# SCALING-UP REDUNDANT FEATURES USING FUZZY-ROUGH DATA ANALYSIS TECHNIQUES.

**K. Anitha**
Department of Mathematics, SRM Institute of Science and Technology- Ramapuram, Chennai, India, anithamaths2019@gmail.com, anithak1@srmist.edu.in

Abstract
Feature Selection is an important task in data analysis especially in high dimensional data like health care data, genome data and financial data. A decision can be taken from the available information which should be certain and precise. In that case if the data is imprecise or noisy data, we cannot perform the optimal selection and the decision will not be accurate. For handling such types of inconsistent information, the theory proposed by Pawlak in 1982 performs very well named as Rough set or Indefinable set since it does not require additional information other than the available knowledge base. It is expressed by its set approximations known as lower and upper approximation and these approximations are constructed through equivalence relations between the variables (feature). The feature $x_i$ is said to be consistent if the removal of $x_i$ will affect the quality of data. Mathematically speaking the set contains these types of variables is called Reduct. An information system contains $n$ number of reduct set. Identification of stable reducts are essential which will enable the data scientists to bring optimal decision making. These type of reducts are known as Dynamic reducts. This paper proposes the method of finding Dynamic reducts using Fuzzy-Rough set model and it is being implemented for Heart Disease Data set.

Keywords: Information System, Set approximations, Fuzzy-Rough Set.
**AMS Subject Classification:** 03E20, 03E72, 03E75, 54C05, 54D05

## I. Introduction

With the advancement of computer technology and cloud computing lots of data repositories are increasing rapidly. In this large volume of data, a greater number of features may be inconsistent or redundant. For identifying these types of variables is a challenging task for data scientists. The data become meaningful when it contains only relevant attributes. For removing irrelevant variables, it is essential to identify the pattern among the data. Many technical tools are available to do this task. But Pawlak Rough set [5] gives the mathematical approach to construct the pattern hidden in the data with using the available information and it does not require additional information. Redundant features can be removed through reduct generation of rough sets. Reducts are the subsets which describes the entire information system. Kopczynski [26] proposed FPGA supported reduct generation hardware-based technique using classical rough sets which is tested for the data sets having 1000 objects to 10,00,000 objects. Zhouming Ma et al [27] gave the extension of variable precision rough set model named CVPRS model which focus on boundary region without changing the set approximation. It is efficient that Rough hybridization technique brings optimal output while handling continuous

**Submitted**: 16/11/2022 **Accepted**: 05/12/2022

valued data when compared with classical rough set model. In this paper Fuzzy-Rough set model is being proposed to handle redundant variables. Fuzzy-rough set covers more number of variables which are having uncertainty and redundancy through its membership values. Mohammed Atef [28] et al constructed certain types of covering for fuzzy-rough sets and also brought relationship among this model with numerical examples. The measure of uncertainty can be increased through fuzzy positive region which is the union of fuzzy lower approximation which gave the lead to fuzzy-rough set approximations. This paper proposes an algorithm for resolving redundancy through dynamic reduct calculation by using fuzzy-rough set model. Section II describes literature survey, Section III focusses preliminary concepts, Section IV exhibits proposed work followed by experimental results and conclusions in consecutive sections.

Decision Making is the act of electing optimal choice from the set of feasible choices. Innumerable mathematical methods have been discussed so far [1-4]. Even though in these numerous techniques Pawlak set theory named rough set plays prominent role in decision making through its indiscernible and discernible relations [5,6]. Certain topological properties of rough set were discussed in [18]. Felix and Ushio proposed binary discernibility matrix for attribute Reduction [10]. In the updated version of [10] Nan Zhang et al proposed finite automata-based method to simplify the Discernibility matrix. Basically, this discernibility matrix was introduced by Skowron and Rauszer [9] in the year 1992. The entries of this matrix are defined through discernible relation which states that two objects are discernible if their attribute values are unequal. Anitha et al [17] identified metric dimension for graphical structure of rough sets. Yao and Zhao proposed simplification process of this matrix for the purpose of calculating the reducts [10]. Anitha and Thangeswari proposed rough set-based optimization techniques in [16]. The term reduct is the key term of rough set. Two objects are said to be discernible if atleast one of the attributes values is different. There can be many subsets constructed for an attribute set. The subset with minimum cardinality and which describes the whole concepts is known as reducts. The information system has many numbers of reduct sets. But removal of these reducts will definitely affect the quality of data. The term core is most influential than reduct since intersection of all reducts forms core. The attributes in the core set will be the decision-making attribute of the entire Information System.

From an Information System we can procreate many reducts and these reducts will contain redundant attribute which will not be stable with respect to the decision rules. The set of reduct which is completely stable with its decision rules is called Dynamic reduct. Generation of Dynamic reducts was proposed in [11] through Object oriented rough set model. In this work they brought into existence Dynamic core through class labels. Walid et al [12] fabricated Dynamic reduct through dynamic programming method and they exhibited their method to Retail Business data. Rahul Kumar gave various fuzzy-rough attribute reduction techniques in [30]. This paper proposes the mathematical model based on fuzzy-rough sets for removing the redundant variables from the data set with uncertainty and ambiguity. This work elicit the novel technique to find the stable reducts known as dynamic reduct based attriubute reduction based on fuzzy-rough intelligent system.

## II. Mathematical Preliminaries of Rough set.

**Definition 3.1: Information Table [10]**

An Information table $I_T$ is a tuple $I_T = \{U_o, A_t, (V_x|x \in A_t), (I_x|x \in A_t)\}$. Where $U_o$, $A_t$, $V_x$ are non-empty finite set of objects, attributes and its values respectively. An information function $I_x: U_o \rightarrow V_x$ that maps an object to exactly one value of $V_x$.

**Definition 3.2: Indiscernibility Relation [10]**

The subset $S \subseteq A_t$, an Indiscernibility relation $IND(S) \subseteq U_o \times U_o$ which is defined by

$$IND(S) = \{(a, b), \forall x \in S, I_x(a) = I_x(b)\}$$

**Definition 3.3: Reduct [10]**

From a given Information Table $I_T$, an Attribute set $R_e \subseteq A_t$ is called Reduct of $I_T$ provided it should satisfy the following axioms

(i) $IND(R_e) = IND(A_t)$

(ii) For any attribute $a \in R_e, IND(R_e - \{a\}) \neq IND(A_t)$ which states that removal of any element from $R_e$ will affect the quality of data.

(iii) $R_e$ should be the set with minimum cardinality with the above two properties.

**Definition 3.4: Discernibility Matrix [10]**

For a given Information Table $I_T$, its Discernibility Matrix $M_{x,y}$ is defined by

$$M_{x,y} = \{a \in A_t, I_a(x) \neq I_a(y)\}$$

Example 3.1

For the following Information and Decision system, the corresponding Discernibility Matrix is reported here

*Tab.3.1: Information System*

| Patient | Temperature (T) | Muscle-Pain(M) | Cough & Cold(C) | loss of taste or smell(L) |
|---|---|---|---|---|
| $p_1$ | High | Severe | yes | yes |
| $p_2$ | High | no | yes | no |
| $p_3$ | no | Severe | no | yes |
| $p_4$ | High | no | no | yes |
| $p_5$ | no | Severe | yes | no |
| $p_6$ | High | no | no | yes |

*Tab.3.2: Decision System*

| Patient | Temperature (T) | Muscle-Pain(M) | Cough & Cold(C) | loss of taste or smell(L) | Covid + |
|---|---|---|---|---|---|
| $p_1$ | High | Severe | yes | yes | Positive |
| $p_2$ | High | no | yes | no | Positive |
| $p_3$ | no | Severe | no | yes | Negative |
| $p_4$ | High | no | no | yes | Positive |
| $p_5$ | no | Severe | yes | no | Negative |
| $p_6$ | High | no | no | yes | Positive |

*Tab.3.3 Discernibility Matrix for the Information System*

| Patient | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|---|
| $p_1$ | - | - | - | - | - | - |
| $p_2$ | M, L | - | - | - | - | - |
| $p_3$ | T, C | T, M, C, L | - | - | - | - |
| $p_4$ | M, C | C, L | T, M | - | - | - |
| $p_5$ | T, L | T, M, L | C, L | T, M, C, L | - | - |
| $p_6$ | M, C | C, L | T, M | - | T, M, C, L | - |

**Definition 3.4: Relative Discernibility Matrix [10]**

Tab.3.3 refers the Discernibility Matrix with respect to the conditional attributes. But for the construction of Reducts it is important to concentrate on Decision attributes. Yao et al gave the definition for Relative Discernibility Matrix as

$$M_{x,y} = \{a \in A_t, [I_a(x) \neq I_a(y)] \wedge [[I_{desion}(x) \neq I_{desion}(y)]\}$$

The following Tab.3.4 represents the Relative discernibility Matrix of the above Decision system

*Tab.3.4 Discernibility Matrix for the Decision System*

| Patient | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|---|
| $p_1$ | - | - | - | - | - | - |
| $p_2$ | - | - | - | - | - | - |
| $p_3$ | T, C | T, M, C, L | - | - | - | - |
| $p_4$ | - | - | T, M | - | - | - |
| $p_5$ | T, L | T, M, L | - | T, M, C, L | - | - |
| $p_6$ | - | - | T, M | - | T, M, C, L | - |

**Definition 3.5: Discernibility function [10]**

The function $f_M$ which is derived from a Discernibility Matrix is called Discernibility Function defined as follows

$$f_M = \bigwedge \left\{ \bigvee M_{x,y} | \ x, y \in U_o, M_{x,y} \ is \ non - empty \right\}$$

The prime implicants of $f_M$ identify the minimum subset of $A_t$ which dominates the decision class. $f_M$ calculates all discernible pairs from $I_T$ which are called Boolean function or implicants. The discernibility function $f_M$ for the Information system $I_T$ contains the propositions of $r$ Boolean variables related to the attributes $a_1, a_2, \ldots\ldots a_r$ such that

$$f_M(a_1, a_2, \ldots\ldots a_r) = \bigwedge \left\{ \bigvee d_{ij} | \quad d_{ij} = [M]_{x,y} \right\}$$

The set of all prime implicants from the discernibility functions forms reducts.

Reduct sets for the above Decision system is

$$R_1 = \{Temperature, Muscle\ Pain, Loss\ of\ Taste\ and\ Smell\}$$
$$R_2 = \{Temperature, Cold\}$$
$$R_3 = \{Temperature, Muscle\ Pain, Cold\}$$
$$R_4 = \{Temperatue, Loss\ of\ Taste\ and\ Smell\}$$

From these set of reducts decision rules can be done.

**Definition 3.6: Fuzzy- Rough Set [19]**

Information system is the basic building block for the construction of rough set. If the conditional attributes are certain but decision attributes are uncertain then the boundary region may be over lapped. Rough hybridization technique with fuzzy set resolves this issue. Fuzzy-rough set construction is the transformation process of equivalence relation in to fuzzy similarity relation which will accommodate a greater number of variables in each cluster. Dubois and Prade [20] gave insight to the construction of fuzzy-rough set.

Let $\mathcal{K}$ be the set, $\mathcal{R}, \mathcal{F}$ be the equivalence relation and fuzzy set defined on $\mathcal{K}$. Then set approximations are defined as follows

$$\mu_{\overline{\mathcal{R}(\mathcal{F})}}(\mathcal{K}_i) = Sup\{\mu_{\mathcal{F}}(k) | \ w(\mathcal{K}_i) = [k]_{\mathcal{R}}\}$$
$$\mu_{\underline{\mathcal{R}(\mathcal{F})}}(\mathcal{K}_i) = Inf\{\mu_{\mathcal{F}}(k) | \ w(\mathcal{K}_i) = [k]_{\mathcal{R}}\}$$

Let $\psi$ be the fuzzy partition on $\mathcal{K}$. Then the set approximations are exhibited as

$$\overline{\psi(\mathcal{F})} = Sup_k\ Min\{\mu_{\mathcal{F}}(k), \mu_{\mathcal{F}_i}(k)\}$$
$$\underline{\psi}(\mathcal{F}) = Inf_k\ Max[1 - \{\mu_{\mathcal{F}}(k), \mu_{\mathcal{F}_i}(k)\}]$$

The pair $\left(\overline{\psi(\mathcal{F})}, \underline{\psi}(\mathcal{F})\right)$ is called Fuzzy-Rough set.

The extensive properties of fuzzy-rough sets are discussed in [21,22,23]. Mohammed and Azzam [24] proposed variable precision covering for fuzzy-rough sets. They introduced four models of variable precision covering for fuzzy-rough set and implemented these 4 models in MADM decision making.

Similarity informative measure between the elements of fuzzy-rough set is given by[25]

$$S(x,y) = 1 - \frac{1}{2}\{|x_{Low} - y_{Low}| + |x_{Upp} - y_{Upp}|\}$$

The more the information measure value refers more uncertainty.

**Definition 3.7: Fuzzy- Rough Decision System**

Let $I_T$ be the Information System, $\mathcal{F}_S$ be the set of fuzzy-similarity relation on conditional attributes then $(I_T, \mathcal{F}_S \cup D)$ is called Fuzzy decision system. Fuzzy-rough discernibility matrix is defined as

$$F_Z D = \begin{pmatrix} \emptyset & \cdots & \{a,,..\} \\ \vdots & \ddots & \vdots \\ \{a,b,\dots\} & \cdots & \emptyset \end{pmatrix}$$

$a, b$, ... are values of conditional attributes. This paper mainly focus on reduct generation using fuzzy-rough set model.

*Tab.3.5 Fuzzy-Rough Discernibility Matrix for the Decision System*

| Patient | $p_1$ | $p_2$ | $p_3$ | $p_4$ | $p_5$ | $p_6$ |
|---|---|---|---|---|---|---|
| $p_1$ | - | - | - | - | - | - |
| $p_2$ | - | - | - | - | - | - |
| $p_3$ | - | T, M, C, L | - | - | - | - |
| $p_4$ | - | - | T, M | - | - | - |
| $p_5$ | - | T, M, L | - | T, M, C, L | - | - |
| $p_6$ | - | - | T, M | - | T, M, L | - |

## III. Proposed Method

This paper proposes the algorithm for finding dynamic reduct by using Fuzzy-rough set model. The data set may possess many number of subset called reducts. The set which is having most stable features is called dynamic reduct. Fuzzy-rough set handles two types of uncertainty like both vagueness and indiscernibility. In this paper fuzzy-rough membership function is being proposed. Let $(\mathbb{U}, F_R)$ be the fuzzy approximation space where $F_R$ is the fuzzy relation on $\mathbb{U}$ . For any set $\mathcal{K} \subseteq F_R$, fuzzy lower and upper approximations of $\mathcal{K}$ is defined by

$$\underline{\mathcal{R}}(\mathcal{K})(t) = \bigwedge_{s\in\mathbb{U}}\{[\mathcal{K}(s) \vee (1-\mathcal{R}(t,s)] \quad t \in \mathbb{U}\}$$

$$\overline{\mathcal{R}}(\mathcal{K})(t) = \bigvee_{s\in\mathbb{U}}\{[\mathcal{K}(s) \wedge (\mathcal{R}(t,s)] \quad t \in \mathbb{U}\}$$

The pair $\left(\underline{R}(\mathcal{K}), \overline{R}(\mathcal{K})\right)$ is known as Fuzzy-Rough set of $\mathcal{K}$ with respect to the fuzzy similarity relation $F_R$.

Theorem 1: For $\mathbb{U}$ and $\emptyset$ be fuzzy universal and null set, let $\mathcal{K}, \mathbb{Q}_1, \mathbb{Q}_2 \subseteq F_R$, then

(i) $\underline{\mathcal{R}}(\mathbb{Q}_1) = \emptyset = \overline{\mathcal{R}}(\mathbb{Q}_1)$ & $\underline{\mathcal{R}}(\mathbb{Q}_1) = \mathbb{U} = \overline{\mathcal{R}}(\mathbb{Q}_1)$

Lower and Upper approximations of $\mathbb{U}$ and $\emptyset$ hold this property.

(ii) If $\mathbb{Q}_1 \subseteq \mathbb{Q}_2$, then $\underline{\mathcal{R}}(\mathbb{Q}_1) \subseteq \underline{\mathcal{R}}(\mathbb{Q}_2)$ & $\overline{\mathcal{R}}(\mathbb{Q}_1) \subseteq \overline{\mathcal{R}}(\mathbb{Q}_2)$

$$\underline{\mathcal{R}}(\mathbb{Q}_1)(t) = \bigwedge_{s\in\mathbb{U}}\{[\mathbb{Q}_1(s) \vee (1-\mathcal{R}(t,s)] \quad t \in \mathbb{U}\} \subseteq \underline{\mathcal{R}}(\mathbb{Q}_2)(t)$$

$$\overline{\mathcal{R}}(\mathbb{Q}_1)(t) = \bigvee_{s\in\mathbb{U}}\{[\mathcal{K}(s) \wedge (\mathcal{R}(t,s)] \quad t \in \mathbb{U}\} \subseteq \overline{\mathcal{R}}(\mathbb{Q}_2)(t)$$

Here Fuzzy-rough membership function for each attribute $a_i$ is defined as

$$\mu_{R(\mathcal{K})}(a_i) = \frac{\left|[a]_{R(\mathcal{K})} \cap R(\mathcal{K})\right|}{\left|[a]_{R(\mathcal{K})}\right|}$$

From the above expression positive region, dependency measures are described as follows

$$\mu_{POS}(R(\mathcal{K})) = Sup_{R(\mathcal{K})}[\mu_{R(\mathcal{K})}\underline{R}(\mathcal{K})(t)]$$

$$\vartheta_{R(\mathcal{K})}(D) = \frac{\sum \mu_{POS}(R(\mathcal{K}))}{|R(\mathcal{K})|}$$

The reduct which appears most frequently in all forms of schematic procedures is considered as most stable and this reduct is known as dynamic reduct $Dy_R$. The calculation of $Dy_R$ from an Information system is proposed in [14,15]. The purpose of selecting dynamic reduct is to get the stable reduct set. The process is completely comparative technique to all possible sets of relative reduct. The information system is consistent only when $\vartheta_{R(\mathcal{K})}(D) = \vartheta_{R(\mathcal{K})}(C) = 1$

Theorem 2:

Let $\vartheta_1, \vartheta_2$ are dependency measures of Fuzzy-rough sets $\mathcal{A}, \mathcal{B}$ with $\vartheta_1 \leq \vartheta_2$ then

$\underline{\mathcal{R}_{\vartheta_1}}(\mathcal{A}) \leq \underline{\mathcal{R}_{\vartheta_2}}(\mathcal{B})$ and $\overline{\mathcal{R}_{\vartheta_1}}(\mathcal{A}) \geq \overline{\mathcal{R}_{\vartheta_2}}(\mathcal{B})$

Proof: Since $\vartheta_1 \leq \vartheta_2$, $\vartheta_{R(\mathcal{K})}(\mathcal{A}) \leq \vartheta_{R(\mathcal{K})}(\mathcal{B})$

$$\frac{\sum \mu_{POS}(R(\mathcal{A}))}{|R(\mathcal{A})|} \leq \frac{\sum \mu_{POS}(R(\mathcal{B}))}{|R(\mathcal{B})|}$$
$$1 - \vartheta_{R(\mathcal{K})}(\mathcal{A}) \geq 1 - \vartheta_{R(\mathcal{K})}(\mathcal{B})$$

By the property of infimum and supremum $\underline{\mathcal{R}_{\vartheta_1}}(\mathcal{A}) \leq \underline{\mathcal{R}_{\vartheta_2}}(\mathcal{A})$ and $\overline{\mathcal{R}_{\vartheta_1}}(\mathcal{A}) \geq \overline{\mathcal{R}_{\vartheta_2}}(\mathcal{A})$

The following algorithm is proposed based on the above properties of fuzzy-rough set model for finding Dynamic reduct $Dy_R$

(i) Let $\mathbb{C}, \mathcal{D}$ be condition and decision attributes of an information system
(ii) $\mathfrak{R} \leftarrow 0$ $\vartheta_{R(\mathcal{K})}(D)(Best) = \vartheta_{R(\mathcal{K})}(D)(Previous) = 0$
(iii) $Q \leftarrow \mathfrak{R}$
(iv) $\forall a \in (\mathbb{C} - \mathfrak{R})$
(v) Add the features one by one until best and previous dependency measures are equal.
(vi) Let $\mathfrak{R}_i = \{R_1, R_2 \ldots . R_n\}$ be the family of reducts of given Information system from $A_t$
(vii) For each subsystem $\mathfrak{R}_i$ compute Reduct of each $A_t$
(viii) Calculate the Stability Coefficient of each Reduct by using

$$Dy_R = \left\{ \mathfrak{R}_i \middle| \frac{|\mathfrak{R}_i \subseteq A_t|}{|A_T|} < \varepsilon \right\}$$

Where $\varepsilon > 0$ is comparatively very small value.

(ix) If the Stability coefficient of $n^{th}$ attribute set is lesser than $(n+1)^{th}$ attribute set then update $(n+1)^{th}$ attribute set. Continue the process until all consistent attribute sets are selected with maximum stability coefficient. This algorithm selects one subset at a time.

## IV. Results and Discussions

For implementing the above algorithm, we have considered heart disease data set from UCI Repository heart disease data with 75 attributes like age, name, blood group, chest pain location … with following class distribution

Class Distribution:

| Knowledge base | 0 | 1 | 2 | 3 | 4 | Total |
|---|---|---|---|---|---|---|
| Australian: | 164 | 55 | 36 | 35 | 13 | 303 |
| Rome | 188 | 37 | 26 | 28 | 15 | 294 |
| South Africa: | 8 | 48 | 32 | 30 | 5 | 123 |
| Canada: | 51 | 56 | 41 | 42 | 10 | 200 |

Dynamic reducts for the above data set has been identified by using ROSETTA Technical toolkit.

The discernibility function for the above information system is

f(0) =

(age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,target
63,1,3,145,233,1,0,150,0,2.3,0,0,1,137,1,2,130,250,0,1,187,0,3.5,0,0,2,141,0,1,130,204,0,0,1
72,0,1.4,2,0,2,156,1,1,120,236,0,1,178,0,0.8,2,0,2,157,0,0,120,354,0,1,163,1,0.6,2,0,2,157,1,
0,140,….
end
f(1) =
(age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,target
63,1,3,145,233,1,0,150,0,2.3,0,0,1,137,1,2,130,250,0,1,187,0,3.5,0,0,2,141,0,1,130,204,0,0,1
72,0,1.4,2,0,2,156,1,1,120,236,0,1,178,0,0.8,2,0,2,157,0,0,120,354,0,1,163,1,0.6,2,0,2,157,1)
end

Attribute Ranking is done through Principal Component Attribute transformer which is the unsupervised learning algorithm. The corresponding correlation matrix is displayed as follows

*Tab.4.1 Correlation Matrix*

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -0.1 | -0.07 | 0.28 | 0.21 | 0.12 | -0.12 | -0.4 | 0.1 | 0.21 | -0.17 | 0.28 | 0.07 | -0.23 |
| -0.1 | 1 | -0.05 | -0.06 | -0.2 | 0.05 | -0.06 | -0.04 | 0.14 | 0.1 | -0.03 | 0.12 | 0.21 | -0.28 |
| -0.07 | -0.05 | 1 | 0.05 | -0.08 | 0.09 | 0.04 | 0.3 | -0.39 | -0.15 | 0.12 | -0.18 | -0.16 | 0.43 |
| 0.28 | -0.06 | 0.05 | 1 | 0.12 | 0.18 | -0.11 | -0.05 | 0.07 | 0.19 | -0.12 | 0.1 | 0.06 | -0.14 |
| 0.21 | -0.2 | -0.08 | 0.12 | 1 | 0.01 | -0.15 | -0.01 | 0.07 | 0.05 | 0 | 0.07 | 0.1 | -0.09 |
| 0.12 | 0.05 | 0.09 | 0.18 | 0.01 | 1 | -0.08 | -0.01 | 0.03 | 0.01 | -0.06 | 0.14 | -0.03 | -0.03 |
| -0.12 | -0.06 | 0.04 | -0.11 | -0.15 | -0.08 | 1 | 0.04 | -0.07 | -0.06 | 0.09 | -0.07 | -0.01 | 0.14 |
| -0.4 | -0.04 | 0.3 | -0.05 | -0.01 | -0.01 | 0.04 | 1 | -0.38 | -0.34 | 0.39 | -0.21 | -0.1 | 0.42 |
| 0.1 | 0.14 | -0.39 | 0.07 | 0.07 | 0.03 | -0.07 | -0.38 | 1 | 0.29 | -0.26 | 0.12 | 0.21 | -0.44 |
| 0.21 | 0.1 | -0.15 | 0.19 | 0.05 | 0.01 | -0.06 | -0.34 | 0.29 | 1 | -0.58 | 0.22 | 0.21 | -0.43 |
| -0.17 | -0.03 | 0.12 | -0.12 | 0 | -0.06 | 0.09 | 0.39 | -0.26 | -0.58 | 1 | -0.08 | -0.1 | 0.35 |
| 0.28 | 0.12 | -0.18 | 0.1 | 0.07 | 0.14 | -0.07 | -0.21 | 0.12 | 0.22 | -0.08 | 1 | 0.15 | -0.39 |
| 0.07 | 0.21 | -0.16 | 0.06 | 0.1 | -0.03 | -0.01 | -0.1 | 0.21 | 0.21 | -0.1 | 0.15 | 1 | -0.34 |
| -0.23 | -0.28 | 0.43 | -0.14 | -0.09 | -0.03 | 0.14 | 0.42 | -0.44 | -0.43 | 0.35 | -0.39 | -0.34 | 1 |

*Tab.4.2 Eigen Values*

| eigenvalue | proportion | cumulative | Decision Rule |
|---|---|---|---|
| 3.30148 | 0.23582 | 0.23582 | 0.438target-0.37oldpeak+0.366thalach 0.336exang+0.325slope... |
| 1.57216 | 0.1123 | 0.34812 | 0.445ïage+0.443trestbps- 0.392sex+0.359chol+0.306fbs... |
| 1.23203 | 0.088 | 0.43612 | 0.56 sex+0.468fbs+0.305ca+0.267thalach+0.241thal... |
| 1.20664 | 0.08619 | 0.52231 | -0.509chol-0.498slope+0.379cp+0.37oldpeak- 0.236thal... |
| 1.02203 | 0.073 | 0.59531 | -0.432ca 0.397restecg+0.337thal+0.322thalach+0.321chol.. |
| 0.97017 | 0.0693 | 0.66461 | 0.668restecg+0.5thal- 0.253fbs0.237exang+0.198cp... |
| 0.86323 | 0.06166 | 0.72627 | -0.515fbs-0.46exang-0.394restecg+0.332ca- 0.326trestbps... |
| 0.77609 | 0.05544 | 0.7817 | -0.605trestbps+0.435fbs+0.375chol- 0.261slope+0.256ca... |
| 0.72919 | 0.05209 | 0.83379 | -0.429ca+0.396age-0.388thalach+0.346thal- 0.342trestbps... |
| 0.62272 | 0.04448 | 0.87827 | -0.536thal+0.53 sex+0.502chol+0.279restecg+0.182cp... |
| 0.5344 | 0.03817 | 0.91644 | 0.649exang+0.536cp+0.348ca- 0.239sex+0.164slope... |
| 0.43003 | 0.03072 | 0.94716 | -0.556thalach-0.525age-0.379oldpeak- 0.277target-0.236exang... |
| 0.37188 | 0.02656 | 0.97372 | -0.636slope- 0.615oldpeak+0.334thalach+0.192ï»¿age+0.149ex ag... |

*Tab.4.3 Eigen Vectors*

| V 1 | V 2 | V 3 | V 4 | V 5 | V 6 | V 7 | V 8 | V 9 | V10 | V11 | V12 | V13 | Attr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| -0.254 | 0.4447 | -0.0669 | -0.0664 | -0.3067 | 0.1264 | 0.2171 | -0.2518 | 0.3961 | -0.0077 | -0.1305 | -0.5248 | 0.1923 | age |
| -0.122 | -0.3916 | 0.5596 | 0.0812 | 0.0561 | -0.062 | 0.1509 | -0.1665 | 0.2301 | 0.53 | -0.2392 | -0.0831 | 0.0464 | sex |
| 0.2746 | 0.2676 | 0.2273 | 0.3788 | 0.1594 | 0.1984 | 0.219 | 0.0515 | 0.3123 | 0.1822 | 0.536 | 0.2227 | 0.0234 | cp |
| -0.1474 | 0.4434 | 0.1989 | 0.0898 | 0.1869 | 0.1813 | -0.3257 | -0.6045 | -0.3419 | 0.0593 | -0.049 | 0.2098 | 0.0093 | trestbps |
| -0.0922 | 0.359 | -0.1807 | -0.5093 | 0.3214 | 0.1034 | -0.0524 | 0.3746 | 0.1293 | 0.5023 | -0.1022 | 0.154 | -0.0141 | chol |
| -0.0529 | 0.306 | 0.468 | 0.1583 | -0.231 | -0.2534 | -0.5151 | 0.4354 | 0.1598 | -0.1418 | -0.1553 | -0.0174 | -0.1251 | fbs |
| 0.1119 | -0.2133 | -0.203 | 0.1902 | -0.3971 | 0.6684 | -0.3944 | 0.0968 | 0.0054 | 0.2789 | -0.1012 | 0.0135 | 0.0767 | restecg |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.3664 | -0.0022 | 0.2674 | -0.1265 | 0.3225 | 0.1234 | -0.0942 | 0.1329 | -0.3879 | 0.0218 | -0.0171 | -0.5558 | 0.3336 | thalach |
| -0.3357 | -0.2051 | -0.1223 | -0.1004 | 0.0391 | -0.2374 | -0.4598 | -0.1064 | 0.1046 | 0.1737 | 0.6486 | -0.2357 | 0.1488 | exang |
| -0.3698 | 0.026 | -0.0885 | 0.3699 | 0.2494 | 0.1712 | 0.1126 | 0.1897 | -0.2141 | 0.0921 | 0.0292 | -0.3794 | -0.6153 | oldpeak |
| 0.3255 | -0.0395 | 0.1958 | -0.4978 | -0.2444 | 0.0611 | -0.0569 | -0.2607 | 0.0357 | 0.0546 | 0.1637 | -0.0948 | -0.6359 | slope |
| -0.2616 | 0.0932 | 0.3046 | -0.1925 | -0.4321 | 0.1772 | 0.332 | 0.2565 | -0.4293 | -0.0048 | 0.3484 | 0.0671 | 0.1399 | ca |
| -0.2226 | -0.1921 | 0.2411 | -0.2358 | 0.3375 | 0.5001 | -0.073 | 0.0487 | 0.3458 | -0.5355 | 0.0484 | 0.035 | 0.0072 | Thal |
| 0.4383 | 0.1511 | -0.1212 | 0.1164 | 0.0077 | -0.013 | -0.0264 | 0.0178 | 0.1449 | -0.0566 | 0.1377 | -0.2771 | -0.0496 | Target |
| … . | … . . | … . . | … . | … | … | … | … | … | … | … | … | … | . . |

*Tab.4.4 Ranked Attributes*

| | | |
|---|---|---|
| 0.7642 | 1 | 0.438target-0.37oldpeak+0.366thalach-0.336exang+0.325slope... |
| 0.6519 | 2 | 0.445age+0.443trestbps-0.392sex+0.359chol+0.306fbs... |
| 0.5639 | 3 | 0.56 sex+0.468fbs+0.305ca+0.267thalach+0.241thal... |
| 0.4777 | 4 | -0.509chol-0.498slope+0.379cp+0.37 oldpeak-0.236thal... |
| 0.4047 | 5 | -0.432ca-0.397restecg+0.337thal+0.322thalach+0.321chol... |
| 0.3354 | 6 | 0.668restecg+0.5 thal-0.253fbs-0.237exang+0.198cp... |
| 0.2737 | 7 | 0.515fbs-0.46exang-0.394restecg+0.332ca-0.326trestbps... |
| 0.2183 | 8 | -0.605trestbps+0.435fbs+0.375chol-0.261slope+0.256ca... |
| 0.1662 | 9 | -0.429ca+0.396age-0.388thalach+0.346thal-0.342trestbps... |
| 0.1217 | 10 | -0.536thal+0.53 sex+0.502chol+0.279restecg+0.182cp... |
| 0.0836 | 11 | 0.649exang+0.536cp+0.348ca-0.239sex+0.164slope... |
| 0.0528 | 12 | -0.556thalach-0.525ï»¿age-0.379oldpeak-0.277target-0.236exang... |
| 0.0263 | 13 | -0.636slope-0.615oldpeak+0.334thalach+0.192ï»¿age+0.149exang... |
| …… | …….. | ………… |

k-Means clustering:
Number of iterations: 5
Initial starting points (random):
Cluster 0: 53,0,2,128,216,0,0,115,0,0,2,0,0,1
Cluster 1: 61,1,0,140,207,0,0,138,1,1.9,2,1,3,0

*Tab.4.5 Final Cluster Centroids*

| Attribute | Full Data (303.0) | Cluster #0 (163.0) | Cluster #1 (140.0) |
|---|---|---|---|
| age | 54.3663 | 52.4724 | 56.5714 |
| sex | 0.6832 | 0.5583 | 0.8286 |
| cp | 0.967 | 1.3926 | 0.4714 |
| trestbps | 131.6238 | 129.4663 | 134.1357 |
| chol | 246.264 | 242.3129 | 250.8643 |
| fbs | 0.1485 | 0.1411 | 0.1571 |
| restecg | 0.5281 | 0.5951 | 0.45 |
| thalach | 149.6469 | 158.8589 | 138.9214 |
| exang | 0.3267 | 0.1288 | 0.5571 |
| oldpeak | 1.0396 | 0.5706 | 1.5857 |
| slope | 1.3993 | 1.6012 | 1.1643 |
| ca | 0.7294 | 0.362 | 1.1571 |
| thal | 2.3135 | 2.1166 | 2.5429 |
| target | 0.5446 | 1 | 0.0143 |
| …… | …….. | ………… | ………….. |

Classical reducts

*Tab.4.6 Classical reducts from Discernibility Matrix*

| S.No | Reduct | Support | Length |
|---|---|---|---|
| 1 | age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,target 63,1,3,145,233…….. | 100 | 1 |
| 2 | age,sex,cp,trestbps,chol,fbs,restecg,thalach,exang,oldpeak,slope,ca,thal,target63,1,3,145,233,1,0,………………………. | 100 | 1 |

## 4.1 Construction of dynamic reducts for this Information System

Dynamic reducts formed through the process of comparative analysis process by taking sub-tables from an input. The reduct which occurs frequently in all sub-tables is called stable or dynamic reduct.

In this experiment first conventional reducts of fuzzy-rough set are being calculated. After the creation of conventional reducts possible number of sub information table is created. For each sub information table reducts have to be calculated. Then stability coefficient for each sub-table reduct is calculated and the reducts with maximum stability coefficient are dynamic and stable reducts which are displayed as follows

*Tab 4.7 Stability co-efficient of*

| Attribute | Stability Co-efficient (303.0) |
|---|---|
| age | 0.734 |
| sex | 0.658 |
| cp | 0,987 |
| trestbps | 0.9443 |
| chol | 0.989 |
| fbs | 0.978 |
| restecg | 0.928 |
| thalach | 0.956 |
| exang | 0.945 |
| oldpeak | 0.740 |
| slope | 0.756 |
| ca | 0.973 |
| thal | 0.961 |
| target | 0.856 |
| …… | …….. |

*Tab.4.8 Dynamic Reducts by using fuzzy-rough set model*

| S.No | Dynamic Reduct | Support | Co-efficient of Stability |
|---|---|---|---|
| 1 | cp,trestbps, chol, fbs, restecg, thalach, exang, ca | 100 | 0.96 |

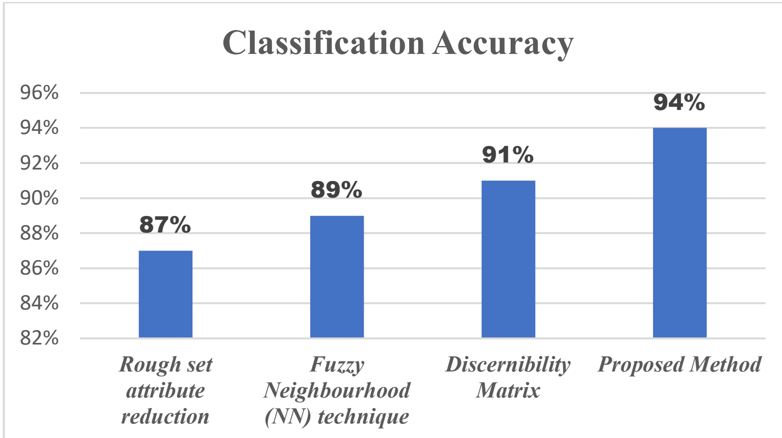
Hence the most efficient variables are cp,trestbps, chol, fbs, restecg, thalach, exang, ca (8 Attributes)

The following table exhibits the comparative results between existing and proposed method

*Tab.4.9 Comparative Results*

| **Name of the Intelligent System** | **Number of Subsets constructed** | **Number of attribues selected** |
|---|---|---|
| Rough set attribute reduction | 768 | 14 |
| Fuzzy Neighbourhood (NN) technique | 645 | 13 |
| Discernibility Matrix | 867 | 13 |
| Proposed Method | 1056 | 8 |

The following graph explains the classification accuracy




## V. Results and Conclusion

The work presented in this paper brings out the importance of dynamic reduct -the attributes with more stability through the dependency measure of fuzzy-rough sets. These reducts are constructed by applying absorption and distributive laws in dicernibility function. Origination of reducts plays significant lead in medical domain to remove the irrelevant attributes. When compared with classical reduct generating algorithm Fuzzy-Rough set model dynamic reduct generation gives minimum number of reducts which contributes automatic approach in medical data science in decision making. In future this work may be extended to derive new approximation space based on fuzzy-rough distance metric and this metric may be defined in this algorithm as well as implemented for some benchmark data sets.

## References

1. Garg, H.; Kumar, K: Distance measures for connection number sets based on set pair analysis and its applications to decision-making process. Appl. Intell., 48, 3346–3359. (2018)
2. Chuan-Yang Ruan: Hesitant Fuzzy Decision-Making Method Based on Correlation Coefficient under Confidence Levels with Application to Multi sensor Electronic Reconnaissance", *Mathematical* Problems in Engineering, vol. 2020, Article ID 9315769, https://doi.org/10.1155/2020/9315769. (2020)
3. Chen, Shu-Jen and Ching-Lai Hwang. Fuzzy Multiple Attribute Decision Making - Methods and Applications. Lecture Notes in Economics and Mathematical Systems. (1992)
4. P.K. Maji, A.R. Roy, R. Biswas, An application of soft sets in a decision making problem, Computers & Mathematics with Applications, Volume 44, Issues 8–9, Pages 1077-108. (2002)
5. Z. Pawlak, Rough Sets: Theoretical aspects of reasoning about data, Kluner Academic, Boston, MA. (1991).

6. Z. Pawlak, Hard set and soft sets, ICS Research Report. Poland, (1994).
7. Felix, R.; Ushio, T. Rough sets-based machine learning using a binary discernibility matrix. In Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials, Honolulu, HI, USA, 10–15 July.1999; pp. 299–305. (1999)
8. Nan Zhang, Baizhen Li, Zhongxi Zhang, Yanyan Guo, A Quick Algorithm for Binary Discernibility Matrix Simplification using Deterministic Finite Automata, Information 2018, 9, 314; doi:10.3390/info9120314, (2018)
9. Skowron, A.; Rauszer, C. The discernibility matrices and functions in information systems. In Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory; Słowi'nski, R., Eds.; Kluwer Academic Publishers: Dordrecht, Holland, pp. 331–362. (1992)
10. Y.Y. Yao and Y. Zhao, Discernibility matrix simplification for constructing attribute reducts, Information Sciences, Vol. 179, No. 5, 867-882. (2009)
11. M.Srivenkatesh, P.V.G.D.Prasad reddy, Y.Srinivas, Dynamic Reduct and Its Properties In the Object-Oriented Rough Set Models, International Journal of Advanced Computer Science and Applications, Vol. 1(4). (2010)
12. Walid Moudani, Ahmad Shahin, Fadi Chakik, Félix Mora-Camino, Dynamic Rough Sets Features Reduction, International Journal of Computer Science and Information Security, Vol. 9(4). (2011)
13. Carine Pierrette Mukamakuza, Jiayang Wang, Li Li, Dynamic Reducts Computation Analysis Based on Rough Sets, 10th International Conference on Natural Computation, 978-1-4799-5151-2/14, IEEE. (2014)
14. Jan G. Bazan. A comparison of dynamic and non-dynamic rough set methods for extracting laws from decision tables. In Polkowski and Skowron [169], chapter 17, pages 321–365. (1998)
15. K. Anitha and M. Thangeswari, Rough Set Based Optimization Techniques, 2020 6th International Conference on Advanced Computing and Communication Systems *(ICACCS)*, pp. 1021-1024, doi: 10.1109/ICACCS48705.2020.9074212.(2020)
16. Anitha, K, Aruna Devi, R, Munir, M, Nisar, K.S, Metric dimension of rough graphs, International Journal of Nonlinear Analysis and Applications, Vol.12, pp: 1793-1806.(2021)
17. Anitha, K. Rough set theory on topological spaces, Lecture Notes in Computer Science 2014, Vol.8818, pp:69-74 (2014)
18. D. Dubois and H. Prade, Putting rough sets and fuzzy sets together. Intelligent Decision Support, Kluwer Academic Publishers, Dordrecht (1992)
19. Weidong Tang, Jinzhao Wu, Dingwei Zheng, On Fuzzy Rough sets and their topological structures, Mathematical Problems in Engineering, Vol.2014, Article ID 546372. (2014)
20. D. Dubois and H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General Systems, vol. 17, pp. 191–208 (1990)

21. S. Nanda and S. Majumdar, Fuzzy rough sets, Fuzzy Sets and Systems, vol. 45, no. 2, pp. 157–160 (1992)
22. A. M. Radzikowska and E. E. Kerre, A comparative study of fuzzy rough sets, *Fuzzy Sets and Systems*, vol. 126, no. 2, pp. 137–155. (2002)
23. W.-Z. Wu, J.-S. Mi, and W.-X. Zhang, Generalized fuzzy rough sets, *Information Sciences*, vol. 151, pp. 263–282 (2003)
24. Mohammed Atef, A. A. Azzam, Covering Fuzzy Rough Sets via Variable Precision, Journal of Mathematics, vol. 2021, Article ID 5525766 (2021)
25. Seema Singh, D.S. Hooda, S.C. Malik, Fuzzy Rough Information Measures and Their Applications, International Journal of Computational Science and Information Technology (IJCSITY) Vol. 9, No. 4. (2021).
26. Kopczynski, M., Grzes, T. FPGA supported rough set reduct calculation for big datasets. J Intell Inf Syst. https://doi.org/10.1007/s10844-022-00725-5. (2022)
27. Zhouming Ma, Jusheng Mi, Yiting Lin, Jinjin Li, Boundary region-based variable precision covering rough set models, Information Sciences, Vol. 608, pp:1524-1540. (2022).
28. Mohammed Atef, José Carlos R. Alcantud, Hussain AlSalman, Abdu Gumaei, On Some Types of Covering-Based $(\mathcal{J},\mathcal{F})$Fuzzy Rough Sets and Their Applications, Journal of Mathematics, vol. 2021, Article:ID 5240501. (2021).
29. Olha Kaminska, Chris Cornelis, Veronique Hoste, LT3 at SemEval-2022 Task 6: Fuzzy-Rough Nearest Neighbor Classification for Sarcasm Detection, Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022), pages 987- 992 July 14-15, 2022 ©Association for Computational Linguistics
30. Rahul Kumar, Sandeep Kumar Gupta, A Survey On Attributes Reduction Techniques Based On Fuzzy Rough Sets, International Journal of Education, Modern Management, Applied Sciences and Social Sciences, Vol.4(1),pp: 13-26, (2022).