ATTRIBUTE SUBSET SELECTION BASED ON FUZZY ROUGH SET AND RANKING APPROACH

G. Uma Mageswari¹, Dr. M. Indra Devi²

Address for Correspondence

¹Assistant Professor, CSE, Bharath Niketan Engineering College, Aundipatti, Theni, Tamil Nadu, India.
²Professor & Head of the Department / CSE, Kamaraj College of Engineering & Technology, Virudhunagar.

ABSTRACT

In the real world, the database contains a large number of attributes and data. The attribute reduction is one of key processes for knowledge acquisition. Attribute reduction or feature selection is employed for dimensionality reduction and the intention is to select a subset of the unique features of a data set, which have more useful information. A rough set feature selection technique uses the information from both the lower approximation dependency value and a distance metric. It also considers the number of objects in the boundary region and the distance of those objects from the lower approximation. Though it is more efficient in mining, the rough set suffers from intensive computation of either discernibility functions or the positive region to find attribute reduction. To overcome the aforementioned problem this paper, proposes an attribute subset selection based on the Fuzzy rough set and distance based ranking approach, which works progressively. Fuzzy rough set feature selection technique uses the information from both the lower and upper approximations in fuzzy preference relation. Additionally, authors discuss a fuzzy preference relation algorithm to create relations from data and integrated fuzzy relation with fuzzy rough sets. They have constructed an algorithm for analyzing the significance and dependency. Furthermore, lower and upper approximations in fuzzy preference investigation can be interpreted as the optimistic decision and pessimistic.

INDEX TERMS — Data Mining, Rough sets, Attribute Reduction, Classification, Fuzzy Sets, Mahalanobis Distance and Boundary Region.

I. INTRODUCTION

In recent years, attribute reduction has become the focus and hotspot of research in the field of Rough Set. It has been widely used in data mining, intelligent control, machine learning, and in numerous fields. The fundamental aspects of rough set theory for knowledge acquisition involves the searching for some particular subsets of condition attributes. Through one subset the information for classification purposes, provided is the same as the condition attribute set done. Such subsets are called reductions. Attribute reduction is an important issue of rough set theory, moreover imperative preprocessing step in data mining, which aimed at finding a small set of rules from the data set with determined target. Subsequently, it concerned an important position in intellectual information processing. In this paper, we propose an attribute reduction based on the fuzzy rough set and ranking based methods in which the knowledge base under the principle of the same classification ability is used to remove the irrelevant and redundant attribute properties.

Rough set theory is one of the effective methods for attribute selection, which conserved the meaning of the attributes and was proposed in [1]. The entity of rough set approaching to attribute reduction is the process of selecting a subset of attributes from the original set of features forming patterns in a given dataset. The subset should be essential and sufficient to describe the concepts and maintaining a suitably high accuracy in representing the original attributes. Moreover, different dimensionality reduction approach minimal attributes reduction conserve the original meaning of the attributes after reduction. It provides a mathematical foundation for feature selection. This feature selection is called reduce. Moreover, there is a drawback in rough set theory, and to overcome the aforementioned problem this paper, an attribute subset selection based on the Fuzzy rough set and distance based ranking approach is proposed, which works progressively. Fuzzy rough set is used to avoid loss of information due to discretization. It reduces the search space and improve efficiency. Furthermore, the distance metric effort to qualify the objects in the boundary region with regard to their adjacency to the lower approximation. Therefore, all the distances of objects in the boundary region are calculated. Hence, the significance value for a subset can be acquired. The authors [2] discuss a fuzzy preference relation algorithm to create relations from data and integrated fuzzy relation with fuzzy rough sets. They have constructed an algorithm for analyzing the significance and dependency. Furthermore, lower and upper approximations in fuzzy preference investigation can be interpreted as the optimistic decision and pessimistic.

Additionally, this paper proposes a Mahalanobis distance based on the correlations between variables by that different patterns can be recognized and analyzed. The advantage of Mahalanobis Distance, takes into consideration the correlations between the variables and this consideration is very significant in pattern analysis. It measures a similarity of an unknown sample set to an identified one. The work of attribute reduction is to remove redundant or unrelated attributes while maintaining the classification ability of knowledge base. It reduces the investigation space and progress the effectiveness of data mining algorithm.

The remainder of this paper is structured as follows, Section II confers about the related works and Section III describes our proposed attribute selection on fuzzy rough set feature selection, Section IV describes the experiments and results achieved. Finally, we present the conclusion and future enhancements of the added work.

II. RELATED WORK

Some of the notable research efforts have been made in previous works. The authors [3] discussed about a rough set based methodology, which efficiently inducted the recruitment rules. Furthermore, the weight of all input attributes are included in the proposed approach so as to improve the quality of the resulting rules. Two classical measures, namely, consistency, degree and approximation accuracy can be employed to estimate the decision table. These two measures cannot complicate the depictions of certainty and consistency. To conquer this shortcoming, authors[4] classified the measures into three types according to their consistencies and three effectual measures ($\alpha$, $\beta$, $\gamma$) introduced for evaluating consistency, certainty and support of a decision rule.
These three new measures depends on the decision granulation and condition granulation. Paper [5], described the attribute reduction in decision theoretic rough set model concerning different classification properties as confidence, coverage, decision monotocity, cost and generality. Hence, it is important for several properties that can be truthfully reflected by a unique measure in the Pawlak rough set model. Furthermore, it can be considered discretely in a probabilistic model. A non-deterministic value in the system is a set of potential values of the attribute for the entity. Certainly such an incomplete information system can be transformed into a set-valued information system was given in [6]. Furthermore, paper attributes reduction the Dempster Shafer assumption of confirmation in incomplete information systems and incomplete decision systems.

The authors [7, 8] discussed about a rough set framework and was called the positive approximation. The main benefit of this approach stems from the fact that this framework is able to differentiate the granulation structure of a rough set using a granulation order. It was based on the positive approximation and they developed an ordinary accelerator for improving the time efficiency of a heuristic attribute reduction that provides a vehicle of production algorithms of rough set based on feature selection methods. Through incorporating the accelerator into each of the above four representative heuristic attribute reduction methods. Author[9] proposed an easy and effective hybrid attribute reduction algorithm based on a generalized fuzzy rough model. A theoretic framework of fuzzy rough model based on fuzzy relations is accessible, which contain a foundation for algorithm construction. Though, several attributes derive significance measures based on the proposed fuzzy-rough model and create a forward greedy algorithm for hybrid attribute reduction.

![Fig. 1. Attribute Selection Process](image)

In [2] proposed a weighted distance learning algorithm for feature selection via maximizing fuzzy dependency. It maximized the fuzzy dependency between features and decide by distance learning and subsequently evaluate the quality of features with the learned weights vector. The features deriving greater weights are considered to be useful for classification learning. A feature selection method based on mutual information is proposed in [10] to select a set of genes from microarray gene expression data. It employed a fuzzy rough set theory to calculate the significance and implication of the genes that enormously decreases the computational complication while keeping high predictive accuracy. Moreover, the proposed approach is compared with that of conventional fuzzy rough approach and rough set method using the predictive accuracy of support vector machine and 1-nearest neighbor rule on dissimilar microarray data sets. A structure for fuzzy-rough set based feature selection built up around the formal notion of a fuzzy decision reduct was given in [11]. The attribute subset should retain the quality of the feature set to a certain extent, which are able to create a shorter attribute subsets. Moreover, they have provided a comprehensive classification evaluation measures that can be used to define fuzzy decision reduct. Jensen [12] summarized the new approaches to fuzzy rough feature selection that are capable of dealing with imprecision and uncertainty. Consequently, it is desirable to hybridize and extend the data imperfection. Such developments offer a high degree of flexibility and provide robust solutions and advanced tools for data analysis. Fuzzy-rough set-based feature selection was exposed to be extremely useful at reducing data dimensionality, but possesses numerous problems that render it ineffective for large datasets. Moreover, the paper proposed three new methods for fuzzy rough feature selection based on fuzzy similarity relations. Richard Jensen and Qiang Shen [13] summarized the approach based on fuzzy rough set and fuzzy rough feature selection, which can address the retain dataset semantics and problems. Additionally, it can be applied two challenging domains in feature reduction, namely, complex system monitoring and classification. Consequently, this approach was demonstrated and compared with numerous dimensionality reducers. In [14] proposed a wasp swarm optimization algorithm for attribute reduction based on rough set and the consequence of the feature. Moreover, it's based on the mutual information between decision attributes and conditional attributes. Then, the algorithm dynamically calculates heuristic information based on the implication of feature to guide search. Another algorithm, namely, Genetic Algorithms and swarm-based approaches have been tried out for feature selection in order for these intentions. A reduction relative to discernibility set is considered to reduce the time complexity of the attribute reduction algorithm based on a traditional algorithm was given in[15]. Iterative knowledge is used to select the required condition attributes to this reduction relative. Moreover, a new approach named Particle Swarm Optimization (PSO) has been confirmed to be competitive with genetic algorithm in numerous tasks mostly in optimal areas[16]. In addition, it optimized the reduction relative by complete optimization theory and resulted in a Pawlak reduction. Though, there are several shortcomings in PSO such as premature convergence. Furthermore, the new algorithm has been proposed, namely, Intelligent Dynamic Swarm (IDS) that is a customized particle swarm optimization.

**III. PROPOSED WORK**

In our proposal, we present an attribute subset selection technique based on the Fuzzy rough set and distance based ranking approach that works gradually. Attribute reduction continue the knowledge base under the principle of the same classification ability to remove irrelevant and redundant attribute properties. Fig 2 depicts the overflow of proposed work. The datasets used in the proposed approach is the UCI credit card dataset[17]. This dataset is interesting because there is a good mix of attributes. The dataset characteristics are multivariate in nature and the characteristics of
attributes are categorical, integer, and Real sets. There are several dataset in the literature, which includes the Wisconsin diagnostic breast cancer (Wdbc), Wisconsin prognostic breast cancer (Wpbc). The proposed Fuzzy Rough set based feature selection uses the credit card dataset for processing the information.

### 3.1 Attribute Reduction

Attribute reduction is a data reduction technique, which is frequently used as a preprocessing step in data mining. It is used to remove the irrelevant and redundant attributes, properties, thereby reducing the search space and improves the efficiency. In addition, attribute reduction is the fundamental concepts of rough set theory. To solve the problem of attribute selection or attribute reduction, the rough set theory. To solve the problem of attribute reduction is the fundamental concepts of

### 3.2 Approximations of Sets

Approximation of sets concepts in rough set theory recommends $X$ is a finite and nonempty set is called the universe and equivalence relation $Z$ partitions the set $X$ into disjoint subsets. Moreover, there are two operations defined for assigning the values to every set called $Y_{-lower}$ approximation and $Y_{-upper}$ approximation of $A$ .

Let $Z=(X,Y)$ be an approximation space and $A$ be a subset of $X$ .

The lower approximation of $A$ by $Y$ in $Z$ is defined as, given in equation 1

$$Y_A = \{e \in X | [e] \subseteq A\}$$  (1)

The upper approximation of $A$ by $Y$ in $Z$ is defined as shown in the equation 2

$$\overline{Y}_A = \{e \in X | [e] \cap A \neq \emptyset\}$$  (2)

Where $[e]$ denotes the equivalence class containing $e$ . A subset $A$ of $X$ is said to be $Y$ definable in $Z$ if and only if $Y_A = \overline{Y}_A$ . The boundary sets $SN_i(A)$ are

---

*Fig. 2. Architecture of Proposed Methodology*

- The lower approximation of $A$ by $Y$ in $Z$ is defined as, given in equation 1
- The upper approximation of $A$ by $Y$ in $Z$ is defined as shown in the equation 2

Where $[e]$ denotes the equivalence class containing $e$ . A subset $A$ of $X$ is said to be $Y$ definable in $Z$ if and only if $Y_A = \overline{Y}_A$ . The boundary sets $SN_i(A)$ are...
defined as \( \overline{Y_A} \). Moreover, in rough set, the pair \( (Y_A, \overline{Y_A}) \) are called \( Y - lower \) approximation and \( Y - upper \) approximation with respect to \( Y \), respectively. The ordered pair \( (Y_A, \overline{Y_A}) \) is a rough set of \( A \) with respect to the approximation relation \( Y \).

A common notion described by a set is always characterized through the so-called upper and lower approximations under static granulation in rough set theory and a constant boundary region of the concept is induced by the upper and lower approximations.

### 3.3 Attribute Reduction Based on Positive Region

A positive region reduction is required to maintain the similar classification ability as the whole attribute set. Then, it is superfluous to consider that the positive-region may increase with the deletion of attributes. In attribute reduction based on positive region, a decision system \( T = \{ U, A \cup B, S, V \} \).

Where, \( A \) is defined as the set of condition attribute. Moreover, the positive region of decision attributes under the condition attribute can be defined as \( \text{POT}(B) \). Then, \( P \) is a subset of condition attributes set \( A \). In an attribute reduction conditions there are two steps,

- \( \forall x \in P \), the \( x \) is excluded from \( P \) if the attributes need to be reduced.

- \( \text{POT}_p(B) = \text{POT}_A(B) \), subsequently \( P \) is the attribute reduction of \( A \).

In rough set theory, the concept of relative core and reduction are both based on the positive region. The comparative attribute dependency degree can be intended by counting the different instances of the subset of the data set, generating positive regions. Subsequently, the computational efficiency of finding minimum reduce is highly enhanced. The main objective of the rough set based feature selection is used to find subset of attributes, which has some particular properties as that of the original data and it is without redundancy. There are several issues for a decision table and is proven to be an NP hard problem. In order to overcome the drawbacks of the decision table, the heuristic search algorithms are proposed, which includes the greedy and forward search algorithms. The below algorithm describes the improved feature selection algorithm, which is based on the forward approximation.[19]

A fuzzy equality relation would accomplish by a real valued fuzzy attribute, as an alternative of crisp equivalence relation. The fuzzy-rough set model is fitted for both the object subset and relation to be approximated are fuzzy. The fuzzy rough set method is an effective approach to data mining and knowledge discovery from hybrid data, which includes the categorical values and numerical values. It is used for feature selection to support the efficient attribute reduction and many heuristic attribute reduction methods is also developed. The attribute reduction strategy use the forward greedy searching technique. In this kind of attribute reduction technique, the important measures of attributes are used for heuristic functions, which is used in the forward feature selection. The significance measures of attributes provide some heuristic algorithms and the reduced attributes are needed. These algorithms are very time consuming. The main objective of the rough set based feature selection is used for finding the subset attributes, which retained some particular properties as the original data and without redundancy.

The fuzzy-rough set is defined by two sets namely, the fuzzy lower approximations and fuzzy upper approximations, obtained by extending the parallel crisp rough set. Moreover, crisp case elements that belong to the lower approximation are said to belong to the approximated set with complete certainty. During the fuzzy rough case, elements may have a membership in the range and allowing greater flexibility in handling ambiguity. The relation and partition induced by these attributes are fuzzy. A set of fuzzy input and output attributes \( \{ T_1, T_2, \ldots, T_n \} \) and \( d \) is the objects of \( S = \{ a_1, a_2, \ldots, a_m \} \). All attributes are restricted to a small set of fuzzy semantic term \( \lambda(T_i) = \{ G_j | i = 1,2,3,\ldots,C \} \). Every object \( a_j \in S \) is classified by a set of classes \( A(R) = \{ G_j | i = 1,2,3,\ldots,C_k \} \).

Where, \( R \) is decision attribute.
$G_j$ is the fuzzy set.

Moreover, the fuzzy partition is defined as $S_k^{(a)} = \{G_j, j = 1, \ldots, P, k = 1, \ldots, C\}$. In our proposal, fuzzy membership functions can be defined and equation given in 3 and 4.

$$SX(G_j) = \text{sup}_{a \in A} \text{max} \{1 - SG_j(a), SX(a)\}$$

(3)

$$ST(R) = \sum_{i=1}^{P} ST_{s}(a)$$

(4)

Then, the positive region of fuzzy set $G_j$ is the maximal membership degree, which is ordinary and class can be classified as fuzzy set $G_j$, shown in equation 5.

$$SPOT(G_j) = \sup_{a \in A} \text{max} \{SG_j(a), SX(a)\}$$

(5)

Hence, the membership of $A$ to the fuzzy positive region and equations are given in equation 6. Furthermore, the fuzzy positive region can calculate the confidence function as shown in equation 7.

$$SPOT(a) = \sup_{a \in A} \text{min} \{SG_j(a), SX(a)\}$$

(6)

$$ST(R) = \sum_{i=1}^{P} ST_{s}(a)$$

(7)

### 3.5 Mean Positive Regions and Distance Metric

The distance metric effort is to qualify the objects in the boundary region with regard to their adjacency to the lower approximation. From this point-of-view, the closer the proximity of an object in the boundary region to objects of the lower approximation. Therefore, all the boundaries of objects in the boundary region are calculated. Hence, the significance value for a subset can be acquired. Since, calculating the lower approximation for a n-dimensional space involves the computation efforts and all the mean value in positive region $POT_p$ is calculated from equation 8.

$$\gamma_p = \frac{\text{pot}_{p}(a)}{|U|}$$

(8)

Where, $|U|$ denotes the cardinality of a set and 0< $\gamma_p$<1. The dependency function reflects the granulation order P’s power to approximate D. The dependency function is used to measure the significance of attributes for designing an attribute reduction algorithm.

The significance measure of a in B is defined by the following equation (9).

$$SG_{s}(x, B, DT, US) = yB \cup \{x\}(DT) - \gamma_p(DT)$$

(9)

Where, $\gamma_p(DT) = \frac{|\text{pot}_{p}(a)|}{|U|} = \frac{\sum_{x \in A} \text{pot}_{p}(a)}{|U|}$.

In a heuristic fuzzy-rough feature selection algorithm, and based on the above said definitions one can gradually add the selected attributes.

Moreover, to measure the quality of the boundary region a significant value $U_p$ for the subset P is calculated by the sum of all objects’ distances are given in equation 10.

$$U_p(S) = \sum_{a \in P} \delta_a (\text{pot}_{\text{new}}, B)$$

(10)

This consequence measure is used in combination with the rough set dependency value to gauge the efficacy of attribute subsets in a comparable way to that of the rough set dependency measure. Since one measure only operates on the objects in the lower approximation and the other only on the objects in the boundary, then both entities are considered individually and then combined to create a new evaluation measure $N_r$, and is given in equation 11.

$$N_r(S) = \frac{U_p(S) + \gamma_p(S)}{2}$$

### 3.5.1 Distance Metric based on Mahalanobis Distance

Mahalanobis distance is a recognized statistical distance function, and a measure of variability can be combined into the distance metric directly. It is a distance measure between two points in the space distinct by two or more interrelated variables. Obviously, Mahalanobis distance takes the correlations within a data set between the variable into consideration. If there are two non-correlated variables, then the Mahalanobis distance between the points of the variable in a 2D scatter plot is similar as Euclidean distance.

It is based on the correlations between variables by which different patterns can be analyzed and identified. The Mahalanobis distance is the distance between an opinion and the center of each group in m-dimensional space defined by m variables and their covariance. It measures similarity of an unidentified sample set to a known one shown in equation 12.

$$D_{\text{Mahalanobis}} = \sqrt{D} = \sqrt{(X - \mu) \Sigma^{-1} (X - \mu)}$$

(12)

In calculating terms, the Mahalanobis distance is equivalent to the Euclidean distance when the covariance matrix is the unit matrix. The Mahalanobis distance is consequently a weighted Euclidean distance, where the weighting is determined by the range of variability of the sample point.

Where, -1 i represent the inverse of the covariance matrix of class (I) $\Sigma^{-1}$ is the inverse of the covariance matrix. The Mahalanobis metric is defined in the independence of the data matrix. It is a distance in geometrical intelligence because the covariance matrix and its inverse are positive definite matrices. The metric defined by the covariance matrix provides a normalization of the data relative to their spread. An important property of the Mahalanobis distance is normalized. Therefore, it is not necessary to regularize the data, provided rounding errors in the covariance matrix.

Moreover, the Mahalanobis distance weighted is expressed by the covariance matrix. The covariance matrix is classified into three different types.

**Spherical:** The covariance matrix is a scalar multiple of the identity matrix $\sum_{i=1}^{m} \sigma_i^2 J$

**Diagonal:** The covariance matrix is diagonal $\sum_{i=1}^{m} \sigma_i^2 \sigma_j^2 \ldots \sigma_i^2$

**Full:** The covariance matrix is concealed to be several positive definite matrix with rank $\times x$.
covariance matrix applies to accurate the effects of cross-covariance between two components of random variable.

### 3.6 Attribute Subset Selection

Even though the individual distances may be useful in identifying the objects that are similar to lower approximation. A technique of achieving this measure is to compute the sum of all the distances and giving a significant value to all subsets considered for selection. The significance value is real-valued and has membership in the range \([0, 1]\) for the purpose of dealing with crisp data.

\[
Z_{\text{sort}}(\{i\}) = (1+1+1+1) = 0.2 \quad (13)
\]

Though, the implication measure alone can be used to investigate for subsets. The results from some initial investigation indicated that are not equal quality as those returned by fuzzy rough set feature selection. Consequently, the significance value was connected with the rough set dependency value.

The final algorithm of the proposed method is described in the form of Pseudocode as below.

**Pseudocode for the proposed Fuzzy Rough Set based Feature Selection**

**Input:** Training dataset, Test Dataset  
**Output:** N, Rel, FRel, IFRel  
**Procedure:**

For \((j=1,2,3,\ldots, D)\)  
\(X_i \leftarrow X_i + \text{Ran}(0,1)(x_{i}^{\text{max}} - x_{i}^{\text{min}})\)

Solution Set \(X=X_1, X_2, X_3, \ldots, X_n\)

Begin Sort(X)

For Cycle 1 to \(N\) do

For each employed bee I do/ applying bee colony

Rand(X) in Neighbour(X)

\(X_i \leftarrow X_i \cap X_n\)

\(X_i \leftarrow X_i \cup X_n\)

Evaluate Rel+Summation\(m(X, c)\) // computing relevancy  
FRel=\(\text{Summation}\(f_m(X, c)\)\)/Computing fuzzy relevancy  
\(X \leftarrow \text{AddRel, FRel}\) // updating Solution Set

End Sort

Begin Sort on X.

Best(S N) \(\leftarrow X(\text{Rank}, \text{Distance})\)

For each onlooker bee I do

Select food \(X_i \leftarrow P_i\) // selecting Set based on Probability  
Rand(X) in Neighbour(X)

\(X_i \leftarrow X_i \cap X_n\)

\(X_i \leftarrow X_i \cup X_n\)

Evaluate Rel+Summation\(m(X, c)\) // compute Relevancy  
FRel=\(\text{Summation}\(f_m(X, c)\)\)/Computing fuzzy relevancy  
\(X \leftarrow \text{AddRel, FRel}\)

End sort

Begin sort(X).

Best(S N) \(\leftarrow X(\text{Rank}, \text{Distance})\) // select best Solution  
If(abandoned solution)

\(X_i \leftarrow x_{i}^{\text{min}} + \text{Ran}(0,1)(x_{i}^{\text{max}} - x_{i}^{\text{min}})\)

New\(X_i \leftarrow \text{Scout bee}\)

End if

End sort

CA=\(\text{Acc}(F(X))\) // compute classification accuracy  
Return \(X_i, CA\) // selected best solution

In the Pseudocode said above, the random variable is chosen and assigned to the variable \(X_i\) and the solution set contains the set of variables. From the solution set, the sorting is applied and a bee colony optimization is also applied in order to get optimized values from the solution sets. The random values are chosen from the neighbor values. Then, the relevancy and the fuzzy relevancy values are calculated, then the solution set is updated based on the rank. The solution set is based on the probability values and relevancy is computed. From the relevancy, the best solution is selected. The two objective functions of the solutions are evaluated and the non-dominated sorting is applied to the solutions. For each and every set of solutions, a random set and the neighborhood is selected. The crossover and the mutation operations are applied and the two objective functions are evolved. The steps are repeated and the best solutions are optimized using the Ant Colony Optimization (ACO) techniques. At the end of the bee colony optimization the non-dominated sorting is applied to the parent solutions and their mutants. The probabilistic values are based on the fitness values for selecting the solution for evolution using the equation (14).

\[
\text{fitness } i = \frac{1}{A(i) - TS(i) - c(i)} \quad (14)
\]

Where, \(A(i)\) is the Pareto front rank value of the solution \(i\), \(T>0\) is the temperature and \(c\) is the crowding distance and \(S(i)\) is given by the equation (15).

\[
S(i) = -aT(i) \log ap(r(i)) \quad (15)
\]

\[
aT(i) = \left(\frac{1}{Z}\right) \exp(-\frac{A(i)}{T}) \quad (16)
\]

\[
Z = \sum^N \exp\left(-\frac{R(i)}{T}\right) \quad (17)
\]

Where, \(AT(i)\) is the Gibbs distribution and \(N\) is the size of the population.

### IV. PERFORMANCE ANALYSIS

In order to evaluate the effectiveness of our proposed work, we organized several executions on the synthetic sequence of tasks. We proposed the methodology with the model of fuzzy rough set feature selection. Additionally, fuzzy rough set based on feature selection (FRFS) is compared with several existing algorithms such as Serial algorithm, Map reduce algorithm, Fuzzy conditional Entropy (FSCE), Fuzzy Accelerated FSCE algorithm (FA-FSCE), Distance metric approach to rough set boundary region for attribute reduction (DMRSAR). Fig 3 illustrates processing accuracy comparatively with existing and proposed work, the FRFS proposed value equivalent increases the accuracy value than existing approaches.

FRFS value leads to more elaborate approach. Furthermore, proposed work increases the number of process accuracy and gives it better accuracy than the existing works such as Serial algorithm, Map reduce, FSCE, FA-FSCE and DMRSAR.

**Fig 3. Comparison of Accuracy for Existing vs Proposed algorithms**

Fig 3 represents the comparison graph depicting the accuracy of the proposed methods to that of the existing methods. In existing there are several algorithms such as Serial Algorithm, Map Reduce, FSCE, FA-FSCE, DMRSAR, and the proposed FRFS. From graph it is clear that the proposed algorithm for attribute reduction achieves high accuracy than the existing algorithms. There are three datasets used for analyzing the performance of the proposed algorithm, which includes, Breast cancer dataset, Wdbc dataset, and credit card dataset. The
Breast cancer, Wdcb, Credit card data set achieved 84%, 86%, and 73% accuracy respectively, when compared to that of the existing methods.

Fig 4 characterizes the time required for the proposed FRFS techniques and existing techniques. Execution time of the approach plays an essential role in the approximation of the efficacy of the feature selection techniques. It exposes that the proposed method consumes less time than the existing method. This obviously denotes that proposed method is faster than the existing methods.

![Fig 4. Execution Time for the Existing and Proposed algorithms](image)

**Fig 4. Execution Time for the Existing and Proposed algorithms**

Fig 5 depicts the memory usage to execute the existing and the proposed techniques. It symbolizes that FRFS requires lesser main memory than the existing methods.

![Fig 5. Memory Consumption for existing and proposed algorithms](image)

**Fig 5. Memory Consumption for existing and proposed algorithms**

FRFS is a good starting point for further work based on the distance metric for exploring the boundary region of rough sets. FRFS also showed better classification accuracy, execution time and memory consumption results.

![Fig 6. Before Reduction and After Reduction of features for existing and proposed algorithms](image)

**Fig 6. Before Reduction and After Reduction of features for existing and proposed algorithms**

In this paper, we derived an apt methodology for an attribute subset selection for the boundary region of rough set. Attribute reduction is one of the core issues in the rough set theory. Attribute reduction maintains the knowledge base under the principle of the same classification ability to remove irrelevant and redundant attribute properties. However, it

![Fig 7. Number of Instances vs. Attributes](image)

**Fig 7. Number of Instances vs. Attributes**

**Precision and Recall**

Precision and recall are the two basic metrics for accessing the performance of the proposed feature selection algorithms[20]. The Precision is the metric, which refers to the datamining algorithm that is used to select the features from the reduced set of attributes. It indicates the 100% accuracy of the proposed methods. The Recall is the percentage of information, which is relevant to the selected features. The equation (18) and (19) describes the formula to calculate the precision and recall values.

\[
\text{Precision} = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FP}} \tag{18}
\]

Where, TP is the True Positive rate and FP is the False Positive rate.

\[
\text{Recall} = \frac{\text{Number of TN}}{\text{Number of TN} + \text{Number of FN}} \tag{19}
\]

Where, TN is the True Negative rate and FN is the False Negative rate.

The F-measure is the ratio of the Precision to the recall values. It is given by the following equation (20).

\[
F - \text{Measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \tag{20}
\]

![Fig 8. Precision, Recall and F-measure values for the existing and proposed dataset](image)

**Fig 8. Precision, Recall and F-measure values for the existing and proposed dataset**

This conspicuously denotes that proposed method is faster, accurate, minimum number of subsets are achieved than the existing methods. Therefore, the results express that our proposed method (FRFS) comparably increase efficiently and effectively than the existing methods such as Serial algorithm, Map reduce, FSCE, FA-FSCE and DMRSAR.

**V. CONCLUSION**

In this paper, we derived an apt methodology for an attribute subset selection for the boundary region of rough set. Attribute reduction is one of the core issues in the rough set theory. Attribute reduction maintains the knowledge base under the principle of the same classification ability to remove irrelevant and redundant attribute properties. However, it
reduces the search space and improves efficiency. Mahalanobis Distance considering the correlations, works better. The effect of distance basis becomes less imperative for generating dataset, since there is no correlation. Moreover, this paper proposes an attribute selection based on fuzzy rough set and ranking approach. Experimental results have shown that our proposed method predicts the relevant memory requirement and also reveals that it consumes less execution time than the existing method. Our proposed algorithm outperforms the existing method.

In future work, investigating the applicability of the suggested procedure attribute reduction is of enormous attention. Improving the number of the attributes in the reduction set will concern the simplicity of decision-making rules directly and it is still a hot research issue for scholars to find efficient attribute reduction algorithms.

REFERENCES


