The Correlative of Fuzzy-Rough Nearest Neighbour Classifiers Using Heart Disease Prediction System

Dr. V. Krishnaiah, Associate Professor, Dept. of CSE, Neil Gogte Institute of Technology (NGIT), Hyderabad, TS, India.

Abstract— Fuzzy-Rough approaches can play a significant role in data mining, since they give understandable outcome. In addition, the approaches considered in data mining contain mostly be learning at greatly ordered and accurate data. In this work, the performance is done of three classifiers using heart disease data prediction system. The fusion of K-NN, Fuzzy K-NN, and Fuzzy-Rough (FRNN) were used to calculate the accuracy of event of heart disease data sets. The experiments are passed out using heart disease data set of Uel machine learning repository, and it is implemented on through the process of using the fuzzy-rough tools in Python and experimenter.

Index Terms— Data Mining, Fuzzy Sets (logic), Rough sets, K-NN, Fuzzy K-NN, Fuzzy-Rough NN, Membership function, Ownership function, Heart disease.

1 Introduction

THIS section discusses about data mining, fuzzy logic & fuzzy sets, rough sets, fuzzy-rough nearest neighbor algorithm, and heart disease.

1.1 Data Mining

Data mining techniques in history of medical data establish with huge investigations establish that the prediction of heart disease is very significant in medical science. In medical history it is observed that the unstructured data as heterogeneous data and it is observed that the data created with dissimilar attributes must be analysed to predict and give information for making diagnosis of a heart patient. Different techniques in data mining have been applied to predict the heart disease patients. Data mining is vital step in finding of knowledge from large data sets. In recent years, Data mining has found its important siege in each field as well as health care. Mining procedure is more than the data examination allow into a group of Classification, Clustering, Association rule mining and Prediction. Data mining helps to extract information from large data repositories. The techniques of data mining are used to achieve huge databases to discover original and helpful patterns. Data mining has evolved as a tool for providing solutions to analysts' problems.

This proposes well organized to find the Fuzzy-Rough logic classifier, which is used as a successful tool to develop the classification accuracy. Our previous research work implemented by using Fuzzy K-NN (F-KNN) classifier with the comparison of simple K-NN classifier and reached 97% [5], and the extension of in this research work is compare to our previous work by using Fuzzy-Rough NN (FRNN) classifier getting more accuracy. The main objective of this research work is to give an approach for diagnosing the heart disease of the patients with Fuzzy-Rough Set (FRS) approach, which uses a test object's nearest neighbours to make the lower and upper approximations of each decision class, and then computes the membership and ownership of the test object to these approximations.

1.2 Fuzzy Logic & Fuzzy Sets

Fuzzy Set is every set to permit its members to contain dissimilar grades of membership in interval [0, 1]. Fuzzy classification offers another to crisp logic by calculating data set build on their membership into every category. Fuzzy membership accept that membership to a agreed class can resolve the range as of complete membership (100%) to non-membership (0%), and that the data set may be classified as partial members into two or more categories [1].

1.3 Rough Sets

The rough sets theory (RST) [2, 3, 4] is based on the research of information system logical properties, and uncertainty in it is expressed by a boundary region. A training set can represent by a table where each row represents objects and every column constitute an attribute is called Information System. Classical definitions Lowe and upper approximations were formerly presented by suggestion to an indiscernible relation which assumed to be an equivalence relation. Every investigated object is associated to a particular part of information, to specific data. The objects which are characterized by the same part of information are commonly undistinguishable from the point of view of the reachable part of information. This expressed in RST by the indiscernibity relations.

1.4 Fuzzy-Rough Nearest Neighbor (FRNN) Algorithm

By compare to the latter, our technique uses the nearest neighbors to lower and upper approximations o decision classes and classifies check specimen position on their membership to these approximations. In the experimental analysis, in this paper calculate the approach with both classical fuzzy-

• *Dr. V. Krishnaiah, Associate Professor, Department of Computer Science And Engineering, Neil Gogte Institute of Technology (NGIT), Kachawanisingaram Village, Peerzadiguda, Uppal, Hyderabad, Telangana State, India, 500058, Email: varkala.krish@gmail.com.*

rough approximations (based on an implicator and a t-norm), as well as newly introduced vaguely quantified rough sets. Bearing results are very good, in general FRNN outperforms FRNN–O, additionally the conventional fuzzy nearest neighbor (FNN) algorithm. In this paper, to predict the heart disease patients, introduced other precedes towards which uses a test objects nearest neighbors to make the lower and upper approximations of. Further, an attempt was made classify the patients based on the attributes collected from medical field. As well as the vaguely quantified rough sets (VQRS) model, this is more robust in the presence of noisy data. FRNN classifier was designed to classify the training and testing data is associated with dissimilar classes. It was established that FRNN classifier suits well as compared with other classifiers of parametric techniques.

1.5 Heart Disease

Heart disease is the world's most important killer, accounting for 29.2% of total global deaths in 2003. The World Health Organization in 2009 projected that nearly 20 million deaths happen yearly from CVD and that by 2030 that figure could rise to approximately 24 million. The WHO expected that 60% of the world's cardiac patients are Indian. Prediction of this disease will help to avoid it in its early stage. Various hospitals manage the clinical data built with health care information system, as system updates a data on usual basis extracting the information position on decision measures is a small complexity for suitable prediction.

2 RELATED WORK

In the year 2019, [5], presented classification over numerous real-world datasets has a peculiar drawback called unstructured class problem. A dataset is said to be unstructured when the majority of the class has more samples insignificantly than the minor class. Such drawbacks result in an ineffective performance of data classification techniques. Classification is a supervised learning method which acquires a training dataset to form its model for classifying unseen examples. As a result, the false-negative rate can be excessively high. The researches focus on the unstructured data classification using uncertain Nearest Neighbor (NN) decision rule and also found the major issues face by k-Nearest Neighbor (k-NN). In any case, given a dataset, prediction of accuracy is a monotonous task to improve the execution of k-NN by tuning. This work addresses the issues faced by k-NN by developing Adaptive-Condensed Nearest Neighbor (Ada-CNN). The Ada-CNN classifier utilizes the distribution and density of test point's neighborhood and learn an appropriate point-explicit by using artificial neural systems. Ada-CNN performed well compared to k-NN and other well-known classifiers. The experimental results showed that Ada-CNN achieved nearly 94% accuracy.

In the year 2018, [6], presented a novel algorithm for computing a training set consistent subset for the nearest neighbor decision rule. The algorithm, called FCNN rule, has some desirable properties. Indeed, it is order independent and has sub quadratic worst case time complexity, while it requires little iteration to converge, and it is likely to select points very close to the decision boundary. And compared the FCNN rule with state of the art competence preservation algorithms using large multi-dimensional training sets. Showing that it out performs existing methods in terms of learning speed and learning scaling behavior, and in terms of size of the model. While it guarantees comparable prediction accuracy.

In the year 2017, [7]: *Methodology Employed:* The decision scheme depends on the incorporation of ANN and Fuzzy Analytic Hierarchy Process (FAHP). Fuzzy-AHP system was employed to calculate the global weights for the attributes depending on their individual involvements. The global weights symbolize the roles of the attributes which were taken up to train ANN Classifier for the forecast of Heart Failure (HF) risks. *Advantage:* The hybrid method and ANN classifier avoid the inappropriate problem and progress the simplification capability. The training factor, such as, an amount of hidden layers and epochs were cautiously experimented over numerous runs and the optimal factor set was chosen. *Limitation:* The ANN method of global weights was used without regularization, but was not suitable for generalization of new data.

In the year 2016, [8]: *Methodology Employed:* A Neighborhood Rough Set Classification (NRSC) Algorithm was harnessed to categorize the ECG signals into usual and abnormal heart beats. These heart beats include Right Bundle Branch Block (RBBB), Left Bundle Branch Block (LBBB), premature ventricular contractions and Paced Rhythm (PR) beat. *Advantage:* By assistance of neighborhood approximation, two kinds of decisions were generated. They are deterministic policies and non-deterministic policies. The first policies are generated by applying lower approximation, and the second policies are spawned by applying upper approximation of the NRS. The categorization process begins by acquiring data set (input–output data pairs) and splitting it into a training set and validation data set. The cross-validation method was used for better reliability of test results. *Limitation:* The collection of the data set is very little. Medical categorization model is utilized as a viable aid to categorize numerous medical data sets.

In the year 2015, [9]: *Methodology Employed:* A data mining obsessed in Coronary Heart Disease (CHD) forecast mold applying fuzzy logic and decision tree was projected. A decision tree routine was exercised to the rules. The projected mold was useful for a pruning leaf node, which augments the simplification capability of the trees, having merges steps and leaf nodes. *Advantage:* The fuzzy logic is a multi-valued logic helpful in resolving ambiguity problems. It can deal with the measure of membership and measure of truth. The use of projected mold increases the CHD prediction accuracy. *Limitation:* Decision trees are mainly in decision analysis, which generates the rules and they are easily understood and perform huge data sets, capable to deal both numerical and categorical data. But, it creates much difficult trees that do not oversimplify the training data.

In the year 2014, [10]: *Methodology Employed:* In this review, the fuzzy rule depend on support system model is applied to forecast heart disease efficiently. This proposed system planned a fuzzy rule that depends on a skilled system and also applied data mining practice for tumbling the whole numeral of attributes. *Advantage:* The system has two main stages, one that performs classification and diagnosis very efficiently, and the other recognizes the threats of respiratory syndromes. In the Fuzzy Rule Base System, it takes part in the forecast of the heart disease. *Limitation:* The fuzzy rule assigns individual membership functions. This approach faces a loss of information.

In the year 2014, [11]: *Methodology Employed:* In this review, an Adaptive Network based Fuzzy Inference System (ANFIS) and Linear Discriminant Analysis (LDA) prediction method were proposed for capability of predicting the Coronary Heart disease. The projected mold might assist the doctors to diagnose CHD to predict the risk of CHD during diagnosis. *Advantage:* The projected mold helped to reduce the clinical data uncertainty and improves the prediction accuracy. The fuzzy rules express the multidimensional input gap configured with the input variables alienated into several sub spaces. Once the separation of the fuzzy space is completed, a fuzzy rule consequent to every fuzzy space is mandatory. The numeral of fuzzy rules and the rule arrangement are critical features for formative classification performance. *Limitation:* This Fuzzy Interference System and LDA prediction mold predicted the Heart diseases but the information loss was raised while predicting the disease.

In the year 2014, [12]: *Methodology Employed:* An Adaptive Neuro-Fuzzy Interference System (ANFIS) was utilized for categorization of heart diseases for serving patients and to predict early and consistent diagnosis. There are two stages for the proposed algorithm. They are advance stage and rearward stage. In the advance stage, a node results move further awaiting layer 4 and the subsequent factors are computed with Least Square Estimation (LSE). Then, error measure is calculated for each node. In the rearward stage, the error signs dispense with rearward stage to revise basis limits by gradient decline. *Advantage:* The ANFIS method in training error and tolerance error was very less. Three Gaussian membership functions were employed for measuring blood pressure and cholesterol. K-fold cross validation was employed, data set is separated into k equal size subsets randomly and the process is continual k times. All subsets are used for testing data and remaining of them are used for training data. *Limitation:* It can be unsafe in small training data collections, where an insignificance and unnecessary data is complex to estimate.

In the year 2014, [13], a decision support scheme was performed. It predicts the likelihood of heart disease of patients for coming years applying fuzzy logic and decision tree which extends their living period. Recommended strategies of blood pressure, total cholesterol, and LDL cholesterol efficiently predict CHD risk in old people. A heart disease prediction technique permits physicians to predict multi variety CHD pace in patients devoid of explicit CHD with 97.67% anticipated accuracy applying Fuzzy logic and decision tree with 1230 training data set.

In the year 2014, [14], "Fuzzy Rule Based Support System (FRBSS)" was proposed and modeled to predict heart disease intelligently and efficiently, and to replace manual efforts. Thus, the results could to more accurate. The system of work is modeled for diagnosis and perceives heart diseases. The scheme has two main stages. It presents classification and diagnosis. The authors tested and compared Neural Network and J48 Decision Tree mold to check presentation of the system.

In the year 2014, [15], in the case of a HDP, there were seven signs extracted from the client and furnished to the fuzzy scheme, and it illustrates the membership function of result chart. This fuzzy scheme gives about 90% of precise outcome. This investigation is not only for common patients, but also for the doctor to make accurate diagnosis and investigation of the CHF.

In the year 2014, [16], a fuzzy authority scheme was used based on Genetic Algorithm to diagnose the CAD disease condition. Genetic Algorithm is applied to optimize the membership function parameters. The future scheme is validated upon the CAD data set and achieved an accuracy of 88.79%. Detection of important attributes and fuzzy rules were obtained by applying the decision tree techniques. Pruning helps in falling the numeral of policy and the final set of rules provides better interpretability. The strength of this system was analyzed using the algorithms similar to classification accuracy, sensitivity and specificity and confusion matrix, and the comparison of the classification accuracy with the existing systems was made.

In the year 2014, [17]: *Methodology Employed:* Sets that, a new classifier, namely, rough-fuzzy classifier by blending rough set theory with the fuzzy set for predicting heart disease was proposed. Majorly, two essential steps are taken in the process of rule generation: core analysis and the indiscernibility matrix formation. Fuzzy scheme is intended by assistance of fuzzy rules of rough set theory and membership functions. So far, the occurrence of heart disease is identified by inputting the data to the fuzzy system. *Advantage:* The projected algorithm can be extended by including the associative analysis to find the related attribute and the rule strength computation. They can also be extended by including statistical measure. Here, the testing data with attributes were certain to the fuzzy logic system, where the test data was converted to the fuzzified value depending on the fuzzy membership function. The rule inference procedure is applied to obtain the linguistic value. Then, it is converted to fuzzy score using average weighted method. From the obtained fuzzy score, the decision is generated whether the test data belongs to the heart disease or not. *Limitation:* The fuzzy rough sets were experimented and by that all declining factors can be attained by applying the discernibility matrix. The discernibility matrix is to run all the attributes and the dimensions are expanded. The algorithm to run reduction is slow.

In the year 2013, [18]: *Methodology Employed:* In this review, FCM clustering algorithm was proposed for locating the peril of heart disease patient. In this FCM, the membership measure is relative to the values of the features in the clusters to the centroids rather than being proportional to the patterns in every cluster. *Advantage:* The FCM is an unsupervised clustering procedure, which allocates one fraction of data to fit into two or more clusters. The projected scheme will help the physicians to diagnosis the syndrome in a proficient way. This clustering algorithm can forecast the probability of patients to have heart attack in a more inexpensive way. It is incredibly effortless to measure the actual patients with abnormality (possibility of heart attack) while limiting the amount of counterfeit alarms with an assistance of FCM. *Limitation:* The FCM clustering technique locates centroid in the neighborhood of heart disease patient's data and misses the small, well-separated cluster.

In the year 2012, [19], a biased fuzzy rule was presented. It depends on Clinical Decision Support System (CDSS) for computer-aided diagnosis of the heart disease. The mechanical process to make the fuzzy laws is a benefit of the planned scheme and the biased process launched in the projected effort is an additional advantage for effective learning of the fuzzy system. The medical decision support scheme for risk prediction of the heart patients includes two phases, namely, (1) Creation of weighted fuzzy rules and (2) budding of a fuzzy rule-based decision support system. The appropriate attributes were produced by affecting the mining process and those attributes were applied to produce the fuzzy rules that were then weighted depending knowledge data sets. These weighted fuzzy rules were applied to construct the medical decision support scheme applying Mamdani fuzzy deduction scheme. Lastly, the carrying out tests were passed out on the UCI machine learning repository and the outcomes in risk prediction make sure that the projected clinical decision support scheme enhanced considerably contrasted by the network depending scheme in the conditions of accuracy, sensitivity and specificity.

3 DATA SOURCE

The data sets of Cleveland and Statlog are utilized. Cleveland data set comprises 303 patient records of heart disease data [20], and the Statlog data set comprises 270 patient records of heart disease data [20]. A total of 573 patient records are chosen for valuation of the proposed prediction scheme.

The data from the patient records will possess 13 input attributes, namely, age, sex, cp, t-rest-bps, chol, Rest-ecg, fbs, thalach, exang, old-peak, solpe, Ca, thal etc.. The detail descriptions of these attributes are provided in given below Table 1. As demonstrated in this document, the numbering for sections upper case Arabic numerals, then upper case Arabic numerals, separated by periods. Initial paragraphs after the section title are not indented. Only the initial, introductory paragraph has a drop cap.

3.1 Data Sets and Attributes

The data set was categorized with 3 attributes: Key, Predictable and Input attributes as given below:

A. *Key Attribute:* Patient Id: Patient's Recognition Number.
B. *Predictable Attribute:*
Diagnosis: Value 1 ≤ 50 % (no heart disease);
Value 0 ≥ 50 % (has heart disease)
C. *Input Aattributes* (Heart Disease): Table 1

Here, a key attribute of a patient record may be patient identification number (Patient ID), which is specified for every patient and can be easily distinguishable.

The below 13 attributes of data of a patient are said to be as input attributes which may have some personal information. Along with these attributes there is one more attribute for every record and can be helpful to identify it even among a number of similar records also termed as key attributes.

This database contains 76 attributes, but all published experiments refer to using a subset of 13 of them. In particular, the Cleveland, Statlog databases are used by ML researchers.. The "goal" field refers to the presence of heart disease in the patient. It is integer valued from 0 (no presence) to 4. Experiments with this database have concentrated on simply attempting to distinguish presence (values 1, 2, 3, 4) from absence (value 0). The names and social security numbers of the patients were recently removed from the database, replaced with dummy values.

Table: 1 Input Attributes for Prediction System

Variable Name	Attribute	Description	Values
f1	Age	Age in years	Continuous
f2	Sex	Male or Female	1 = male; 0 = female
f3	CP	Chest Pain Type	1 = typical type I 2 = typical type angina 3 = non-angina pain 4 = asymptomatic
f4	T rest bps	Resting blood pressure	Continuous value in mm hg.
f5	Chol.	Serum cholesterol	Continuous value in mg/dl.
f6	Restecg	Resting electrographic results	0 = normal 1 = containing ST_T sign odd 2 = left ventricular hypertrophy
f7	Fbs	Fasting blood sugar	1 ≥ 120 mg/dl. 0 ≤ 120 mg/dl.
f8	Thalach	Maximum heart rate attained	Continuous value
f9	Exang	Exercise make angina	0 = no; 1 = yes
f10	Oldpeak	ST despair make by exercise virtual to rest	Continuous value
f11	Solpe	Slope of the crest exercise ST fragment	1 = un sloping 2 = fat 3 = down sloping
f12	Ca	Number of main vessels tinted by floursopy	0 - 3 value
f13	Thal	Defect type	3 = normal 6 = fixed 7 = reversible defect

4 CLASSIFICATIONS AND PREDICTION

In data mining Classification and Prediction are the common types of applications. Classification is a process used to predict collection of data instances. A model of classes is formed from a set of samples with a number of attributes and a class label. It aim is to classify things into categories or classes. The model allots suitable class label to new data. The worth of the model is evaluated by means of classification accuracy measure. The term Prediction can estimate all classified instances. The process of separating a data set into mutually exclusive groups. Such that the members of each group are as "close" as feasible to one another, and different groups are as "far" as possible from one another.

The foundation of the algorithm is to allocate membership and ownership as a function of the vector's distance from its *Fuzzy-Rough Nearest Neighbors' Classification* and those memberships and ownerships of their neighbors' in the feasible classes. Further than obtaining these K samples, the procedures be different significantly. Jensen and Chris Cornelis [22] have proposed *FRNN Classification Algorithm* and a brief description of algorithm is mentioned below.

While rough set theory handles only one type of deficiency model in data, it is corresponding to other concepts for the purpose, such as fuzzy set theory. Two fields might be considered related in the logic that both can stand inconsistency and uncertainty, the variation being the type of uncertainty and their approach to it, fuzzy sets are concerned by vagueness, rough sets are concerned by indiscernibility. A simple (non-fuzzy) *K-nearest neighbor classifier (KNN)* to facilitate the Euclidean distance to calculate the nearby neighbor (or neighbors' if more than one object has the closest distance) in the training data, and outputs this object's assessment as its prediction.

4.1 Fuzzy Nearest Neighbour Classification

To classify the test objects based on their relationship to a given number K of neighbors' (along with the training objects), and along with neighbors' membership degree to (crisp or fuzzy) class labels presented the modified *fuzzy K-Nearest Neighbor (F-KNN) algorithm* to given below.

Table 2: The Fuzzy K-NN Algorithm

FNN (A, M, a, K):
Input:'A' is a set of training data;
'M' is a set of decision classes;
'b' is a object to be classified;
'K' is the number of nearest neighbours.
Begin
N ← call procedure to get nearest neighbors (b, K)
for all $M \in m$ do
M(b) = $\sum_{a \in N} R[a, b]\, M(a)$ //the range of M(b) to wherever an unclassified object "b' be a member of class 'M'//
End for
Print $arg\, max_{M \in m}(M(b))$
Return 'b'
End

Intended for the cause of *FNN*, the range of M(b) to wherever an unclassified object "b' be a member of class 'M' is evaluated as:

$$M(b) = \sum_{a \in N} R(a, b) M(a)$$

Where N is the set of object 'b's K nearest neighbour, achieved by computing the fuzzy sameness among y and all training objects, and selecting the K objects that have maximum sameness degree, and R(a, b) is the [0, 1], valued similarity of 'a' and 'b' by the standard approach.

$$\text{Where } R(a, b) = \|b - a\|^{-2/(p-1)} / \sum_{i \in N} \|b - j\|^{-2/(p-1)}$$

Here ||·|| indicates the Euclidean norm, and 'p' is a parameter that command the whole weighting of the sameness. FNN algorithm illustrate that classify a test object 'b' to the class with the maximum resulting membership.

The main idea of this algorithm is that the degree of familiarity of neighbours should influence the impact that their class membership has on developing the class membership for the test object. This algorithm complexity for the classification of one test pattern is O (|A|+K·|A|).

4.2 Fuzzy-Rough Ownership

A fuzzy-rough ownership function is built that efforts to grip both "fuzzy uncertainty" (caused by overlapping classes) and "rough uncertainty" (caused by deficient information, i.e., attributes, about the objects).

The fuzzy-rough ownership function 'τ_m'of class 'M' is defined as, for an object 'b' as follows,

$$\tau_m = \sum_{a \in U} R(a, b) M(a) / |\mathbb{A}|$$

In this, the fuzzy relation R is evaluated by:

$$R(a, b) = \exp(-\sum_{x \in A} K_x (x(b) - x(a))^{2/(p-1)})$$

Where 'p' is weighting of the similarity (as in FNN) and K_x is a parameter that chooses the bandwidth of the membership, defined as:

$$K_x = |\mathbb{A}| / 2 \sum_{a \in U} \|x(b) - x(a)\|^{2/(p-1)}$$

The confidence of 'b' can be classified to class M evaluated by $\tau_m(b)$. The consistent crisp classification algorithm is known as FRNN-O. Primarily, the parameter K_x is intended for every attribute and entirely memberships of decision classes for test object 'b' are set to 0.

Then, the weighted distance of 'b' from every object in the creation is calculated and recycled to modify the class memberships of 'b'. Finally every training object can be considered, and this algorithm gives heights membership in class. The algorithm's complexity is O (|X|.|A|+|A|·(|X|+|M|)).

Table 3: The Fuzzy-Rough Ownership Nearest neighbor Algorithm

```
FRNN - O (A,𝔸, M, b):
Input: 'A' is a set of training data;
       '𝔸' is a set of uncertain characteristics;
       'M' is set of conclusion classes;
       'b' is the object to be classified.
Begin
   for allσ ∈ 𝔸 do
      K_x = |𝔸|/2 Σ_{a∈U} ||x[b] − x[a]||^{2/p−1}     //K_x is a parameter
           that decides thebandwidth of the membership//
   End for
   N ← |𝔸|
   for all M∈m do
      τ_m[b] = 0;        //τ_m(b)is interpreted as the confidence
           with which b can be classified to class M//
   End for
   for all a∈N do
      z = Σ_{x∈M} K_x(x[b] − x[a])^2
      for all M∈m d
         τ_m[b] = M[a].exp(−z^{(p−1)})/|N|
      End for
   End for
   Print arg max_{M∈m} τ_m[b]
   Return 'b'
End
```

4.3 Fuzzy-Rough Nearest Neighbour

Proposed algorithm merged a fuzzy-rough approximation among the ideas of the classical. The foundation of following algorithms are that the means of the nearest neighbours of a test object 'b' can evaluate the lower and upper approximation of a decision class gives excellent clues to predict the membership of the test object to the class. Moreover, the algorithm is based on the choice of the fuzzy tolerance relation R.

Table 4: The Fuzzy-Rough Nearest neighbour Algorithm – Classification

```
FRNN - 1 (A, M, b):
Input: 'A' is a set of training data;
       'M' is a set of decision classes;
       'b' is the object to be classified.
Begin
   N ← call procedure to get nearest neighbors (b, K)
   τ ← 0, Class ← ϕ
   for all M∈m do
      if ((R↓M)(b) + (R↑M)(b))/ 2 ≥ τ then
         Class ← M
         τ ←[(R↓M)(b) + (R↑M)(b)]/ 2
      end if
   end for
   Print Class
   Return 'b'  //Class for b//
End
```

A general way of establishing R is as follows given the set of conditional attributes 'M', R is defined by:

$$R(a, b) = \min_{x \in M} R_x(a, b)$$

Where R_x (a, b) is the degree for attribute 'x' to which objects 'a' and 'b' are similar. Possible options include.

$$R_x(a,b) = 1 - [|x(a) - x(b)|] / [|x_{max} - x_{min}|]$$

$$R_x^1(a, b) = \exp[-((x(a) - x(b))^2 / 2\sigma^2]$$

In which x_{max} and x_{min} are the maximal, minimal occurring values of that attribute, and σ^2 is the variance of attribute 'x'

The foundation after the algorithm is gives to the lower and the upper approximation of a decision class, evaluated via means of the nearest neighbours of a test object 'b', deliver worthy hints to predict the membership of the test object to that class. In specific, if (R↓M)(b) is maximum, it imitates the all of 'b's neighbours belong to M, however a maximum value of (R↑M)(b) means is at least one neighbour goes to that class.

Here, the lower and upper approximations are evaluated from the following equations:

$$(R{\downarrow}R_{z^D})(a) = inf_{b \in N} \tau(R(a, b), R_z(b, d)$$

$$(R{\uparrow}R_{z^D})(a) = sup_{b \in N} \tau(R(a, b), R_z(b, d)$$

Where R_z is the fuzzy tolerance relation for the decision feature 'z'.

Table 5: The Fuzzy-Rough Nearest neighbour Algorithm – Prediction

```
FRNN - 2 (A, z, b):
Input: 'A' is a set of training data;
       'z' is a set of decision feature;
       'b' is the object to be classified.
Begin
   N ← call procedure to get nearest neighbors (b, K)
   τ_1 ← 0, τ_2 ← 0
   for all d ∈ N do
      C← [(R↓R_zD))(b) + (R↑R_zD))(b)]/ 2
      τ_1 ← τ_1 + C *z(b)
      τ_2 ← τ_2 + M
   End for
   If τ_2> 0 then
      Return τ_1/τ_2
   Else
      Return Σ_{a∈N} z(a)/|N|
   End if
End
```

Every *NN* approach is run two times, the first time location K=10 and the second time with K set to the occupied set of training objects. This is calculated via 2x10 fold cross validation. C is set to 2 for *FNN* and *FRNN-O*. The new approaches were selected from the fuzzy relation of $R_?^2(a,b)$. While using FRNN – FRS, the use of K is not required it proposition as R(a, b) gets smaller; 'a' tends to have only have a minor control on (R↓M)(b) and (R↑M)(b).

5 CORRELATIVE EXAMINATION OF CLASIFIERS

The dataset made of 573 records among a set of 13 attributes. During the collection of a few missing values are found because of manual write down by doctors. By means of data missing Weka 3.6.6 tool the missing values are replaced by an appropriate data by way of mean, mode method by Replaced Missing Values filter option available in the tool.

To get better prediction of classifiers, Fuzzy-Rough logic combined. The classifiers such as *K-NN, Fuzzy K-NN, and Fuzzy-Rough NN (FRNN)* are applied for diagnosis of patients getting heart disease. The classifiers were giving with data set by fuzzy-rough sets. A fuzzy-rough ownership function is formulated that attempts to handle mutually fuzzy uncertainty (caused by overlapping classes) and rough uncertainty (caused by insufficient knowledge, i.e., attributes, about the objects). After merged fuzzy–rough technique results are shown in Table. 6, observations display that the *Fuzzy–Rough NN (FRNN)* data mining technique exceeded other two data mining techniques.

Table 6: Comparative of three classifiers

Data Mining Techniques	**Accuracy**	**Mean Absolute Error**
K-NN	56.32	0.059
Fuzzy K-NN	97.86	0.008
Rough-Fuzzy NN	99.85	0.0016

A confusion matrix is obtained to compute the accuracy of classification. A confusion matrix shows how many instances have been assigned to every class.. In this present work, these two sets are training and testing and it has a 2×2 confusion matrix.

Table 7: Class A = YES (has heart disease);
Class B = NO (No heart disease)

	A (has heart disease)	**B (no heart disease)**
A (has heart disease)	True Positive	False Negative
B (no heart disease)	False Positive	True Negative

The data set made of 573 records and it has been divided into on 25 classes, where each class consists of 23 records. Training and Testing sets are divided into equal amount to predict the exactness of the system. To measured certainty of 13 attributes are chosen to classify the system by a symbolic learning approach by interval method, i.e., (μ-σ, μ+σ) as indicate in the Table 1.

In this research work trained the classifiers to classify the clinical data set as either ''has heart disease'' or ''no heart disease''. The accuracy of a classifier can be calculated using precision and recall. The general and particular of confusion matrixes are two classes (i.e. healthy and sick) of three classifiers are shown in Tables (8, 9, 10). It is a significant measure for inspecting the how well classifier can recognize dissimilar classes of tuples. Examine in expressions of positive tuples (eg. diagnosis = healthy) against from negative tuples (eg. diagnosis = sick) for the given two classes.

Table 8: Confusion matrix access from *k-nn* classifier

	A	B
A	234	25
B	17	254

Table 9: Confusion matrix access from *f-knn* classifier

	A	B
A	253	11
B	12	265

Table 10: Confusion matrix access from *frnn* classifier

	A	B
A	273	4
B	3	298

The symbolic classifies for dissimilar values of ''K'' are shown in Table 11.

Table 11: Overall accuracy access for data set among unreliable K (*k-nn, fuzzy k-nn, fuzzy-rough c*lassifiers)

K	**K-NN Accuracy**	**Fuzzy K-NN Accuracy**	**Fuzzy-Rough NN Accuracy**
1	0.55	0.97	0.99
3	0.48	0.87	0.95
5	0.45	0.84	0.93
7	0.44	0.81	0.91
9	0.42	0.80	0.89

To show the difference between the performances of three classifiers, it is visualized in Figure 1.

Figures 1, shows the performance comparison of all three classifiers among 10 fold cross validation and fold cross validation to 10 fold cross validation. It is obviously states that the evaluated measurement of the performance based on 13 attributes are accurate in natural history and the whole accuracy accessed from the dataset among unreliable with different classifiers.

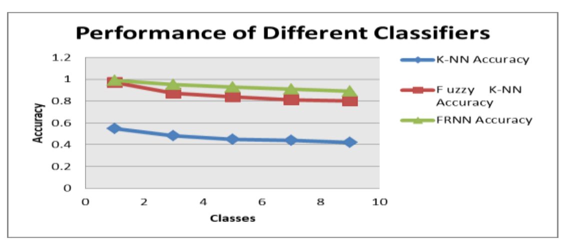

Figure 1: Performance assessment of all the classifiers among 10 fold cross validation accuracy

It was experimental that from Figure 1, the accuracy of predicting a heart disease patient with the 13 attributes states in Table 1 is extended up to 99% and *Fuzzy-Rough NN (FRNN)* is improved performance than other classifiers, while *K-NN* shows dispense performance than other classifiers.

6 CONCLUSION

The intention of work is to predict the more accurate presence of heart disease. Initially, 13 attributes were engaged in predicting the heart disease. In this work, the classifiers *K-NN, Fuzzy K-NN, Fuzzy-Rough NN (FRNN)* were handled to predict the heart disease. The both fuzzy uncertainty (caused by overlapping classes) and rough uncertainty (caused by insufficient knowledge, i.e. attributes, about the objects) handle by a fuzzy-rough ownership function is constructed for those uncertainty. Also, the observations presented that the *FRNN* technique is outperformed. Further extend our work applying the other data mining techniques, with Rough set to evaluate the strength of heart disease and also provide for data compression.

ACKNOWLEDGMENT

I would like sincerely to express my gratitude for my management of *Neil Gogte Institute of Technology (NGIT)*, and also the Department of Computer Science and Engineering.

REFERENCES

[1] Gopal. S., and C. Woodcock, "THEORY AND METHODS FOR ACCUR CY ASSESSMENT THEMATIC MAPS USING FUZZY SETS", Photo Grammatik Engineering and Remote Sensing 60: 1994, Page no. 181- 188.

[2] Komorowski. J, Pawalak, Z. Polkowski. I, Skowron.A."Rough sets: A tutorial", In: S.k.paland A. Skowron (Eds). "Rough-Fuzzy Hybridization", A New Trend in Decision-Making. Singapore, Springer-Verlag. 1998, Page no. 3- 98.

[3] Pawlak, Z. "Rough sets" In: Int. J. of Information and Computer Sciences, 11, 5, 1982, Pg. no. 341 - 356.

[4] Pawlak. Z. " A primer on Rough Sets: A New Approach to Drawing Conclusions from Data". In: Cardozo Law Review. Volume 22, issue 5-6, July '2001, Page no. 1407-1415.

[5] Nijaguna Gollara Siddappa, Thippeswamy Kampalappa, 'Adaptive Condensed Nearest Neighbor for Imbalance Data Classification", International Journal of Intelligent Engineering and Systems (IJIES), Volume-12, Number 2, 2019 DOI: 10.22266 / ijies 2019. 0430.11, Pp. 104 - 113.

[6] Kiyong Noh, Heon Gyu Lee, Ho-Sun Shon, Bum Ju Lee, and Keun Ho Ryu, "Associative Classification Approach for Diagnosing Cardiovascular Disease", Springer 2006, Vol. 345, Pp. 721- 727.

[7] Samuel, O. W., et., al., "An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction", Expert Systems with Applications, Vol. 68, Pp. 163 - 172, 2017.

[8] Kumar, et., all, "Neighborhood rough set based ECG signal classification for diagnosis of cardiac diseases", Soft Computing, Pp. 1 - 13, 2016.

[9] Kim, Jaekwon, et., al., "Data-mining-based coronary heart disease risk prediction model using fuzzy logic and decision tree", Healthcare informatics research, Vol. 21, Issue. 3, Pp. 167 - 174, 2015.

[10] Oad, Kantesh Kumar, XuDeZhi, and Pinial Khan Butt, "A fuzzy rule based approach to predict risk level of heart disease", Global Journal of Computer Science and Technology (GJCST), Vol. 14, Issue. 3-C, Pp. 17-26, 2014.

[11] Yang, J. G., Kim, J. K., Kang, U. G., & Lee, Y. H. "Coronary heart disease optimization system on adaptive-network-based fuzzy inference system and linear discriminant analysis (ANFIS–LDA)", Personal and ubiquitous computing, 18(6), Pp. 1351-1362, 2014.

[12] Ziasabounchi, et., all, "ANFIS based classification model for heart disease prediction", International Journal of Electrical & Computer Sciences-Network Security) (IJECS-IJENS), 14.02, Pp. 7-12, 2014.

[13] K. Cinetha, et., al., "Decision Support System for Precluding Coronary Heart Disease (CHD) Using Fuzzy Logic", International Journal Computer Science and Technology (IJCST) – Volume 2, Issue 2, ISSN: 2347-857, Pp.102 - 107, Mar-Apt 2014.

[14] Kantesh Kum ar Oad & Xu DeZhi, " A Fuzzy Rule based Approach to Predict Risk Level of Heart Disease". Global Journal of Computer Science and Technology (GJCST), (0975-4350), Volume No. 14, Issue 3, Version 1, Pp.17 - 22, 2014.

[15] Sharat Chandra, Ripu Ranjan Sinha, "Fuzzy based Congestive Heart Failure Diagnosis and Analysis", International Journal of Computer Application (IJCA), (0975 – 8887), Volume No.105 – No. 6, Pp. 5 - 8, November 2014.

[16] Y. Niranjana Devi, S. Anto, "An Evolutionary-Fuzzy Expert System for the Diagnosis of Coronary Artery Disease", International Journal of Advanced Research and Computer Engineering and Technology (IJARCET), ISSN: 2278 – 1323, Volume 3, Issue 4, Pp.1478 - 1484, April 2014.

[17] Srinivas, K, et. all, "Rough-Fuzzy classifier: A system to predict the heart disease by blending two different set theories", Arabian Journal for Science and Engineering (AJSE): 39.4, Pp. 2857 - 2868, 2014.

[18] Chitra R, et., al., "Heart attack prediction system using Fuzzy C Means classifier", IOSR JCE, Vol. 14, Pp. 23 - 31, 2013.

[19] P. K. Anooj,"Clinical decision supports system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University (JKSU), Computer and Information Sciences: 24, Pp, 27 – 40, 2012.

[20] Clevelanddatabase:http://archive.ics.uci.edu/ml/datsets/Heart+disease

[21] Stalogdatabase:http://archive.ics.uci.edu/ml/machine-learning_databases/stalog/heart/.

[22] Richard Jensen, ChrisCornelis, "Fuzzy Rough NearestNeighbour Classification", IEEE.

BIOGRAPHIES

I, Dr. V. Krishnaiah having 20 years of teaching experience in reputed Engineering Institutions and presently working as Associate Professor, Dept. of CSE, *Neil Gogte Institute of Technology (NGIT)*, Hyderabad, Telangana State, India. I received Ph.D.(Computer Science & Engineering), and M.Tech.(Computer Science & Engineering) both from College of Engineering (CEH), Jawaharlal Nehru Technological University Hyderabad (JNTUH), Hyderabad, Telangana State, India; M.Phil.(Maths) from Alagappa University (AU), Karaikudi, Tamilnadu State, India; and M.Sc.(Maths) from Osmania University (OU), Hyderabad, Telangana State, India.

Published various publications in International / National Journals and also presented papers in National / International conferences. And my Google Scholar: Citations – **460+**, H Index – **06**; i10 Index – **06**.https://scholar.google.co.in/citations?user=mP3O0AYAAAAJ&hl=en.

My research interests include Data Mining, Network Security, Image Processing, Machine Learning, and Data Science, AIML.

Awarded as "*Certificate of Excellence in Reviewing*" in, 'In recognition of an outstanding contribution to the quality of journal", from '*South Asian Journal of Social Studies and Economics' (SAJSSE)*, UK, Landon.