Computers, Materials & Continua

Tech Science Press

DOI: 10.32604/cmc.2024.057383

ARTICLE

Attribute Reduction on Decision Tables Based on Hausdorff Topology

Nguyen Long Giang[1], Tran Thanh Dai[2], Le Hoang Son[3], Tran Thi Ngan[4], Nguyen Nhu Son[1] and Cu Nguyen Giap[5,*]

[1]Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, 10000, Vietnam

[2]Faculty of Information Technology, University of Economics-Technology for Industries, Hanoi, 10000, Vietnam

[3]VNU Information Technology Institute, Vietnam National University, Hanoi, 10000, Vietnam

[4]International School, Vietnam National University, Hanoi, 10000, Vietnam

[5]Center of Science and Technology Research and Development, Thuongmai University, Hanoi, 10000, Vietnam

*Corresponding Author: Cu Nguyen Giap. Email: cunguyengiap@tmu.edu.vn

Received: 16 August 2024 Accepted: 16 October 2024 Published: 18 November 2024

ABSTRACT

Attribute reduction through the combined approach of Rough Sets (RS) and algebraic topology is an open research topic with significant potential for applications. Several research works have introduced a strong relationship between RS and topology spaces for the attribute reduction problem. However, the mentioned recent methods followed a strategy to construct a new measure for attribute selection. Meanwhile, the strategy for searching for the reduct is still to select each attribute and gradually add it to the reduct. Consequently, those methods tended to be inefficient for high-dimensional datasets. To overcome these challenges, we use the separability property of Hausdorff topology to quickly identify distinguishable attributes, this approach significantly reduces the time for the attribute filtering stage of the algorithm. In addition, we propose the concept of Hausdorff topological homomorphism to construct candidate reducts, this method significantly reduces the number of candidate reducts for the wrapper stage of the algorithm. These are the two main stages that have the most effect on reducing computing time for the attribute reduction of the proposed algorithm, which we call the Cluster Filter Wrapper algorithm based on Hausdorff Topology. Experimental validation on the UCI Machine Learning Repository Data shows that the proposed method achieves efficiency in both the execution time and the size of the reduct.

KEYWORDS

Hausdorff topology; rough sets; topology from rough sets; attribute reduction

1 Introduction

The rapid development of technologies in data collection and integration increases the complexity of data dimensions and noisy attributes [1,2]. Attribute reduction is indeed critical to selecting attributes according to the most significant contribution in the dataset [3,4]. Some applications of attribute reduction include data classification [1,2], and recommendation systems [5,6]. Recently, there have been some effective applications of RST in decision making support, including the diagnosis support in heart disease [7], COVID-19 (Corona Virus Disease 2019) [8], Chikungunya disease [9],

Dengue fever [10] and data reduction [11]. From the original attribute set, attribute reduction aims to find out a subset that satisfies given constraints. Rough set theory (RST) is an effective tool for solving the attribute reduction problem for a long time [12–14]. RST-based attribute reduction methods often define a measure as the criteria for selecting the attributes. Some measures include fuzzy POS measure [15–17], intuitionistic fuzzy POS measure [18,19], fuzzy entropy measure [20–22], intuitionistic fuzzy entropy measure [23], fuzzy distance [24], intuitionistic fuzzy distance [25]. These measures are all based on approximate space for definition, so these approaches are expensive in storage and computation time.

Topological space is a powerful mathematical tool that is widely applied in many different areas of life [26–28]. The structure of a topology not only expresses the relationships among the objects in a set but also the relationships among the groups of objects in a set [29–31]. The reduct topology concept, proposed by Lashin et al. [29], grasped the attention of researchers in recent years [30,31]. Based on the mechanical similarity between rough sets and topologies [29], researchers have proposed many methods for building topologies according to the rough set approach, such as topology based on the RST covering approach [31], β-covering based rough sets [32], by the fuzzy rough set approach [33,34], by the intuitionistic fuzzy rough set approach [35,36] and by intuitionistic fuzzy approximate space [37]. The researchers also pointed out when the operations in the inner and outer regions of the topology are equivalent to the approximation operations of the rough set [38,39]. From the relationship between approximation operations of RST and topology, new RST models on topological space have been proposed [40]. The researchers also showed that the RST and topology are equivalent in the case of approximate space. Fig. 1 shows the approaches to attribute reduction in a decision table.

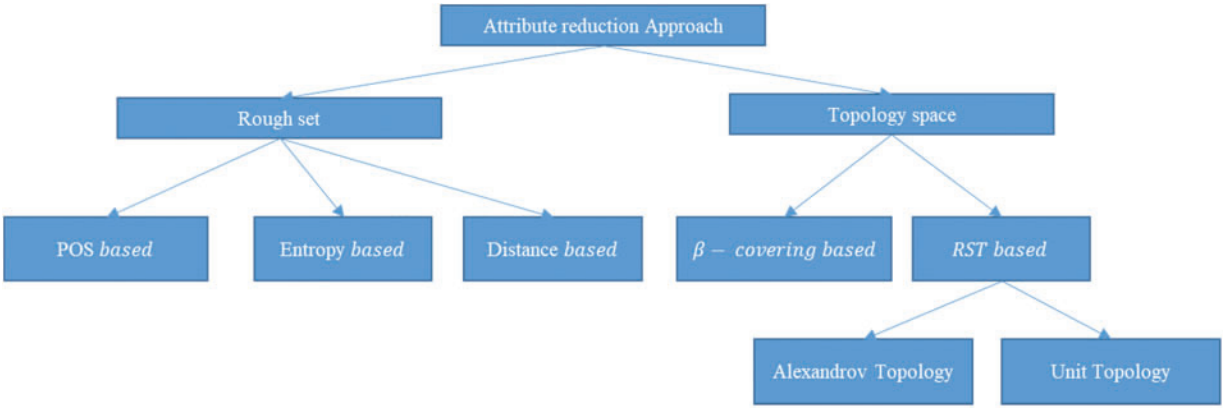

Figure 1: Attribute reduction approaches

Apart from the research on the relationship between rough set theory models and topologies, the methods used to build the Alexandrov topology structure regarding the rough set approach are also concerned and developed, including Alexandrov topologies based on the fuzzy rough set approach [33] and the intuitionistic fuzzy rough set approach [35].

Current attribute reduction methods based on algebraic topology approaches do not consider the separability of the attribute space. When separating attributes into non-intersecting groups, it will greatly reduce the number of candidate attribute sets that need to be processed. In this paper, we propose a Hausdorff topology structure using the RST approach to construct a novel attribute reduction method for decision tables. Unlike the structure of Alexandrov topologies, each element in a Hausdorff topology always has an opposite element, i.e., two different objects can be always

distinguished through their neighbors. Then, we can use the Hausdorff topological structure as the standard structure for selecting the attributes. Furthermore, in 2014, Yun et al. [38] showed that two covers of attributes are different, but their topologies can be the same. Then, we can group attributes with the same topological structure. Based on those observations, a novel attribute reduction model, using the Hausdorff topology approach, is introduced. This model has the following main steps:

(1) Selecting: Filter attributes related to the reduct based on the criteria of Hausdorff's proposed topological structure.

(2) Partitioning: Group the filtered attributes into groups based on the proposed co-structure concept. It should be noted that these groups of attributes do not intersect. Therefore, the number of groups will not be larger than the number of attributes obtained at the selection step.

(3) Wrapping: Wrapper the attribute groups obtained from the partitioning step to select the attribute group with the highest accuracy and consider this group as the reduct.

The main contributions of this paper include:

(1) Propose the Hausdorff topological structure based on the RST approach on the attribute space.

(2) Propose a novel attribute reduction method on topological space.

Apart from the introduction and conclusion sections, the paper's structure includes: Section 2 recalls some basic knowledge of topology and rough sets. Section 3 proposes the method to construct a topology based on beta fuzzy approximation space. Section 4 studies the separability properties of the Hausdorff topology. Section 5 presents the attribute reduction model according to the Hausdorff topology structure. Section 6 presents some experimental results. The discussions are presented in Section 7.

2 Preliminaries

This section recalls some basic knowledge of RS and topology [29,38] that is essential background knowledge will be used in the next sections of this study.

The numeric decision table (shown in Table 1) is represented by the tuple $DT = (U, C, D, f)$, where $C \cap D = \varnothing$, U is a non-empty set of objects, C is a non-empty set of conditional attributes, and D is a decision attribute. The function f determines a value for each $u \in U$ corresponding to $c \in C$ or D. For convenience, we denote $DT = (U, C, D)$.

Table 1: The numeric decision table

U	a	b	c	d	e	f	D
u1	1.0	0.4	0.8	0.2	1.0	0.0	0
u2	1.0	0.4	0.2	0.4	0.2	0.8	1
u3	0.8	0.6	1.0	0.0	0.6	0.4	0
u4	0.2	0.6	0.8	0.2	0.0	1.0	1
u5	0.2	0.8	0.8	0.2	0.0	1.0	1
u6	0.2	0.8	0.2	0.8	0.0	1.0	0

Definition 1. Given a decision table $DT = (U, C, D)$ and an equivalence relation R on U. The pair (U, R) is an approximation space. Then the upper and lower approximation operations of $X \subseteq U$ are

defined by:

$$\underline{R}(X) = \left\{ x \in U | [x]_R \subseteq X \right\} \tag{1}$$

$$\overline{R}(X) = \left\{ x \in U | [x]_R \cap X \neq \varnothing \right\} \tag{2}$$

where $[x]_R$ is an equivalence class of $x \in U$ according to the relation R or a subset of U that are the same with x according to the R. Then $\underline{R}(X)$ is the set of objects in U that belong to X, and $\overline{R}(X)$ is the set of objects in U that could belong to X.

Definition 2. The topological space is defined as a pair $(U, \mathcal{T})$ consisting of a family of U subsets labelled $\mathcal{T}$ that satisfy the following conditions:

(1) $\varnothing \in \mathcal{T}$ and $U \in \mathcal{T}$;
(2) $\mathcal{T}$ is closed under the finite intersection operation;
(3) $\mathcal{T}$ is closed under any union operation.

Definition 3. Let R be an equivalence relation on U. Then R is a fuzzy equivalence relation if:

(1) Reflexive: For all $x \in U$, $R(x, x) = 1$;
(2) Symmetric: For all $x, y \in U$, $R(x, y) = R(y, x)$;
(3) Min-transitive: For all $x, y, z \in U$, $\min(R(x, y), R(y, z)) \leq R(x, z)$.

Definition 4. Let R be a fuzzy equivalence relation on U, then R can be represented by the fuzzy equivalence relation matrix $\mathcal{M}$, where $\mathcal{M} = [i, j]_{|U| \times |U|}$.

Then, to get the lower approximation and the upper approximation of $X \subseteq U$, instead of using the approximation space (U, R) through the relation R, we can redefine the upper and lower approximation as follows:

$$\underline{\mathcal{M}}(X) = \{ x \in U | M[x] \subseteq X \} \tag{3}$$

$$\overline{\mathcal{M}}(X) = \{ x \in U | M[x] \cap X \neq \varnothing \} \tag{4}$$

Definition 5. Let $\mathcal{A}, \mathcal{B}$ be fuzzy equivalence relation matrices on U corresponding to $A, B \subseteq C$, then the intersection of two matrices is defined as follows:

$$\mathcal{A}[i, j] \cap \mathcal{B}[i, j] = min(\mathcal{A}[i, j], \mathcal{B}[i, j]), \forall i, j \in [1, |U|]. \tag{5}$$

Definition 6. Let $\mathcal{A}, \mathcal{B}$ be fuzzy equivalence relation matrices on U corresponding to $A, B \subseteq C$, then the union of two matrices is defined as follows:

$$\mathcal{A}[i, j] \cup \mathcal{B}[i, j] = max(\mathcal{A}[i, j], \mathcal{B}[i, j]), \forall i, j \in [1, |U|]. \tag{6}$$

Proposition 1. [38] *Let $\mathcal{M}$ be a relation matrix on U. For all $X, Y \subseteq U$*

(1) $\underline{\mathcal{M}}(X) = \left(\overline{\mathcal{M}}(X^c)\right)^c, \overline{\mathcal{M}}(X) = \left(\underline{\mathcal{M}}(X^c)\right)^c$ *where* $X^c = U \setminus X$
(2) $\underline{\mathcal{M}}(U) = U, \overline{\mathcal{M}}(\varnothing) = \varnothing$
(3) $\underline{\mathcal{M}}(X \cap Y) = \underline{\mathcal{M}}(X) \cap \underline{\mathcal{M}}(Y), \overline{\mathcal{M}}(X \cup Y) = \overline{\mathcal{M}}(X) \cup \overline{\mathcal{M}}(Y)$
(4) *If* $X \subseteq Y$ *then* $\underline{\mathcal{M}}(X) \subseteq \underline{\mathcal{M}}(Y), \overline{\mathcal{M}}(X) \subseteq \overline{\mathcal{M}}(Y)$
(5) $\underline{\mathcal{M}}(X \cup Y) \supseteq \underline{\mathcal{M}}(X) \cup \underline{\mathcal{M}}(Y), \overline{\mathcal{M}}(X \cap Y) \subseteq \overline{\mathcal{M}}(X) \cap \overline{\mathcal{M}}(Y)$

Proposition 2. [38] *Let $\mathcal{M}$ be a relation matrix on U. For all $X, Y \subseteq U$*

(1) *$\mathcal{M}$ is serial if for any $X \subseteq U$: $\underline{\mathcal{M}}(X) \subseteq \overline{\mathcal{M}}(X)$ or $\overline{\mathcal{M}}(U) = U$*
(2) *$\mathcal{M}$ is reflexive if for any $X \subseteq U$: $\underline{\mathcal{M}}(X) \subseteq X$ or $X \subseteq \overline{\mathcal{M}}(U)$*
(3) *$\mathcal{M}$ is symmetric if for any $X \subseteq U$: $X \subseteq \underline{\mathcal{M}}\left(\overline{\mathcal{M}}(X)\right)$ or $\overline{\mathcal{M}}\left(\underline{\mathcal{M}}(X)\right) \subseteq X$*
(4) *$\mathcal{M}$ is transitive if for any $X \subseteq U$: $\underline{\mathcal{M}}(X) \subseteq \underline{\mathcal{M}}\left(\underline{\mathcal{M}}(X)\right)$ or $X \subseteq U$: $\overline{\mathcal{M}}\left(\overline{\mathcal{M}}(X)\right) \subseteq \overline{\mathcal{M}}(X)$*

Proposition 3. [38] *Let $\mathcal{M}$ be a serial relation matrix on U. Then $\mathcal{T} = \left\{X \subseteq U | \underline{\mathcal{M}}(X) = \overline{\mathcal{M}}(X)\right\}$ is a topology on U.*

3 Topologies Induced by β Fuzzy Approximation

Based on the fundamental knowledge presented in Section 2, in this section, we will discuss how to construct the topology from the β fuzzy approximation space. On that basis, we construct the algebraic structure of the proposed topology. We first define the concepts of β fuzzy equivalence relation, β fuzzy approximation space as below:

Definition 7. Let U be a non-empty set of objects in the interval [0, 1], the β fuzzy equivalence relation on U denoted by R^β is defined as the following:

For all $x, y \in U$

$$R^\beta(x,y) = \begin{cases} 1 - |x-y| & if \ \ 1 - |x-y| \geq \beta \\ 0 \ if \ \ 1 - |x-y| < \beta. \end{cases} \tag{7}$$

Definition 8. Given $DT = (U, C, D)$, the β fuzzy approximation space is denoted by (U, R^β), where $R^\beta = \{R^\beta(x,y) | x, y \in U\}$.

For each $c \in C$, we have U different values. We denote the relation R^β with respect $c \in C$ by R_c^β.

Definition 9. Given the decision table $DT = (U, C, D)$ and $\mathcal{M}_p^\beta$, $\mathcal{M}_q^\beta$ are β fuzzy equivalence relation matrices corresponding to $R_p^\beta, R_q^\beta, p, q \in C$. Then:

(1) $\mathcal{M}_p^\beta$ is less than $(\prec)$ $\mathcal{M}_q^\beta$ if for all $x, y \in U$, $R_p^\beta(x,y) < R_q^\beta(x,y)$;
(2) $\mathcal{M}_p^\beta = \mathcal{M}_q^\beta$ if for all $x, y \in U$ then $R_p^\beta(x,y) = R_q^\beta(x,y)$.

Proposition 4. *Let (U, R^β) be a fuzzy approximation space. Then $\mathcal{T} = \left\{X \subseteq U : \underline{R^\beta}(X) = \overline{R^\beta}(X)\right\}$ is a topology on U with respect R^β.*

Proof:

(1) Based on the second property of Proposition 1, we have $\underline{R^\beta}(U) = U$ and $\overline{R^\beta}(\varnothing) = \varnothing$, then $\varnothing \in \mathcal{T}$ and $U \in \mathcal{T}$. This satisfies the first condition in Definition 2.
(2) Assume that $X, Y \in \mathcal{T}$, then $\underline{R^\beta}(X) = \overline{R^\beta}(X)$, and $\underline{R^\beta}(Y) = \overline{R^\beta}(Y)$, and more $\underline{R^\beta}(X \cap Y) \subseteq \overline{R^\beta}(X \cap Y)$ and $\overline{R^\beta}(X \cap Y) \subseteq \overline{R^\beta}(X) \cap \overline{R^\beta}(Y) = \underline{R^\beta}(X) \cap \underline{R^\beta}(Y) = \underline{R^\beta}(X \cap Y)$. Then $\overline{R^\beta}(X \cap Y) = \underline{R^\beta}(X \cap Y)$. It follows that $X \cap Y \in \mathcal{T}$. This satisfies the second condition in Definition 2.
(3) Assume that $X_k \in \mathcal{T} | k \in K$. Then $\underline{R^\beta}(X_k) = \overline{R^\beta}(X_k)$, so for all $k \in K$, we have $\overline{R^\beta}\left(\underset{k\in K}{\cup} X_k\right) = \underset{k\in K}{\cup} \overline{R^\beta}(X_k) = \underset{k\in K}{\cup} \underline{R^\beta}(X_k) \subseteq \underline{R^\beta}\left(\underset{k\in K}{\cup} X_k\right)$. In more $\underline{R^\beta}\left(\underset{k\in K}{\cup} X_k\right) \subseteq \overline{R^\beta}\left(\underset{k\in K}{\cup} X_k\right)$ so $\underline{R^\beta}\left(\underset{k\in K}{\cup} X_k\right) = \overline{R^\beta}\left(\underset{k\in K}{\cup} X_k\right)$. This satisfies the third condition in Definition 2.

From (1)–(3), we can conclude that $\mathcal{T}$ is a topology on U with respect R^β. □

Definition 10. Let $DT = (U, C, D)$ and R_a^β be a β fuzzy equivalence relation on U with respect $a \in C$. Then topology of the attribute $a \in C$ denoted by $\mathcal{T}_a$ is defined by:

$$\mathcal{T}_a = \left\{ X \subseteq U | \underline{R_a^\beta}(X) = \overline{R_a^\beta}(X) \right\}. \tag{8}$$

Definition 11. Let $\mathcal{T}_p, \mathcal{T}_q$ be two topologies with respect R_p^β, R_q^β. Then $\mathcal{T}_p \subseteq \mathcal{T}_q$ if for all $e \in \mathcal{T}_p$ then $e \in \mathcal{T}_q$.

Proposition 5. *Let* $\mathcal{T}_p = \left\{ X \subseteq U | \underline{R_p^\beta}(X) = \overline{R_p^\beta}(X) \right\}$ *and* $\mathcal{T}_q = \left\{ X \subseteq U | \underline{R_q^\beta}(X) = \overline{R_q^\beta}(X) \right\}$ *be two topologies on* U. *Then* $\mathcal{T}_p \subseteq \mathcal{T}_q$ *if* $R_q^\beta \prec \overline{R_p^\beta}$.

Proof:

Assume that $R_p^\beta \prec R_q^\beta$, so for all $x \in U$ we have $[x]_p^\beta \subseteq [x]_q^\beta$. Then for all $X \subseteq U$, if $[x]_q^\beta \subseteq X$ then $[x]_p^\beta \subseteq X$, or in other words if $X \in T_p$ then $X \in T_q$. Based on Definition 11, the Proposition has been proved. □

Definition 12. Let R^β be a β fuzzy equivalence relation on U, then R^β is the smoothest if and only if for all $x, y \in U$, $R^\beta(x, y) = 1$ if $x = y$ and $R^\beta(x, y) = 0$ if $x \neq y$. We denote the smoothest β fuzzy equivalence relation as R_1^β.

Proposition 6. $\mathcal{T}$ *is the largest topology* $(\mathcal{T}_1)$ *if* $R^\beta = R_1^\beta$.

Proof:

It is clear that the topology defined on R_1^β is the family $S \subseteq U$ such that $[x]_1^\beta \subseteq S$. Furthermore, since $[x]_1^\beta$ is the non-subset smallest of U then for all $x \in U$, the number of subset $S \subseteq U$ such as $[x]_1^\beta \subseteq S$ are $\left|\left\{2^{(U-x)}\right\} \cup \{x\}\right| = \left|\left\{2^{(U-x)}\right\}\right| = 2^{|U-1|}$, that is the family of subset largest. The Proposition has been proved. □

Definition 13. Let R^β be a β fuzzy equivalence relation on U, then R^β is said to be coarsest relation if and only if for all $x, y \in U$, $R^\beta(x, y) = 1$. We denote the coarsest β fuzzy equivalence relation as R_0^β.

Proposition 7. $\mathcal{T}$ *is the smallest topology denoted by* $\mathcal{T}_0$ *if* $R^\beta = R_0^\beta$.

Proof:

It is clear that the topology defined on R_0^β is the family $S \subseteq U$ such that $[x]_0^\beta \subseteq S$. Furthermore, since $[x]_0^\beta = U$ is the non-subset largest of U then for all $x \in U$, the number of subset $S \subseteq U$ such as $[x]_0^\beta \subseteq S$ are $\left|\left\{2^{(U-x)}\right\} \cup \{U\}\right| = |\{U\}| = 1$, that is the family of subset *smallest*. The Proposition has been proved. □

Definition 14. Let $DT = (U, C, D)$ and $R_{\{p,q\}}^\beta$ be a β fuzzy equivalence relation of $p, q \in C$ on U, $R_{\{p,q\}}^\beta = R_p^\beta \cap R_q^\beta$. Then topology on $R_{\{p,q\}}^\beta$ is defined by:

$$\mathcal{T}_{\{p,q\}} = \left\{ X \subseteq U | \underline{R_{\{p,q\}}^\beta}(X) = \overline{R_{\{p,q\}}^\beta}(X) \right\} \tag{9}$$

4 Hausdorff Topology from Topology Space

Based on the topological structure proposed in Section 3 of the article. In this section, we study the separability property of the Hausdorff topology. On that basis, we also examine the relationships of objects in U to determine in which cases $\mathcal{T}$ is a Hausdorff topology on U.

Corollary 1. *Given a topology* $\mathcal{T}$ *base on Proposition 4. If* $X \in \mathcal{T}$, *then* $\exists Y \in \mathcal{T}: Y = U \setminus X$.

Definition 14. Given a topology $\mathcal{T}$ with respect to R^β. Then the topology $\mathcal{T}$ is said to be Hausdorff topology if every $x \neq y \in U$ always exists two open sets $V_x, V_y \in \mathcal{T}$ such that $V_x \cap V_y = \varnothing$.

Proposition 8. *Given a topology $\mathcal{T}$ base on Proposition 4. If $\mathcal{T}$ is non-trivial topology then $\mathcal{T}$ is a Hausdorff topology.*

Proof:

Assume that $\mathcal{T}$ is non-trivial topology, which mean that $\mathcal{T}$ is not an empty set. Then, based on Corollary 1, if $X \in \mathcal{T}$ then there always exists Y such that $X \cap Y = \varnothing$. The proposition is proven. □

Thus, $\mathcal{T}$ is called Hausdorff topology if $\mathcal{T}$ is not empty in relation R^{β}. The next question is how to make a relation R^{β}, so that $\mathcal{T}$ is not empty.

Definition 15. Let $DT = (U, C, D)$ and R^{β} be a β fuzzy equivalence relation on U. Then topology $\mathcal{T}$ with respect R^{β} is non-trivial topology if for all $x \in U$, exists $y \neq x \in U$ such that $[x]_{R^{\beta}} \cap [y]_{R^{\beta}} = \varnothing$.

Proposition 9. *Let R^{β} be a β fuzzy equivalence relation on U. If $x \neq y \in U$ are two the smallest elements and $|[x]_{R^{\beta}} \cap [y]_{R^{\beta}} = \varnothing$ then topology $\mathcal{T}$ with respect R^{β} is non-trivial topology.*

Proof:

(1) Suppose $x > y$ and $[x]_{R^{\beta}} \cap [y]_{R^{\beta}} = \varnothing$, then according to Eq. (7) we have $R^{\beta}(x, y) = 0$ or $1 - |x - y| < \beta$.
(2) For all $z \in U - \{x, y\}$, if $z > x$ then $z > y$, so $1 - |z - y| < \beta$ then $R^{\beta}(z, y) = 0$.

From (1) and (2), we have $\forall z \in U : z > x, [z]_{R^{\beta}} \cap [y]_{R^{\beta}} = \varnothing$. The Proposition has been proved. □

5 Attribute Reduction Based on Hausdorff Topology Space

Based on the topological structure proposed by Hausdorff, in this section, we propose an attribute reduction model according to the Hausdroff topological approach. Differing from current attribute reduction models, the proposed model includes three stages. In the first stage, we remove redundant attributes based on the definition of redundant attributes in the Hausdorff topological approach. In this stage, only properties with Hausdorff topological structure (based on Lemma 1 and Definition 16) are selected. In the second stage, we group condition attributes with the same structure as the decision attribute D based on the concept of D-homomorphism (based on Definition 16 and Definition 17). The third stage is used to choose the best attribute group according to the method of wrapping the attribute groups. To illustrate these stages, the evaluation of the proposed algorithm and illustrative numerical examples are presented. Table 2 details the differences between the Hausdorff topology attribute reduction method and the traditional approach.

Table 2: Comparing attribute reduction approaches

	Hausdorff topology approach	Traditional approach
Attribute search method	Consider only attributes with Hausdorff topological structure	Consider all possible attributes
Attribute evaluation method	Group condition attributes that have the same structure as the decision attribute	Use dependency measures to construct attribute importance measures
Method of determining the reduct	Use the wrapper method to determine the subset with the best classification accuracy	Use the conservative property of the measure to complete the attribute selection process

Lemma 1. *Let PD be a partition of the decision attribute D and $\mathcal{T}$ be a Hausdorff topology. Then for every $X \in PD$, there always exists $Y \in \mathcal{T}$ such that $X \cap Y \neq \varnothing$.*

Proof:

Based on Corollary 1 we have $\mathcal{T} \neq \varnothing$. For all $X \in PD$, suppose that $\exists G \in \mathcal{T}: G \cap X = \varnothing$, based on Corollary 1 $\exists \sim G = U \setminus G \in \mathcal{T}: \sim G \cap X \neq \varnothing$. □

Definition 16. Given a decision table $DT = (U, C, D)$. Then $c \in C$ is called a relative attribute if the topology of c on U is a Hausdorff topology.

Therefore, the attributes that have a Hausdorff topological structure are all attributes that control the decision set D, which is always contained in candidate reducts and is often called relative attributes. The above property can help us deliver an attribute filtering step that is effected in the attribute reduction model with a much more optimal computational cost than traditional attribute filtering methods. Next, to wrap relative attributes effectively, we define some concepts to optimize candidate reduct sets.

Definition 17. Given a decision table $DT = (U, C, D)$ and H^β is a family Hausdorff topology of C base on R^β. Then $p \in H^\beta$ is called homomorphic with $q \in H^\beta$ if $R_p^\beta = R_q^\beta$.

Definition 18. Given a decision table $DT = (U, C, D)$ and H^β is a family Hausdorff topology of C base on R^β. Then $p \in H^\beta$ is called D-homomorphic with $q \in H^\beta$ if $R_{pD}^\beta = R_{qD}^\beta$. Where $R_{pD}^\beta = R_p^\beta \cap R_D^\beta$, $R_{qD}^\beta = R_q^\beta \cap R_D^\beta$.

Based on Definitions 16, 17, and 18, we propose an attribute reduction model using the Hausdorff topology approach. The proposed model includes three main independent phases. 1) Filter phase for relative attributes, 2) Phase grouping relative attributes, and 3) Phase wrapper groups of relative attributes.

Algorithm 1: Cluster filter wrapper based on hausdorff topology (CFW)

Input Decision Table $DT = (U, C, D)$, $\Delta = \{0.1, 0.2, \ldots, 0.8, 0.9\}$ and classification *Model*
Output: The reduct *Red*
Begin

```
Red ← ∅;
for all β ∈ Δ do
    H^β ← ∅;                                   // Init Hausdorff attributes
    CH^β ← ∅;                                  // Init Hausdorff attribute groups
    Red^β ← ∅;                                 // Init a reduct at beta
    for all c ∈ C do
        T_c = {X ⊆ U | R_c^β(X) = closure(R_c^β)(X)};
        if T_c ≠ ∅ then
            H^β ← H^β ∪ {c};
        end if
    end for
    for all p ∈ {H^β − CH^β} do
        G_p ← ∅;                               //Init Hausdorff attributs group
        for all q ∈ {H^β − CH^β − {p}} do
```

(Continued)

Algorithm 1 (continued)

15. **if** $R_p^{\beta} \cap \mathcal{R}_D = R_q^{\beta} \cap \mathcal{R}_D$ **then**
16. $G_p \leftarrow G_p \cup \{q\}$; // Grouping relative attributes
17. **end if**
18. **end for**
19. $CH^{\beta} \leftarrow CH^{\beta} \cup G_p$;
20. $Red^{\beta} = G : Accuracy\,(Model, G) = \max\limits_{G \in CH^{\beta}} \{Accuracy\,(Model, G)\}$;
21. **end for**
22. **if** Accuracy(Model, R_{β}) > Accuracy(Model, R) **then**
23. $Red \leftarrow Red^{\beta}$; //Wrapper β Groups
24. **end if**
25. **end for**
26. **return** Red

End.

In this algorithm, we denote H^{β} as Hausdorff attributes selected from the original attribute set C at β threshold. As mentioned, H^{β} is a relative reduct set. CH^{β} are groups of attributes clustered from H^{β}, where each item G_p in CH^{β} is a group of homomorphic attributes with attribute p. Therefore, the Hausdorff attribute wrapper phase is used to find the best reduct candidate at each β threshold. In reality, the Hausdorff attributes may not exist. Therefore, we use the threshold β in the threshold set Δ for adjustment. Then, for each value β, a different candidate reduct Red^{β} can exist. Therefore, the β wrapper stage determines the best candidate reduct for R. Next, we evaluate the complexity of the Algorithm 1 (CFW algorithm). Let $|U|$ be the number of objects, $|C|$ be the number of condition attributes, $|H^{\beta}|$ is the number of relative attributes, and $|CH^{\beta}|$ is the number of groups of relative attributes. Then the complexity of 1) Steps 6–11 is $\mathcal{O}\,(2|U||C|)$; 2) Steps 14–18 is $\mathcal{O}\left(|U|^2|H^{\beta}|\right)$. Let $\mathbb{T}$ is the execution time of the *Model* classification, the complexity of Steps 20–22 is $\mathbb{T}$. Then the complexity of Steps 12–21 is $\mathcal{O}\left(|U|^2|H^{\beta}|\right) + \mathcal{O}\,(|CH^{\beta}|\mathbb{T})$. With $|\Delta|$ is very small, from 1) and 2), we have the complexity of the algorithm is $\mathcal{O}\,(2|U||C|) + \mathcal{O}\left(|U|^2|H^{\beta}|\right) + \mathcal{O}\,(|CH^{\beta}|\mathbb{T})$. The details of CFW algorithm are presented in Fig. 2 below.

Fig. 2 shows the progress of the CFW algorithm. There are three stages in this algorithm. The first stage is performed to filter Hausdorff attributes H^{β} from the attribute set C in decision table DT. In the second stage, a group of candidates CH^{β} is generated. In the last stage, the wrapper method is used in order to define Red^{β} where Red^{β} is the reduct with the highest accuracy among the candidates in CH^{β}. CFW algorithm stops when all values of beta in Δ are taken and the final reduct Red is found.

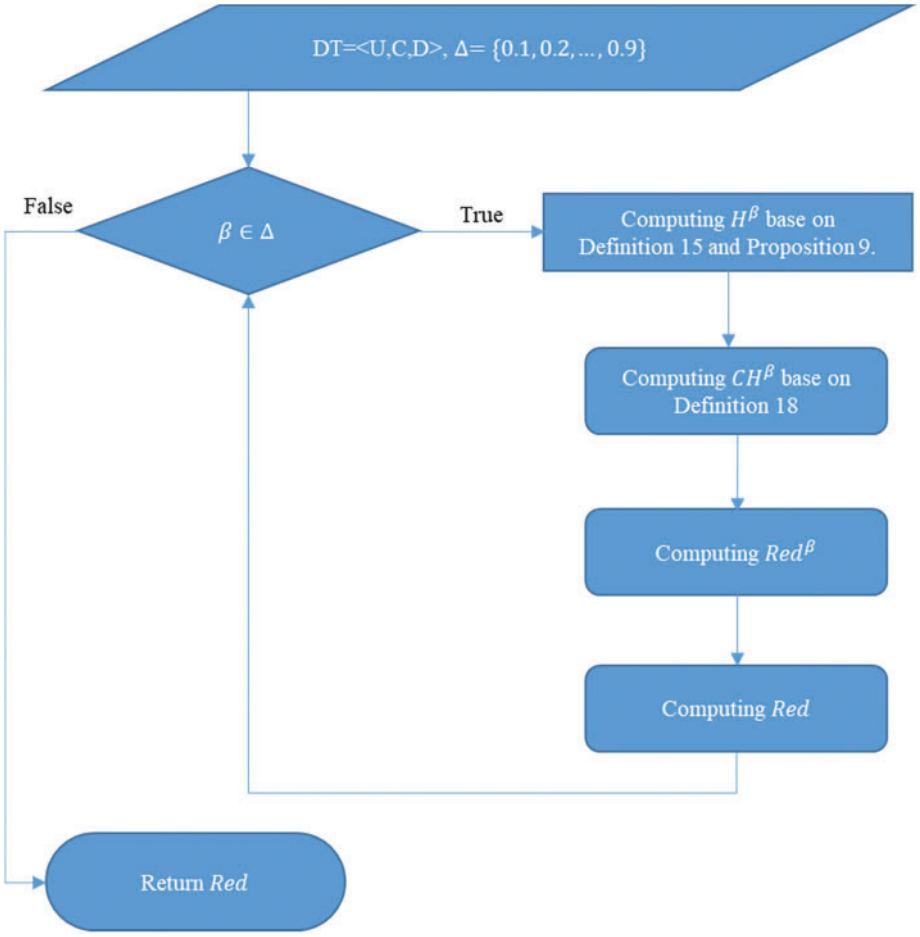

Figure 2: The diagram of the CFW algorithm

Example

Let $DT = (U, C, D)$ as Table 1 and $\beta = 0.7$ we have:

- Initialization steps: $Red \leftarrow \varnothing$; $H^{\beta} \leftarrow \varnothing$; $CH^{\beta} \leftarrow \varnothing$; $Red^{\beta} \leftarrow \varnothing$;
- Filtering relative attributes:

1) The attribute $a \in C$ has $u_3 = 0.6 \neq u_4 = 0.0$ and $\begin{matrix}[u_3]\, R_a^{0.7} = [0.8 & 0.8 & 1 & 0 & 0 & 0] \\ [u_4]\, R_a^{0.7} = [0 & 0 & 0 & 1 & 1 & 1]\end{matrix}$

2) The attribute $b \in C$ has $u_1 = 0.2 \neq u_3 = 0.4$ and $\begin{matrix}[u_1]\, R_b^{0.7} = [1 & 1 & 0.8 & 0.8 & 0 & 0] \\ [u_3]\, R_b^{0.7} = [0.8 & 0.8 & 1 & 1 & 0.8 & 0.8]\end{matrix}$

3) The attribute $c \in C$ has $u_1 = 0.6 \neq u_2 = 0.0$ and $\begin{matrix}[u_1]\, R_c^{0.7} = [1 & 0 & 0.8 & 1 & 1 & 0] \\ [u_2]\, R_c^{0.7} = [0 & 1 & 0 & 0 & 0 & 1]\end{matrix}$

4) The attribute $d \in C$ has $u_1 = 0.4 \neq u_3 = 0.2$ and $\begin{matrix}[u_1]\, R_d^{0.7} = [1 & 0.8 & 0.8 & 1 & 1 & 0] \\ [u_3]\, R_d^{0.7} = [0.8 & 0 & 1 & 0.8 & 0.8 & 0]\end{matrix}$

5) The attribute $e \in C$ has $u_2 = 0.2 \neq u_4 = 0.0$ and $\begin{matrix}[u_2]\, R_e^{0.7} = [0 & 1 & 0 & 0.8 & 0.8 & 0.8] \\ [u_4]\, R_e^{0.7} = [0 & 0.8 & 0 & 1 & 1 & 1]\end{matrix}$

6) The attribute $f \in C$ has $u_2 = 0.2 \neq u_4 = 0.0$ and $\begin{matrix}[u_1]\, R_f^{0.7} = [1 & 0 & 0 & 0 & 0 & 0] \\ [u_3]\, R_f^{0.7} = [0 & 0 & 1 & 0 & 0 & 0]\end{matrix}$

7) Base one the Definition 15 and Proposition 9 show that, only $\{a, c, f\}$ attributes are topology Hausdorff, so $H^\beta = \{a, c, f\}$

– Group relative attributes:

1) Calculating the relational matrices for attributes in H^β according to Eq. (7) we have:

$$R_a^{0.7} = \begin{bmatrix} 1 & 1 & 0.8 & 0 & 0 & 0 \\ 1 & 1 & 0.8 & 0 & 0 & 0 \\ 0.8 & 0.8 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix};$$

$$R_c^{0.7} = \begin{bmatrix} 1 & 0 & 0.8 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0.8 & 0 & 1 & 0.8 & 0.8 & 0 \\ 1 & 0 & 0.8 & 1 & 1 & 0 \\ 1 & 0 & 0.8 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix};$$

$$R_f^{0.7} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0.8 & 0.8 & 0.8 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 1 & 1 & 1 \\ 0 & 0.8 & 0 & 1 & 1 & 1 \\ 0 & 0.8 & 0 & 1 & 1 & 1 \end{bmatrix};$$

$$R_D = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}.$$

2) According to Definition 18 we have:

Considering attribute a, we see $R_a^{0.7} \cap R_D = R_c^{0.7} \cap R_D = \begin{bmatrix} 1 & 0 & 0.8 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0.8 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

Then $G_a = \{a, c\}$, so $CH^\beta = CH^\beta \cup G_a = \{\{a, c\}\}$.

$G_f = \{f\}$, so $CH = CH \cup G_f = \{\{a, c\}, \{f\}\}$; Assuming the attribute group $\{a, c\}$ has the highest classification accuracy on *Model*, we get $Red^{0.7} = \{a, c\}$. Suppose the candidate reduct at value $\beta = 0.7$ is the candidate with the highest classification accuracy on *Model*. Then we obtain the reduct $Red = Red^{0.7} = \{a, c\}$.

The Hausdorff topology does not always exist for every fuzzy approximation space. Therefore, the beta value is used to change the smoothness of the fuzzy partitions. According to Propositions 5 and 6, we can see the relationship between the topology structure and the smoothness of the fuzzy partition. The smoother the fuzzy partition, the larger the size of the topology, which means the possibility of the Hausdorff topology appearing is greater. In the experiment, we set the step value to 0.1 to clarify the difference after each change in beta value.

6 Experiments

6.1 Experimental Scenarios

The goal to evaluate the algorithm's performance when applied to real datasets and to be the basis for affirming the theoretical foundation we propose is entirely reasonable. The following is the experimental plan for the proposed algorithm. Step 1: Choose the best β value for the algorithm on each dataset. This experimental process aims to choose the most suitable β value for each dataset, where β is in the [0, 1] range with each step of 0.1. Step 2: Compare the proposed algorithm with other attribute reduction algorithms. After choosing the best beta value of the proposed algorithm for each dataset, compare the proposed algorithm with some attribute reduction algorithms: (1) Attribute reduction algorithm according to the rough set (VPRS) approach [41]; (2) Attribute reduction algorithm according to the fuzzy rough set (FRS) approach [42]; (3) Attribute reduction algorithm according to the fuzzy information entropy (IFE) approach [21]; (4) Attribute reduction algorithm according to the fuzzy distance (FD) approach [24].

Each algorithm experiments on 90% of randomly selected data from each dataset for evaluation. Perform this evaluation 10 times with Support Vector Machine classification (SVM) and k-Nearest Neighbors models (kNN, $k = |D|$) for finding the best reduct. Each model uses precision measurement and the 10-fold cross-validation method. The result of evaluating the classification accuracy of the reduct is the average accuracy (expressed as a percentage) of 10 different times. Before performing the attribute reduction algorithm, we standard all datasets to the range [0, 1]. We use Python for all algorithms on the Anaconda 3.6 platform running on a Windows 10 operating system with an i5 Processor and 8 GB of RAM (Random Access Memory). All algorithms were performed on 16 datasets from UCI [43].

All the datasets are numeric data types. The details of the experimental data sets are in Table 3, with $|U|$ as a sample number and $|C|$ as the number of condition attributes arranged in ascending order. $|D|$ as the class number of decision attributes. Especially in the **UFDC** and **Sonar** datasets that contain noise data, the model's classification accuracy is low when used to train the model. Criteria to evaluate the effectiveness of algorithms include the computation time (in seconds), the size of the reduct (number of attributes), and classification accuracy (in percentage). Besides, the ability of the algorithms to handle noise is also considered.

Table 3: Dataset description

ID	Data	Describe	$\|U\|$	$\|C\|$	$\|D\|$
1	Wine	Wine	178	13	3
2	Heart	Statlog (Heart)	270	13	2
3	CMSC	Climate model simulation crashes data set	540	18	2
4	PDS	Parkinsons data set	196	22	2

(Continued)

Table 3 (continued)

ID	Data	Describe	$\|U\|$	$\|C\|$	$\|D\|$
5	BCWD	Breast cancer wisconsin (Diagnostic)	569	30	2
6	BCWP	Breast cancer wisconsin (Prognostic)	198	33	2
7	IS	Ionosphere	351	34	2
8	UFDC	Ultrasonic flowmeter diagnostics (C)	181	43	4
9	UFDD	Ultrasonic flowmeter diagnostics (D)	181	43	4
10	SHDC	SPECTF heart data set	267	44	2
11	UFDB	Ultrasonic flowmeter diagnostics (B)	92	51	3
12	DPDS	Divorce predictors data set	170	54	2
13	Sona	Connectionist bench	208	60	2
14	Musk	Musk	476	166	2
15	VRB	Voice rehabilitation (Binary)	126	310	2
16	VRG	Voice rehabilitation (Gender)	126	310	2

6.2 Analysis the β Value for the Algorithm on Each Dataset

Before comparing the reduct with other algorithms, we need to find the β value such that the CFW algorithm gives the best reduct. For each experimental dataset, we perform the algorithm with different β values. We can get the best result for each β value in terms of size or classification accuracy. In this study, we choose the option that balances classification accuracy and the size of the reduct. The reduct obtained with the algorithms may differ for each classification model. Moreover, each β value may exist or not the reduct and these reducts have the size and classification accuracy different.

In Fig. 3, we can see the best β value for each dataset on the SVM classification model. Fig. 4 shows the best β value for each dataset on the kNN classification model. The results show that there is no linear relationship between the size and classification accuracy of the reduct with β, there is no linear relationship between the size and classification accuracy of the reduct.

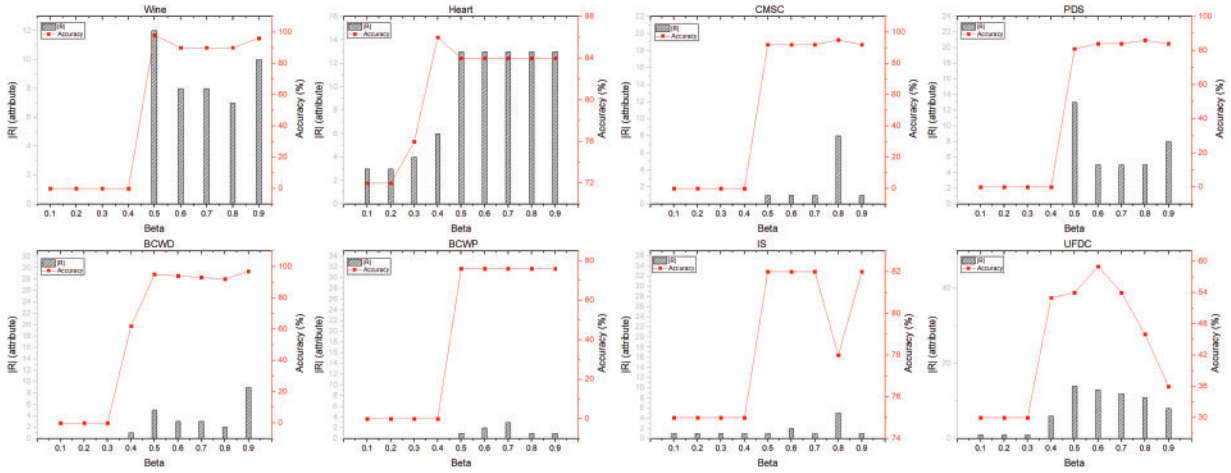

Figure 3: (Continued)

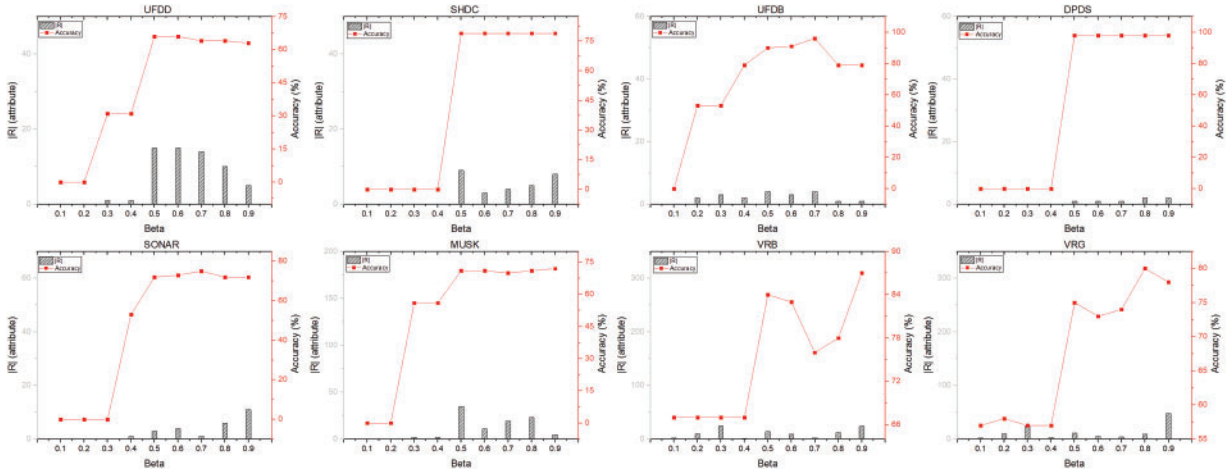

Figure 3: The chart analyzes the relationship between the size and classification accuracy of the reduct at each specified value of β on the SVM classification model

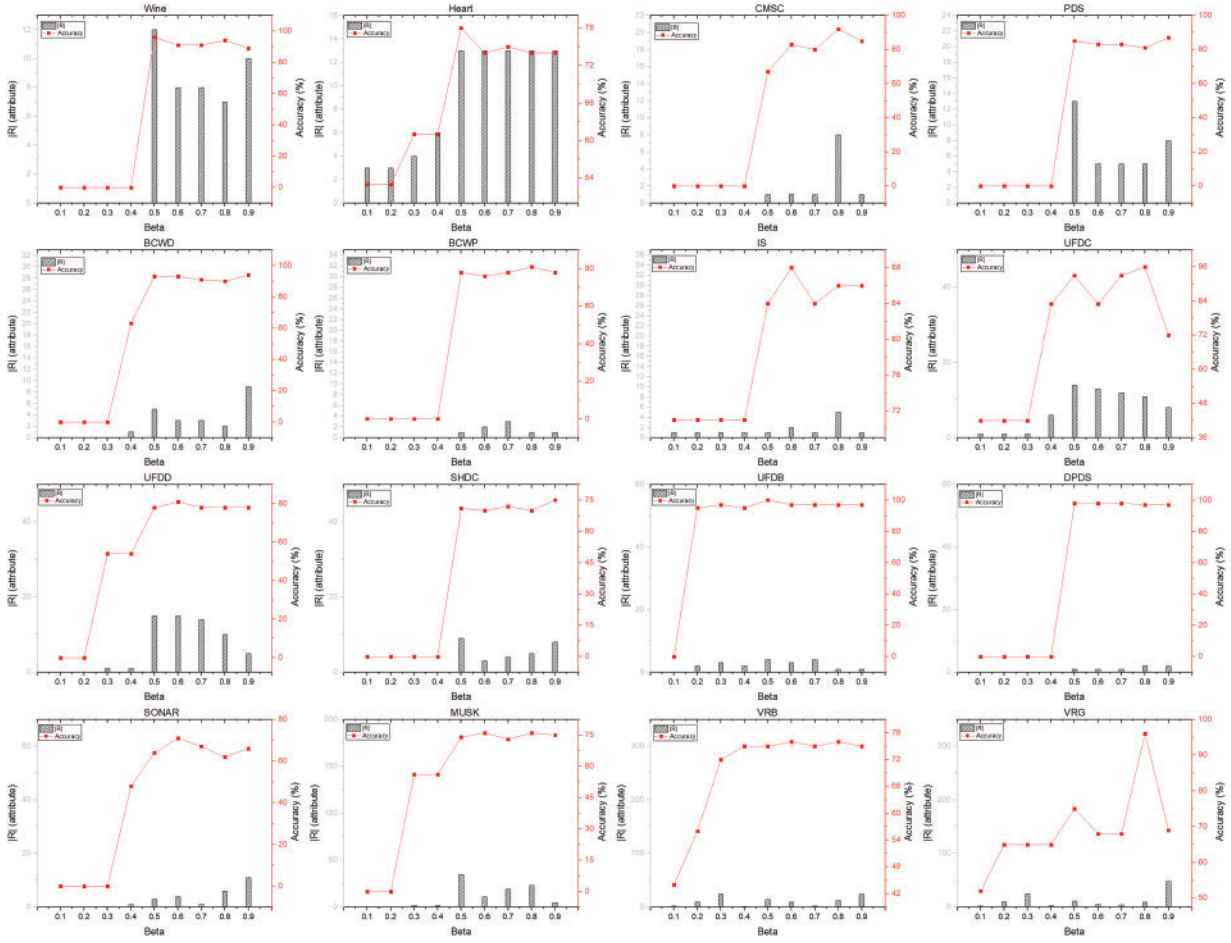

Figure 4: The chart analyzes the relationship between the size and classification accuracy of the reduct at each specified value of β on the kNN classification model

Tables 4 and 5 describe the effects of the beta parameter on the size of the reduction, the classification accuracy and the running time of the CFW algorithm.

Table 4: Analysis beta on SVM classification model

Dataset	Analysis	Beta								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Wine	Reduct size (|R|)	0	0	0	0	12	8	8	7	10
	Accuracy (%)	0	0	0	0	98	90	90	90	96
	Computing time (s)	0	0	0	0	0.05	0.05	0.05	0.05	0.05
Heart	|R|	3	3	4	6	13	13	13	13	13
	Accuracy (%)	72	72	76	86	84	84	84	84	84
	Computing time (s)	0.05	0.05	0.05	0.05	0.08	0.08	0.1	0.1	0.1
CSMC	Reduct size (|R|)	0	0	0	0	1	1	1	8	1
	Accuracy (%)	0	0	0	0	92	92	92	95	92
	Computing time (s)	0	0	0	0	0.28	0.28	0.38	0.4	0.42
PDS	|R|	0	0	0	0	13	5	5	5	8
	Accuracy (%)	0	0	0	0	81	84	84	86	84
	Computing time (s)	0	0	0	0	0.17	0.2	0.27	0.27	0.27
BCWD	Reduct size (|R|)	0	0	0	1	5	3	3	2	9
	Accuracy (%)	0	0	0	62	95	94	93	92	97
	Computing time (s)	0	0	0	0.67	0.67	0.7	0.77	0.77	0.77
BCWP	Reduct size (|R|)	0	0	0	0	1	2	3	1	1
	Accuracy (%)	0	0	0	0	76	76	76	76	76
	Computing time (s)	0	0	0	0	0.71	0.71	0.71	0.81	0.81
IS	Reduct size (|R|)	1	1	1	1	1	2	1	5	1
	Accuracy (%)	75	75	75	75	82	82	82	78	82
	Computing time (s)	0.2	0.2	0.2	0.2	0.6	0.6	0.7	0.73	0.73
UFDC	Reduct size (|R|)	1	1	1	6	14	13	12	11	8
	Accuracy (%)	30	30	30	53	54	59	54	46	36
	Computing time (s)	0.05	0.05	0.05	0.05	0.11	0.11	0.22	0.22	0.22
UFDD	Reduct size (|R|)	0	0	1	1	15	15	14	10	5
	Accuracy (%)	0	0	31	31	66	66	64	64	63
	Computing time (s)	0	0	0.07	0.07	0.18	0.18	0.2	0.3	0.3
SHDC	Reduct size (|R|)	0	0	0	0	9	3	4	5	8
	Accuracy (%)	0	0	0	0	79	79	79	79	79
	Computing time (s)	0	0	0	0	0.45	0.45	0.55	0.55	0.55
UFDB	Reduct size (|R|)	0	2	3	2	4	3	4	1	1
	ACC	0	53	53	79	90	91	96	79	79
	Computing time (s)	0	0.2	0.2	0.2	0.56	0.56	0.66	0.7	0.7
DPDS	Reduct size (|R|)	0	0	0	0	1	1	1	2	2
	Accuracy (%)	0	0	0	0	98	98	98	98	98
	Computing time (s)	0	0	0	0	0.82	0.82	0.9	09	0.9

(Continued)

Table 4 (continued)

Dataset	Analysis	Beta								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Sonar	Reduct size (|R|)	0	0	0	1	3	4	1	6	11
	Accuracy (%)	0	0	0	53	72	73	75	72	72
	Computing time (s)	0	0	0	0.63	0.63	0.63	0.75	0.75	0.75
Musk	Reduct size (|R|)	0	0	2	2	35	11	19	23	5
	Accuracy (%)	0	0	56	56	71	71	70	71	72
	Computing time (s)	0	0	1.15	1.15	2.26	2.26	3.2	3.2	3.3
VRB	Reduct size (|R|)	2	10	25	1	14	9	2	12	24
	Accuracy (%)	67	67	67	67	84	83	76	78	87
	Computing time (s)	0.85	0.85	0.85	0.85	0.85	1.7	2.2	2.2	2.2
VRG	Reduct size (|R|)	2	10	25	3	11	5	4	9	48
	Accuracy (%)	57	58	57	57	75	73	74	80	78
	Computing time (s)	0.92	0.92	0.92	0.92	0.92	1.93	2.3	2.3	2.3

Table 5: Analysis beta on kNN classification model

Dataset	Analysis	Beta								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Wine	Reduct size (|R|)	0	0	0	0	12	8	8	7	10
	Accuracy (%)	0	0	0	0	96	91	91	94	89
	Computing time (s)	0	0	0	0	0.07	0.07	0.07	0.07	0.07
Heart	|R|	3	3	4	6	13	13	13	13	13
	Accuracy (%)	53	53	61	61	78	74	75	74	74
	Computing time (s)	0.05	0.05	0.05	0.05	0.1	0.1	0.1	0.1	0.1
CSMC	Reduct size (|R|)	0	0	0	0	1	1	1	8	1
	Accuracy (%)	0	0	0	0	67	83	80	92	85
	Time	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26
PDS	|R|	0	0	0	0	13	5	5	5	8
	Accuracy (%)	0	0	0	0	85	83	83	81	87
	Computing time (s)	0	0	0	0	0.09	0.09	0.09	0.09	0.09
BCWD	Reduct size (|R|)	0	0	0	1	5	3	3	2	9
	Accuracy (%)	0	0	0	63	93	93	91	90	94
	Time	0	0	0	0.49	0.49	0.49	0.49	0.49	0.49
BCWP	Reduct Size (|R|)	0	0	0	0	1	2	3	1	1
	Accuracy (%)	0	0	0	0	78	76	78	81	78
	Computing time (s)	0	0	0	0	0.57	0.57	0.57	0.57	0.57
IS	Reduct size (|R|)	1	1	1	1	1	2	1	5	1
	Accuracy (%)	71	71	71	71	84	88	84	86	86
	Computing time (s)	0.21	0.21	0.21	0.21	0.21	0.43	0.43	0.43	0.43

(Continued)

Table 5 (continued)

Dataset	Analysis	Beta								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
UFDC	Reduct size (|R|)	1	1	1	6	14	13	12	11	8
	Accuracy (%)	42	42	42	83	93	83	93	96	72
	Computing time (s)	0.12	0.12	0.12	0.12	0.49	0.49	0.49	0.49	0.49
UFDD	Reduct size (|R|)	0	0	1	1	15	15	14	10	5
	Accuracy (%)	0	0	54	54	78	81	78	78	78
	Computing time (s)	0	0	0.2	0.2	0.62	0.62	0.62	0.62	0.62
SHDC	Reduct size (|R|)	0	0	0	0	9	3	4	5	8
	Accuracy (%)	0	0	0	0	71	70	72	70	75
	Computing time (s)	0	0	0	0	0.85	0.85	0.85	0.85	0.85
UFDB	Reduct size (|R|)	0	2	3	2	4	3	4	1	1
	ACC	0	95	97	95	100	97	97	97	97
	Computing time (s)	0	0.25	0.25	0.25	0.25	0.77	0.77	0.77	0.77
DPDS	Reduct size (|R|)	0	0	0	0	1	1	1	2	2
	Accuracy (%)	0	0	0	0	98	98	98	97	97
	Computing time (s)	0	0	0	0	1.06	1.06	1.06	1.06	1.06
Sonar	Reduct size (|R|)	0	0	0	1	3	4	1	6	11
	Accuracy (%)	0	0	0	48	64	71	67	62	66
	Computing time (s)	0	0	0	0.52	0.52	0.52	0.52	0.52	0.52
Musk	Reduct size (|R|)	0	0	2	2	35	11	19	23	5
	Accuracy (%)	0	0	56	56	74	76	73	76	75
	Computing time (s)	0	0	0.07	0.07	1.55	1.55	1.55	1.55	1.55
VRB	Reduct size (|R|)	2	10	25	1	14	9	2	12	24
	Accuracy (%)	44	56	72	75	75	76	75	76	75
	Computing time (s)	1.02	1.02	1.02	1.02	2.12	2.12	2.12	2.12	2.12
VRG	Reduct size (|R|)	2	10	25	3	11	5	4	9	48
	Accuracy (%)	52	65	65	65	75	68	68	96	69
	Computing time (s)	3.25	3.25	3.25	3.25	7.67	7.67	7.67	7.67	7.67

Table 4 shows the results of using SVM. Time consuming of CFW depends on the existence of reduct ($|R| \neq 0$ or $|R| = 0$). When $|R| = 0$, it means that we cannot find the reduct by Topo Hausdorff approach or we cannot find the attributes satisfying Hausdorff properties by Definition 15 and Proposition 9. In this situation, the time consumption of the proposed algorithm equals zero. On selected datasets, the reducts are almost not defined when the beta is less than 0.5. When beta $\geq$0.5, the size of final reducts is different. The values of beta also affect the smoothness of partitions. The larger value of beta is, the higher the smooth in partitions is.

The results obtained by applying kNN are presented in Table 5 with similar characteristics as in Table 4.

From the results in Tables 4 and 5, the accuracy and size of reduct depend on the classification model applied in CFW algorithm.

6.3 Compare the Proposed Algorithm with Other Attribute Reduction Algorithms

After choosing β for the proposed algorithm, which has the best size and classification accuracy of the reduct. We compare the results of the proposed algorithm to other algorithms. Comparison standards include classification accuracy, the size of reducts, and the computation time. The average value of accuracy, the reduct's size, and the computation time of the algorithms are denoted by $|-|$.

6.3.1 Evaluation of Algorithms on SVM Classification Model

The experimental results of the algorithms on the SVM classification model are presented in Tables 6, 7 and 9. Observing the results in the tables, we can see that the average classification accuracy on 16 datasets is not significantly different from the original dataset and the VPRS and FRS algorithms. However, the reduct's size and the computation time are much better than the VPRS and FRS algorithms. The VPRS and FRS algorithms use the dependency measure according to the POS approach of the extended rough set model. The remaining IFE and FD algorithms using entropy and distance measures give the weakest average results regarding computation time and classification accuracy.

Table 6: The size of the reduct obtained for each algorithm

ID	|C|	|R|					
		CFW-SVM	CFW-kNN	VPRS	FRS	FIE	FD
1	13	10.8	7.6	11.8	10.4	10.6	7.1
2	13	6.7	5.5	11.5	13.9	10.2	6.7
3	20	8.2	8.7	9.5	20.3	20.1	3.5
4	22	5.2	4.4	9.4	8.5	10.8	4.3
5	30	3.2	3.6	14.8	7.6	12.1	4.1
6	32	2.9	2.2	8.9	12.6	12.4	5.8
7	34	2.1	2.1	20.9	11.3	19.6	6.1
8	43	13.9	4.3	15.3	8.7	11.7	5.2
9	43	5.1	3.6	19.9	6.6	8.3	3.3
10	44	3.1	2.2	44.3	10.3	14.7	5.9
11	51	4.1	3.4	8.9	5.8	11.9	5.2
12	54	2.5	1.6	8.4	15.7	24.4	4.4
13	60	4.6	7.4	44.3	17.5	25.2	7.6
14	166	5.7	11.4	86.6	23.9	29.5	8.8
15	310	9.1	4.3	56.6	18.9	35.8	7.5
16	310	9.6	2.1	72.4	16.5	36.4	10.6
$\|-\|$	77.81	6.05	**4.65**	27.71	13.03	18.35	6.02

Table 7: The classification accuracy of each algorithm on each dataset with the SVM classification model

ID	|C|	Classification accuracy (%)				
		CFW-SVM	VPRS	FRS	FIE	FD
1	98 ± 0.7	96 ± 0.9	99 ± 0.6	99 ± 0.3	93 ± 0.1	96 ± 0.8
2	84 ± 0.8	**86 ± 0.6**	84 ± 0.3	84 ± 0.3	82 ± 0.9	80 ± 0.7
3	95 ± 0.8	95 ± 0.4	92 ± 0.4	95 ± 0.1	95 ± 0.8	92 ± 0.6
4	84 ± 0.7	**86 ± 0.6**	84 ± 0.7	85 ± 0.9	84 ± 0.7	75 ± 0.8
5	98 ± 0.6	94 ± 0.7	94 ± 0.2	96 ± 0	96 ± 0.8	94 ± 0.7
6	77 ± 0.3	76 ± 0.3	76 ± 0.6	76 ± 0.2	76 ± 0.8	76 ± 0
7	88 ± 0.5	82 ± 1	88 ± 0.9	87 ± 0.5	87 ± 0.3	89 ± 0.6
8	44 ± 0.8	**59 ± 0.7**	45 ± 0.5	49 ± 0.1	49 ± 0.6	50 ± 1
9	68 ± 0.8	63 ± 0.5	68 ± 0.1	64 ± 1	63 ± 0.7	62 ± 0.5
10	79 ± 0.5	79 ± 1	79 ± 0	79 ± 0	79 ± 0.6	79 ± 0.3
11	100	96 ± 0.9	100	100	92 ± 0.8	100
12	98 ± 0.6	98 ± 0.3	98 ± 0.3	98 ± 0.6	98 ± 0.4	98 ± 0.5
13	65 ± 0.8	**73 ± 0.2**	65 ± 0.2	70 ± 0.7	64 ± 0	58 ± 0
14	75 ± 0.3	72 ± 0.2	74 ± 0.8	61 ± 0.4	61 ± 0.1	55 ± 0.4
15	83 ± 0.1	83 ± 0.2	88 ± 0.6	91 ± 0.4	80 ± 0.8	86 ± 1
16	85 ± 0.9	80 ± 0.2	91 ± 0.7	82 ± 0.5	67 ± 0.2	68 ± 0.4
| − |	83 ± 0.4	82 ± 0.3	83 ± 0.4	82 ± 0.6	79 ± 0.2	79 ± 0.8

Fig. 5 explains the algorithms' performance with the SVM classification model. Each figure illustrates the relationship between the reduct size and the classification ability of each algorithm on each dataset. We can see the difference in the reduct's size from each algorithm and its classification accuracy. The proposed algorithm is superior in time to the other algorithms. The classification accuracy of the reduct from the algorithm method is also better on datasets **Heart**, **CMC**, **PDS**, **BCWP**, **UFDC**, **SHDC**, **DPDS**, **Sonar**, classification accuracy increased from 44% to 59% for noisy datasets **UFDS**, **Sonar**.

In general, on the SVM classification model, the proposed algorithm for the reduct is significantly better than the remaining algorithms, and the computation time is superior. Moreover, the proposed algorithm improves the classification accuracy for noise datasets.

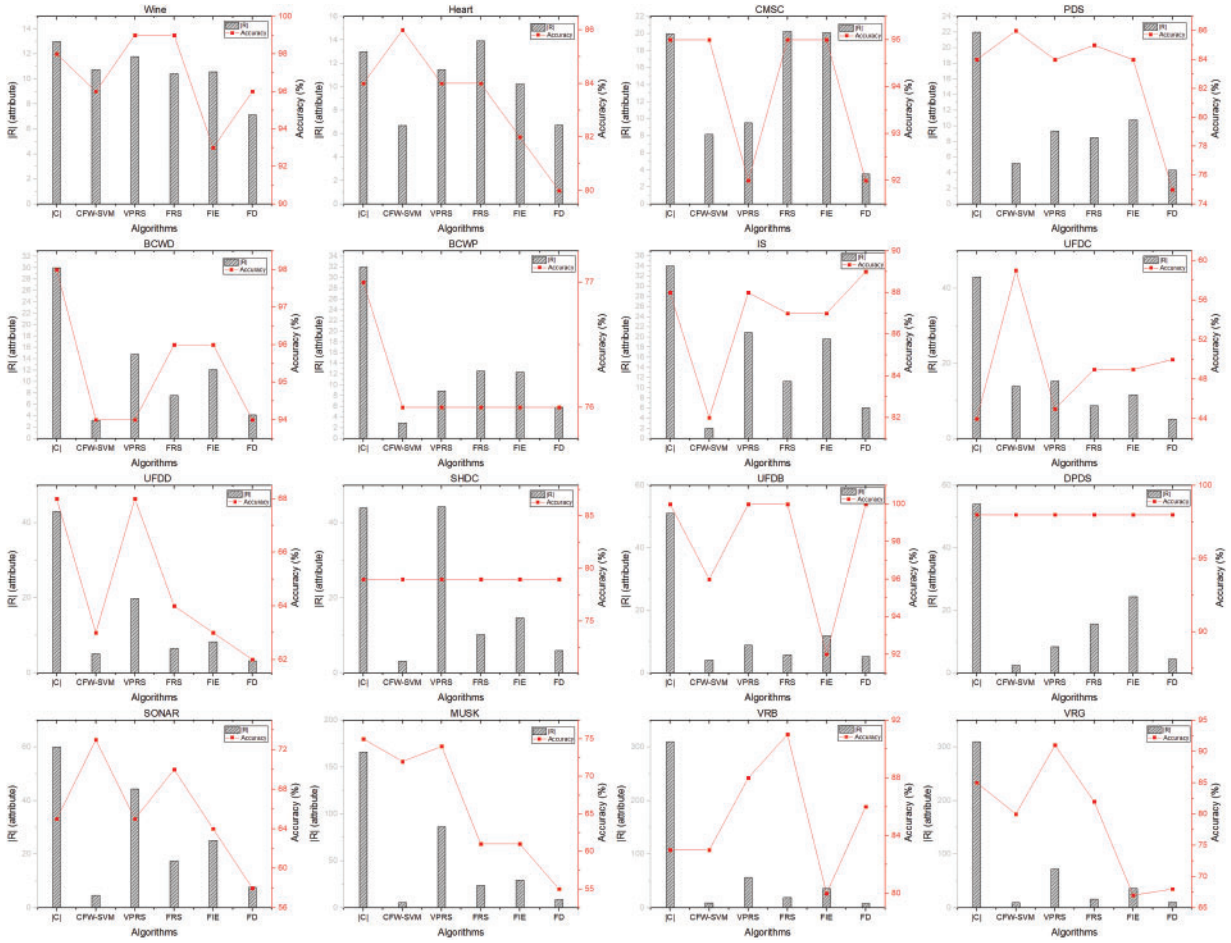

Figure 5: The diagram analyzes the relationship between the reduct's size and classification accuracy at each algorithm on the SVM classification model

6.3.2 Evaluation of Algorithms on kNN Classification Model

The experimental results of the kNN classification model of the algorithms are described in detail in Tables 6, 8 and 9. Observing the effects on each dataset with each evaluation criterion showed that 12 out of 16 datasets gave significantly better results than SVM classification model.

Table 8: The classification accuracy of each algorithm on each dataset with the kNN classification model

ID	C	Classification accuracy (%)				
		CFW-kNN	VPRS	FRS	FIE	FD
1	96 ± 0.2	94 ± 0.1	94 ± 0.1	96 ± 0.9	91 ± 0.4	94 ± 0.6
2	77 ± 0.5	$\mathbf{78 \pm 0.1}$	77 ± 0.3	77 ± 0.3	76 ± 0.2	69 ± 0.7

(Continued)

Table 8 (continued)

ID	C	Classification accuracy (%)				
		CFW-kNN	VPRS	FRS	FIE	FD
3	84 ± 0.1	**92 ± 0.1**	86 ± 0.2	84 ± 0.6	84 ± 0.9	71 ± 0.1
4	85 ± 0.7	85 ± 0.3	88 ± 0.9	87 ± 0.1	84 ± 0.3	74 ± 0.5
5	95 ± 0.2	93 ± 0.1	93 ± 0.9	93 ± 0.9	94 ± 0.7	93 ± 0.7
6	78 ± 0.8	**81 ± 0.9**	74 ± 0.6	79 ± 0.6	79 ± 0.6	75 ± 0.6
7	85 ± 0.6	88 ± 0.6	86 ± 0.9	88 ± 0.7	88 ± 0.4	89 ± 0.4
8	82 ± 0.1	**96 ± 0.2**	82 ± 0.1	74 ± 0.9	78 ± 0.1	76 ± 0.2
9	81 ± 0.5	81 ± 0.9	77 ± 0.9	77 ± 0.5	82 ± 0.6	72 ± 0.7
10	66 ± 0.1	**75 ± 0.7**	66 ± 0.5	69 ± 0.8	67 ± 1	72 ± 0.6
11	98 ± 0.8	100	100	100	98 ± 0	98 ± 0.5
12	98 ± 0.4	98 ± 0	98 ± 0.3	98 ± 0.4	96 ± 0.9	98 ± 0.2
13	68 ± 0.3	**71 ± 0.3**	64 ± 0.5	62 ± 0.7	60 ± 0.7	55 ± 0.3
14	77 ± 0.5	76 ± 0.7	77 ± 0.1	75 ± 1	69 ± 0.4	64 ± 0.3
15	68 ± 0.3	76 ± 0.6	77 ± 0.1	81 ± 0.3	65 ± 0.1	73 ± 0.1
16	70 ± 0.8	**96 ± 0.4**	75 ± 0.8	76 ± 0.9	61 ± 1	60 ± 0.9
| − |	82 ± 0.7	**86 ± 0.6**	82 ± 0.6	82 ± 0.4	80 ± 0.2	77 ± 0

Table 9: The computation time of each algorithm

ID	Computation time (s)					
	CFW-SVM	CFW-kNN	VPRS	FRS	FIE	FD
1	0.05	0.07	0.7	1.27	0.58	0.3
2	0.08	0.1	1.08	0.79	1.03	0.63
3	0.28	0.26	6.19	18.28	6.14	3.79
4	0.17	0.09	1.24	3.09	1.23	0.6
5	0.67	0.49	11.19	23.01	10.69	7.3
6	0.71	0.57	1.81	4.13	2.14	1.04
7	0.6	0.43	6.13	12.17	6.08	3.17
8	0.11	0.49	4.23	6.73	2.49	1.17
9	0.18	0.62	4.84	5.43	2.11	1.12
10	0.45	0.85	2.41	9.16	4.8	2.29
11	0.56	0.77	1.22	1.39	1.22	0.38
12	0.82	1.06	2.41	6.42	4.34	1.31
13	0.63	0.52	7.8	11.63	6.65	2.25
14	2.26	1.55	124.48	216.6	73.32	32.31
15	1.7	2.12	36.25	28.67	28.59	4.1
16	1.93	7.67	131.24	26.09	29.11	3.63
| − |	**0.70**	**1.10**	21.45	23.43	11.28	4.09

The remaining datasets have similar classification accuracy results, but the reduct's size and computation time are superior. Fig. 6 provides an in-depth breakdown of the performance of several methods when applied to the kNN classification model. The scaler on each graph illustrates the relationship between the reduct's size and the classification accuracy. We can notice the difference in classification accuracy and the difference in the reduct's size achieved by each algorithm for each dataset. The proposed attribute reduction method is superior to the other algorithms.

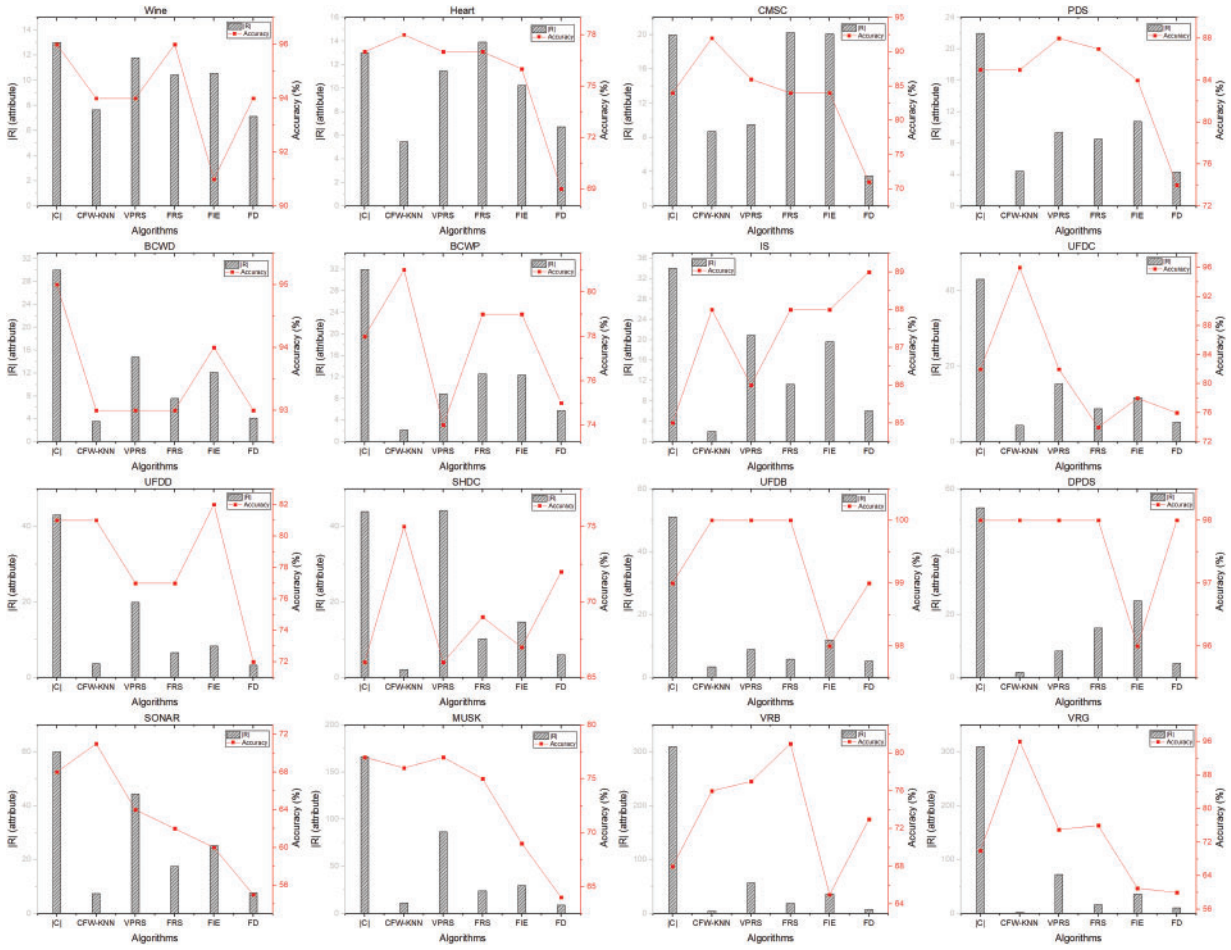

Figure 6: The diagram analyzes the relationship between the reduct's size and classification accuracy at each algorithm on the kNN classification model

In addition, the reduct from the proposed method has a higher level of accuracy **Heart**, **CMCS**, **BCWP**, **IS**, **UFDC**, **UFDD**, **SHDC**, **UFDB**, **DPDS**, **Sonar**, **Musk**, **VRG**. In particular, performance has been improved for the noise dataset **VRG**, with an improvement in accuracy from 70 to 96 percent.

In general, on the kNN classification model, the proposed algorithm for the reduct is significantly better than the remaining algorithms. In addition, the computation time and classification accuracy are superior to those of the compared algorithms that are depicted in Fig. 7.

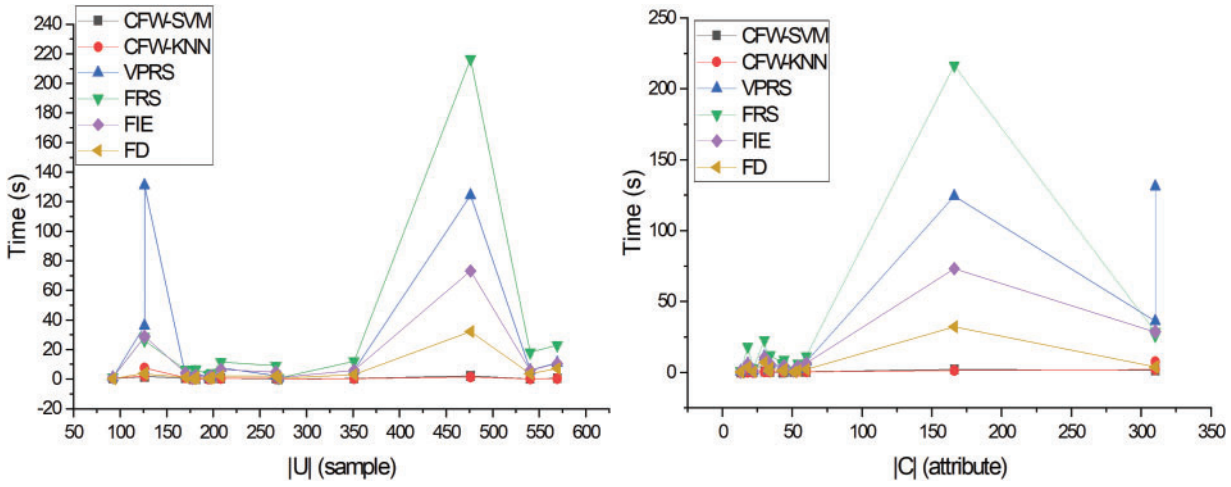

Figure 7: The diagram analyzes the relationship between the computation time and $|U|$ (left), between the computation time and $|C|$ (right) of algorithms

7 Discussion

The experimental results presented in Section 6 of this study show that the proposed algorithm is superior to others. But the question is, what factors make the algorithm efficient regarding the computation time, the classification accuracy, and the reduct's size? We will first analyze the computation time.

7.1 The Computation Time of the Proposed Algorithm

As presented in the introduction of this study, the proposed algorithm is significantly better than other algorithms in theoretical computation time. Most traditional attribute reduction methods use a filter with one-by-one attributes to create the reduct, which is ineffective for large-dimensional datasets.

Furthermore, traditional attribute reduction methods are based on granular computation, which has to use approximation spaces to process. Then, large datasets will have a significant computational load on this approximation space. Most of all have a complexity of $\mathcal{O}\left(|U|^2|C|^2\right)$. In contrast, the complexity of the proposed algorithm is $\mathcal{O}\left(2|U||C|\right)+\mathcal{O}\left(|U|^2|H^\beta|\right)+\mathcal{O}\left(|CH^\beta|\mathbb{T}\right)$. Two special phases significantly improve the computation time, which the filtering phase of Hausdorff attributes with complexity is $\mathcal{O}\left(2|U||C|\right)$ and the cluster phase clustering the same structure Hausdorff attributes with complexity $\mathcal{O}\left(|U|^2|H^\beta|\right)$, where $|H^\beta|$ is much smaller than the original $|C|$. The effectiveness of time running of algorithms is presented in Fig. 7.

7.2 The Classification Accuracy of the Reduct from the Proposed Algorithm

Most of the traditional attribute reduction methods based on the rough set approach use the measure to evaluate the significance of the attribute. But those measures still have the disadvantage of only considering the similarity of two subsets based on the total components without considering the similarity for each element. Meanwhile, the topology approach allows us to evaluate the component's similarity through the structure of that components. In this study, we use the Hausdorff topology structure because of distinguishable components in the Hausdorff topology. It is also why the traditional rough set model works so well on categorical datasets. So, we use the distinguishable concept in the rough set model for the topology space.

7.3 Size of the Reduct from the Proposed Algorithm

As mentioned above, the topology approach to distinguish attributes is better than the traditional approach. Yu et al. [38] have also shown that two different granular can have the same topology based on the rough set method. Therefore, the higher the similarity, the more attributes in the same group and *vice versa*. The attribute wrapper method is especially suitable for attribute reduction based on the topology approach. Suppose that the cluster phase divides the subset of the original attribute into many clusters. In that case, the resulting reduct will have a small size and a higher ability to choose a group of attributes with the best accuracy, and *vice versa*.

7.4 Two-Way ANOVA Analysis

After getting the experiment results from applying the proposed structure using 9 different values of beta in the range of [0, 1], two-way ANOVA (Analysis of Variance) is used in order to demonstrate the significance of the differences in performance metrics across various datasets. The significant level in this analysis is alpha = 0.05. Hypotheses, in this case, include Null hypothesis (H0): The means of accuracy on the datasets are the same. Alternative hypothesis (H1): There is at least one different value.

The summary results on kNN are presented in Table 10 below.

Table 10: Two-way ANOVA with different values of beta on kNN

ANOVA						
Source of variation	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Values of beta	10.7290	8	1.3412	35.1660	4.85157E-28	2.0164
Datasets	2.6419	15	0.1761	4.6183	6.74521E-07	1.7505
Error	4.5765	120	0.0381			
Total	17.9480	143				

From two first rows in Table 10, the values of F are larger than F crit in both *Values of beta* and *Datasets*. This means that the different values of beta will lead to different levels of accuracy. Thus, in our experiments, the best value of beta is selected and applied to the topology. Moreover, the different datasets will get different values in accuracy.

The summary results of ANOVA analysis from applying SVM are presented in Table 11.

Table 11: Two-way ANOVA with different values of beta on SVM

ANOVA						
Source of variation	*SS*	*df*	*MS*	*F*	*p-value*	*F crit*
Values of beta	10.5067	8	1.3133	33.3907	4.02614E-27	2.0164
Datasets	2.42462	15	0.1616	4.1096	5.00269E-06	1.7505
Error	4.71988	120	0.0393			
Total	17.6512	143				

Due to the values of F being larger than F crit in both aspects in Table 11, we can state that by applying SVM in our topology, the different values of beta will lead to different levels of accuracy. Thus, the best value of beta is selected and applied to the topology.

8 Conclusion

As mentioned in the introduction, attribute reduction is an important problem widely applied in many fields related to knowledge processing. However, most current attribute reduction algorithms using the measure approach have the complexity of $\mathcal{O}\left(|U|^2|C|^2\right)$, so it is unsuitable for large datasets. The proposed method has a theoretical complexity of $\mathcal{O}\left(2|U||C|\right) + \mathcal{O}\left(|U|^2|H^\beta|\right) + \mathcal{O}\left(|CH^\beta|\mathbb{T}\right)$ that much faster than the traditional attribute reduction method. The proposed algorithm is entirely new in both approach and design method, including:

1) Use the Hausdorff topology as a criterion to select relative attributes. This stage generates H^β of relative reducts and has complexity $\mathcal{O}\left(2|U||C|\right)$, significantly improving the selection time of relative attributes.

2) Use the concept of dependent D-isomorphism to generate candidate reducts. This stage generates CH^β groups of candidate reducts that do not intersect and have complexity $\mathcal{O}\left(|U|^2|H^\beta|\right)$, significantly minimizing the number of candidate reducts.

In the future, we can develop new algorithms based on the distinguishing properties of Hausdorff topology, such as the cross-exchange method, cluster computation method, core computation method, dependent topological structure selection method, and some hybrid methods.

Acknowledgement: The authors are grateful to all the editors and anonymous reviewers for their comments and suggestions.

Funding Statement: This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant Number 102.05-2021.10.

Author Contributions: Nguyen Long Giang and Le Hoang Son are responsible for idea convey and revised & review of the manuscript; Tran Thanh Dai is responsible for 1st draft writing and experiments; Tran Thi Ngan and Nguyen Nhu Son are responsible for data collection and preparation with analysis; Cu Nguyen Giap is responsible for all processes of research. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Data is available from the corresponding author on reasonable request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. Bashir, I. U. Khattak, A. Khan, F. H. Khan, A. Gani and M. Shiraz, "A novel feature selection method for classification of medical data using filters, wrappers, and embedded approaches," *Complexity*, vol. 2022, no. 1, pp. 1–12, 2022. doi: 10.1155/2022/8190814.

[2] L. Meenachi and S. Ramakrishnan, "Differential evolution and ACO based global optimal feature selection with fuzzy rough set for cancer data classification," *Soft Comput.*, vol. 24, no. 24, pp. 18463–18475, 2020. doi: 10.1007/s00500-020-05070-9.

[3] M. Baldomero-Naranjo, L. I. Martínez-Merino, and A. M. Rodríguez-Chía, "A robust SVM-based approach with feature selection and outliers detection for classification problems," *Expert. Syst. Appl.*, vol. 178, no. 12, 2021, Art. no. 115017. doi: 10.1016/j.eswa.2021.115017.

[4] M. Riahi-Madvar, A. Akbari Azirani, B. Nasersharif, and B. Raahemi, "A new density-based subspace selection method using mutual information for high dimensional outlier detection," *Knowl.-Based Syst.*, vol. 216, no. 2, 2021. doi: 10.1016/j.knosys.2020.106733.

[5] A. J. Fernández-García, L. Iribarne, A. Corral, J. Criado, and J. Z. Wang, "A recommender system for component-based applications using machine learning techniques," *Knowl.-Based Syst.*, vol. 164, no. 4, pp. 68–84, 2019. doi: 10.1016/j.knosys.2018.10.019.

[6] B. Saravanan, V. Mohanraj, and J. Senthilkumar, "A fuzzy entropy technique for dimensionality reduction in recommender systems using deep learning," *Soft Comput.*, vol. 23, no. 8, pp. 2575–2583, 2019. doi: 10.1007/s00500-019-03807-9.

[7] D. I. Taher, R. Abu-Gdairi, M. K. El-Bably, and M. A. El-Gayar, "Decision-making in diagnosing heart failure problems using basic rough sets," *AIMS Math.*, vol. 9, no. 8, pp. 21816–21847, 2024. doi: 10.3934/math.20241061.

[8] R. Abu-Gdairi and M. K. El-Bably, "The accurate diagnosis for COVID-19 variants using nearly initial-rough sets," *Heliyon*, vol. 10, no. 10, 2024, Art. no. e31288. doi: 10.1016/j.heliyon.2024.e31288.

[9] M. K. El-Bably, R. Abu-Gdairi, and M. A. El-Gayar, "Medical diagnosis for the problem of Chikungunya disease using soft rough sets," *AIMS Math.*, vol. 8, no. 4, pp. 9082–9105, 2023. doi: 10.3934/math.2023455.

[10] R. A. Hosny, R. Abu-Gdairi, and M. K. El-Bably, "Enhancing Dengue fever diagnosis with generalized rough sets: Utilizing initial-neighborhoods and ideals," *Alexandria Eng. J.*, vol. 94, no. 3, pp. 68–79, 2024. doi: 10.1016/j.aej.2024.03.028.

[11] R. Abu-Gdairi, A. A. El-Atik, and M. K. El-Bably, "Topological visualization and graph analysis of rough sets via neighborhoods: A medical application using human heart data," *AIMS Math.*, vol. 8, no. 11, pp. 26945–26967, 2023. doi: 10.3934/math.20231379.

[12] D. S. Truong, L. Thanh Hien, and N. Thanh Tung, "An effective algorithm for computing reducts in decision tables," *J. Comput. Sci. Cybern.*, vol. 38, no. 3, pp. 277–292, Sep. 2022. doi: 10.15625/1813-9663/38/3/17450.

[13] P. Viet Anh, V. Duc Thi, and N. Ngoc Cuong, "A novel algorithm for finding all reducts in the incomplete decision table," *J. Comput. Sci. Cybern.*, vol. 39, no. 4, pp. 313–321, Nov. 2023. doi: 10.15625/1813-9663/18680.

[14] P. M. Ngoc Ha, T. D. Tran, N. M. Hung, and H. T. Dung, "A novel extension method of VPFRS mode for attribute reduction problem in numerical decision tables," *J. Comput. Sci. Cybern.*, vol. 40, no. 1, pp. 37–51, Mar. 2024. doi: 10.15625/1813-9663/19696.

[15] J. Ye, J. Zhan, W. Ding, and H. Fujita, "A novel fuzzy rough set model with fuzzy neighborhood operators," *Inform. Sci.*, vol. 544, no. 1, pp. 266–297, 2021. doi: 10.1016/j.ins.2020.07.030.

[16] L. Zhang, J. Zhan, and J. C. R. Alcantud, "Novel classes of fuzzy soft β-coverings-based fuzzy rough sets with applications to multi-criteria fuzzy group decision making," *Soft Comput.*, vol. 23, no. 14, pp. 5327–5351, 2019. doi: 10.1007/s00500-018-3470-9.

[17] J. He, L. Qu, Z. Wang, Y. Chen, D. Luo and C. F. Wen, "Attribute reduction in an incomplete categorical decision information system based on fuzzy rough sets," *Artif. Intell. Rev.*, vol. 55, no. 7, pp. 5313–5348, 2022. doi: 10.1007/s10462-021-10117-w.

[18] X. Zhang, B. Zhou, and P. Li, "A general frame for intuitionistic fuzzy rough sets," *Inform. Sci.*, vol. 216, no. 5, pp. 34–49, 2012. doi: 10.1016/j.ins.2012.04.018.

[19] A. Tan, W. Z. Wu, Y. Qian, J. Liang, J. Chen and J. Li, "Intuitionistic fuzzy rough set-based granular structures and attribute subset selection," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 3, pp. 527–539, 2019. doi: 10.1109/TFUZZ.2018.2862870.

[20] Z. Li, Y. Chen, G. Zhang, L. Qu, and N. Xie, "Entropy measurement for a hybrid information system with images: An application in attribute reduction," *Soft Comput.*, vol. 26, no. 21, pp. 11243–11263, 2022. doi: 10.1007/s00500-022-07502-0.
[21] J. Xu, Y. Wang, H. Mu, and F. Huang, "Feature genes selection based on fuzzy neighborhood conditional entropy," *J. Intell. Fuzzy Syst.*, vol. 36, no. 1, pp. 117–126, 2019. doi: 10.3233/JIFS-18100.
[22] X. Ma, J. Wang, W. Yu, and Q. Zhang, "Attribute reduction of hybrid decision information systems based on fuzzy conditional information entropy," *Comput. Mater. Contin.*, vol. 79, no. 2, pp. 2063–2083, 2024. doi: 10.32604/cmc.2024.049147.
[23] A. Tan, S. Shi, W. Z. Wu, J. Li, and W. Pedrycz, "Granularity and entropy of intuitionistic fuzzy information and their applications," *IEEE Trans. Cybern.*, vol. 52, no. 1, pp. 192–204, 2022. doi: 10.1109/TCYB.2020.2973379.
[24] N. L. Giang *et al.*, "Novel incremental algorithms for attribute reduction from dynamic decision tables using hybrid filter-wrapper with fuzzy partition distance," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 5, pp. 858–873, 2020. doi: 10.1109/TFUZZ.2019.2948586.
[25] T. T. Nguyen *et al.*, "A novel filterwrapper algorithm on intuitionistic fuzzy set for attribute reduction from decision tables," *Int. J. Data Warehousing Min.*, vol. 17, no. 4, pp. 67–100, 2021. doi: 10.4018/IJDWM.
[26] S. E. Han, "Digital topological rough set structures and topological operators," *Topol. Appl.*, vol. 301, no. 2, 2021. doi: 10.1016/j.topol.2020.107507.
[27] S. Mishra and R. Srivastava, "Fuzzy topologies generated by fuzzy equivalence relations," *Soft Comput.*, vol. 22, no. 2, pp. 373–385, 2018. doi: 10.1007/s00500-016-2458-6.
[28] P. K. Singh and S. Tiwari, "Topological structures in rough set theory: A survey," *Hacettepe J. Math. Stat.*, vol. 49, no. 4, pp. 1270–1294, 2020. doi: 10.15672/hujms.662711.
[29] E. F. Lashin and T. Medhat, "Topological reduction of information systems," *Chaos Solitons Fractals*, vol. 25, no. 2, pp. 277–286, 2005. doi: 10.1016/j.chaos.2004.09.107.
[30] Z. M. Ma and B. Q. Hu, "Topological and lattice structures of L-fuzzy rough sets determined by lower and upper sets," *Inform. Sci.*, vol. 218, no. 5, pp. 194–204, 2013. doi: 10.1016/j.ins.2012.06.029.
[31] L. Su and W. Zhu, "Dependence space of topology and its application to attribute reduction," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 4, pp. 691–698, 2018. doi: 10.1007/s13042-016-0598-8.
[32] S. E. Han, "Topological properties of locally finite covering rough sets and K-topological rough set structures," *Soft Comput.*, vol. 25, no. 10, pp. 6865–6877, 2021. doi: 10.1007/s00500-021-05693-6.
[33] C. Y. Wang, "Topological characterizations of generalized fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 312, pp. 109–125, 2017. doi: 10.1016/j.fss.2016.02.005.
[34] K. Qin and Z. Pei, "On the topological properties of fuzzy rough sets," *Fuzzy Sets Syst.*, vol. 151, no. 3, pp. 601–613, 2005. doi: 10.1016/j.fss.2004.08.017.
[35] Z. Bashir, M. G. Abbas Malik, S. Asif, and T. Rashid, "The topological properties of intuitionistic fuzzy rough sets," *J. Intell. Fuzzy Syst.*, vol. 38, no. 1, pp. 795–807, 2020. doi: 10.3233/JIFS-179449.
[36] S. M. Yun, Y. S. Eom, and S. J. Lee, "Topology of the redefined intuitionistic fuzzy rough sets," *Int. J. Fuzzy Logic Intell. Syst.*, vol. 21, no. 4, pp. 369–377, 2021. doi: 10.5391/IJFIS.2021.21.4.369.
[37] T. T. Dai, N. L. Giang, V. D. Thi, T. T. Ngan, H. T. M. Chau and L. H. Son, "A new approach for attribute reduction from decision table based on intuitionistic fuzzy topology," *Soft Comput.*, vol. 28, no. 20, pp. 11799–11822, 2024. doi: 10.1007/s00500-024-09910-w.
[38] H. Yu and W. R. Zhan, "On the topological properties of generalized rough sets," *Inf. Sci.*, vol. 263, no. 1, pp. 141–152, 2014. doi: 10.1016/j.ins.2013.09.040.
[39] M. K. El-Bably, K. K. Fleifel, and O. A. Embaby, "Topological approaches to rough approximations based on closure operators," *Granul. Comput.*, vol. 7, no. 1, pp. 1–14, 2022. doi: 10.1007/s41066-020-00247-x.
[40] W. Yao and S. E. Han, "A topological approach to rough sets from a granular computing perspective," *Inf. Sci.*, vol. 627, no. 3–4, pp. 238–250, 2023. doi: 10.1016/j.ins.2023.02.020.
[41] S. Zhao, E. C. C. Tsang, and D. Chen, "The model of fuzzy variable precision rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 2, pp. 451–467, 2009. doi: 10.1109/TFUZZ.2009.2013204.

[42] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, 2009. doi: 10.1109/TFUZZ.2008.924209.
[43] UCI, "Machine learning repository," 2021. Accessed: Mar. 26, 2023. [Online]. Available: https://archive.ics.uci.edu/ml/index.php