

# **An Optimized Classifier Frame Work based on Rough Set and Random Tree**

Nidhi Patel

Department of Computer Science & engineering  
Samrat Ashok Technological Institute  
Vidisha, MP, India

## **ABSTRACT**

Over the past two decades, Machine Learning has become one of the mainstays of information technology. Machine learning is concerned with the development of algorithms and achieves optimization classification of attributes. Classification under the decision tree is the prediction approach of data mining techniques. In the decision tree, classification algorithm has the most common classifier to build tree. This research work proposes an optimized classifier framework based on rough set and random tree classifier. Therefore, this paper puts forward a new algorithm, which combined with rough set theory and random Tree, here rough set theory used to reduce the attributes in the decision system, and uses the reduct data, as the input of decision tree. Random tree algorithm is increase high accuracy rate of the result. This article has put new concepts into practice, and the result of these concepts shows that rough set with random tree classifier have high accuracy and low time consumption compared over the rough set based J48 classifier.

## **Keywords**

Data mining; Rough Set; Reduce Attributes; Decision Tree; Random Tree; MATLAB;WEKA.

## **1. INTRODUCTION**

Data mining, the extraction of hidden predictive information from large databases is a powerful new technology. Data mining is a multi disciplinary field, which is combination of machine learning, statistics, database technology and artificial intelligence. Data mining has proved to very useful in the field of medical analysis, as it increases diagnostic accuracy, reduces costs of patient treatment and save human resources. There are various data mining techniques such as Association, Classification, Clustering, Neural Network and Regression. Data mining have two tasks one is Descriptive task (Clustering rule, Association rule, and Sequential Pattern) and other one is Predictive task (Classification, Regression and Deviation Diction).Machine learning is concerned with design and development of algorithms. It is a programming approach to computers to achieve optimization classification is the prediction approach in data mining techniques. Machine learning has four tasks these are supervised learning, unsupervised, reinforcement and evolutionary learning. In machine learning and data mining field, such databases constructed with a considerable number of attributes are often encountered. A considerable number of attributes are actually similar to find some of these attributes are irrelevant or redundant, which not only occupy extensive computing resources, but also seriously collision the decision making process. For these reasons, it becomes natural to remove the irrelevant or redundant information and make the data set compact. Attribute reduction, also known as feature selection, which is performed to process an information system. Rough set theory, first designed by Pawlak (1982), can serve to deal

with data classification problems by adopting the view of equivalence classes [6]. The role of rough set is to process an information system through removing redundant attributes [4]. Many attribute reduction methods in rough set theory have been propose for achieving reduct and have wide applications in many fields [7].

The reducts of information systems usually may be not one and only and all the reducts can be capture with the top-down attribute selection algorithm. Fortunately, in most of the real-world applications, it is dispensable to find all the reducts. Generally, the reduct with the least attributes is select as the optimal one in no show of other sources of information. A growing number of attribute reduction methods are developing to acquire only one reduct. They are usually implemented through a confident measure to evaluate the significance of attributes.

The attribute significance measures can be dependence and consistency based [3]. However, such approaches probably converge to a local optimum, and the acquired reducts may be not the minimum one, but local optimal reducts. In case a local optimal reduct is treated as the optimal reduct, the complexity of encourage rules may increase and false decisions may be made [2]. Therefore, the issue about how to acquire the optimal reduct of a decision table needs to be check into thoroughly. According to the above analysis, the aim of this paper is to propose rough set with attribute significant measures such as dependency [5]. An enhancement heuristic searching strategy is used to find the optimal reduct more effectively with random tree classifier. Decision Tree have two types one is classification other is regression. Decision Tree classification is a technology commonly used in data mining. Decision Tree has various advantages, such as effectiveness in classification, high speed, easy to understand and so on. There are many decision tree algorithms, such as ID3, C4.5, J48, Random Tree etc [17]. Random Tree can deal with both classification and regression problem. Regression Tree is a data-mining task of predicting the value of target [18]. In proposed works, we are using new concepts rough set with random tree to find result with higher accuracy and low time consumption [16].

## **2. CONCEPTS**

Here we using rough set based random tree classier. In rough set, removes the redundant attributes of the decision system by using the approximations and then the reduct attribute be the input of the decision tree and then use the random tree classier under the classification tree to introduce set of rules for decision.

### **2.1 Rough Set Theory**

Rough set theory is a continuous of conventional set theory that supports approximations in decision-making [1]. It provides a powerful tool to achieve reducts of information

system, such reducts do not contain redundant data so the problem starts when more than one attribute has the equal greatest contribution, that is, the significances of different attributes are given as equally largest in some domical especially for the small data sets. In such cases, any one of such attributes is in any case selected .In order to check whether the obtained reduct is the optimal one, the top-down attribute selection algorithm is using in foremost to acquire each reduct. A rough set is itself the approximation of a vague concept (set) a pair of precise concepts, called lower and upper approximations. The lower approximation is an explication of the domain objects, which are know with certainty to belong to the subset of interest. The upper approximation is an explication of the objects, which not impossibly belong to the subset and other region from apart of above the boundary region is the set of objects that cannot impossibly, but not certainly [8,9]. The rough set theory assumes that the data is a true and accurate reflection of the real world [12].

## 2.2 Information and Decision Systems

An information system represents data in the form of table with rows (objects in the table) and columns (attributes). In medical datasets, for example, patients represented as objects and their reading such as blood pressure, form attributes. The attributes values for an exact patient are their specific reading for that their calculation. An information system may be increased by the insertion of decision attributes. The terms attribute feature and variable are used together. For example, the medical information system discussed previously could be extending to involve patient classification information, such as whether a patient is fit or not. A more example of a decision system can be seen in table 1. Here, the table consists of four conditional attributes (a; b; c; d), a decision attribute (e) with eight objects. A decision information system is consistent for every set of objects. Moreover, those objects values are identical and the corresponding decision objects are not identical. More formally,  $Q = (U, A)$  is an information system, where  $U$  is a non-empty set of finite objects and  $A$  is a non-empty finite set of attributes. In such a way that  $a : U \rightarrow V_a$  for every  $a \in A$ .  $V_a$  is the set of values that an attribute may take. For decision systems,  $A = \{C \cup D\}$  where  $C$  is the set of input features and  $D$  is the set of class indices. Here, a class ordering  $d \in D$  is itself a variable  $d : U \rightarrow \{0,1\}$  such that for a  $a \in U$ ,  $d(a) = 1$  if a has class  $d$  and  $d(a) = 0$  otherwise.

## 2.3 Indiscernibility

A core concept of Rough set theory is that of equivalence between objects called indiscernibility. The relation  $IND(p)$  is called a  $P$  indiscernibility relation. The division of  $U$  is a family of all similarity classes of  $IND(p)$  and is denoted by,  $U/IND(P)$ , with any  $P \subseteq A$  there is an associated similarity relation  $IND(p)$ :

$$IND(P) = \{(i, j) \in U^2 \mid \forall a \in P, a(x) = a(j)\} \quad (1)$$

Note that this corresponds to the similarity relation for which two objects are identical if we have identical vectors of attribute values for the attributes in  $P$ . The division of  $U$ , determined by  $IND(P)$  is express by  $U/IND(P)$  that is simply the set of equivalence classes produced by  $IND(P)$ :

$$U/IND(P) = \otimes \{U/IND(a) \mid a \in P\} \quad (2)$$

**Table 1. An Example Dataset**

$i \in U$	a	b	c	d	e
0	S	R	T	T	R
1	R	S	S	S	T
2	T	R	R	S	S
3	S	S	R	T	T
4	S	R	T	R	S
5	T	T	R	S	S
6	T	S	S	S	T
7	R	S	S	R	S

Where

$$A \otimes B = \{X \cap Y \mid \forall X \in A, \forall Y \in B, X \cap Y \neq \phi\} \quad (3)$$

If  $(i; j) \in IND(P)$ , then  $j$  and  $i$  are indiscernible by attributes from  $P$ . The equivalence classes of the indiscernibility relation with consideration to  $P$  are denoted  $[i]_P$ ,  $i \in U$ . For the explanatory example, if  $P = \{b, c\}$  then objects 1, 6 and 7 are indiscernible; as are objects 0 and 4.  $IND(P)$  creates the following division of  $U$ :

$$\begin{aligned} U/IND(P) &= U/IND(b) \otimes U/IND(c) \\ &= \{\{0,2,4\}, \{1,3,6,7\}, \{5\}\} \\ &= \otimes \{\{2,3,5\}, \{1,6,7\}, \{0,4\}\} \\ &= \{\{2\}, \{0,4\}, \{3\}, \{1,6,7\}, \{5\}\} \end{aligned}$$

## 2.4 Lower and Upper Approximations

Let  $X \subseteq U$ ,  $X$  can be approximated using only the information include within  $P$  by build the  $P$ -lower and  $P$ -upper approximations of the classical crisp set  $X$ :

$$\underline{P}X = \{i \mid [i]_P \subseteq X\} \quad (4)$$

$$\overline{P}X = \{i \mid [i]_P \cap X \neq \phi\} \quad (5)$$

It is such a collection of rows  $\{\underline{P}X, \overline{P}X\}$  that is termed a rough set. Consider the approximation of  $X$  concept in figure 1. Each square in the diagram represents a similar class; produced by indiscernibility between objects values. Using the features in set  $B$ , via these similarity classes, the lower and upper approximations of  $X$  can be included. Similarities classes include within  $X$  exist to the lower approximation. Objects lying within this region can be said to exist certainly to concept  $X$ . Similarity classes within  $X$  and ahead its boundary form the upper approximation. Those objects in this region can only be said to possibly exist to the concept.

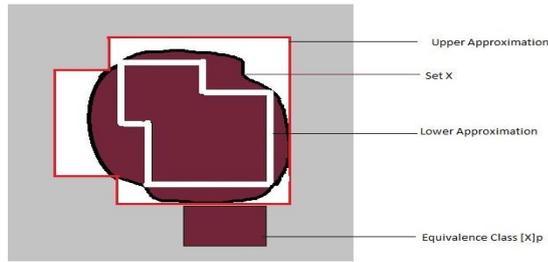


Fig 1 A Rough Set

## 2.5 Positive, Negative and Boundary Regions

Let P and I be equivalence relations over U, then the positive (lower approximation), negative (upper approximation) and boundary regions are define as following:

$$POS_p(I) = \bigcup_{i \in U/I} \underline{P}X \quad (6)$$

$$NEG_p(I) = U - \bigcup_{i \in U/I} \bar{P}X \quad (7)$$

$$BND_p(I) = \bigcup_{i \in U/I} \bar{P}X - \bigcup_{i \in U/I} \underline{P}X \quad (8)$$

The lower approximation consists of all objects of U that can be classified into classes of U/I using the information included within attribute P. The upper approximation,  $NEG_p(I)$ , is the set of objects that cannot be classified into classes of U/I. The boundary approximation,  $BND_p(I)$ , is the set of objects that can possibly, but not surly, be classified in this way. For example, let  $(P = \{b, c\}) \& I = \{e\}$ , then

$$POS_p(I) = U \setminus \{\phi, \{2, 5\}, \{3\}\} = \{2, 3, 5\}$$

$$NEG_p(I) = U - U \setminus \{\{0,4\}, \{2,0,4,1,6,7,5\}, \{3,1,6,7\}\} \\ = \phi$$

$$BND_p(I) = U \setminus \{\{0,4\}, \{2,0,4,1,6,7,5\}, \{3,1,6,7\}\} \\ - \{2, 3, 5\} \\ = \{0, 1, 4, 6, 7\}$$

This way (method) that objects 2, 3 and 5 can be surly classified as belonging to a class in attribute e, when considering attributes b and c. The rest of the objects cannot be classified, as the information that would make them distinct is not present.

## 2.6 Feature Dependency and Significance

The valuable issue in data analysis is determining dependencies between attributes. Seemingly, a set of attributes I depends totally on a set of attributes P, expressed by  $P \Rightarrow I$ , if all attribute values from I are specially determine by values of attributes from P. If there exist a functional dependency between values of I and P then I fully depend on P. In rough set theory, dependency is expressed in the following way (method).

For  $P, I \subset A$ , said that I depends on P in a degree q ( $0 \leq q \leq 1$ ), denoted  $P \Rightarrow_q I$ , if

$$q = \gamma_p(I) = \frac{|POS_p(I)|}{|U|} \quad (9)$$

Where  $|S|$  stands for the cardinality of set S. If  $q = 1$ , I depends totally on P, if  $0 < q < 1$ , I depends partially (in a degree q) on P, and if  $q = 0$  then I does not depend on P. In the example, the degree of dependency of attribute  $\{e\}$  from the attributes  $\{b, c\}$  is:

$$\gamma_{\{b,c\}}(\{e\}) = \frac{|POS_{\{b,c\}}(\{e\})|}{|U|} \\ = \frac{| \{2,3,5\} |}{| \{0,1,2,3,4,5,6,7\} |} = \frac{3}{8}$$

Here calculating the change in dependency when a feature is removed from the set of considered possible features, an approximated calculation of the given feature significance can be acquired. The higher the change in dependency, the more significant the feature is. If the significance is 0, then the feature is dispensable. More formally, given P, I and a feature  $i \in P$ , the significance of feature i upon I is express by

$$\sigma_p(I, a) = \gamma_p(I) - \gamma_{p \setminus \{a\}}(I) \quad (10)$$

For example, if we have  $P = \{a, b, c\}$  &  $I = e$ , then the following results will seen :

$$\gamma_{\{a,b,c\}}(\{e\}) = | \{2,3,5,6\} | / 8 = 4/8$$

$$\gamma_{\{a,b\}}(\{e\}) = | \{2,3,5,6\} | / 8 = 4/8$$

$$\gamma_{\{b,c\}}(\{e\}) = | \{2,3,5\} | / 8 = 5/8$$

$$\gamma_{\{a,c\}}(\{e\}) = | \{2,3,5,6\} | / 8 = 4/8$$

Calculating the significance of the three attributes give:

$$\sigma_p(I, a) = \gamma_{\{a,b,c\}}(\{e\}) - \gamma_{\{b,c\}}(\{e\}) = 1/8$$

$$\sigma_p(I, b) = \gamma_{\{a,b,c\}}(\{e\}) - \gamma_{\{a,c\}}(\{e\}) = 0$$

$$\sigma_p(I, c) = \gamma_{\{a,b,c\}}(\{e\}) - \gamma_{\{a,b\}}(\{e\}) = 0$$

From this follows that attribute a can be an indispensable, but attributes b and c can be allocate with when considering the dependency between the given individual conditions attribute.

## 2.7 Reducts

For many application problems, it is often important to maintain a compact form of the information system [10,11]. One process to implement this is to search for a minimal representation of the original dataset. For this, the conception of a reduct is way out and defined as a minimal subset R of the initial feature set C such that for a given set of features D [13].

## 2.8 Decision Tree

Reduct attributes put as an input of decision tree. A decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute, each branch represents the outcome of the test and each leaf node represents a class label. The paths from root to leaf represents classification rule. Classification is possibly most frequently used in data mining technique. It is a process of finding a set of models that describes and differentiates data classes and concepts, for being able to use the model to predict the classes whose label is unknown. There are many algorithms that can be used for classification, such as decision tree, neural network, logistic regression etc. Various decision algorithms used in classification like C4.5, J48, random tree etc. A random tree is a collection of tree predictors that is called forest. It can deal with both classification and regression problems [15]. In classification, random tree classifier that takes the input

feature vector, classifies it with every tree in the forest, and out puts the class label that received the majority of “vote”. In case of a regression the classification, response is the average of the response over all the trees in the forest [24].

### 3. PROPOSED METHODOLOGY

In proposed methodology, works focus on the finding the optimal reduct from the redundant data and apply as input for random tree classifier to achieve high accuracy in decision-making system. In which rough set works on finding the optimal reduct as input for random tree classifier in classification to achieve high accuracy with low time consumption. So completely, process is divide into two steps first is pre-processing of redundant data with lower and upper approximation to find optimal reduct. While in the next step testing is done as input of reduct data for random tree, it is produce n number of decision rules for n number of class label in decision tree and introduce high accuracy and low time consumption. Planned procedure has been applied here for, finding an optimal reduct from the redundant attribute with better accuracy in the field of medical analysis as it increases diagnostic accuracy, to reduce costs of patient treatment using a hybrid combination of rough set and random tree classifier.

The planned procedure implemented shown in fig 2, here works in the following way:

In the open dataset discipline, dataset is the unit to measure the information released in a public open data repository. Here using medical field related dataset from uci machine learning Repository. Section 4.3 shows the medical field related dataset. Dataset construct with a considerable number of attributes are often encountered. It is common to find some of these attributes quite redundant. It seriously affects the decision-making process system. In an information and decision system is in a pair  $Q=(U,A)$ . Where U and A, are finite, non empty sets called the universe and attributes set. In pair  $Q=(U,A)$  each attribute  $a \in A$  and  $V_a$  holds the value of attributes, called the domain of a. Any subset of A determines a binary relation on U, After taking dataset, we find the indiscernibility classes that are a core concept of Rough set theory is that of equivalence between objects called indiscernibility which will be called an Indiscernibility Classes. After finding the indiscernibility classes, apply lower & upper approximation because Rough set is itself the approximation of a vague concept (set) a pair of precise concepts, called positive and negative regions. The lower approximation is an explication of the domain objects, which are know with certainty to belong to the subset of interest. The upper approximation is an explication of the objects, which not impossibly belong to the subset.

In positive region find the reduct attributes and remove redundancy.

$$PX = \{i \mid [i]_p \subseteq X\}$$

Moreover, apply Degree of Dependency and significance on attributes based on positive region. In rough set theory, dependency is defined by  $q$ . Here if degree of dependency vary  $0 < q < 1$ . It means partially dependency of attribute, if  $q=1$ , it means fully dependency and if  $q=0$  it means does not depend of any other attributes. By calculating the dependency, we estimate significance of attribute. If the significance is 0, then the attribute is removed otherwise it is indispensable.

$$q = \gamma_p(I) = \frac{|POS_p(I)|}{|U|}$$

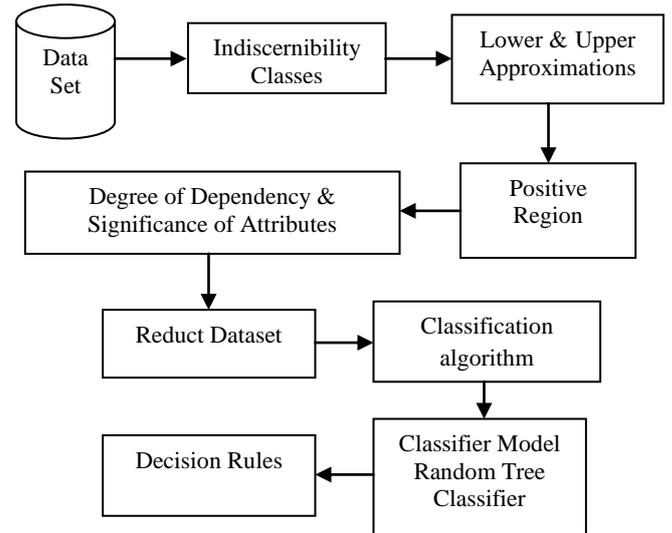


Figure 2: The Proposed System

It is often important to maintain a short form of the information system. This paper aims to find a minimal representation of the original dataset. For this, the concept of reduct attributes define as a minimal subset R of the initial attribute set C. Apply classification algorithm in reduct dataset. We are using decision tree algorithm for classification, a decision tree is a structure that includes a root node, branches, and leaf nodes. At last, we apply decision rules on classification attributes that are representing the knowledge in the form of IF-THEN rules .One rule is created for each path from the root to a leaf. Each attribute-value pair along a path forms a conjunction. The leaf node holds the class prediction. Rules are easier for humans to understand. Class prediction value will increase with the accuracy of R by decision rule

#### 3.1 Dependency based Attribute Reduction Algorithm

Dependency based attribute reduction algorithm described in 3.1. In which algorithm removes the redundant attributes and gives the optimal reduct. An information system expressed as a tuples  $IS=(U,V,f,C,D)$ . Where U denotes a non-empty finite set of objects, A denotes the entire attribute set including the condition attribute set including the condition attribute set C and the decision attribute set D, V denotes a non-empty set of attribute values ,and the function f is a Cartesian product of C,D and U into V.

Input: An information system  $IS = (U, C \cup D, V, f)$ .

Output: One reduct R of the information system IS

Step 1: Compute the consistency  $\gamma_p(D)$  based on Rough Set positive region.

Step 2: Let  $R = \emptyset$

Step 3: Compute the core set Core(C) and  $R=R \cup Core(C)$

Step 4: To each attribute  $a \in C - R$ , compute SGF (a,R,D),

let a1 represent the one that maximize SGF(a,R,D)

Step5: Add the selected attribute a1 to the subset R, i.e.

$R=R \cup a_1$

Step 6: If  $\gamma_R(D) = \gamma_p(D)$ , continue; otherwise, go to step4;

Step7: If  $R \neq \text{Core}(C)$ , remove or take off redundant attribute if exists;

Step 8: R Output.

End

### 3.2 RSBRT (rough set based random tree)

#### Classifier algorithm:

RSBRT classifier algorithm describe in 3.2. An algorithm to generate an n number of decision rules for decision tree and give the high accuracy and low time consumption in the following way:

Input: An information system  $IS = (U, C \cup D, V, f)$

**The random tree algorithm (for both classification and regression problems) is as follows:**

1. Make n tree bootstrap samples from the original reduct data.
2. For each of the bootstrap samples, grow an unshorted classification or regression tree, with the following modification: at each node, rather than selecting the best split among all predictors, randomly sample m attempt of the predictors and select the best split from among those variables. (Bagging can be thought of as the special case of random forests obtained when m attempt = p, the number of predictors.)
3. Predict new data by aggregating the predictions of the n tree trees (i.e., majority votes and average vote for classification, average for regression).

**An estimate of the error rate can be obtained, based on the training data, by the following:**

1. At every bootstrap iteration or in redundancy, predict the data not in the bootstrap sample (what Breiman calls “out-of-bag”, or OOB data) using the tree grown with the bootstrap sample.
2. Now aggregate the OOB predictions. Calculate the error rate sample, and call it the OOB approximate calculation of error rate. Our experience has been that the OOB approximate calculation of error rate is completely accurate, given that enough trees have been grown (in another way the OOB estimate can bias upward)[19, 20].

## 4. EXPERIMENTAL RESULTS AND ANALYSIS

The implementation of the proposed Rough Set based Random Tree works in the case of both classification and the regression problems according to dataset. Therefore, first, the required tools and techniques are discussed then after the code implementation and development of the system is provided. The following software and hardware required for the implementation of the proposed system.

Hardware Requirement: 2.0 GHz Processor required (Pentium 4 and above), Minimum 2 GB Random Access Memory 40 GB hard disk space.

Software Requirement: Operating System (Windows 7 and above), MATLAB R2013a, Weka 3.7.4JDK1. If dataset is numerical case, it will be implemented by regression process.

## 4.1 Implementation of the Rough set based Random Tree using WEKA toolkit

This proposed algorithm uses WEKA, as API in MATLAB.WEKA is a comprehensive open source Machine Learning toolkit, written in Java. These functions provide a basic MATLAB interface to WEKA to allow the transformation of data back and forth and to use most of the features available in WEKA, such as training classifiers. By doing so, the accuracy rate of the Random tree algorithm has increased with rough set to large extent as compared to the accuracy of the same algorithm in Weka.

## 4.2 The Datasets

Medical field is an appropriate field for data mining technology due to a number of reasons. A data set is a collection of data. Most commonly, a data set corresponds to the contents of a single database table, or a single statistical data matrix, where each column in the table represents a particular variable, and each row corresponds to a given member of the data set.

There are four data sets that we are using [14].

**1. Lung Cancer:** Lung cancer can cause blockages in your major airways. Hong and Young to demonstrate the power of the optimal discriminate plane even in ill-posed settings used this data. There are 56 attributes and 32 instances in dataset [21].

**2. Hepatitis:** Hepatitis is a liver disease and can cause jaundice, aches in joints and muscles, fatigue, and loss of appetite. There are 19 attributes and 155 instances in dataset [22].

**3. Banknote Authentication Dataset:** Data were extracted from images that were taken from genuine and forged banknote-like specimens. For digitization, an industrial camera usually used for print inspection was used. Wavelet Transform tool were used to extract features from images. There are 5 attributes and 1372 instances in dataset.

**4. Lymphography Dataset:** Lymphography is a medical imaging technique in which a radio contrast agent is injected, and then an X-ray picture is taken to visualize structures of the lymphatic system, including lymph nodes, lymph ducts, lymphatic tissues, lymph capillaries and lymph vessels .There are 18 attributes and 148 instances in dataset [23].

## 4.3 Results Analysis

Results Analysis based on accuracy and time consumption in which Rough set +RT (random tree) is better performed rather than rough set +J48.

### 4.3.1 Accuracy

Accuracy of proposed classification algorithm is a measurement of total accurate identified instances over the given samples. The accuracy of the classification can be evaluated on following datasets [15].

**Table 2. Accuracy Comparisons between Rough set +J48 and Rough set + R T Classifier**

Datasets	Instanc es	Attribut es	Rough set +J48  Accura cy (%)	Rough set+RT  Accura cy (%)
Lung-Cancer	32	56	93.75%	99.35%

Hepatitis	155	19	82.58%	100%
Banknote Authentication	1372	5	85.34%	100%
Lymphography	148	18	91.89%	99.92%

The comparative accuracy of two algorithms has given in Table 2 and figure 3 shows the better accuracy of Rough set+ based Random Tree Classifier than the Rough set based J48 Classifier [18]. The classify accuracy rate has been increased in following way from rough set based J48 to rough set based Random Tree respectively lung-Cancer from 93.75% to 99.35%,Hepatitis from 82.58% to 100%,Banknote Authentication from 85.34% to 100% and Lymphography

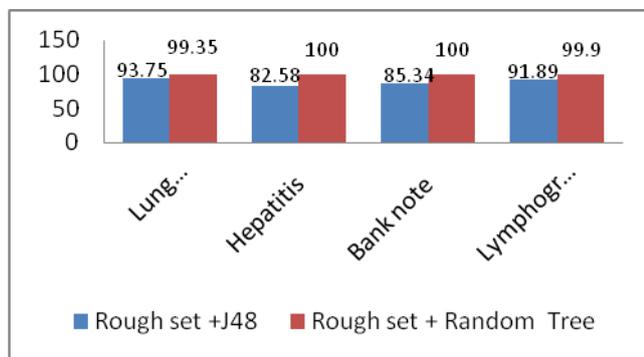


Figure 3 Comparing accuracy using graphical representation

from 91.89% to 99.92%. According to the evaluated results the accuracy performance of the proposed algorithm is much better as compared to rough set based J48.

#### 4.3.2 Time Consumption

The amount of time consumed in developing data model using proposed algorithm is as on following datasets. Time consumption means time complexity of the algorithm on various datasets.

Table 3. Time Consumption of Rough set + J48 and Rough set + RT Classifier

Datasets	Instances	Attributes	Rough set +J48 Time Consumption(In Seconds)	Rough set +RT Time Consumption(In Seconds)
Lung-Cancer	32	56	0.01	0.01
Hepatitis	155	19	0.02	0.02
Banknote Authentication	1372	5	0.05	0.04
Lymphography	148	18	0.02	0.02

The comparative time complexity of algorithms has given in Table 3 shows the better performance of Rough set based Random Tree Classifier than Rough set based J48 Classifier. The time consumption has been decreased in following way from rough set based J48 to rough set based Random Tree respectively lung-Cancer 0.01(no change), Hepatitis 0.02(no change), Banknote Authentication from 0.05 to 0.04 and Lymphography 0.02 (no change). According to the evaluated results, the time consumption performance of the proposed algorithm is much better as compared to rough set based J48.

Table 4 Comparison among size of tree, correct and incorrect instance

Datasets	Size of Tree Rough set +J48	Size of Tree Rough set +Random Tree	Correctly classified instance / incorrect classified instance(J48)	Correctly classified instance / incorrect classified instance(Random Tree)
Lung-Cancer	7	26	30/2	32/0
Hepatitis	23	159	128/27	145/1
Banknote Authentication	3	559	1171/201	1371/1
Lymphography	25	83	136/12	148/0

The comparative among size of tree, correct and incorrect instance have given in Table 4 shows that size of tree is increased from rough set based J48 classifier to rough set based random tree classifier respectively. Lung Cancer from 7 to 26, Hepatitis from 23 to 159, Banknote Authentication from 3 to 559 and Lymphography from 25 to 83. The size of tree is directly proportional to correctly classifier instance because of using rough set based random tree. Correctly classified instance increased from rough set based J48 classifier to Rough set based random tree classifier respectively lung Cancer from 30 to 32, Hapatitis from 128 to 145,Banknote Authentication from 1171 to 1371 and Lymphography from 136 to 148.

## 5. CONCLUSION AND FUTURE WORKS

In the entire study, shows that used rough set attribute reduction algorithm to calculate the indiscernibility relation of attribute, the positive region and random Tree classifier in Classification to increases the accuracy of decision. Decision tree algorithmic program is a classical approach of supervised machine learning and data processing. In proposed work, rough set applied to finding an optimal reduct from the redundant data in the field of medical analysis. Uses the reduct data as an input of decision tree then random tree classifier in the classification, introduce set of rules then with the help of rules make n number of class labels in decision trees, which are helpful to take correct decisions. The algorithms based on decision tree are able to develop a transparent and reliable data model in order to maintain the transparency and relativity between attributes. Some decision tree algorithms are computationally expensive in terms of accuracy and time. Therefore, a number of approaches are

developed in recent years by which the classifiers are claimed to provide much efficient classification accuracy in less time complexity to overcome these computationally expensive in our proposed approach.

The proposed algorithm is enhancing attribute reduction by using rough set and classification accuracy of decision tree by random tree, with low time consumption.

Used model is implemented using WEKA 3.7.4 and MATLAB R2013a, the comparative study is performed with respect to the Rough set based J48 Classifier, and Rough set based random Tree Classifier. The comparison among these algorithms is performed among accuracy and time complexity.

**Table 4.Performance Summary of Our Proposed Approach**

S. No.	parameters	Proposed Rough set +Random Tree Classifier	Rough set +J48 classifier
1	Accuracy	High	low
2	Time Consumed	Low	high

The proposed classifier, rough set based random Tree produces high accuracy, low error rate and consumes less time as compared with rough set based J48 classifier. The proposed classifier is efficient and accurate which provides effective results as compared to the traditional algorithms. In future, we will optimize the performance of classification in terms of memory consumption and training time and we will use other data parallel with this classifier for analysis of big data.

## 6. REFERENCES

- [1] Z. Pawlak.: Rough Sets.International Journal of Computer and Information Sciences, vol. 11 pp. 341–356, pp 341–356; (1982).
- [2] Dash, M., & Liu, H.: Consistency-based search in feature selection. Artificial Intelligence, vol.151, pp 155–176, (2003).
- [3] Dai, J. H.:set approach to incomplete data. Information Sciences, consistency –based search in feature 241.vol. 43,–572002, pp 43-57 (2013).
- [4] I.D.untsch, G. Gediga. Rough Set Data Analysis.In: A. Kent & J. G.Williams (Eds.Encyclopedia of Computer Science and Technology), vol.43, pp. 281–301, (2000).
- [5] H. Sever.: The status of research on rough sets for knowledge discovery in databases. In: Proceedings of the Second International Conference on Nonlinear Problems in Aviation and Aerospace (ICNPAA98), vol. 2, pp.673–680,( 1998).
- [6] Ahmad, A., & Dey, L.: A feature selection technique for classificatory analysis. Pattern Recognition Letters, 26(1)pp, 43–56, (2005).
- [7] Chai, J. Y., & Liu, J. N. C.: A novel believable rough set approach for supplier selection. Expert Systems with Applications, Vol. 41, Issue Pp 92–104,1, January (2014).
- [8] A. Skowron, Z. Pawlak, J. Komorowski, L. Polkowski. :A rough set perspective on data and knowledge. Handbook of data mining and knowledge discovery, pp. 134–149, Oxford University Press, (2002).
- [9] A. Skowron, S. K. Pal. Special issue: Rough sets, pattern recognition and data mining. Pattern Recognition Letters, vol. 24, pp 829–933, (2003).
- [10] Hu, Q. H., Zhao, H., Xie, Z. X., & Yu, D. R.: Consistency based attribute reduction. In Z.-H. Zhou, H. Li, & Q. Yang (Eds.), PAKDD LNCS (LNAI) Vol. 4426.pp 96-107 (2007). Heidelberg: Springer.
- [11] Deng, T. Q., Yang, C. D., & Wang, X. F. A: reduct derived from feature selection.Pattern Recognition Letters, vol.33,pp 1628–1646; (2012)
- [12] Z. Pawlak. Rough Sets: Theoretical Aspects of Reasoning About Data Kluwer Academic Publishing, 1991.
- [13] Kai Zheng<sup>a</sup>, Jie Hu<sup>a, ·</sup>, Zhenfei Zhan<sup>b</sup>, Jin Ma, :An enhancement for heuristic attribute reduction algorithm in roughset,vol 41,pp 6748-6754( 2014)
- [14] <https://archive.ics.uci.edu/ml/datasets.html>.
- [15] Loh WY. Regression by parts: fitting visually interpretable models with GUIDE. In: Chen C, Hordle W, Unwin A, eds. Handbook of Data Visualization. New York: Springer,vol.28, pp447–469; (2008).
- [16] Chaudhuri P, Lo WD, Loh WY, Yang CC.: Generalized regression trees.vol 5:pp641–666; (1995).
- [17] Loh WY.: Regression tree models for designed experiments. IMS Lecture Notes-Monograph Series, vol49,pp210–228;( 2006).
- [18] Xiang Zhuoyuan and Zhang Lei ,:Research on an optimized C 4.5 Algorithm Based on Rough set Theory.DOI:10.1109/ICMeCG.2012.74Beijing, pp. 72-274International Conference on Management of e-Commerce and e-Government;(2012).
- [19] L. Breiman. :Random forests. Machine Learning, vol45,pp5–32, (2001).
- [20] Liaw and Matthew Wiener: Classification and Regression by randomForest Andy. Vol. 2/3, (December 2000).
- [21] Aeberhard's second ref. above, or email to stefan '@' coral.cs.jcu.edu.au.
- [22] G.Gong (Carnegie-Mellon University) via BojanCestnik Jozef Stefan Institute Jamova3961000Ljubljana Yugoslavia(tel.: (38)(+61) 214-399 ext.287) }
- [23] Igor Kononenko, University E.Kardelj Faculty for electrical engineering Trzaska25 61000 Ljubljana (tel.: (38)(+61) 265-161
- [24] Purva Sewaiwar, Kamal Kant Verma :Comparative Study of Various Decision Tree Classification Algorithm Using WEKA. International Journal of Emerging Research in Management &Technology ,Vol 4, pp 2278-9359(2015).