Miskolc Mathematical Notes
Vol. 23 (2022), No. 2, pp. 787–800

HU e-ISSN 1787-2413
DOI: 10.18514/MMN.2022.3855

USING FUZZY-ROUGH SET EVALUATION FOR FEATURE SELECTION AND NAIVE BAYES TO CLASSIFY THE PARKINSON DISEASE

NAIYER MOHAMMADI LANBARAN, ERCAN ÇELİK, AND ÖZGÜR KOTAN

Received 26 May, 2021

Abstract. Feature selection is one of the issues in machine learning as well as statistical pattern recognition. This is important in many fields (such as classification) because there are many features in these areas, many of which are either unused or have little information load. Not eliminating these features does not make a problem in terms of information, but it does increase the computational burden for the intended application. Besides, it causes to store of so much useless information along with useful data. A problem for machine learning research occurs when there are many possible features with few attributes of training data. One way is to first specify the best attributes for prediction and then to classify features based on a measure of their dependence. In this study, the Fuzzy- Rough subset evaluation has been used to take features in core of similar features. Fuzzy-rough set-based feature selection (FS) has been demonstrated to be extremely advantageous at reducing dataset size but has various problems that yield it unproductive for big datasets. Fuzzy- Rough subset evaluation algorithm indicates that the techniques greatly decrease dimensionality while keeping classification accuracy. This paper considers classifying attributes by using fuzzy set similarity measures as well as the dependency degree as a relatedness measure. Here we use Artificial Neural Network, Naive Bayes as classifiers, and the performance of these techniques are compared by accuracy, precision, recall, and F-measure metrics.

2010 *Mathematics Subject Classification:* 303B52; 68T27;68T37;94D05

Keywords: Fuzzy- Rough subset evaluation, Feature selection, Artificial Neural- Network, Naive Bayes, Parkinson's disease, models

1. INTRODUCTION

These days with the increasing of computer and database technology, a large number of features can be obtained and saved in databases for various real-world implementations. Some of the features could be irrelevant or redundant for classifying learning; they could significantly lead to reduces efficiency and accuracy of classifiers and causes utmost computational confusion. Consequently, for employing a data collection it is needed to prepossess the data to extract inessential attributes. Attribute choosing or feature selection, a significant method for declining the number of unneeded attributes, is applied to figure out an ideal attribute subclass to carry out classification on the assumption that keeping categorization accuracy. Of late years, attribute extraction has been comprehensively utilizing in data procedure, template identification, and machine learning [9], [12], [14], [24], [25].

The standard of the machine-legible input that works on it usually contributes to the success of machine learning algorithms. These items comprise whether there is indifferent, needless, untrustworthy, or noisy data. Data mining, which has started to gain great importance recently, has achieved great success in interpreting big data by processing; It makes predictions effective in solving major problems in many sectors, especially finance, health, communication, and education [15].

The use of feature selection, artificial intelligence, and machine learning applications in the field of health is carried out in many sub-fields such as medical diagnosis and disease tracking, cost estimation, imaging analysis, resource planning, and emergency management, processing of unstructured data [3, 4, 16, 18]. Artificial intelligence models, which are also used to make high-scale patient data functional, play an active role in increasing data reliability and quality [6], [23]. However, regarding the use of artificial intelligence in the field of health; The accuracy of clinical data, data management, legal and ethical processes regarding data protection limit the use of artificial intelligence in the field of health. Machine learning techniques are most commonly used in the field of health sciences for predicting, diagnosis, and determining post-illness complications, and it is aimed to provide patients with better quality healthcare services by saving time and workload [5].

Indeed, it is stated that machine learning-based classification methods play a decisive role in both decision support systems and disease diagnosis in today's medical research [13].

Recently, there has been a great deal of interest in evolving methods that able to deal with inaccuracies and uncertainties, and a significant amount of them is in the field of fuzzy and rough sets. Rough set theory's success is owing to some particulars of the theory. Merely the facts covered in the data are investigated. No further information about the inputs is needed for data analysis such as thresholds or proficient knowledge on a specific field. And it detects the minutest knowledge representation for data. While rough set theory controls only one type of handicap that exists in data, it is supplementary to other concepts for the goal, like fuzzy set theory. Fuzzy sets are deal with ambiguity, and rough sets are cover the undiscriminating [17]. Parkinson's disease (PD) appears as the decease of dopaminergic neurons in the substantiate nigra pars condensed inside the midbrain. This neuro damage caused a domain of symptoms containing coordination outcomes, bradykinesia, vocal changes, and toughness. Dysarthria is also detected in PD patients; it is specified by laxity, paralysis, and deficiency of coordination in the motor-speech system: affecting respiration, phonation, articulation, and prosody. whereas symptoms and the disease period change, PD is mostly not distinguished for many years. accordingly, there is a necessity for more accurate diagnostic implements for PD finding because, as the disease moves forward, more symptoms become apparent that make PD more difficult to treat. Therefore, a big deal of endeavor has been made to extend methods for early detection, mostly at pre-symptomatic phases in order to slow or stop disease forward movement. The rapid advance in machine learning techniques has made it challenging to combine large-scale, high-dimensional objects. Thus, it has extended quickly in computer-aided machine learning approaches for Combined analysis. Well-known pattern analysis methods, such as Artificial Neural Network (ANN), Naive Byes (NB), have been used for early detection of PD and the prediction of PD progression [8], [19].

2. MATERIALS AND METHODS

Fuzzy sets were presented and expressed using membership functions by L.A. Zadeh in 1965 and have many convenient utilizations [27]. For the aim to implement fuzzy set similarity measures to specify concurrence between two distinct features, each feature is expressed as a fuzzy set over the patient's data sets. The dataset must be normalized to specify a degree of membership in $[0,1]$. The patient's membership degree in the fuzzy set specifies the level of certainty for that patient's data. Suppose U is a reference set (finite and non-null collection of items) and R is a non-null set of finite features.

Definition 1. In the classical set theory an element must belong or not belong to a set. In fuzzy theory, an element can belong to a set by k degree $(0 \leq k \leq 1)$. The fuzzy belonging function is shown such as $\mu_A(x) \in (0,1)$ where A is an element collection, x is an object, and A is a fuzzy set which is given below [21].

Definition 2. A quaternary (U,R,K_S,S) gives a fuzzy information system (FIS). K_S is the set of entire fuzzy numbers and S is a knowledge operation which F (x, r) $=$ $\mu_r(x), \forall x \in U, r \in R$ and $S : U \times R \to K_S$ and $\mu_r(x)$ is membership degree.

Definition 3. A rough set concept is another way to deal with ambiguity. Unlike the fuzzy set, the uncertainty in the rough set is determined with a boundary area, not through a partial membership function. Internal topological functions and closures as estimation can define a rough set. U is a given universe and $R \subseteq U \times U$ is an indiscernibility connection which demonstrates our information shortage about members of U. Let R is equivalence intercourse and $X \subseteq U$. Now determine the set X with about R through a primary hypothesis of rough set theory.

- R-`lower approximation of` X: $\underline{R}(x) = \bigcup \{R(x) : R(x) \subseteq X\}$
- R-`upper approximation of` X: $\overline{R}(x) = \bigcup \{R(x) : R(x) \cap X \neq \varnothing\}$
- R-`boundary approximation of` X: $RnR(x) = \overline{R}(x) - \underline{R}(x)$

The rough set membership function is described as:

$\mu_X R : U \to (0,1)$ where $\mu_X R(x) = \dfrac{|X \cap R(x)|}{|R(x)|}$ and $|R(x)|$ denotes the cardinality of x.

Definition 4. A quaternary (U, R, K_S, S) gives a fuzzy information system (FIS). K_S is the set of all fuzzy numbers and S is an information function which F (x, r) $=$ $\mu_r(x), \forall x \in U, r \in R$ and $S : U \times R \rightarrow K_S$ and $\mu_r(x)$ is membership degree.

Definition 5. Let $R = (U, A)$ be an information framework, that U is a not empty collection of limited objects (the universe of discourse), and A is a nonempty limited collection of features which $a : U \rightarrow V_a$; $\forall a \in A$. V_a is the collection of values that criteria a can receive.

$$IND(P) = \left\{(x, y) \in U^2 | \forall a \in P,\ a(x) = a(y)\right\}$$

The division of U , which is produced by $IND(P)$, is demonstrated $U/IND(P)$ (or U/P for simplicity) and can be computed as:

$$U/IND(P) = \bigotimes \left\{U/IND(\{a\}) | a \in P\right\}$$

where $\bigotimes$ is particularly defined as follows for sets A and B:

$$A \bigotimes B = \{X \cap Y | X \in A, Y \in B, X \cap Y \neq \varnothing\}$$

If $(x, y) \in IND(P)$, then x and y are indistinguishable by criteria from P. The equation classes of the P indistinguishable relationship are demonstrated $[x]_P$.

Let $X \subseteq U$. X can be approximated by employing only the knowledge contained within P by setting up the $P-lower$ and $P-upper$ approximations of X:

$$\underline{P}X = \{x \in U \mid [x]_P \subseteq X\}$$
$$\overline{P}X = \{x \in U \mid [x]_P \cap X \neq \varnothing\}$$

$\langle \underline{P}X, \overline{P}X \rangle$ is labelled as a rough set. Let P and Q be sets of elements inducing equivalence relationship over U, then the positive, negative, and boundary regions can be described as:

$$POS_P(Q) = \bigcup_{X \in U/Q} \underline{P}X$$

$$NEG_P(Q) = U - \bigcup_{X \in U/Q} \overline{P}X$$

$$BND_P(Q) = \bigcup_{X \in U/Q} \overline{P}X - \bigcup_{X \in U/Q} \underline{P}X$$

The positive region comprises all items of U that can be categorized into classes of U/Q by employing the information in attribute P. The boundary region is the set of attributes that can possibly, but not sure, be categorized in this method. The negative region is the set of elements that cannot be categorized into classes of U/Q. A significant subject in data mining is finding dependencies between features. intuitionally, a set of features Q depends entirely on a set of attributes P, which is demonstrated by $P \Rightarrow Q$, if all data values from Q are uniquely specified by values of objects from P . If there is an operative dependency between values of Q and P, then Q depends entirely on P. $\forall P, Q \subset A$, Q depends on P in a degree $k(0 \leq k \leq 1)$, which is denoted $P \Rightarrow k\ Q$, if:

$$k = \gamma_P = \frac{|POS_P(Q)|}{|U|}$$

- If $k = 1$, Q depends entirely on P;
- if $0 < k < 1$,Q depends partially (in a degree k) on P;
- if $k = 0$, then Q does not depend on P.

By computing the change in dependency when a feature is eliminated from the set of observed conditional features, a scale of the importance of the attribute can be acquired. The vast the change in dependency, the more important the attribute is. If the importance is 0, then the attribute is dispensable. More officially, given P, Q, and an attribute $a \in P$ [2]:

$$\sigma_P(Q,a) = \gamma_P(Q) - \gamma_{P-a}(Q)$$

Definition 6. A description for fuzzy $P-$lower and $P-$upper approximations was identified as follows:

$$\mu_{\underline{P}X}(F_i) = \inf_x \max\{1 - \mu_{F_i}(x), \mu_x(x)\} \qquad \forall i$$
$$\mu_{\overline{P}X}(F_i) = \sup_x \min\{\mu_{F_i}(x), \mu_x(x)\} \qquad \forall i$$

where F_i is a fuzzy equivalency class, and X is the (fuzzy) notion to be approximated.

The $\langle \mu_{\underline{P}X}, \mu_{\overline{P}X} \rangle$ is described as a fuzzy-rough set. These definitions disunite a concise from the crisp upper and lower approximations, as the memberships of single attributes to the approximations are not clearly accessible. Thereupon, the fuzzy lower and upper approximations are described as follows in [7].

$$\mu_{\underline{P}X}(x) = \sup_{F \in U/P} \min\left(\mu_F(x),\ \inf_{y \in U} \max\{1 - \mu_F(y), \mu_X(y)\}\right)$$
$$\mu_{\overline{P}X}(x) = \sup_{F \in U/P} \min\left(\mu_F(x),\ \sup_{y \in U} \min\{\mu_F(y), \mu_X(y)\}\right)$$

Definition 7. Fuzzy-Rough Reduction procedure: Fuzzy-rough set-based FS constructs on the idea of the fuzzy lower approximation to provide a reduction of datasets comprising real-valued features. The procedure will be the same with a crisp approach when the attributes are well-defined nominal attributes. The membership of an attribute $x \in U$ be the property of the fuzzy positive region can be determined by:

$$\mu_{POS_P(Q)}(x) = \sup_{X \in U/Q} \mu_{\overline{P}X}(x)$$

The fuzzy- rough dependency function can be determined as come after:

$$\gamma'_P(Q) = \frac{|\mu_{POS_P(Q)}(x)|}{|U|} = \frac{\sum_{x \in U} \mu_{POS_P(Q)}(x)}{|U|}$$

If the fuzzy-rough diminution process is to be a utility, it should be able to struggle with various features by calculating the dependency between the multiple subsets of the original attribute set. In the fuzzy condition, objects could belong to various equivalence classes, and thus, the Cartesian product of $U/IND(\{a\})$ and $U/IND(\{b\})$ must be noted in assigning U/P .

$$(U/P) = \bigotimes \{U/IND(a) | a \in P\}$$

where

$$A = \{X \cap Y | X \in A, Y \in B, X \cap Y \neq \varnothing\}$$

Each set in U/P indicate an equivalence class [22].

Definition 8. A fuzzy knowledge framework is determined as (U, B, V_F, F), which V_F is the set of whole fuzzy data and F is an knowledge assignment determined as $F : U \times B \to V_F$.

Such that F(x, b) $= \mu_b(x), \forall x \in U$ and $b \in B$, where $\mu_b(x)$ is a membership grade of the criteria x for the feature b .

Definition 9. Let FDS be a fuzzy decision system and fuzzy resemblance among two criteria for every feature could be determined as [10]:

$$SIM_a(x_i, x_j) = 1 - \frac{|\mu_a(x_i) - \mu_a(x_j)|}{|\mu_{a_{\max}} - \mu_{a_{\min}}|}$$

That, $\mu_a(x_i)$ and $\mu_a(x_j)$ are membership function of criteria x_i, x_j separately and $\mu_{a_{\max}}$ and $\mu_{a_{\min}}$ are maximum and minimum membership function for a criterion "a" respectively [20].

For a subset of attributes P,

$$(x_i, x_j) \in SIM_P^{\delta} \text{ iff } SIM_a(x_i, x_j) \geq \delta$$

where δ is a resemblance entrance and tolerance classes are produced by fuzzy similarity connection as:

$$SIM_P^{\delta}(x_i) = \{x_j \in U | (x_i, x_j) \in SIM_P^{\delta}\}$$

So, lower and upper approximations of $X \subseteq U$ are expressed as:

$$\underline{P}^{\delta} X = \{x_i | SIM_P^{\delta}(x_i) \subseteq X\}$$

$$\overline{P}^{\delta}X = \{x_i | SIM_P^{\delta}(x_i) \cap X \neq \varnothing\}$$

The pair ($\underline{P}^{\delta}X$, $\overline{P}^{\delta}X$) is named tolerance fuzzy rough set. The positive region and dependency degree determined as before.

In each phase, we append one feature in the decrement collection and compute the grade of dependency, when there is not increase in the amount of dependency, the algorithm ends.

3. THE RESEARCH FINDINGS AND DISCUSSION

Fuzzy logic employs linguistic variables, determined as fuzzy sets, to approach human reasoning. The used feature selection technic gets pre-processed data set as an input and generates ranked attributes based on the got-together approach of $F-$Score and Accuracy on data set statement values normalized by fuzzy Gaussian membership function. we used fuzzy- rough subset evaluation to feature selection. These top- n selected data are used by the Fuzzy- discernibility and Neural- a network for classification. The process is explained with the following steps:

a. `Normalization` Before doing anything with the dataset, it must be normalized. Occasionally, some values may be missing in the real valued dataset, there are operations such as filling or eliminating these values in preparation. And sometimes the scalars of the data are not the same, so it is necessary to bring them to the same proportions, and this is one of the works done in preparation as normalization or standardization. Sometimes there are nominal values that must be change to numeric values. Fuzzification is one of those normalization which based on fuzzy membership function. Fuzzification is the procedure of turning crisp inputs into fuzzy utilities. The membership functions supplied membership values indicate the degree of membership of a linguistic term.

   The datasets used in this study comprise expression data for a set of attributes existing at https://www.kaggle.com/c/parkinsons-detection. This work is applied using Weka version 3.9.5.

b. `Attribute Selection` After normalized the dataset it is time to attribute selection. In this study as said before we applied a fuzzy rough subset evaluation technique for attribute selection and the searching method is Hill Climber. In this method, the fuzzy- rough set similarity is implemented for every feature and then select the $n-$top attributes.

c. `Classifying` In this study, Artificial Neural Network (ANN) and Naive Bayes (NB) are utilized as classifiers to determine the performance of attribute selection technique. Artificial Neural Network (ANN) and Naive Bayes (NB) are employed as classifiers to determine the performance of attribute selection techniques in this work. Naive Bayes is an uncomplicated learning algorithm that makes use of Bayes principle each other by a powerful supposition which the criteria are circumstantially individualistic, dedicated the class. Meanwhile this independence supposition is frequently disturbed, Naive Bayes against all odds constantly gives passionate classification correctness. Unified by its algorithmic performance and plenty alternative favourable features, it guides to Naive Bayes becoming extensively used virtually [26].

An ANN is adaptive in disposition because it changes its anatomy and modifies its weight to keep down the error. An adaptation of weight is based on the knowledge that moves internally and externally by network within the learning period. The vantages of ANN are it needs less formal statistical training, indirectly uncover complex nonlinear connections betwixt dependent and independent variables, figure out all probable interplays among predictor variables, and the presence of various training algorithms [1].

d. `Measurement Metrics` As mentioned in this paper, classification and accuracy, precision, recall, $F-$measure, and computational time metrics have been employed to analyze these techniques. The classifier performance is evaluated using training set validation. The average accuracy for the datasets is determined as the measure of correctly classified test samples. Then, the performances of every technique are evaluated with each other. The measurement metrics definitions are given below:

d-1. `Accuracy` This evaluation parameter is used to determine how close the measurements of a value are to the true value.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP}.$$

Here, TP is a true positive indicator that is accurately identified, TN represents a true negative that has been properly rejected, FP false positive that is misidentified, and likewise, FN represents a false negative that has been wrongly rejected.

d-2. `Precision` This is determined by the proximity of two or more measurements to each other. Precision is also expressed as a positive predictive measure.

$$\text{Precision}\,(p) = \frac{TP}{TP+FP}.$$

d-3. `Recall` is also noted as the actual positive proportion or sensibility that is retrieved to measure a division of the relevant samples.

$$Recall\,(r) = \frac{TP}{TP+FN}.$$

d-4. `Recall-Precision` metric is a useful measure of success of prediction when the classes are very imbalanced. A large domain below the graph displays both great recall and big precision, that high precision shows fallen false-positive rate, and high recall relates to a fallen false-negative rate.

d-5. `F- Measure` This is an evaluation of test carefulness. It considers both p and r in the test to account for the measure.

$$F - Measure = \frac{2pr}{p+r}.$$

d-6. `Computing Period` The interval indispensable to accomplish computational progress by assessing the classification implementation time [11].

Sum up Table 1 shows the Artificial Neural Network classification performance for Parkinson detection before selection attribute and after that. As can be seen from the table, a relative increase in performance is visible, and this means that the reduction of dimensions in the dataset while raising performance has caused a significant reduction in the calculated time.

Table 1. ANN Performances of Classification for Parkinson

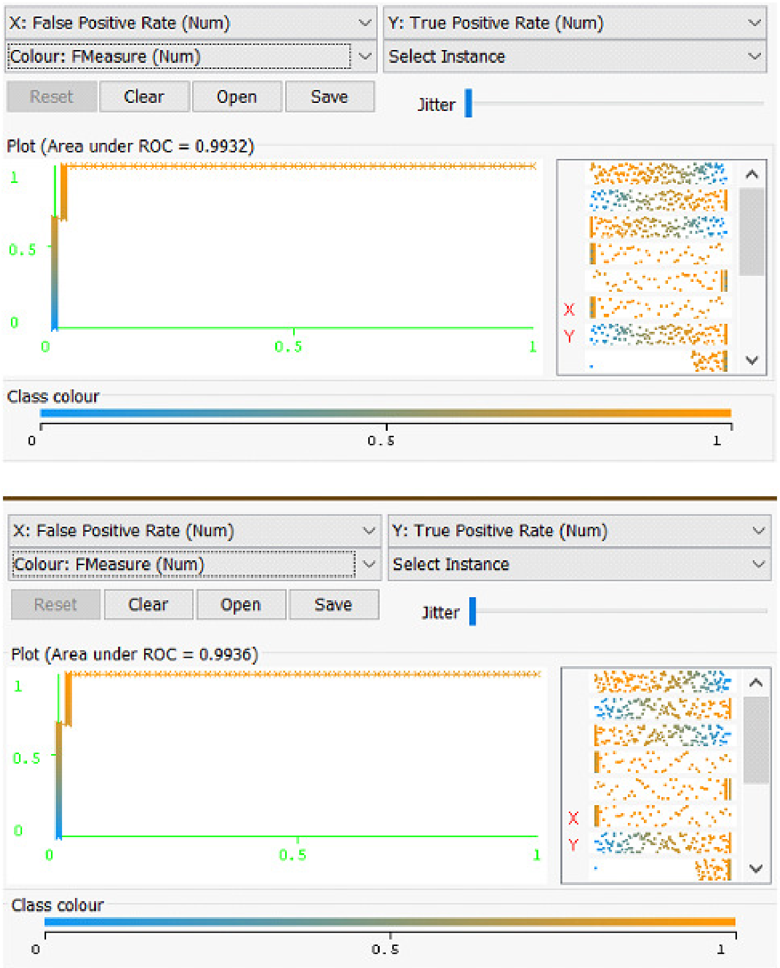
	ANN-Before Changing		ANN-After Changing	
TP Rate	1	0.979	1	0.979
FP Rate	0.021	0	0.021	0
Precision	0.993	1	0.993	1
Recall	1	0.979	1	0.979
F- Measure	0.997	0.989	0.997	0.989
ROC Area	0.993	0.993	0.994	0.994
Calculation- Time(S)	0.39	0.39	0.01	0.01
Class	Diagnosis	Not- Diagnosis	Diagnosis	Not- Diagnosis

In Table 2, you can find the Naive Bayes classification performance for Parkinson detection before selection attribute and after that. As table 2 demonstrates while there is a small change in the time of calculating, the performance of the algorithm has not been changed. So, the dimensionality reduction had not been affected by Naive Bayes' act. Hence, it can be said that Naive Bayes performed better than the Artificial Neural Network technique.

Table 2. NB Performances of Classification for Parkinson

	NB- Before Changing		NB-After Changing	
TP Rate	0.626	0.938	0.701	0.917
FP Rate	0.063	0.374	0.083	0.299
Precision	0.968	0.45	0.963	0.5
Recall	0.626	0.938	0.701	0.917
F- Measure	0.76	0.608	0.811	0.647
ROC Area	0.871	0.878	0.881	0.881
Calculation- Time(S)	0.019	0.019	0.018	0.018
Class	Diagnosis	Not- Diagnosis	Diagnosis	Not- Diagnosis

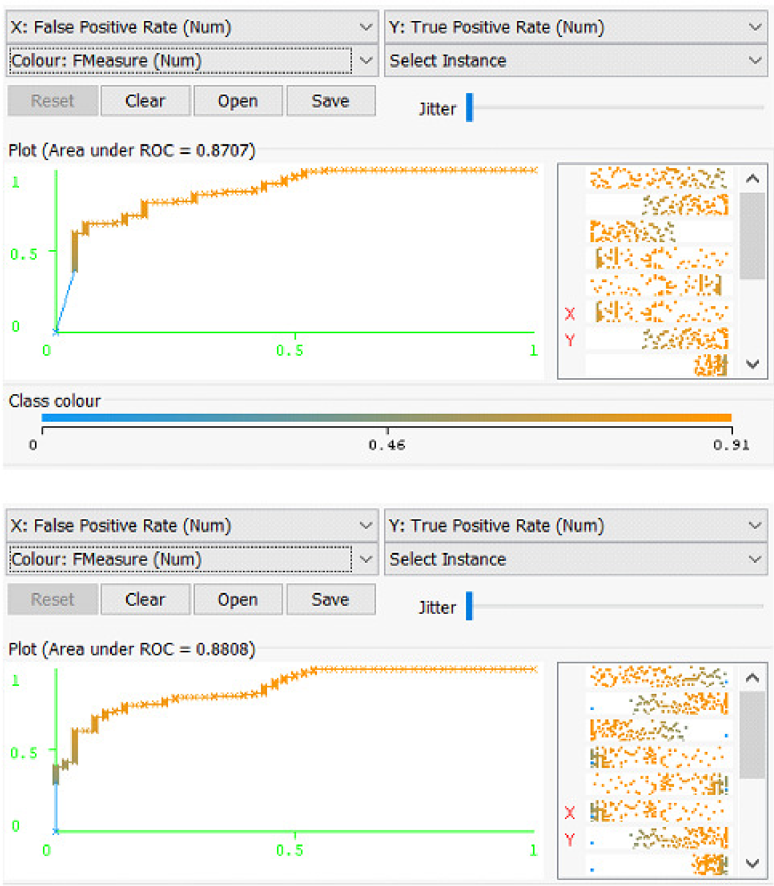
The ROC curve demonstrates the equilibration between susceptibility (or TPR) and specificity (FPR). Those curves are nearby to the top-left corner display better performance. A random classifier is awaited to have points over the diagonal (FPR = TPR). Consider that the ROC independent of the class distribution. This makes it fruitful for evaluating classifiers anticipating scarce cases such as diseases or disasters. Conversely, interpreting performance using accuracy would have preferred classifiers that regularly predict a negative result for infrequent cases.



(a) ANN Before

(b) ANN After

FIGURE 1. ANN ROC area before and after attribute selection



(a) NB Before

(b) NB After

FIGURE 2. The NB ROC- Area for Parkinson disease

The area under the curve (AUC), To compare distinct classifiers, it can be gainful to outline the performance of each classifier into an individual evaluation. One communal approach is to figure out the area under the ROC curve, which is called AUC. Although a classifier with a high AUC can sometimes give a worse point in a specific area than another classifier with a lower AUC, the AUC acts well as a general scale of predictive accuracy. As it can be seen Figure 1 and Figure 2 illustrate the changes after the dimension reduction. As it seems from Figures 1 and 2 while in the ANN algorithm a slight improvement in the AUC area is visible, the AUC area NB has become better, and this means reducing the dimensions in the Naive Bayes method has had a good effect. On the other hand, the ROC curve of ANN totally has been acted better than the NB ROC curve. But here is the significant thing that should be noticed is changing the ROC and AUC before and after feature reduction. Therefore, instead of comparing Figures 1 and 2 with each other, the changes in Figures 1-a to 1-b should be considered and respectively for Figures 2-a to 2-b.

4. CONCLUSION

In this study, Fuzzy- Rough Subset Evaluation algorithm has been used for feature selection that can select a small set of datasets to prepare a highly precise classification of the instances. The proposed dataset normalized by fuzzy Gaussian membership function. The F-Score and Fuzzy-Rough Subset Evaluation are exploited on the normalized dataset to rank the objects. F-score is utilized to recognize relevant attributes and FRSE is applied to remove the redundancy among the features. In all feature selection results in Parkinson's data, the data selected may or may not be a subset of disease progression signature. So, the top n attributes are selected for classification in ANN and NB. The train set validation is used to detect the average classification accuracy. It provides 0.04% average classification accuracy for the ANN method and a 1.01% percentage for NB classifiers. It also gives the highest average AUC accuracy for NB compared to the ANN algorithms. In summary, the normalization of a dataset by using the Fuzzy Gaussian membership function can modify the classification accuracy with the suggested measure. The performance of the Naive Bayes is totally better than the Artificial Neural Network algorithm. So, the proposed method is effective and consistent for the detection of Parkinson with a small number of features.

REFERENCES

[1] S. Agatonovic-Kustrin and R. Beresford, "Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research," *Journal of pharmaceutical and biomedical analysis*, vol. 22, no. 5, pp. 717–727, 2000, doi: 10.1016/S0731-7085(99)00272-1.

[2] R. Babuška and H. Verbruggen, "Constructing fuzzy models by product space clustering," in *Fuzzy model identification*. Springer, 1997, pp. 53–90, doi: 10.1007/978-3-642-60767-7-2.

[3] C. Bianca, F. Pappalardo, M. Pennisi, and M. Ragusa, "Persistence analysis in a kolmogorov-type model for cancer-immune system competition," in *AIP Conference Proceedings*, vol. 1558, no. 1, doi: 10.1063/1.4825874. American Institute of Physics, 2013, pp. 1797–1800.

[4] C. Bianca, M. Pennisi, S. Motta, and M. A. Ragusa, "Immune system network and cancer vaccine," in *AIP Conference Proceedings*, vol. 1389, no. 1, doi: 10.1063/1.3637764. American Institute of Physics, 2011, pp. 945–948.

[5] D. S. Char, N. H. Shah, and D. Magnus, "Implementing machine learning in health care-addressing ethical challenges," *The New England journal of medicine*, vol. 378, no. 11, p. 981, 2018, doi: 10.1056/NEJMp1714229.

[6] S. L. Cichosz, M. D. Johansen, and O. Hejlesen, "Toward big data analytics: review of predictive models in management of diabetes and its complications," *Journal of diabetes science and technology*, vol. 10, no. 1, pp. 27–34, 2016, doi: 10.1177/1932296815611680.

[7] Q. Feng and R. Li, "Discernibility matrix based attribute reduction in intuitionistic fuzzy decision systems," in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*," doi: 10.1007/978-3-642-41218-9-16. Springer, 2013, pp. 147–156.

[8] D. Heisters, "Parkinson's: symptoms, treatments and research," *British Journal of Nursing*, vol. 20, no. 9, pp. 548–554, 2011, doi: 10.12968/bjon.2011.20.9.548.

[9] T. K. Ho and M. Basu, "Complexity measures of supervised classification problems," *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 3, pp. 289–300, 2002, doi: 10.1109/34.990132.

[10] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Transactions on fuzzy systems*, vol. 17, no. 4, pp. 824–838, 2008, doi: 10.1109/TFUZZ.2008.924209.

[11] N. M. Lanbaran and E. Çelik, "Prediction of breast cancer through tolerance-based intuitionistic fuzzy-rough set feature selection and artificial neural network," *Gazi University Journal of Science*, pp. 1–1, doi: 10.35378/gujs.857099.

[12] J. Li, H. Zhao, and W. Zhu, "Fast randomized algorithm with restart strategy for minimal test cost feature selection," *International Journal of Machine Learning and Cybernetics*, vol. 6, no. 3, pp. 435–442, 2015, doi: 10.1007/s13042-014-0262-0.

[13] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K.-W. Low, S.-F. Newman, J. Kim *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature biomedical engineering*, vol. 2, no. 10, pp. 749–760, 2018, doi: 10.1038/s41551-018-0304-0.

[14] J.-M. Ma, Y. Leung, and W.-X. Zhang, "Attribute reductions in object-oriented concept lattices," *International Journal of Machine Learning and Cybernetics*, vol. 5, no. 5, pp. 789–813, 2014, doi: 10.1007/s13042-013-0214-0.

[15] R. R. Mendes, F. B. de Voznika, A. A. Freitas, and J. C. Nievola, "Discovering fuzzy classification rules with genetic programming and co-evolution," in *European Conference on Principles of Data Mining and Knowledge Discovery*," doi: 10.1007/3-540-44794-6-26. Springer, 2001, pp. 314–325.

[16] S. Narula, K. Shameer, A. M. Salem Omar, J. T. Dudley, and P. P. Sengupta, "Reply: deep learning with unsupervised feature in echocardiographic imaging," *Journal of the American College of Cardiology*, vol. 69, no. 16, pp. 2101–2102, 2017, doi: 10.1016/j.jacc.2017.01.062.

[17] Z. Pawlak, *Rough sets: Theoretical aspects of reasoning about data*. Springer Science & Business Media, 2012, vol. 9, doi: 10.1007/978-94-011-3534-4.

[18] M. A. Ragusa and G. Russo, "Odes approaches in modeling fibrosis: Comment on" towards a unified approach in the modeling of fibrosis: A review with research perspectives" by martine ben amar and carlo bianca." *Physics of life reviews*, vol. 17, pp. 112–113, 2016, doi: j.plrev.2016.05.012.

[19] L. C. Shukla, J. Schulze, J. Farlow, N. D. Pankratz, J. Wojcieszek, and T. Foroud, "Parkinson disease overview," *GeneReviews®[Internet]*, 2019.

[20] A. Skowron and J. Stepaniuk, "Tolerance approximation spaces," *Fundamenta Informaticae*, vol. 27, no. 2, 3, pp. 245–253, 1996, doi: 10.3233/FI-1996-272311.

[21] T. Terano, K. Asai, and M. Sugeno, *Applied fuzzy systems*. Academic Press, 2014.

[22] A. K. Tiwari, S. Shreevastava, T. Som, and K. K. Shukla, "Tolerance-based intuitionistic fuzzy-rough set approach for attribute reduction," *Expert Systems with Applications*, vol. 101, pp. 205–212, 2018, doi: 10.1016/j.eswa.2018.02.009.

[23] B. X. Tran, C. A. Latkin, G. T. Vu, H. L. T. Nguyen, S. Nghiem, M.-X. Tan, Z.-K. Lim, C. S. Ho, and R. Ho, "The current research landscape of the application of artificial intelligence in managing cerebrovascular and heart diseases: a bibliometric and content analysis," *International journal of environmental research and public health*, vol. 16, no. 15, p. 2699, 2019, doi: 10.3390/ijerph16152699.

[24] E. C. Tsang, Q. Hu, and D. Chen, "Feature and instance reduction for pnn classifiers based on fuzzy rough sets," *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 1, pp. 1–11, 2016, doi: 10.1007/s13042-014-0232-6.
[25] C. Wang, Q. He, D. Chen, and Q. Hu, "A novel method for attribute reduction of covering decision systems," *Information sciences*, vol. 254, pp. 181–196, 2014, doi: 10.1016/j.ins.2013.08.057.
[26] G. I. Webb, "Naïve bayes." *Encyclopedia of machine learning*, vol. 15, pp. 713–714, 2010, doi: 10.1007/978-1-4899-7502-7-581-1.
[27] L. A. Zadeh, "Fuzzy sets," in *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*. World Scientific, 1996, pp. 394–432, doi: /10.1142/2895.

Authors' addresses

Naiyer Mohammadi Lanbaran
Atatürk University, Faculty of Science, Department of Mathematics, Erzurum, Turkey
E-mail address: `naiyer.mohammadi.lanbaran13@ogr.atauni.edu.tr`

Ercan Çelik
(corresponding) Atatürk University, Faculty of Science, Department of Mathematics, Erzurum, Turkey; Department of Applied Mathematics and Informatics, Kyrgyz-Turkish Manas University, Bishkek/ Kyrgyzstan
E-mail address: `ercelik@atauni.edu.tr; ercan.celik@manas.edu.kg`

Özgür Kotan
Atatürk University, Faculty of Science, Department of Mathematics, Erzurum, Turkey
E-mail address: `Ozgurkotan2525@gmail.com`