This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3381993

Fuzzy Rough Attribute Reduction Based on Fuzzy Implication Granularity Information

Jianhua Dai, Zhilin Zhu, Xiongtao Zou

***Abstract*— Fuzzy rough set model is a powerful tool for handling attribute reduction tasks for complex data. While the fuzzy rough set model commonly employs fuzzy information entropy to measure attribute uncertainty, utilizing fuzzy conditional information entropy for measuring attribute relationships presents a drawback due to its lack of monotonicity, impacting attribute reduction results. Furthermore, entropy computations involve numerous logarithmic function computations, resulting in a significant computational burden. Moreover, the results obtained from logarithmic functions are unbounded. To address these problems, this paper presents the concept of Fuzzy Implication Granularity Information (FIGI) for measuring attribute information. Additionally, we introduce several related generalizations, such as fuzzy conditional implication granularity information, fuzzy mutual implication granularity information, and fuzzy joint implication granularity information, aiming to measure the relationships between attributes. Notably, the introduced fuzzy conditional implication granularity information to measure the relationship between attributes demonstrates the desirable property of monotonicity. Crucially, all the metrics proposed in this paper are bounded, ensuring that computed values within the range of 0 to 1. Finally, we propose a forward greedy attribute reduction algorithm based on the monotonic fuzzy conditional implication granularity information (MFIGI), and the performance of our MFIGI algorithm was compared against six different attribute reduction algorithms using three classifiers across 15 different datasets, the experimental results demonstrate the excellence of our MFIGI algorithm.**

***Index Terms*—Fuzzy rough set; attribute reduction; fuzzy implication granularity information; granular computing.**

I. INTRODUCTION

WITH the development of the times, data dimensions continue to grow, and dataset sizes also experience exponential growth. However, data often contains redundant information or noise. Therefore, in many fields such as pattern recognition, machine learning, preprocessing of data is necessary. Attribute reduction, as a preprocessing step, can eliminate redundant information from the dataset while retaining important information, attracting considerable attention [1], [2], [3], [4], [5].

The rough set model proposed by Pawlak [6] offers a mathematical tool for handling uncertainty issues. This theory has been successfully applied in various fields such as machine learning, decision analysis, process control, pattern recognition, and data mining [7], [8], [9], [10], [11]. Pawlak's rough set model employs equivalence relations to discern between samples, rendering it well-suited for processing symbolic data [4], [12], [13]. It requires a discretization step when using Pawlaks rough set model to handle real-valued data. However, this discretization process might result in the loss of useful information, potentially impacting the accuracy of the analysis [14], [15], [16]. Therefore, Dubois et al. [1] proposed the fuzzy rough set model for real-valued data processing tasks. The fuzzy rough set model has been widely applied in attribute reduction [10], [17], [18], [19] and rule reasoning [20], [21], [22]. In the fuzzy rough set model, attribute reduction tasks are performed by constructing fuzzy similarity relationships between samples, avoiding the loss of information caused by discretization. Hu et al. [23] also proposed the neighborhood rough set model for heterogeneous attribute reduction tasks. Meanwhile, many researchers have conducted additional studies based on the neighborhood rough set model [24], [25], [26], [27]. The neighborhood rough set model constructs neighborhood relationships between samples by setting a neighborhood radius parameter [23]. When the distance between samples is less than the neighborhood radius parameter, samples within the neighborhood are considered equivalent. However, because the setting of the neighborhood radius is done before the start of attribute reduction tasks, the rationality of setting the neighborhood radius parameter needs further consideration[23], [24], [28]. Therefore, the viewpoint of considering samples within the neighborhood as equivalent is unreasonable.

Meanwhile, when applying the fuzzy rough set model to tasks, the measurement of information content within attributes and the quantification of relationships between attributes have become crucial issues, drawing the attention of numerous researchers [3], [5], [13], [19], [29]. This research primarily focuses on three aspects: firstly, extending Pawlak's rough set theory in the fuzzy rough set model as done by Jensen et al. [5], [29]. However, classical fuzzy dependency measures may not effectively reflect the classification ability. Therefore, Wang et al. [18] proposed a method of calculating fuzzy dependency by fuzzying the decision enabling a better fit to the data. Because Jensen et al. [5], [29] primarily focused on the lower approximation space and overlooked considerations for the upper approximation space, Wang et al. [30] developed the concept of fuzzy self-information based on their fitting fuzzy rough set model, which considers both the upper and lower fuzzy approximation spaces for attribute reduction.

This work was partially supported by the National Natural Science Foundation of China (62376093, 61976089), the Major Program of the National Social Science Foundation of China (20&ZD047), the Natural Science Foundation of Hunan Province (2021JJ30451, 2022JJ30397), the Hunan Provincial Science & Technology Project Foundation (2018TP1018, 2018RS3065). (Corresponding author: Jianhua Dai.)

Jianhua Dai, Zhilin Zhu and Xiongtao Zou are with the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, the College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China, (e-mail: jhdai@hunnu.edu.cn).

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3381993

The second direction involves extending information entropy, proposing concepts like fuzzy joint information entropy and fuzzy conditional information entropy, such as the research by Hu et al. [31], [32]. The third direction involves measuring the relationships between attributes based on granularity [33], [34], [35], [36]. For instance, Qian et al. [37] introduced the concept of fuzzy information granularity within the fuzzy binary granules. This concept aims to measure the information contained in attributes and to measure the relationships or influences between attributes. Liang et al. [36], based on Pawlak's rough set theory, proposed the concept of knowledge granularity, measuring the amount of knowledge [34].

The concept of information entropy is widely employed across various attribute reduction tasks [10], [38], [39], [28], [40]. To measure the discriminative ability or information content contained in attribute subsets, Yager [41] introduced the concept of fuzzy information entropy into the fuzzy rough set model, mainly as an extension of Shannon's entropy [42]. Hernandez and Recasens [43] further extended Yager's work by formulating the concepts of fuzzy joint entropy and fuzzy conditional entropy. Hu et al. [31] redefined fuzzy joint information entropy, fuzzy conditional information entropy, etc., to measure the relationships between attributes. Dai et al. [44] furthered the concept of fuzzy mutual information and introduced the information gain ratio for attribute reduction tasks. The most widely used are the fuzzy information entropy and fuzzy conditional information entropy proposed by Hu et al. [31], [32]. However, the fuzzy information conditional entropy faces a significant problem: its variation during attribute reduction is not monotonic. This leads to a challenge in determining when to halt the attribute reduction process, ultimately impacting the rationality of the attribute reduction subset. Moreover, the computation process of fuzzy information entropy involves numerous logarithmic function computations, which are computationally burdensome and the logarithmic function computation result is unbounded. Dai et al. [45], [38], Qian et al. [37], [33] and Liang et al. [36], [34], then proposed a series of methods to measure relationships between attributes from a granular perspective, avoiding some of the drawbacks associated with entropy. However, the measurement methods lack correlation and relatively poor generalization. When measuring the relationship between conditional attributes and the decision, they often lack rationality.

To address these problems, we propose in this paper the concept of Fuzzy Implication Granularity Information (FIGI) through the analysis of fuzzy granule structures. Simultaneously, we construct a series of extensions based on fuzzy implication granularity information, such as fuzzy conditional implication granularity information and fuzzy joint implication granularity information, to measure the relationships between attributes. It is worth noting that our fuzzy conditional implication granularity information, used to measure the relationship between decision attribute and conditional attributes, exhibits monotonicity. Therefore, building attribute reduction algorithm based on fuzzy conditional implication granularity information is quite reasonable. Furthermore, the various forms proposed in this paper for measuring relationships between attributes are bounded, with computed values falling within the range of 0 to 1. The contributions of this paper can be summarized as follows:

(1) We introduce a bounded metric called Fuzzy Implication Granularity Information (FIGI) to measure the information content of attributes, avoiding the unbounded problem of logarithmic functions in entropy operations. Additionally, present a series of generalized forms based on our fuzzy implication granularity information to measure relationships between attributes. Furthermore, these forms possess bounded property, ensuring that their computed values within the range of 0 to 1.

(2) We analyze that the fuzzy conditional entropy lacks monotonicity when measuring the relationship between decision attribute and conditional attributes. As a solution, we introduce a metric called fuzzy conditional implication granularity information, which exhibits the property of monotonicity.

(3) We construct a forward greedy attribute reduction algorithm based on the monotonic fuzzy conditional implication granularity information (MFIGI) and conduct attribute reduction experiments on 15 datasets, comparing it with six other algorithms. We compare the classification accuracy of attribute reduction results across three classifiers.

The rest of this paper is outlined as follows. In the section II, we provide a concise overview of fundamental concepts within fuzzy rough sets, including knowledge granules, fuzzy granularity, and related forms. In the section III, we construct new forms of fuzzy implication granularity information to measure relationships between attributes or the information content within attributes. Additionally, based on our measurements, a greedy forward attribute reduction algorithm (MFIGI) is proposed. Finally, in the section IV, we evaluate the superiority of our algorithm through extensive experimentation and statistical analysis.

II. Preliminary Knowledge

In this section, we primarily introduce fundamental concepts in fuzzy rough sets and existing measures of information content for attributes in the fuzzy rough set model.

A. Fuzzy Rough Sets

In contrast to the equivalence relationship in Pawlak's rough sets model [6], [13], fuzzy similarity relationship is utilized in fuzzy rough sets [1], [5], [46] to describe the relationship between samples. For a given nonempty finite universe U, $\Re$ is a fuzzy similarity relation, $\Re$ satisfies the following properties:
(1) Reflexivity: $\Re\left(x_i, x_i\right)=1$;
(2) Symmetry: $\Re\left(x_i, x_j\right)=\Re\left(x_j, x_i\right)$;
(3) Transitivity: $\Re\left(x_i, x_j\right) \geqslant T\left\{\Re\left(x_i, x_k\right), \Re\left(x_k, x_j\right)\right\}$.

Definition 1. *[46] The T is a T-norms, for $\forall \mu, \nu, z, \omega \in P$, $T: P \times P$ satisfies the following properties:*
(1) Boundary condition: $T(1, \mu)=\mu$;
(2) Commutativity: $T(\mu, \nu)=T(\nu, \mu)$;
(3) Associativity: $T(\mu, T(\nu, \omega))=T(T(\nu, \omega), \omega)$;
(4) Monotonicity: $\omega \geqslant \mu, z \geqslant \nu \Rightarrow T(\mu, v) \leqslant T(\omega, z)$.

The commonly used T-norms are shown below:

(1) $T_M(\nu,\omega)=\min\{\nu,\omega\}$;

(2) $T_{\cos}(\nu,\omega)=\max\left\{\nu\omega-\sqrt{1-\nu^2}\sqrt{1-\omega^2},0\right\}$.

Proposition 1. *[47] For* $\forall\nu,\omega\in[0,1], T_M(\nu,\omega)\geqslant T_{\cos}(\nu,\omega)$.

For a given universe U, $\Re$ is a fuzzy similarity relation on U, which is commonly represented by the following matrix:

$$M(\Re)=\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}, \tag{1}$$

where r_{ij} can be expressed as the degree of fuzzy similarity between sample x_i and sample x_j, if r_{ij} is equal to 1 it means that x_i and x_j are completely equivalent and if it is 0 it means that x_i and x_j are completely unequal, the similarity relation between samples using the measure: $\Re(x_i,x_j)=\max\left(\min\left(\frac{f(x_j)-f(x_i)+\sigma}{\sigma},\frac{f(x_i)-f(x_j)+\sigma}{\sigma}\right),0\right)$ where σ denotes the standard deciation.

Some operations of relation $\Re$ are defined as:

(1) $\Re_1=\Re_2\Leftrightarrow\Re_1(x_i,x_j)=\Re_2(x_i,x_j)$;

(2) $\Re=\Re_1\cup\Re_2=\max\{\Re_1(x_i,x_j),\Re_2(x_i,x_j)\}$;

(3) $\Re=\Re_1\cap\Re_2=\min\{\Re_1(x_i,x_j),\Re_2(x_i,x_j)\}$;

(4) $\Re_1\subseteq\Re_2\Leftrightarrow\Re_1(x_i,x_j)\leqslant\Re_2(x_i,x_j)$.

Definition 2. *[47]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, $B\subseteq C$ is a attribute subset. The nonempty set U can be divided into as follows:

$$U/\Re_B=\{[x_i]_B\}_{i=1}^n, \tag{2}$$

where $[x_i]_B=\frac{\Re_B(x_i,x_1)}{x_1}+\frac{\Re_B(x_i,x_2)}{x_2}+\cdots+\frac{\Re_B(x_i,x_n)}{x_n}$ is a fuzzy similarity class generated by x_i.

B. Fuzzy Information Measure

In this section, we primarily introduce fuzzy information granularity proposed by Liang et al. [36] and Qian et al. [37], fuzzy information entropy and its generalization defined by Hu et al. [31] based on fuzzy rough sets, as well as relevant measures of fuzzy granularity provided by Dai et al. [45].

Definition 3. *[31], [37]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, for an attribute subset $B\subseteq C$, the fuzzy similarity relation on B is expressed as $\Re_B$, then the fuzzy-information granularity of $\Re_B$ is defined as:

$$GK(\Re_B)=\frac{1}{n}\sum_{i=1}^n\frac{|[x_i]_B|}{n}, \tag{3}$$

where $|[x_i]_B|$ is the cardinality of the fuzzy similarity class $[x_i]_B$.

If the fuzzy similarity relationship degrades to the equivalence relation in Pawlak's rough set model, the fuzzy information granulation will degenerate into information granulation [36].

Furthermore, Hu et al. [31] extended the information entropy to measure the information quantity contained in attributes in a fuzzy rough set model. They provided a measure of information quantity for attribute subsets in the fuzzy rough set model.

Definition 4. *[31]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, for an attribute subset $B\subseteq C$, information quantity of attribute subset B is defined as:

$$H(B)=-\frac{1}{n}\sum_{i=1}^n\log\frac{|[x_i]_B|}{n}. \tag{4}$$

Definition 5. *[31]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, for the attribute subsets $B,E\subseteq C$, the conditional entropy E to B is defined as:

$$H(E|B)=-\frac{1}{n}\sum_{i=1}^n\frac{|[x_i]_B\cap[x_i]_E|}{|[x_i]_B|}, \tag{5}$$

where $[x_i]_B,[x_i]_E$ are the fuzzy similarity classes generated by B and E respectively. $|[x_i]_B|$,$|[x_i]_B\cap[x_i]_E|$are the cardinality of fuzzy similarity classes $[x_i]_B$ and $[x_i]_B\cap[x_i]_E$.

Definition 6. *[45]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, for an attribute subset $B\subseteq C$, letting $U/\Re_B=\{[x_1]_B,[x_2]_B,\cdots,[x_n]_B\}$, the fuzzy granularity of $\Re_B$ is defined as:

$$Disc(\Re_B)=\frac{1}{n}\sum_{i=1}^n\left(1-\frac{|[x_i]_B|}{n}\right), \tag{6}$$

where $|[x_i]_B|$ is the cardinality of fuzzy similarity class $[x_i]_B$. The relation can degrade as crisp relation if $\forall x_i,x_j\in U,\Re_B(x_i,x_j)=\{0,1\}$.

Definition 7. *[45]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, for the attribute subsets $B,E\subseteq C$, $\Re_B$ and $\Re_E$ are fuzzy similarity relation on B and E respectively. The joint fuzzy granularity of B and E is defined as:

$$Disc(\Re_B\Re_E)=\frac{1}{n}\sum_{i=1}^n\left(1-\frac{|[x_i]_B\cap[x_i]_E|}{n}\right) \tag{7}$$

Proposition 2. *[45]* For the given $\Re_B$ and $\Re_E$ are fuzzy similarity relation, $Disc(\Re_B)+Disc(\Re_E)\geqslant Disc(\Re_B\Re_E)$.

Definition 8. *[45]* Given an $FD\mathrm{T}=\langle U,C\cup D,V\rangle$, for the attribute subsets $B,E\subseteq C$, $\Re_B$ and $\Re_E$ are fuzzy similarity relation on B and E respectively. The mutual fuzzy granularity of B and E is defined as:

$$Disc(\Re_B;\Re_E)=\frac{1}{n}\sum_{i=1}^n\left(1-\frac{\left|[x_i]_B^{\mathbb{C}}\cap[x_i]_E^{\mathbb{C}}\right|}{n}\right), \tag{8}$$

where $[x_i]_B^{\mathbb{C}}$ means the complement of $[x_i]_B$, $[x_i]_B^{\mathbb{C}}=\frac{1-r_{i1}}{x_1}+\frac{1-r_{i2}}{x_2}+\cdots+\frac{1-r_{in}}{x_n}$.

III. Fuzzy Implication Granularity Information Metrics

In this section, we introduce the concept of fuzzy implication granularity information to measure attribute significance.

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3381993

Furthermore, we present several forms based on fuzzy implication granularity information to evaluate relationships between attributes. Additionally, we construct a greedy forward attribute reduction algorithm based on our fuzzy implication granularity information measure.

A. Fuzzy Implication Granularity Information

This section mainly discusses the issues with the fuzzy conditional information entropy proposed by Hu et al [31]. It lacks monotonicity when measuring the relationship between the decision attribute and conditional attributes. Additionally, the computation of entropy involves a significant computational burden due to the inclusion of numerous logarithmic functions, and the logarithmic function computation is unbounded. These problems will influence the selection of attribute subsets.

To address these two main problems, we introduce the concept of Fuzzy Implication Granularity Information (FIGI) and its extensions, such as fuzzy conditional implication granularity information, fuzzy mutual implication granularity information, and fuzzy joint implication granularity information. We also highlight some of their properties. It's worth noting that the fuzzy conditional implication granularity proposed in this section possesses monotonicity, ensuring the rationality of stopping criteria when selecting attributes. Moreover, all the metrics we introduce do not require logarithmic function computations, making their computations relatively simple. Furthermore, all the metrics we propose are bounded, maintaining values between 0 and 1.

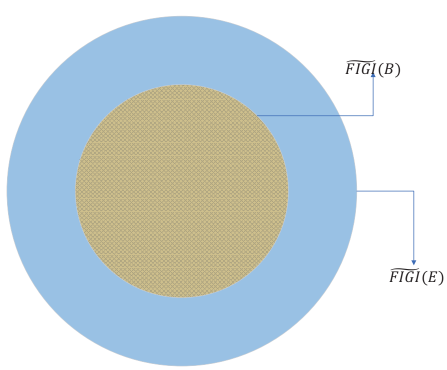


Fig. 1: The relationship between attribute subsets B and E.

Definition 9. Given a fuzzy decision table $FDT = \langle U, C \cup D, V \rangle$, for an attribute subset $B \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The fuzzy implication granularity information of B is defined as follows:

$$\widetilde{FIGI}(B) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\left|[x_i]_B\right|}{n}\right), \tag{9}$$

where $[x_i]_B$ represents the fuzzy similarity class generated by x_i under the attribute subset B, $|\bullet|$ represents the cardinality of the fuzzy similarity class.

Because the cardinality of the fuzzy similar class $|x_i| \leqslant n$, it is evident that the computation results of fuzzy implication granularity information fall within the range of 0 to 1.

The fuzzy implication granularity information can be used to measure the implicit information within attributes. A higher value indicates a larger amount of information contained within the attributes, whereas a lower value signifies less information.

Proposition 3. Given a fuzzy decision table $FDT = \langle U, C \cup D, V \rangle$, for an attribute subset $B \subseteq C$, the fuzzy implication granularity information of B $0 \leqslant \widetilde{FIGI}(B) < 1$.

In analyzing granularity structures, the lower bound of the interaction information quantity between attributes is primarily determined by attributes with coarser granularity structures. Therefore, we provide a definition for mutual implication granularity information.

Definition 10. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The fuzzy mutual implication granularity information of B and E is defined as follows:

$$\widetilde{MFIGI}(B; E) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\left|[x_i]_B \cup [x_i]_E\right|}{n}\right), \tag{10}$$

where $[x_i]_B \cup [x_i]_E$ = max $\{\Re_B(x_i, x_j), \Re_E(x_i, x_j)\}$, $|\bullet|$ represents the cardinality of the fuzzy similarity class.

Proposition 4. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, if $B \subseteq E$ then $\widetilde{MFIGI}(B; E) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\left|[x_i]_B \cup [x_i]_E\right|}{n}\right) = \widetilde{FIGI}(B)$.

Proof: Assume that $\Re_B$, $\Re_E$ are fuzzy relations induced by B, E, because of $B \subseteq E, \forall x_i, x_j \in U$ we can obtain $\Re_B(x_i, x_j) \geqslant \Re_E(x_i, x_j)$ means fuzzy similarity class $[x_i]_B \cup [x_i]_E = [x_i]_B$ then $\left|[x_i]_B \cup [x_i]_E\right| = \left|[x_i]_B\right|$, therefore $\widetilde{MFIGI}(B; E) = \widetilde{FIGI}(B) = \frac{1}{n}\sum_{i=1}^{n}\left(1 - \frac{\left|[x_i]_B\right|}{n}\right)$. ■

Proposition 4 means that if for the attribute subsets $B, E \subseteq C$, $B \subseteq E$ the fuzzy mutual implication granularity information value is equivalent to the fuzzy implication granularity information value of B. The relationship between B and E is illustrated in Fig 1.

Proposition 5. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, the fuzzy mutual implication granularity information between B and E, $0 \leqslant \widetilde{MFIGI}(B; E) < 1$.

Proof: Assume that $\Re_B$, $\Re_E$ are fuzzy relations induced by B, E, we can obtain fuzzy similarity class $\left|[x_i]_B \cup [x_i]_E\right| \leqslant |U| \leqslant n$, then $0 < \frac{\left|[x_i]_B \cup [x_i]_E\right|}{n} \leqslant 1$ therefore $0 \leqslant \widetilde{MFIGI}(B; E) < 1$. ■

Proposition 5 implies that the fuzzy mutual implication granularity information, as defined by us to measure the relationship of mutual influence between two attribute subsets, is bounded, with values falling within the range of 0 to 1.

Definition 11. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The B relative to E fuzzy conditional implication granularity information is defined as follows:

$$\widetilde{FIGI}\left(B|E\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|[x_i]_B \cup [x_i]_E\right|}{n} - \frac{\left|[x_i]_B\right|}{n}\right). \quad (11)$$

Fuzzy conditional implication granularity information of B based on E, $\widetilde{FIGI}\left(B|E\right)$ can be regard as the complementary fuzzy implication granularity information for E when the subset of B is added to the subset of E, given that E is known.

Therefore, when $\widetilde{FIGI}\left(B|E\right)$ is smaller, it can be understood as E imply more information from B, or interpreted as the complement of information from B to E being smaller.

It is worth noting that our fuzzy conditional implication granularity, when used to measure relationships between attributes, possesses bounded property, with a value range between 0 and 1.

There is a relationship between the fuzzy conditional implication granularity information and the fuzzy joint implication granularity information as follows:
(1) $\widetilde{FIGI}\left(E\right) + \widetilde{FIGI}\left(B|E\right) = \widetilde{FIGI}\left(B, E\right)$.
(2) $\widetilde{FIGI}\left(B\right) + \widetilde{FIGI}\left(E|B\right) = \widetilde{FIGI}\left(B, E\right)$.

Proposition 6. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, if $B \subseteq E$ then $\widetilde{FIGI}\left(B|E\right) = 0$.

Proof: Assume that $\Re_B$, $\Re_E$ are fuzzy relations induced by B, E, because of $B \subseteq E, \forall x_i, x_j \in U$ we can obtain $\Re_B\left(x_i, x_j\right) \geqslant \Re_E\left(x_i, x_j\right)$ means fuzzy similarity class $[x_i]_B \cup [x_i]_E = [x_i]_B$ then $\left|[x_i]_B \cup [x_i]_E\right| = \left|[x_i]_B\right|$, therefore $\frac{\left|[x_i]_B \cup [x_i]_E\right|}{n} - \frac{\left|[x_i]_B\right|}{n} = 0$, so $\widetilde{FIGI}\left(B|E\right) = 0$. ■

Proposition 6 means that E contains all the information possessed by B, and B is completely implied by E, as shown in Fig 1.

Proposition 7. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, if $E \subseteq F$ then $\widetilde{FIGI}\left(B|F\right) \leqslant \widetilde{FIGI}\left(B|E\right)$.

Proof: Assume that $\Re_B$, $\Re_E$, $\Re_F$ are fuzzy relations induced by B, E, F, because of $E \subseteq F$, $\forall x_i, x_j \in U$ we can obtain $\Re_E\left(x_i, x_j\right) \geqslant \Re_F\left(x_i, x_j\right)$ means fuzzy similarity class $[x_i]_E \cup [x_i]_F = [x_i]_E$, then $\left|[x_i]_B \cup [x_i]_E\right| \geqslant \left|[x_i]_B \cup [x_i]_F\right|$. So, we have proven that $\widetilde{FIGI}\left(B|E\right) \geqslant \widetilde{FIGI}\left(B|F\right)$. ■

Proposition 7 implies that the fuzzy conditional implication granularity information is monotonic and decreases monotonically.

We also provide corresponding examples, illustrating situations where our fuzzy conditional implication granularity information exhibits monotonicity while fuzzy conditional information entropy does not, as shown in Example III.1.

Example III.1. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$. The fuzzy relationship matrix generated by B, E, and the decision attribute D is as follows:

$$\Re_B = \begin{bmatrix} 1 & 1 & 0.94 & 1 & 0.13 & 0.13 \\ 1 & 1 & 0.95 & 1 & 0.14 & 0.14 \\ 0.94 & 0.95 & 1 & 0.94 & 0.25 & 0.25 \\ 1 & 1 & 0.94 & 1 & 0.14 & 0.13 \\ 0.13 & 0.14 & 0.25 & 0.14 & 1 & 1 \\ 0.13 & 0.14 & 0.25 & 0.13 & 1 & 1 \end{bmatrix},$$

$$\Re_E = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0.05 \\ 1 & 1 & 1 & 1 & 1 & 0.05 \\ 1 & 1 & 1 & 1 & 1 & 0.05 \\ 1 & 1 & 1 & 1 & 1 & 0.05 \\ 1 & 1 & 1 & 1 & 1 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 1 \end{bmatrix},$$

$$\Re_D = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix},$$

The computation results of fuzzy conditional information entropy and fuzzy conditional implication granularity information are as follows: According to Equation 5, $H\left(D|B\right) = 0.6988$, $H\left(D|B, E\right) = 0.7142$, Obviously fuzzy conditional information entropy is not monotonic. Conversely, according to Equation 11, $\widetilde{FIGI}\left(D|B\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|[x_i]_D \cup [x_i]_B\right|}{n} - \frac{\left|[x_i]_D\right|}{n}\right) = 0.2450$, $\widetilde{FIGI}\left(D|B, E\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\left|[x_i]_D \cup [x_i]_{B \cup E}\right|}{n} - \frac{\left|[x_i]_D\right|}{n}\right) = 0.2361$.

Due to the lack of monotonicity in fuzzy conditional information entropy, it cannot guarantee the rationality of attribute subset reduction in subsequent processes, potentially leading to suboptimal reduction subset. In contrast, the fuzzy conditional implication granularity information proposed in this paper does not encounter such problems.

Definition 12. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The definition of the fuzzy joint implication granularity information of B and E is as follows:
$\widetilde{FIGI}\left(B, E\right) =$

$$\frac{1}{n}\sum_{i=1}^{n}\left(1 + \frac{\left|[x_i]_B \cup [x_i]_E\right|}{n} - \frac{\left|[x_i]_B\right| + \left|[x_i]_E\right|}{n}\right). \quad (12)$$

Proposition 8. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, if $B \subseteq E$ then $\widetilde{FIGI}\left(B, E\right) = \widetilde{FIGI}\left(E\right)$.

Proof: Assume that $\Re_B$, $\Re_E$ are fuzzy relations induced by B, E, because of $B \subseteq E, \forall x_i, x_j \in U$ we can obtain $\Re_B\left(x_i, x_j\right) \geqslant \Re_E\left(x_i, x_j\right)$ means fuzzy similarity class $[x_i]_B \cup [x_i]_E = [x_i]_B$ then $\left|[x_i]_B \cup [x_i]_E\right| = \left|[x_i]_B\right|$, therefore $\frac{\left|[x_i]_B \cup [x_i]_E\right|}{n} - \frac{\left|[x_i]_B\right|}{n} = 0$, so $\widetilde{FIGI}\left(B, E\right) = \widetilde{FIGI}\left(E\right)$. ■

Proposition 9. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E \subseteq C$, the fuzzy joint implication granularity information of B and E, $0 \leqslant \widetilde{FIGI}(B, E) < 1$.

For the attribute subsets $B, E \subseteq C$, based on the above definitions 9 to 12 the following relationships can be obtained.
(1) $M\widetilde{FIGI}(B; E) = \widetilde{FIGI}(B) + \widetilde{FIGI}(E) - \widetilde{FIGI}(B, E)$
(2) $M\widetilde{FIGI}(B; E) = \widetilde{FIGI}(B) - \widetilde{FIGI}(B|E)$
(3) $M\widetilde{FIGI}(B; E) = \widetilde{FIGI}(E) - \widetilde{FIGI}(E|B)$

Definition 13. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The fuzzy joint conditional implication granularity information of B under the conditions of knowing both E and F is defined as follows:
$\widetilde{FIGI}(B|E, F) =$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{|[x_i]_B \cup [x_i]_{E\cup F}|}{n} - \frac{|[x_i]_B|}{n}\right), \tag{13}$$

where $[x_i]_{E\cup F} = [x_i]_E \cap [x_i]_F$.

Proposition 10. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. We can obtain $\widetilde{FIGI}(B|E, F) \leqslant \widetilde{FIGI}(B|F)$ and $\widetilde{FIGI}(B|E, F) \leqslant \widetilde{FIGI}(B|E)$.

Proof: Clearly, as evident in the proof of Proposition 7. ■

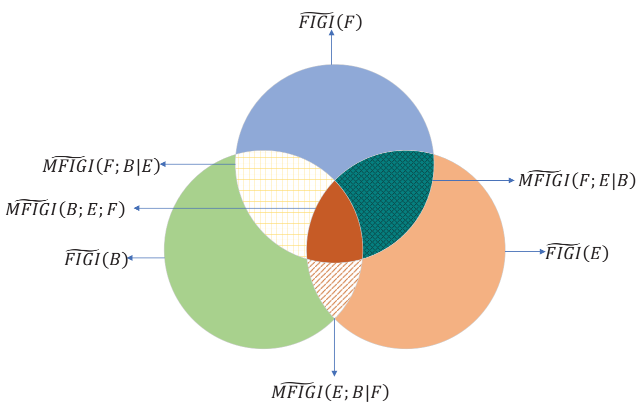


Fig. 2: The relationship among attribute subsets B, E and F.

In Fig. 2, we can also observe that there exists an additive and subtractive relationship among fuzzy implication granularity information, fuzzy conditional implication granularity information, and fuzzy mutual implication granularity information. Additionally, the fuzzy conditional implication granularity information we defined changes monotonically with the variation of attribute subsets, unlike the uncertain changes seen with fuzzy conditional entropy.

Definition 14. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The conditional mutual implication granularity between B and F under the knowledge of E, i.e., the fuzzy conditional mutual implication granularity information, is defined as follows:
$M\widetilde{FIGI}(F; B|E) =$

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{|[x_i]_B \cup [x_i]_E|}{n} - \frac{|[x_i]_B \cup [x_i]_{E\cup F}|}{n}\right), \tag{14}$$

where $[x_i]_{E\cup F} = [x_i]_E \cap [x_i]_F$.

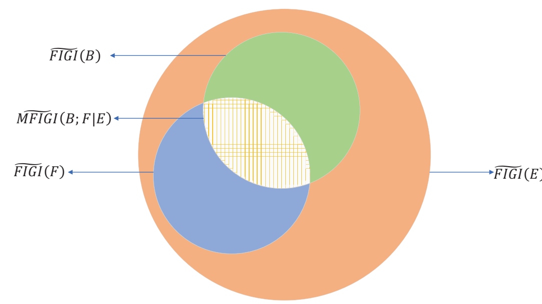


Fig. 3: E contains all the information of both B and F.

Proposition 11. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. If $E \supseteq B \wedge E \supseteq F$ it indicates that E has stronger discriminatory power than both B and F, meaning that E implies all the information in B and F as shown in Fig 3. Therefore, $M\widetilde{FIGI}(B; F|E) = 0$.

Proof: Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, if $E \supseteq B \wedge E \supseteq F$, and since $E \supseteq F$, then $[x_i]_{E\cup F} = [x_i]_E \cap [x_i]_F = [x_i]_E$, because $E \supseteq B$ we can obtain $[x_i]_B \cup [x_i]_E = [x_i]_B$. It is evident that $M\widetilde{FIGI}(B; F|E) = 0$. ■

Based on the Definition 11 and Definition 13, we can establish the relationships between the fuzzy conditional mutual implication granularity information, fuzzy conditional implication granularity information, and fuzzy joint conditional implication granularity information as follows:
(1) $M\widetilde{FIGI}(B; F|E) = \widetilde{FIGI}(F|E) - \widetilde{FIGI}(F|E, B)$
(2) $M\widetilde{FIGI}(B; F|E) = \widetilde{FIGI}(B|E) - \widetilde{FIGI}(B|E, F)$

Definition 15. Given an $FDT = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. The three-dimensional fuzzy mutual implication granularity information between B, E, and F is defined as follows:

$$\begin{aligned} M\widetilde{FIGI}(B; F; E) &= M\widetilde{FIGI}(B; F) - M\widetilde{FIGI}(B; F|E) \\ &= M\widetilde{FIGI}(B; E) - M\widetilde{FIGI}(B; E|F) \\ &= M\widetilde{FIGI}(E; F) - M\widetilde{FIGI}(E; F|B). \end{aligned} \tag{15}$$

Proposition 12. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. $\widetilde{MFIGI}\,(B;F;E) \geqslant 0$.

Proof: Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, for the attribute subsets $B, E, F \subseteq C$, conditional attribute set $C = \{a_1, a_2, \cdots, a_m\}$ consists of m attributes. $\widetilde{MFIGI}\,(B;F;E) = \widetilde{MFIGI}\,(B;F) - \widetilde{MFIGI}\,(B;F|E) = \frac{1}{n}\sum_{i=1}^{n}\left(1 + \frac{\left|[x_i]_F \cup [x_i]_{E\cup B}\right|}{n} - \frac{\left|[x_i]_B \cup [x_i]_F\right| + \left|[x_i]_E \cup [x_i]_F\right|}{n}\right)$, where $\left|[x_i]_F \cup [x_i]_{E\cup B}\right| = \left|([x_i]_F \cup [x_i]_E) \cap ([x_i]_F \cup [x_i]_B)\right|$. We assume that a is defined as $([x_i]_F \cup [x_i]_E)$ and b is defined as $([x_i]_F \cup [x_i]_B)$. In this case, the expression can be represented as $1 + \frac{\min\{a,b\}}{n} - \frac{a+b}{n}$. Additionally, it's worth noting that a and $b \leqslant n$, then $\frac{a}{n}$ and $\frac{b}{n} \leqslant 1$. So, Proposition 12 is proven. ■

Three-dimensional interaction information defined by information entropy can be positive or negative [48]. Under the measurement of our fuzzy implication granularity information, the three-dimensional interaction between attributes either has no interaction information or has interaction information, meaning that the value of the interaction information is bigger than or equal to 0. The fuzzy three-dimensional mutual implication granularity information value defined in our paper is non-negative as shown in Fig 2.

B. Attribute Reduction Algorithm Based on Fuzzy Implication Granularity Information

Based on the previous analysis, our fuzzy conditional implication granularity information possesses the property of monotonicity. In this section, we constructing a forward greedy algorithm based on fuzzy conditional implication granularity information.

Based on the analysis of Definition 11, we can comprehend that fuzzy conditional implication granularity information can be regarded as a supplement to the information about known attributes. Additionally, since the value of fuzzy implication granularity information for decision is fixed, we can construct a greedy forward attribute reduction algorithm based on monotonic fuzzy conditional implication granularity information.

Definition 16. Given an $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, $\forall a_i \in C$, we employ fuzzy conditional implication granularity information to measure the significance of attributes. The definition of attribute significance is as follows:

$$SIG\,(S, a_i, D) = \widetilde{FIGI}\,(D|S) - \widetilde{FIGI}\,(D|a_i \cup S)\,, \quad (16)$$

where S represents the current attribute subsets, a_i is the attribute under consideration for significance evaluation, and a higher computed value indicates greater attribute significance. Additionally, if the current reduced attribute subset space $S = \emptyset$, then $\widetilde{FIGI}\,(D|S) = \widetilde{FIGI}\,(D)$.

Therefore, we can construct a monotonic greedy forward attribute reduction algorithm based on equation 16.

This algorithm utilizes $SIG\,(red, a_i, D)$ to measure the significance of attributes. Since our fuzzy conditional implication granularity information is monotonic. We set the stopping criterion to halt the selection of attributes when the significance of adding any attribute to the red decreases below a small threshold. This implies that the addition of new attribute to the subset does not significantly enhance discriminative power.

Algorithm 1 A forward greedy attribute reduction algorithm based on the monotonic fuzzy conditional implication granularity information (MFIGI).

Input: $FD\mathrm{T} = \langle U, C \cup D, V \rangle$, attribute significance parameter ω.
Output: One reduct red.
Initialize $red = \emptyset$;
for each $a_i \in C$ **do**
  Compute the fuzzy relation R_{a_i};
end for
Compute the fuzzy conditional implication granularity information $\widetilde{FIGI}\,(D|C)$;
while $\widetilde{FIGI}\,(D|red) \neq \widetilde{FIGI}\,(D|C)$ **do**
  for each $a_i \in C - red$ **do**
    Compute $SIG\,(red, a_i, D)$;
  end for
  Find $a_k \in C - red$, where $SIG\,(S, a_k, D)$ is the biggest value;
  if $SIG\,(red, a_k, D) > \omega$ **then**
    $red \leftarrow red \cup \{a_k\}$;
    $C \leftarrow C - a_k$;
  else
    return red;
  end if
end while
return red

Assuming the attribute space C contains M attributes, the worst-case search time for reduction will require $\left(M^2 + M\right)/2$ attribute evaluations. The overall time complexity of the algorithm is $\left(M^2 + M\right)/2$.

IV. Experimental analysis

In this section, we compare our constructed attribute reduction algorithm (MFIGI) with six different algorithms capable of handling complex attribute reduction tasks. These include an attribute reduction algorithm based on fuzzy granularity, namely GFMRI [45], a algorithm based on fuzzy conditional entropy (FRSCE) [31], [32], a algorithm based on fuzzy information gain ratio (Gain Ratio) [44], a algorithm based on fuzzy self-information (FSI) [30], and a algorithm based on fitting fuzzy rough set model (NFRS) [18]. Additionally, neighborhood rough set model is a important tool for attribute reduction. So we also compared an advanced attribute reduction algorithm based on neighborhood combination entropy (FScNCE) [11].

We compared the classification accuracy on CART, KNN, and NB classifiers separately, demonstrating the advantages of our attribute reduction results. We demonstrate the advantages of our algorithm through the following three aspects:

(1) The classification performance of attribute reduction subsets on different classifiers;

(2) The length of attribute reduction subsets;

(3) The statistical analysis among the seven algorithms;

All experiments are implemented by MATLAB R2022a, and run on a personal computer with an Intel Core i9-12900H CPU at 5.00 GHz, 32.0 GB RAM.

A. Experimental design

In our experimental section, prior to attribute reduction, we first normalize the attributes using the following method:

$$F(x_i, a_k) = \frac{f(x_i, a_k) - \min_{a_k}}{\max_{a_k} - \min_{a_k}} \tag{17}$$

Here, $\max_{a_k}$ represents the maximum value in attribute a_k, $\min_{a_k}$ is the minimum value in attribute a_k, and $f(x_i, a_k)$ is the attribute value of sample x_i under attribute a_k.

The detailed information for the datasets used in our study is presented in Table I.

TABLE I: Description of datasets.

No.	Dataset	Samples	Attributes	Classes
1	Ecoli	336	7	8
2	CLL-SUB	111	11340	3
3	Prostate	102	5966	2
4	Wine	178	13	3
5	Solar-flare	1066	12	3
6	CNS	60	7129	2
7	Iranian Churn	3150	12	2
8	Colontumor	62	2000	2
9	Lung	203	3312	5
10	Seismic	2584	18	2
11	Glioma	50	4434	4
12	Winequality	4898	11	7
13	Lung-michigan	96	7129	2
14	ILPD	583	10	2
15	Ovarian	253	15154	2

These three classifiers, encompassing Decision Tree (CART), K-Nearest Neighbors (KNN with K = 3), and Naive Bayes (NB), stand as prevalent tools frequently used in measuring the classification accuracy of diverse datasets.

CART: a tree-based algorithm for classification and regression tasks. It creates a tree structure in attribute space, splitting the dataset into subsets at nodes based on attributes. It progressively selects optimal attributes for splitting, forming a tree used for predicting classes or values for new data.

KNN: An instance-based learning method for classification and regression. It predicts by measuring distances in attribute space.

NB: A set of probability classifiers based on Bayes' theorem. It assumes attribute independence, meaning each attribute's effect on classification is treated as independent. Despite this assumption not always holding in real data, Naive Bayes performs well in tasks like text classification and spam filtering, even with limited data.

To comprehensively measure the performance of these classifiers on datasets, ten-fold cross-validation has been employed.

B. Analysis of Experimental Results

Table II to Table IV respectively present the classification accuracy results of our MFIGI algorithm and the other six algorithms across three different classifiers.

As the FScNCE algorithm is based on neighborhood rough set model and involve a neighborhood radius parameter. In this paper, to traverse the neighborhood radius from 0.05 to 0.3 with a step size of 0.05, obtaining the optimal classification accuracy for comparison in this paper.

Because our MFIGI algorithm 1 is a monotonic forward greedy attribute reduction algorithm, we use a stopping parameter ω to halt the attribute reduction process when measuring the significance of attributes. When the significance of an attribute below ω, that the current attribute's contribution to enhancing the discriminative power of the attribute subset is almost negligible, prompting us to stop the attribute reduction process. For high-dimensional datasets, we set the parameter as $\omega = 0.005$, while for low-dimensional datasets, we set $\omega = 0.002$. In this paper, high-dimensional datasets refer to datasets with more than 1000 attributes, while low-dimensional datasets comprise datasets with attributes ranging from 0 to 1000.

Table V displays the length of the attribute subsets obtained by all seven algorithms across all fifteen datasets. It's important to note that the FScNCE algorithm is an attribute reduction algorithm based on the neighborhood rough set model. Therefore, the attribute subset lengths for the three classifiers may not be consistent. Hence, we use the average length of the attribute subset obtained by the FScNCE algorithm across the three classifiers for comparison. We can see that the average length of the attribute subset obtained using our MFIGI algorithm, 6.07, is the shortest among all seven algorithms. For the attribute reduction accuracy results in Tables II, III, and IV, we have highlighted the optimal results in bold font and underscored the second-best results for display.

As shown in Table II, our MFIGI algorithm achieved the top subset classification accuracy on ten datasets and secured the second-best classification accuracy on four datasets when using the NB classifier. Across the total of 15 datasets, our algorithm demonstrated excellent classification accuracy on 14 datasets. Simultaneously, our average classification accuracy of 85.98% surpassed the average accuracy of the second-ranked FRSCE algorithm, which stood at 81.95%, by 4.03%. Notably, our classification accuracy outperformed the original dataset's accuracy across all datasets, indicating that our MFIGI algorithm improved classification accuracy while reducing the dimensionality of attributes.

As shown in Table III, on the CART classifier, our MFIGI algorithm achieved the top subset classification accuracy on eight datasets and secured the second-best classification accuracy on four datasets. Across the total of 15 datasets, our algorithm demonstrated excellent classification accuracy on 12 datasets. Simultaneously, our average classification accuracy of 85.69% surpassed the average accuracy of the second-ranked Gain Ratio algorithm, which stood at 83.00%, by 2.69%. Notably, our classification accuracy was higher than

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3381993

TABLE II: Classification accuracies of attributes subsets with NB (%).

Dataset	ALL	NFRS	FSI	Gain Ratio	GFMRI	FRSCE	FScNCE	MFIGI
CLL-SUB	59.47±14.92	69.24±10.00	68.48±12.22	61.36±9.88	72.12±22.35	**75.91±16.33**	65.83±15.67	75.76±11.07
Ovarian	88.19±8.71	95.63±4.41	98.49±3.60	96.89±5.11	96.86±3.59	**100±0.00**	99.60±1.26	**100±0.00**
Lung-michigan	94.78±5.52	93.67±8.92	**98.89±3.51**	93.78±7.15	95.67±5.60	97.89±4.46	**98.89±3.51**	97.78±4.68
Colontumor	80.48±21.67	77.14±17.68	83.81±15.75	**89.05±12.86**	80.48±16.87	83.81±13.03	84.05±15.23	85.71±16.03
Prostate	64.18±17.32	82.00±14.76	91.18±12.86	71.64±13.42	**92.09±9.17**	85.27±9.69	81.18±7.36	**92.09±10.31**
Glioma	26.33±16.21	47.17±26.15	47.33±31.69	64.50±16.89	65.33±20.86	50.17±17.51	63.83±21.55	**72.67±19.99**
Lung	85.14±9.00	91.70±5.88	81.21±9.02	86.40±8.08	77.50±4.16	92.32±7.79	86.79±8.36	**95.16±6.39**
Wine	97.68±4.08	96.54±5.58	96.60±2.93	97.68±4.08	96.63±3.91	90.42±3.82	97.68±4.08	**98.27±3.93**
Solar-flare	98.87±0.74	**99.53±0.66**	99.16±0.69	98.50±1.19	**99.53±0.66**	98.69±0.91	98.59±0.91	**99.53±0.66**
Seismic	41.25±7.90	**93.42±0.01**	91.64±0.97	79.33±4.60	**93.42±0.01**	**93.42±0.01**	91.45±1.01	**93.42±0.01**
ILPD	66.37±4.24	66.88±3.12	69.84±4.17	66.37±4.24	**71.51±2.23**	66.36±4.02	68.79±4.33	71.20±1.27
Iranian Churn	83.55±2.16	88.86±1.08	86.03±1.39	83.55±2.16	89.30±1.39	87.17±1.55	84.38±1.85	**89.94±1.14**
CNS	69.05±13.75	68.95±14.16	75.29±18.02	73.81±19.44	65.95±20.05	80.24±10.65	65.05±3.48	**83.24±11.19**
Winequality	48.40±1.98	**52.55±2.64**	44.65±1.62	48.40±1.98	48.40±1.98	49.75±1.97	49.77±2.23	49.77±1.71
Ecoli	85.21±6.23	**85.21±6.23**	84.45±7.17	**85.21±6.23**	38.90±7.68	77.79±5.15	**85.21±6.23**	**85.21±6.23**
Average	72.60±8.96	80.78±7.47	81.15±7.69	79.77±6.99	78.91±7.18	81.95±6.16	81.87±6.69	**85.98±5.96**

TABLE III: Classification accuracies of attributes subsets with CART (%).

Dataset	ALL	NFRS	FSI	Gain Ratio	GFMRI	FRSCE	FScNCE	MFIGI
CLL-SUB	57.80±15.37	72.05±8.02	72.80±14.95	63.94±13.59	70.45±15.34	68.56±13.35	67.50±12.40	**74.77±18.07**
Ovarian	98.83±1.89	98.03±2.81	**98.80±1.93**	98.40±2.80	95.69±3.79	98.40±2.80	**98.80±1.93**	98.40±2.80
Lung-michigan	96.67±5.37	**98.89±3.51**	**98.89±3.51**	97.78±4.68	97.78±4.68	**98.89±3.51**	**98.89±3.51**	**98.89±3.51**
Colontumor	72.38±15.92	**82.62±11.11**	71.19±15.33	80.48±10.75	70.71±17.39	79.29±10.29	74.52±12.65	82.38±16.20
Prostate	82.45±9.98	87.10±13.40	86.27±12.29	88.09±10.35	**90.10±12.48**	88.09±10.35	86.18±10.65	89.09±11.02
Glioma	77.17±14.57	58.67±18.20	73.50±17.29	**87.00±15.29**	78.83±16.22	59.83±20.85	69.00±21.83	86.50±12.43
Lung	90.11±7.68	**93.46±4.19**	80.35±9.36	86.79±6.29	75.78±5.61	90.15±4.61	84.19±8.26	89.67±7.21
Wine	91.56±5.51	92.64±6.61	91.41±6.34	91.01±6.06	91.00±6.01	83.19±6.27	91.56±5.51	**92.67±6.01**
Solar-flare	99.35±0.99	**99.53±0.66**	99.35±0.99	99.35±0.99	**99.53±0.66**	99.35±0.99	99.35±0.99	**99.53±0.66**
Seismic	89.32±0.38	**93.42±0.01**	90.94±1.65	89.51±1.39	**93.42±0.01**	93.30±0.19	90.13±0.81	**93.42±0.01**
ILPD	64.73±5.98	66.94±7.22	66.08±6.30	65.93±5.89	66.73±4.53	66.40±4.16	67.09±5.67	**69.97±4.12**
Iranian Churn	93.74±1.43	90.29±1.54	**94.16±1.61**	93.65±1.50	92.92±1.91	90.95±1.57	93.93±1.51	92.45±2.06
CNS	56.71±13.52	64.48±21.13	63.29±14.90	63.38±14.80	78.38±15.26	65.62±16.33	65.62±16.33	**78.48±15.42**
Winequality	59.31±2.90	54.72±1.82	50.12±2.78	**59.53±3.03**	59.29±3.04	49.61±2.11	59.31±1.92	58.78±2.47
Ecoli	80.10±8.65	80.12±9.05	80.14±7.62	80.14±9.02	55.32±7.12	77.41±7.61	80.10±8.65	**80.41±8.88**
Average	80.68±7.34	82.20±6.30	81.15±7.28	83.00±6.54	81.06±7.06	80.60±6.35	81.75±6.88	**85.69±6.82**

the original dataset's accuracy across 11 datasets, indicating that our MFIGI algorithm improved classification accuracy while reducing the dimensionality of attributes.

As shown in Table IV, on the KNN classifier, our MFIGI algorithm achieved the top subset classification accuracy on nine datasets, with an average accuracy of 85.70%, which is 0.53% higher than the average accuracy of the second-ranked FScNCE algorithm, standing at 85.16%. Across 12 datasets, our classification accuracy was higher than the original dataset's accuracy. Additionally, the subset obtained based on the measurement through fuzzy conditional implication granularity information showed consistent classification accuracy on the KNN classifier, as depicted in detail in Fig 5.

We obtained the reduced subset using our MFIGI algorithm and then plotted the changes in subset classification accuracy alongside the variations in our fuzzy conditional implication granularity information, as detailed in Fig 5. The blue line represents the trend of fuzzy conditional implication granularity information values. It can be observed that the attribute subsets obtained through our monotonic fuzzy conditional implication granularity information exhibit a generally monotonic trend in classification accuracy across different classifiers. We observe that the values of Fuzzy Conditional Implication Granularity Information (FCIGI), monotonically decrease as the number of attributes increases. The declining curve is notably smooth, with all values ranging between 0 and 1. This indicates that our Fuzzy Conditional Implication Granularity Information is an effective and reasonable measure the relationships between attributes and decision.

C. Statistical Analysis

This section demonstrates the ranking of the seven algorithms used in our paper based on their classification accuracies across different classifiers. Firstly, Friedman test [49] was conducted to obtain statistical results for comparing the classification accuracies resulting from these seven algorithms. With $\alpha = 0.1$, the Friedman test rejects the null hypothesis of all algorithms having the same performance, indicating a significant difference in performance among these attribute reduction algorithms.

$$F_F = \frac{(N-1)\,\chi_F^2}{N\,(k-1) - \chi_F^2},$$

TABLE IV: Classification accuracies of attributes subsets with KNN (%).

Dataset	ALL	NFRS	FSI	Gain Ratio	GFMRI	FRSCE	FScNCE	MFIGI
CLL-SUB	53.26±15.98	73.03±16.44	67.50±15.67	65.83±14.46	75.00±13.08	72.95±10.52	72.20±11.92	**80.23±11.03**
Ovarian	93.29±2.66	99.22±1.65	99.63±1.17	98.49±3.60	96.09±4.03	**100±0.00**	99.60±1.26	**100±0.00**
Lung-michigan	98.89±3.51	97.78±4.68	94.67±5.64	94.56±7.78	**98.89±3.51**	**98.89±3.51**	**98.89±3.51**	**98.89±3.51**
Colontumor	72.62±15.07	72.38±8.26	77.38±11.46	84.05±13.04	77.38±15.96	79.05±13.46	**87.38±9.53**	84.29±12.41
Prostate	81.27±15.04	86.09±11.78	87.45±12.02	90.18±6.67	91.09±11.97	92.00±7.89	**94.00±10.75**	**94.00±8.43**
Glioma	80.50±15.68	64.83±10.96	58.83±25.84	70.50±17.85	**77.50±11.26**	68.17±20.55	74.50±16.18	66.83±19.44
Lung	95.99±3.89	94.09±5.15	83.01±9.27	90.21±5.47	75.17±7.81	**95.06±6.96**	86.20±9.72	94.69±4.79
Wine	94.34±3.82	95.49±3.54	94.47±4.30	94.34±3.82	93.75±4.28	89.87±5.91	**97.71±2.96**	96.04±2.74
Solar-flare	99.53±0.66	**99.53±0.66**	**99.53±0.66**	**99.53±0.66**	**99.53±0.66**	**99.53±0.66**	**99.53±0.66**	**99.53±0.66**
Seismic	92.69±0.77	**93.42±0.01**	92.57±0.79	92.57±0.75	92.26±0.99	92.45±0.86	92.76±1.05	**93.42±0.01**
ILPD	66.19±5.68	67.53±7.28	65.33±3.03	66.19±5.68	68.44±3.47	69.11±2.91	66.37±5.42	**69.99±5.37**
Iranian Churn	95.24±1.19	90.32±1.32	95.08±1.28	95.24±1.19	93.81±1.18	91.39±1.92	**95.26±1.24**	92.10±1.72
CNS	57.76±21.27	70.95±20.32	80.48±12.69	71.57±8.01	69.71±14.22	**84.90±9.58**	72.29±17.16	74.81±14.55
Winequality	56.25±2.02	53.92±1.69	47.08±2.04	56.25±2.02	56.25±2.02	37.44±2.30	**56.35±2.81**	**56.35±1.44**
Ecoli	84.29±4.62	**84.29±4.62**	81.04±7.07	**84.29±4.62**	48.48±4.29	78.60±5.67	**84.29±4.62**	**84.29±4.62**
Average	81.47±7.46	82.86±5.57	81.60±7.16	83.59±6.26	80.89±6.04	83.30±5.99	85.16±5.83	**85.70±5.44**

TABLE V: Sizes of attribute subsets.

Dataset	ALL	NFRS	FSI	Gain Ratio	GFMRI	FRSCE	FScNCE	MFIGI
CLL-SUB	11340	20	24	27	12	150	4.33	9
Ovarian	15154	6	19	10	2	6	2.33	5
Lung-michigan	7129	62	18	7	2	20	1	2
Colontumor	2000	71	20	12	13	70	8	8
Prostate	5966	91	21	19	7	118	3	8
Glioma	4434	71	29	12	16	78	3.33	9
Lung	3312	55	25	15	6	76	3.67	6
Wine	13	12	10	13	7	2	10.67	11
Solar-flare	12	1	4	9	2	10	6	1
Seismic	18	1	12	14	2	7	8.67	1
ILPD	10	8	7	10	2	7	8.67	1
Iranian Churn	12	4	10	12	9	7	11	6
CNS	7129	93	20	48	12	75	3.67	8
Winequality	11	3	3	11	11	1	10.33	10
Ecoli	7	7	5	7	3	6	6	6
Average	3769.8	33.67	15.13	12.87	7.07	41.93	6.31	**6.07**

where

$$\chi_F^2 = \frac{12N}{k\left(k+1\right)}\left(\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right).$$

Given k comparison algorithms and N datasets. R_j represents the ranking of the j-th algorithm for all dataset. The calculated significance value CV = 1.8455. The performance differences between our algorithm and others are reflected in the Friedman statistic (F_F), with a value of 2.3781 for the CART classifier, 3.7712 for the KNN classifier, and 3.5059 for the NB classifier.

Additionally, a Nemenyi post-hoc test [50] was conducted to compare the differences among the seven algorithms.

$$CD_\alpha = q_\alpha \sqrt{\frac{k\left(k+1\right)}{6N}}, \tag{18}$$

where q_α is the critical tabulated value, k and N are the number of datasets and algorithms. If the difference in average ranks between two algorithms exceeds the critical difference (CD), the hypothesis of equal performance between the two algorithms is rejected at the respective confidence level. For a confidence level of $\alpha = 0.1$, the CD value for the seven algorithms across 15 datasets is 2.1241.

Tables VI, VII, and VIII respectively showcase the rankings of subset accuracy results obtained from experiments with seven attribute reduction algorithms across 15 datasets on the NB, CART, and KNN classifiers. It is observable that our MFIGI algorithm achieves the best average ranking across all three classifiers. Notably, on the NB classifier, our MFIGI algorithm's average ranking across all datasets is 2, which is 1.83 points higher than the second-best FRS attribute reduction algorithm's average ranking of 3.83. Additionally, based on the results of the experimental classification accuracy rankings, we conducted a Nemenyi test to further analyze the experimental results, as illustrated in Fig.4.

Fig 4 (a), (b), and (c) respectively represent the results of the Nemenyi test across three classifiers.

In conclusion, the experimental results show that the MFIGI algorithm obtains the optimal classification accuracy of the seven algorithms on CART, KNN, and NB classifiers while also obtaining the shortest subset of attributes among the seven algorithms. This shows that our MFIGI algorithm is indeed

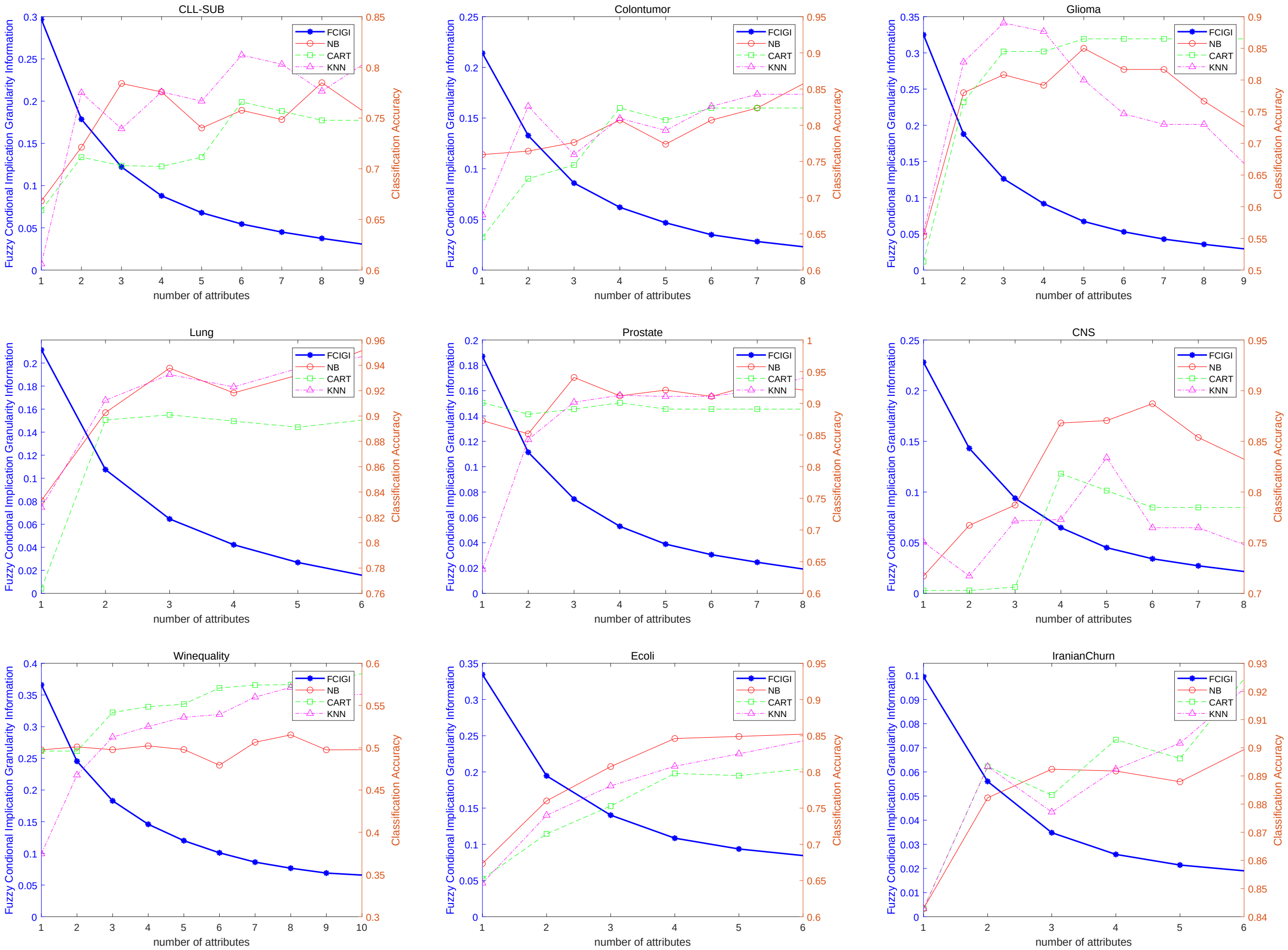
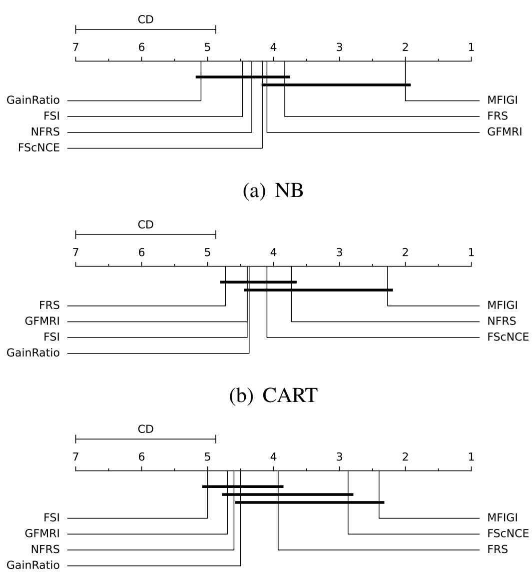


Fig. 5: Fuzzy Conditional Implication Granularity Information and Subset Accuracy Variation



(a) NB

(b) CART

(c) KNN

Fig. 4: Accuracy comparison with seven algorithms on three classifiers.

TABLE VI: T

he rankings of seven algorithms on NB.

Dataset	NFRS	FSI	Gain Ratio	GFMRI	FRS	FScNCE	MFIGI
CLL-SUB	4	5	7	3	1	6	2
Ovarian	4	5	6	7	1.5	3	1.5
Lung-michigan	7	1	6	5	2	3.5	3.5
Colontumor	7	3.5	1	6	3.5	2	5
Prostate	6	3.5	7	1.5	5	3.5	1.5
Glioma	7	6	3	2	5	4	1
Lung	3	6	5	7	2	4	1
Wine	6	5	2.5	4	7	2.5	1
Solar-flare	2	4	7	2	5	6	2
Seismic	2.5	5	7	2.5	2.5	6	2.5
ILPD	5	3	6	1	7	4	2
Iranian Churn	3	5	7	2	4	6	1
CNS	5	3	4	6	2	7	1
Winequality	1	7	5.5	5.5	4	2.5	2.5
Ecoli	2.5	5	2.5	7	6	2.5	2.5
Average	4.33	4.47	5.1	4.1	<u>3.83</u>	4.17	**2**

better than the other algorithms.

V. CONCLUSIONS AND FUTURE WORK

In this paper, a measurement called Fuzzy Implication Granularity Information and its generalized forms to measure relationships between attributes. While fuzzy conditional information entropy lacks monotonicity, our proposed Fuzzy Conditional Implication Granularity Information exhibits monotonicity. Fuzzy entropy involves a large number of logarithmic

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2024.3381993

TABLE VII: T

he rankings of seven algorithms on CART.

Dataset	NFRS	FSI	Gain Ratio	GFMRI	FRS	FScNCE	MFIGI
CLL-SUB	3	2	7	4	5	6	1
Ovarian	6	1.5	4	7	4	1.5	4
Lung-michigan	3	3	6.5	6.5	3	3	3
Colontumor	1	6	3	7	4	5	2
Prostate	5	6	3.5	1	3.5	7	2
Glioma	7	4	1	3	6	5	2
Lung	1	6	4	7	2	5	3
Wine	2	4	5	6	7	3	1
Solar-flare	2	5.5	5.5	2	5.5	5.5	2
Seismic	2	5	7	2	4	6	2
ILPD	3	6	7	4.5	4.5	2	1
Iranian Churn	7	1	3	4	6	2	5
CNS	5	7	6	2	3.5	3.5	1
Winequality	5	6	1	3	7	2	4
Ecoli	4	3	2	7	6	5	1
Average	<u>3.73</u>	4.4	4.37	4.4	4.73	4.1	**2.27**

TABLE VIII: T

he rankings of seven algorithms on KNN.

Dataset	NFRS	FSI	Gain Ratio	GFMRI	FRS	FScNCE	MFIGI
CLL-SUB	3	6	7	2	4	5	1
Ovarian	5	3	6	7	1.5	4	1.5
Lung-michigan	5	6	7	2.5	2.5	2.5	2.5
Colontumor	7	5.5	3	5.5	4	1	2
Prostate	7	6	5	4	3	1.5	1.5
Glioma	6	7	3	1	4	2	5
Lung	3	6	4	7	1	5	2
Wine	3	4	5	6	7	1	2
Solar-flare	4	4	4	4	4	4	4
Seismic	1.5	4.5	4.5	7	6	3	1.5
ILPD	4	7	6	3	2	5	1
Iranian Churn	7	3	2	4	6	1	5
CNS	6	2	5	7	1	4	3
Winequality	5	6	3.5	3.5	7	1.5	1.5
Ecoli	2.5	5	2.5	7	6	2.5	2.5
Average	4.6	5	4.5	4.7	3.93	<u>2.87</u>	**2.4**

function computations, and value of the result is unbounded, which could potentially affect the attribute reduction process. However, all measurement based on fuzzy implication granularity information avoid logarithmic function computations and the value of the result within the range of 0 to 1. Extensive experiments and statistical analysis demonstrate that our algorithm achieves better attribute subset lengths while maintaining good classification accuracy across various datasets. As shown in Fig 5, our fuzzy conditional implication granularity information, depicted by the blue line, exhibits a decreasing trend. Simultaneously, the overall trend of subset classification accuracy is increasing.

Future work aims to extend the application of fuzzy implication granularity information to multi-label data in attribute reduction tasks. Based on our fuzzy implication granularity information metric, we comprehensively consider the interaction between attributes and labels, optimizing the selection of attribute subsets. Furthermore, due to the proposition of our fuzzy implication granularity information in measuring the relationships between attributes, it avoids the computational burden of logarithmic operations typically present in information entropy, resulting in a lesser computational burden. Additionally, in our paper, we introduced the concepts of Fuzzy Conditional Mutual Implication Granularity Information and Three-Dimensional Fuzzy Mutual Implication Granularity Information, which can be utilized to measure the relationships among multiple attributes. Therefore, the measurement approach of our fuzzy implication granularity information could be considered for handling more complex data tasks in the future. Due to the use of tolerance relations to determine relationships among samples in incomplete information systems, the fuzzy implication granularity information measurement approach proposed in our paper demonstrates good generalizability and robustness. Therefore, it can also be applied in incomplete information systems to measure the relationships between attributes as well as the relationship between attributes and the decision. Utilizing our fuzzy implication granularity information metric in tolerance relation environments to carry out attribute reduction tasks.

REFERENCES

[1] D. Dubois and H. Prade, “Rough fuzzy sets and fuzzy rough sets,” *International Journal of General Systems*, vol. 17, no. 2-3, pp. 191–209, 1990.

[2] J. Dai, J. Chen, Y. Liu, and H. Hu, “Novel multi-label feature selection via label symmetric uncertainty correlation learning and feature redundancy evaluation,” *Knowledge-Based Systems*, vol. 207, p. 106342, 2020.

[3] J. Dai, H. Hu, W.-Z. Wu, Y. Qian, and D. Huang, “Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets,” *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 4, pp. 2174–2187, 2018.

[4] J. Dai, Q. Hu, H. Hu, and D. Huang, “Neighbor inconsistent pair selection for attribute reduction by rough set approach,” *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 2, pp. 937–950, 2018.

[5] R. Jensen and Q. Shen, “Fuzzy-rough sets assisted attribute selection,” *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 1, pp. 73–89, 2007.

[6] Z. Pawlak, “Rough rets,” *International Journal of Computer & Information Sciences*, vol. 11, no. 5, pp. 341–356, 1982.

[7] J. Qian, D. Q. Miao, Z. H. Zhang, and W. Li, “Hybrid approaches to attribute reduction based on indiscernibility and discernibility relation,” *International Journal of Approximate Reasoning*, vol. 52, no. 2, pp. 212–230, 2011.

[8] E. C. C. Tsang, D. Chen, D. S. Yeung, X.-Z. Wang, and J. W. T. Lee, “Attributes reduction using fuzzy rough sets,” *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 5, pp. 1130–1141, 2008.

[9] J. Dai, Q. Hu, J. Zhang, H. Hu, and N. Zheng, “Attribute selection for partially labeled categorical data by rough set approach,” *IEEE Transactions on Cybernetics*, vol. 47, no. 9, pp. 2460–2471, 2017.

[10] J. Wan, H. Chen, T. Li, Z. Yuan, J. Liu, and W. Huang, “Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures,” *IEEE Transactions on Cybernetics*, vol. 53, no. 2, pp. 1208–1221, 2023.

[11] P. Zhang, T. Li, Z. Yuan, C. Luo, K. Liu, and X. Yang, “Heterogeneous feature selection based on neighborhood combination entropy,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 3, pp. 3514–3527, 2022.

[12] Y. Qian, J. Liang, W. Pedrycz, and C. Dang, “Positive approximation: an accelerator for attribute reduction in rough set theory,” *Artificial Intelligence*, vol. 174, no. 9, pp. 597–618, 2010.

[13] Z. Pawlak and A. Skowron, “Rudiments of rough sets,” *Information Sciences*, vol. 177, no. 1, pp. 3–27, 2007.

[14] Y. Yang, D. Chen, H. Wang, and X. Wang, “Incremental perspective for feature selection based on fuzzy rough sets,” *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 3, pp. 1257–1273, 2018.

[15] H. Wang, “Nearest neighbors by neighborhood counting,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 942–953, 2006.

[16] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, “A novel algorithm for finding reducts with fuzzy rough sets,” *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 385–389, 2012.

[17] J. Dai and H. Tian, “Fuzzy rough set model for set-valued data,” *Fuzzy Sets and Systems*, vol. 229, pp. 54–68, 2013.

[18] C. Wang, Y. Qi, M. Shao, Q. Hu, D. Chen, Y. Qian, and Y. Lin, “A fitting model for feature selection with fuzzy rough sets,” *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 741–753, 2017.

[19] C. Wang, Y. Qian, W. Ding, and X. Fan, “Feature selection with fuzzy-rough minimum classification error criterion,” *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 8, pp. 2930–2942, 2022.

[20] G. Liu, Y. Yang, and B. Li, “Fuzzy rule-based oversampling technique for imbalanced and incomplete data learning,” *Knowledge-Based Systems*, vol. 158, pp. 154–174, 2018.

[21] W. Xu, K. Yuan, W. Li, and W. Ding, "An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 1, pp. 76–88, 2023.

[22] W. Li, H. Zhou, W. Xu, X.-Z. Wang, and W. Pedrycz, "Interval dominance-based feature selection for interval-valued ordered data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 6898–6912, 2023.

[23] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, 2008.

[24] L. Sun, X.-Y. Zhang, Y.-H. Qian, J.-C. Xu, S.-G. Zhang, and Y. Tian, "Joint neighborhood entropy-based gene selection method with fisher score for tumor classification," *Applied Intelligence*, vol. 49, no. 4, pp. 1245–1259, 2019.

[25] L. Sun, T. Yin, W. Ding, Y. Qian, and J. Xu, "Feature selection with missing labels using multilabel fuzzy neighborhood rough sets and maximum relevance minimum redundancy," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 5, pp. 1197–1211, 2022.

[26] P. Zhu and Q. Hu, "Adaptive neighborhood granularity selection and combination based on margin distribution optimization," *Information Sciences*, vol. 249, pp. 1–12, 2013.

[27] B. Sang, W. Xu, H. Chen, and T. Li, "Active antinoise fuzzy dominance rough feature selection using adaptive k-nearest neighbors," *IEEE Transactions on Fuzzy Systems*, vol. 31, no. 11, pp. 3944–3958, 2023.

[28] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 7, pp. 2986–2999, 2018.

[29] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on Fuzzy Systems*, vol. 17, no. 4, pp. 824–838, 2009.

[30] C. Wang, Y. Huang, W. Ding, and Z. Cao, "Attribute reduction with fuzzy rough self-information measures," *Information Sciences*, vol. 549, pp. 68–86, 2021.

[31] Q. Hu, D. Yu, Z. Xie, and J. Liu, "Fuzzy probabilistic approximation spaces and their information measures," *IEEE Transactions on Fuzzy Systems*, vol. 14, no. 2, pp. 191–201, 2006.

[32] Q. Hu, D. Yu, and Z. Xie, "Information-preserving hybrid data reduction based on fuzzy-rough techniques," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 414–423, 2006.

[33] Y. Qian, J. Liang, and C. Dang, "Knowledge structure, knowledge granulation and knowledge distance in a knowledge base," *International Journal of Approximate Reasoning*, vol. 50, no. 1, pp. 174–188, 2009.

[34] J. Liang and Y. Qian, "Information granules and entropy theory in information systems," *Science in China Series F: Information Sciences*, vol. 51, no. 10, pp. 1427–1444, 2008.

[35] J. T. Yao, A. V. Vasilakos, and W. Pedrycz, "Granular computing: perspectives and challenges," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1977–1989, 2013.

[36] J. Liang and Z. Shi, "The information entropy, rough entropy and knowledge granulation in rough set theory," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 12, no. 1, pp. 37–46, 2004.

[37] Y. Qian, J. Liang, W. Wu, and C. Dang, "Information granularity in fuzzy binary grc model," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 2, pp. 253–264, 2011.

[38] J. Dai and H. Tian, "Entropy measures and granularity measures for set-valued information systems," *Information Sciences*, vol. 240, pp. 72–82, 2013.

[39] L. Sun, L. Wang, W. Ding, Y. Qian, and J. Xu, "Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 19–33, 2021.

[40] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.

[41] R. R. Yager, "Entropy measures under similarity relations," *International Journal of General Systems*, vol. 20, no. 4, pp. 341–358, 1992.

[42] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[43] E. Hernández and J. Recasens, "A reformulation of entropy in the presence of indistinguishability operators," *Fuzzy Sets and Systems*, vol. 128, no. 2, pp. 185–196, 2002.

[44] J. Dai and Q. Xu, "Attribute selection based on information gain ratio in fuzzy rough set theory with application to tumor classification," *Applied Soft Computing*, vol. 13, no. 1, pp. 211–221, 2013.

[45] J. Dai and J. Chen, "Feature selection via normative fuzzy information weight with application into tumor classification," *Applied Soft Computing*, vol. 92, p. 106299, 2020.

[46] N. N. Morsi and M. M. Yakout, "Axiomatics for fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 100, no. 1, pp. 327–342, 1998.

[47] D. Yeung, D. Chen, E. Tsang, J. Lee, and W. Xizhao, "On the generalization of fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 3, pp. 343–361, 2005.

[48] A. Jakulin, "Machine learning based on attribute interactions," Ph.D. dissertation, Univerza v Ljubljani, 2005.

[49] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.

[50] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.