Research Article GU J Sci 34(4): 1064-1075 (2021) DOI: 10.35378/gujs.857099

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

Prediction of Breast Cancer through Tolerance-based Intuitionistic Fuzzy-rough Set Feature Selection and Artificial Neural Network

Naiyer Mohammadi LANBARAN *, Ercan CELIK

Ataturk University, Faculty of Science, 25240 Erzurum, Turkey

Highlights

- This paper focuses on feature selection process by tolerance-based intuitionistic fuzzy-rough set.
- Two hybrid approaches are proposed for classification in the study.
- A highly precise and more efficient classification accuracy were obtained.

Article Info

Received: 09 Jan 2021
Accepted: 16 Feb 2021

Keywords

Intuitionistic
Fuzzy-Rough set
Feature- selection
Breast cancer

Abstract

The importance of diagnosing breast cancer is one of the most significant issues in medical science. Diagnosing whether the cancer is benign or malignant is extremely essential in ascertaining the type of cure, moreover, to bringing down bills. This study aims to use the tolerance-based intuitionistic fuzzy-rough set approach to pick attributes and data processing with help of machine learning for the classification of breast cancer. The main purpose of selecting a feature is to make a subset of input variables by removing irrelevant variables or variables that lack predictive information. This study shows how to eliminate redundant data in big data and achieve more efficient results. Rough set theory has already been used successfully to set down attributes, but this theory is insufficient to reduce the properties of a real- value dataset because it will possibly drop knowledge through the decomposition procedure. and this prevents us from getting the right results. In this study, we used the tolerance based intuitive fuzzy rough method for attribute selection. In this technique, lower and upper approaches are used to intuitive fuzzy sets from rough sets to remove uncertainty due to having simultaneous membership, non-membership, and hesitation degrees and obtain better results. The used method is demonstrated to be better performing in the shape of chosen attributes.

1. INTRODUCTION

Feature selection is comprehensively employed in a variety of data processing implementations in particular textuality mining, genetic investigation, and data mining. Attribute Selection is very useful for applications where the main features are important for the model interpretation and knowledge extraction because the main features of the data set are preserved during this process. Feature selection can be defined as the process of identifying related features and removing unrelated and duplicate features to view a subset of features that describe the problem well and with minimal downtime. There are several benefits to this, some of which are outlined below [1]:

- Improving the performance of machine learning algorithms,
- Understand data, gain knowledge about the process, and help visualize it,
- Reduce overall data, limit storage requirements, and possibly help reduce costs,
- Reduce feature sets, store resources in the next round of data collection or during operation,

*Corresponding author, e-mail: naiyer.mohammadi.lanbaran13@ogr.atauni.edu.tr

- Simplicity and the ability to use simpler models and gain speed.

For all these reasons, feature selection plays a key role in "big data analysis" scenarios. Dimension reduction methods are often split into two categories: feature selection and feature extraction, and each of them have its own special properties. From one side, feature extraction methods reduce dimensions by combining key features. Hence, they can build a set of new features that are usually more compact and more distinctive. These methods are preferred in applications such as image analysis, image processing, and information retrieval because in these cases the accuracy of the model is more important than its interpretability. On the other hand, feature selection reduces dimensions by eliminating unrelated and duplicate features [2-4]. In recent years, Rough set theory has become one of the most powerful solutions to artificial intelligence problems such as data mining. One of the great strategies is called basic used to extract knowledge and discover hidden patterns of information systems, which is the same as feature selection. Data mining feature reduction is one of the most important topics in rough set theory, but the classical version of rough set theory is not very suitable for discussing feature loss in incomplete information systems [5-7].

A method that can diminish the dimensional and keep the meaning of the attributes by using the information included in the data set is plainly desirable. Rough Set Theory (RST) [8] can be used as a mechanism to detect data affiliations and reduce the number of attributes in a data set using data alone. Nevertheless, traditional RST methods are commonly infirm of handling real-valued data straightly. Already, discretization techniques were exerted beforehand to convert the data into distinct values, but this may end in information loss. As a result of this, various extensions to the original theory have been suggested. Two critical developments in this field have been fuzzy-rough sets [8] and similarity or tolerance-based rough set theory [9,10]. Hence, it is favorable to develop methods to enable the tools of data reduction for crisp and real-value attributed data sets that employ this additional knowledge. A new view of the fuzzy rough set to deal with ambiguity and incompatibility in the data set has been come up [11]. To handle the shortcomings of the rough classical set theory, the rough fuzzy set approach has been implemented from different sides. Despite the undeniable obstacles, fuzzy set theory is still a strong instrument to handle the ambiguity. Many ideas have been proposed to deal with ambiguous and complicated knowledge structures. Among the most popular derivatives of fuzzy sets are vague and intuitive fuzzy sets in which the interval-valued membership function is employed. Membership functions defined based on an interval are more efficient in resolving the ambiguity of the data set. [12,13]. The intuitive fuzzy set has better performance due to the concurrent use of membership, non-membership, and hesitation functions to tackle ambitiousness. Therefore, in this method, there is optimal control over uncertainty compared to fuzzy perspectives. So, it capable of tackle with data structures that give a better sight of the unstable real-world ambiguities. In some cases, the degree of non-membership is more practical than the degree of membership. In such cases, intuitive fuzzy set theory is a more effective choice to simulate human decision-making which is often associated with erroneous and unreliable results. The success of the concept of intuitive fuzzy sets has been repeatedly seen on decision-making and pattern identification issues.

Protecting the exact degree of uncertainty in intuitive fuzzy sets is an undeniable advantage over fuzzy sets. Different from fuzzy sets, intuitionistic fuzzy sets protect an accurate degree of uncertainty[14-19]. In this thesis, the tolerance degree of similarity between two attributes according to upper and lower estimations in intuitive fuzzy set theory is used for feature selection. Then we apply artificial neural networks and K- nearest neighbor algorithms for analysis of our result for prediction of Breast Cancer.

2. MATERIAL METHOD

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. İn a classical set theory a component 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 and x is an object. and $A = \{\langle x, \mu_X(x)\rangle | x \in U\}$ is a fuzzy set.

Definition 2. 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 : $\underline{R}(x) = \cup\{R(x): R(x) \subseteq \mathrm{X}\}$

$R - upper$ approximation of X: $\bar{R}(x) = \cup\{R(x): R(x) \cap \mathrm{X} \neq \emptyset\}$

$R - boundary$ approximation of X: $RnR(x) = \bar{R}(x) - \underline{R}(x)$

The rough set membership function is described as:

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

Definition 3. $A \subseteq \mathrm{U}$ is a set of attributes that $A = \{\langle x, n_A(x), m_A(x)\rangle | x \in U\}$ here $n_A: U \rightarrow [0,1]$ and $m_A: U \rightarrow [0,1]$ that $0 \leq n_A(x) + m_A(x) \leq 1, \forall x \in U$ and $n_A(x)$ and $m_A(x)$ are membership and non-membership function and A is named intuitionistic fuzzy set. And $\forall x \in U$: $\pi_A(x) = 1 - n_A(x) - m_A(x)$ demonstrates the grade of the hesitancy of *x in A*, and $0 \leq \pi_A(x) \leq 1$. The arranged pair $\langle n_A, m_A\rangle$ is the value of an intuitionistic fuzzy set.

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 $F(x, r) = \mu_r(x), \forall x \in U \; and \; r \in R$ and $S: U \times R \rightarrow K_S$ and $\mu_r(x)$ is membership degree.

Definition 5. A quaternary (U, R, K_{IS}, IS) gives a fuzzy information system (IFIS). K_{IS} is the set of all intuitive fuzzy numbers and IS is an information function which $IF(x, r) = \langle n_r(x), m_r(x)\rangle, \forall x \in U \; and \; r \in R$ and $iS: U \times R \rightarrow K_{IS}$.

The information system of real issues is on an immense scale. so, in this paper, we use a technique for reducing dimensionality supported by an intuitionistic fuzzy rough set to attribute selection.

Here in this method, by considering one or a subset of several features, we determine the similarity between the two features [20].

Definition 6. Similarity relationship:

$$S_r(x_i, x_j) = 1 - \sqrt{\theta\left(\mu_r(x_i) - \mu_r(x_j)\right)^2 + \lambda\left(\omega_r(x_i) - \omega_r(x_j)\right)^2 + \varphi(\pi_r(x_i) - \pi_r(x_j))^2} \tag{1}$$

where, $\mu_r(x_i)$, $\omega_r(x_i)$ and $\pi_r(x_i)$ are membership, non-membership and hesitancy functions of an item x_i according to feature r respectively, and $\theta, \lambda \; and \; \varphi$ show weighted operators. These factors are determined according to the user needs and pre-determined conditions below

a. $\theta \geq \lambda > \varphi$,
b. $\theta + \lambda + \varphi = 1$,
c. $0 \leq \theta, \lambda, \varphi \leq 1$.

Q is a subcategory of features, the resemblance relationship between two items is described as follows:

$$(x_i, x_j) \in sim_Q^\sigma \; iff \; \prod_{r \in T} sim_r(x_i, x_j) \geq \gamma. \tag{2}$$

Here σ is a resemblance threshold. Tolerance rate of an item x_i:

$$sim_Q^\sigma(x_i) = \{x_j \in U | (x_i, x_j) \in sim_Q^\sigma\}. \tag{3}$$

Definition 7. To determine the positive area and dependency function we need lower and upper approximations of $X \subseteq U$ are [21, 22]:

$$\underline{apx}P^{\sigma}X = \{x_i | sim_Q^{\sigma} \subseteq X\} \tag{4}$$

$$\overline{apx}P^{\sigma}X = \{x_i | sim_Q^{\sigma} \cap X \neq \emptyset\}. \tag{5}$$

The ($\underline{apx}P^{\sigma}X$, $\overline{apx}P^{\sigma}X$) ordered pair is an intuitive fuzzy tolerance rough set. Let T be a set of properties that produce the equivalence ratio to U. Hence, the positive area is as:

$$Pos_Q^{\sigma}(T) = \cup_{X \in U/T} \underline{apx}P^{\sigma}X. \tag{6}$$

And dependency function shown as:

$$\Psi_Q^{\sigma}(Q) = \frac{|Pos_Q^{\sigma}(T)|}{|U|}. \tag{7}$$

3. THE RESEARCH FINDINGS AND DISCUSSION

This breast cancer database was obtained from the University of Wisconsin Hospitals, Madison from Dr. William H. Wolberg which available in https://archive.ics.uci.edu/ml/machine-learning-databases/breast-cancer-wisconsin/. Features are computed from a digitized image of a fine needle aspirate (FNA) of a breast mass.

They describe the characteristics of the cell nuclei present in the image. A few of the images can be found at http://www.cs.wisc.edu/~street/images/. Python 3.7 software has been used to evolve the classification of data circulation. This dataset comprises 569 samples; 357 of them correspond to class one (Benign) and 212 samples belong to class two (Malignant). Before making anything like feature selection, feature extraction, and classification, we start with data normalization. To normalize, we first fuzzify the data. Hereby, we use a Gaussian fuzzifier. Gaussian fuzzy membership functions are fully well- known in the fuzzy logic literature, as they are the foundation for the connectivity between fuzzy systems and Radial Basis Function (RBF) neural networks. The Gaussian membership function performed by using the formula given below:

$$\mu_i(x) = exp^{\left(-\frac{(C_i - x)^2}{2\sigma_i^2}\right)} \tag{8}$$

where C_i and σ_i are mean and the standard deviation of $i.th$ attribute in a dataset.

Now we have a fuzzy decision system (FDS) that given in Table 1. (For convenience, only the five rows of each table are displayed).

Table 1. *Fuzzy Decision System (FDS)*

Row	0	1	2	3	4
Radius- Mean	0.547839	0.187473	0.287072	0.954271	0.531737
Texture- Mean	0.116559	0.939387	0.901177	0.999989	0.590671
Perimeter- Mean	0.446478	0.241419	0.293181	0.928388	0.528471
Area- Mean	0.616007	0.161769	0.296692	0.936184	0.62139
Smoothness- Mean	0.29228	0.710395	0.641542	0.533136	0.479387
Compactness- Mean	0.004559	0.888146	0.574459	0.452497	0.903814
Concavity- Mean	0.029633	0.999716	0.394737	0.542362	0.619401
Concave Points- Mean	0.040489	0.860509	0.125537	0.513633	0.649474
Symmetry- Mean	0.085547	0.999999	0.643069	0.999637	0.003084
Fractal Dimension- Mean	0.078536	0.685725	0.92385	0.543688	0.628169
Radius- Se	0.045077	0.882826	0.470094	0.999991	0.923382
Texture- Se	0.852346	0.681199	0.737665	0.999987	0.920633

Perimeter- Se	0.018078	0.965924	0.696255	0.991038	0.975928
Area- Se	0.045319	0.759131	0.49769	0.986914	0.99212
Smoothness- Se	0.977362	0.832581	0.956853	0.988811	0.987645
Compactness- Se	0.420184	0.78657	0.717422	0.64064	0.720481
Concavity- Se	0.769429	0.907426	0.977555	0.58558	0.7242
Concave Points- Se	0.803851	0.966724	0.362378	0.464995	0.842991
Symmetry- Se	0.516944	0.722979	0.972298	0.986579	0.036091
Fractal Dimension- Se	0.662722	0.995068	0.957827	0.954711	0.716645
Radius- Worst	0.168672	0.195795	0.318901	0.914699	0.553883
Texture- Worst	0.396992	0.934115	0.999713	0.855783	0.4076
Perimeter- Worst	0.070419	0.3078	0.403393	0.879396	0.537661
Area- Worst	0.135001	0.167466	0.346323	0.903351	0.666964
Smoothness- Worst	0.425274	0.931888	0.870159	0.414672	0.977483
Compactness- Worst	0.032599	0.911521	0.556343	0.473002	0.612677
Concavity- Worst	0.108062	0.98929	0.693856	0.431246	0.485689
concave Points- Worst	0.071649	0.553842	0.14793	0.288886	0.400737
Symmetry- Worst	0.022755	0.970697	0.514867	0.756526	0.569179
Fractal Dimension- Worst	0.1532	0.961237	0.979925	0.506242	0.978686
Diagnose	1	1	1	0	0

Then, we convert the FDS to an intuitionistic fuzzy decision system (IFDS) by the equation given in Definition 3. Hesitancy degree is considered constant as $\pi = 0.2$ [23] (IFDS system is given in Table 2).

Table 2. *Intuitionistic Fuzzy Decision System (IFDS)*

Row	0	1	2	3	4
Radius- Mean	0.252161	0.612527	0.512928	-0.15427	0.268263
Texture- Mean	0.683441	-0.13939	-0.10118	-0.19999	0.209329
Perimeter- Mean	0.353522	0.558581	0.506819	-0.12839	0.271529
Area- Mean	0.183993	0.638231	0.503308	-0.13618	0.17861
Smoothness- Mean	0.50772	0.089605	0.158458	0.266864	0.320613
Compactness- Mean	0.795441	-0.08815	0.225541	0.347503	-0.10381
Concavity- Mean	0.770367	-0.19972	0.405263	0.257638	0.180599
Concave Points- Mean	0.759511	-0.06051	0.674463	0.286367	0.150526
Symmetry- Mean	0.714453	-0.2	0.156931	-0.19964	0.796916
Fractal Dimension- Mean	0.721464	0.114275	-0.12385	0.256312	0.171831
Radius- Se	0.754923	-0.08283	0.329906	-0.19999	-0.12338
Texture- Se	-0.05235	0.118801	0.062335	-0.19999	-0.12063
Perimeter- Se	0.781922	-0.16592	0.103745	-0.19104	-0.17593
Area- Se	0.754681	0.040869	0.30231	-0.18691	-0.19212
Smoothness- Se	-0.17736	-0.03258	-0.15685	-0.18881	-0.18765
Compactness- Se	0.379816	0.01343	0.082578	0.15936	0.079519
Concavity- Se	0.030571	-0.10743	-0.17755	0.21442	0.0758
Concave Points- Se	-0.00385	-0.16672	0.437622	0.335005	-0.04299
Symmetry- Se	0.283056	0.077021	-0.1723	-0.18658	0.763909
Fractal Dimension- Se	0.137278	-0.19507	-0.15783	-0.15471	0.083355

Radius- Worst	0.631328	0.604205	0.481099	-0.1147	0.246117
Texture- Worst	0.403008	-0.13412	-0.19971	-0.05578	0.3924
Perimeter- Worst	0.729581	0.4922	0.396607	-0.0794	0.262339
Area- Worst	0.664999	0.632534	0.453677	-0.10335	0.133036
Smoothness- Worst	0.374726	-0.13189	-0.07016	0.385328	-0.17748
Compactness- Worst	0.767401	-0.11152	0.243657	0.326998	0.187323
Concavity- Worst	0.691938	-0.18929	0.106144	0.368754	0.314311
concave Points- Worst	0.728351	0.246158	0.65207	0.511114	0.399263
Symmetry- Worst	0.777245	-0.1707	0.285133	0.043474	0.230821
Fractal Dimension- Worst	0.6468	-0.16124	-0.17993	0.293758	-0.17869
Diagnose	1	1	1	0	0

The next step is to find a similarity between every two items and calculate the dependency degree for each attribute. Then obtain a subset of selected features for use in machine learning methods (Table 3).

Here we set $\gamma = 0.8$ and $\theta = 0.4, \lambda = 0.4, and\ \phi = 0.2$.

Table 3. *Selected Features*

Row	0	1	2	3	4
Texture- Mean	0.116559	0.939387	0.901177	0.999989	0.590671
Area Mean	0.616007	0.161769	0.296692	0.936184	0.62139
Smoothness- Mean	0.29228	0.710395	0.641542	0.533136	0.479387
Concavity- Mean	0.029633	0.999716	0.394737	0.542362	0.619401
Symmetry- Mean	0.085547	0.999999	0.643069	0.999637	0.003084
Fractal Dimension- Mean	0.078536	0.685725	0.92385	0.543688	0.628169
Texture- Se	0.852346	0.681199	0.737665	0.999987	0.920633
Area- Se	0.045319	0.759131	0.49769	0.986914	0.99212
Smoothness- Se	0.977362	0.832581	0.956853	0.988811	0.987645
Concavity- Se	0.769429	0.907426	0.977555	0.58558	0.7242
Symmetry- Se	0.516944	0.722979	0.972298	0.986579	0.036091
Fractal Dimension- Se	0.662722	0.995068	0.957827	0.954711	0.716645
Smoothness- Worst	0.425274	0.931888	0.870159	0.414672	0.977483
Concavity- Worst	0.108062	0.98929	0.693856	0.431246	0.485689
Symmetry Worst	0.022755	0.970697	0.514867	0.756526	0.569179
Fractal Dimension- Worst	0.1532	0.961237	0.979925	0.506242	0.978686
Diagnose	1	1	1	0	0

The results are calculated and discussed below.

3.1. Measurement Metrics

In this paper, we use Artificial Neural Network (ANN) and k-Nearest Neighbors (KNN) methods for classification and accuracy, precision, recall, F-measure, and computational time metrics have been employed to analyze these techniques.

Then, the performances of every technique are evaluated with each other. The measurement metrics are demonstrated in Figure 1 and definitions are given below:

Accuracy: This evaluation parameter is used to determine how close the measurements of a value are to the true value

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}\,. \tag{9}$$

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.

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

$$Precision(p) \ = \frac{TP}{TP+FP}\,. \tag{10}$$

Recall: is also noted as the actual positive proportion or sensibility that is retrieved to measure a division of the relevant samples

$$\text{Recall(r)} \ = \frac{TP}{TP+FN}\,. \tag{11}$$

Recall- Precision metric is a useful measure of success of prediction when the classes are very imbalanced. A high area under the curve represents both high recall and high precision, where high precision relates to a low false-positive rate, and high recall relates to a low false-negative rate.

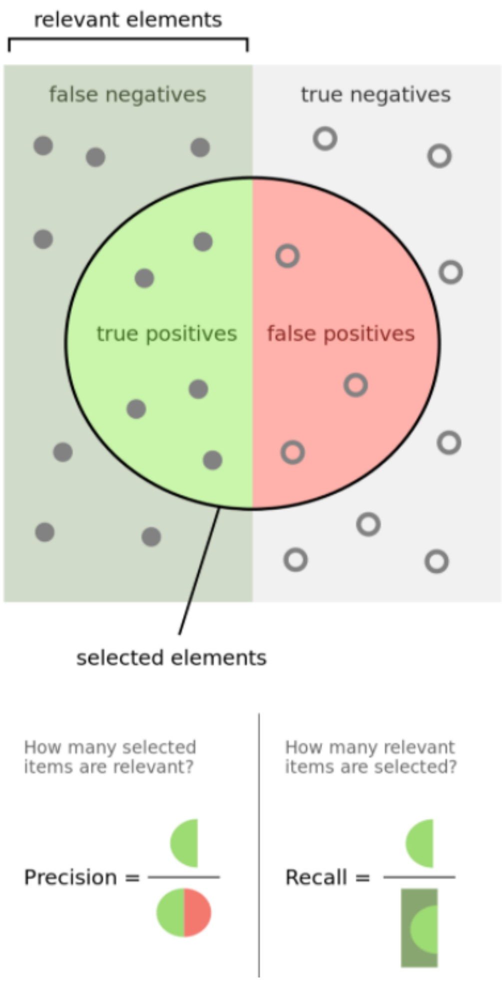

Figure 1. *Recall- Precision metric*

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

$$\mathrm{F-\ Measure} = \frac{2pr}{p+r} \tag{12}$$

Computing Period: The interval indispensable to accomplish computational progress by assessing the classification implementation time.

3.2. Accurate Metric Analysis

The efficiency analysis according to accuracy is given in Figure 2. The accurate metric is explained as the scale of being true or exact. Figure 2a shows the accuracy performance of two machine learning methods for both databases of breast cancer, I mean with the original dataset and subset of selected features.

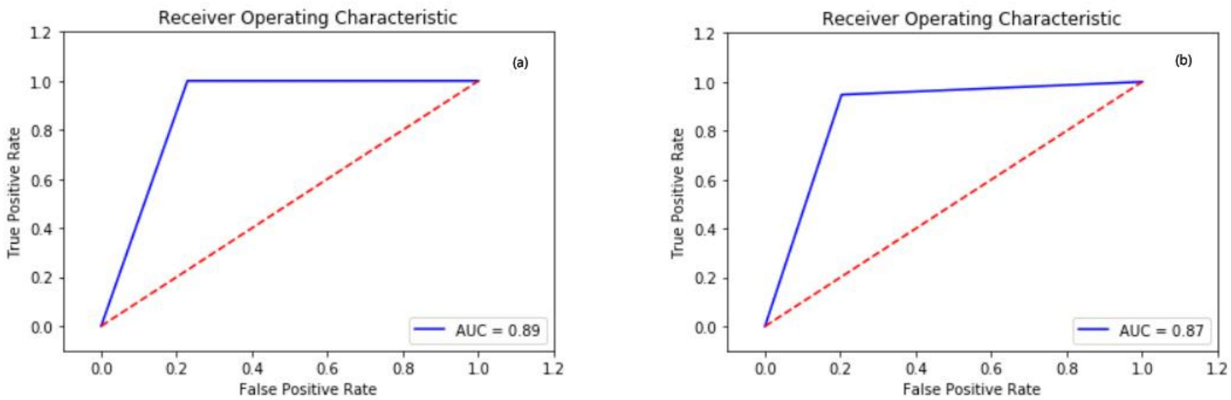

Figure 2. *Artificial Neural Network Accuracy Analysis; a) New Dataset b) Original Dataset*

As can be seen from the Figure2. the accuracy of the ANN methods in both datasets is compared and the accuracy changes of Artificial neural networks in both before and after feature selection. As shown in Figure 2. the accuracy of the model after applying its algorithm to the dataset, has improved by 2% compared to the previous situation. And this shows that model ANN shows a better situation in the classification.

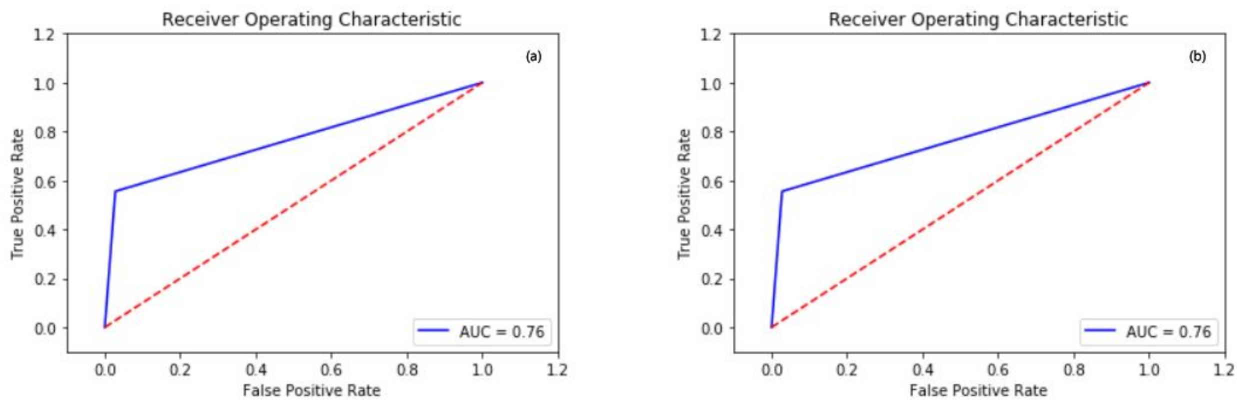

Figure 3. *K- Nearest Neighbor Accuracy Analysis; a) New Dataset b) Original Dataset*

Here in Figure 3. you can see the correctness has not changed in K- Nearest Neighbors for both before and after feature selection. Therefore, it can be clearly said that this reduction in dimensions does not affect improving the performance of this model.

3.3. Analysis Using Recall and Precision

In this case, the recall performance of the breast cancer database with dimensionality reduction is 0.95 for ANN classifier and with the normal dataset is 0.925. Precision analysis for ANN method is about 0.587 percent for the normal dataset while it reached 0.6032. For the K-NN model, before applying reduction on the dataset we got Recall about 0.556, and after selecting a feature and exert dimensionality reduction on the dataset it's become 0.460 while the precision for both states is 0.92 and 0.91.

3.4. Analysis Using F- Measure

Usually, there is an exclusive criterion that considers both precision and recall, and hence, you can aim to maximize this quantity to build your model better. This metric is specified as the F1-score, which clearly shows the harmonic mean of precision and recall. Unfortunately, it is not conceivable to maximize both these metrics simultaneously, because one is another cost. For some issues where both precision and recall are significant, one can pick a method that maximizes this F-1 score. In this work we obtained these results for both method:

ANN for the original dataset is about 0.718 while after changing the dataset is improves up to 0.7378. In KNN method for the original dataset is about 0.69 while after changing the dataset is down to 0.6105. So, as it can be seen in this work ANN model works well rather than the KNN model which given in Table 4.

Table 4. *The Result of Recall, Precision and F- Measure Metrics Analysis*

	Recall		Precision		F- Measure	
	ANN	K-NN	ANN	K-NN	ANN	K-NN
Original Dataset	0.925	0.556	0,587	0,92	0.718	0,69
New Dataset	0.95	0,460	0,6032	0,91	0.7378	0,6105

3.5. Analysis Using Train and Validation

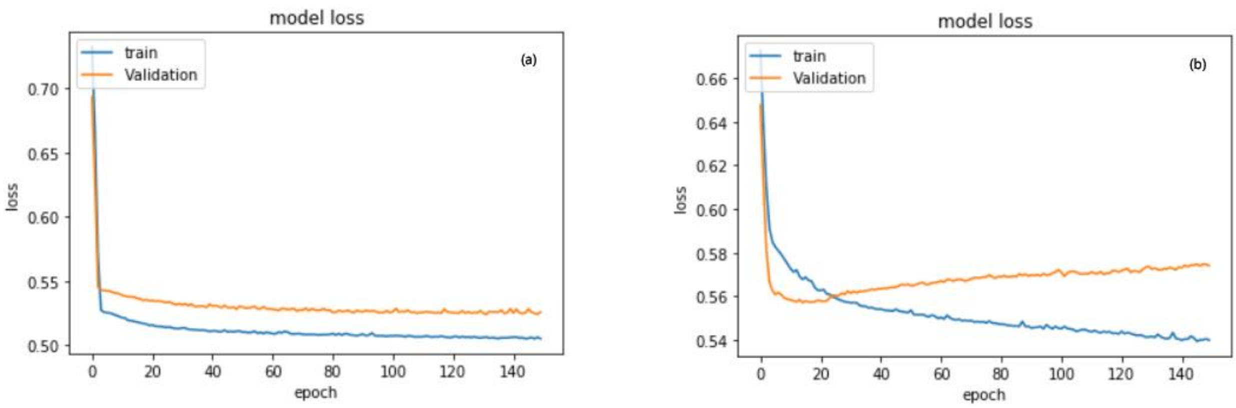

Figure 4. *Train and Validation Analysis for ANN Model; a) New Dataset b) Original Dataset*

In Figure 4. (b) we notice that the training loss and validation loss aren't correlated. This means the as the training loss is decreasing, the validation loss remains the same as increases over the iterations. This means that the model is not exactly improving, but is instead overfitting the training data. This isn't what we are looking for. Graph for the model (a) in this case, there is clearly a health correlation between training loss and the validation loss. They both seem to reduce and stay at a constant value. This means that the model is well trained and is equally good on the training data as well as the hidden data.

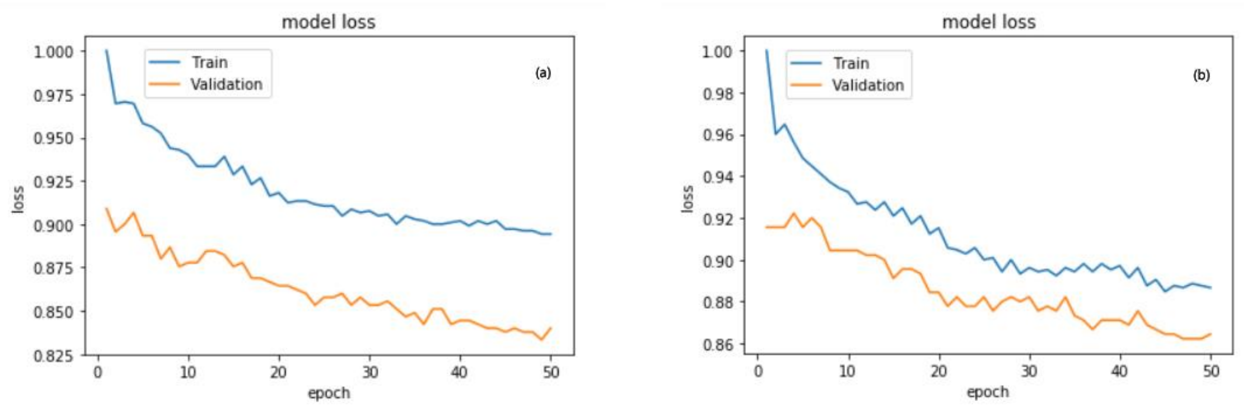

Figure 5. *Train and Validation Analysis for KNN Model; a) New Dataset b) Original Dataset*

In the Figure 5. for both (a) and (b) Here we find the validation loss is much better than the training one, which reflects the validation dataset is easier to predict than the training dataset. An explanation could be the validation data is scarce but widely represented by the training dataset, so the model performs extremely well on these few examples. Anyway, this means the validation dataset does not represent the training dataset, so there is a problem with representativeness.

Therefore, as can be seen from the results, the ANN model performs better in training and validation data.

4. RESULTS

Artificial intelligence methods such as ANN and fuzzy logic are very powerful tools that can be used by researcher to analyze, model and understand complex clinical data in various medical application areas. Further clinical trials are needed before these currently evolving methods can be implemented in real clinical settings. It is an inevitable fact that ANN will open a new era in the field of medicine and shed light on more advanced diagnosis and treatment methods, considering the gains and successes of studies in the field of health. It has been shown to be a useful tool in diagnosing complex diseases at risk of misdiagnosis. Therefore, it can be said that the ANN technique can reduce unnecessary research, unfavorable surgery rates and potential costs. In this study we employ an intuitionistic fuzzy apparatus applying rough set concept and tolerance - based on developing knowledge extraction. With this technic, we can select more efficient feature for modelling in machine learning. The artificial Neural Network and K- Nearest Neighbors' systems were utilized to classify the data. After fuzzifying data to normalize them and apply mentioned technique to select attributes then separated into train and test which were used to construct the artificial neural-network system and K- nearest neighbors. Finally, the dataset was classified in both dimensionality reduction dataset and original dataset. The experiential outcomes of the applied mechanism were measured for the breast cancer dataset. Hereafter, the efficiency of these two systems was examined using metrics such as precision, recall, F-measure, computation time, and accuracy. The preferable accuracy of 89% was obtained by the Artificial Neural Network method for the better classification of a data stream. ANN technique can productively distinguish malignant irregularities from benign ones and accurately predict the risk of breast cancer for singular abnormalities. In this study, we have evaluated whether attribute selection affect an artificial neural network (ANN) trained on a large prospectively collected dataset can discriminate between benign and malignant diseases and accurately predict the probability of breast cancer for individual patients. Also, the minimum computation time was 0.16058 that obtained by the ANN model for the new dataset. In this study, we tried to show how the feature selection and machine learning is the right decision to speed up the process and get more accurate results.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Agarwal, R., “https://towardsdatascience.com/the-5-feature-selection-algorithms-every-data-scientist-need -to-know-3a6b566efd2”. Access date: 27.07.2019.

[2] Jensen, R., Shen, Q., “Fuzzy-Rough Sets Assisted Attribute Selection”, IEEE Transactions on Fuzzy Systems, Feb; 15(1): 73, (2007).

[3] Jensen, R., Shen, Q., “Fuzzy-Rough Attribute Reduction with Application to Web Categorization”, Fuzzy Sets and Systems, 141(3): 469-485, (2004).

[4] Jensen, R., Shen, Q., “Computational intelligence and feature selection”, rough and fuzzy approaches, Hoboken, New Jersey, John Wiley & Sons, 8: (2008).

[5] Pawlak, Z., “Rough sets”, International Journal of Parallel Programming, 11(5): 341–356, (1982).

[6] Pawlak, Z., “Rough sets, theoretical aspects of reasoning about data”, Dordrecht, Netherlands: Springer Science, Business Media, 9: (2007).

[7] Pawlak, Z., Skowron, A., “Rough sets: Some extensions. Information Sciences”, 177(1): 28–40, (2012).

[8] Slowinski, R., Vanderpooten, D., "Similarity Relation as a Basis for Rough Approximations", Advances in Machine Intelligence and Soft Computing, P. Wang, ed., Duke University Press, (IV): 17-33, (1997).

[9] Han, J., Lee, J.G., Kamber, M., "An overview of clustering methods in geographic data analysis", Geographic Data Mining and Knowledge Discovery 2, 149-170, (2009).

[10] Tan, K.P., Liu, H.G., Liang, P.H., "Synthesis, docking studies, and evaluation of pyrimidines as inhibitors of SARS-CoV 3CL protease", Bioorganic and Medicinal Chemistry Letters, 20(12): 3569-3572, (2010).

[11] Dubois, D., Prade, H., "Putting Rough Sets and Fuzzy Sets Together", Intelligent Decision Support, 203-232, (1992).

[12] Dubois, D., Prade, H., Farreny, H., Martin, R., Clouaire, Testemale, C., “Possibility Theory an Approach to Computerized Processing of Uncertainty”, Plenum Press, New York, (1988).

[13] Dubois, D., Prade, H., “Twofold fuzzy sets and rough sets-Some issues in knowledge representation”, Fuzzy Sets and Systems, 23: p3-ndash, 18, (1987).

[14] Atanassov, K. T., “Intuitionistic fuzzy sets”, In Intuitionistic fuzzy sets, Physica, Heidelberg, 1-137, (1999).

[15] Bustince, H., Burillo, P., “Vague sets are intuitionistic fuzzy sets”, Fuzzy Sets and Systems, 79 (3): 403-405, (1996).

[16] Elena, E., Castiñeira, Torres-Blanc, C., Cubillo, S., “Measuring contradiction on A-IFS defined in finite universes”, Knowledge-Based Systems, 24 (8): 1297-1309, (2011).

[17] Deschrijver, G., Kerre, E.E., “On the position of intuitionistic fuzzy set theory in the framework of theories modelling imprecision”, Information Sciences, 177 (8): 1860-1866, (2017).

[18] Dymova, L., Sevastjanov, P., “An interpretation of intuitionistic fuzzy sets in terms of evidence theory: Decision making aspect”, Knowledge-Based Systems, 23 (8): 772-782, (2010).

[19] Mukherjee, S., Basu, K., “Solution of a class of Intuitionistic Fuzzy Assignment Problem by using similarity measures”, Knowledge-Based Systems, 27: 170-179, (2012).

[20] Feng, Q., Li, R., “Discernibility Matrix Based Attribute Reduction in Intuitionistic Fuzzy Decision Systems”, Paper presented at the RSFDGrC, (2013).

[21] Tiwari, A.K., Shreevastava, S., Som, T., Shukla, K.K., “Tolerance-based intuitionistic fuzzy-rough set approach for attribute reduction”, Expert Systems with Applications, 101(C July): 205–212, (2018).

[22] Skowron, A., Stepaniuk, J., "Tolerance Approximation Spaces", Fundamenta Informaticae, 27 (2): 245-253, (1996).

[23] Jurio, A., Paternain, D., Bustince, H., Guerra, C., Beliakov, G., “A construction method of Atanassov's intuitionistic fuzzy sets for image processing”, IEEE International Conference Intelligent Systems, London, 337-342, (2010).