomly selected,

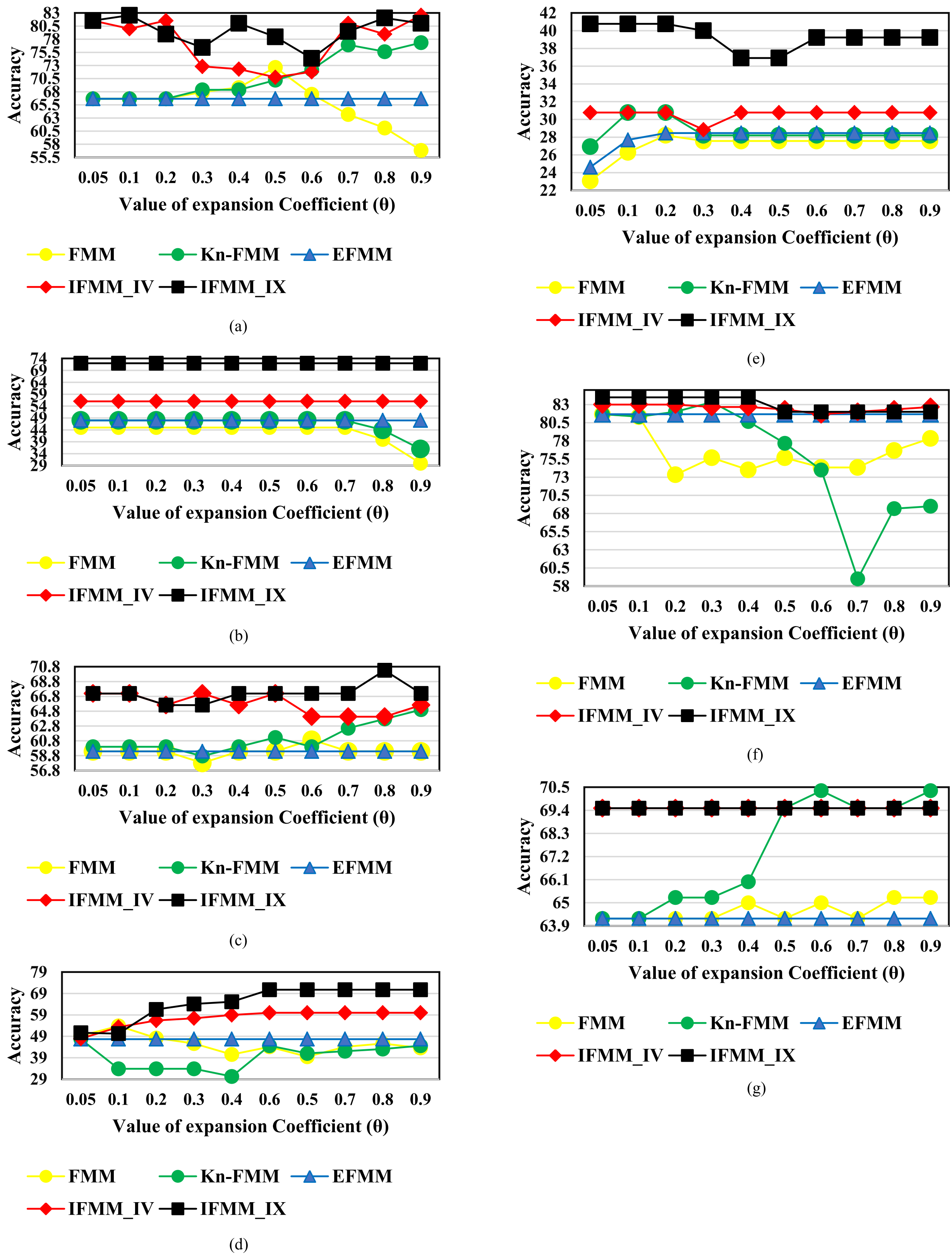

Fig. 7. Accuracy rate of various FMM networks for (a) Wisconsin Breast cancer (Prognostic), (b) lung cancer, (c) hepatitis, (d) Parkinson's, (e) primary tumor, (f) optical recognition of handwritten digits, and (g) statlog (Landsat Satellite) database.

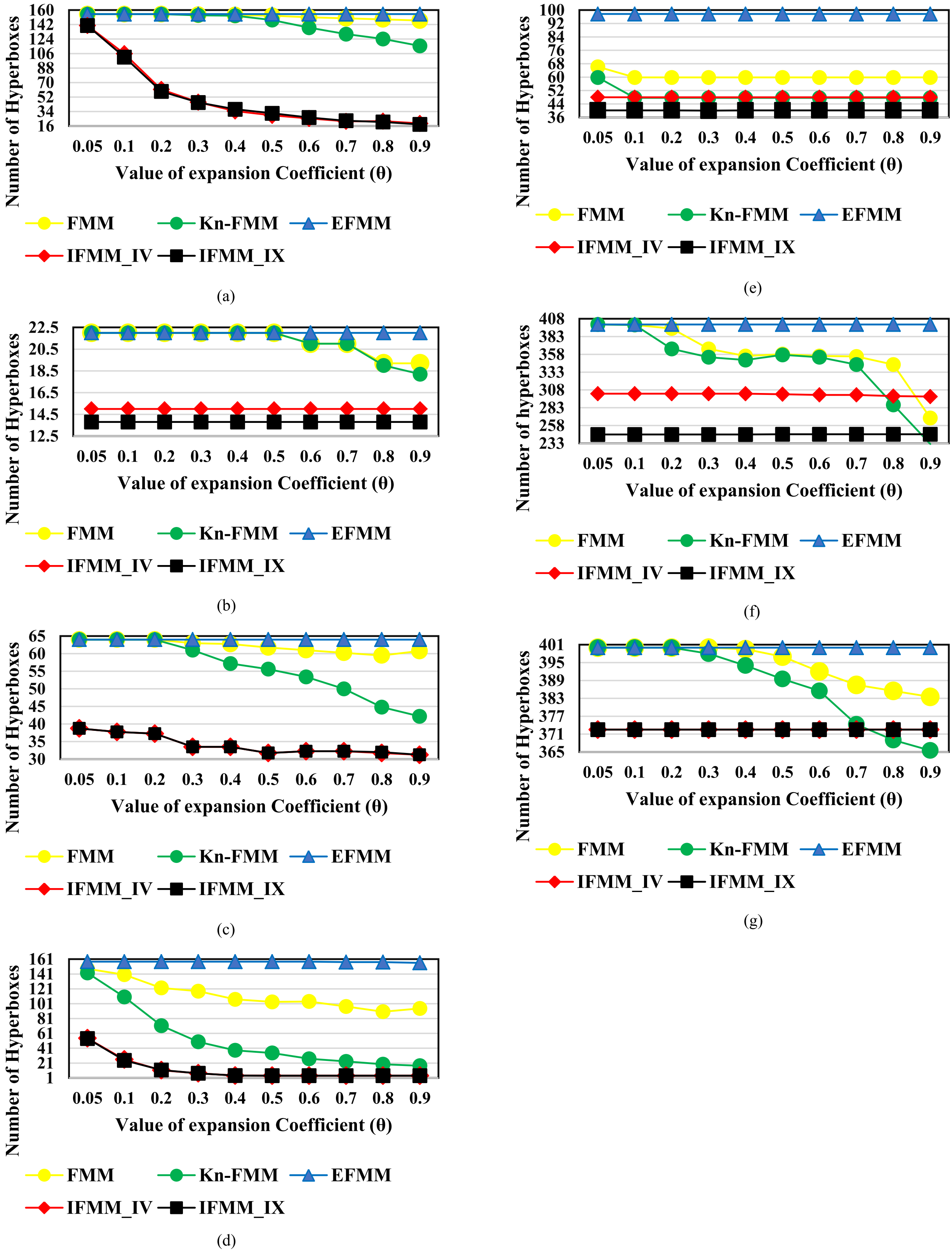

Fig. 8. Number of generated hyperboxes for classifying various datasets using different FMM networks. (a) Wisconsin Breast cancer (Prognostic). (b) Lung cancer. (c) Hepatitis. (d) Parkinson's. (e) Primary tumor. (f) Optical recognition of handwritten digits. (g) Statlog (Landsat Satellite) database.

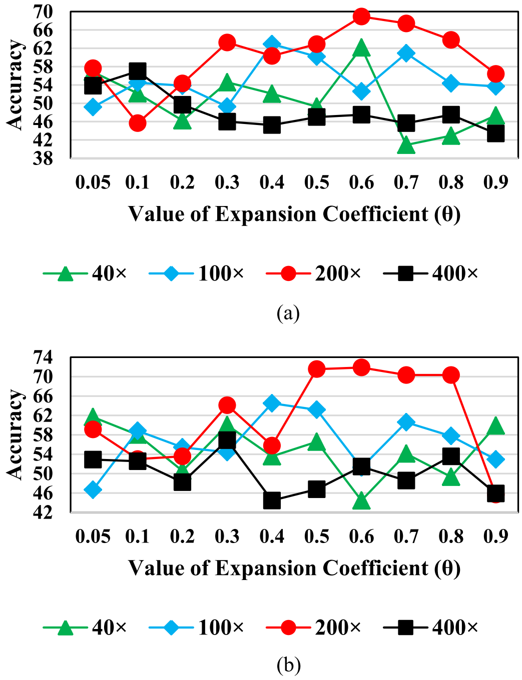

Fig. 9. Accuracy rate of various magnification factors of histopathological images using (a) IFMM_IV and (b) IFMM_IX networks.

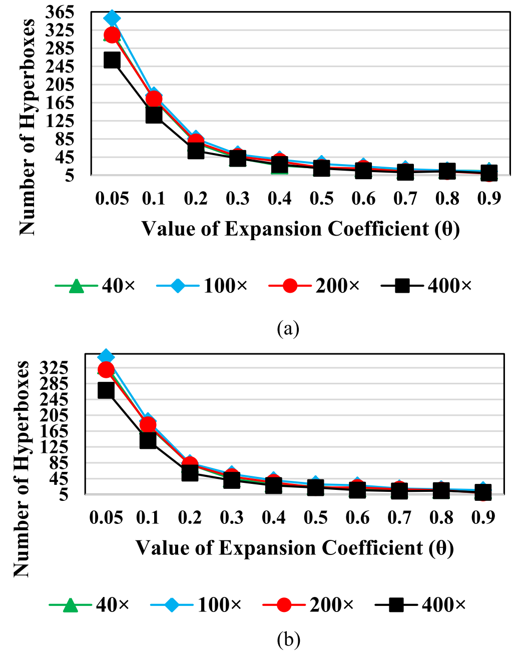

Fig. 10. Number of generated hyperboxes for various magnification factors of histopathological images using (a) IFMM_IV and (b) IFMM_IX.

TABLE IV
AVERAGE ACCURACY RATE FOR VARIOUS MAGNIFICATION FACTORS OF HISTOPATHOLOGICAL IMAGES

Magnification	Average Accuracy Rates	
	IFMM_IV	IFMM_IX
40×	50.45	54.81111
100×	55.13333	56.55
200×	**60.075**	**61.53333**
400×	48.28333	50.12222

which are at that corresponding magnification rate. Hence, four magnification rates are used which are 40×, 100×, 200×, and 400×. For each magnification rate, 500 images belong to benign, and rest 500 are of malignant. Corresponding to each magnification factor, 70% of its images are selected for training and 30% for testing. Gray level cooccurrence matrix (GLCM) is used to extract the features from images, which are then fed into IFMM_IV and IFMM_IX network to classify the dataset. Adjacency direction of 0° is used to calculate the GLCM features. All 13 haralick texture features are calculated, which can be found in [57]. At last, a 13-dimensional (13-D) feature vector can be obtained by averaging the haralick texture feature values for three channels of a color image. The average of eight trials are presented in this paper. Fig. 9(a) and (b) shows the accuracy rates of IFMM_IV and IFMM_IX corresponding to four magnification factors, respectively.

From Fig. 9(a) and (b), it is concluded that magnification factor of 200× gives best result for both IFMM_IV and IFMM_IX networks. Fig. 10(a) and (b) shows the number of hyperboxes generated by IFMM_IV and IFMM_IX networks, respectively, for various magnification factor corresponding to histopathological images. From Fig. 10(a) and (b), it is concluded that the magnification factor of 100× and 400× generates the maximum and minimum number of hyperboxes, respectively. Table IV shows the average accuracy rates for various magnification factors of histopathological images. From Table IV, it is clear that the magnification factor of 200× gives best result for both IFMM_IV and IFMM_IX networks.

V. CONCLUSION

Hyperbox classifier has attracted many researchers for data classification and has made considerable progress in the past few decades. Hyperbox classifier is efficiently implemented using FMM network. In this paper, two improved FMM networks are proposed by interpreting the drawbacks of FMM and its two well-known modifications, i.e., k-NFMM and EFMM. First, the expansion process of FMM increases the overlapping region between different classes; on the other hand, the expansion process of EFMM creates a large number of hyperboxes, which are not actually required. k-NFMM tries to reduce the creation of a large number of small sized hyperbox by using k-highest mechanism, but like FMM it also increases the overlapping between hyperboxes. Second, during the expansion process, when more than one hyperbox has the same membership function value, FMM, k-NFMM, and EFMM randomly select a hyperbox to be expanded in order to accommodate the current training patterns. Third, the contraction process of FMM, k-NFMM, and EFMM are sometimes biased toward a particular hyperbox, and sometimes evenly partitions the overlapping region irrespective of the hyperbox size, which is undesirable. The accuracy of FMM, k-NFMM, and EFMM were also analyzed, and it was found that the accuracy of abovementioned methods can be improved. Hence, two improved FMM networks were proposed to provide a better accuracy rate by solving shortcomings of FMM, *k*-NFMM, and EFMM.

The main contributions of this paper are the three modifications applied to FMM, *k*-NFMM, and EFMM for accurately classifying a testing pattern. First, a combined strategy was adopted using the *k*-nearest neighbor mechanism and semiperimeter of hyperboxes to fix the expandable hyperboxes. Second, a weighted formula was used to get the probable hyperboxes that could be expanded to include a training pattern. Third, weighted contraction rules were applied to remove the overlapping regions of two hyperboxes due to the drawbacks of equal priority syndrome and entire priority syndrome. The performance of the proposed methods are compared with the FMM, *k*-NFMM, and EFMM using seven benchmark datasets namely WBC (Prognostic), lung cancer, hepatitis, Parkinson's, primary tumor, optical recognition of handwritten digits, and statlog (Landsat Satellite) data base. It was found that the proposed methods provide more accuracy rate than FMM, *k*-NFMM, and EFMM. Moreover, as compared to other existing methods, the proposed methods generate a lesser number of hyperboxes, while classifying the datasets. Thus, less maintenance cost is required. Due to its combined strategy and weighted contractions rules, the proposed methods have a larger network complexity. The proposed methods are also used to find the best magnification factor for classifying the histopathological images. From the experiments, it was concluded that the magnification factor of 200× gives best accuracy rate. However, the proposed methodology can be improved by eliminating some less-efficient overlap test rules, which unnecessarily increases the network complexity. Moreover, hyperbox statistics like centroid, surface area, volume, and distance between min and max points can be used to generate the new set of contraction rules. Simultaneously, training phase may generate few or more inefficient hyperboxes, which affect the performance of the system. Hence, some rules can be applied to overcome such inefficient hyperboxes. The proposed methods can also be applied to unsupervised learning and various real-world applications like gesture recognition, face detection, text categorization, remote sensing, bioinformatics, business, medicine, industrial fault detection, and diagnosis.

References

[1] D. Graupe, *Principles of Artificial Neural Networks*. Singapore: World Scientific, 1997.

[2] M. Wu and P. Rastgoufard, "Optimum decision by artificial neural networks for reactive power control equipment to enhance power system stability and security performance," in *Proc. IEEE Power Eng. Soc. Gen. Meeting*, vol. 2. Jun. 2004, pp. 2120–2125.

[3] P. J. G. Lisboa, "A review of evidence of health benefit from artificial neural networks in medical intervention," *Neural Netw.*, vol. 15, no. 1, pp. 11–39, 2002.

[4] Y. S. Xia, F. Gang, and W. Jun, "A primal-dual neural network for online resolving constrained kinematic redundancy in robot motion control," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 35, no. 1, pp. 54–64, Feb. 2005.

[5] Y. Li and W. Ma, "Applications of artificial neural networks in financial economics: A survey," in *Proc. Int. Symp. Comput. Intell. Des.*, 2010, vol. 1, pp. 211–214.

[6] M.-Y. Chow and S. O. Yee, "Methodology for on-line incipient fault detection in single-phase squirrel-cage induction motors using artificial neural networks," *IEEE Trans. Energy Convers.*, vol. 6, no. 3, pp. 536–545, Sep. 1991.

[7] D. Sánchez, P. Melin, O. Castillo, "A grey wolf optimizer for modular granular neural networks for human recognition," *Comput. Intell. Neurosci.*, vol. 2017, 2017, Art. no. 4180510.

[8] D. Sánchez, P. Melin, and O. Castillo, "Optimization of modular granular neural networks using a firefly algorithm for human recognition," *Eng. Appl. Artif. Intell.*, vol. 64, pp. 172–186, 2017.

[9] D. Sánchez, P. Melin, and O. Castillo, "Optimization of modular granular neural networks using a hierarchical genetic algorithm based on the database complexity applied to human recognition," *Inf. Sci.*, vol. 309, pp. 73–101, 2015.

[10] P. Melin, J. Amezcua, F. Valdez, and O. Castillo, "A new neural network model based on the LVQ algorithm for multi-class classification of arrhythmias," *Inf. Sci.*, vol. 279, pp. 483–497, 2014.

[11] J. H. D G. P. Zhang, "Neural networks for classification: A survey," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 30, no. 4, pp. 451–462, Nov. 2000.

[12] P. Melin, F.Olivas, O. Castillo, F. Valdez, J. Soria, and M. Valdez, "Optimal design of fuzzy classification systems using PSO with dynamic parameter adaptation through fuzzy logic," *Expert Syst. Appl.*, vol. 40, no. 8, pp. 3196–3206, 2013.

[13] A. Quteishat, C. P. Lim, J. Tweedale, and L. C. Jain, "A neural networkbased multi-agent classifier system," *Neurocomputing*, vol. 72, no. 7, pp. 1639–1647, Mar. 2009.

[14] M. McCloskey and N. J. Cohen, "Catastrophic interference in connectionist networks: The sequential learning problem," in *The Psychology of Learning and Motivation*, G. H. Bower, Ed. New York, NY, USA: Academic, 1989, pp. 109–165.

[15] R. Ratcliff, "Connectionist models of recognition memory: Constraints imposed by learning and forgetting functions," *Psychol. Rev.*, vol. 97, no. 2, pp. 285–308, 1990.

[16] R. Polikar, L. Upda, S. S. Upda, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 31, no. 4, pp. 497–508, Nov. 2001.

[17] P. K. Simpson, "Fuzzy min-max neural networks. I. Classification," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 776–786, Sep. 1992.

[18] P. K. Simpson, "Fuzzy min-max neural networks–Part 2: Clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 1, pp. 32–45, Feb. 1993.

[19] S. Grossberg, "Adaptive pattern classification and universal recording: I. Parallel development and coding of neural feature detectors," *Biol. Cybern.*, vol. 23, no. 3, pp. 121–134, Jul. 1976.

[20] S. Grossberg, "Adaptive pattern classification and universal recording: II. Feedback, expectation, olfaction, illusions," *Biol. Cybern.*, vol. 23, no. 4, pp. 187–202, Dec. 1976.

[21] A. Quteishat and C. P. Lim, "Application of the fuzzy min-max neural networks to medical diagnosis," in *Proc. Int. Conf. Knowl.-Based Intell. Inf. Eng. Syst.*, 2008, pp. 548–555.

[22] G. Bortolan, I. I. Christov, and W. Pedrycz, "Hyperbox classifiers for ECG beat analysis," in *Proc. Comput. Cardiology*, 2007, pp. 145–148.

[23] H.-J. Kim, J. Lee, and H.-S. Yang, "A weighted FMM neural network and its application to face detection," in *Proc. Int. Conf. Neural Inf. Process.*, 2006, pp. 177–186.

[24] A. M. Quteishat and C. P. Lim, "A modified fuzzy min-max neural network and its application to fault classification," in *Proc. Soft Comput. Ind. Appl.*, 2007, pp. 179–188.

[25] K. Rao J. and M. Okade, "Role of hyperbox classifiers for color recognition," in *Proc. 10th Indian Conf. Comput. Vis., Graph., Image Process.*, IIT Guwahati, India, Dec. 2016, pp. 1–2.

[26] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, Jun. 1965.

[27] S. Yilmaz and Y. Oysal, "Fuzzy wavelet neural network models for prediction and identification of dynamical systems," *IEEE Trans. Neural Netw.*, vol. 21, no. 10, pp. 1599–1609, Oct. 2010.

[28] J. E. Moreno, O. Castillo, J. R. Castro, L. G. Martínez, and P. Melin, "Data mining for extraction of fuzzy IF-THEN rules using Mamdani and Takagi-Sugeno-Kang FIS," *Eng. Lett.*, vol. 15, no. 1, pp. 82–88, 2007.

[29] D. Hidalgo, O. Castillo, and P. Melin, "Type-1 and type-2 fuzzy inference systems as integration methods in modular neural networks for multimodal biometry and its optimization with genetic algorithms," *Inf. Sci.*, vol. 179, no. 13, pp. 2123–2145, Jun. 2009.

[30] D. Sanchez, P. Melin, O. Castillo, and F. Valdez, "Modular granular neural networks optimization with multi-objective hierarchical genetic algorithm for human recognition based on iris biometric," in *Proc. IEEE Congr. Evol. Comput.*, 2013, pp. 772–778.

[31] P. Melin, O. Mendoza, and O. Castillo, "Face recognition with an improved interval type-2 fuzzy logic Sugeno integral and modular neural networks," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 41, no. 5, pp. 1001–1012, Sep. 2011.

[32] C.-F. Juang, T.-C. Chen, and W.-Y. Cheng, “Speedup of implementing fuzzy neural networks with high-dimensional inputs through parallel processing on graphic processing units,” *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 4, pp. 717–728, Aug. 2011.

[33] K. S. Yap, C. P. Lim, and M. T. Au, “Improved GART neural network model for pattern classification and rule extraction with application to power systems,” *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2310–2323, Dec. 2011.

[34] M. Davanipoor, M. Zekri, and F. Sheikholeslam, “Fuzzy wavelet neural network with an accelerated hybrid learning algorithm,” *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 3, pp. 463–470, Jun. 2012.

[35] M. Pratama, S. G. Anavatti, and E. Lughofer, “GENEFIS: Towards an effective localist network,” *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 3, pp. 547–562, Jun. 2014, doi: 10.1109/TFUZZ.2013.2264938.

[36] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, “Identification and prediction of dynamic systems using an interactively recurrent self-evolving fuzzy neural network,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 2, pp. 310–321, Feb. 2013.

[37] B. Gabrys and A. Bargiela, “General fuzzy min-max neural network for clustering and classification,” *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 769–783, May 2000.

[38] A. V. Nandedkar and P. K. Biswas, “A general reflex fuzzy min-max neural network,” *Eng. Lett.*, vol. 14, no. 1, pp. 195–205, Feb. 2007.

[39] A. Likas, “Reinforcement learning using the stochastic fuzzy min-max neural network,” *Neural Process. Lett.*, vol. 13, no. 3, pp. 213–220, 2001.

[40] A. Rizzi, M. Panella, and F. M. F. Mascioli, “Adaptive resolution min-max classifiers,” *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 402–414, Mar. 2002.

[41] A. Bargiela, W. Pedrycz, and M. Tanaka, “An inclusion/exclusion fuzzy hyperbox classifier,” *Int. J. Knowl.-Based Intell. Eng. Syst.*, vol. 8, no. 2, pp. 91–98, Aug. 2004.

[42] H. J. Kim and H. S. Yang, “A weighted fuzzy min-max neural network and its application to feature analysis,” in *Advances in Natural Computation* (Lecture Notes in Computer Science), vol. 3612, L. Wang, K. Chen, and Y. Ong, Eds., Berlin, Germany: Springer-Verlag, 2005, pp. 1178–1181.

[43] A. V. Nandedkar and P. K. Biswas, “A fuzzy min-max neural network classifier with compensatory neuron architecture,” *IEEE Trans. Neural Netw.*, vol. 18, no. 1, pp. 42–54, Jan. 2007.

[44] H. Zhang, J. Liu, D. Ma, and Z. Wang, “Data-core-based fuzzy min-max neural network for pattern classification,” *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2339–2352, Dec. 2011.

[45] A. Quteishat and C. P. Lim, “A modified fuzzy min-max neural network with rule extraction and its application to fault detection and classification,” *Appl. Soft Comput.*, vol. 8, no. 2, pp. 985–995, Mar. 2008.

[46] A. Quteishat, C. P. Lim, and K. S. Tan, “A modified fuzzy min-max neural network with a genetic-algorithm-based rule extractor for pattern classification,” *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 40, no. 3, pp. 641–650, May 2010.

[47] M. Seera, C. P. Lim, D. Ishak, and H. Singh, “Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid FMM-CART model,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 1, pp. 97–108, Jan. 2012.

[48] M. Seera and C. P. Lim, “Online motor fault detection and diagnosis using a hybrid FMM-CART model,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 4, pp. 806–812, Apr. 2014.

[49] M. F. Mohammed and C. P. Lim, “Improving the fuzzy min-max neural network with a k-nearest hyperbox expansion rule for pattern classification,” *Appl. Soft Comput.*, vol. 52, pp. 135–145, 2017.

[50] M.F. Mohammed and C.P. Lim, “An enhanced fuzzy min-max neural network for pattern classification,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 3, pp. 417–429, Mar. 2015.

[51] M. F. Mohammed and C. P. Lim, “A new hyperbox selection rule and a pruning strategy for the enhanced fuzzy min–max neural network,” *Neural Netw.*, vol. 86, pp. 69–79, 2017.

[52] T. Kohonen, *Self-Organization and Associative Memory*. Berlin, Germany: Springer-Verlag, 1988.

[53] M. A. Litte *et al.*, “Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection,” *Biomed. Eng. Online*, vol. 6, no. 1, 2007, Art. no. 23.

[54] K. Bache and M. Lichman. (2013). *UCI Machine Learning Repository*. Irvine, CA, USA: School Inf. Comput. Sci., Univ. California. [Online]. Available: http://archive.ics.uci.edu/ml

[55] F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, “A dataset for breast cancer histopathological image classification,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1455–1462, Jul. 2016.

[56] A. L. Mescher, *Junqueiras Basic Histology: Text and Atlas*. New York, NY, USA: McGraw-Hill, 2013.

[57] R. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

Santhos Kumar A. received the M.E. degree in electronics and communication engineering (ECE) from S.V.P. Engineering College affiliated to Andhra University, Visakhapatnam, India, in 2012. He is currently working toward the Ph.D. degree in electronics and communication engineering with the Indian Institute of Information Technology, Design and Manufacturing (IIITDM) Jabalpur, Jabalpur, India.

His research interests include biomedical image processing and pattern classification.

Anil Kumar (M’15) received the B.E. degree in electronic and telecommunication engineering from the Army Institute of Technology, Pune University, Pune, India, in 2002, and the M.Tech. and Ph.D. degrees in electronic and telecommunication engineering from the Indian Institute of Technology Roorkee, Roorkee, India, in 2006 and 2010, respectively.

He is currently an Assistant Professor with the Electronics and Communication Engineering Department, Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Jabalpur, India. His research interests include the design of digital filters and filter banks, biomedical signal processing, image processing, and speech processing.

Varun Bajaj (M’17) received the B.E. degree in electronics and communication engineering (ECE) from Rajiv Gandhi Technological University, Bhopal, India, in 2006, the M.Tech. (Hons.) degree in microelectronics and very large scale integration (VLSI) design from the Shri Govindram Seksaria Institute of Technology and Science, Indore, India, in 2009, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology Indore, Indore, India, in 2014.

He is currently an Assistant Professor of ECE at the Indian Institute of Information Technology, Design and Manufacturing Jabalpur, Jabalpur, India. His research interests include biomedical signal processing, image processing, and time–frequency analysis.

Girish Kumar Singh received the B.Tech. degree in electrical engineering from the G.B. Pant University of Agriculture and Technology, Pantnagar, India, in 1981, and the Ph.D. degree in electrical engineering from Banaras Hindu University, Varanasi, India, in 1991.

He worked in the industry for nearly five and a half years. He is currently a Professor with the Electrical Engineering Department, Indian Institute of Technology Roorkee, Roorkee, India. His academic and research interests include the design and analysis of electrical machines and biomedical signal processing, power quality, and renewable energy.