Scalable Fuzzy Rough Set Reduct Computation Using Fuzzy Min–Max Neural Network Preprocessing

Anil Kumar and P. S. V. S. Sai Prasad, *Member, IEEE*

***Abstract*—A fuzzy rough set (FRS) is a hybridization of rough sets and fuzzy sets and provides a framework for reduct (feature subset selection) computation for hybrid decision systems. However, the existing FRS-based feature selection approaches are intractable for large decision systems due to the space complexity of the FRS methodology. We propose a novel fuzzy min–max neural network (FMNN)-FRS reduct computation approach utilizing the FMNN to enhance the scalability of FRS approaches. The FMNN provides a single pass epoch learning of arriving at granules of objects in the form of fuzzy hyperboxes for multiple decision classes. In the proposed approach, the FMNN model is used to reconstruct the object-based decision system into a fuzzy hyperbox-based interval-valued decision system. Then, a novel way of constructing the fuzzy discernibility matrix (FDM) from the interval-valued decision system is introduced. A fuzzy rough approximate reduct computation algorithm is developed with the induced FDM. The FMNN-FRS approach reduces the space complexity of FRS reduct computation significantly and results in enhanced scalability. Comparative experimental analysis has been done with the existing FRS reduct approaches on benchmark hybrid decision systems and established the relevance of the FMNN-FRS approach. The FMNN-FRS approach obtained the exact reduct in most of the datasets in much lesser computational time than existing FRS approaches while preserving similar classification accuracy. The FMNN-FRS method achieved enhanced scalability to such large decision systems, at which it is not possible to obtain reduct by existing FRS approaches.**

***Index Terms*—Discernibility matrix, feature subset selection, fuzzy min–max neural network (FMNN), fuzzy rough sets (FRSs), granular computing, hyperbox, reduct, rough sets.**

I. Introduction

MAKING a decision under imprecision and uncertainty is one of the most challenging topics in the field of data analysis. The objective of data analysis is to find or learn hidden patterns in a dataset, which is beneficial to find dependencies. Feature selection plays an essential role in analyzing the datasets when some of the features might be redundant/irrelevant degrading the performance and increasing the computational complexity of the model [1]. The selection of essential features after discarding irrelevant features is always a challenging task that preserves the discernment knowledge of datasets.

In the 1980s, Pawlak [2] introduced the concept of classical rough set theory (RST) as a mathematical tool for classification and analysis of incomplete and uncertain information. RST gave new momentum to data mining [3] and knowledge discovery [4] and provided a unique insight into artificial intelligence and cognitive sciences both in practical and theoretical views [1], [5].

Application of classical rough sets to numeric decision systems will produce feature subsets with finer granularity. Hence, the induced rules from the selected features suffer from poor generalizability to test datasets. So, one of the solutions is to discretize the dataset beforehand and produce a new dataset with categorical values [6]. However, the discretization method is often inadequate and causes essential information loss that can hamper the quality of subsequent feature subset selection. Lately, Dubois *et al.* [7], [8] generalized the RST that deals with symbolic and real-valued conditional attributes without the need for domain specific knowledge with fuzzy rough set (FRS) theory. The FRS can approximate the crisp decision concepts in the fuzzy approximation space.

The first pioneering work on feature selection based on the FRS was introduced by Jensen and Shen [9]. It performed well in terms of retaining fewer attributes with higher classification accuracy than RST-based reduction on web dataset, which aided in web categorization. In [9], the authors proposed an algorithm to compute close-to-minimal reduct based on the dependence function and also measure the quality of attributes. Subsequently, several aspects of improvement [10]–[12] based on feature selection and computation time were done for [9]. In [11], the authors introduced three robust techniques based on the fuzzy similarity relation and also developed the fuzzy discernibility matrix (FDM) for computing the feature selection. In particular, these techniques have shown high flexibility and reduced the complexity of computing the Cartesian product of fuzzy equivalence classes in [10]. This approach [11] received the consideration of researchers in [13]–[15] and became an effective approach for reduct computation.

Skowron and Rauszer [16] introduced a feature selection mechanism based on the concept of a crisp discernibility matrix in the context of Pawlak's RST. Jensen and Shen [11] further extended into the FDM to determine the FRS reducts. Though finding all/minimal reducts with these techniques is an NP-Hard problem, these methods provide a crucial mathematical

Manuscript received May 14, 2019; revised October 1, 2019 and December 11, 2019; accepted December 31, 2019. Date of publication January 13, 2020; date of current version May 4, 2020. This work was supported by the All India Council for Technical Education, Government of India, under RPS project under Grant File No. 8-47/RIFD/RPS/POLICY-1/2016-17. *(Corresponding author: Anil Kumar.)*

The authors are with the School of Computer and Information Sciences, University of Hyderabad, Hyderabad 500046, India (e-mail: anilhcu@uohyd.ac.in; saics@uohyd.ernet.in).

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

Digital Object Identifier 10.1109/TFUZZ.2020.2965899

foundation for reduct computation [11], [17]–[19]. Even though these approaches can be guaranteed to obtain the exact reduct, they require a substantial computational complexity for large datasets.

To mitigate the processing overhead on FRS approaches, in 2015, Jensen and Parthalin [20] presented two efficient approaches for reduct computation to improve the scalability on the large dataset by reducing the computation complexity. In [20], the authors reformulated the method of calculation for membership degree of an object to the k-nearest neighbor (KNN) objects of different classes in the dataset. A similar idea was developed in 2018 by Zhang *et al.* [21] by using representative instances. The idea is to utilize minimal or representative instances that can describe the objective of the entire object space. Thus, it presented the aim of increasing the classification accuracy and alleviate the computational complexity in search of reduct in large datasets. Both approaches [20], [21] result in an approximate reduct computation. The concept of the approximate reduct is introduced by Slezak [22] that contains the potential attributes to achieving near to exact reduct capability.

In 2018, Zhang *et al.* [23] provided a fast and efficient algorithm based on the existing FRS-based information entropy feature selection method [24]. The corresponding algorithm [23] computes in significantly less time for acquiring the same features as in [24], which is time consuming.

In 2018, Dai *et al.* [18] presented a diverse approach that transforms the property of FDM into the crisp discernibility matrix by restricting entries to only those attributes which have maximal discernibility. Then, Johnson's algorithm [25] was applied on the crisp discernibility matrix to achieve the reduct. In [18], the authors proposed two algorithms: RMDPS and WRMDPS.

The existing FRS approaches [9]–[12], [20] involve object-based computations but do not impose the granular or subgranular aspect in FRS-based reduct computations, which help in reducing the complexity of computation as in the FRS. The existing approximate reduct computation approaches require a prior generation of fuzzy similarity matrices having a memory requirement of $O(|U|^2|C|)$, where $|U|$ is the size of the object space and $|C|$ is that of attribute space. Hence, these approaches are still not scalable to large decision systems. The utility of approximate reduct computation in generating the useful feature subset with reduced time complexity is established in [20] and [21]. The need for reducing the space complexity for enhancing the scalability of FRS reduct computation is the motivation for the proposed approach.

Granular computing has an important role in computing for feature subset selection. A granule is a subset of data points that cannot be discerned among themselves. This process comprises partitioning of objects of a class into granules, which are indistinguishable [26]. In this article, such an intuitive idea is introduced for a solution to FRS feature subset selection by using the concept of hyperbox utilizing the fuzzy min–max neural network (FMNN) as a preprocessor.

In 1992, Simpson [27] proposed a new concept, "Fuzzy Min–Max Neural Network," which integrates fuzzy logic and neural networks for pattern classification. During the learning process, the FMNN creates hyperboxes, which are characterized by minimum point and maximum point with the corresponding fuzzy membership function. These min–max points are computed or adjusted using the FMNN learning process. The FMNN evokes a dynamic network structure with several salient learning features, such as online learning, nonlinear separability, and nonparametric classification. The main advantage of the FMNN is that it has the potential to learn approximate decision concepts through single pass training. The unique blend of one epoch learning combined with adaptable to incremental learning has made the FMNN suitable for current scenarios of building intelligent systems in an online environment. The FMNN is further enhanced in the direction for better representation of overlapping region among hyperboxes [28]–[30], optimization and refinement of resulting hyperboxes [31], [32], and in the development of other hybrid soft computing models efficiently [33], [34].

In this article, the proposed approach is formed of two steps. First is the construction of fuzzy hyperboxes with FMNN preprocessor results in the formation of an interval-valued decision system (IDS). Second is the creation of fuzzy discernibility relation for inducing an FDM based on the IDS to find the approximate reduct.

The first step is to construct the justifiable IDS in terms of fuzzy hyperboxes, which is semantically well-defined information. The main advantage of creating an IDS using an FMNN is that it retains the boundary information of overlapping intervals to each attribute in the original dataset. Unlikely, in the IDS resulting from discretization, the induced decision system will only contain the nonoverlapping interval values generated from arrived cut points having more information loss. Hence, the proposed IDS based on the FMNN is a novel contribution, as it achieves a significant reduction of the size of the dataset while retaining overlapping interval information of each attribute. In our article, we have adopted a simplified version of the FMNN preserving the requisite information for FDM construction in the second step.

The second step of the proposed approach is to define the fuzzy discernibility relation based on the IDS to induce the FDM. Each clause in the FDM contains a set of every feature that discerns to a certain degree. The extent of a feature belonging on the particular clause is determined by using the standard negation of Jaccard's similarity measure (JS) [35]. This similarity measure is used to estimate the similarity between intervals. Each clause in the FDM is a fuzzy discernibility between the intervals. Here, the selection of FDM is motivated from [11] to provide the approximate reduct. Then, the degree of satisfiability of all clauses is used in the FDM to find a single approximate reduct.

The whole motivation behind this proposed article, FMNN-FRS, is to use IDS-based FDM construction to reduce the significant computational time and compress the space complexity of the FDM for finding an approximate reduct. The quality of the resulting approximate reduct from the FMNN-FRS is assessed extensively based on the obtained gamma measure and generalizability of induced different classifiers. The approximate reduct computed in this manner has empirically been demonstrated to achieve 0.90–0.99 gamma measure of the total reduct. In this article, the size of the IDS is much smaller than the cardinality for the original dataset. Hence, we ascertain that FMNN-based

FRS reduct computation is furthermore scalable to those datasets on which existing approaches for reduct computation are not tractable.

The robustness of the proposed algorithm FMNN-FRS is demonstrated based on comparison with RMDPS [18], WR-MDPS [18], and FRS-Entropy [23]. And, a comparative study is also conducted to validate the effectiveness of the FMNN-FRS using two classifiers: kNN and CART.

The remaining part of this article is organized as follows. Section II briefly introduces the basics of FRSs, FDM, and FMNN. Section III presents the proposed FMNN-FRS reduct algorithm. Section IV presents the experiments and analysis of comparative results with other FRS approaches. Section V concludes this article.

II. Preliminaries

In this section, we briefly review some basic notions about FRSs with its corresponding reduct computation method and the FDM. Also, we define the concepts of the FMNN.

A. Fuzzy Rough Sets

Let the decision system DT $= (U, C \cup \{d\})$, where U is a nonempty finite set of objects, $D = \{d\}$ is a set of decision attributes, and C is a nonempty finite set of quantitative attributes. The approximate similarity between two objects w.r.t "a" attribute is obtained through the fuzzy similarity relation R_a [11]. Few examples of R_a for $(x, y) \in U$, where σ_a is the standard deviation of "a" attribute, are given as

$$\mu_{R_a}(x, y) = \exp\left(-\frac{(a(x) - a(y))^2}{2\sigma_a^2}\right) \tag{1}$$

$$\mu_{R_a(x,y)} = \max\left(\min\left(\frac{a(y) - a(x) + \sigma_a)}{\sigma_a}, \frac{a(x) - a(y) + \sigma_a)}{\sigma_a}\right), 0\right). \tag{2}$$

The similarity relation is extended to any subset of $B \subseteq C$ by

$$\mu_{R_B}(x, y) = \Gamma_{a \in B}\{\mu_{R_a}(x, y)\} \tag{3}$$

where Γ represents a t-norm. Γ is an increasing, commutative, and associative property with $[0, 1]^2 \to [0, 1]$ mapping satisfying $\Gamma(x, 1) = x, \forall x \in [0, 1]$. In the FRS, the concept of lower and upper approximations on the fuzzy concept X is given in Radzikowska–Kerry's FRS model [36] based on the fuzzy similarity relation R as follows:

$$\mu_{\underline{R_B}X}(x) = \inf_{y \in U} \Im(\mu_{R_B}(x, y), \mu_X(x)) \tag{4}$$

$$\mu_{\overline{R_B}X}(x) = \sup_{y \in U} \Gamma(\mu_{R_B}(x, y), \mu_X(x)) \tag{5}$$

where $\forall y \in U$, $\Im$ and Γ denote the fuzzy implicator the and t-norm, respectively. A fuzzy implicator $\Im$ is a $[0, 1]^2 \to [0, 1]$ mapping that satisfies left ($\Im(x, .)$ decreases, $\forall x \in [0, 1]$) and right monotonic ($\Im(x, .)$ increases, $\forall x \in [0, 1]$), which satisfies $\Im(0, 0) = \Im(0, 1) = \Im(1, 1) = 1$ and $\Im(1, 0) = 0$. Equations (4) and (5) are the generalized version of lower and upper approximations in a classical rough set model. For $x \in U$, the membership of an object x belonging to the fuzzy positive region defined is as

$$\mu_{\text{POS}_B(D)}(x) = \sup_{X \in U/D} \mu_{\underline{R_B}X}(x). \tag{6}$$

Using the fuzzy positive region, the fuzzy dependence function γ on the subset B can be defined as

$$\gamma_B(D) = \frac{\sum_{x \in U} \mu_{\text{POS}_B(D)}(x)}{|U|}. \tag{7}$$

The fuzzy dependence function evaluates the significance of the subset of features that preserve the dependence degree of the complete dataset. B is said to be a reduct of the decision system DT, if and only if $\gamma_B(D) = \gamma_C(D)$ and, for any $B' \subset B$, $\gamma_{B'}(D) < \gamma_C(D)$.

The time and space complexities of the FRS-based approach for reduct computation are $(O|C|^2|U|^2)$ and $(O|C||U|^2)$, respectively, and hence are not scalable to large quantitative decision systems.

B. Decision-Relative Fuzzy Discernibility Matrix

Jensen and Shen [11] extended the concept of crisp discernibility matrix [16] based on RST into the FDM for use in FRS reduct computation. Each entry (known as clauses) in the FDM, denoted by M_{ij}, stores a set of all conditional attributes with certain degrees of discernibility, meaning that each entry is a fuzzy set. A feature "a" belongingness to the fuzzy clause M_{ij} is determined by the fuzzy discernibility measure:

$$\mu_{M_{ij}}(a) = N(\mu_{R_a}(i, j)). \tag{8}$$

From (8), each entry in the FDM is determined as

$$M_{i,j} = \{a_x |\ a \in C, x = N(\mu_{R_a}(i, j))\} \quad \forall i, j \in U \tag{9}$$

where N is a fuzzy negator, and $\mu_{R_a}(i, j)$ is the fuzzy indiscernibility measure between i and j objects w.r.t attribute "a." $N(\mu_{R_a}(i, j))$ is the measure of the fuzzy discernibility w.r.t "a" attribute. The entry in M_{ij} means that objects i and j are discerning with a certain degree in range between 0 and 1 with regard to all attributes. For example, an entry in the FDM might be $\{a_{0.34}, b_{0.45}, c_{0.67}\}$. Using (8) and (9), the fuzzy discernibility function f'_D w.r.t the FDM is defined as follows:

$$f'_D(a_1, \ldots, a_n) = \wedge\{\vee M_{i,j} \leftarrow d_{N(\mu_{R_D}(i,j))} \mid \forall i, j \in U\}. \tag{10}$$

Here, $\leftarrow$ is the fuzzy implication. For considering the decision attributes, the construction allows only those clauses that have different decision values for achieving reduction. If $\mu_{R_D}(i, j) = 1$, then objects i and j have the similar decision class; otherwise, objects i and j have different decision classes.

For finding reducts, the degree of satisfaction of a clause $M_{i,j}$ for a given subset of features B w.r.t the decision attribute "d" is defined as

$$\text{SAT}_{B,d}(M_{ij}) = S_{a \in B}\{\mu_{M_{i,j}}(a)\} \leftarrow \mu_{M_{i,j}}(d) \tag{11}$$

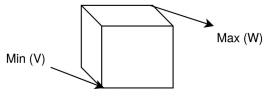

Fig. 1. Hyperbox.

where S is a t-conorm. Hence, the total satisfiability of all clauses for B is defined as

$$\mathrm{SAT}(B) = \frac{\sum_{i,j\in U, i\neq j} \mathrm{SAT}_{B,d}(M_{i,j})}{\sum_{i,j\in U, i\neq j} \mathrm{SAT}_{C,d}(M_{i,j})} \tag{12}$$

where C is the set of all attributes. A subset of conditional attributes $B \subseteq C$ is called as reduct if and only if it satisfies the following conditions:

1) $\mathrm{SAT}(B) = \mathrm{SAT}(C) = 1$; and
2) $\forall B' \subset B, \mathrm{SAT}(B') < \mathrm{SAT}(B)$.

C. Fuzzy Min–Max Neural Network

In 1992, Simpson [27] proposed a supervised learning neural network classifier known as FMNN having salient features as online learning, nonlinear separability, and nonparametric classification. The FMNN evokes a single pass dynamic network structure to deal with pattern classification using n-dimensional hyperbox fuzzy sets to represent the pattern spaces.

Each hyperbox B_j, as shown in Fig. 1, represents a subregion in the n-dimensional space that is characterized by minimum point and maximum point with the corresponding fuzzy membership function that shows the membership degree of the sample concerning particular fuzzy hyperbox, defined as

$$B_j = \{X, V_j, W_j, f(X, V_j, W_j)\} \quad \forall X \in I^n \tag{13}$$

where X is the input pattern, V_j and W_j are the minimum and maximum points of hyperbox B_j, and I^n is the n-dimensional unit pattern space. The fuzzy membership function is defined as

$$b_j(X_h) = \frac{1}{2n}\sum_{i=1}^{n}[\max(0, 1 - \max(0, \gamma \min(1, x_{hi} - w_{ji}))) + \max(0, 1 - \max(0, \gamma \min(1, v_{ji} - x_{hi})))] \tag{14}$$

where $X_h = (x_{h1}, x_{h2}, \ldots, x_{hn})$ is the hth input pattern in the n-dimensional space, and $V_j = (v_{j1}, v_{j2}, \ldots, v_{jn})$ and $W_j = (w_{j1}, w_{j2}, \ldots, w_{jn})$ are the corresponding minimum point and maximum point for hyperbox B_j. γ is the sensitive parameter that regulates how fast the membership decreases as the distance between X_h and B_j increases.

FMNN training is a single epoch process involving three stages for acquiring knowledge: 1) expansion process; 2) overlap test; and 3) contraction process. During the training phase, each input pattern enters into the network and tries to accommodate one of the existing same-class hyperboxes that provide the full fuzzy membership degree.

Otherwise, the network finds the hyperbox corresponding to the same decision class in the pattern that having the highest degree of membership for the input pattern class. The selected hyperbox is expanded to accommodate the input pattern subject to the following criterion:

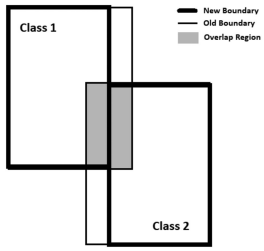

Fig. 2. Overlapped region by two hyperboxes.

$$\sum_{i=1}^{n}(\max(w_{ji}, x_{hi}) - \min(v_{ji}, x_{hi})) \leq n\theta \tag{15}$$

where theta (θ) is the user-defined parameter with a range for $(0 < \theta < 1)$ controlling the size of a hyperbox.

If the expansion criterion in (15) is satisfied, then the hyperbox expands to incorporate the input pattern by adjusting the minimum and maximum points by using the following:

$$v_{ji}^{\text{new}} = \min(v_{ji}^{\text{old}}, x_{hi}) \quad \forall i = 1, 2, 3, \ldots, n \tag{16}$$

$$w_{ji}^{\text{new}} = \max(w_{ji}^{\text{old}}, x_{hi}) \quad \forall i = 1, 2, 3, \ldots, n. \tag{17}$$

If the condition (15) is not satisfied, then a point hyperbox is created with minimum and maximum points set to the input pattern.

After the expansion process, the overlap test and the contraction step are computed as described in [27]. The overlap test tries to determine the overlap among a different class of hyperboxes during the expansion. If there is an overlap in each dimension (two hyperboxes have some common region), the test memorizes the smallest overlap along any dimension. The same is then used during a contraction step to result in nonoverlapping hyperboxes.

In Fig. 2, both hyperboxes H_1 and H_2 have an overlapping region (shaded region). The overlap test determines the smallest overlap along the horizontal dimension, and then, the contraction step adjusts the hyperboxes along the smallest overlap dimension (shown in bold outline).

FMNN learning achieves nonlinear separability among patterns of multiple decision classes by visiting each training pattern once. The approach seamlessly extends for incremental learning as the incoming training points go through these steps and hyperboxes are altered appropriately, incorporating the new information. The main advantage of the FMNN is that it produces approximate classification very fast, and the knowledge represented through fuzzy granules of hyperboxes can be utilized in the construction of other soft computing models efficiently.

III. Proposed FMNN-FRS Reduct Algorithm

The proposed FMNN-FRS reduct approach aims to compute the approximate fuzzy rough reduct efficiently in terms of space and time complexities. In the FMNN-FRS reduct approach, we

Algorithm 1: Creating Hyperboxes of Different Classes.

Input : DT: Training Samples, γ, θ
Output: HBS: Collection of hyperboxes of different classes, Learning model FMN.

```
Let HBS = {φ};
for every x in DT do
    if FMN.Belongs(x) ≠ ∅ then
        Next;
    else
        H = FMN.HMemb(x);
        if H exist then
            if Exp_H(x) == True then
                FMN.Expand(H,x);
                FMN.Update(HBS,H);
            else
                H_N = FMN.Create(x);
                FMN.Save(HBS,H_N);
            end
        else
            H_N = FMN.Create(x);
            FMN.Save(HBS,H_N);
        end
    end
end
Return HBS
```

have introduced a solution for FRS feature subset selection, utilizing the FMNN learning model [27] as a preprocessor.

In this section, we describe the creation of the IDS system based on fuzzy hyperboxes that are generated from the FMNN learning model. We will then elaborate on the construction of an FDM based on the IDS to find an approximate reduct. Each of these stages involved in the FMNN-FRS is given in Algorithms 1 and 2, respectively.

A. Creation of Fuzzy Hyperboxes

The traditional FMNN algorithm, as described in Section II-C, has a three-step approach of expansion, overlap, and contraction for each training pattern. The overlapping and contraction steps result in nonoverlapping between the pair of hyperboxes belonging to different decision classes. This disambiguation helps in crisp decision making for classification but also results in crucial information loss of boundary region between decision classes. In this article, our objective of the FMNN preprocessor is to aid in the construction of the FDM. But, following the traditional procedure of the FMNN may lose valuable information for the representation of discernibility among the objects of different classes. Hence, we have a simplified FMNN training procedure to contain only the expansion step for preserving the naturally overlapping regions among the hyperboxes of multiple decision classes.

Algorithm 1 gives the simplified FMNN training process for arriving at hyperboxes with possible overlap among multiple decision classes.

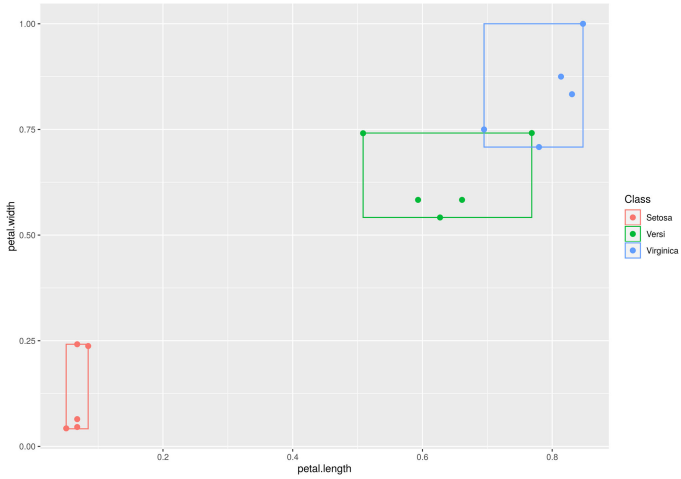

Fig. 3. Example of FMNN-FRS reduct computation on the toy dataset.

Let DT represent the set of the training patterns, HBS is the set of hyperboxes (initially empty), and FMN represents the FMNN learning model. For every input pattern x belonging to the DT, Belongs(x) tries to find the fuzzy membership value of x with all hyperboxes of the same decision class of x using (14) and determine the hyperboxes giving full membership of one to x. Memb$_H(x)$ denotes membership value of x on H

$$\text{Belongs}(x) = \{H \in \text{HBS} |\ \text{Memb}_H(x) == 1\ \&\ \text{class}(x) == \text{class}(H)\}. \tag{18}$$

If Belongs(x) is not empty, then x is added to the hyperbox giving the full membership without any modification of HBS. Otherwise, HMemb(x) obtains the hyperbox H that provides the highest degree membership among the hyperboxes in HBS of the same class. In that case, the input pattern x strives to accommodate within hyperbox H using the Expand(H, x) method based on (15). If the expansion criteria (15) are satisfied, the hyperbox H is expanded using (16) and (17) to include x. If the expansion criteria are not being met, then a point hyperbox is created using Create(x), and x is included in the point hyperbox and the resulting H_N is added to HBS.

1) Illustration of Fuzzy Hyperboxes Creation Using a Toy Dataset: We are illustrating the FMNN algorithm with a sample of the iris dataset [37] (15 objects, two attributes (petal.length and petal.width), and three decision classes). We have taken five objects belonging to each decision classes in construction of the toy decision system. Applying Algorithm 1 on the toy dataset resulted in three hyperboxes, as shown in Fig. 3. Class setosa, class versi, and class virginica represented in red, green, and blue colors, respectively, in Fig. 3. Here, we constructed the 2-D geometry visualization between petal.length and petal.width attributes, and the hyperboxes were created using Algorithm 1.

B. FMNN-Preprocessor-Based FDM

FMNN preprocessing results in hyperboxes, where each hyperbox represents granular objects of the corresponding decision class, which are similar to each other. Objects of the hyperbox

H are the objects having absolute membership of 1. As the objects are in the nearby vicinity, there are chances that most of them represent single decision class, but some exceptions can exist, as overlapping among hyperboxes cannot be avoided, as described in Section III-A. Here, we are constructing the IDS based on hyperboxes, which can retain the boundary information of overlapping intervals to each attribute in the decision system DT. The hyperbox is bounded by V (minimum point) and W (maximum point) that represents the area in space belonging to a particular decision class. This representative hyperbox is taken as a single entity for representing the member objects and becomes an object in the resulting IDS.

Let IDS $=$ (HBS, $C \cup \{d\}$) be an IDS, where HBS $= \{H_1, H_2, \ldots, H_r\}$ represents the universe of hyperboxes. Let $[V^H, W^H]$ represent the minimum and maximum points of hyperbox H. In the IDS, the value of an hyperbox $H \in$ HBS over an attribute $a \in C$ is represented by the interval v_a^H to w_a^H ($[v_a^H, w_a^H]$), where, v_a^H is the component of the minimum point V^H and w_a^H is the component of the maximum point W^H corresponding to the attribute a. The value of the decision attribute d is taken as per the decision class to which H belongs.

Here, we are constructing the FDM based on the IDS. Each clause in the FDM corresponds to a pair of hyperboxes. Based on (9), an entry corresponding to hyperboxes H_i and H_j is a vector of fuzzy discernibility measure for all attributes. In the construction of the FDM for the decision system, the valid entries are defined for a pair of hyperboxes belonging to different decision classes. To find fuzzy discernibility measure, we require a fuzzy similarity measure applicable to interval-valued data. Several similarity measures are defined in the literature for interval-valued data [35], [38]. Out of these, JS [35] is used for the proposed article.

JS [35] introduced the concept of similarity measure for interval-valued data based on real numbers. It satisfies the boundness, symmetry, reflexivity, and transitivity properties of a similarity measure. Hence, Jaccard's similarity is a fuzzy equivalence relation defined over the universe of interval-valued data objects. Let I_x and I_y represent two overlapping intervals. The Jaccard's similarity measure JS(I_x, I_y) is defined as

$$\text{JS}(I_x, I_y) = \frac{|I_x \cap I_y|}{|I_x \cap I_y| + |I_x \setminus I_y| + |I_y \setminus I_x|} \tag{19}$$

where $|I_x \cap I_y|$ is the size of intersection between I_x and I_y. $|I_x \setminus I_y|$ is the size of the interval segment of I_x that are not overlapping with I_y. Similarly, $|I_y \setminus I_x|$ is the size of the interval segment of I_y that are not overlapping with I_x. If I_x and I_y are not overlapping to each other, then JS$(I_x, I_y) = 0$ indicates that both intervals are completely different from each other. Likewise, if I_x and I_y are fully overlapping to each other, then JS$(I_x, I_y) = 1$ indicates that both intervals are completely identical to each other.

Using JS, the FDM entry between H_i and H_j belonging to different classes is defined as

$$\begin{aligned}&\text{FDM}(H_i, H_j) = \{\ a_s \mid a \in C,\\ &s = N(\text{JS}([v_a^{H_i}, w_a^{H_i}], [v_a^{H_j}, w_a^{H_j}]))\}\ H_i, H_j \in \text{HBS and}\\ &d(H_i) \neq d(H_j).\end{aligned} \tag{20}$$

TABLE I
IDS BASED ON THE TOY DATASET

Hyperboxes	petal.length	petal.width	Class
H_1	[0.0508,0.0847]	[0.0427,0.2417]	setosa
H_2	[0.5085,0.7685]	[0.5417,0.7415]	versi
H_3	[0.6949,0.8475]	[0.7083,1.0000]	virginica

TABLE II
FDM BASED ON THE IDS

	H_1	H_2	H_3
H_1	$\emptyset$	$\{a_1, b_1\}$	$\{a_1, b_1\}$
H_2	$\{a_1, b_1\}$	$\emptyset$	$\{a_{0.78}, b_{0.93}\}$
H_3	$\{a_1, b_1\}$	$\{a_{0.78}, b_{0.93}\}$	$\emptyset$

a : petal.length, b : petal.width.

TABLE III
DETAILS OF BENCHMARK DATASETS

Dataset	Attributes	Objects	Decision classes
Ionosphere	32	351	2
Movement & libras	90	360	15
Sonar	60	208	2
LSVT	310	126	2
Musk1	166	476	2
WDBC	30	569	2
Segment	16	2310	7
Steel	27	1941	7
Page block	10	5472	5
Vehcile	18	846	4
Water	38	521	3
Waveform	21	5000	3
Colon	2000	62	2
Leukemia	7129	72	2

The component corresponding to $a \in C$ is

$$\text{FDM}^a(H_i, H_j) = N\left(\text{JS}\left(\left[v_a^{H_i}, w_a^{H_i}\right], \left[v_a^{H_j}, w_a^{H_j}\right]\right)\right) \tag{21}$$

where N denotes the fuzzy negation.

The FDM constructed in this manner is an approximation to the FDM constructed at an object level, i.e., an FDM entry between pair of objects of different classes is always a superset of the corresponding FDM entry between the hyperboxes containing these objects. Hence, the fuzzy rough reduct computed using the FDM for the IDS results as an approximate reduct.

1) Illustration of FDM Construction: The IDS is constructed using the fuzzy hyperboxes created in Section III-A1 and is given in Table I. Based on the IDS, the FDM is constructed corresponding to hyperboxes (H_1, H_2, H_3) of different decision classes, as shown in Table II. In Table II, a and b represent the petal.length and petal.width attributes, respectively.

C. Approximate Reduct Computation Based on the FDM

In this section, we provide the approximate reduct computation algorithm using the FDM constructed on the IDS, as given in Section III-B. Algorithm 2 gives the procedure for computing an approximate reduct based on the FDM. Algorithm 2 follows the sequential forward selection control strategy and starts with reduct Red initialize to an empty set. In each iteration, SAT measure is computed (12) for (Red $\cup \{a\}$) $\forall a \in C -$ Red. The

Algorithm 2: Finding an Approximate Reduct.

Input : FDM :Fuzzy Discernibility Matrix of IDS, C: Conditional Attributes
Output: Reduct: Red

```
Red = ∅, SAT(Red) = 0, SAT(C) = 1 ;
while SAT(Red) ≠ SAT(C) do
    BestSat = 0, BestAttribute = {};
    for each a ∈ C − Red do
        S_a = SAT(Red ∪ {a}) from (12);
        if S_a > BestSat then
            BestSat = S_a;
            BestAttribute = {a};
        end
    end
    Red = Red ∪ BestAttribute;
end
Return Red
```

attribute achieving maximum SAT measure is included in the reduct set Red. The algorithm terminates when SAT(Red) becomes equal to SAT(C) (i.e., 1) and returns the obtained reduct Red.

The relevance of obtained approximate reduct by the FMNN-FRS with the aspects of computational time, reduct size, reduct quality, and utility of reduct for classification is studied in depth through experimentation and comparative study in Section IV.

1) Illustration of FMNN-FRS Reduct Computation: Algorithm 2 is applied on the FDM given in Table II. After the first iteration, SAT values for a and b are

$$\text{SAT}(\{a\}) = \frac{2.78}{3} = 0.926, \quad \text{SAT}(\{b\}) = \frac{2.93}{3} = 0.9766.$$

Attribute b is added into the reduct set on the basis of maximal satisfiability. After the second iteration, a is added to reduct set for achieving SAT(C). Hence, the approximate reduct for the toy dataset is $\{a, b\}$.

IV. Experiments

A. Environment and Objectives of Experimentation

The hardware configuration of the system used for experiments is CPU: Intel(R) i5-7500, Clock Speed: 3.40 GHz $\times$ 4, RAM: 8-GB DDR4, OS: Ubuntu 16.04 LTS 64 bit, and Software: MATLAB R2017a. The detailed experimental evaluation is conducted on 14 benchmark numeric decision systems, details given in Table III, from UCI machine learning repository [37], which include two high-dimensional datasets (Colon and Leukemia) from Bioinformatics research group [39] repository. The proposed algorithm FMNN-FRS is implemented in MATLAB environment. In our experiments, we set the sensitive parameter γ value equal to 4, as recommended [27], [30]. Through empirical experiments, we deduced that $theta$ values of 0.2 and 0.3 are appropriate in computation FMNN-FRS reduct. In the FMNN-FRS experiment, Lukasiewicz t-conorm $(S(x, y) = \min\{x + y, 1\}$ for (11) and fuzzy standard negation $(N(x) = 1 - x)$ for (20) are used.

TABLE IV
Gaussian Kernel FRS-Based Gamma Measure for FRS Reducts

S. No	Datsets	Gamma Meausre					
		Unred	FMNN-FRS(0.2)	FMNN-FRS(0.3)	RMDPS	WRMDPS	FRS-Entropy
1	Ionosphere	1	0.94	0.98	1	1	1
2	Move. & Libra	1	0.95	0.98	0.99	0.99	0.99
3	Sonar	1	0.97	0.99	1	1	1
4	LSVT	1	0.84	0.73	1	1	1
5	Musk1	1	0.90	0.87	1	1	1
6	WDBC	1	0.94	0.90	1	1	1
7	Segment	0.98	0.97	0.95	0.98	0.98	0.97
8	Steel	0.99	0.99	0.99	0.99	0.99	0.99
9	Page Block	0.87	0.87	0.85	0.87	0.87	0.87
10	Vehicle	0.99	0.99	0.99	0.99	0.99	0.99
11	Water	1	0.99	0.99	1	1	1
12	Waveform	1	1	1	1	1	1
13	Colon	1	0.92	0.83	1	1	1
14	Leukemia	1	0.98	0.98	1	1	1

The performance of the proposed algorithm FMNN-FRS is assessed by comparing it with some of the recent approaches developed for FRS reduct computation in 2018, such as RMDPS [18], WRMDPS [18], and FRS-Entropy [23]. The comparative experiments are conducted in the same system using MATLAB environment. The performance of the FMNN-FRS is examined through a comparative evaluation with respect to the following objectives:

1) To measure the quality of approximate reduct computation by the FMNN-FRS approach.
2) To determine the significance of FMNN-FRS reduct for aiding in the construction of classifier through tenfold cross-validation.
3) To assess the enhanced scalability of the FMNN-FRS algorithm for large datasets.

B. To Evaluate the Quality of Approximate Reduct Computed by the FMNN-FRS Approach

Reduct computation in the FMNN-FRS is based on an FDM. Since the FDM based on the IDS is a fuzzy approximation of the original decision system, the FMNN-FRS theoretically results in an approximate reduct.

The objective of this section is to assess the quality of approximate reduct obtained based on validation by computing the obtained gamma measure by the approximate reduct over the original decision system. In the experiments, the representative algorithms RMDPS, WRMDPS, and FRS-Entropy follow their own FRS model with a specific fuzzy similarity relation, t-norm, and t-conorm in the case of the discernibility matrix approach. To validate the relevance of reduct quality comparison, we need to utilize a different FRS model so that the comparisons of gamma measures are the same for each method. Hence, the Gaussian kernel FRS [40] is used for computation of gamma measure by reducts from the compared algorithms.

Table IV contains the resulting gamma measure of reducts obtained by applying the proposed algorithm as well as the compared algorithms on the entire datasets. Also, Table IV represents the gamma measure obtained from the unreduced

TABLE V
TENFOLD CROSS-VALIDATION EXPERIMENT RESULTS FOR CLASSIFICATION USING CART

Datasets	FMNN-FRS(0.3)	FMNN-FRS(0.2)		RMDPS		WRMDPS		FRS-Entropy		Unred	
	Mean ± Std	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val
Ionosphere	87.75 ± 4.03	88.01 ± 4.24	0.90^o	88.03 ± 4.22	0.89^o	88.03 ± 4.00	0.88^o	86.03 ± 6.67	0.46^o	87.74 ± 5.05	1.00^o
Mov.&Libras	52.22 ± 5.67	53.88 ± 6.82	0.40^o	62.77 ± 7.76	0.02^+	62.77 ± 7.19	0.01^+	64.16 ± 10.51	0.02^+	63.61 ± 6.86	0.00^+
Sonar	72.28 ± 13.02	68.64 ± 6.73	0.50^o	75.64 ± 10.32	0.38^o	70.50 ± 8.64	0.69^o	70.64 ± 10.12	0.56^o	72.42 ± 7.98	0.97^o
LSVT	76.11 ± 11.43	75.83 ± 8.28	0.95^o	72.50 ± 9.66	0.47^o	79.16 ± 9.82	0.51^o	74.16 ± 13.29	0.73^o	75.27 ± 8.82	0.88^o
Musk1	78.75 ± 5.32	77.75 ± 7.06	0.73^o	78.47 ± 5.23	0.91^o	76.88 ± 6.45	0.29^o	82.60 ± 7.11	0.19^o	80.35 ± 8.30	0.62^o
WDBC	91.90 ± 3.05	93.67 ± 3.04	0.08^o	93.23 ± 3.68	0.25^o	93.06 ± 3.84	0.34^o	92.88 ± 3.19	0.44^o	92.75 ± 2.79	0.40^o
Segment	95.41 ± 2.05	95.67 ± 1.30	0.70^o	95.49 ± 1.26	0.89^o	95.80 ± 1.27	0.57^o	95.54 ± 1.15	0.84^o	95.45 ± 1.04	0.95^o
Steel	91.55 ± 2.39	91.24 ± 2.18	0.62^o	91.34 ± 2.28	0.68^o	91.34 ± 2.15	0.66^o	90.16 ± 2.26	0.04^-	91.55 ± 1.95	1.00^o
Page Block	96.41 ± 0.37	96.36 ± 0.63	0.75^o	96.36 ± 0.55	0.73^o	96.32 ± 0.55	0.54^o	96.19 ± 0.56	0.14^o	96.36 ± 0.45	0.64^o
Vehicle	69.42 ± 4.91	69.42 ± 5.59	0.99^o	73.24 ± 5.92	0.03^+	73.12 ± 5.10	0.04^+	68.98 ± 5.38	0.75^o	71.38 ± 5.94	0.29^o
Water	81.75 ± 6.24	85.19 ± 7.08	0.19^o	80.22 ± 3.99	0.45^o	79.64 ± 2.95	0.22^o	82.90 ± 4.33	0.47^o	83.47 ± 4.10	0.35^o
Waveform	76.26 ± 2.18	73.90 ± 2.41	0.04^-	74.72 ± 2.54	0.26^o	74.68 ± 2.65	0.18^o	74.76 ± 2.33	0.18^o	74.96 ± 2.49	0.17^o
Colon	72.91 ± 19.76	67.91 ± 12.43	0.39^o	72.08 ± 18.00	0.87^o	72.08 ± 18.00	0.90^o	77.50 ± 22.23	0.53^o	69.58 ± 17.35	0.50^o
Leukemia	88.88 ± 13.08	86.03 ± 11.70	0.17^o	87.14 ± 14.20	0.66^o	84.28 ± 17.10	0.23^o	91.42 ± 9.98	0.55^o	80.31 ± 12.25	0.05^-
Average	80.82	80.25		81.51		81.26		81.99		81.08	
Lose/Win/Tie		1/0/13		0/2/12		0/2/12		1/1/12		1/1/12	

Each column entry is $\mu \pm \sigma$ of classification accuracy.

decision system (mention as "Unred" in Table IV) to validate the relevance of resulted reducts through checking whether the obtained reduct is satisfying or reaching near to (Unred) gamma measure or not.

Analysis of Results: In Table IV, especially in waveform, vehicle, and steel datasets, it is observed that the FMNN-FRS (0.3 and 0.2) has achieved an equal gamma measure as obtained by "Unred" satisfying the required reduct property fully. In the remaining datasets, the FMNN-FRS (0.3 and 0.2) has indeed achieved near to expected gamma measure w.r.t the entire dataset gamma value with a range of 0.95–0.99 except in LSVT, musk1, and colon datasets.

Overall, it can be seen that the approximate reduct from the FMNN-FRS method is not resulting in any significant loss in the quality of reduct except in a few datasets. The compared algorithms have also achieved the relevant or approximate gamma measure in given datasets, but even that approximation is negligible as in the case of the FMNN-FRS method. Hence, empirically, we have established the FMNN-FRS results in quality reduct with an insignificant reduction in the gamma measure.

Section IV-C explores the relevance of obtained approximate reduct of the FMNN-FRS in achieving the construction of the classification learning model, which is the primary objective of the feature subset selection. Moreover, the comparative analysis with reduct length and computational time will be elaborated as part of Section IV-C in tenfold cross-validation.

C. Relevance of the Proposed Approach in Construction of Classifiers

This section contains the comparative experiments conducted among algorithms for reduct computation, i.e., among FMNN-FRS(0.3), FMNN-FRS(0.2), RMDPS, WRMDPS, and FRS-Entropy algorithms. The relevance of reduct in inducing a classification model is studied through tenfold cross-validation experiments. In each fold of experiments, one fold is preserved for the testing data, and the remaining nine folds are used for training data. A reduct algorithm is applied to the training data. So, based on the reduct that is obtained, the classification model is constructed for comparison. The classification accuracy of the resulting model is evaluated based on the test data. Two different classifier models are used, namely CART and kNN with default options, and for kNN experiments, k is taken as 3. To examine the relevance of reducts, we are also constructed the classification model with an unreduced dataset (mentioned as "Unred" in the given tables) for comparison.

Tables V and VI present the results of the tenfold experiment for classification accuracies with CART and kNN, respectively. Similarly, Tables VII and VIII illustrate the computational time and the reduct length of the algorithms, respectively. Student's paired t-test is performed in order to evaluate the statistical significance of FMNN-FRS(0.3) algorithm with FMNN-FRS(0.2), RMDPS, WRMDPS, and FRS-Entropy and Unred. Each of these tables contains the mean and standard deviation of the respective measure value for FMNN-FRS(0.3) on all datasets. In the result reported for the remaining algorithms, we present mean and standard deviation along with t-test results of the measured value. The p-value obtained in t-test experiment is provided in the table. If the p-value is less than equal to 0.05 and average performance is less than FMNN-FRS(0.3), then the compared algorithm is performing significantly inferior to FMNN-FRS(0.3) and symbolically reported as a loss ("$-$"). Else, it is represented as a win ("+"). If $p > 0.05$, then both approaches are statistically similar and represented as a tie with the symbol of "$\circ$." For example, In Table V, the p-value column of FRS-Entropy shows the "$-$" sign in the waveform dataset, which means that FMNN-FRS(0.3) is performing significantly inferior to FRS-Entropy. In each table, the last two lines summarize the total average results for all datasets and the count of the number of statistically better ("+"), loss ("$-$"), and equivalent ("$\circ$") for each method in comparison with FMNN-FRS(0.3).

Analysis of Results: Based on the results reported in Table V for CART-based classification, the proposed FMNN-FRS(0.3) obtained statistically similar results with compared algorithms and Unred for all datasets except in movement & libras and vehicle datasets. The compared algorithms RMDPS and WRMDPS have statistical better accuracies than FMNN-FRS(0.3) in both movement & libras and vehicle datasets.

TABLE VI
TENFOLD CROSS-VALIDATION EXPERIMENT RESULTS FOR CLASSIFICATION USING KNN

Datasets	FMNN-FRS(0.3)	FMNN-FRS(0.2)		RMDPS		WRMDPS		FRS-Entropy		Unred	
	Mean ± Std	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val
Ionosphere	87.16 ± 6.08	86.29 ± 9.51	0.77^o	83.47 ± 6.56	0.08^o	83.46 ± 6.98	0.10^o	85.17 ± 7.73	0.37^o	84.32 ± 7.16	0.21^o
Mov.&Libras	69.72 ± 9.66	75.00 ± 9.97	0.16^o	76.94 ± 6.00	0.02^+	79.16 ± 4.58	0.00^+	80.00 ± 6.90	0.00^+	81.11 ± 4.68	0.00^+
Sonar	79.50 ± 8.31	77.35 ± 8.86	0.51^o	82.71 ± 7.85	0.35^o	82.35 ± 11.88	0.50^o	84.35 ± 8.73	0.20^o	82.85 ± 9.54	0.38^o
LSVT	78.33 ± 13.72	74.44 ± 12.12	0.52^o	81.66 ± 14.59	0.57^o	80.27 ± 11.14	0.67^o	81.66 ± 9.46	0.34^o	79.44 ± 7.98	0.77^o
Musk1	81.65 ± 6.07	78.35 ± 4.47	0.13^o	83.64 ± 7.36	0.54^o	85.11 ± 5.67	0.11^o	82.79 ± 5.14	0.56^o	83.05 ± 4.25	0.55^o
WDBC	94.48 ± 4.01	95.38 ± 2.71	0.47^o	96.47 ± 2.38	0.18^o	96.47 ± 2.38	0.18^o	96.81 ± 2.36	0.15^o	96.47 ± 2.66	0.18^o
Segment	95.62 ± 2.52	96.62 ± 0.88	0.20^o	96.66 ± 0.70	0.27^o	96.66 ± 0.70	0.27^o	96.71 ± 0.71	0.20^o	96.71 ± 0.71	0.25^o
Steel	92.68 ± 1.10	91.96 ± 1.49	0.05^-	93.30 ± 1.58	0.13^o	93.30 ± 1.58	0.13^o	93.20 ± 1.84	0.25^o	93.25 ± 1.57	0.11^o
Page Block	95.98 ± 0.88	95.88 ± 0.87	0.50^o	96.07 ± 0.76	0.60^o	96.07 ± 0.76	0.60^o	95.99 ± 0.79	0.92^o	96.07 ± 0.76	0.60^o
Vehicle	66.83 ± 5.25	66.00 ± 4.81	0.50^o	68.88 ± 4.95	0.12^o	68.88 ± 4.95	0.12^o	68.74 ± 5.16	0.15^o	68.99 ± 5.07	0.09^o
Water	81.75 ± 5.77	83.68 ± 3.98	0.36^o	81.95 ± 4.46	0.92^o	82.34 ± 4.12	0.81^o	84.45 ± 2.79	0.14^o	85.21 ± 3.16	0.11^o
Waveform	79.88 ± 1.80	79.44 ± 2.13	0.69^o	80.92 ± 2.62	0.28^o	80.92 ± 2.62	0.28^o	80.92 ± 2.62	0.28^o	80.92 ± 2.62	0.28^o
Colon	76.25 ± 9.22	76.25 ± 16.43	1.00^o	80.83 ± 24.23	0.50^o	74.16 ± 20.95	0.78^o	80.00 ± 18.92	0.50^o	76.25 ± 16.43	1.00^o
Leukemia	91.42 ± 9.98	87.46 ± 10.54	0.30^o	86.03 ± 9.57	0.11^o	81.42 ± 16.56	0.07^o	86.66 ± 13.24	0.20^o	85.23 ± 10.39	0.12^o
Average	83.66	83.15		84.96		84.32		85.53		84.99	
Lose/Win/Tie		1/0/13		0/1/13		0/1/13		0/1/13		0/1/13	

Each column entry is $\mu \pm \sigma$ of classification accuracy.

TABLE VII
TENFOLD CROSS-VALIDATION EXPERIMENT RESULTS FOR COMPUTATIONAL TIME

Datasets	FMNN-FRS(0.3)	FMNN-FRS(0.2)		RMDPS		WRMDPS		FRS-Entropy	
	Mean ± Std	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val
Ionosphere	0.153 ± 0.012	0.188 ± 0.014	0.00^-	0.263 ± 0.009	0.00^-	0.276 ± 0.007	0.00^-	1.75 ± 0.057	0.00^-
Mov.&Libras	0.225 ± 0.019	0.324 ± 0.027	0.00^-	0.721 ± 0.023	0.00^-	0.756 ± 0.021	0.00^-	7.66 ± 0.155	0.00^-
Sonar	0.177 ± 0.025	0.183 ± 0.039	0.38^o	0.127 ± 0.019	0.00^+	0.134 ± 0.019	0.00^+	2.95 ± 0.308	0.00^-
LSVT	0.050 ± 0.016	0.116 ± 0.014	0.00^-	0.080 ± 0.007	0.00^-	0.083 ± 0.005	0.00^-	15.98 ± 0.67	0.00^-
Musk1	0.639 ± 0.071	0.962 ± 0.064	0.00^-	0.897 ± 0.010	0.00^-	0.943 ± 0.012	0.00^-	28.30 ± 0.302	0.00^-
WDBC	0.081 ± 0.005	0.173 ± 0.010	0.00^-	0.689 ± 0.012	0.00^-	0.730 ± 0.011	0.00^-	2.89 ± 0.068	0.00^-
Segment	0.049 ± 0.0176	0.155 ± 0.020	0.00^-	19.34 ± 0.462	0.00^-	20.56 ± 1.517	0.00^-	25.38 ± 2.00	0.00^-
Steel	0.767 ± 0.037	1.425 ± 0.065	0.00^-	2.68 ± 0.056	0.00^-	2.869 ± 0.049	0.00^-	9.09 ± 0.069	0.00^-
Page Block	0.095 ± 0.017	0.157 ± 0.020	0.00^-	24.07 ± 0.404	0.00^-	24.98 ± 0.39	0.00^-	34.26 ± 0.675	0.00^-
Vehicle	0.168 ± 0.020	0.307 ± 0.039	0.00^-	2.603 ± 0.315	0.00^-	2.754 ± 0.266	0.00^-	4.26 ± 0.397	0.00^-
Water	0.160 ± 0.008	0.380 ± 0.016	0.00^-	0.495 ± 0.017	0.00^-	0.523 ± 0.019	0.00^-	3.46 ± 0.091	0.00^-
Waveform	14.92 ± 1.41	34.22 ± 1.93	0.00^-	168.51 ± 96.97	0.00^-	79.49 ± 2.53	0.00^-	1144.8 ± 0.77	0.00^-
Colon	0.166 ± 0.032	0.231 ± 0.048	0.00^-	0.086 ± 0.007	0.00^+	0.090 ± 0.005	0.00^+	78.63 ± 3.56	0.00^-
Leukemia	0.301 ± 0.037	0.372 ± 0.056	0.00^-	0.400 ± 0.025	0.00^-	0.426 ± 0.019	0.00^-	867.44 ± 45.10	0.00^-
Lose/Win/Tie		13/0/1		12/2/0		12/2/0		14/0/0	
Average	1.28	2.80		15.78		9.61		159.00	

Each column entry is $\mu \pm \sigma$ of computational time.

TABLE VIII
TENFOLD CROSS-VALIDATION EXPERIMENT RESULTS FOR REDUCT LENGTH

Datasets	FMNN-FRS(0.3)	FMNN-FRS(0.2)		RMDPS		WRMDPS		FRS-Entropy	
	Mean ± Std	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	p-Val
Ionosphere	6.30 ± 0.67	5.70 ± 0.48	0.08^o	26.60 ± 1.26	0.00^-	27.30 ± 1.56	0.00^-	30.80 ± 0.42	0.00^-
Mov.&Libras	6.50 ± 0.70	5.90 ± 0.31	0.02^+	64.20 ± 2.04	0.00^-	67.20 ± 1.39	0.00^-	61.40 ± 0.84	0.00^-
Sonar	6.20 ± 0.42	5.30 ± 0.48	0.00^+	16.40 ± 1.50	0.00^-	17.00 ± 1.15	0.00^-	54.70 ± 1.24	0.00^-
LSVT	3.40 ± 0.51	4.50 ± 0.52	0.00^-	8.80 ± 0.42	0.00^-	9.90 ± 0.56	0.00^-	90.20 ± 2.69	0.00^-
Musk1	8.20 ± 0.42	7.50 ± 0.52	0.00^+	17.80 ± 0.91	0.00^-	18.20 ± 0.91	0.00^-	103 ± 2.53	0.00^-
WDBC	3.70 ± 0.67	4.60 ± 0.51	0.00^o	23.80 ± 0.63	0.00^-	23.90 ± 0.63	0.00^-	27.10 ± 0.56	0.00^-
Segment	8.10 ± 1.37	9.20 ± 1.13	0.03^-	15.00 ± 0.00	0.00^-	15.00 ± 0.00	0.00^-	12.90 ± 0.31	0.00^-
Steel	12.40 ± 1.26	10.60 ± 0.69	0.00^+	20.20 ± 0.78	0.00^-	20.20 ± 0.78	0.00^-	15.90 ± 0.31	0.00^-
Page Block	7.90 ± 0.87	7.40 ± 0.69	0.24^o	9.90 ± 0.31	0.00^-	9.90 ± 0.31	0.00^-	8.00 ± 0.47	0.75^o
Vehicle	13.10 ± 1.10	11.50 ± 1.18	0.01^+	17.90 ± 0.32	0.00^-	17.90 ± 0.32	0.00^-	13.40 ± 0.52	0.43^o
Water	8.20 ± 1.03	7.30 ± 0.48	0.02^+	18.30 ± 1.16	0.00^-	18.40 ± 1.57	0.00^-	33.80 ± 0.63	0.00^-
Waveform	13.40 ± 0.96	13.30 ± 0.67	0.79^o	21.00 ± 0.00	0.00^-	21.00 ± 0.00	0.00^-	21.00 ± 0.00	0.00^-
Colon	2.90 ± 0.31	2.80 ± 0.42	0.34^o	6.20 ± 0.63	0.00^-	6.80 ± 0.63	0.00^-	129.30 ± 7.40	0.00^-
Leukemia	2.00 ± 0.00	2.00 ± 0.00	1.00^o	4.10 ± 0.316	0.00^-	4.21 ± 0.421	0.00^-	434.30 ± 20.56	0.00^-
Lose/Win/Tie		2/6/6		14/0/0		14/0/0		12/0/2	
Average	7.30	6.90		19.30		19.73		73.98	

Each column entry is $\mu \pm \sigma$ of reduct length.

Furthermore, FMNN-FRS(0.3) has performed significantly better in the steel dataset over the FRS-Entropy algorithm. The average performance of FMNN-FRS(0.3) is 80.82%, which is very near to compared algorithms and Unred (81.08%) in CART. Similarly, based on the results in Table VI for kNN, the proposed FMNN-FRS(0.3) obtained statistically similar results with compared algorithms and Unred for all datasets except in movement & libras. Even the average performance of FMNN-FRS(0.3) is 83.66%, which is also very near to compared algorithms and Unred (84.99%). In conclusion, the FMNN-FRS(0.3) method obtained relevance approximate reduct achieving statistically similar classification accuracies in comparison to existing FRS approaches.

Based on results in Table VII, FMNN-FRS (0.3) incurred significantly less computational time in reduct computation than the compared algorithms for all datasets except in sonar and colon datasets. The compared algorithms RMDPS and WRMDPS performed significantly better than FMNN-FRS(0.3) in sonar and colon datasets. The compared approaches have lost in at least 12 datasets over FMNN-FRS(0.3). These significant results in computational time are due to the substantial reduction in time complexity of the FMNN-FRS over the compared algorithms (traditional FRS time complexity is $O(|U|^2|C|^2)$, whereas the FMNN-FRS achieves $O(|\text{HBS}|^2|C|^2)$. It is observed that $|\text{HBS}| << |U|$. For example, in the waveform and page-block dataset, FMNN-FRS(0.3) outperformed and saved more than 90–99% of the average computational time than other compared algorithms.

The results given in Table VIII establishes that the FMNN-FRS(0.3) computed reduct with statistically significant lesser size than RMDPS, WRMDPS for all datasets. Even FMNN-FRS(0.3) obtained statistically similar reducts with FRS-Entropy on page-block and vehicle datasets, and in remaining datasets, FMNN-FRS(0.3) performed statistically significant than FRS-Entropy. The mixed results are obtained while comparing with FMNN-FRS(0.2). The FMNN-FRS approach has resulted in the least average reduct length over compared approaches achieving 60–90% reduct length gain.

In summary, the proposed approach FMNN-FRS has resulted in a reduct relevant for classification model construction in significantly less computational time with a significantly less reduct length.

D. Enhanced Scalability of the FMNN-FRS Algorithm in Large Datasets

This section demonstrates the significance of the FMNN-FRS algorithm on large decision systems, where existing FRS-based approaches are failing to compute reduct. Table IX exhibits the six benchmark large quantitative datasets from UCI repository [37], in which comparative FRS methods (RMDPS, WRMDPS, and FRS-Entropy) could not compute reduct on the given system/hardware configuration.

The existing FRS-based reduct computation algorithms have the space complexity of $O(|U|^2|C|)$, which hinders the applicability to large decision systems. In the proposed algorithm, the space complexity is reduced to $O(|HBS|^2|C|)$ from $O(|U|^2|C|)$. In order to understand the significance of space complexity reduction, we have compared $|U|^2$ with average $|\text{HBS}|^2$ through tenfold cross validation. The results are given in Table X. In Table X, the average and standard deviation of the number of hyperboxes (NOH) obtained for $\theta = 0.2$ and 0.3 are provided and the value of $|\text{HBS}|^2$ are given using average values

TABLE IX
DETAILS OF LARGE DECISION SYSTEMS

Dataset	Attributes	Objects	Decision classes
Texture	40	5500	11
Shuttle	9	57999	7
Musk2	166	6598	2
WDG2	40	5000	3
Sensorless Drive Diagnosis	48	58509	11
Satimage	36	6435	6

TABLE X
REDUCTION OF LARGE DATASET SIZES WITH THE FMNN AS A PREPROCESSOR

Datsets	EDS $\lvert U\rvert^2$	NOH $\mu \pm \sigma$		PHDS $\lvert HBS\rvert^2$		POR	
		$\theta = 0.2$	$\theta = 0.3$	$\theta = 0.2$	$\theta = 0.3$	$\theta = 0.2$	$\theta = 0.3$
Texture	4950 × 4950	130 ± 6.78	47 ± 1.35	130 × 130	47 × 47	97%	99%
Shuttle	52199 × 52199	19 ± 2.84	15 ± 1.37	19 × 19	15 × 15	99%	99%
WDG2	4500 × 4500	964.40 ± 2.78	172.27 ± 3.24	964 × 964	172 × 172	78%	96%
Musk2	5938 × 5938	1039 ± 0.90	757 ± 2.09	1039 × 1039	757 × 757	82%	87%
Sensorless	52658 × 52658	240.58 ± 17.32	27 ± 0	240 × 240	27 × 27	99%	99%
Satimage	6435 × 6435	541.20 ± 7.45	220.12 ± 3.28	541 × 541	220 × 220	92%	98%

Abbreviation EDS: Estimated FDM sizes, NOH: Number of hyperboxes, PHDS: Proportional FDM sizes, and POR: Percentage of reduction.

TABLE XI
TENFOLD CROSS-VALIDATION EXPERIMENT RESULTS FOR COMPUTATIONAL TIME AND REDUCT LENGTH ON LARGE DATASETS

Datsets	Reduct Length		
	FMNN-FRS(0.3)	FMNN-FRS(0.2)	
	Mean ± Std	Mean ± Std	p-Val
Texture	10.80 ± 1.47	9.10 ± 0.99	0.00^+
Shuttle	5.70 ± 0.82	6.90 ± 1.10	0.01^-
WDG2	13.10 ± 0.56	12.20 ± 0.42	0.02^+
Musk2	14.00 ± 0.81	13.20 ± 0.42	0.01^+
Sensorless	8.70 ± 0.82	14.90 ± 1.60	0.00^-
Satimage	15.30 ± 1.49	16.80 ± 0.91	0.01^-
Lose/Win/Tie		3/3/0	
Average	11.27	12.18	
Datsets	**Computational Time (in secs)**		
	FMNN-FRS(0.3)	FMNN-FRS(0.2)	
	Mean ± Std	Mean ± Std	p-Val
Texture	0.127 ± 0.004	0.581 ± 0.029	0.00^-
Shuttle	0.623 ± 0.006	0.596 ± 0.009	0.00^+
WDG2	35.65 ± 0.708	66.75 ± 1.23	0.00^-
Musk2	41.02 ± 1.55	44.47 ± 0.93	0.00^-
Sensorless	0.46 ± 0.02	2.61 ± 0.15	0.00^-
Satimage	1.64 ± 0.05	6.88 ± 0.30	0.00^-
Lose/Win/Tie		5/1/0	
Average	13.40	20.35	

TABLE XII
TENFOLD CROSS-VALIDATION EXPERIMENT RESULTS FOR CLASSIFICATION ACCURACY RESULTS USING CART AND KNN CLASSIFIERS ON LARGE DATASETS

Datsets	CART classification results					kNN classification results				
	FMNN-FRS(0.3)	FMNN-FRS(0.2)		Unred		FMNN-FRS(0.3)	FMNN-FRS(0.2)		Unred	
	Mean ± Std	Mean ± Std	p-Val	Mean ± Std	p-Val	Mean ± Std	Mean ± Std	p-Val	Mean ± Std	p-Val
Texture	91.94 ± 1.09	91.74 ± 0.92	0.66^o	92.41 ± 0.95	0.32^o	97.85 ± 0.58	97.56±0.93	0.41^o	98.72 ±0.54	0.00^+
Shuttle	99.96 ± 0.30	99.75 ± 0.03	0.04^-	99.97 ± 0.02	0.91^o	99.91 ± 0.03	99.91 ± 0.04	1.00^o	99.90 ± 0.03	1.00^o
WDG2	74.20 ± 1.67	74.18 ± 1.48	0.97^o	74.58 ± 1.81	0.63^o	77.58 ± 1.12	76.76 ± 2.14	0.30^o	77.30 ± 1.794	0.67^o
Musk2	95.31 ± 0.66	95.07 ± 0.72	0.44^o	96.62 ± 0.63	0.00^+	95.34 ± 0.59	95.16 ± 0.67	0.53^o	96.68 ± 0.64	0.00^+
Sensorless	97.70 ± 1.54	98.16 ± 0.86	0.42^o	98.43 ± 0.21	0.15^o	97.00 ± 4.18	97.56 ± 2.51	0.72^o	99.03 ± 0.12	0.14^o
Satimage	85.05 ± 1.54	85.56 ± 1.46	0.45^o	85.85 ± 1.42	0.24^o	89.71 ± 1.34	90.16 ± 1.16	0.43^o	90.83 ± 1.09	0.06^o
Lose/Win/Tie		1/0/5		0/1/5			0/0/6		0/2/4	
Average	90.61	90.74		91.32		92.89	92.85		93.74	

obtained. The percentage of reduction (POR) of $|\text{HBS}|^2$ over $|U|^2$ is in the range of 78–99% across the given datasets. Owing to this significant reduction, the FMNN-FRS could be applied on those datasets where the compared algorithms (RMDPS, WR-MDPS, and FRS-Entropy) would not execute, as the required memory space for these datasets is not available in the given system considered.

As similar to Section IV-C, a comparative analysis is performed by employing tenfold cross-validation experiments. Table XII presents the classification model results using CART and kNN classifiers for aiding in the comparative study. Also, classification models are constructed on the unreduced dataset (mentioned under "Unred" in Table XII) for the comparison perspective. The results of tenfold experiments in aspects of reduct size and computational time are reported in Table XI.

In Table XI, it is observed that the FMNN-FRS(0.3) performed statistically more significant (lesser reduct size) than FMNN-FRS(0.2) in shuttle, sensorless, and satimage datasets. Moreover, the total average reduct size of FMNN-FRS(0.3) is 11.27, which is lesser than 12.18 for FMNN-FRS(0.2). In the aspect of computational time in Table XI, FMNN-FRS(0.3) always incurred statistically less time than FMNN-FRS(0.2) for all datasets except in the shuttle dataset. This is expected because of the lesser number of hyperboxes constructed with an increase in the $theta$ value of 0.3 than 0.2.

In Table XII, it can be seen from the results that the FMNN-FRS(0.3) method achieved statistically equivalent classification accuracies to FMNN-FRS(0.2) and Unred in sensorless, satimage, wdg2, and texture datasets in CART classifiers. Similarly, in the kNN classifier, FMNN-FRS(0.3) performed statistically equivalent classification accuracies to FMNN-FRS(0.2) and Unred in satimage, sensorless, shuttle, and wdg2 datasets.

A similar pattern is observed based on the results given in Section IV-C and in this section between FMNN-FRS(0.2) and FMNN-FRS(0.3) approaches. Both FMNN-FRS(0.2) and FMNN-FRS(0.3) achieved similar classification accuracies and mixed results for reduct lengths, while FMNN-FRS(0.3) obtained significantly better computational times than FMNN-FRS(0.2). Hence, based on these results, the value 0.3 is recommended for $theta$ in FMNN-FRS reduct computation.

The experimental results established that the applicability of the FRS reduct algorithm is enhanced strongly with FMNN preprocessing. But it also to be noted that the FMNN requires the entire data to be loaded in the memory for working. In the case of large data size, it is constrained by the complexity of $O(|U||C|)$ to load in memory. For such large datasets, which cannot be loaded into memory, we require distributed strategies, which will be investigated in the future work.

V. CONCLUSION

This proposed work is a novel hybridization of an FMNN with an FRS for scalable reduct computation of fuzzy rough reduct in hybrid decision systems. The traditional discernibility matrix constructed between a pair of objects is replaced by an FDM constructed between a pair of hyperboxes resulting from FMNN single epoch training. The hyperbox-based FDM is used for computation of fuzzy-rough approximate reduct for the hybrid decision system. The proposed FMNN-FRS reduct approach has resulted in significant computational gains over existing FRS reduct computation approaches while achieving similar or better classification accuracies. The objective of FMNN preprocessing on FRS reduct computation in the proposed FMNN-FRS is achieved in the enhanced scalability demonstrated by the FMNN-FRS. The FMNN-FRS approach could scale to reduct computation in such large decision systems, over which existing FRS approaches are unable to arrive at reduct owing to space constraints. Based on the experimental analysis, $\theta = 0.3$ is recommended for the FMNN-FRS algorithm. More attempts will be done in the future to improve further scalability of the FMNN-FRS approach using distributed and parallel approaches.

ACKNOWLEDGMENT

The authors would like to acknowledge Zhang *et al.* for providing the source code for FRS entropy [23].

REFERENCES

[1] Y. Yao, Y. Zhao, and J. Wang, "On reduct construction algorithms," in *Proc. 1st Int. Conf. Rough Sets Knowl. Technol.*, 2006, vol. 4062, pp. 297–304.
[2] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982.
[3] J. Zhang, T. Li, and H. Chen, "Composite rough sets for dynamic data mining," *Inf. Sci.*, vol. 257, pp. 81–100, 2014.
[4] W. P. Ziarko, "Rough sets and knowledge discovery: An overview," in *Rough Sets, Fuzzy Sets and Knowledge Discovery*, W. P. Ziarko, Ed. New York, NY, USA: Springer, 1994, pp. 11–15.

[5] Z. Pawlak, *Rough Sets—Theoretical Aspects of Reasoning About Data*, vol. 9. Norwell, MA, USA: Kluwer, 1991.
[6] H. S. Nguyen, "Discretization problem for rough sets methods," in *Proc. 1st Int. Conf. Rough Sets Current Trends Comput.*, 1998, pp. 545–552.
[7] D. Dubois and H. Prade, "Putting rough sets and fuzzy sets together," in *Intelligent Decision Support*, vol. 11, R. Slowinski, Ed. New York, NY, USA: Springer, 1992, pp. 203–232.
[8] D. Didier and P. Henri, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, nos. 2/3, pp. 191–209, 1990.
[9] R. Jensen and Q. Shen, "Fuzzy-rough attribute reduction with application to web categorization," *Fuzzy Sets Syst.*, vol. 141, no. 3, pp. 469–485, 2004.
[10] R. Jensen and Q. Shen, "Fuzzy-rough sets assisted attribute selection," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 73–89, Feb. 2007.
[11] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Aug. 2009.
[12] R. B. Bhatt and M. Gopal, "On the compact computational domain of fuzzy-rough sets," *Pattern Recognit. Lett.*, vol. 26, no. 11, pp. 1632–1640, 2005.
[13] C. Cornelis, R. Jensen, G. Hurtado, and D. Ślezak, "Attribute selection with fuzzy decision reducts," *Inf. Sci.*, vol. 180, no. 2, pp. 209–224, 2010.
[14] P. S. V. S. S. Prasad and C. R. Rao, "An efficient approach for fuzzy decision reduct computation," *Trans. Rough Sets*, vol. 17, pp. 82–108, 2014.
[15] Y. Qian, Q. Wang, H. Cheng, J. Liang, and C. Dang, "Fuzzy-rough feature selection accelerator," *Fuzzy Sets Syst.*, vol. 258, pp. 61–78, 2015.
[16] A. Skowron and C. Rauszer, "The discernibility matrices and functions in information systems," in *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, vol. 11, R. Słowiński, Ed. Dordrecht, The Netherlands: Springer, 1992, pp. 331–362.
[17] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 385–389, Apr. 2012.
[18] J. Dai, H. Hu, W. Wu, Y. Qian, and D. Huang, "Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 4, pp. 2174–2187, Aug. 2018.
[19] D. Chen, X. Wang, and S. Zhao, "Attribute reduction based on fuzzy rough sets," in *Rough Sets and Intelligent Systems Paradigms*, vol. 4585. New York, NY, USA: Springer, 2007, pp. 381–390.
[20] R. Jensen and N. M. Parthalin, "Towards scalable fuzzy-rough feature selection," *Inf. Sci.*, vol. 323, pp. 1–15, 2015.
[21] X. Zhang, C. Mei, D. Chen, and Y. Yang, "A fuzzy rough set-based feature selection method using representative instances," *Knowl.-Based Syst.*, vol. 151, pp. 216–229, 2018.
[22] D. Slezak, "Approximate reducts in decision tables," in *Proc. 6th Int. Conf. Inf. Process. Manage. Uncertainty Knowl.-Based Syst.*, 1996, pp. 1159–1164.
[23] X. Zhang, X. Liu, and Y. Yang, "A fast feature selection algorithm by accelerating computation of fuzzy rough set-based information entropy," *Entropy*, vol. 20, no. 10, 2018, Art. no. 788.
[24] X. Zhang, C. Mei, D. Chen, and J. Li, "Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy," *Pattern Recognit.*, vol. 56, pp. 1–15, 2016.
[25] A. Ohrn, "Discernibility and rough sets in medicine: Tools and applications," doctoral dissertation, Dept. Comput. Inf. Sci., Norwegian Univ. Sci. Technol., Trondheim, Norway, 2000, pp. 1–239.
[26] T. Y. Lin, "Granular computing: Fuzzy logic and rough sets," in *Computing with Words in Information/Intelligent Systems 1: Foundations*, vol. 33, L. A. Zadeh and J. Kacprzyk, Eds. Heidelberg, Germany: Physica-Verlag, 1999, pp. 183–200.
[27] P. K. Simpson, "Fuzzy min-max neural networks. I. Classification," *IEEE Trans. Neural Netw.*, vol. 3, no. 5, pp. 776–786, Sep. 1992.
[28] A. Bargiela and W. Pedrycz, "Supervised and unsupervised information granulation: A study in hyperbox design," in *Novel Developments in Granular Computing: Applications for Advanced Human Reasoning and Soft Computation*. Hershey, PA, USA: IGI Global, 2010, pp. 48–68.
[29] R. Davtalab, M. H. Dezfoulian, and M. Mansoorizadeh, "Multi-level fuzzy min-max neural network classifier," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 470–482, Mar. 2014.
[30] 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.
[31] O. F. Reyes-Galaviz and W. Pedrycz, "Granular fuzzy modeling with evolving hyperboxes in multi-dimensional space of numerical data," *Neurocomputing*, vol. 168, pp. 240–253, 2015.
[32] A. Bargiela and W. Pedrycz, "Optimised information abstraction in granular min/max clustering," in *Emerging Paradigms in Machine Learning*, vol. 13, S. Ramanna, L. C. Jain, and R. J. Howlett, Eds. Berlin, Germany: Springer, 2013, pp. 31–48.
[33] F. Pourpanah, C. P. Lim, X. Wang, C. J. Tan, M. Seera, and Y. Shi, "A hybrid model of fuzzy min-max and brain storm optimization for feature selection and data classification," *Neurocomputing*, vol. 333, pp. 440–451, 2019.
[34] P. M. Sonule and B. S. Shetty, "An enhanced fuzzy min-max neural network with ant colony optimization based-rule-extractor for decision making," *Neurocomputing*, vol. 239, pp. 204–213, 2017.
[35] P. Jaccard, "Nouvelles recherches sur la distribution florale," *Bull. Soc. Vaudoise des Sci. Naturelles*, vol. 44, pp. 223–270, 1908.
[36] A. M. Radzikowska and E. E. Kerre, "A comparative study of fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 126, no. 2, pp. 137–155, 2002.
[37] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml
[38] S. Kabir, C. Wagner, T. C. Havens, D. T. Anderson, and U. Aickelin, "Novel similarity measure for interval-valued data based on overlapping ratio," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, 2017, pp. 1–6.
[39] "Bioinformatics research group." [Online]. Available: https://eps.upo.es/bigs/datasets.html, Accessed on: Oct. 4, 2019.
[40] S. Ghosh, P. S. V. S. Sai Prasad, and C. R. Rao, "Third order backward elimination approach for fuzzy-rough set based feature selection," in *Proc. Int. Conf. Pattern Recognit. Mach. Intell.*, 2017, pp. 254–262.

Anil Kumar received the M.Tech. degree in artificial intelligence from the University of Hyderabad, Hyderabad, India, in 2015, where he is currently working toward the Ph.D. degree in computer science.

His current research interests include rough sets, fuzzy rough sets, fuzzy min–max neural networks, neural networks, and machine learning.

P. S. V. S. Sai Prasad (Member, IEEE) received the M.Sc. degree in mathematics and the M.Tech. degree in computer science from Sri Sathya Sai University, Anantapur, India, in 1997 and 2001, and the Ph.D. degree in computer science from the University of Hyderabad, Hyderabad, India, in 2014.

He is currently an Assistant Professor with the School of Computer and Information Sciences, University of Hyderabad. He has authored or coauthored more than 25 publications in national/international journals and conferences. His research interests include rough sets, fuzzy rough sets, soft computing, and big data analytics.